ELG 5163

MACHINE VISION

Image Formation (Geometry)

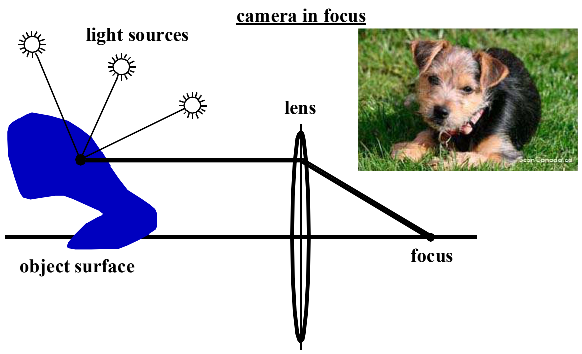
IMAGE FORMATION PARAMETERS

- As image formation is a result of light reflecting or diffusing on a surface and passing through the lens of a camera, it depends on many physical parameters:

- Geometric parameters
    - Position of surfaces
    - Orientation of surfaces
- Photometric parameters
    - Lighting (type, intensity, origin, direction)
    - Reflectance characteristics of surfaces (lambertian, specular)
    - Sensor sensitivity to light and its response to different wavelengths
- Optical parameters
    - Lens characteristics
    - Focal length of the camera
    - Field of view

IMAGE FOCUSING

- In general, several light sources exist in the environment. As such, every point on a surface receives photons (light energy) from several sources.
- In order to get sharp images, it is necessary that the light reflected or diffused by one point on a surface is directed exactly to the same point on the image plane.
- When this is achieved, the camera is in focus.
- If not, one point on the surface will be projected on different locations on the image plane, forming a diffused circular area (blurry patch of light).



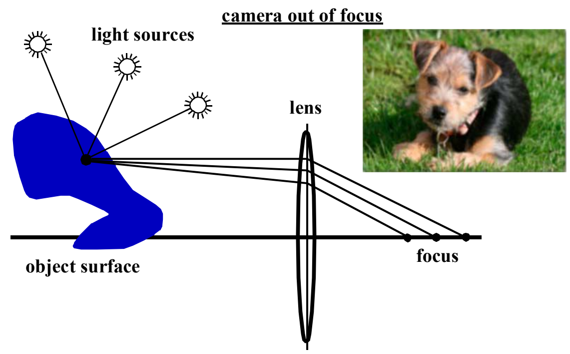


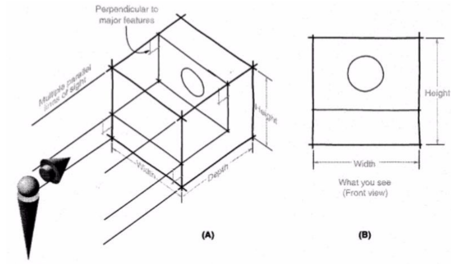
IMAGE FOCUSING

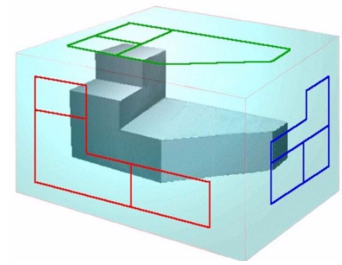
- In practice, camera are equipped with an optical system made of lenses and apertures that make all rays coming from a single surface point to converge exactly to the same point on the image plane, when properly adjusted.

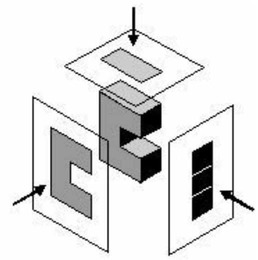
- In theory, we need to develop a simpler model of the sensing device that satisfies the constraint of focusing.
  - The classical model of a camera is based on the use of a virtual pinhole.
  - The pinhole is a very small aperture that theoretically allows only one ray coming from a given point to reach the image plane.
  - This creates a one-to-one correspondence between points on the object surface and image points.

PROJECTIONS

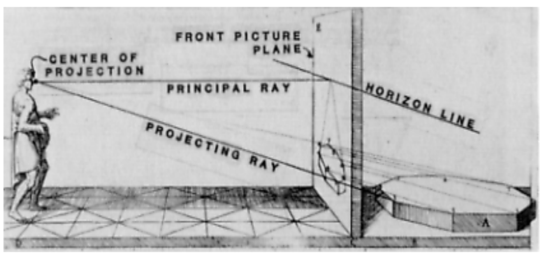
- Orthographic projection
  - Projection at full scale on image planes, independently from depth.

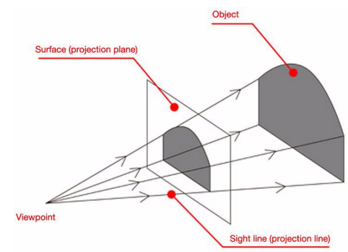






- Perspective projection
  - Projection takes depth into consideration.

