Image Formation

KOM4520 Fundamentals of Robotic Vision

Most of the contents are adopted from
Robotics, Vision and Control: Fundamental Algorithms in MATLAB, Peter Corke, 2nd ed., Springer Verlag. 2017
R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, 2nd edition, Prentice Hall, 2008
Digital Image Processing: An Algorithmic Introduction using Java, W Burger, Mark J. Burge, Springer Verlag, 2008

Today's lecture

- Geometry of Image Formation
  - Pin-hole cameras
  - Cameras with thin lens
  - Depth of Field
- Perspective Projection
  - Modeling a perspective camera
- Discrete Image Plane
- Camera Matrix
- Lens distortion

Geometry of Image Formation

- How images are formed and captured will be discussed today.

- From images we can deduce the size, shape and position of objects in the world as well as other characteristics such as color and texture which ultimately lead to recognition.

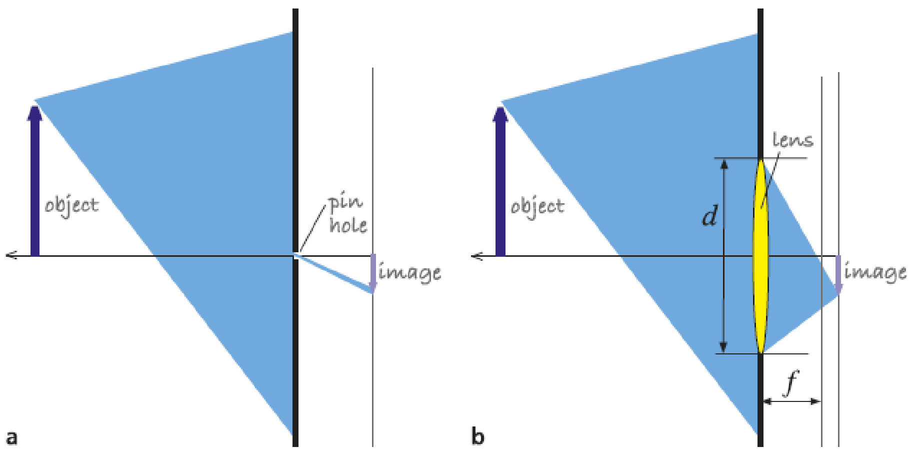
- It has long been known that a simple pin-hole is able to create a perfect inverted image on the wall of a darkened room. The pinhole camera is just a box with very small aperture through which light passes, projecting the image on the photosensitive material (film or photo paper).

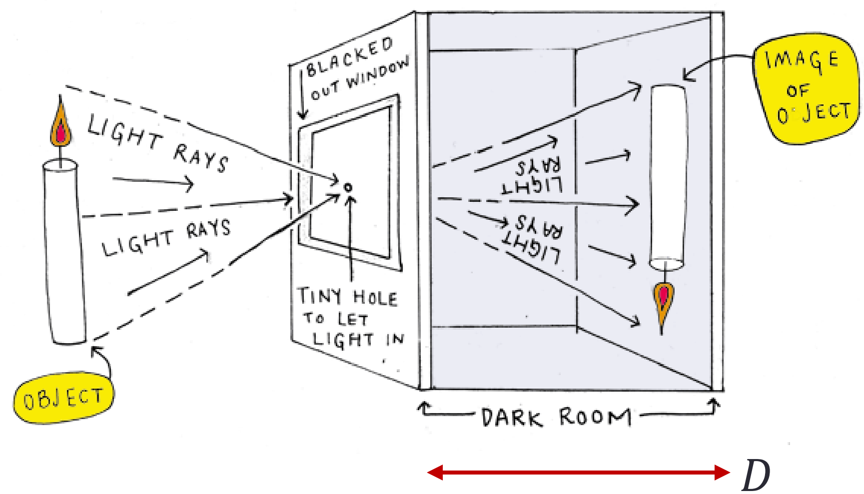
- The process of image formation, in an eye or in a camera, involves a *projection* of the 3-dimensional world onto a 2-dimensional surface.
- The depth information is lost and we can no longer tell from the image whether it is of a large object in the distance or a smaller closer object.
- This transformation from 3 to 2 dimensions is known as perspective projection.

- the topic of camera calibration,
- the estimation of the parameters of the perspective transformation.
- cameras capable of wide-angle, panoramic or light-field imaging.

