CS4243
Computer Vision
&
Pattern Recognition

Camera Projection Models

- Our interest is not in explaining human visual behaviour.   Our interest is in learning the geometry of image formation in cameras.

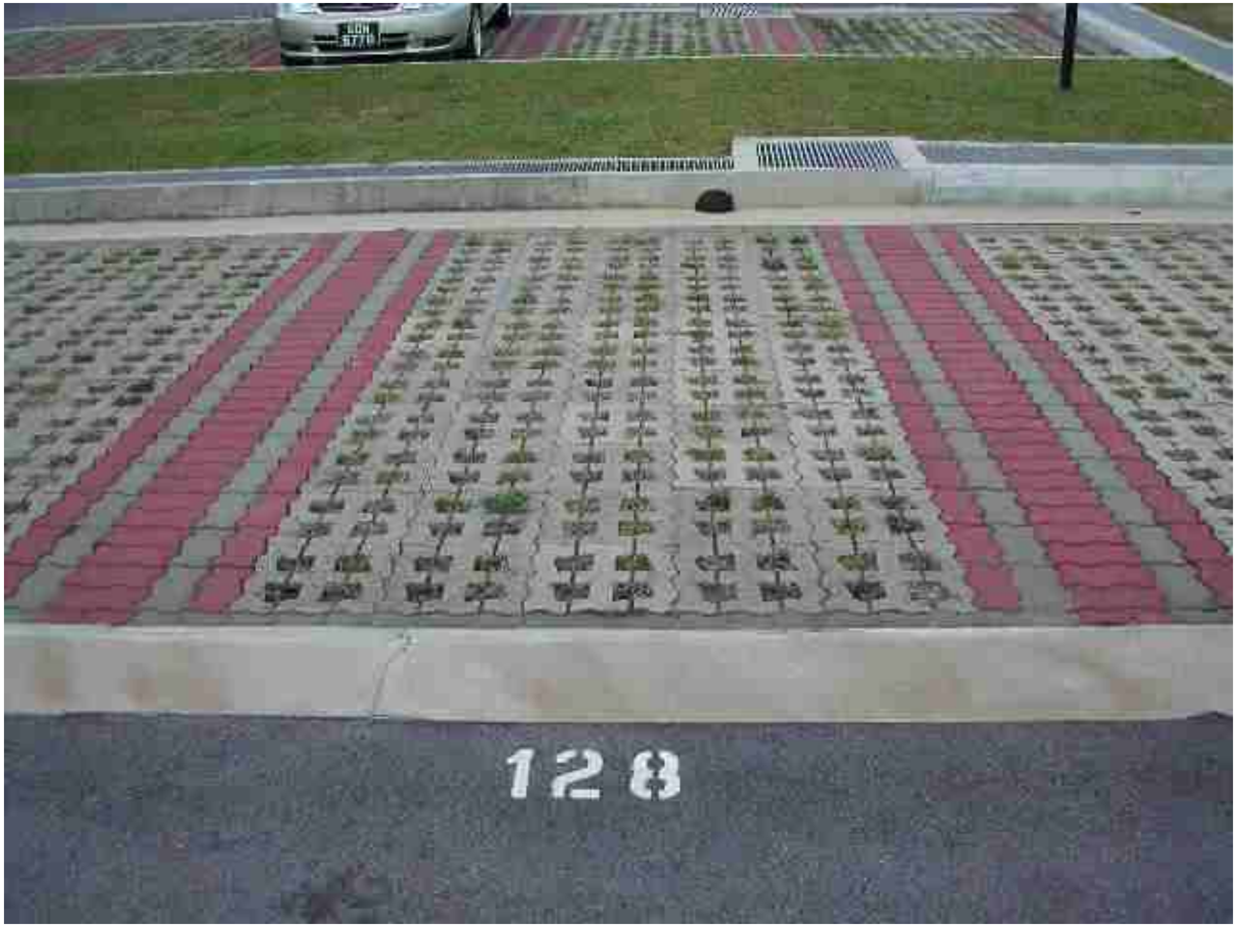
- Understand how images are formed:

    - 3D world projected onto 2D image
    - Distortion, pixellation, quantization

- Camera Projection Models:

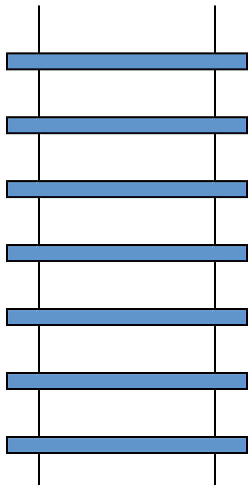
    - Mathematical description of the 3D to 2D projection process.

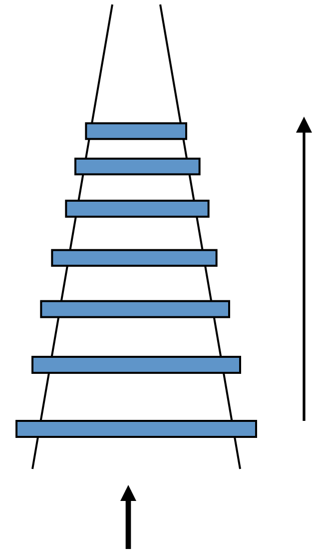

AHEAD
STRIPS


128

Right angles in 3D do not appear as right angles in 2D image

Distortion: Foreshortening





Gaps become smaller and smaller

view direction

Railway track viewed vertically from top

Railway track viewed from ground level.

What else can be said about a single image ?