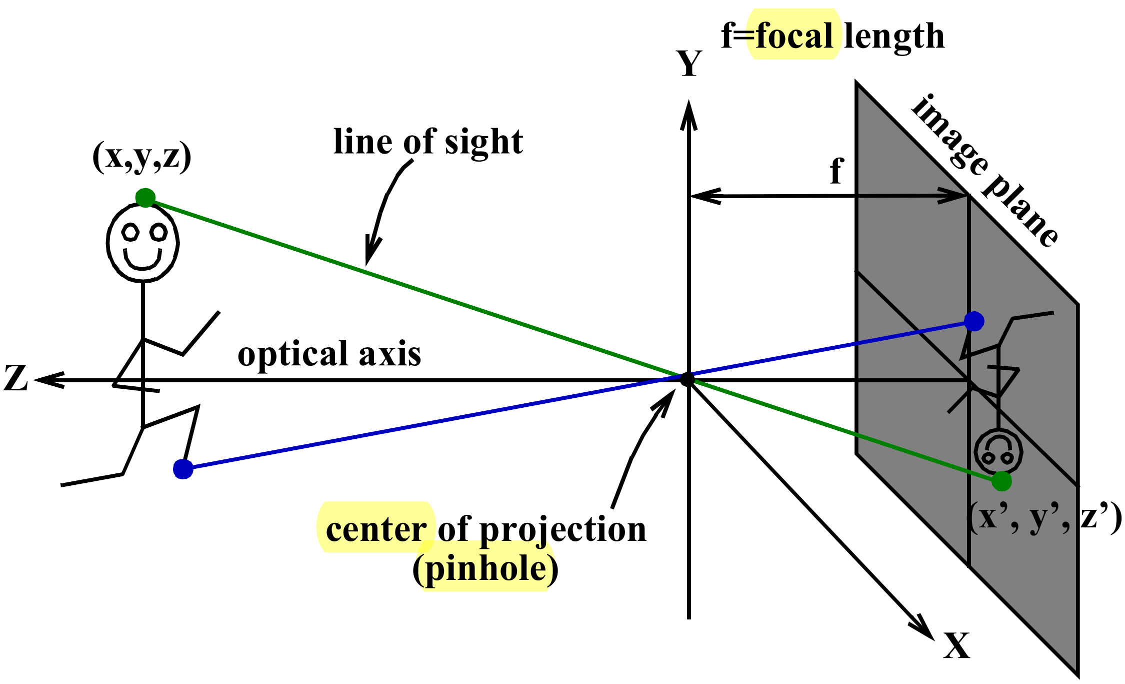




ELG 5163 - MACHINE VISION

PINHOLE CAMERA MODEL

- Inverting model



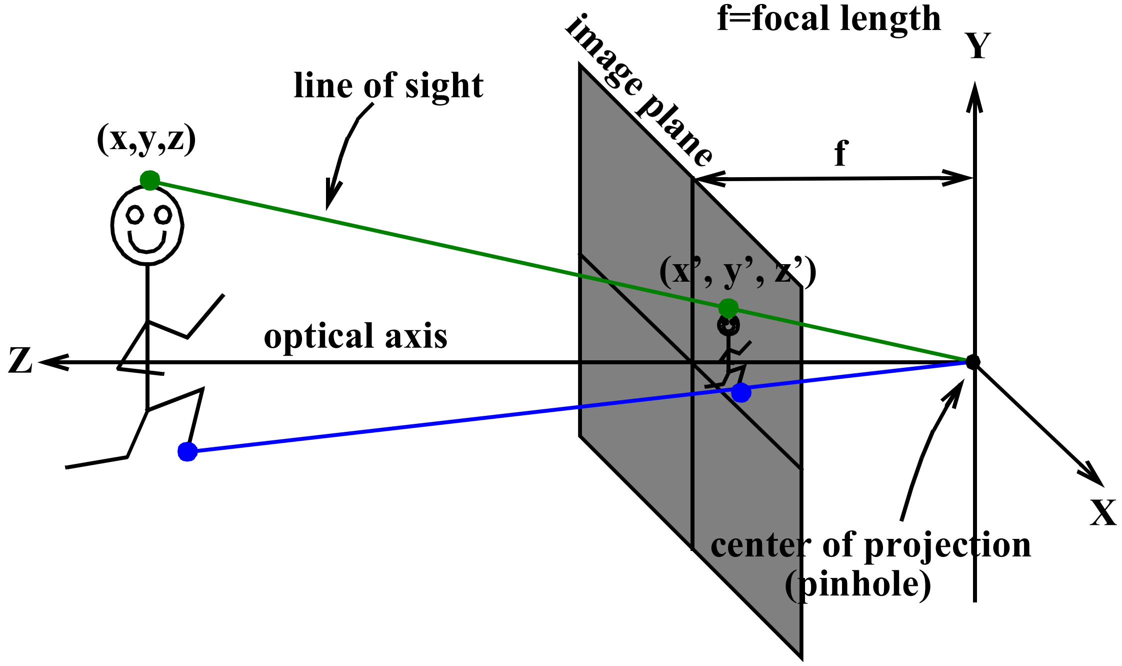
UNIVERSITY OF OTTAWA – COPYRIGHTED MATERIAL

The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to reproduce, distribute, modify, or upload the course materials to any website or media.