Pin-hole camera

Light gathering ability of
a pin-hole camera and **b** a lens

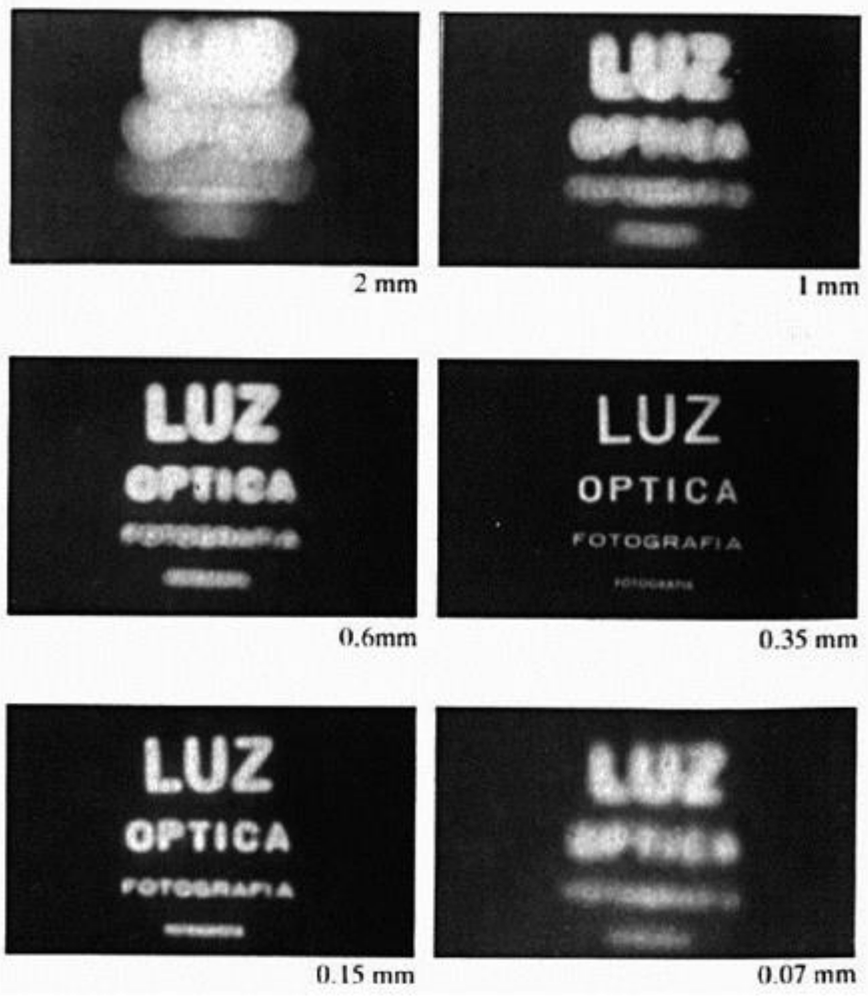




- The pin-hole camera produces a very dim image since its radiant power
- Pinhole Camera: a simple camera without a lens and with a single small aperture, a pinhole.
- A pin-hole camera has no focus adjustments – all objects are in focus irrespective of distance.
- Sharp inverted (upside down) image produced by a small pinhole.
- Bigger image produced when camera is closer to the object.
- Blurred image produced by large pinhole.

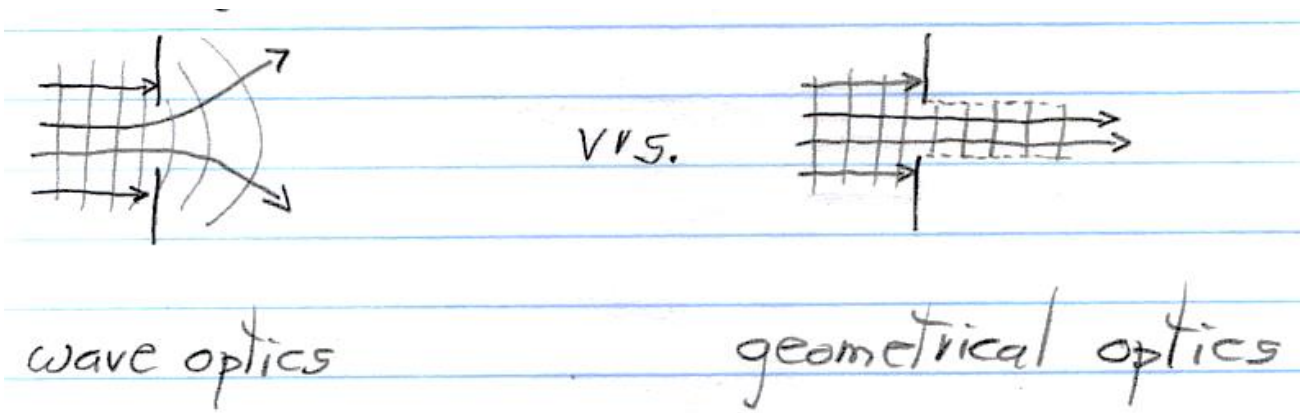
Optimal pinhole diameter : $d = (\lambda D)^{0.5}$ | λ : wavelength of visible light, | D : focal length

Pin-hole camera



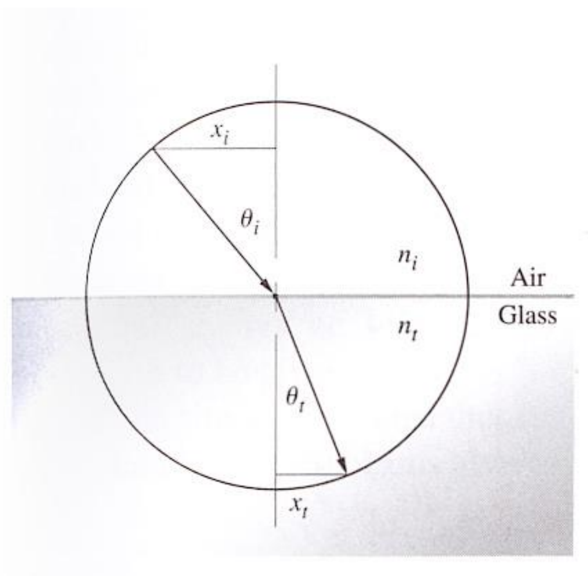
Diffraction of light...

Wave optics vs. Geometrical Optics

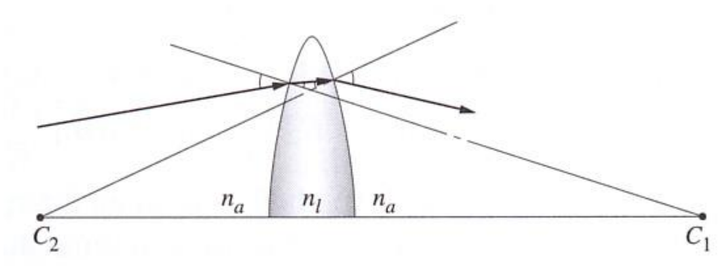


- in geometrical optics, the rays are assumed as that they do not bend as they pass through a narrow passage
- this assumption is valid if the passage is much larger than the wavelength.