- interposition -> occlusion
- perspective scaling →closer (bigger), further (smaller)
- motion parallax → closer (move faster),
  further (move slower)
- infer orientation of known geometrical objects

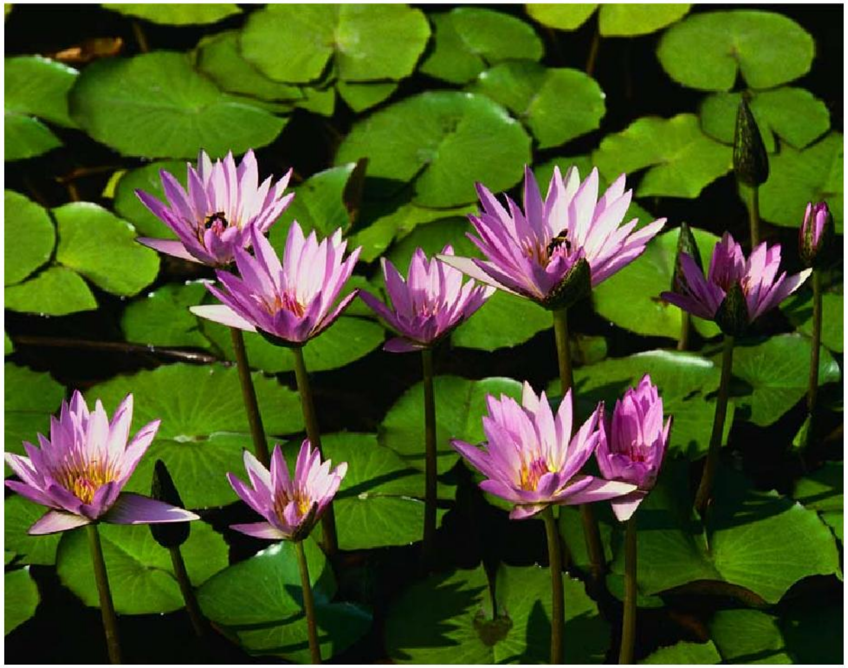
Pixellation – Quantization of Spatial Information

- A 1k by 1k camera sensor plane partitions a scene into 1k by 1k squares, each squared being recorded by a pixel

- The larger the number of pixels, the clearer is the picture   i.e. image spatial resolution is higher

- Lower resolution images can be formed from higher resolution images.   Successive reduction of resolution forms an image pyramid

High Resolution Image
600 x 800 pixels



Low Resolution Image
75 x 100 pixels

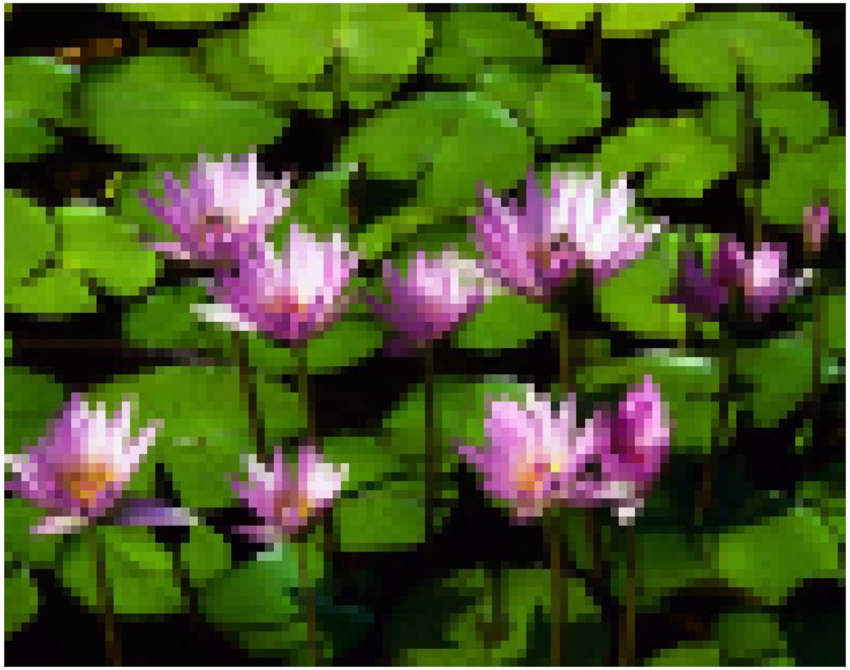
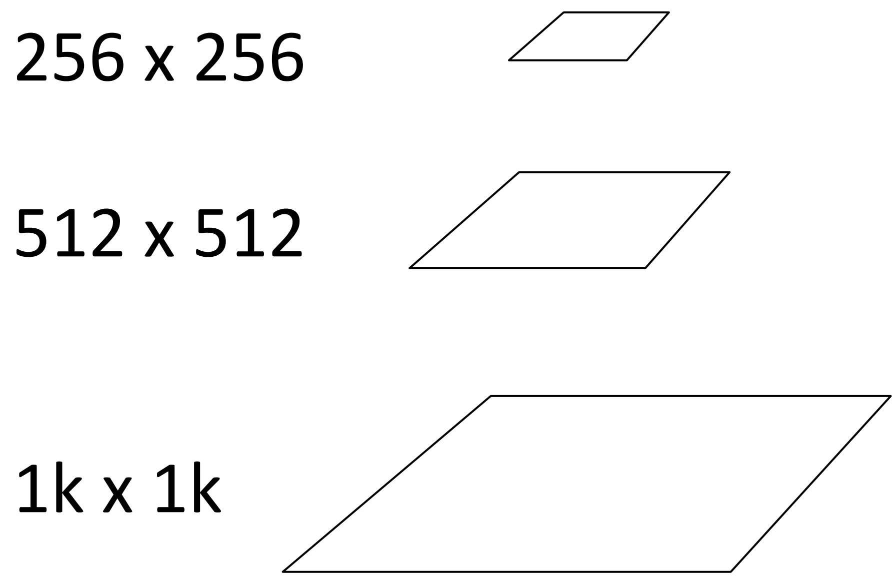