© P. Payeur, 2021

PINHOLE CAMERA MODEL

- Non-inverting model



ELG 5163 - MACHINE VISION

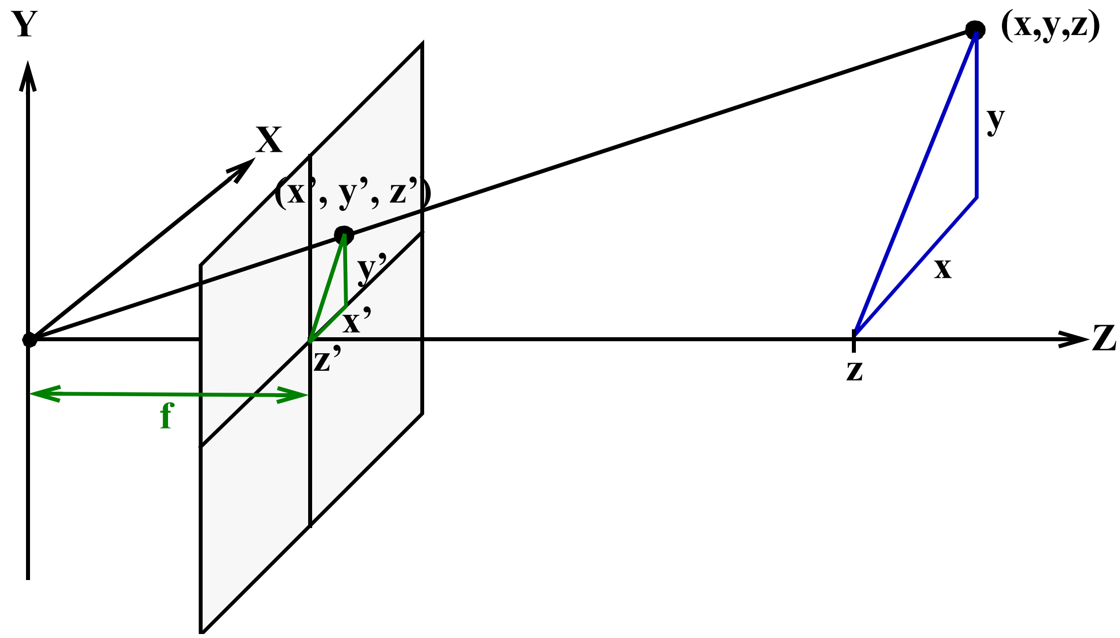
WEAKNESSES OF THE PINHOLE MODEL

- Small aperture:
  - Very limited light can actually reach the image plane.
  - Images would to be dark with very few grey levels.
- If the pinhole is enlarged:
  - Too many light rays enter the camera and the focus is lost.
  - Surface points are projected as blurry areas on the image plane.
- Solution:
  - Real camera use lenses to make all light rays converge.
  - No pinhole is actually used.
  - The model is however very useful in vision systems design, and is worth developing further.

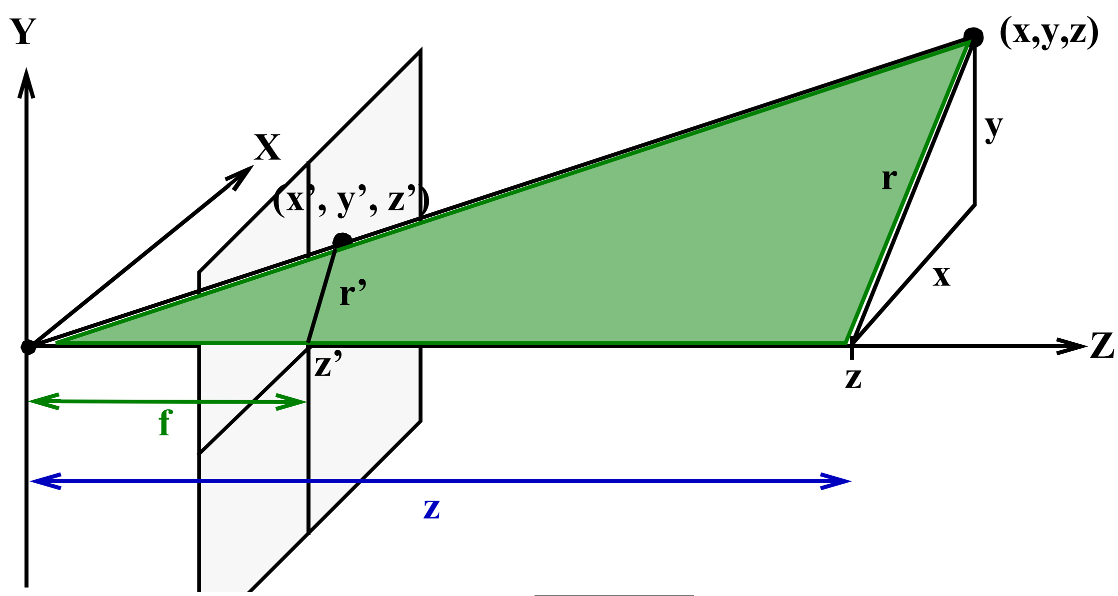
UNIVERSITY OF OTTAWA – COPYRIGHTED MATERIAL
The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to reproduce, distribute, modify, or upload the course materials to any website or media.
© P. Payeur, 2021

PERSPECTIVE PROJECTION

- Using the pinhole camera model, simple equations can be derived to represent the projection of an object surface point on the image plane.

