Feature descriptors and matching

Detections at multiple scales

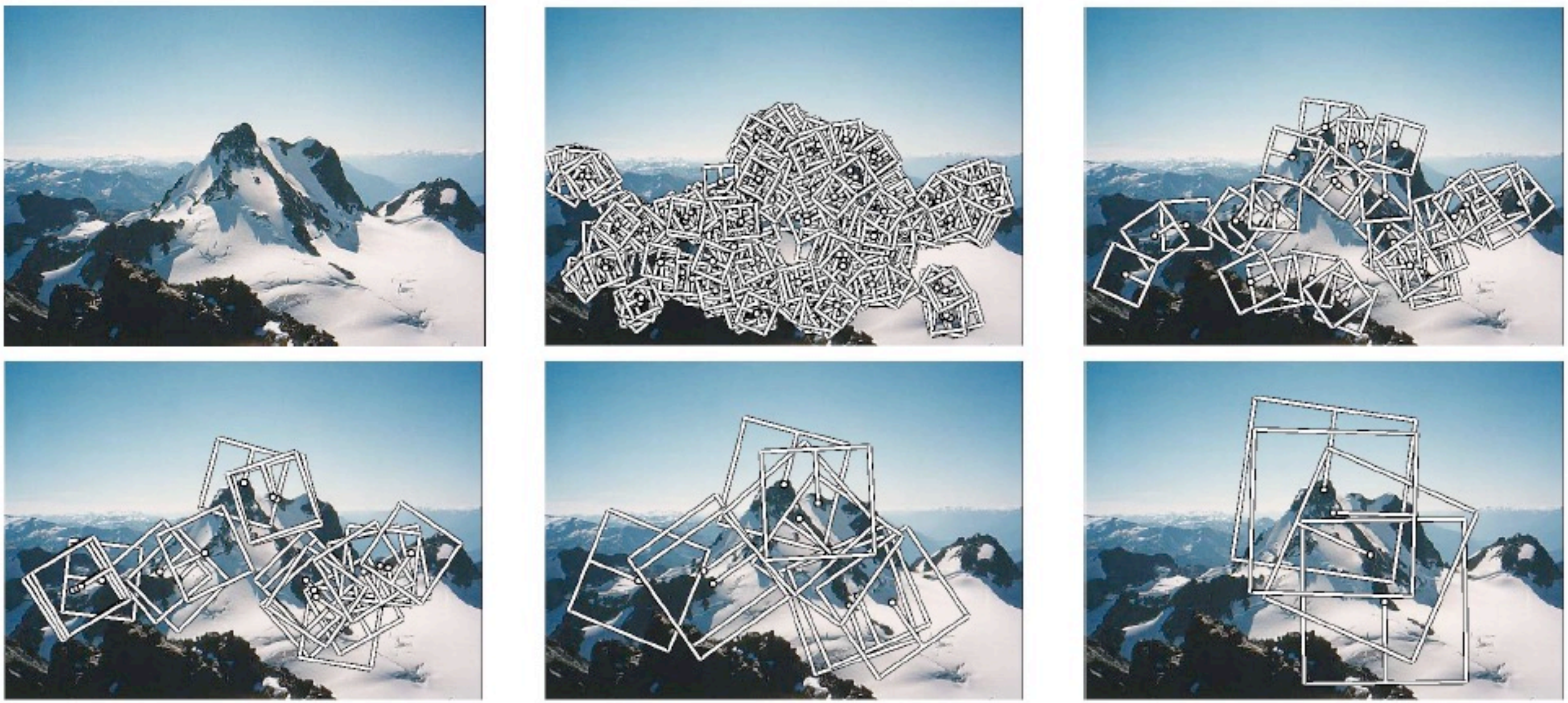


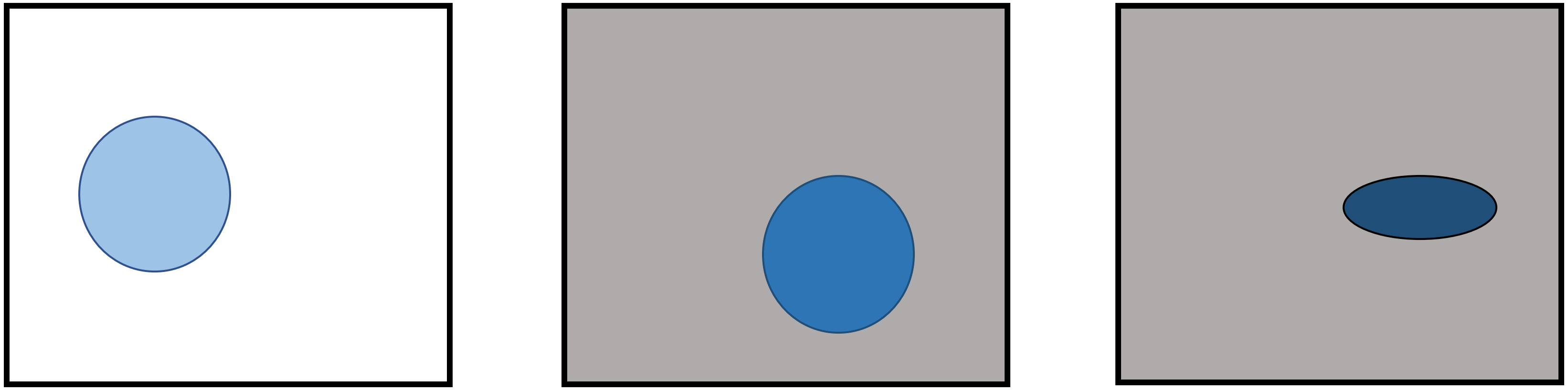
Figure 1. Multi-scale Oriented Patches (MOPS) extracted at five pyramid levels from one of the Matier images. The boxes show the feature orientation and the region from which the descriptor vector is sampled.

Invariance of MOPS

- Intensity
- Scale
- Rotation

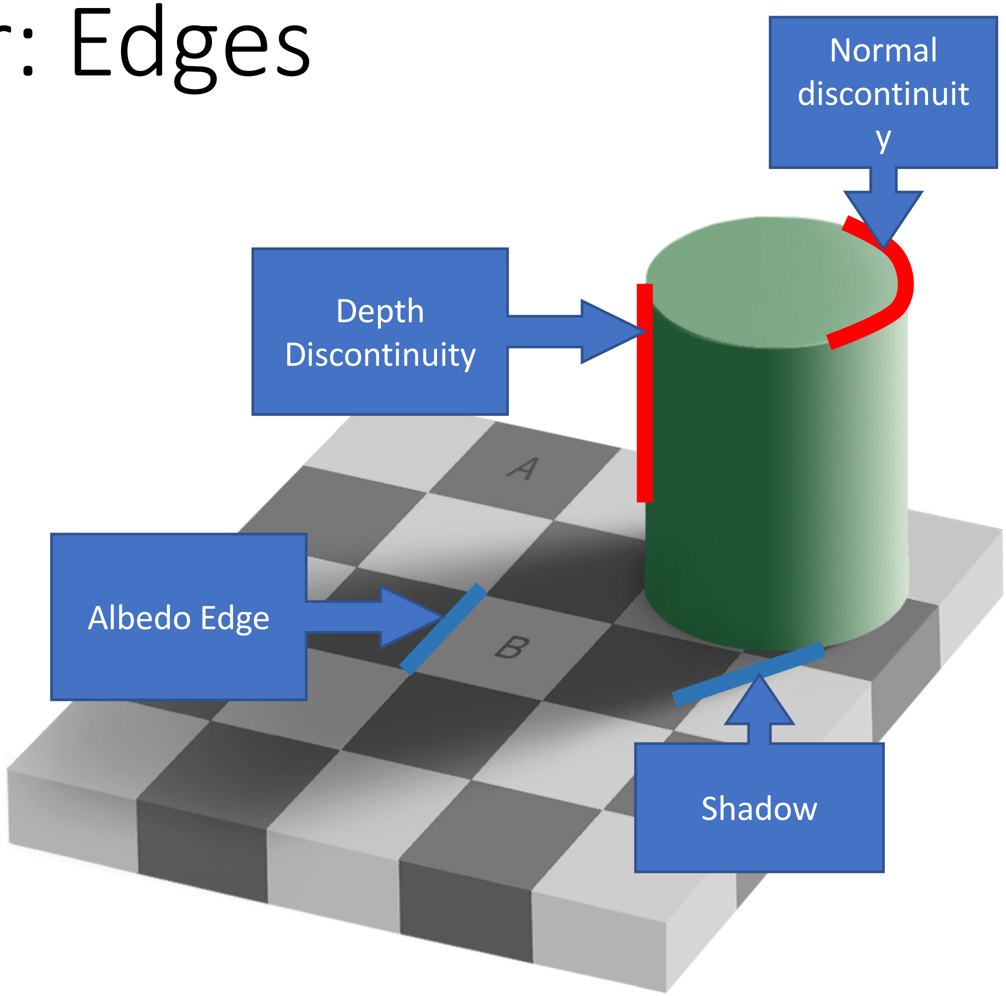
Color and Lighting

Out-of-plane rotation



Out-of-plane rotation

Better representation than color: Edges

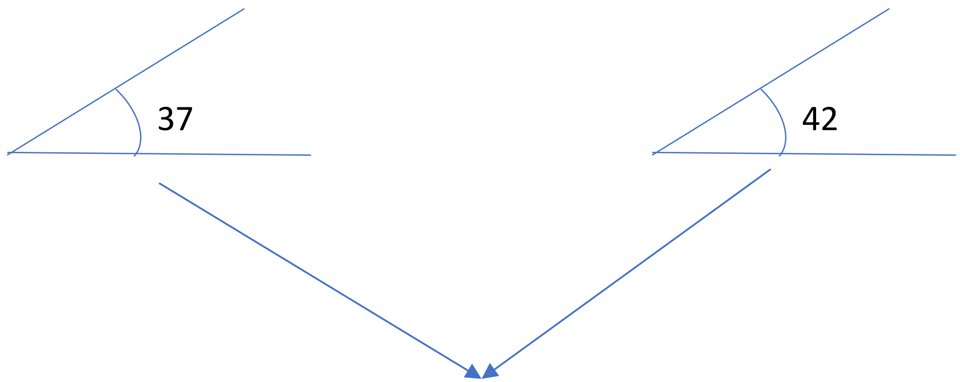


Towards a better feature descriptor

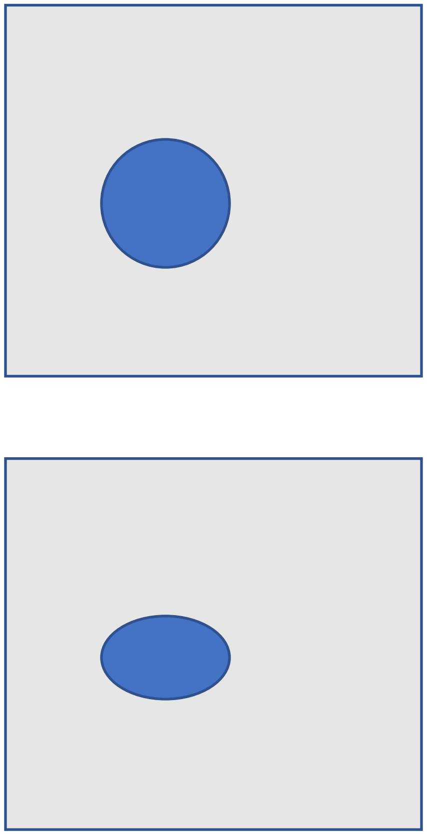
- Match *pattern of edges*
  - Edge orientation – clue to shape

