Introduction to Computer Vision

Michael J. Black

Nov 2009

Perspective projection and affine motion

BROWN

COURSE OFFERING

SEMESTER II, 2009-2010

PROFESSOR THOMAS SERRE

COGS1280 Computational Cognitive Science

A detailed introduction to computational modeling of cognition, summarizing traditional approaches and providing experience with state-of-the-art methods. Covers pattern recognition approaches, shallow and hierarchical networks including Bayesian probabilistic models, and illustrates how they have been applied in several key areas in cognitive science, including visual perception and attention, object and face recognition, learning and memory as well as decision-making and reasoning. Focuses on modeling simple laboratory tasks from cognitive psychology. Connections to contemporary research in computer science will be emphasized highlighting how computational models may motivate the development of new hypothesis for experiment design in cognitive psychology.

Tues., Thurs. 2:30 - 3:50

Goals

- Today
  - Perspective projection
  - 3D motion
- Wed
  - Projects
- Friday
  - Regularization and robust statistics

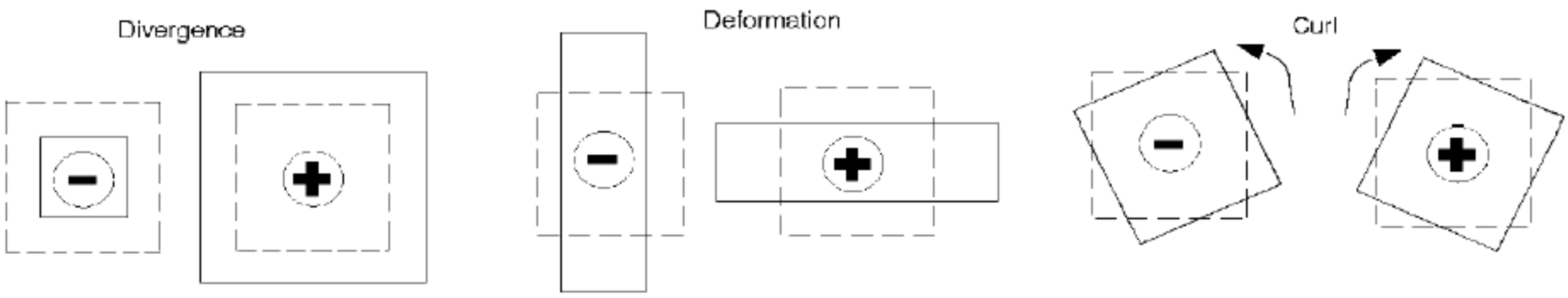
Reading

- Szeliski
  - 2.2.2: 2D tranformations including affine
  - 2.1.3: 3D transformations including rotation matrices (more detailed than we need)
  - 2.1.4: 3D to 2D projections

Affine Flow

$$E(\mathbf{a}) = \sum_{x,y \in R} (\nabla I^T \mathbf{u}(\mathbf{x};\mathbf{a}) + I_t)^2$$

$$\mathbf{u}(\mathbf{x};\ \mathbf{a}) = \begin{bmatrix} u(\mathbf{x};\mathbf{a}) \\ v(\mathbf{x};\mathbf{a}) \end{bmatrix} = \begin{bmatrix} a_1 + a_2 x + a_3 y \\ a_4 + a_5 x + a_6 y \end{bmatrix}$$

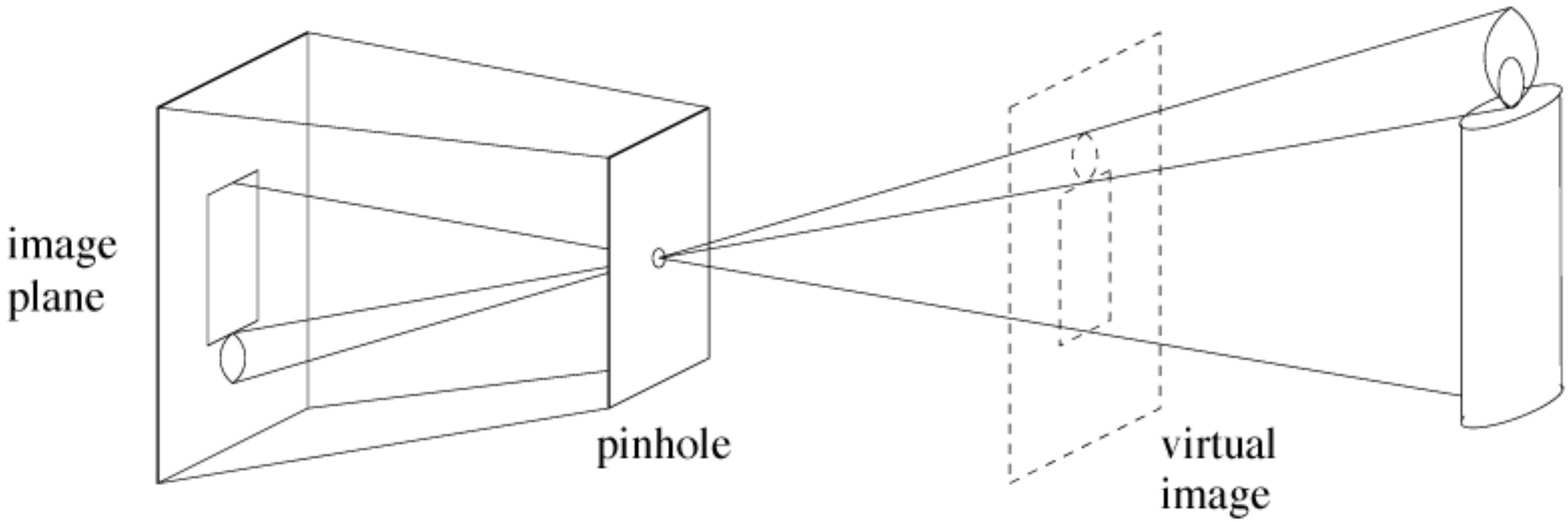


Why Affine?

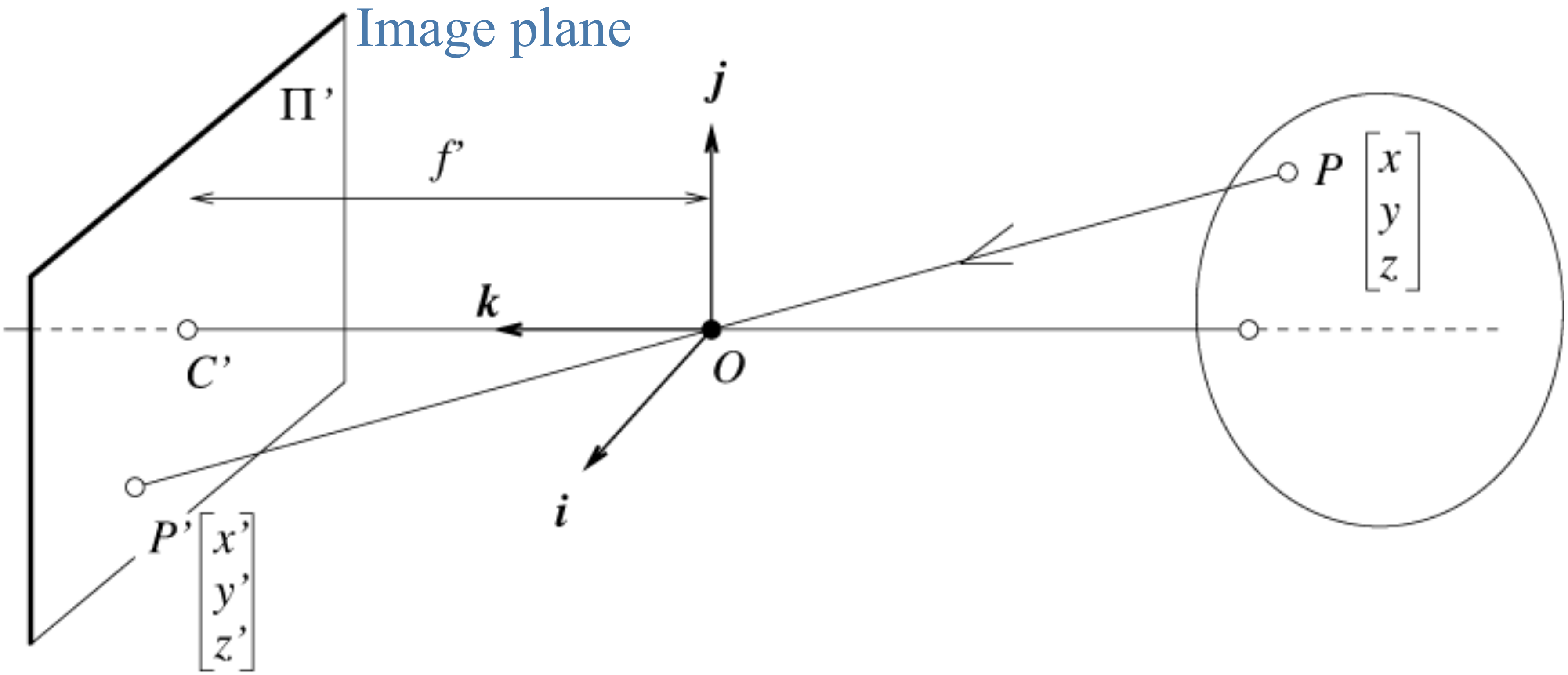
- Where does this affine approximation come from?
- All our models are approximations to the world. What are the assumptions in the affine approximation?
- For this we need some geometry.