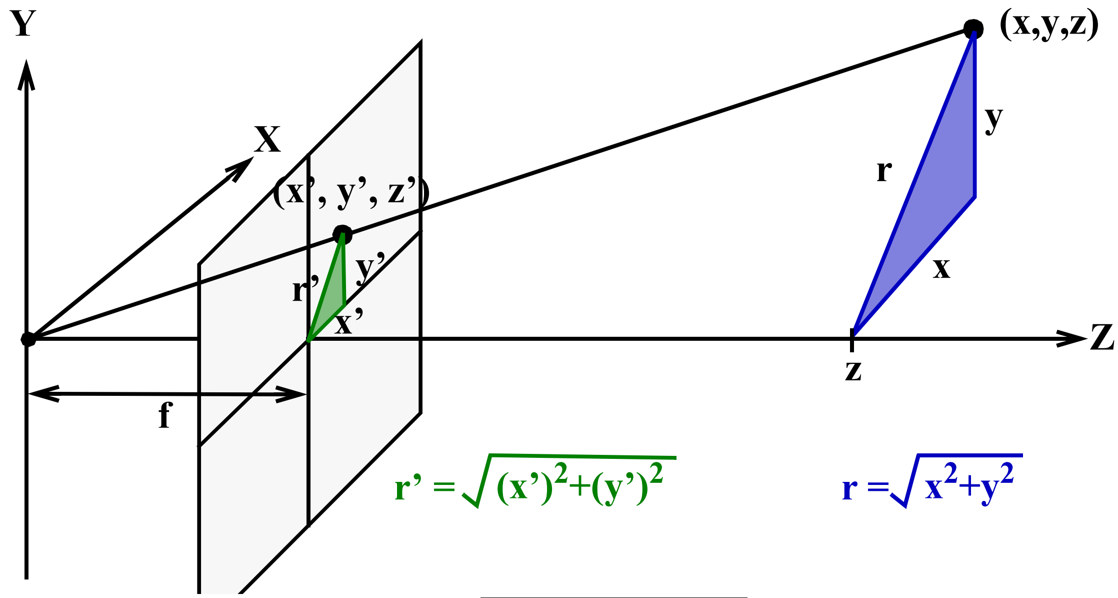


PERSPECTIVE PROJECTION



$$\boxed{\frac{f}{z} = \frac{r'}{r}}$$

PERSPECTIVE PROJECTION



$$r' = \sqrt{(x')^2 + (y')^2} \qquad r = \sqrt{x^2 + y^2}$$

$$\boxed{\frac{x'}{x} = \frac{y'}{y} = \frac{r'}{r}}$$

PERSPECTIVE PROJECTION

- **Combining our equations from similar triangles:**

$$\frac{x'}{x} = \frac{y'}{y} = \frac{r'}{r} = \frac{f}{z}$$

- **we get the perspective projection equations:**

$$\boxed{x' = \frac{f}{z}x}$$

$$\boxed{y' = \frac{f}{z}y}$$

- **and on the image plane, the z' coordinate is always (for this model):**

$$\boxed{z' = f} \text{ (known by construction)}$$

- **Image point forms a nonlinear relationship with the coordinates (x, y) of the object surface point (it depends both on f and z).**

PERSPECTIVE PROJECTION

- **The perspective projection equations can be rewritten as follows:**

$$\frac{x}{x'} = \frac{z}{f} = \frac{y}{y'}$$

- **Placing (x, y, z) in evidence, we get the inverse perspective projection equations:**

$$\boxed{x = \frac{z}{f}x'}$$

$$\boxed{y = \frac{z}{f}y'}$$

$$\boxed{z = \frac{f}{y'}y = \frac{f}{x'}x} \text{ (!?!)}$$

- **The actual distance to the object, z, can be estimated only up to a scale factor (x/x' or y/y', that is real object size / object projection size), unless the object's actual size (x or y) is known.**
- **Reconstructing 3D from 2D is challenging.**

WEAK PERSPECTIVE MODEL

- **In order to linearize the relationship between the coordinates of the image point and the object surface point, an assumption is made that the distance between two points in the scene is much smaller than the average distance, $\bar{z}$, of all points from the object to the camera.**
- **This involves that the shape of the object is relatively smooth and far away from the camera.**

WEAK PERSPECTIVE MODEL

- In this case, we approximate the perpective projection by the following:

$$\boxed{x' \approx \frac{f}{\bar{z}} x}$$

$$\boxed{y' \approx \frac{f}{\bar{z}} y}$$

- As $\bar{z}$ is now a constant, that averages distance to all points on the subject, we obtain an approximate but linear relationship between the image point coordinates (x', y') and the point on the actual scene (x, y). The relationship only depends on the focal length.