Image Pyramid

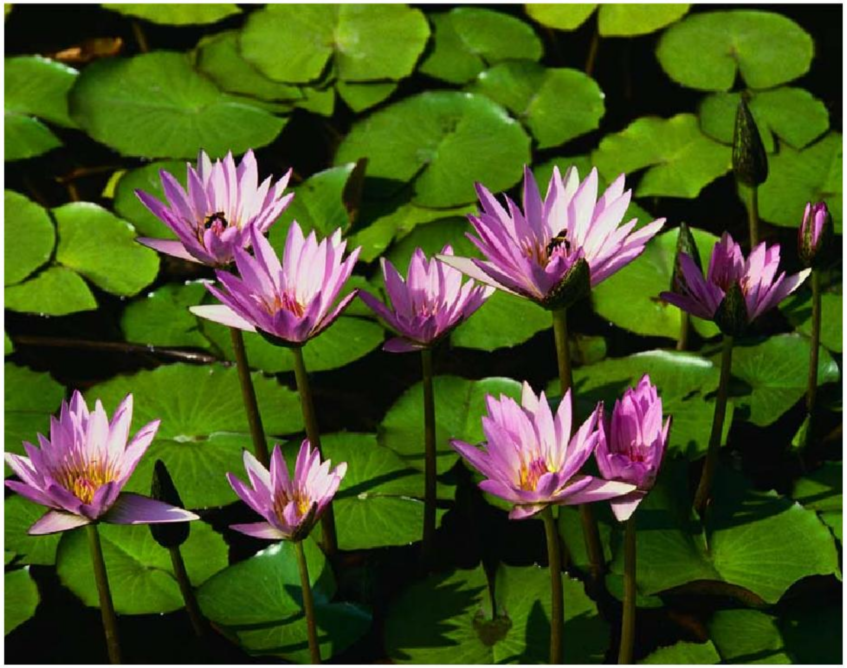


Very important concept for feature tracking (to be covered in future lectures)

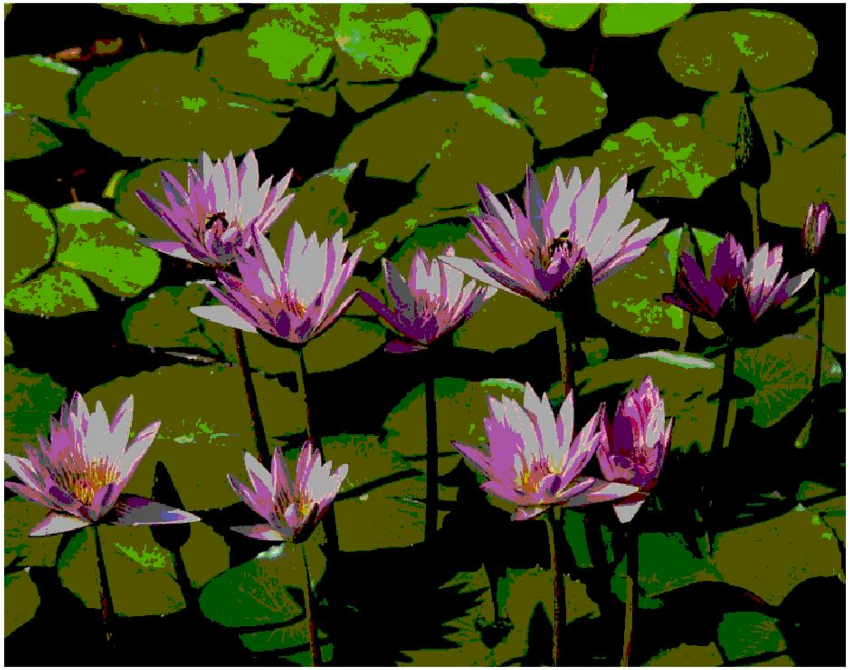
Quantization of intensities

color (or gray scale) is quantized

Example: each pixel is represented by 8 bits, or 256 discrete intensity levels



Each color component (i.e. r,g,b) is quantized to 256 levels



Each color component (i.e. r,g,b) is quantized to 25 levels

Camera Projection Models

- 3D scenes project to 2D images

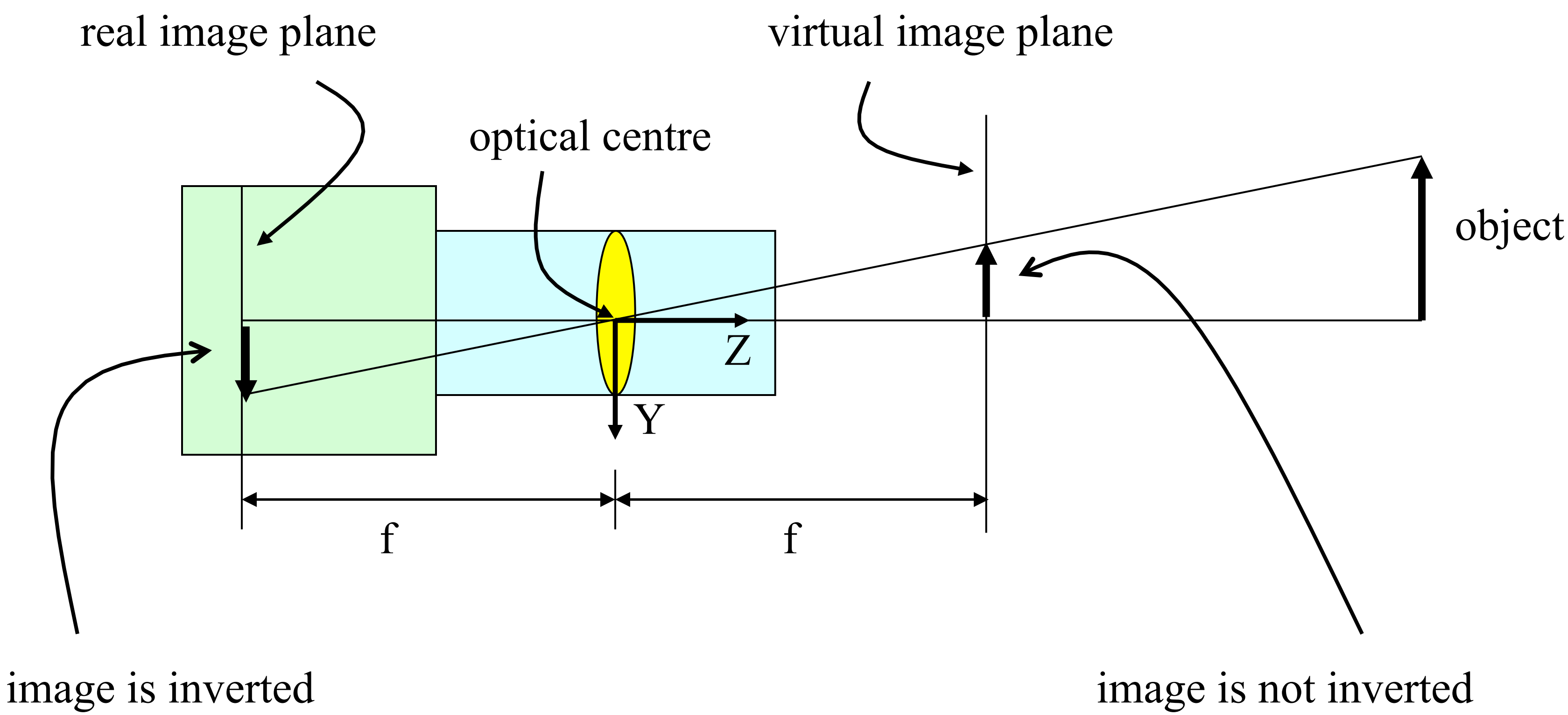
  Common projection models are:

  - orthographic
  - weak-perspective (scaled orthographic)
  - para-perspective
  - perspective (pinhole)

least accurate

most accurate

Where is the Image Formed in a Camera ?