- Be resilient to *small deformations*
  - Deformations might move pixels around, but slightly
  - Deformations might change edge orientations, but slightly

Invariance to deformation by quantization



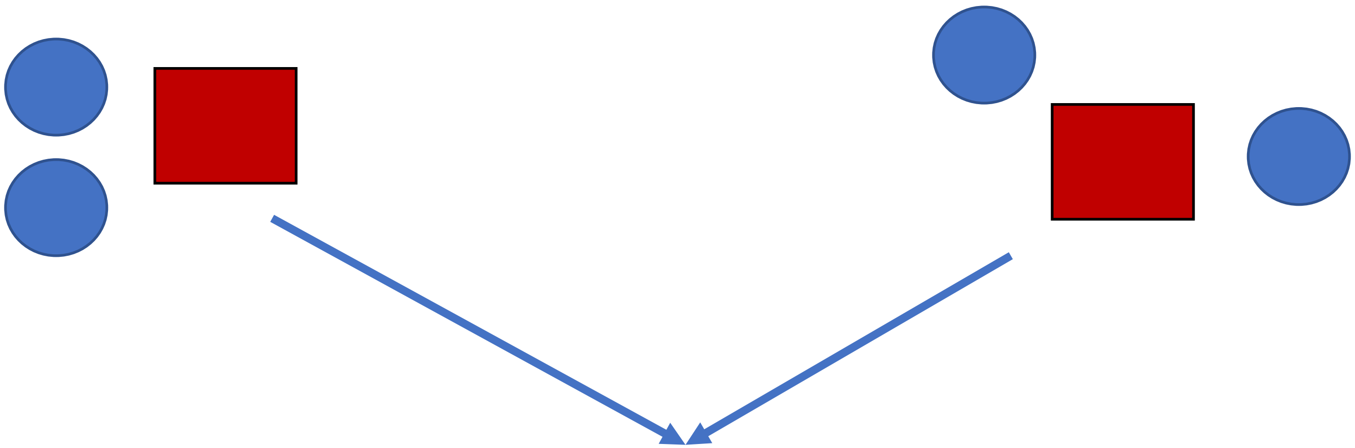
Between 30 and 45



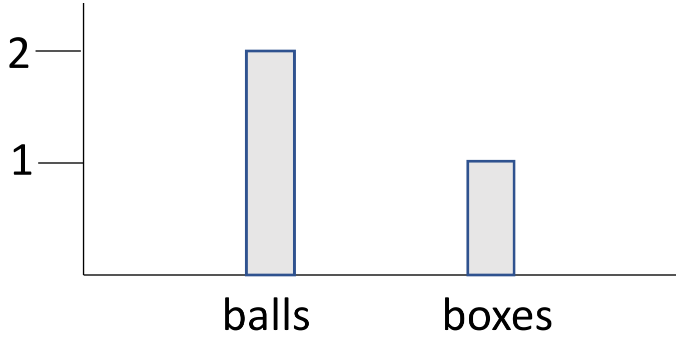
Invariance to deformation by quantization

$$g(\theta) = \begin{cases} 0 & \text{if } 0 < \theta < 2\pi/N \\ 1 & \text{if } 2\pi/N < \theta < 4\pi/N \\ 2 & \text{if } 4\pi/N < \theta < 6\pi/N \\ & \dots \\ N-1 & \text{if } 2(N-1)\pi/N \end{cases}$$

Spatial invariance by histograms



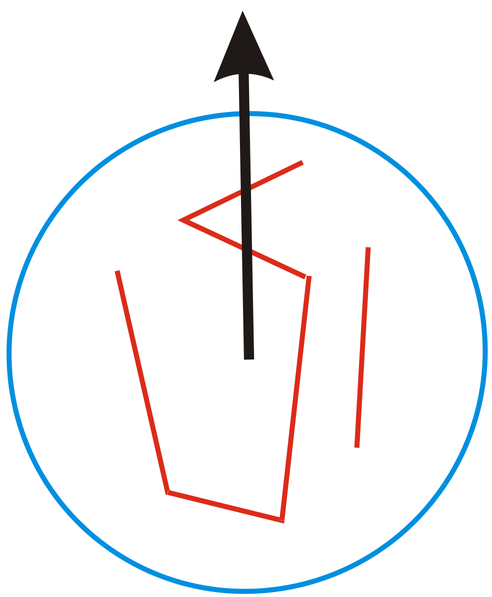
2 blue balls, one red box

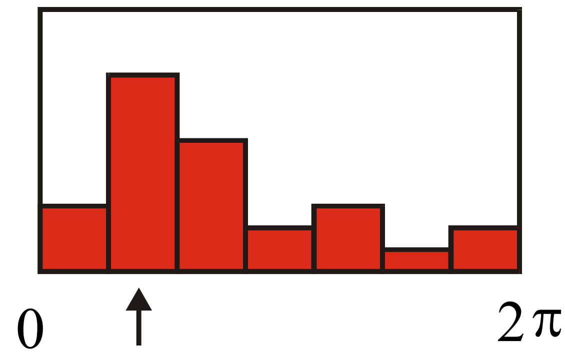


Rotation Invariance by Orientation Normalization

[Lowe, SIFT, 1999]

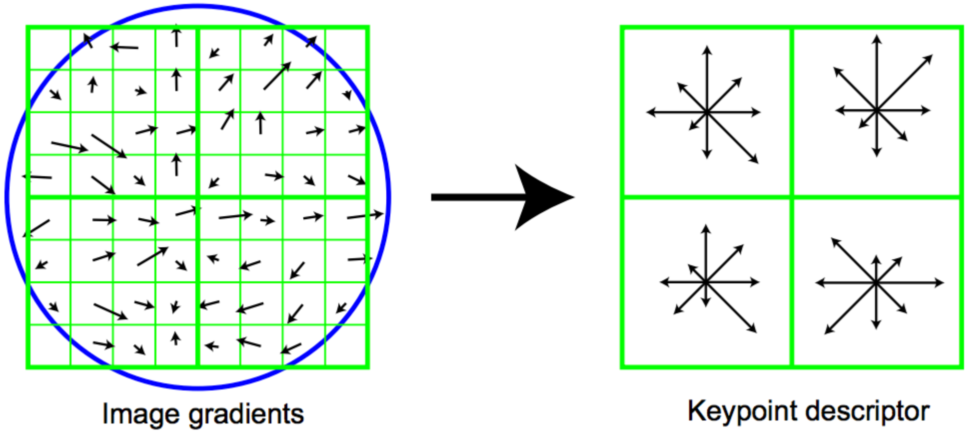
- Compute orientation histogram
- Select dominant orientation
- Normalize: rotate to fixed orientation





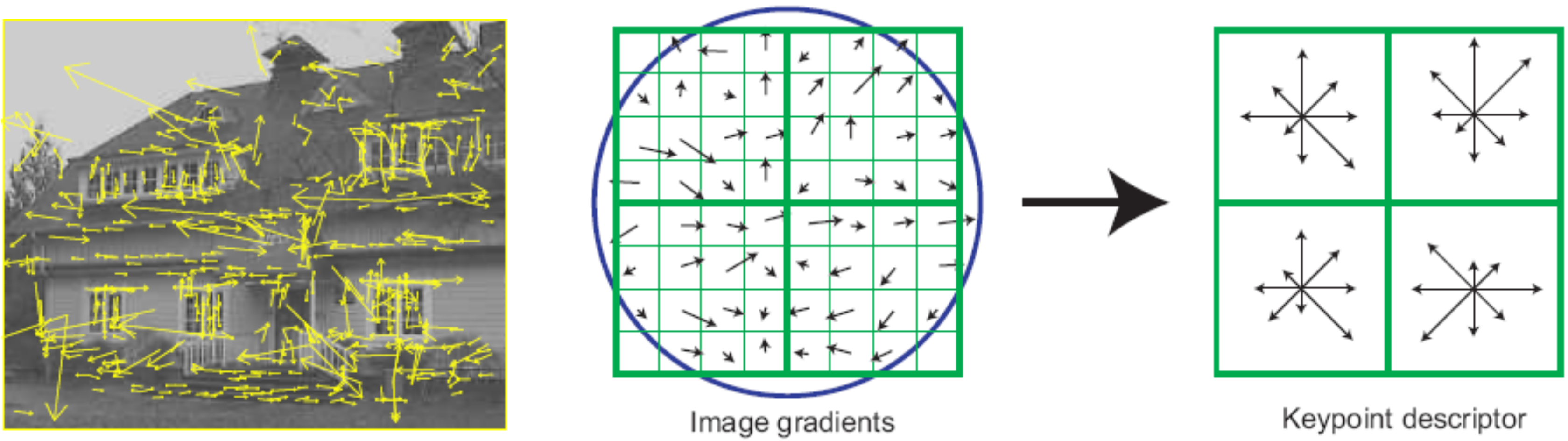
The SIFT descriptor

- Compute edge magnitudes + orientations
- Quantize orientations *(invariance to def)*
- Divide into spatial cells
- Compute orientation histogram in each cell *(spatial invariance)*