ELG 5163 - MACHINE VISION

PERSPECTIVE PROJECTION AS AN HOMOGENEOUS TRANSFORMATION

- If we consider the coordinates of the image point, (x', y', z'), as an homogeneous coordinate vector, $[\tilde{x}', \tilde{y}', \tilde{z}', \tilde{w}']^T$.

- The coordinates of the object surface point, (x, y, z), can also be written as an homogeneous coordinate vector, $[\tilde{x}, \tilde{y}, \tilde{z}, \tilde{w}]^T$, where the weight, $\tilde{w}$, is equal to 1 as this point is in real 3D space (full sccale).
- We can represent the perspective projection as follows:

$$\boxed{\begin{bmatrix} \tilde{x}' \\ \tilde{y}' \\ \tilde{z}' \\ \tilde{w}' \end{bmatrix} = P \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{f} & 0 \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \\ 1 \end{bmatrix}}$$

- This matrix product leads to:

$$\begin{bmatrix} \tilde{x}' \\ \tilde{y}' \\ \tilde{z}' \\ \tilde{w}' \end{bmatrix} = \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \\ \tilde{z}/f \end{bmatrix}$$

- We note that the weight of the projected point, $\tilde{w}'$, is not equal to 1 as this is not a coordinate point in the real 3D space but rather on the image plane (scaled down).
- The projection operation introduces some distortion in the world representation (projecting 3D onto 2D) that appears as a scaling factor.
- Ultimately, we retrieve the perspective projection equations:

$$\tilde{x}' = \frac{\tilde{x}}{(\tilde{z}/f)} = \frac{f\tilde{x}}{\tilde{z}}$$

$$\tilde{y}' = \frac{\tilde{y}}{(\tilde{z}/f)} = \frac{f\tilde{y}}{\tilde{z}}$$

$$\tilde{z}' = \frac{\tilde{z}}{(\tilde{z}/f)} = \frac{f\tilde{z}}{\tilde{z}} = f$$

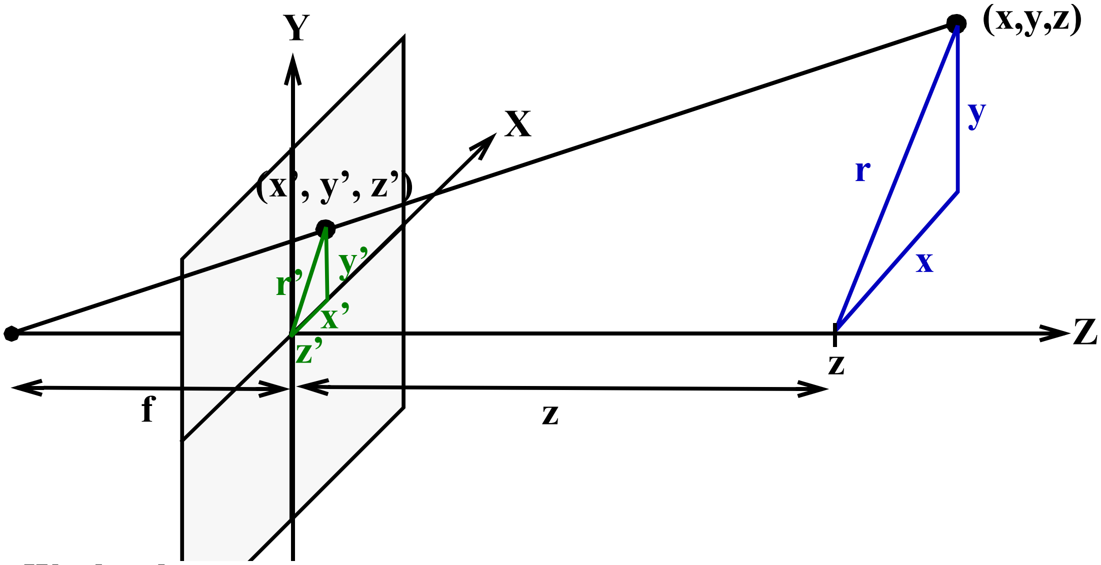
ELG 5163 - MACHINE VISION

- Therefore, the perspective projection operation can be defined as a 4x4 matrix *P*:

$$\boxed{P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{f} & 0 \end{bmatrix}}$$

- Note that the projection matrix has a different structure than that of homogeneous geometrical transformations (where $S \neq \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$ and $I \neq 1$).
- In that, perspective projection is not an homogeneous transformation, as it does not preserve dimensions.
- The *P* matrix cannot be inverted as its determinant equals 0.
- The definition of the projection matrix, *P*, also depends on where the origin of the reference frame (0,0,0) is located.