Pinhole cameras

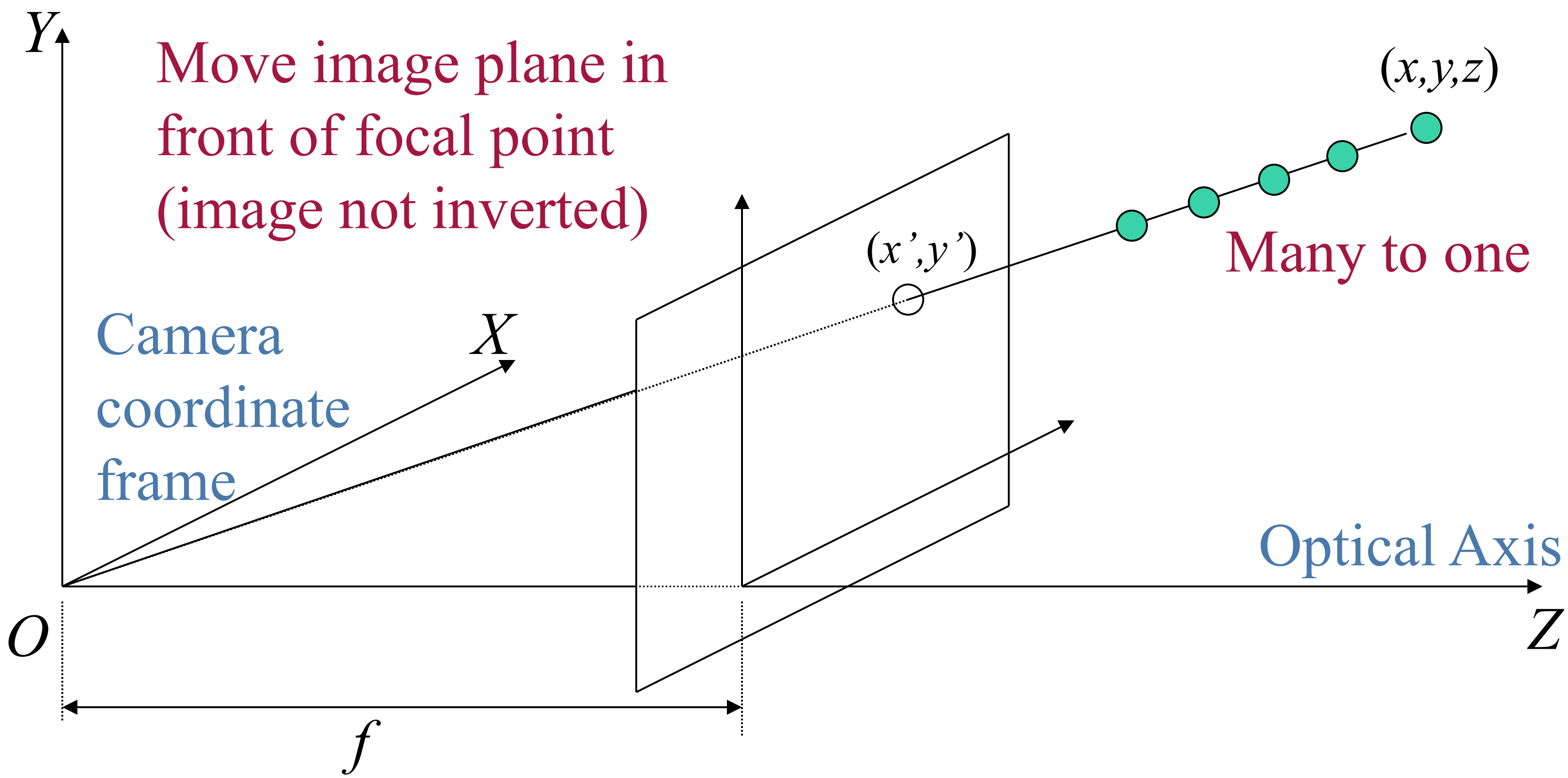
- Abstract camera model - box with a small hole in it.
- Easy to build but needs a lot of light.