Distinctive Image Features from Scale-Invariant Keypoints. Lowe. In IJCV 2004

The SIFT descriptor

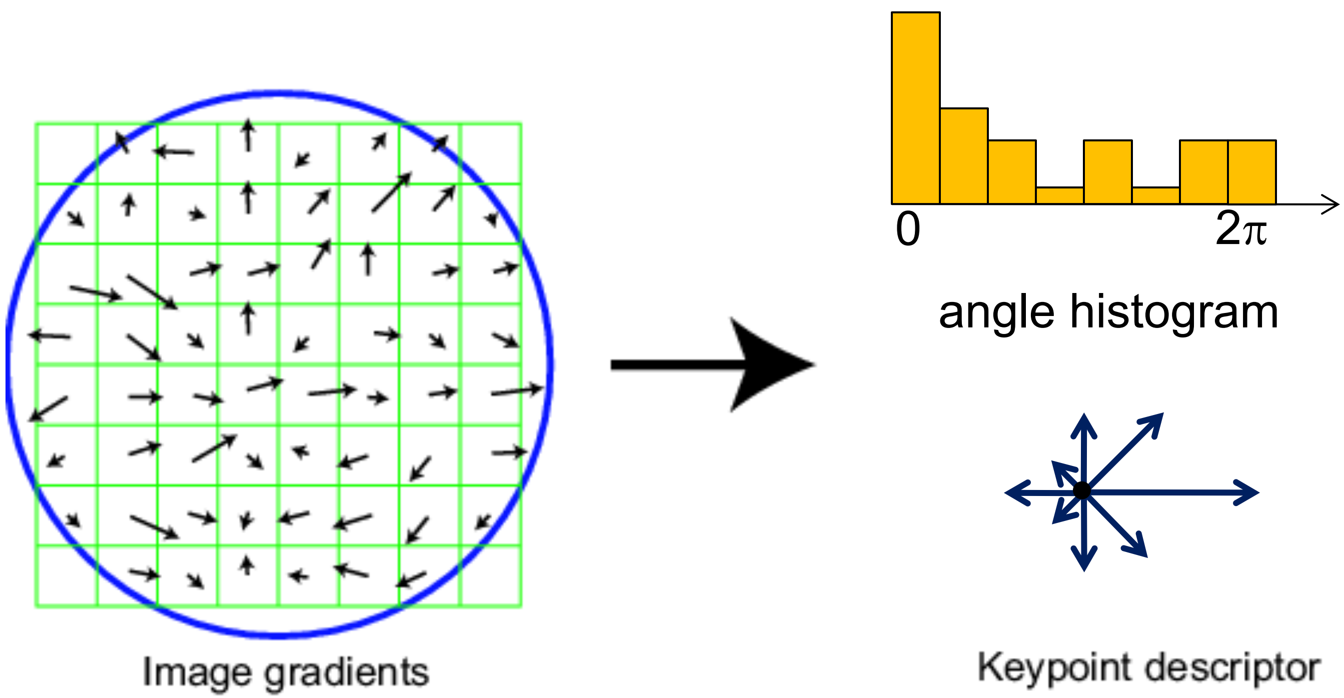


SIFT – Lowe IJCV 2004

Scale Invariant Feature Transform

Basic idea:

- DoG for scale-space feature detection
- Take 16x16 square window around detected feature
  - Compute gradient orientation for each pixel
  - Throw out weak edges (threshold gradient magnitude)
  - Create histogram of surviving edge orientations

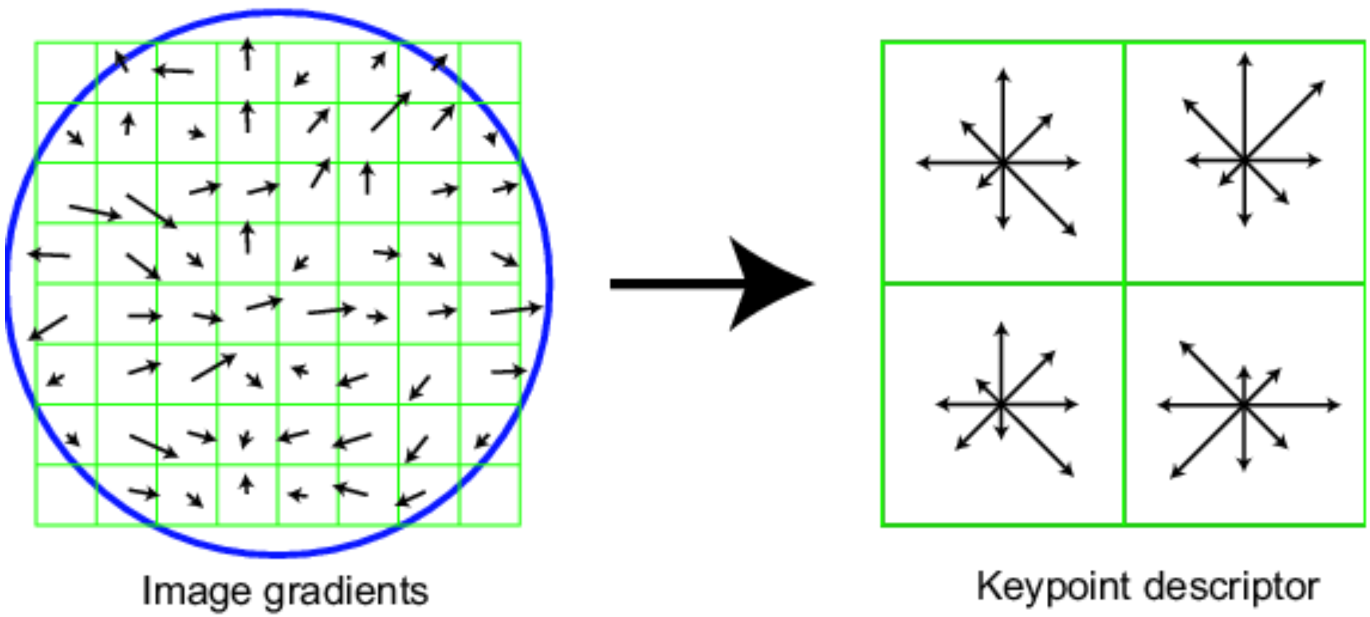


Adapted from slide by David Lowe

SIFT descriptor

Create histogram

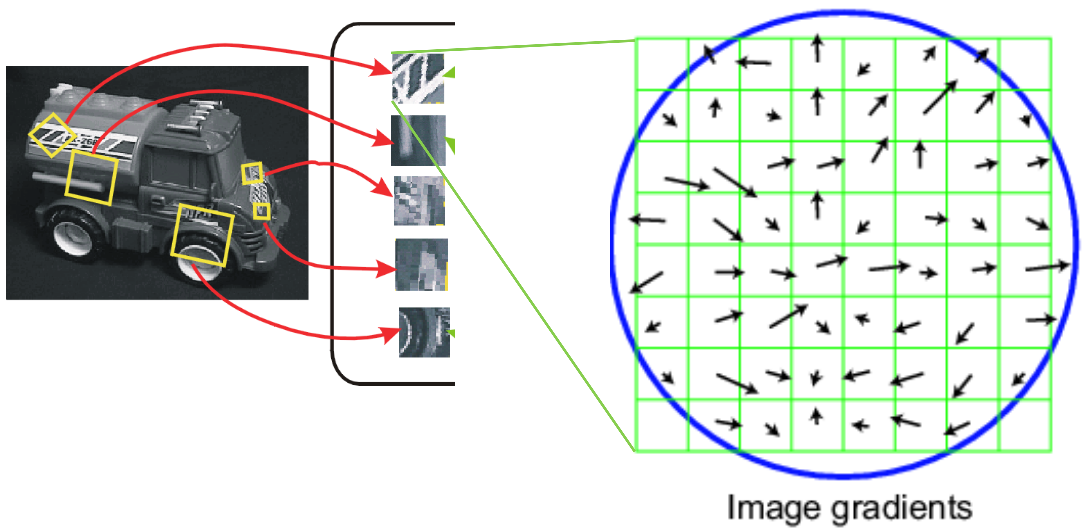
- Divide the 16x16 window into a 4x4 grid of cells (2x2 case shown below)
- Compute an orientation histogram for each cell
- 16 cells * 8 orientations = 128 dimensional descriptor



Adapted from slide by David Lowe

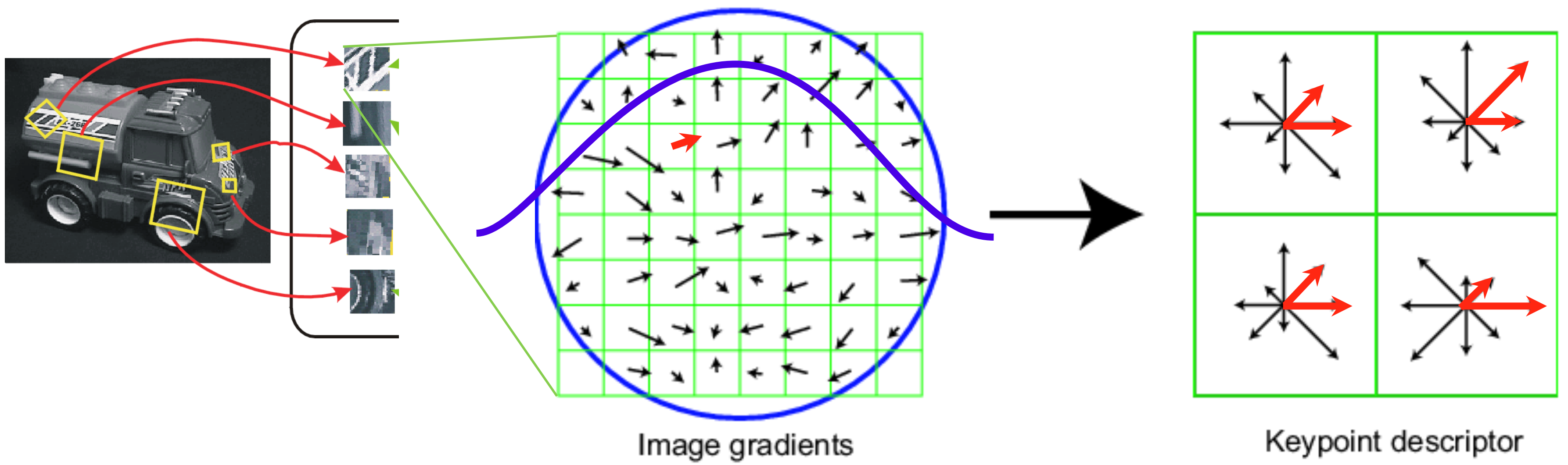
SIFT vector formation

- Computed on rotated and scaled version of window according to computed orientation & scale
  - resample the window
- Based on gradients weighted by a Gaussian