Refraction index



$$\frac{x_i}{x_t} = \frac{\sin\theta_i}{\sin\theta_t} = \frac{n_t}{n_i}$$



- As waves change speed at an interface, they also change direction
- Index of refraction n_r is defined as

$$\frac{\text{speed of light in a vacuum}}{\text{speed of light in medium } r}$$

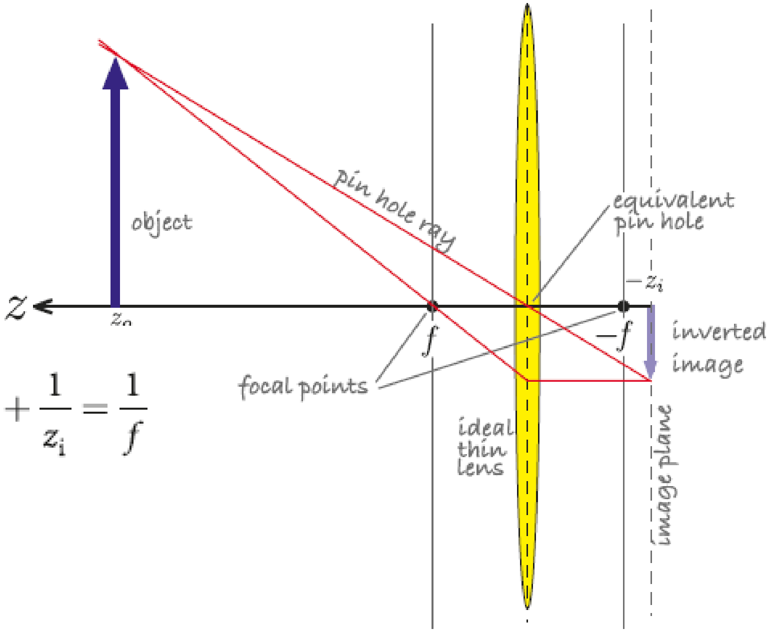
Air = ~1.0

Water = 1.33

Glass = 1.5 – 1.8

- when transiting from air to glass, light bends towards the normal
- when transiting from glass to air, light bends away from the normal
- light striking a surface perpendicularly does not bend.

Thin lens



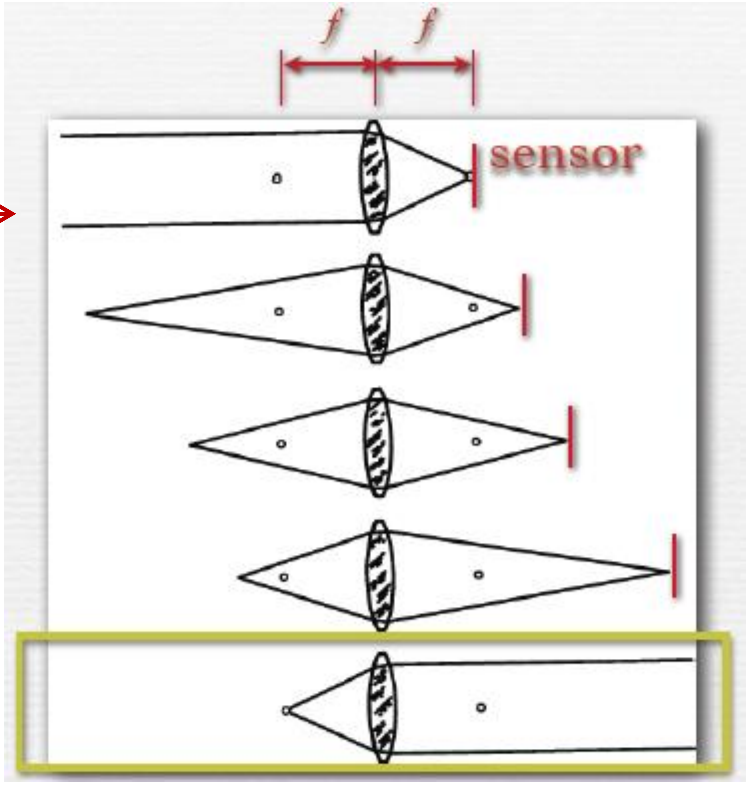
$$\frac{1}{z_o} + \frac{1}{z_i} = \frac{1}{f}$$

- Image formation geometry for a thin convex lens shown in 2-dimensional cross section.
- A lens has two focal points at a distance of *f* on each side of the lens.
- By convention the camera's optical axis is the *z*-axis
- For $z_o > f$ an inverted image is formed on the image plane at $z_i < -f$.
- Lenses have the important advantage of collecting more light than the pinhole admits

- The downside of using a lens is the need to focus.
- Our own eye has a single convex lens made from transparent crystalline proteins, and focus is achieved by muscles which change its shape
- A high quality camera lens is a compound lens comprising multiple glass or plastic lenses.

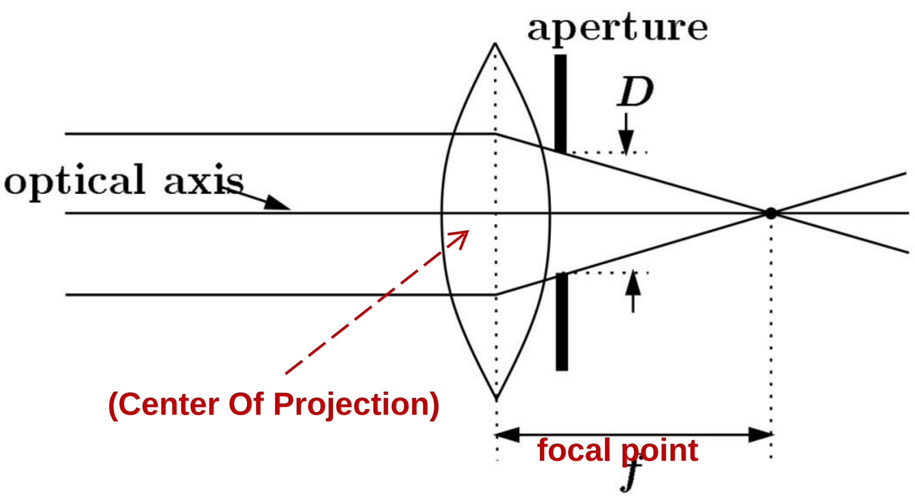
Changing the Focus

- In a camera the image plane is fixed at the surface of the sensor chip so the focus ring of the camera moves the lens along the optical axis so that it is a distance z_i from the image plane.
- for an object at infinity $z_i = f$.
- at $z_o = z_i = 2f$ we have 1:1 imaging,
- In 1:1 imaging, if the sensor is 35mm wide, an object 35mm wide will fill the frame.
- We can’t focus on objects closer to lens than its focal length *f*