Real Image Plane vs Virtual Image Plane

The image is formed on the real image plane physically. The image is vertically and laterally inverted.

We can imagine a virtual image plane at a distance of f in front of the camera optical center, where f is the focal length. The image on this virtual image plane is not inverted but geometrically correct. It is more intuitive to think of images on this virtual image plane. In the sequel, we will consider images on virtual image plane.

Pinhole Camera Model (Perspective Projection Model)

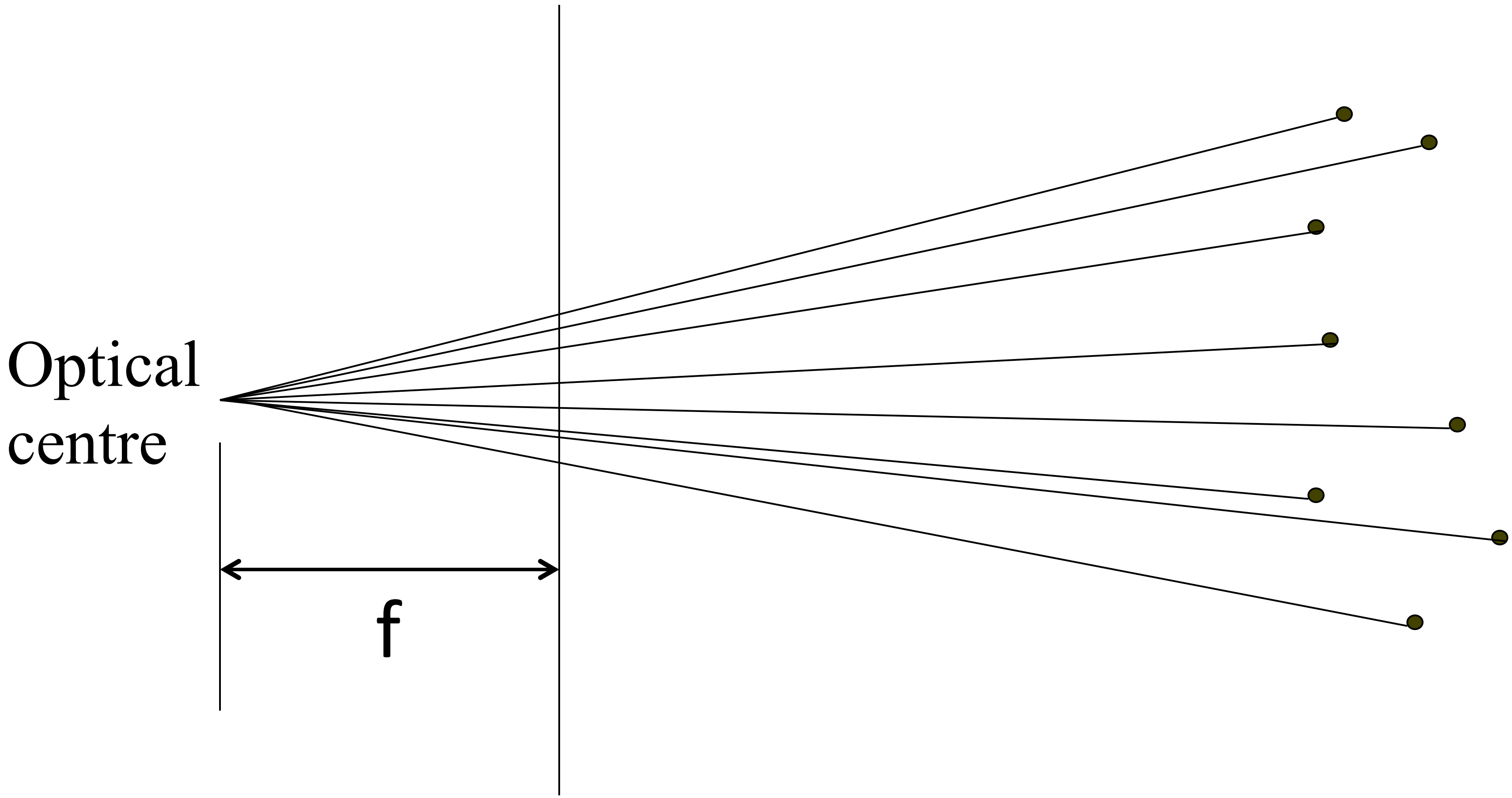
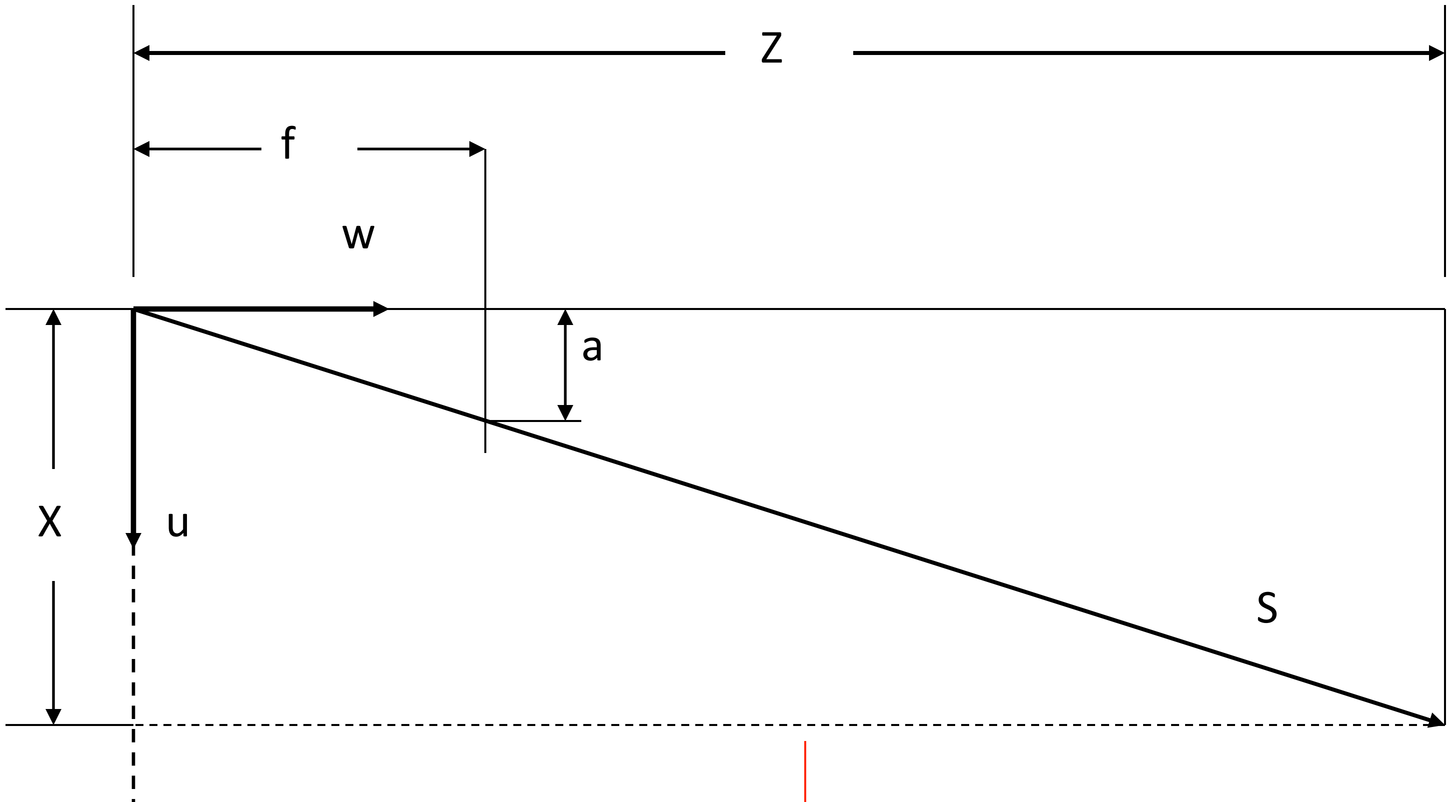