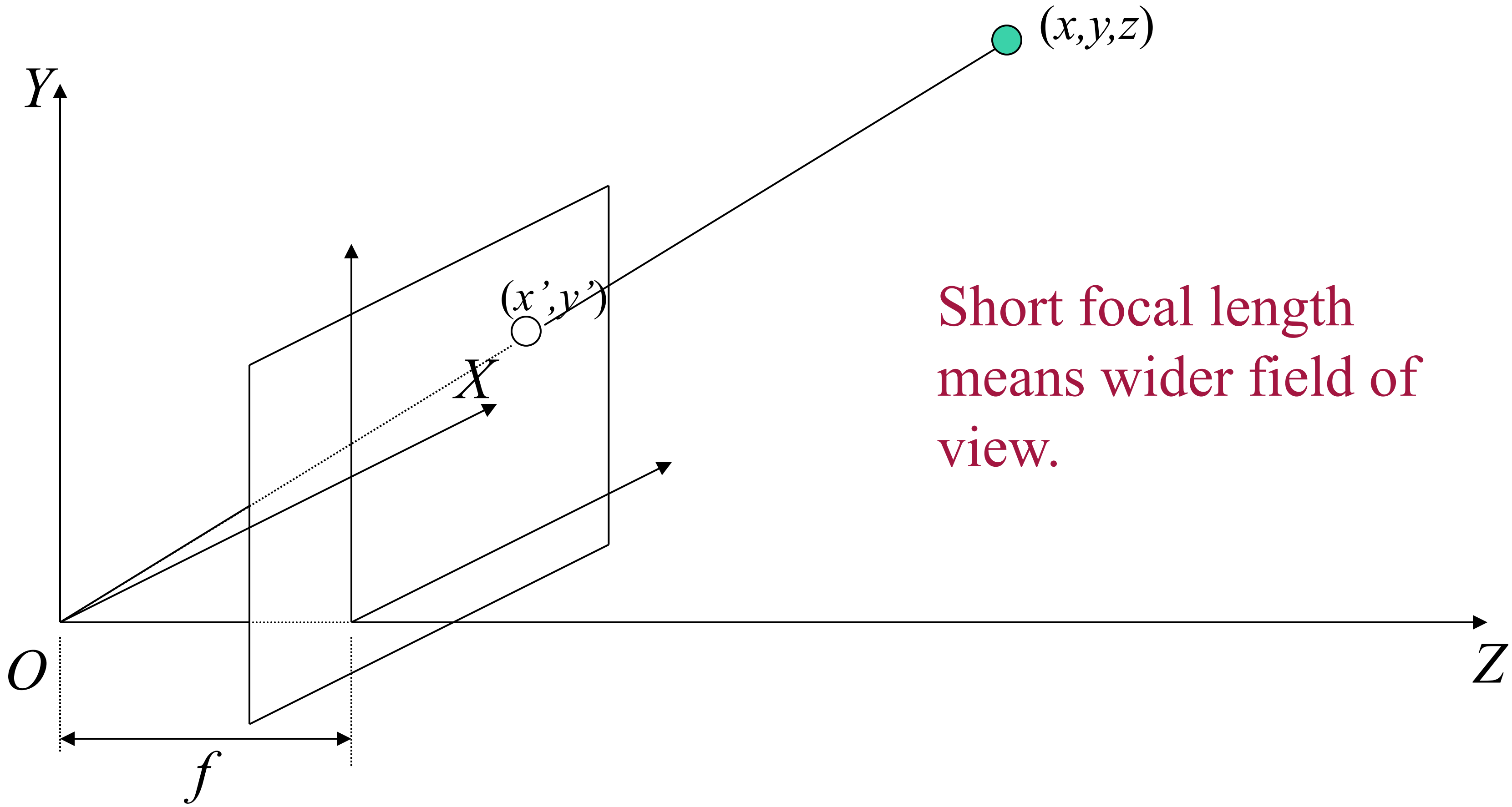
The equation of projection



Perspective Projection

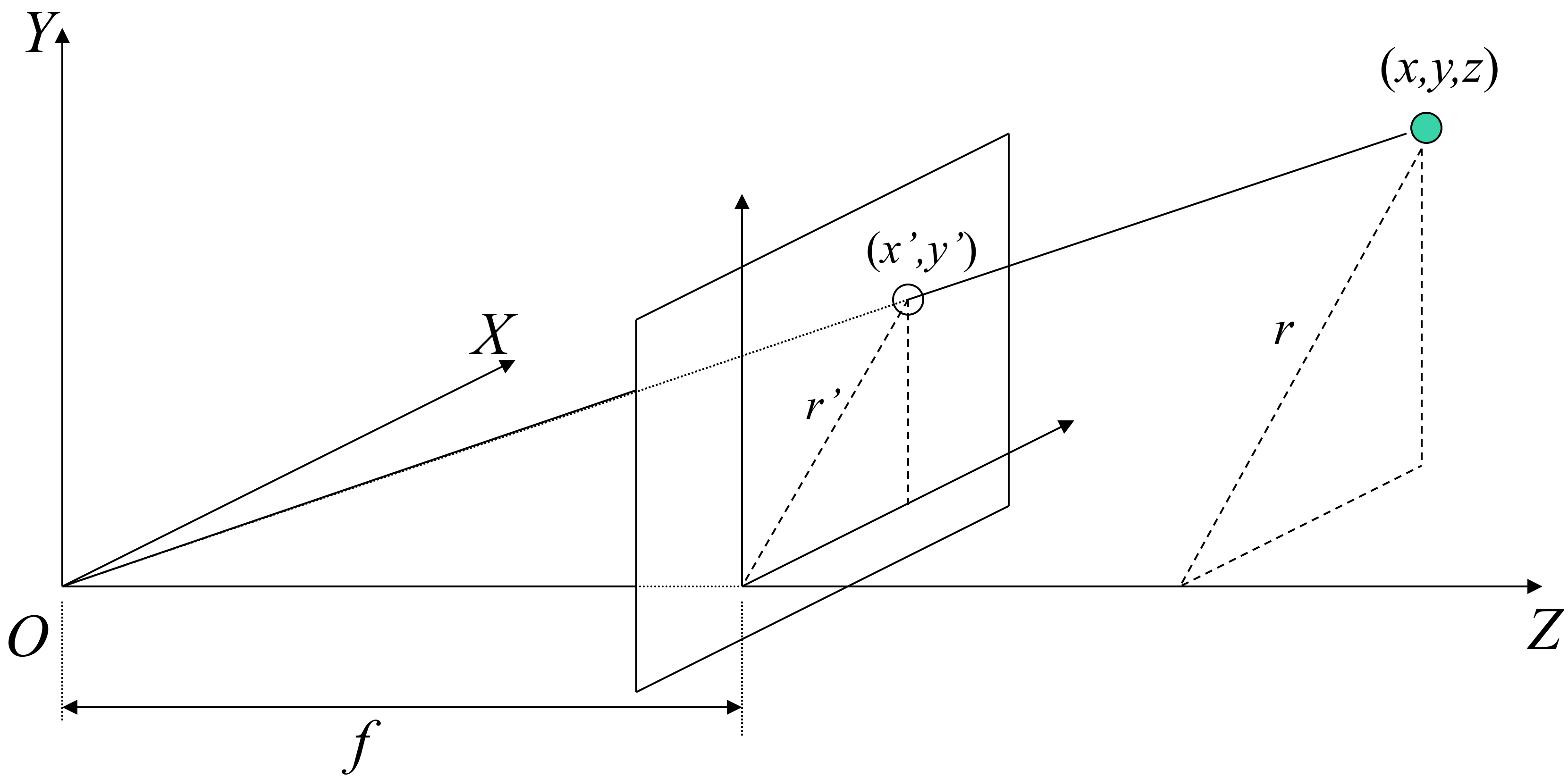


Perspective Projection



Short focal length means wider field of view.