$$\frac{1}{z_o} + \frac{1}{z_i} = \frac{1}{f}$$

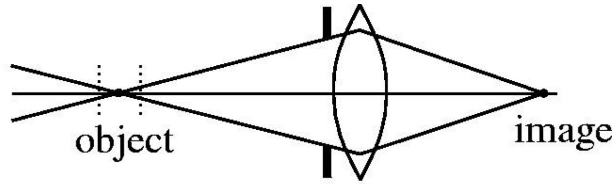
Depth of Field

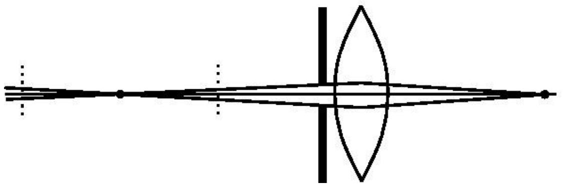


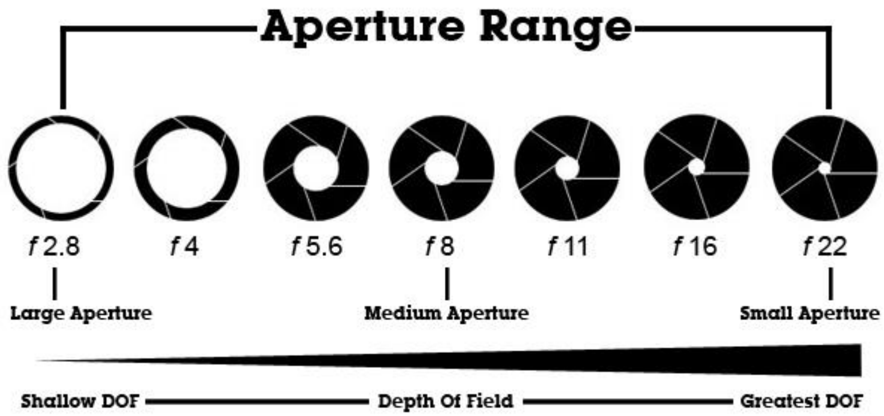
- A lens focuses parallel rays onto a single focal point
  - focal point at a distance f beyond the lens plane

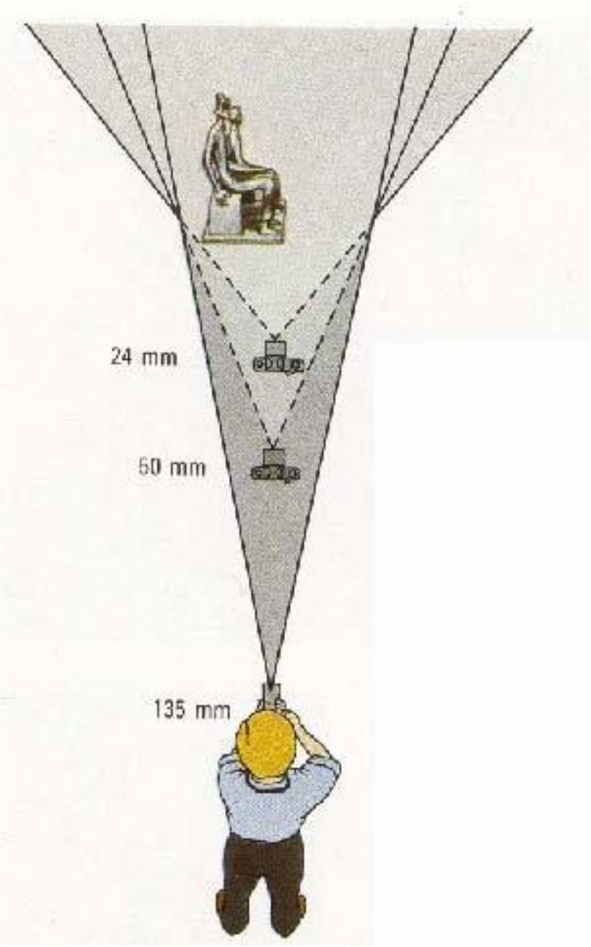
  (f is a function of the shape and index of refraction of the lens)
- Aperture of diameter D restricts the range of ray
  - aperture may be on either side of the lens
- Lenses are typically spherical (easier to produce)

Depth of Field







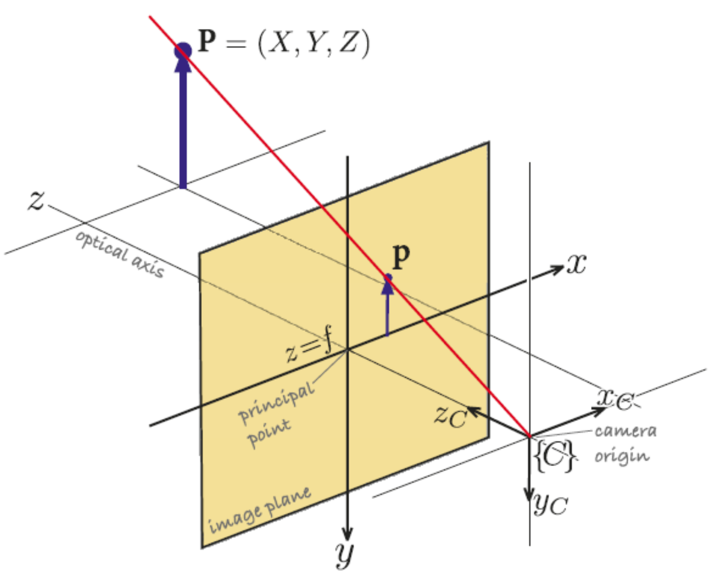


Small f (i.e., large FOV), camera close to car

Large f (i.e., small FOV), camera far from car

Less perspective distortion!

Perspective Projection



The central - perspective model.