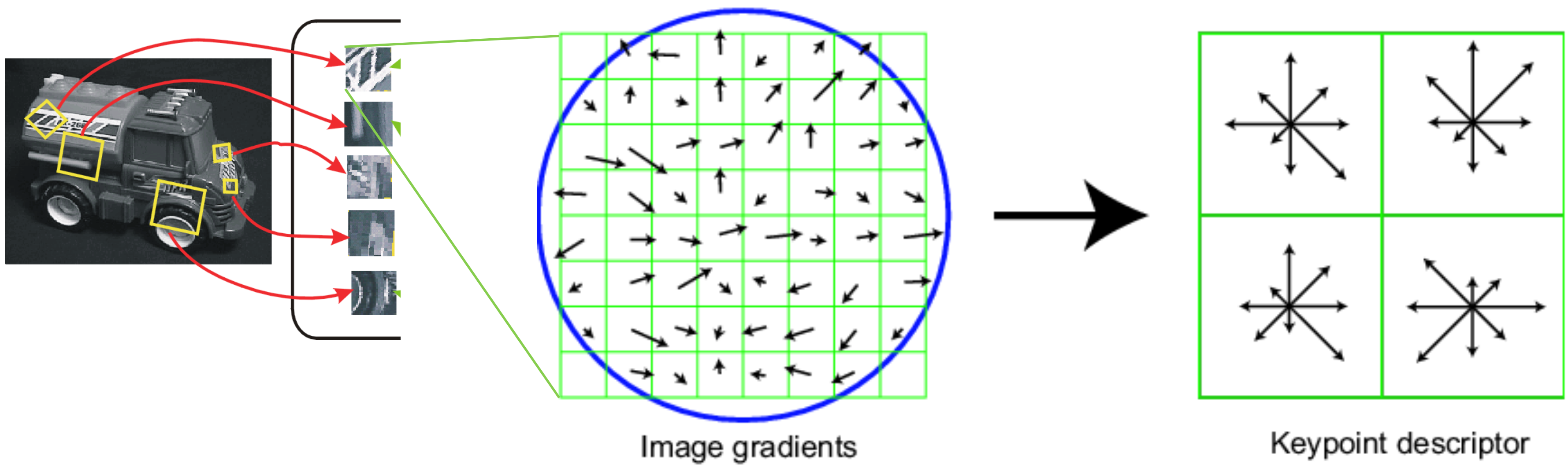
Ensure smoothness

- Trilinear interpolation
  - a given gradient contributes to 8 bins:
    4 in space times 2 in orientation



Reduce effect of illumination

- 128-dim vector normalized to 1
- Threshold gradient magnitudes to avoid excessive influence of high gradients
  - after normalization, clamp gradients >0.2
  - renormalize



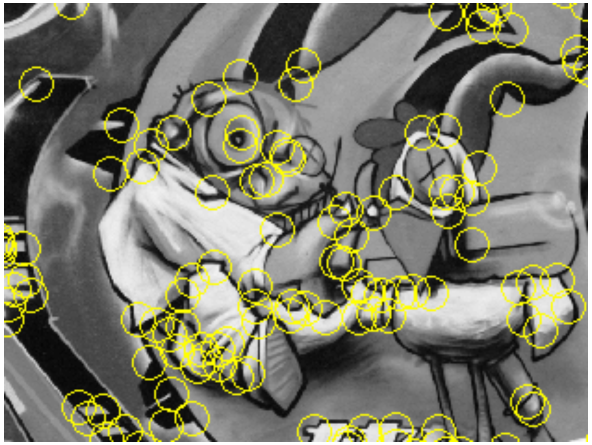
Properties of SIFT

Extraordinarily robust matching technique

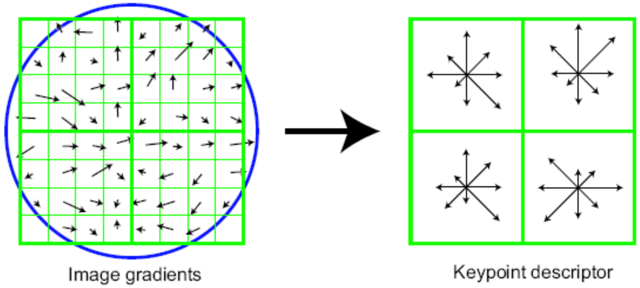
- Can handle changes in viewpoint
  - Up to about 60 degree out of plane rotation
- Can handle significant changes in illumination
  - Sometimes even day vs. night (below)
- Fast and efficient—can run in real time
- Lots of code available:
  http://people.csail.mit.edu/albert/ladypack/wiki/index.php/Known_implementations_of_SIFT

Summary

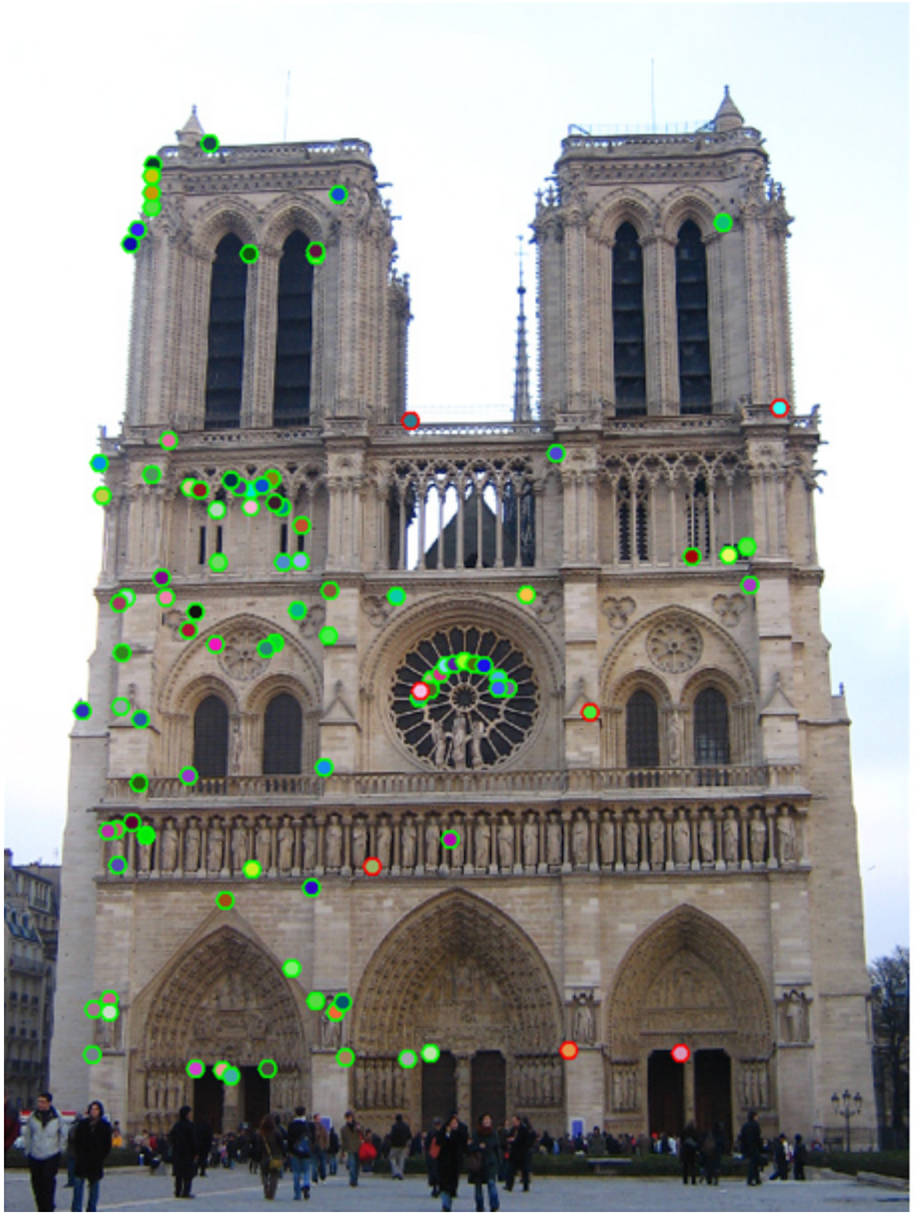
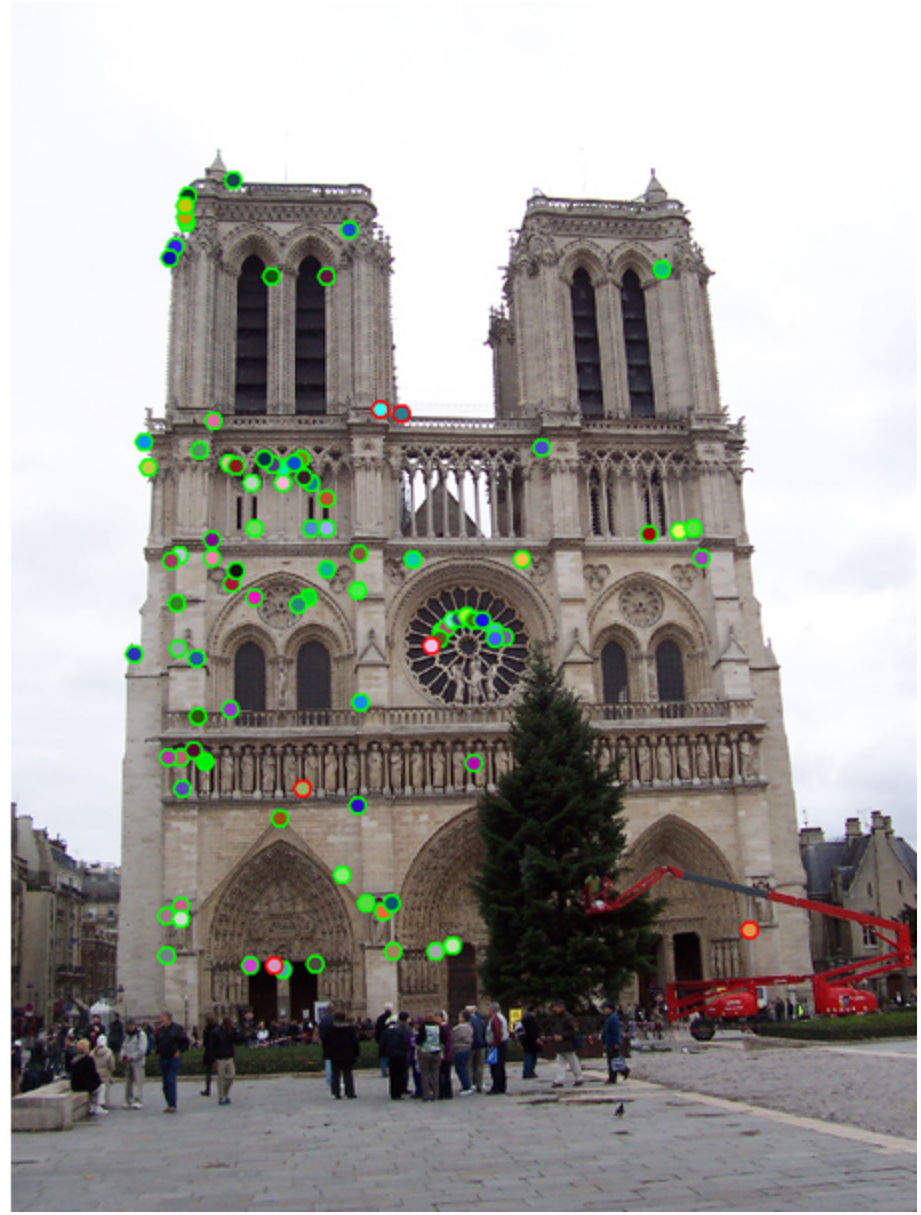
- Keypoint detection: repeatable and distinctive
  - Corners, blobs, stable regions
  - Harris, DoG



- Descriptors: robust and selective
  - spatial histograms of orientation
  - SIFT and variants are typically good for stitching and recognition
  - But, need not stick to one



Which features match?