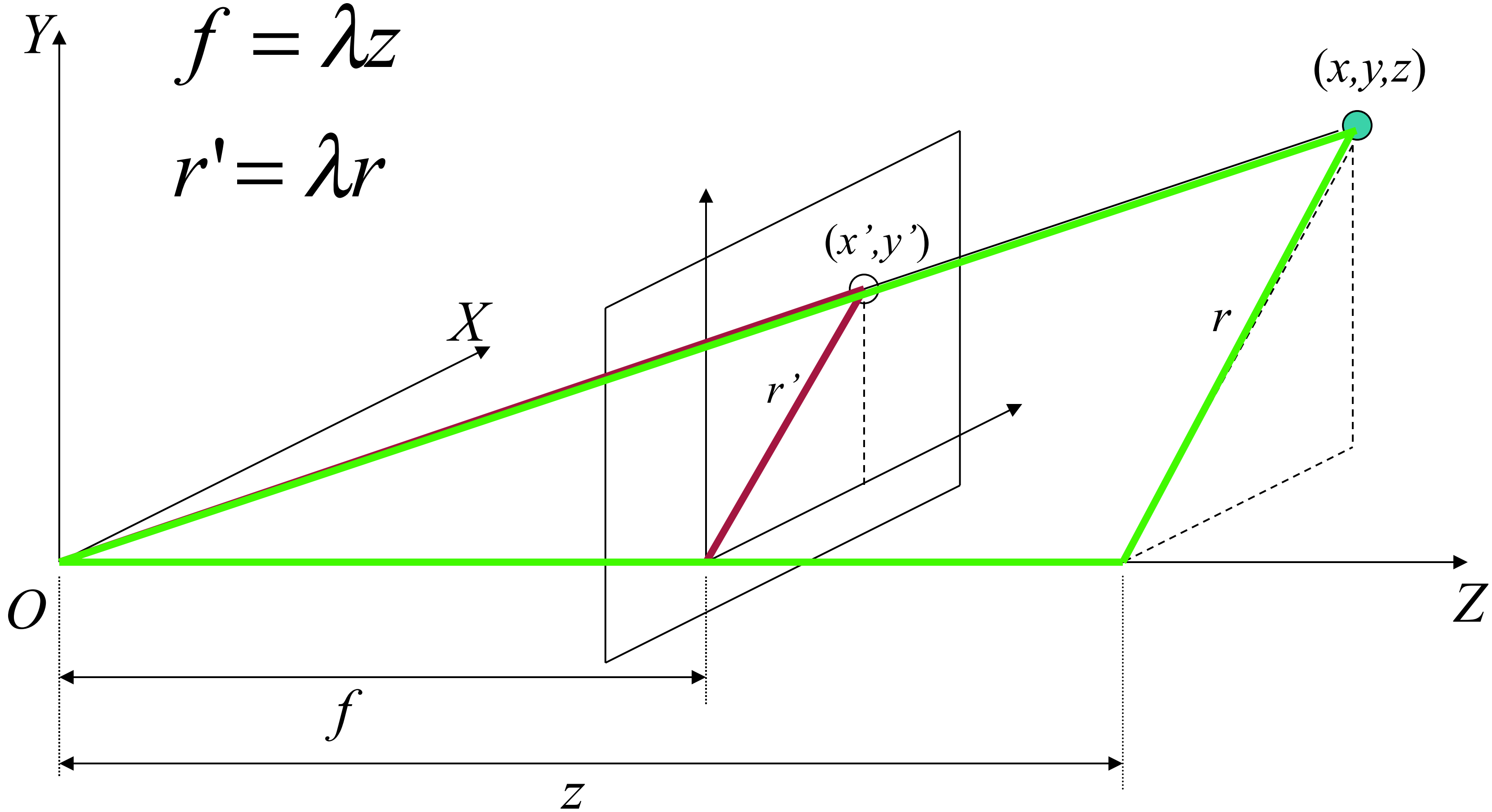
Perspective Projection



Perspective Projection

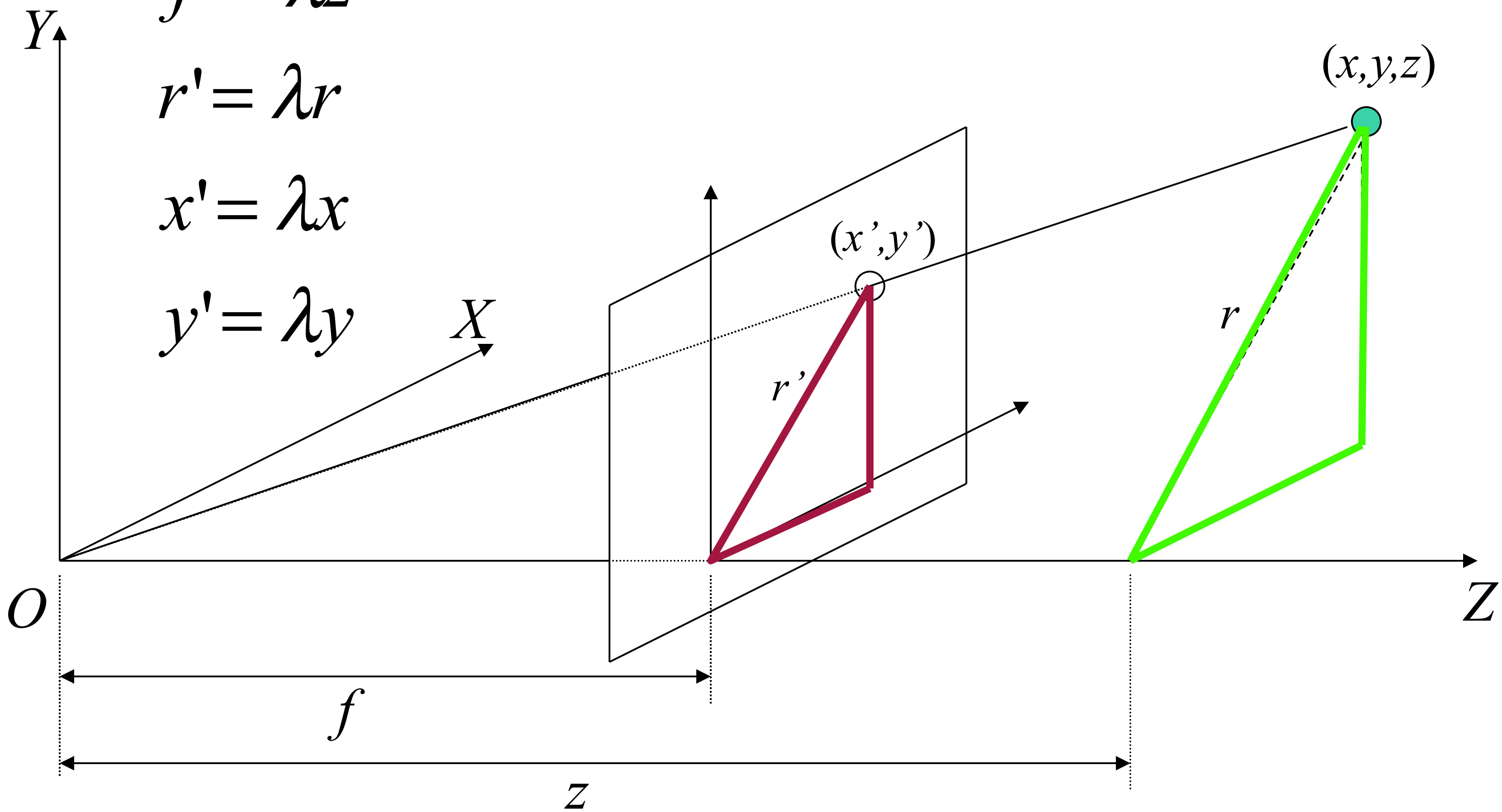
$$f = \lambda z$$

$$r' = \lambda r$$



Perspective Projection

$$f = \lambda z$$
$$r' = \lambda r$$
$$x' = \lambda x$$
$$y' = \lambda y$$



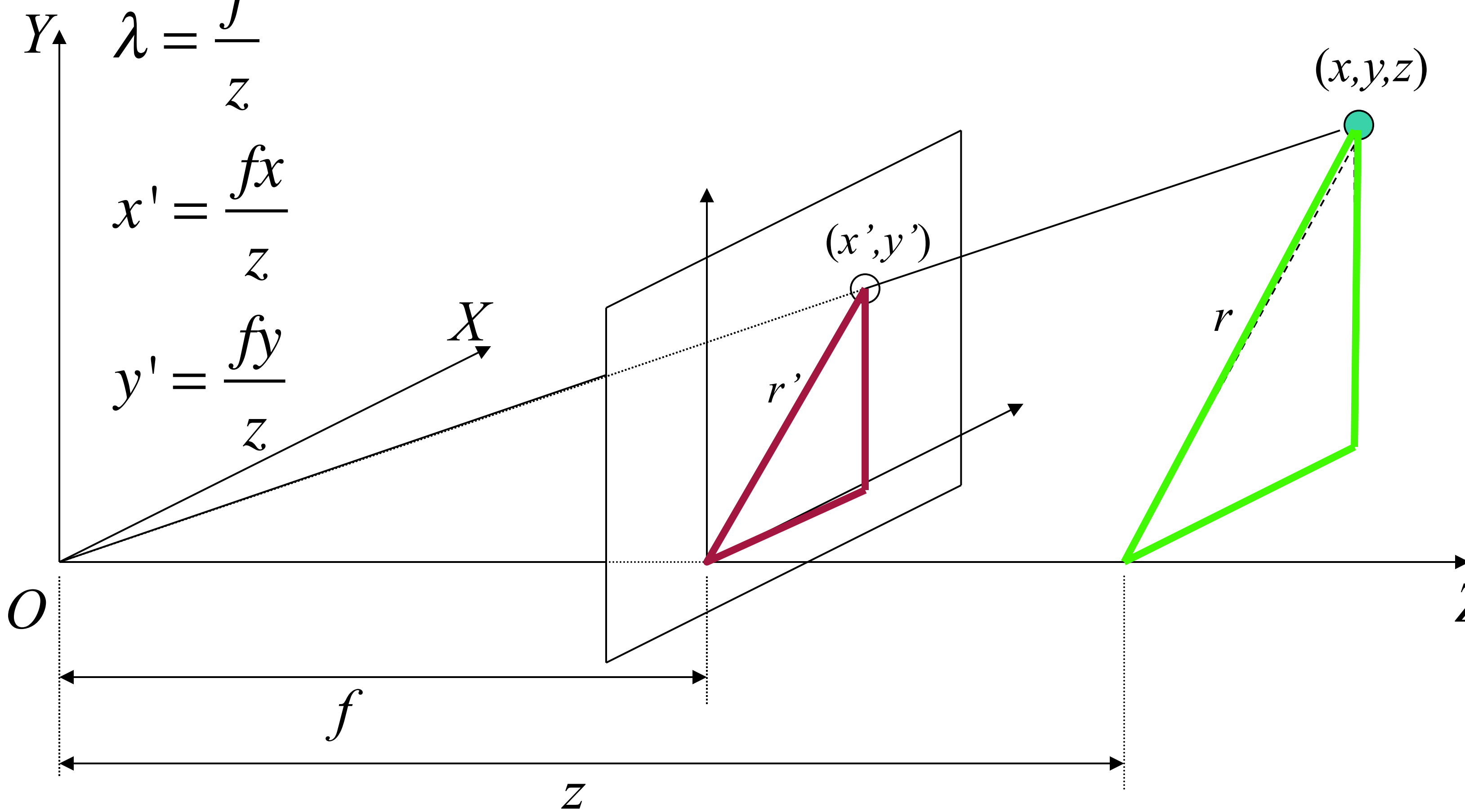
(important slide)