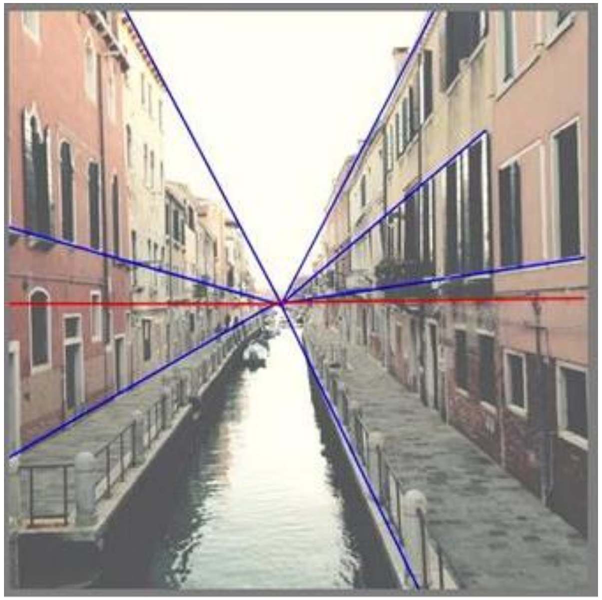
- In computer vision it is common to use the central perspective imaging model.
- The rays converge on the origin of the camera frame {*C*} and a non-inverted image is *projected* onto the image plane located at $z = f$.
- The *z*-axis intersects the image plane at the principal point which is the origin of the 2D image coordinate frame.
- Using similar triangles we can show that a point at the world coordinates $\boldsymbol{P} = (X, Y, Z)$ is projected to the image point $\boldsymbol{p} = (x, y)$ by

$$x = f\frac{X}{Z} \quad y = f\frac{Y}{Z}$$

This is a perspective projection

Perspective Projection

1. Perspective Projection performs a mapping from 3-dimensional space to the 2-dimensional image plane
2. Straight lines in the world are projected to straight lines on the image plane.
3. Parallel lines in the world are projected to lines that intersect at a vanishing point. In drawing, this effect is known as foreshortening. The exception are the lines lying in a plane parallel to the image plane which always remain parallel
4. Conics in the world are projected to conics on the image plane. For example, a circle is projected as a circle or an ellipse.
5. The size (area) of a shape is not preserved and depends on distance.
6. The mapping is not one-to-one and no unique inverse exists. That is, given (*x*, *y*) we cannot uniquely determine (*X*, *Y*, *Z*).
7. The transformation is not conformal – it does not preserve shape since internal angles are not preserved. Translation, rotation and scaling are examples of conformal transformations.



Modeling a Perspective Camera

- We can write the image-plane point coordinates in homogeneous form

  $\tilde{\boldsymbol{p}} = [\tilde{x}, \tilde{y}, \tilde{z}]^{\mathrm{T}}$: homogenous image coordinates

$$\tilde{x} = fX$$
$$\tilde{y} = fY$$
$$\tilde{z} = Z$$
$$\tilde{\boldsymbol{p}} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

  Non-homogenous image coordinates are

$$x = \frac{fX}{Z}, \qquad y = \frac{fY}{Z}$$

- If we write a world coordinate in homogenous for as well

$${}^{C}\tilde{\boldsymbol{P}} = [X, Y, Z, 1]^{\mathrm{T}}$$

${}^{C}\tilde{\boldsymbol{P}}$: the coordinate of the point with respect to the camera frame $\{C\}$

$$\tilde{\boldsymbol{p}} = \boldsymbol{C}\,{}^{C}\tilde{\boldsymbol{P}}$$

$\boldsymbol{C}$: 3x4 camera matrix

$$\tilde{\boldsymbol{p}} = \underbrace{\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\boldsymbol{C}} {}^{C}\tilde{\boldsymbol{P}} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} {}^{C}\tilde{\boldsymbol{P}}$$

Modeling a Perspective Camera

```
cam = CentralCamera('focal', 0.015);
```

returns an instance of a CentralCamera object with a 15 mm lens.

We can define a world point

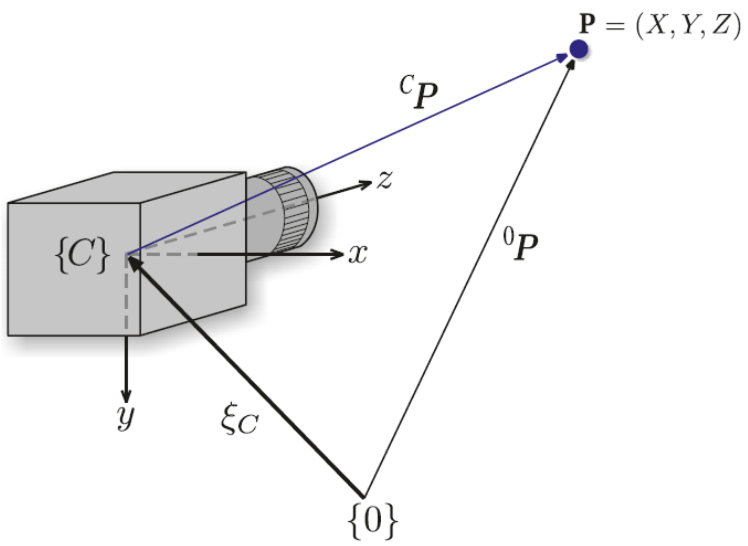
```
P = [0.3, 0.4, 3.0]';
```

in units of meters and the corresponding image-plane coordinates are

```
cam.project(P)

ans =
0.0015
0.0020
```

The point on the image plane is at (1.5, 2.0) mm with respect to the principal point.