Image plane (virtual)

Pinhole Camera Model (Perspective Projection Model)

- it is usually more convenient to use the "front" image plane instead of "real" image plane
- The imaging process is a many-to-one mapping i.e. all points on the 3D ray map to a single image point.

  Therefore, depth information is lost
- by the use of similar triangles, can write down the perspective equation

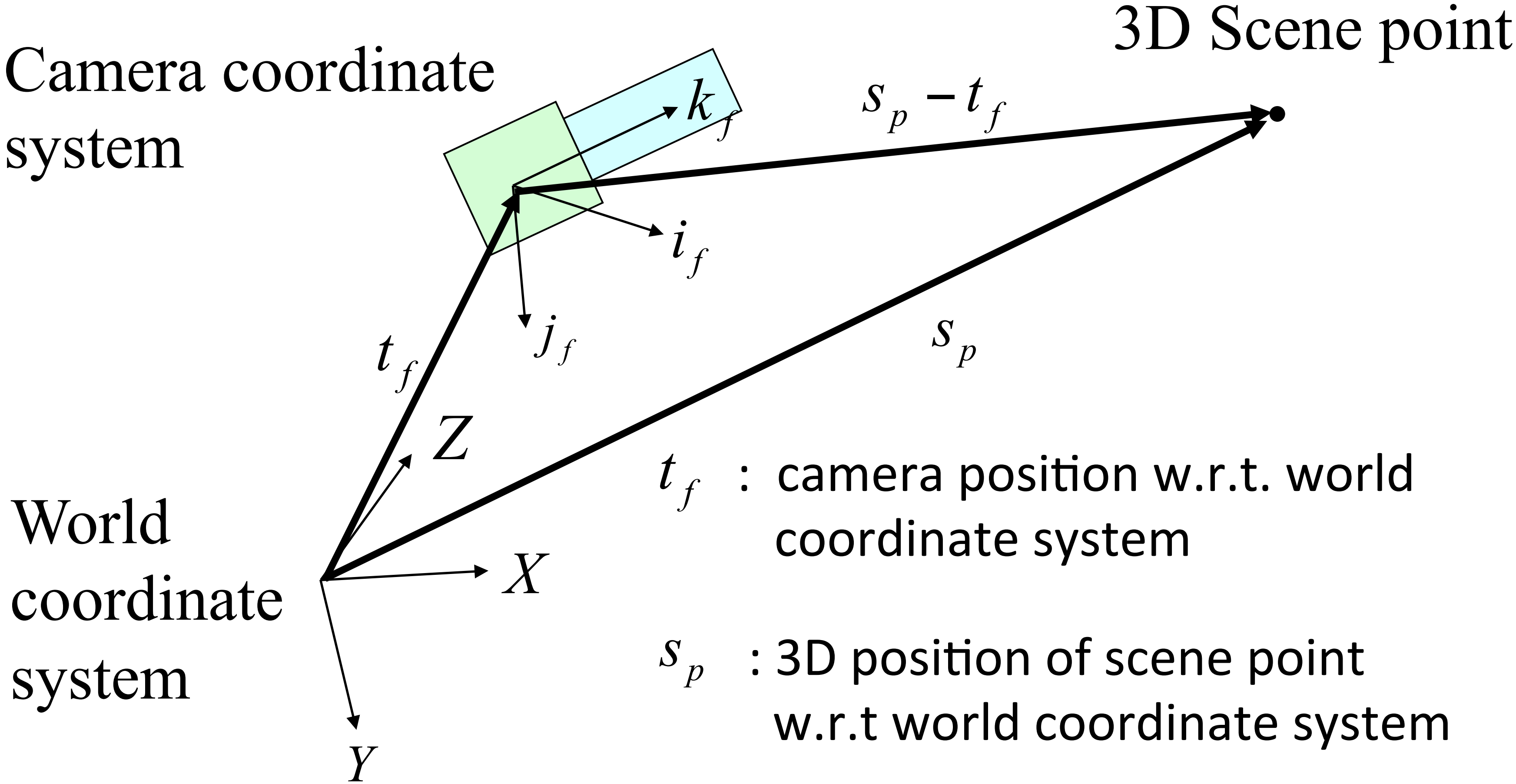


If u and w are unit vectors,

$$S^T u = X$$
$$S^T w = Z$$

Using similar triangles,

$$a = \frac{f\, S^T u}{S^T w}$$



t_f : camera position w.r.t. world coordinate system

s_p : 3D position of scene point w.r.t world coordinate system

$s_p - t_f$: 3D position of scene point w.r.t. camera coordinate system

Perspective Projection Equations