Perspective Projection

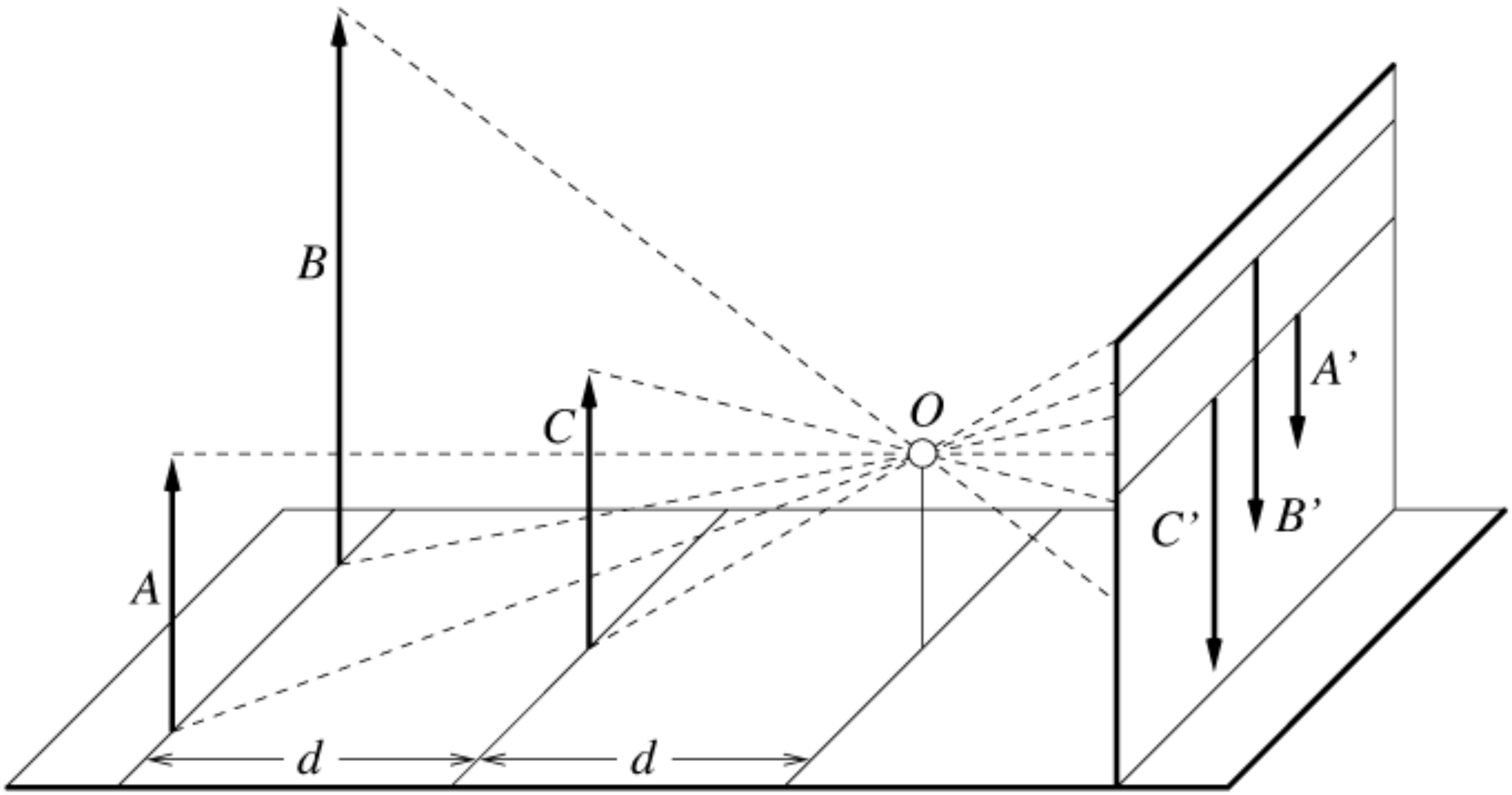
$$\lambda = \frac{f}{z}$$

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

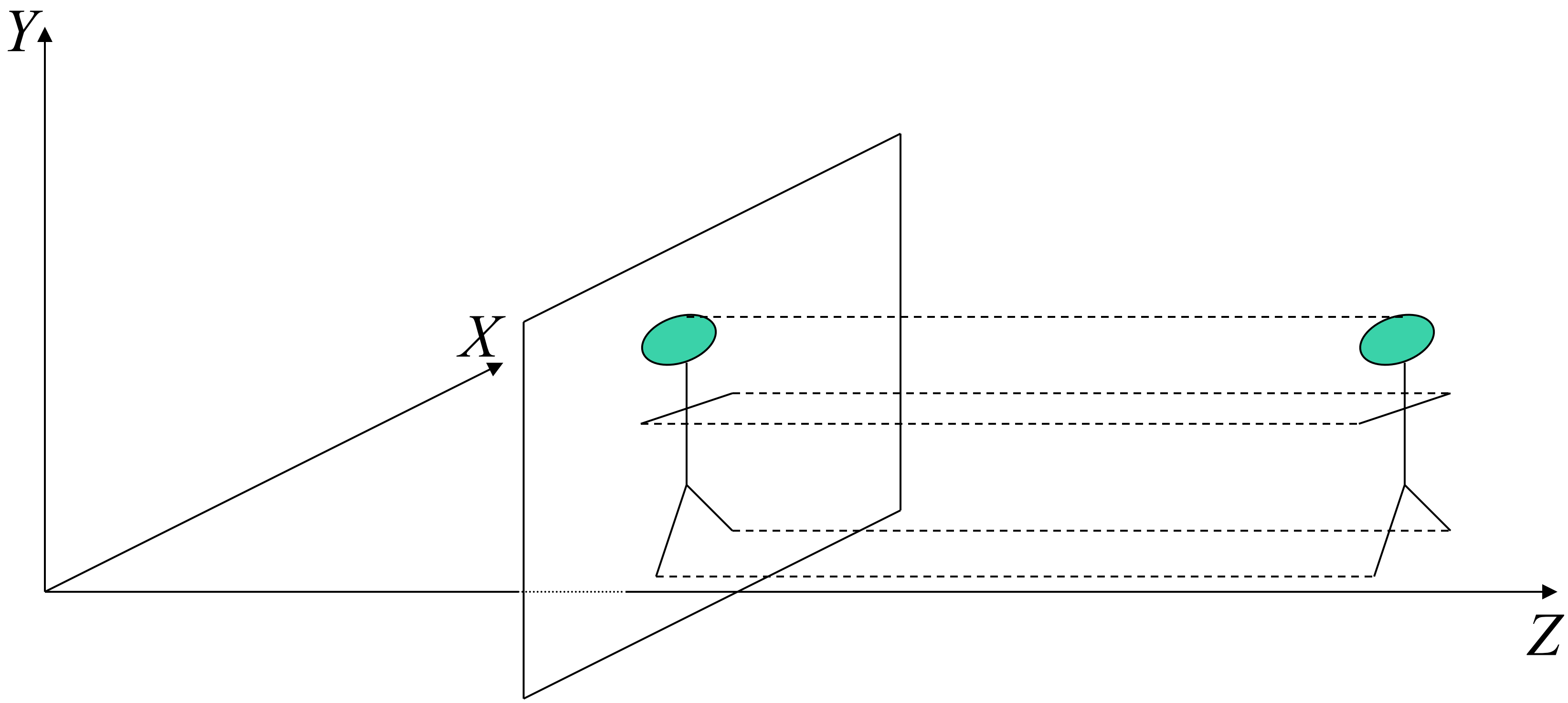
$$y' = \frac{fy}{z}$$



Distant objects are smaller



Orthographic Projection

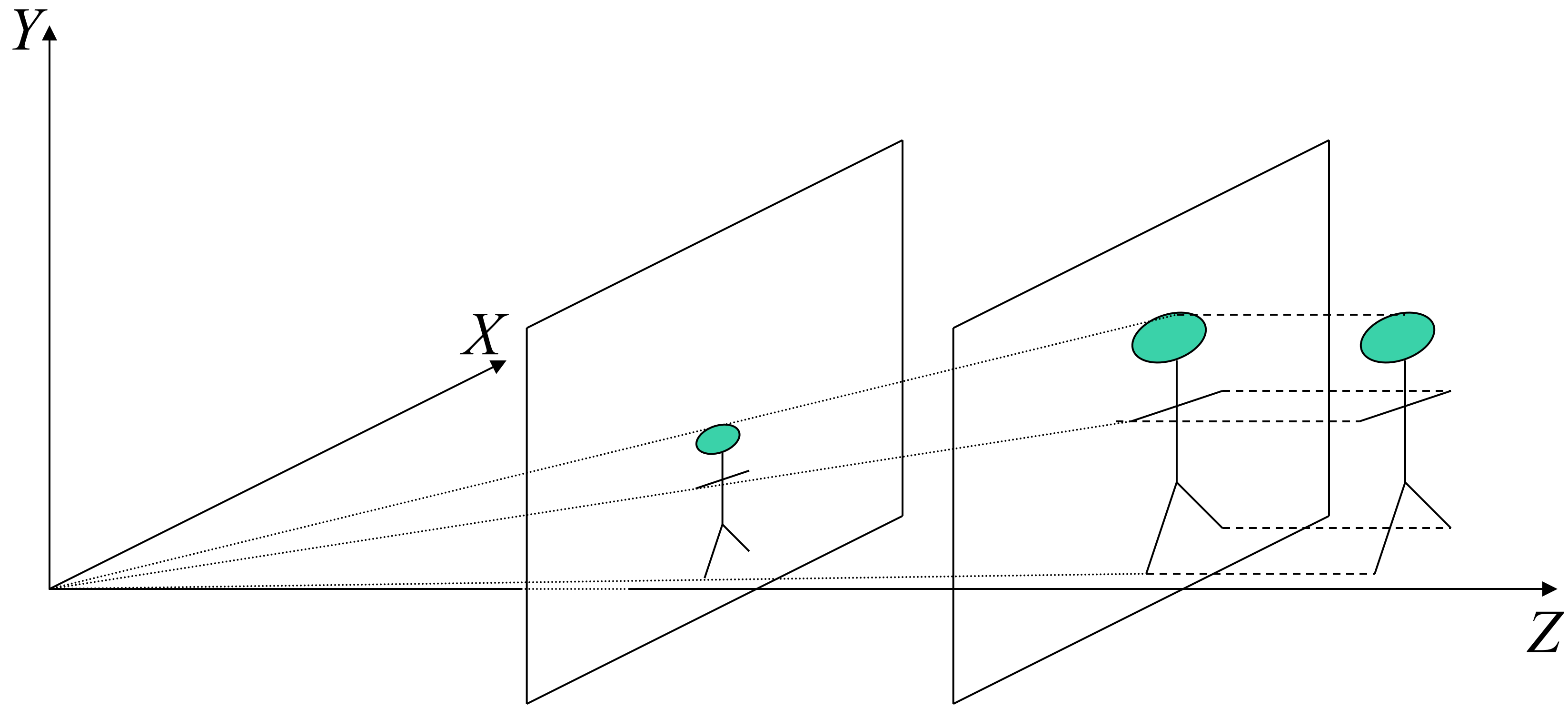


Assume an infinite focal length and that the world is infinitely far away.

$$(x, y, z) \rightarrow (x, y)$$

$$x' = x, y' = y$$