$$^{C}\boldsymbol{P} = (\ominus\xi_C) \cdot {}^{0}\boldsymbol{P}$$



Modeling a Perspective Camera

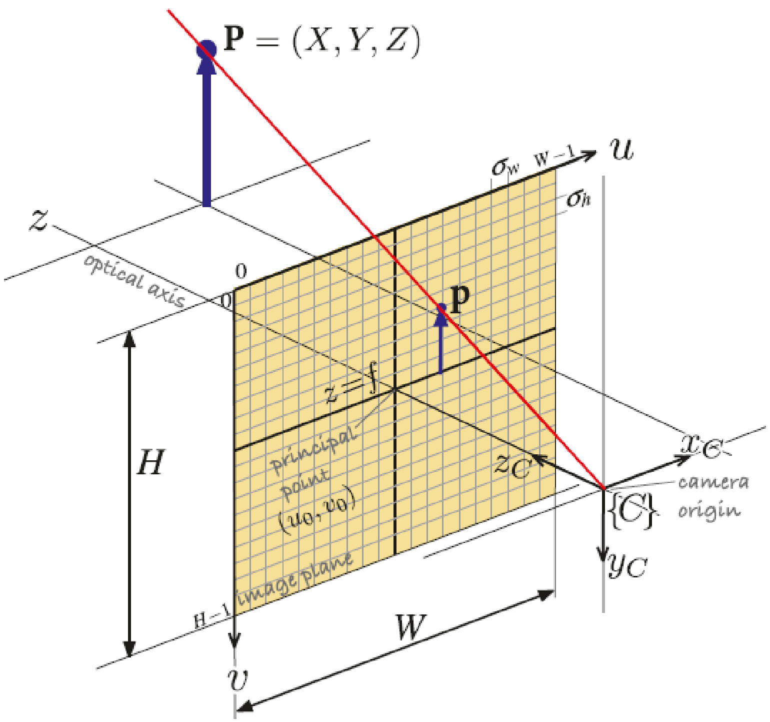
- In general the camera will have an arbitrary pose ξ_C with respect to the world coordinate frame
- The position of the point with respect to the camera is

$$ {}^{C}\boldsymbol{P} = (\boldsymbol{T}_C)^{-1}\, {}^{O}\boldsymbol{P} $$

```
cam.project(P, 'pose', SE3(-0.5, 0, 0) )
ans =
0.0040
0.0020
```

- where the third argument is the pose of the camera ξ_C as a homogeneous transformation.
- We see that the *x*-coordinate has increased from 1.5 mm to 4.0 mm, that is, the image point has moved to the *right*.

Discrete Image Plane



Central projection model showing image plane and discrete pixels

- In a digital camera the image plane is a $W \times H$ grid of light-sensitive elements called photosites that correspond directly to the picture elements (or pixels) of the image.
- The pixel coordinates are a 2-vector (u, v) of nonnegative integers and by convention the origin is at the top-left hand corner of the image plane.

$$u = \frac{x}{\rho_w} + u_o, \qquad v = \frac{y}{\rho_h} + v_o$$

ρ_w: width of the pixel

ρ_h: height of the pixel

u_o, v_o : is the principal point – the pixel cooordinate of the point where the optical axis intersects the image plane.

$$\tilde{\boldsymbol{p}} = \underbrace{\begin{bmatrix} 1/\rho_w & 0 & u_o \\ 0 & 1/\rho_h & v_o \\ 0 & 0 & 1 \end{bmatrix}}_{\boldsymbol{K}} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} {}^{C}\tilde{\boldsymbol{P}}$$

$\tilde{\boldsymbol{p}} = [\tilde{u}, \tilde{v}, \tilde{w}]^{\mathrm{T}}$: homogenous pixel coordinates of the world point $\boldsymbol{P}$

Non-homogenous pixel coordinates are

$$u = \frac{\tilde{u}}{\tilde{w}}, \qquad v = \frac{\tilde{v}}{\tilde{w}}$$

Discrete Image Plane

if the pixels are $10\ \mu m$ square and the pixel array is 1280×1024 pixels with its principal point at image-plane coordinate $(640, 512)$ then

```
cam = CentralCamera('focal', 0.015, 'pixel', 10e-6, 'resolution', [1280 1024],
'centre', [640 512], 'name', 'mycamera')
cam =
name: mycamera [central-perspective]
focal length: 0.015
pixel size: (1e-05, 1e-05)
principal pt: (640, 512)
number pixels: 1280 x 1024
pose: t = (0,0,0), RPY/yxz = (0,0,0) deg
```

which displays the parameters of the camera model including the camera pose $\boldsymbol{T}$. The corresponding nonhomogeneous pixel coordinates of the previously defined world point are

```
>> cam.project(P)
ans =
790
712
```

Camera Matrix

$$\underbrace{\tilde{\boldsymbol{p}} = \begin{bmatrix} f/\rho_w & 0 & u_o \\ 0 & f/\rho_h & v_o \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\text{intrinsic}}\underbrace{\left({}^0\boldsymbol{T}_C\right)^{-1}}_{\text{extrinsic}}\tilde{\boldsymbol{P}}$$

$$\tilde{\boldsymbol{p}} = \boldsymbol{K}\boldsymbol{P}_0\left({}^0\boldsymbol{T}_C\right)^{-1}\tilde{\boldsymbol{P}} = \boldsymbol{C}\tilde{\boldsymbol{P}}$$

- Where all the terms are rolled up into the camera matrix $\boldsymbol{C}$. This is a 3 × 4 homogeneous transformation which performs scaling, translation and perspective projection. It is often also referred to as the projection matrix or the camera calibration matrix.
- We have already mentioned the fundamental ambiguity with perspective projection, that we cannot distinguish between a large distant object and a smaller closer object. We can rewrite the equation as

$$\tilde{\boldsymbol{p}} = (\boldsymbol{C}\boldsymbol{H}^{-1})\left(\boldsymbol{H}\tilde{\boldsymbol{P}}\right) = \boldsymbol{C}'\tilde{\boldsymbol{P}}'$$

$\boldsymbol{H}$**:** is an arbitrary nonsingular 3 × 3 matrix.

- This implies that an infinite number of camera $\boldsymbol{C}'$ and world point coordinate $\tilde{\boldsymbol{P}}'$ combinations will result in the same image-plane projection p.
- This illustrates the essential difficulty in determining 3-dimensional world coordinates from 2-dimensional projected coordinates. It can only be solved if we have information about the camera or the 3-dimensional object.