$$u_{fp} = \frac{f(s_p - t_f)^T i_f}{(s_p - t_f)^T k_f} \beta_u + u_0$$

$$v_{fp} = \frac{f(s_p - t_f)^T j_f}{(s_p - t_f)^T k_f} \beta_v + v_0$$

$$s_p = \begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix}$$

3D scene point

$$t_f = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

Camera translation

$$i_f = \begin{bmatrix} i_x \\ i_y \\ i_z \end{bmatrix}$$

Camera horizontal axis

$$j_f = \begin{bmatrix} j_x \\ j_y \\ j_z \end{bmatrix}$$

Camera vertical axis

$$k_f = \begin{bmatrix} k_x \\ k_y \\ k_z \end{bmatrix}$$

Camera optical axis

Camera intrinsic parameters

f focal length

u_0 Image center horizontal offset

v_0 Image center vertical offset

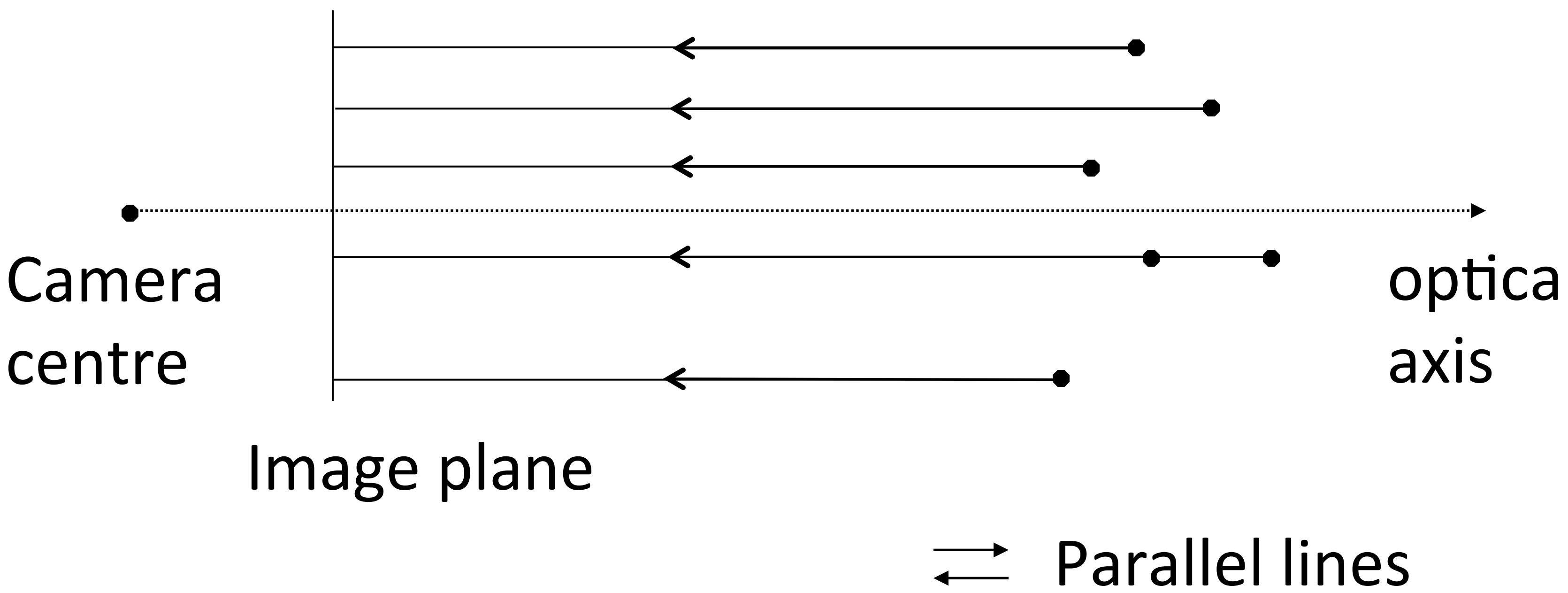
β_u pixel scaling factor in horizontal direction