Feature matching

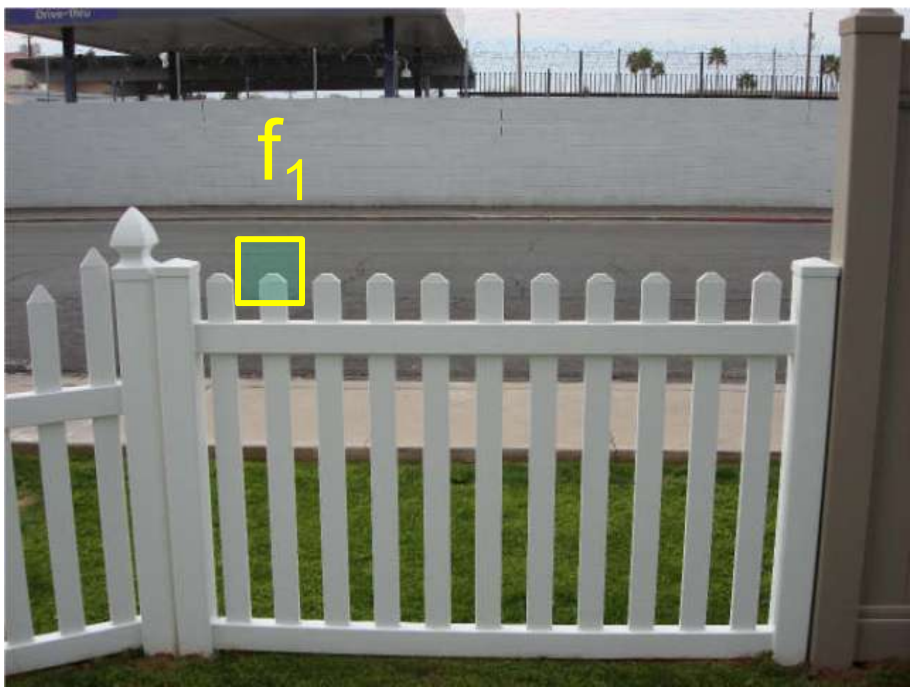
Given a feature in I_1, how to find the best match in I_2?

1. Define distance function that compares two descriptors
2. Test all the features in I_2, find the one with min distance

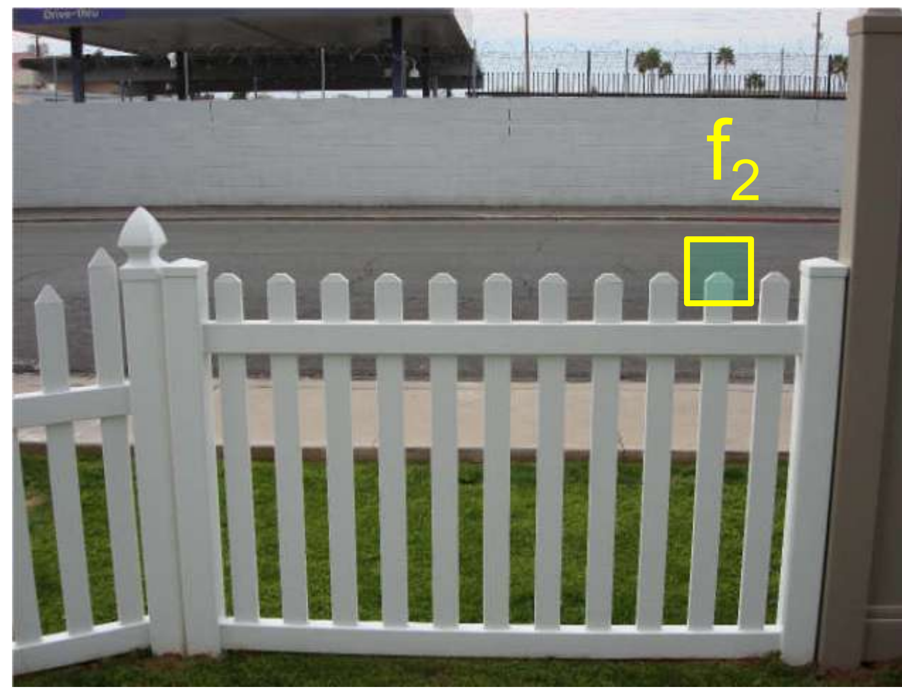
Feature distance

How to define the difference between two features f_1, f_2?

- Simple approach: L_2 distance, $||f_1 - f_2||$
- can give good scores to ambiguous (incorrect) matches



I_1

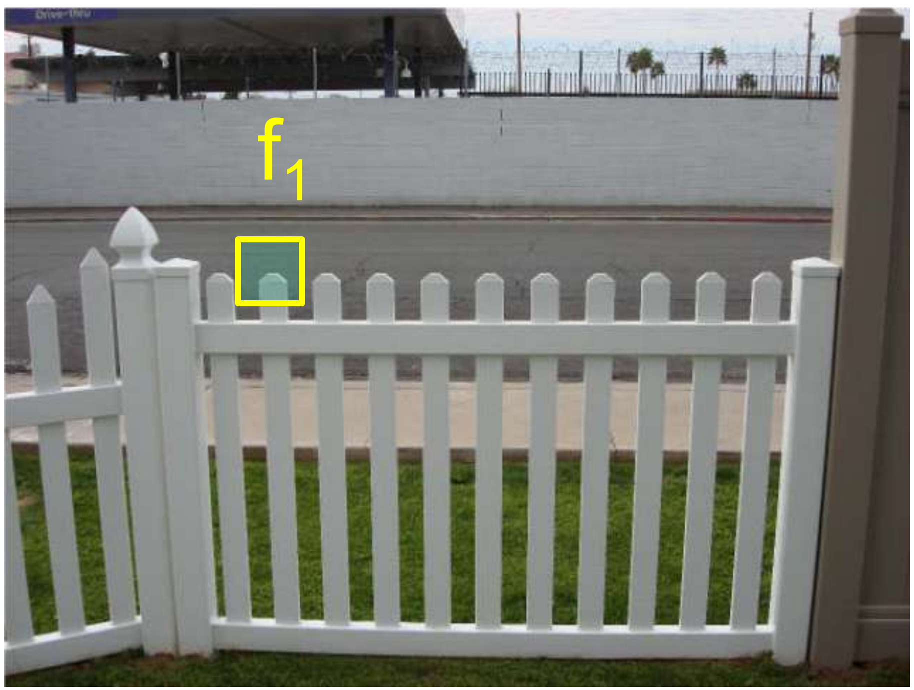


I_2

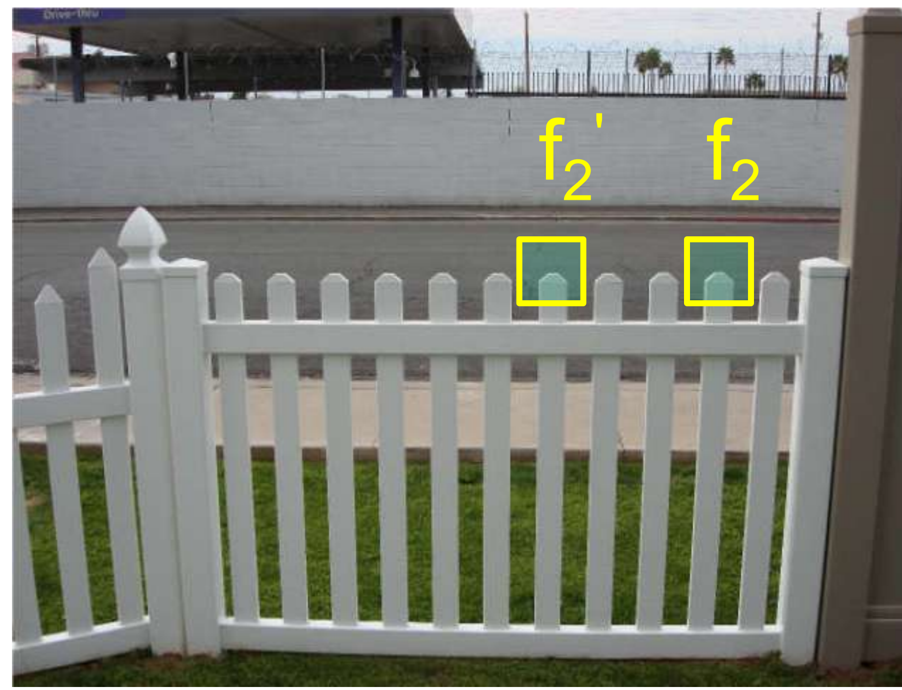
Feature distance

How to define the difference between two features f_1, f_2?

- Better approach: ratio distance = $||f_1 - f_2|| / || f_1 - f_2' ||$
  - f_2 is best SSD match to f_1 in I_2
  - f_2' is 2nd best SSD match to f_1 in I_2
  - gives large values for ambiguous matches