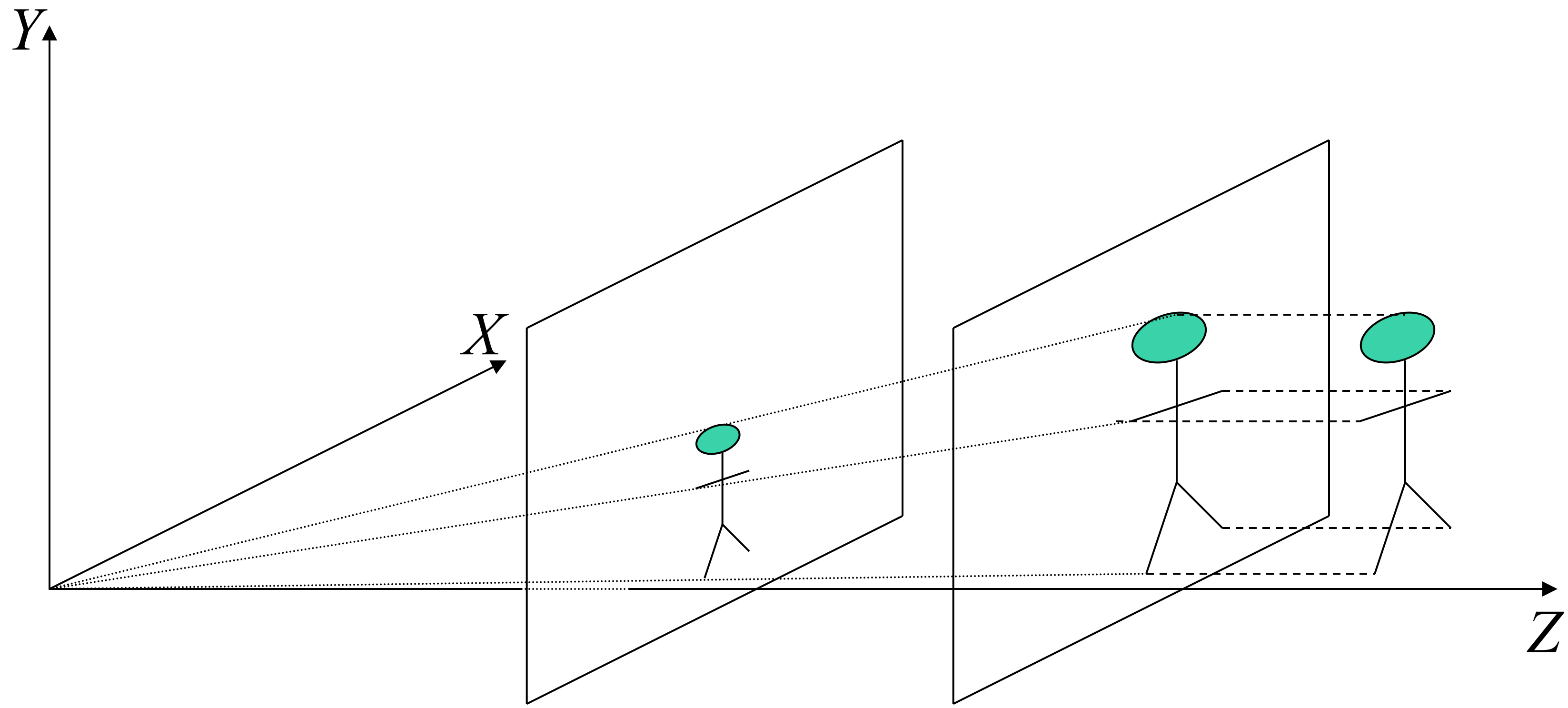
Weak Perspective (scaled orthography)



Assume variation in depth is small relative to the distance from the camera.

Approximate scene as a fronto-parallel plane

Weak Perspective (scaled orthography)



$$x' = sx$$

$$y' = sy$$

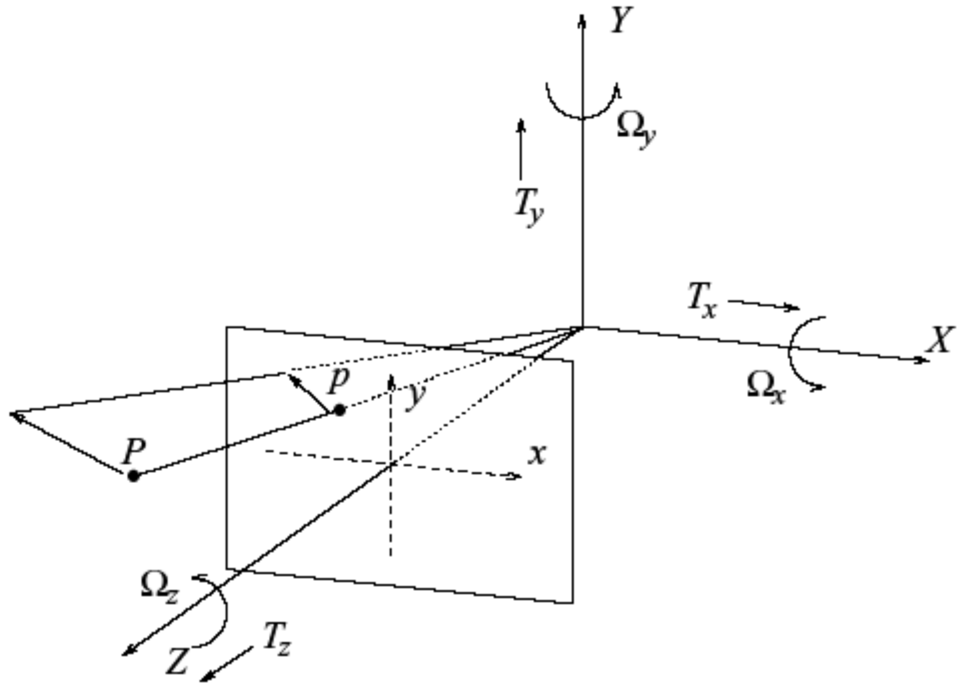
Claim

For small motions, affine flow approximates the motion of a plane viewed under orthographic projection.