β_v pixel scaling factor in vertical direction

Camera intrinsic parameters can be obtained through calibration

Orthographic Projection Model

Projection rays are parallel



Orthographic projection equations

$$u_{fp} = (s_p - t_f)^T i_f \beta_u + u_0$$

$$v_{fp} = (s_p - t_f)^T j_f \beta_v + v_0$$

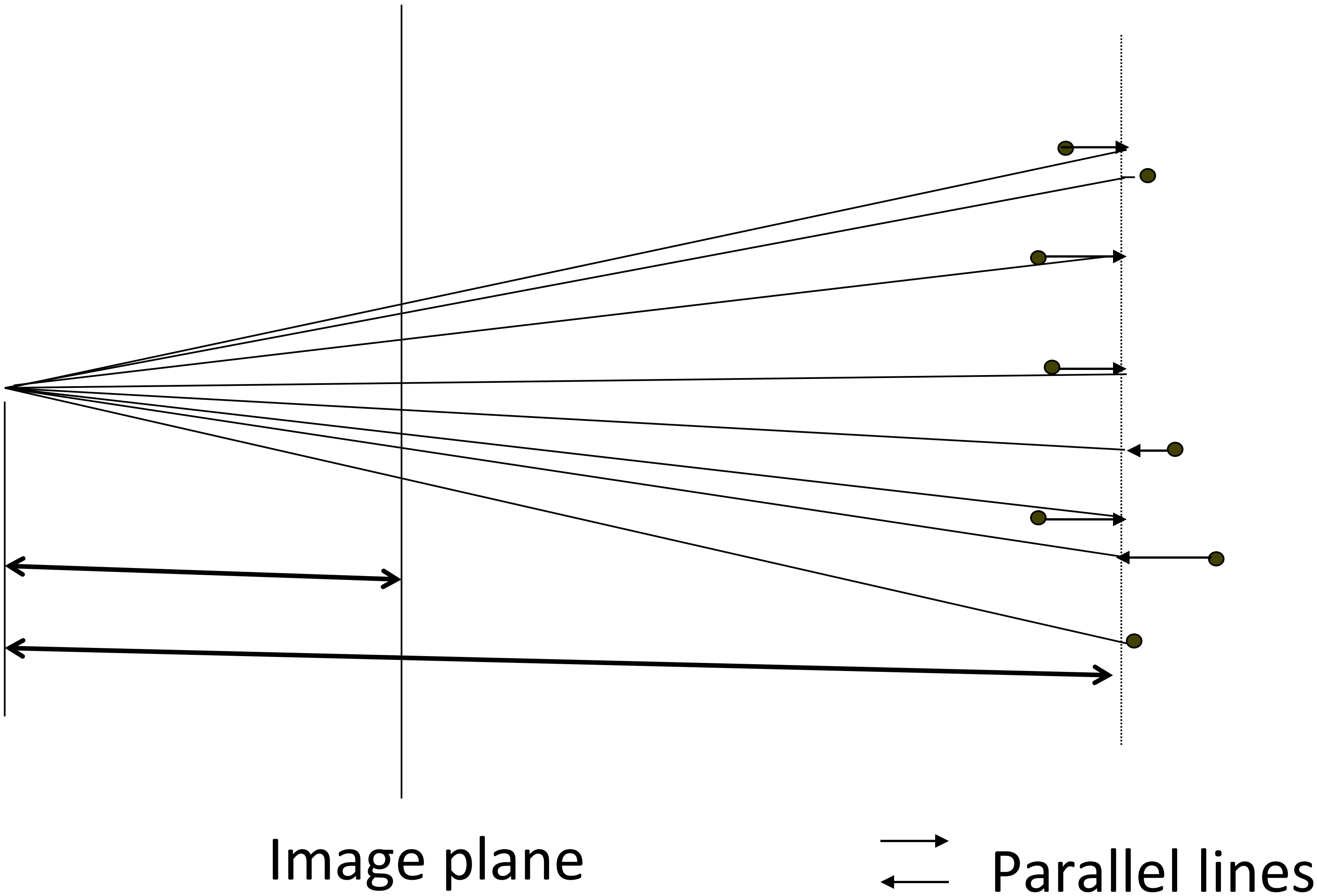
Weak-Perspective Projection Model

- Weak-Perspective Projection is also known as scaled orthographic projection.

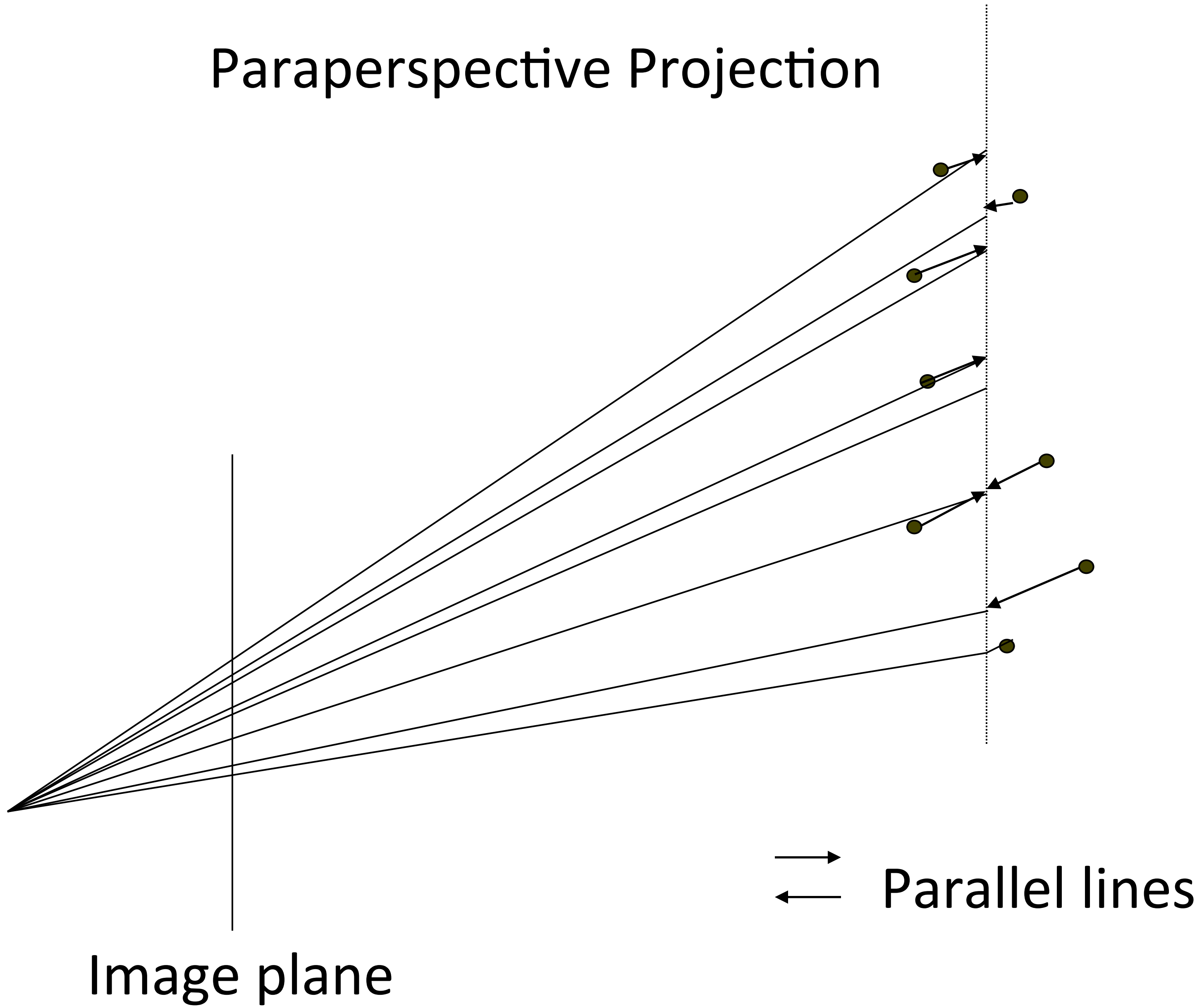
$$u_{fp} = \frac{f(s_p - t_f)^T i_f}{z_f} \beta_u + u_0$$