Camera Matrix

- The projection can also be written in functional form as
$$\boldsymbol{p} = \mathcal{P}(\boldsymbol{P}, \boldsymbol{K}, \xi_C)$$

- $\boldsymbol{P}$: point coordinate vector in the world frame.
- $\boldsymbol{K}$: camera parameter matrix comprises the intrinsic parameters which are the innate characteristics of the camera and sensor such as f, ρ_w, ρ_h, u_0 and v_0.
- ξ_C is the pose of the camera and comprises a minimum of **six** parameters – the extrinsic parameters – that describe camera translation and orientation in **SE**(3).

- There are **5 intrinsic and 6 extrinsic parameters** – a total of 11 independent parameters to describe a camera.
- The camera matrix has 12 elements so one degree of freedom, the overall scale factor, is unconstrained and can be arbitrarily chosen.

- In practice the camera parameters are not known and must be estimated using a camera calibration procedure

Camera Matrix

The camera intrinsic parameter matrix K for this camera is (focal length: 0.015 pixel size: (1e-05, 1e-05))

```
cam.K
ans =
1.0e+03 *
1.5000 0      0.6400
0      1.5000 0.5120
0      0      0.0010
```

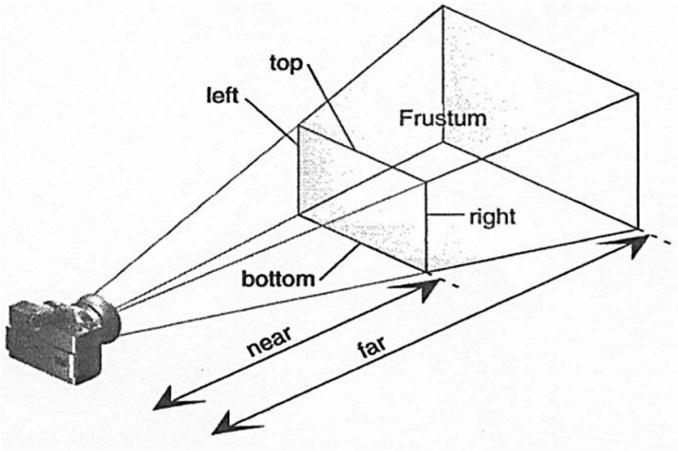
The camera matrix is implicitly created when the Toolbox camera object is constructed and for this example is

```
cam.C
ans =
1.0e+03 *
1.5000 0      0.6400 0
0      1.5000 0.5120 0
0      0      0.0010 0
```

Camera Matrix

- The **field of view** of a lens is an open rectangular pyramid, a frustum, that subtends angles θ_h and θ_v in the horizontal and vertical planes respectively.
- A *normal lens* is one with a field of view around $50°$, while a wide angle lens has a fi eld of view $> 60°$.
- Beyond 110° it is difficult to create a lens that maintains perspective projection, so non-perspective fisheye lenses are required.

- The field of view of a camera is a function of its focal length *f*.
- A wide-angle lens has a small focal length, a telephoto lens has a large focal length, and a zoom lens has an adjustable focal length.



Camera Matrix

- In the horizontal direction the half-angle of view is $\frac{\theta_h}{2} = \tan^{-1}\frac{W/2\rho_w}{f}$
- W: is the number of pixels in the horizontal direction.
- We can then write

$$\theta_h = 2\tan^{-1}\frac{W\rho_w}{2f}, \theta_v = 2\tan^{-1}\frac{H\rho_h}{2f}$$

- We note that the field of view is also a function of the dimensions of the camera chip which
- is $W\rho_w \times H\rho_h$.

The field of view is computed by the fov method of the camera object

```
cam.fov() * 180/pi
ans =
46.2127 37.6930
```

in degrees in the horizontal and vertical directions respectively.

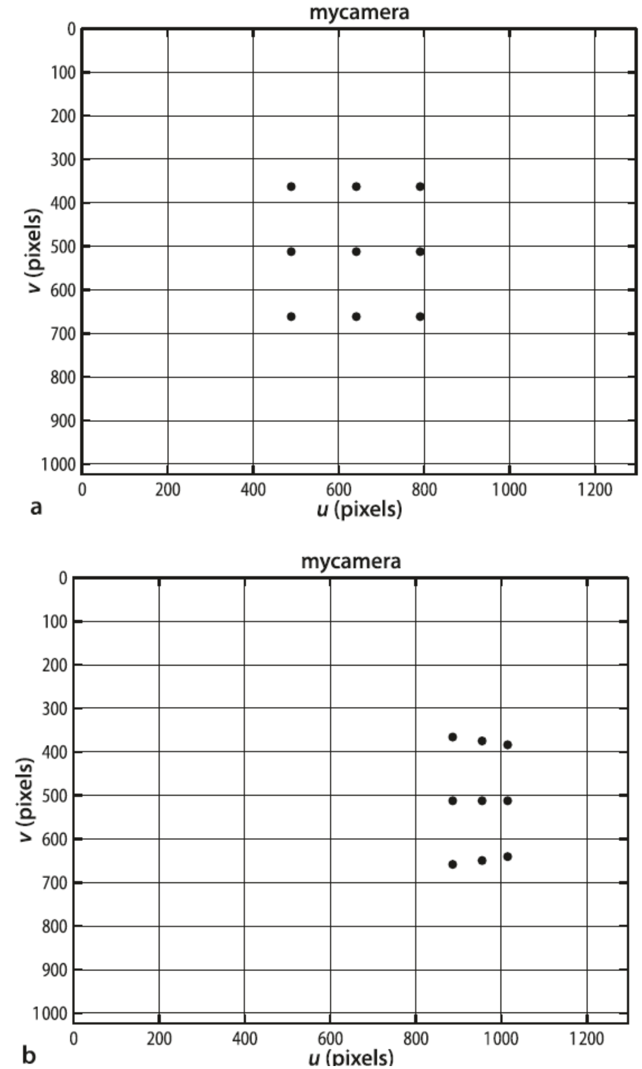
Projecting Points

The CentralCamera class is a subclass of the Camera class and inherits the ability to project multiple points or lines. We can create a 3 × 3 grid of points in the *xy*-plane with overall side length 0.2 m and centered at (0, 0, 1)

```
P = mkgrid(3, 0.2, 'pose', SE3(0, 0, 1.0));
```

which returns a 3 × 9 matrix with one column per grid point where each column comprises the coordinates in *X*, *Y*, *Z* order.