- For example, **a different perspective projection model** can be defined where the origin is located on the image plane rather than being on the center of projection:



- We then have:

$$\frac{x'}{x} = \frac{y'}{y} = \frac{r'}{r} = \frac{f}{f+z}$$

- The **perspective projection equations** are then slightly different:

$$\boxed{x' = \frac{f}{f+z}x}$$

$$\boxed{y' = \frac{f}{f+z}y}$$

- and on the image plane, the z' coordinate is always:

$$\boxed{z' = 0}$$

- The corresponding projection matrix is then:

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{f} & 1 \end{bmatrix}$$

- This projection matrix has the advantage of being **invertible**.

- But this projection matrix does not fully satisfies the perspective projection equations. Let's develop the equations in homogeneous coordinates:

$$\begin{bmatrix} \tilde{x}' \\ \tilde{y}' \\ \tilde{z}' \\ \tilde{w}' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{f} & 1 \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \\ 1 \end{bmatrix}$$

- which corresponds to:

$$\tilde{x}' = \tilde{x}$$

$$\tilde{y}' = \tilde{y}$$

$$\tilde{z}' = \tilde{z}$$

$$\tilde{w}' = \frac{\tilde{z}}{f} + 1 = \frac{\tilde{z} + f}{f}$$

- Extracting the Cartesian coordinates from the homogeneous representation for the image point leads to:

$$x' = \frac{fx}{f+z}$$

$$y' = \frac{fy}{f+z}$$

$$??????\ z' = \frac{fz}{f+z}\ ?????$$

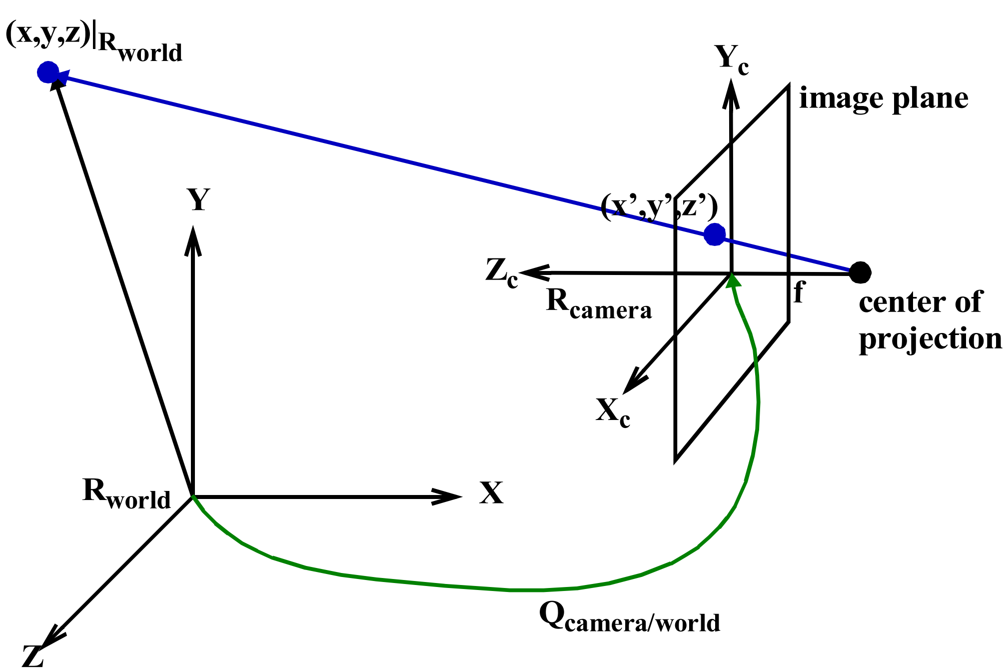
- We see that, with this position of the origin for the projection representation, the Z component of the projection operation does not have the proper physical meaning (z' can only be 0 as the origin of R_{camera} is on the image plane with this model).
- CONCLUSION:
  - The perspective projection operation in terms of homogeneous coordinates can be useful as long as it is used with extra care. One must always make sure to use the proper expression for *P*, to validate each result and to keep in mind the intrinsic limitations of this method at representing the reality.

PERSPECTIVE PROJECTION WITH MULTIPLE REFERENCE FRAMES

- **In the previous examples, the coordinates of the image point and of the object surface point were defined with respect to a same reference frame attached to the camera.**
- **In practice, such a simplification is hardly feasible because the center of projection and the image plane are not accessible on a real sensor. Often, we don't even know precisely where they are inside of the camera. We then need perform some calibration to find out where are the center of projection and the image plane with respect to the casing of the camera.**
- **Moreover, objects in the world are never defined with respect to a sensor.**
- **Assuming a calibrated camera, we need to consider two reference frames:**
  - **One imaging reference frame with respect to which image points are defined (camera centered frame).**
  - **One world reference frame with respect to which everything (objects and the camera reference frame itself) is defined.**