$$u = v_x = a_1 x + a_2 y + a_3$$
$$v = v_y = a_4 x + a_5 y + a_6$$

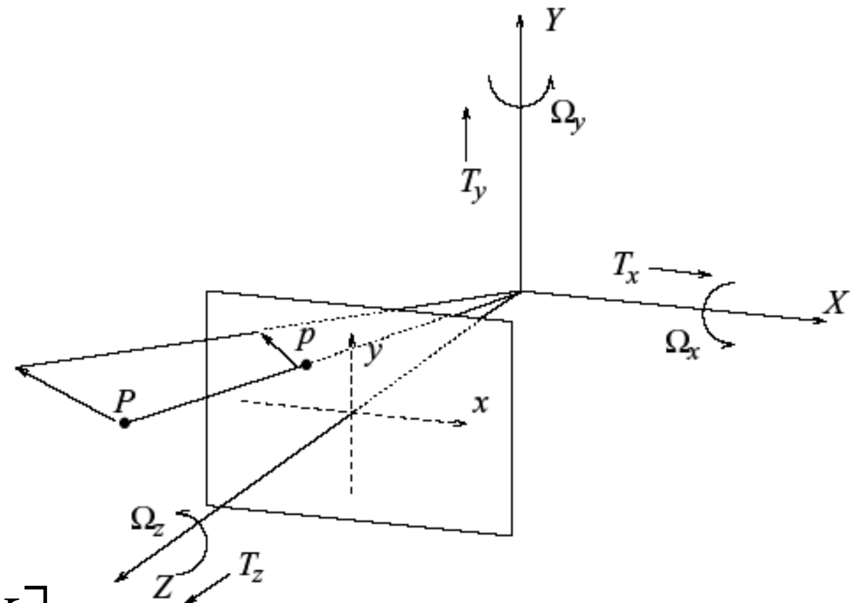
$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} x & y & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & y & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix}$$

Recall: 3D motion



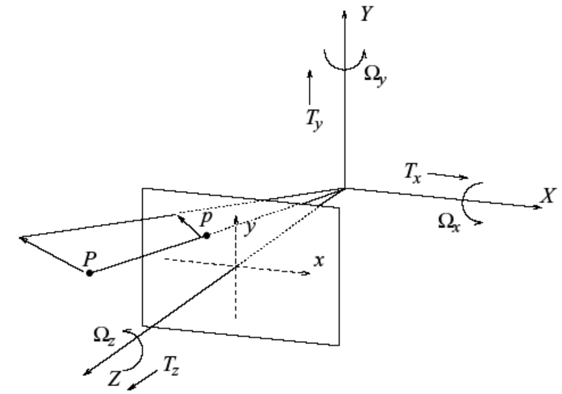
Motion Models

3D Rigid Motion



$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + T = R_Z^{\Omega_Z} R_Y^{\Omega_Y} R_X^{\Omega_X} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + T$$

Rotation



$$R_Z^{\Omega_Z} = \begin{bmatrix} \cos\Omega_Z & -\sin\Omega_Z & 0 \\ \sin\Omega_Z & \cos\Omega_Z & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

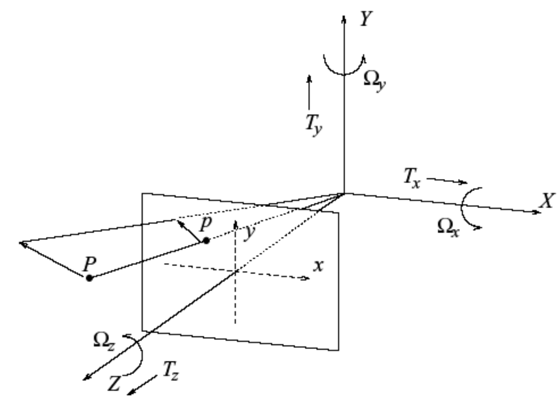
$$R_Y^{\Omega_Y} = \begin{bmatrix} \cos\Omega_Y & 0 & \sin\Omega_Y \\ 0 & 1 & 0 \\ -\sin\Omega_Y & 0 & \cos\Omega_Y \end{bmatrix}$$

$$R_X^{\Omega_X} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\Omega_X & -\sin\Omega_X \\ 0 & \sin\Omega_X & \cos\Omega_X \end{bmatrix}$$

Rotation matrix: orthogonal with determinant =1

$$R^T = R^{-1} \qquad \det(R) = 1$$

Rotation



$$R_Z^{\Omega_Z} = \begin{bmatrix} \cos\Omega_Z & -\sin\Omega_Z & 0 \\ \sin\Omega_Z & \cos\Omega_Z & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$R_Y^{\Omega_Y} = \begin{bmatrix} \cos\Omega_Y & 0 & \sin\Omega_Y \\ 0 & 1 & 0 \\ -\sin\Omega_Y & 0 & \cos\Omega_Y \end{bmatrix}$$

$$R_X^{\Omega_X} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\Omega_X & -\sin\Omega_X \\ 0 & \sin\Omega_X & \cos\Omega_X \end{bmatrix}$$

$$= \begin{bmatrix} \cos\Omega_Y\cos\Omega_Z & \sin\Omega_X\sin\Omega_Y\cos\Omega_Z - \cos\Omega_X\sin\Omega_Z & \cos\Omega_X\sin\Omega_Y\cos\Omega_Z + \sin\Omega_X\sin\Omega_Z \\ \cos\Omega_Y\sin\Omega_Z & \sin\Omega_X\sin\Omega_Y\sin\Omega_Z + \cos\Omega_X\cos\Omega_Z & \cos\Omega_X\sin\Omega_Y\sin\Omega_Z - \sin\Omega_X\cos\Omega_Z \\ -\sin\Omega_Y & \sin\Omega_X\cos\Omega_Y & \cos\Omega_X\cos\Omega_Y \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}$$

Assumption: Small Motion

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} \cos\Omega_Y \cos\Omega_Z & \sin\Omega_X \sin\Omega_Y \cos\Omega_Z - \cos\Omega_X \sin\Omega_Z & \cos\Omega_X \sin\Omega_Y \cos\Omega_Z + \sin\Omega_X \sin\Omega_Z \\ \cos\Omega_Y \sin\Omega_Z & \sin\Omega_X \sin\Omega_Y \sin\Omega_Z + \cos\Omega_X \cos\Omega_Z & \cos\Omega_X \sin\Omega_Y \sin\Omega_Z - \sin\Omega_X \cos\Omega_Z \\ -\sin\Omega_Y & \sin\Omega_X \cos\Omega_Y & \cos\Omega_X \cos\Omega_Y \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}$$

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} \approx \begin{bmatrix} 1 & -\Omega_Z & \Omega_Y \\ \Omega_Z & 1 & -\Omega_X \\ -\Omega_Y & \Omega_X & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}$$

$\cos\theta \approx 1$ (If θ is small)

$\sin\theta \approx \theta$

3D Motion

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} \approx \left(\begin{bmatrix} 0 & -\Omega_Z & \Omega_Y \\ \Omega_Z & 0 & -\Omega_X \\ -\Omega_Y & \Omega_X & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}$$

$$\begin{bmatrix} V_X \\ V_Y \\ V_Z \end{bmatrix} = \begin{bmatrix} X' - X \\ Y' - Y \\ Z' - Z \end{bmatrix} \approx \begin{bmatrix} 0 & -\Omega_Z & \Omega_Y \\ \Omega_Z & 0 & -\Omega_X \\ -\Omega_Y & \Omega_X & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}$$

$$V_X = -\Omega_Z Y + \Omega_Y Z + T_X$$

$$V_Y = \Omega_Z X - \Omega_X Z + T_Y$$

$$V_Z = -\Omega_Y X + \Omega_X Y + T_Z$$

Assumption: Planar World + Orthographic Projection

$$Z = a + bX + cY$$

$$x = X$$

$$y = Y$$

$$u = v_x = -\Omega_Z y + \Omega_Y Z + T_X$$

$$v = v_y = \Omega_Z x - \Omega_X Z + T_Y$$

$$u = v_x = -\Omega_Z y + \Omega_Y (a + bx + cy) + T_X$$

$$v = v_y = \Omega_Z x - \Omega_X (a + bx + cy) + T_Y$$

Assumption: Planar World

$$u = v_x = -\Omega_Z y + \Omega_Y (a + bx + cy) + T_X$$
$$v = v_y = \Omega_Z x - \Omega_X (a + bx + cy) + T_Y$$

$$u = v_x = (\Omega_Y c - \Omega_Z) y + \Omega_Y bx + (\Omega_Y a + T_X)$$
$$v = v_y = (\Omega_Z - \Omega_X b) x - \Omega_X cy + (T_Y - \Omega_X a)$$

$$u = v_x = a_1 x + a_2 y + a_3$$
$$v = v_y = a_4 x + a_5 y + a_6$$

Substitute:

$$a_1 = \Omega_Y b$$
$$a_3 = \Omega_Y a + T_X$$
$$a_2 = \Omega_Y c - \Omega_Z$$
$$a_4 = \Omega_Z - \Omega_X b$$
$$a_5 = -\Omega_X c$$
$$a_6 = T_Y - \Omega_X a$$

Affine Flow

Small motion assumption

* e.g. at video frame rate

Planar surface

* look at only a small region of the scene

Orthographic projection

* surface distant from camera

* long focal length

$$u = v_x = a_1 x + a_2 y + a_3$$

$$v = v_y = a_4 x + a_5 y + a_6$$

Assumptions

What might be wrong with this?