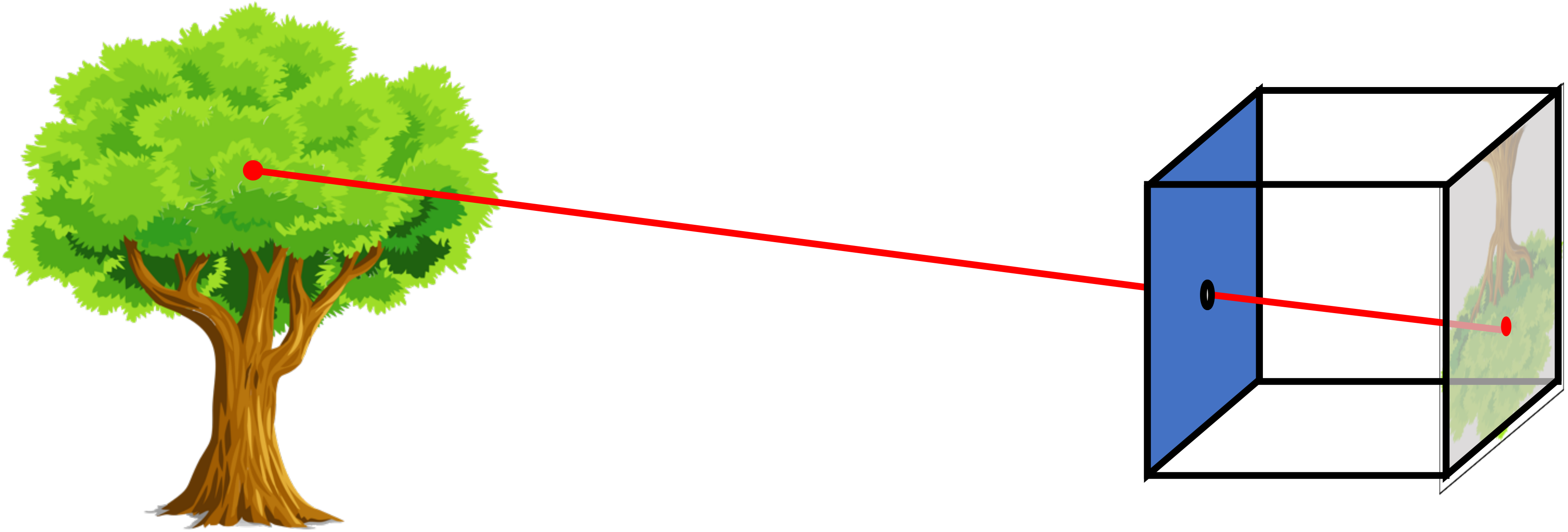
I_1



I_2

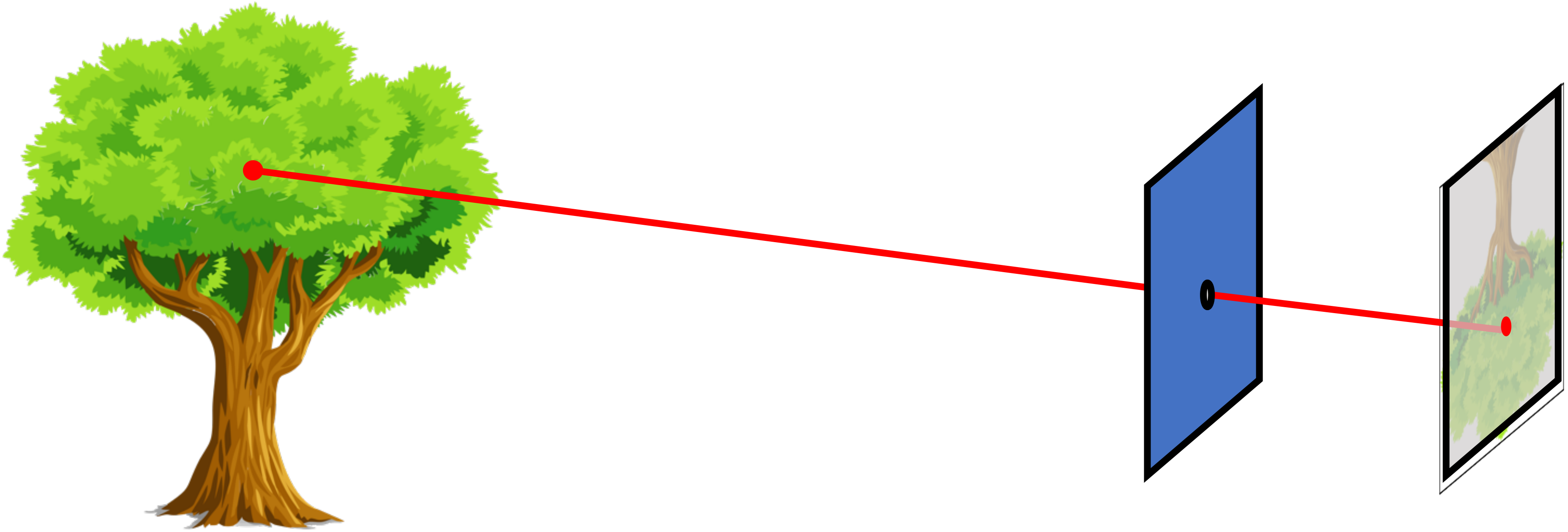
Geometry of Image Formation

The pinhole camera

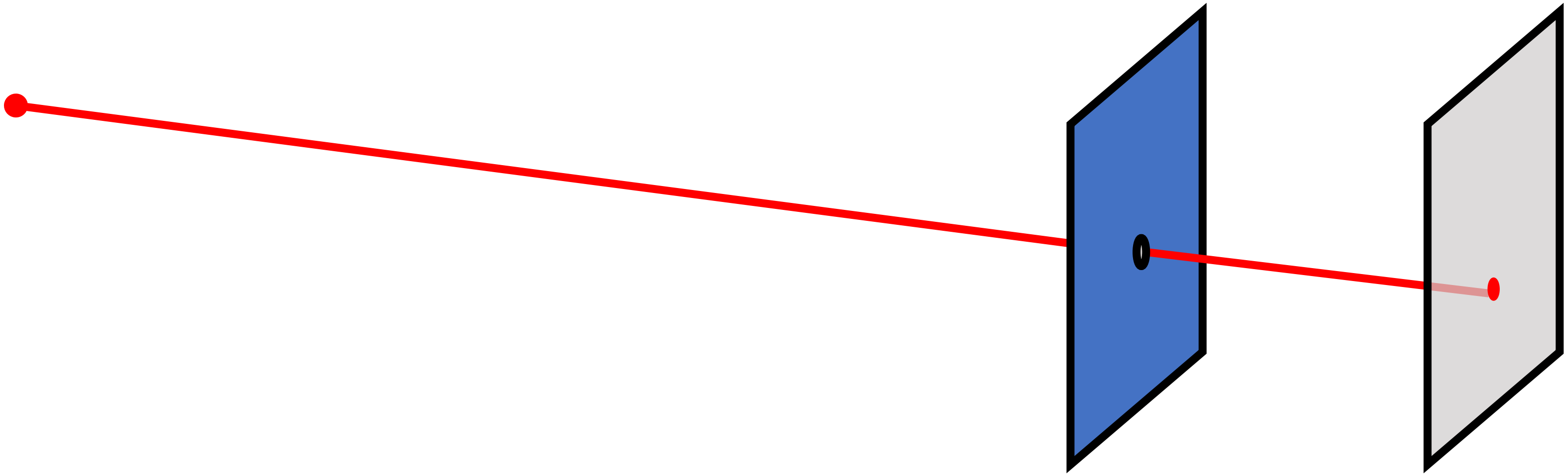


Let's get into the math

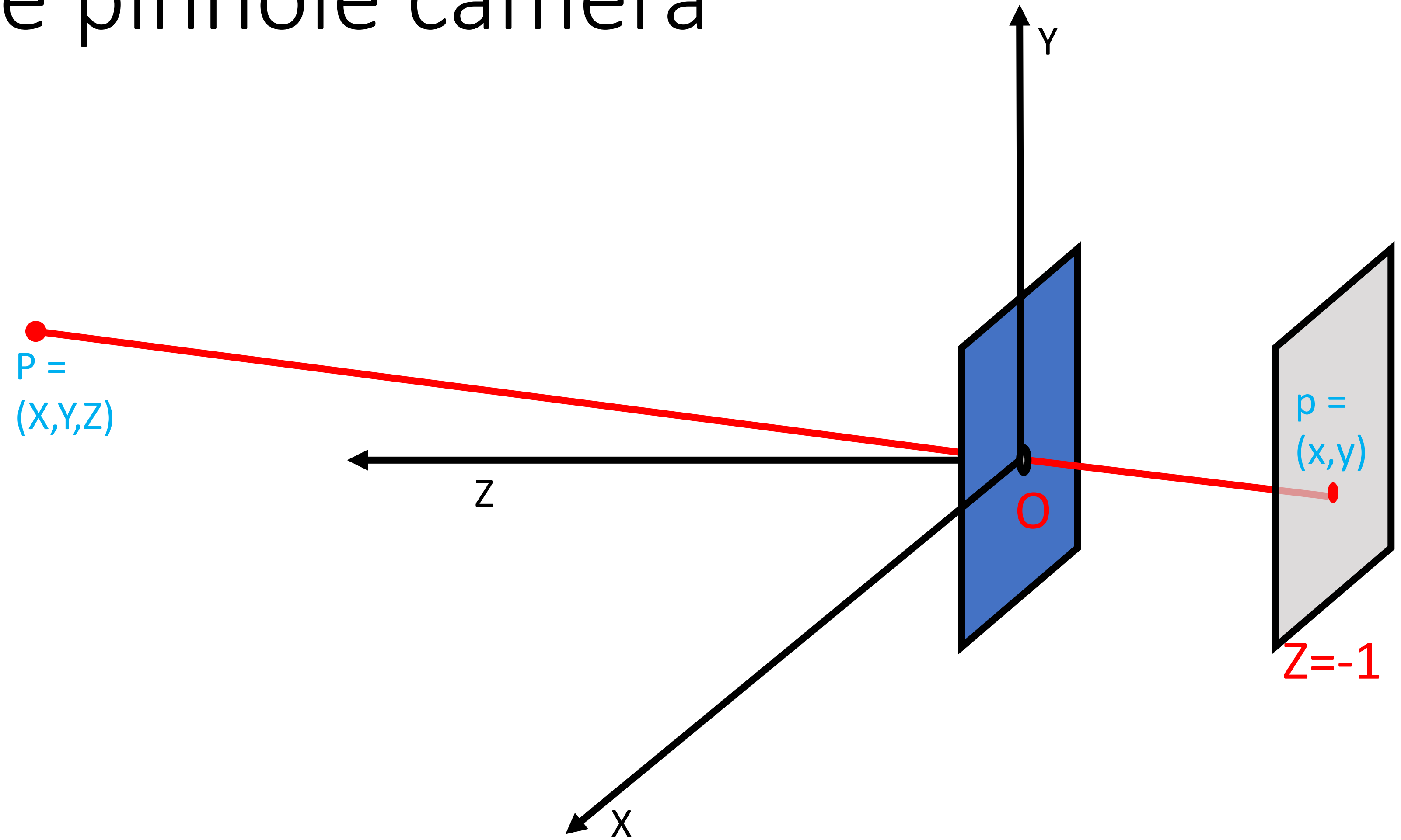
The pinhole camera



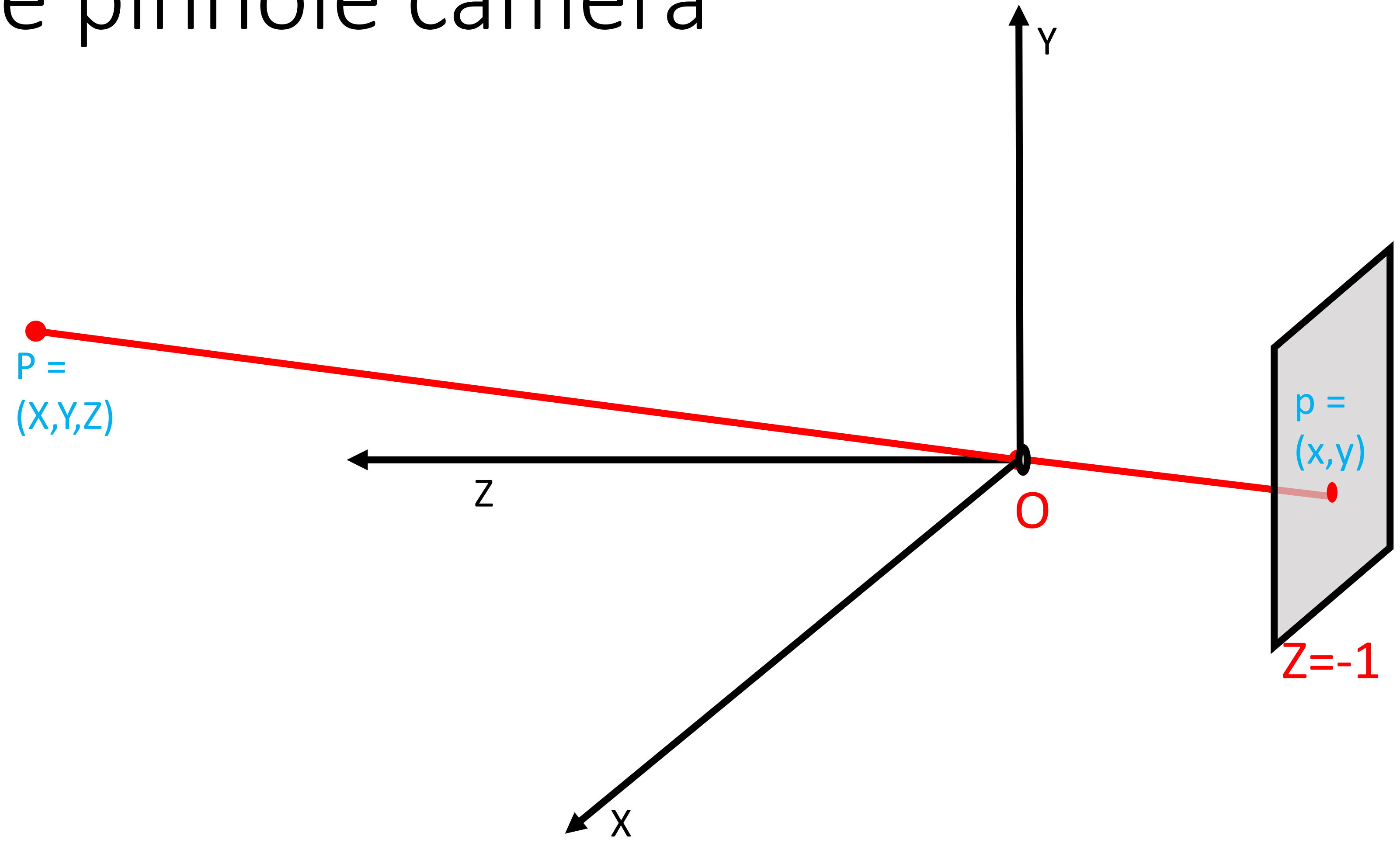
The pinhole camera



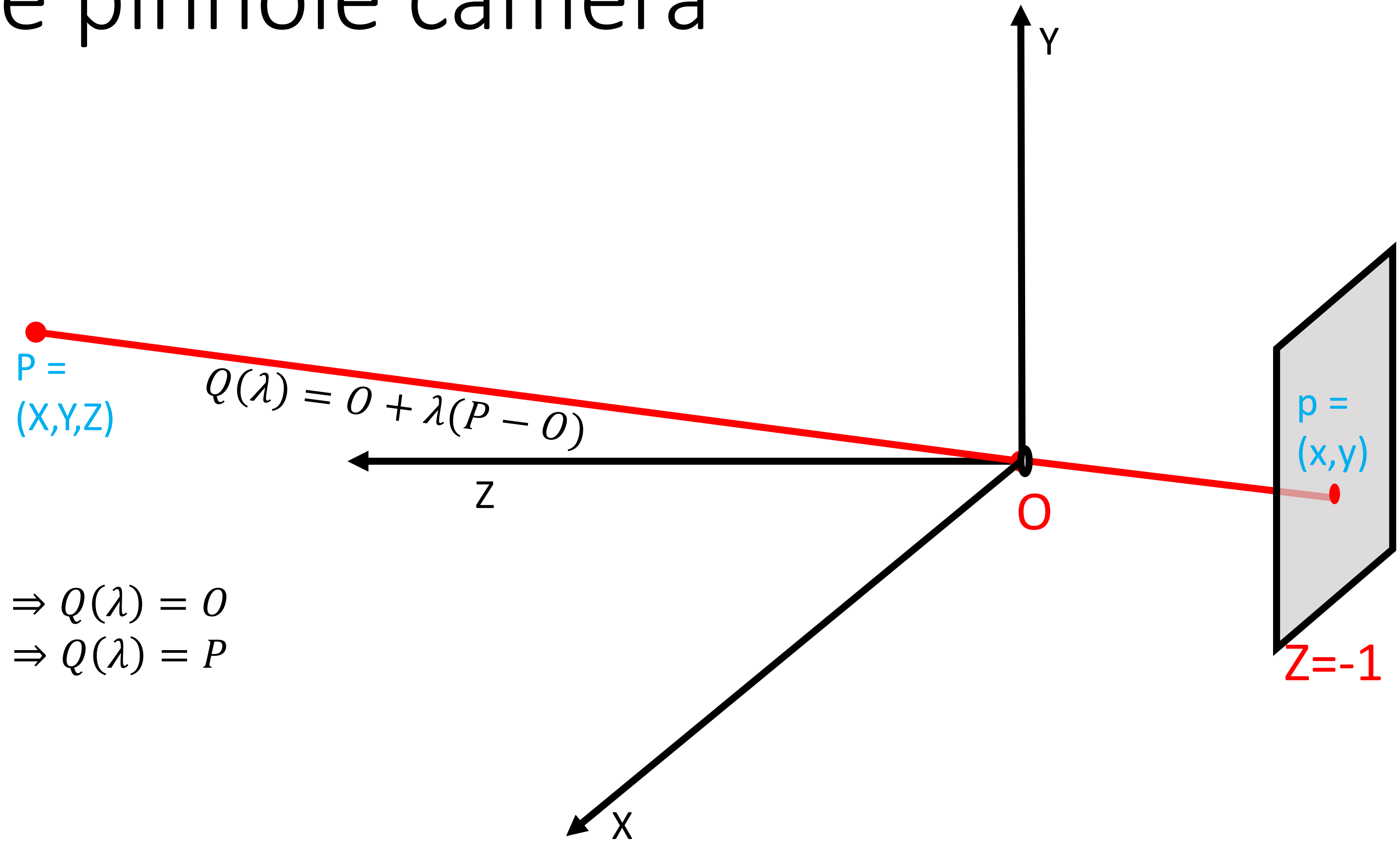
The pinhole camera



The pinhole camera



The pinhole camera



$$\lambda = 0 \Rightarrow Q(\lambda) = O$$