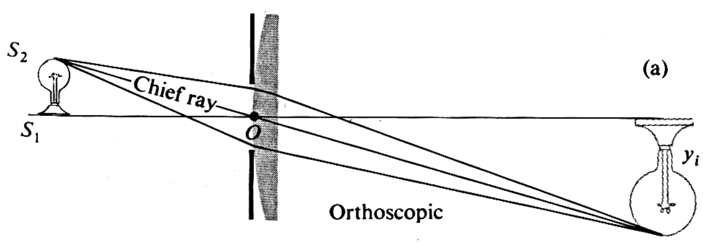
```
cam.project(P)
ans =
490 490 490 640 640 640 790 790 790
362 512 662 362 512 662 362 512 662

cam.plot(P)

Tcam = SE3(-1,0,0.5)*SE3.Ry(0.9);
cam.plot(P, 'pose', Tcam)
```

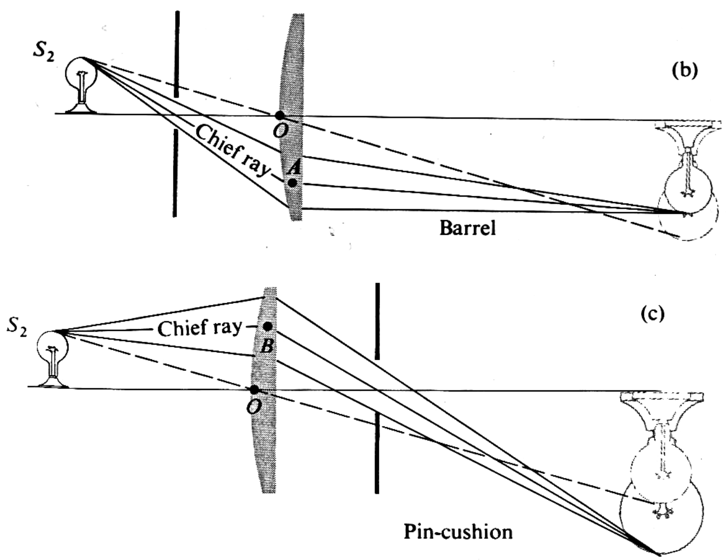


Lens distortion

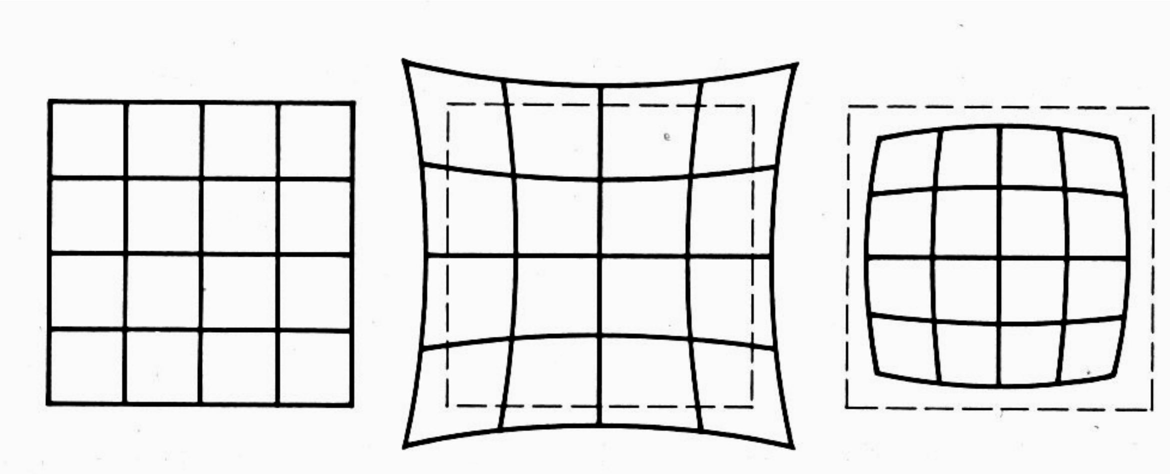
- No lenses are perfect and the low-cost lenses used in many webcams are far from perfect.
- Lens imperfections result in a variety of distortions including chromatic aberration (color fringing), spherical aberration or astigmatism (variation in focus across the scene), and geometric distortions where points on the image plane are displaced.

- Geometric distortion is generally the most problematic effect that we encounter for robotic applications, and comprises two components: radial and tangential. Radial distortion causes image points to be translated along radial lines from the principal point.





Lens distortion



No distortion Pin cushion Barrel

- Radial distortion of the image
  - Caused by imperfect lenses
  - Deviations are most noticeable for rays that pass through the edge of the lens

Lens distortion

- The coordinate of the point (u, v) after distortion is given by

$$u^d = u + \delta_u$$
$$v^d = v + \delta_v$$

$$\begin{bmatrix} \delta_u \\ \delta_v \end{bmatrix} = \underbrace{\begin{bmatrix} u(k_1 r^2 + k_2 r^4 + k_3 r^6 + \cdots) \\ v(k_1 r^2 + k_2 r^4 + k_3 r^6 + \cdots) \end{bmatrix}}_{\text{radial}} + \underbrace{\begin{bmatrix} 2p_1 uv + p_2(r^2 + 2u^2) \\ p_1(r^2 + 2v^2) + 2p_2 uv \end{bmatrix}}_{\text{tangential}}$$

- In practice three coeffi cients are suffi cient to describe the radial distortion and the
- distortion model is parameterized by (k1, k2, k3, p1, p2) which are considered as addi tional intrinsic parameters. Distortion can be modeled by the CentralCamera class using the 'distortion' option, for example

- `cam = CentralCamera('focal', 0.015, 'pixel', 10e-6, 'resolution', [1280 1024], 'centre', [512 512], 'distortion', [k1 k2 k3 p1 p2] )`

Lens distortion

(Radial distortion)
distorted image

Corrected image