ELG 5163 - MACHINE VISION

- **Assuming that we know where the camera is with respect to the world reference frame. These two frames and the relationship between them is illustrated by means of a transformation graph.**



- To compute the coordinates of the image point (x', y', z') on the image plane, one must first find out the coordinates of this point with respect to the camera reference frame $(x_c, y_c, z_c)\big|_{R_{camera}}$ and then perform the perspective projection operation.

$$\boxed{\tilde{v}' = P\tilde{v}_{\text{camera}} = PQ^{-1}_{\text{camera/world}}\tilde{v}_{\text{world}}}$$

- where:

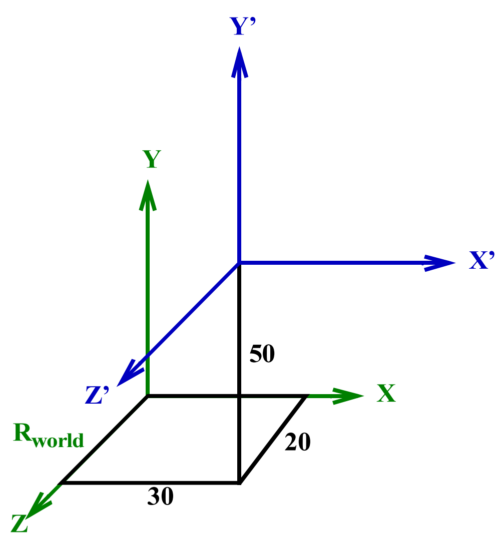
$\tilde{v}' = (\tilde{x}', \tilde{y}', \tilde{z}', \tilde{w}')$ the image point

$\tilde{v}_{camera} = (\tilde{x}_c, \tilde{y}_c, \tilde{z}_c, 1)$ the object point w/r to R_{camera}

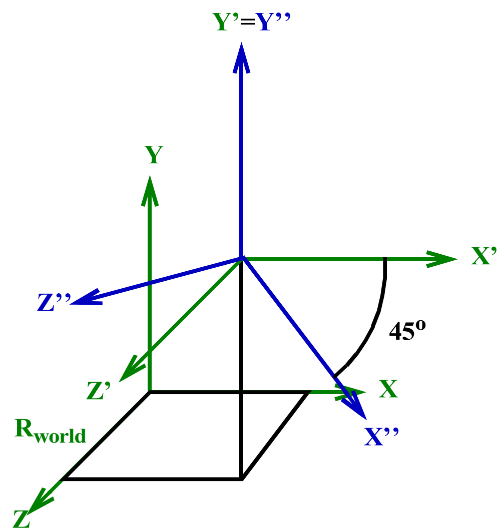
$\tilde{v}_{world} = (\tilde{x}, \tilde{y}, \tilde{z}, 1)$ the object point w/r to R_{world}

AN EXAMPLE

- The position and orientation of the camera reference frame is defined by the following set of elementary transformations (with respect to an evolving reference frame) starting from R_{world}, and obtained through calibration.
  - 1) A translation of (30, 50, 20) along (X, Y, Z)

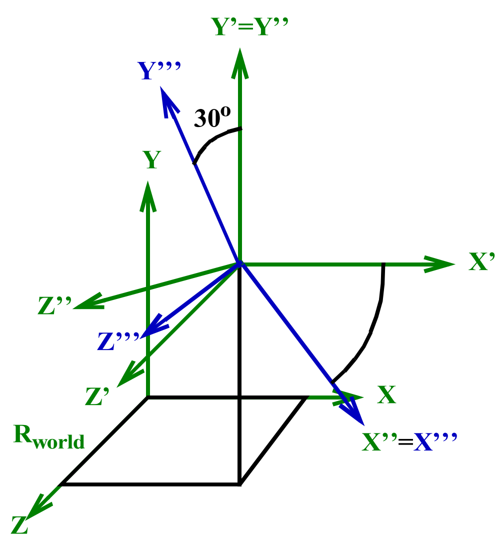


- 2) A rotation of -45^{o} around Y'



ELG 5163 - MACHINE VISION

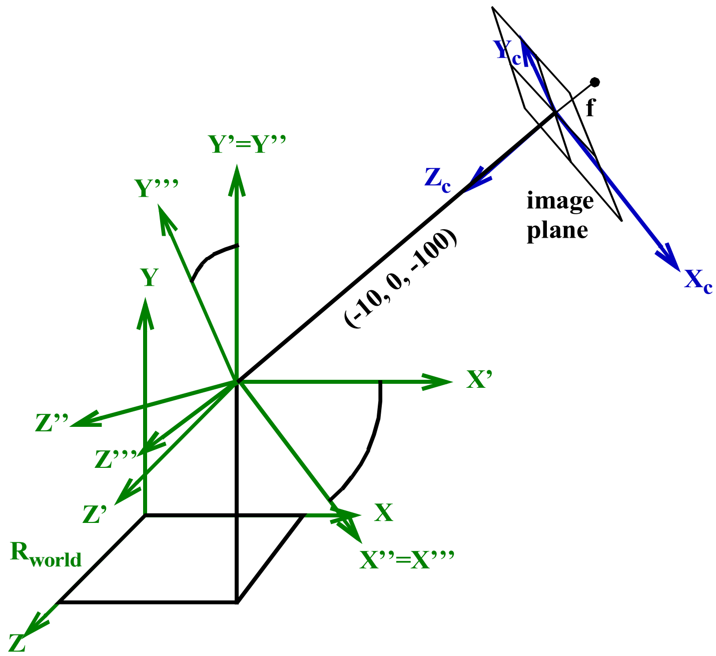
- 3) A rotation of 30^{o} around X''.