$$v_{fp} = \frac{f(s_p - t_f)^T j_f}{z_f} \beta_v + v_0$$

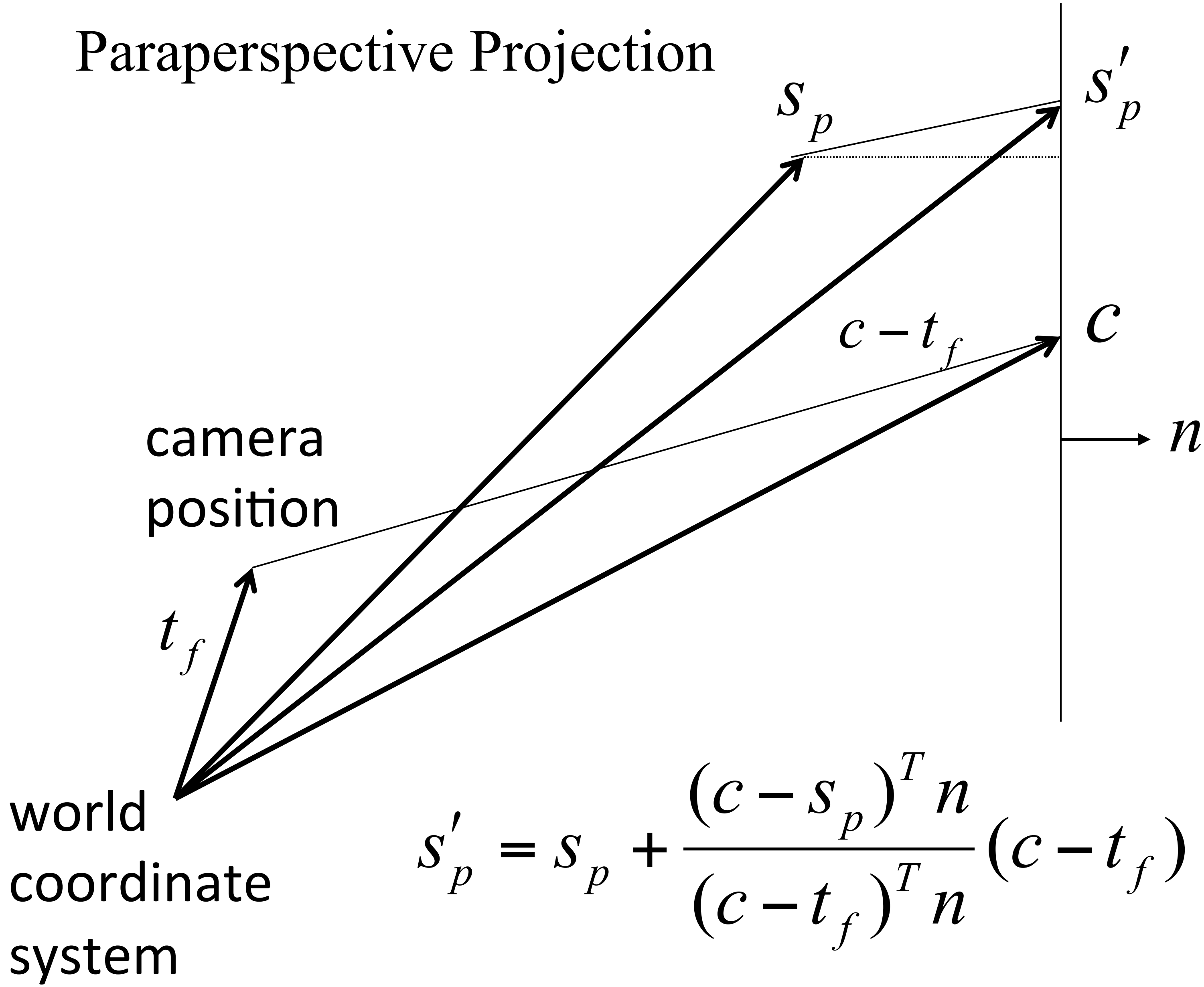
Weak Perspective Projection



Paraperspective Projection



Paraperspective Projection



$$s'_p = s_p + \frac{(c - s_p)^T n}{(c - t_f)^T n}(c - t_f)$$

Choose $n = k_f$

$$s'_p = s_p + \frac{(c - s_p)^T k_f}{(c - t_f)^T k_f}(c - t_f)$$

$$u_{fp} = \frac{(s'_p - t_f)^T i_f}{(c - t_f)^T k_f} f\, k_u + u_0$$

Note that the variable f represents focal length, whereas the subscript f represents frame f.

Similarly,

$$v_{fp} = \frac{(s'_p - t_f)^T j_f}{(c - t_f)^T k_f} f\, k_v + v_0$$

Let $z_f = (c - t_f)^T k_f$

and after some manipulation, get

$$u_{fp} = \frac{1}{z_f}[\, (s_p - c)^T \, (i_f - k_f \frac{(c - t_f)^T i_f}{z_f}) + (c - t_f)^T i_f \,]\, f\, k_u + u_0$$

$$v_{fp} = \frac{1}{z_f}[\, (s_p - c)^T \, (j_f - k_f \frac{(c - t_f)^T j_f}{z_f}) + (c - t_f)^T j_f \,]\, f\, k_v + v_0$$

Additional Comments:

- lens distortion causes distortion in image
- change of focal length (zooming) scales the image (not true if assuming orthographic projection)

camera translation and rotation + 3D scene = image