$$\lambda = 1 \Rightarrow Q(\lambda) = P$$

$$\begin{aligned} &Q(\lambda) \\ &= \big(0 + \lambda(X - 0), 0 + \lambda(Y - 0), 0 + \lambda(Z - 0)\big) \\ &= (\lambda X, \lambda Y, \lambda Z) \end{aligned}$$

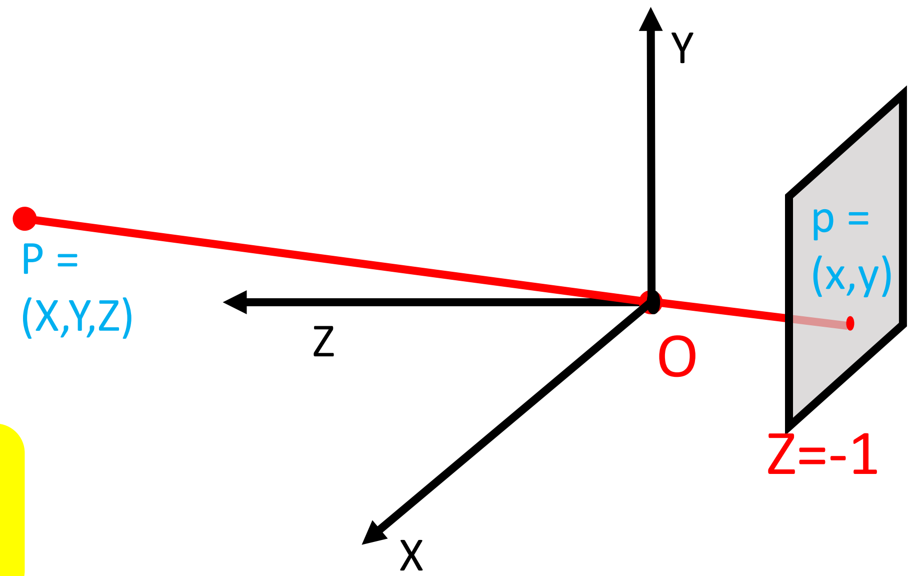
The pinhole camera

- Pinhole camera collapses *ray OP to point p*
- Any point on ray OP = $O + \lambda(P - O) = (\lambda X, \lambda Y, \lambda Z)$
- For this point to lie on Z=-1 plane:

$$\lambda^* Z = -1$$
$$\Rightarrow \lambda^* = \frac{-1}{Z}$$

- Coordinates of point p:

$$(\lambda^* X, \lambda^* Y, \lambda^* Z) = (\frac{-X}{Z}, \frac{-Y}{Z}, -1)$$



The projection equation

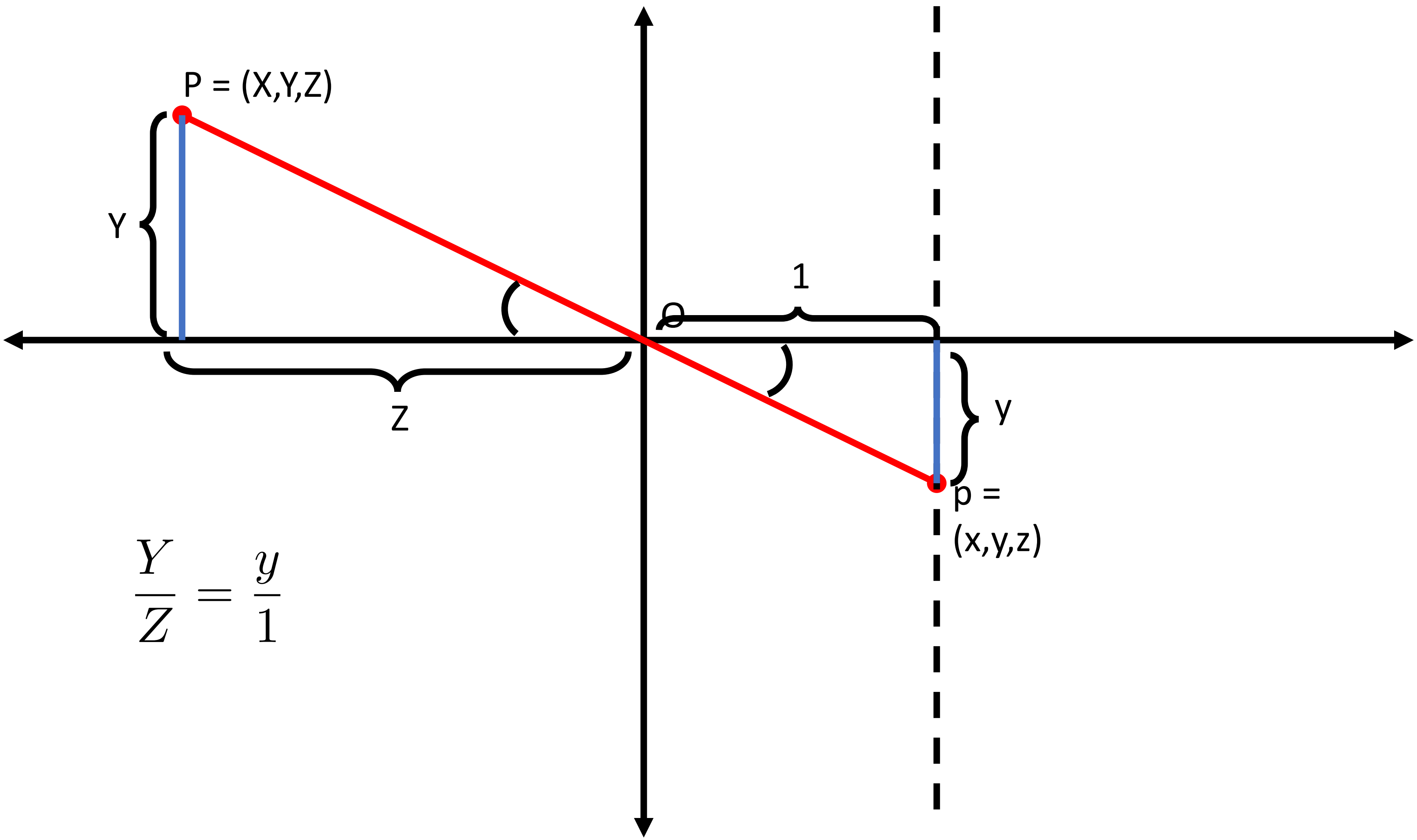
- A point P = (X, Y, Z) in 3D projects to a point p = (x,y) in the image

$$x = \frac{-X}{Z}$$
$$y = \frac{-Y}{Z}$$

- But pinhole camera's image is inverted, invert it back!

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

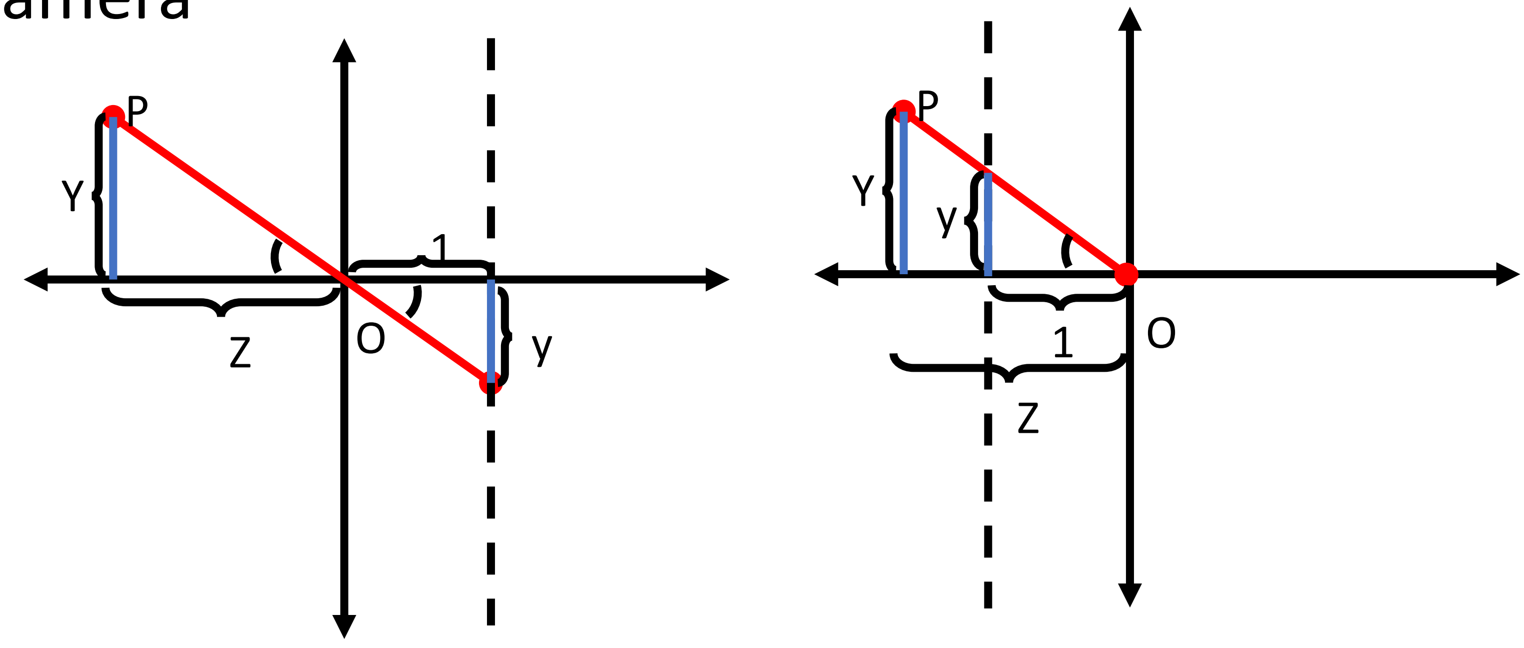
Another derivation



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

A virtual image plane

- A pinhole camera produces an inverted image
- Imagine a ”virtual image plane” in the front of the camera



The projection equation

$$x = \frac{X}{Z}$$

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

Consequence 1: Farther away objects are smaller

Image of foot: $(\frac{X}{Z}, \frac{Y}{Z})$

Image of head: $(\frac{X}{Z}, \frac{Y+h}{Z})$

$$\frac{Y+h}{Z} - \frac{Y}{Z} = \frac{h}{Z}$$

Consequence 2: Parallel lines converge at a point

- Point on a line passing through point A with direction D:
  $$Q(\lambda) = A + \lambda D$$
- Parallel lines have the same direction but pass through different points
  $$Q(\lambda) = A + \lambda D$$
  $$R(\lambda) = B + \lambda D$$
-

Consequence 2: Parallel lines converge at a point

- Parallel lines have the same direction but pass through different points

$$Q(\lambda) = A + \lambda D$$
$$R(\lambda) = B + \lambda D$$

- $A = (A_X, A_Y, A_Z)$
- $B = (B_X, B_Y, B_Z)$
- $D = (D_X, D_Y, D_Z)$

Consequence 2: Parallel lines converge at a point

- $Q(\lambda) = (A_X + \lambda D_X, A_Y + \lambda D_Y, A_Z + \lambda D_Z)$
- $R(\lambda) = (B_X + \lambda D_X, B_Y + \lambda D_Y, B_Z + \lambda D_Z)$
- $q(\lambda) = \left(\frac{A_X+\lambda D_X}{A_Z+\lambda D_Z}, \frac{A_Y+\lambda D_Y}{A_Z+\lambda D_Z}\right)$
- $r(\lambda) = \left(\frac{B_X+\lambda D_X}{B_Z+\lambda D_Z}, \frac{B_Y+\lambda D_Y}{B_Z+\lambda D_Z}\right)$
- Need to look at these points as Z goes to infinity
- Same as $\lambda \to \infty$

Consequence 2: Parallel lines converge at a point

- $q(\lambda) = \left(\frac{A_X+\lambda D_X}{A_Z+\lambda D_Z}, \frac{A_Y+\lambda D_Y}{A_Z+\lambda D_Z}\right)$
- $r(\lambda) = \left(\frac{B_X+\lambda D_X}{B_Z+\lambda D_Z}, \frac{B_Y+\lambda D_Y}{B_Z+\lambda D_Z}\right)$

$$\lim_{\lambda\to\infty} \frac{A_X + \lambda D_X}{A_Z + \lambda D_Z} = \lim_{\lambda\to\infty} \frac{\frac{A_X}{\lambda} + D_X}{\frac{A_Z}{\lambda} + D_Z} = \frac{D_X}{D_Z}$$

$$\lim_{\lambda\to\infty} q(\lambda) = (\frac{D_X}{D_Z}, \frac{D_Y}{D_Z})$$

$$\lim_{\lambda\to\infty} r(\lambda) = (\frac{D_X}{D_Z}, \frac{D_Y}{D_Z})$$

Consequence 2: Parallel lines converge at a point

- Parallel lines have the same direction but pass through different points

$$Q(\lambda) = A + \lambda D$$
$$R(\lambda) = B + \lambda D$$

- Parallel lines converge at the same point $(\frac{D_X}{D_Z}, \frac{D_Y}{D_Z})$
- This point of convergence is called the *vanishing point*
- What happens if $D_Z = 0$?

Consequence 2: Parallel lines converge at a point

What about planes?

$$N_X X + N_Y Y + N_Z Z = d$$

$$\Rightarrow N_X \frac{X}{Z} + N_Y \frac{Y}{Z} + N_Z = \frac{d}{Z}$$

$$\Rightarrow N_X x + N_Y y + N_Z = \frac{d}{Z}$$

Take the limit as Z approaches infinity

$$N_X x + N_Y y + N_Z = 0$$

Vanishing line of a plane

What about planes?

$$N_X X + N_Y Y + N_Z Z = d$$

Normal: (N_X, N_Y, N_Z)

What do parallel planes look like?

$$N_X X + N_Y Y + N_Z Z = d$$

$$N_X x + N_Y y + N_Z = 0$$

$$N_X X + N_Y Y + N_Z Z = c$$

$$N_X x + N_Y y + N_Z = 0$$

Vanishing lines

Parallel planes converge!

Vanishing line

$$N_X X + N_Y Y + N_Z Z = d$$

- What happens if $N_X = N_Y = 0$?
- Equation of the plane: $Z = c$
- Vanishing line?