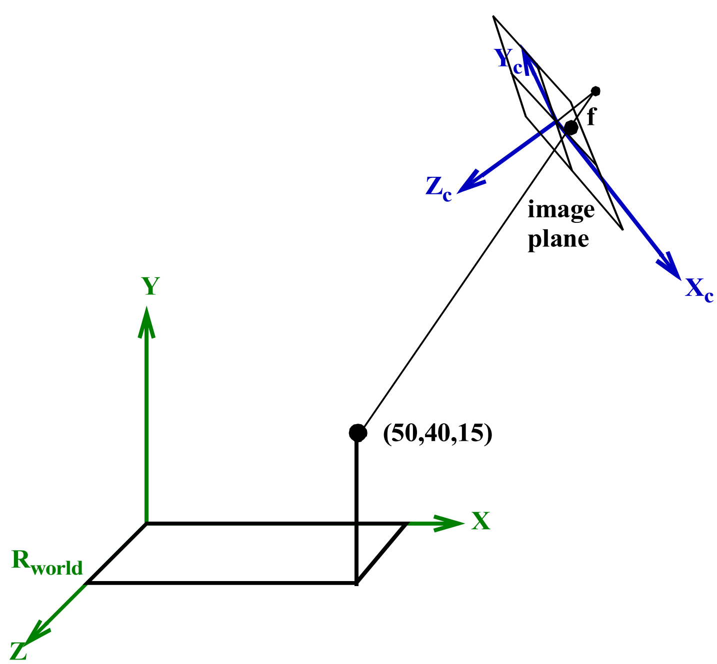
UNIVERSITY OF OTTAWA – COPYRIGHTED MATERIAL
The materials you receive for this course are protected by copyright and to be used for this course only. You do not have permission to reproduce, distribute, modify, or upload the course materials to any website or media.

© P. Payeur, 2021

- 4) A translation of (-10, 0, -100) along (X''', Y''', Z''').



- A point (50, 40, 15) defined with respect to R_{world} will be projected on the image plane of the camera as follows:



- **The geometrical transformation between R_{world} and R_{camera} is given by:**

$$Q_{\text{camera/world}} = Q_T(30, 50, 20)Q_{R_y}(-45°)Q_{R_x}(30°)Q_T(-10, 0, -100)$$

- **Replacing with the definitions of elementary transformation matrices, we have:**

$$Q_{\text{camera/world}} = \begin{bmatrix} 1 & 0 & 0 & 30 \\ 0 & 1 & 0 & 50 \\ 0 & 0 & 1 & 20 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.707 & 0 & -0.707 & 0 \\ 0 & 1 & 0 & 0 \\ 0.707 & 0 & 0.707 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.866 & -0.5 & 0 \\ 0 & 0.5 & 0.866 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -10 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -100 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$Q_{\text{camera/world}} = \begin{bmatrix} 0.707 & -0.354 & -0.612 & 84.2 \\ 0 & 0.866 & -0.5 & 100 \\ 0.707 & 0.354 & 0.612 & -48.3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- **This means that the coordinates of the point defined with respect to the camera coordinates are:**

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = Q^{-1}_{\text{camera/world}} \begin{bmatrix} 50 \\ 40 \\ 15 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.707 & 0 & 0.707 & -25.4 \\ -0.354 & 0.866 & 0.354 & -39.8 \\ -0.612 & -0.5 & 0.612 & 131.1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 50 \\ 40 \\ 15 \\ 1 \end{bmatrix} = \begin{bmatrix} 20.6 \\ -17.5 \\ 89.7 \\ 1 \end{bmatrix}$$

- **Applying the perspective projection operation to this point known in the camera reference frame if the focal length is f=10:**

$$\begin{bmatrix} \tilde{x}' \\ \tilde{y}' \\ \tilde{z}' \\ \tilde{w}' \end{bmatrix} = P \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0.1 & 1 \end{bmatrix} \begin{bmatrix} 20.6 \\ -17.5 \\ 89.7 \\ 1 \end{bmatrix} = \begin{bmatrix} 20.6 \\ -17.5 \\ 89.7 \\ 9.97 \end{bmatrix}$$

- **Finally, the coordinates of the projected point on the image plane are:**

$$\boxed{\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 2.07 \\ -1.76 \\ \mathbf{0} \end{bmatrix}}$$