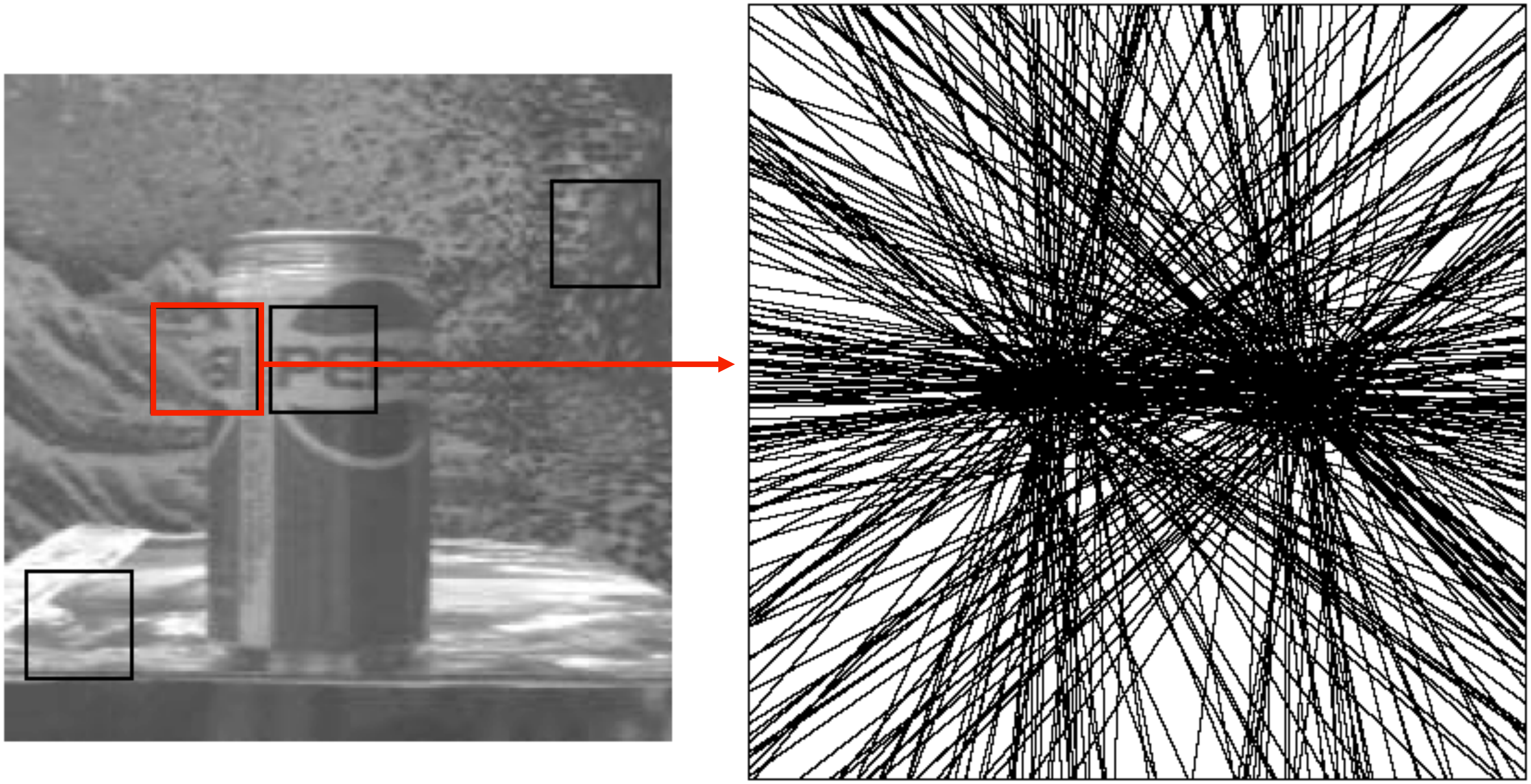
$$E(\mathbf{a}) = \sum_{x,y \in R} (\nabla I^T \mathbf{u}(\mathbf{x};\mathbf{a}) + I_t)^2$$

Is there a probabilistic interpretation?

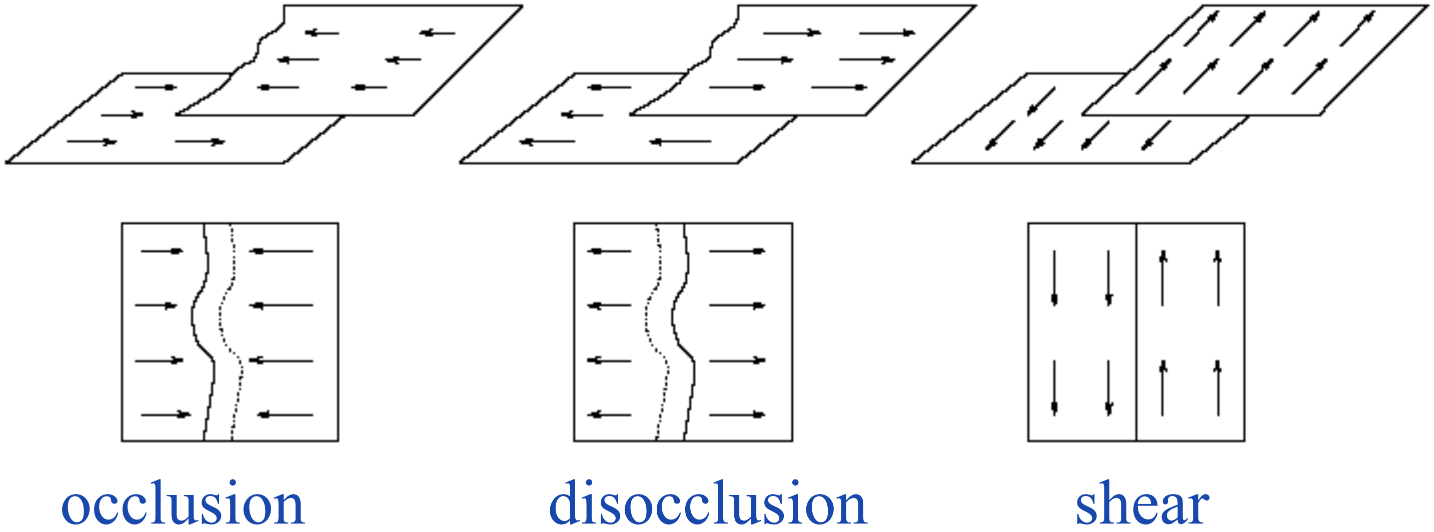
$$\max_{\mathbf{a}} p(I \mid \mathbf{a}) \propto \exp(-\frac{1}{2\sigma^2} \sum_{x,y \in R} (\nabla I^T \mathbf{u}(\mathbf{x};\mathbf{a}) + I_t)^2)$$

Minimize the negative log.

Multiple Motions

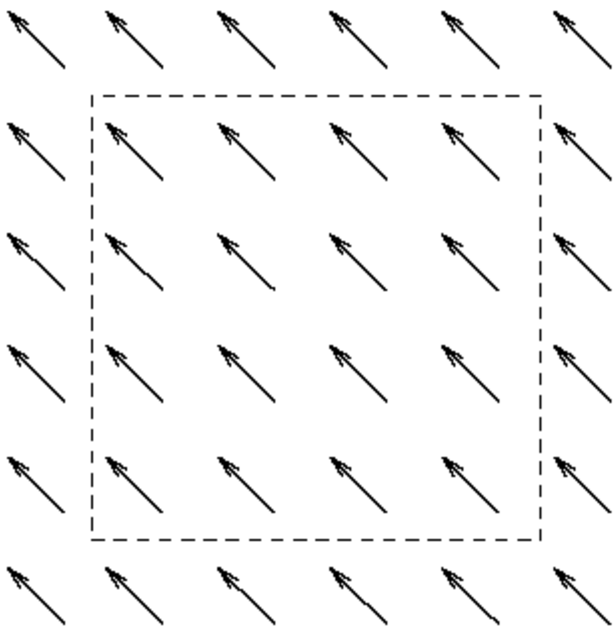


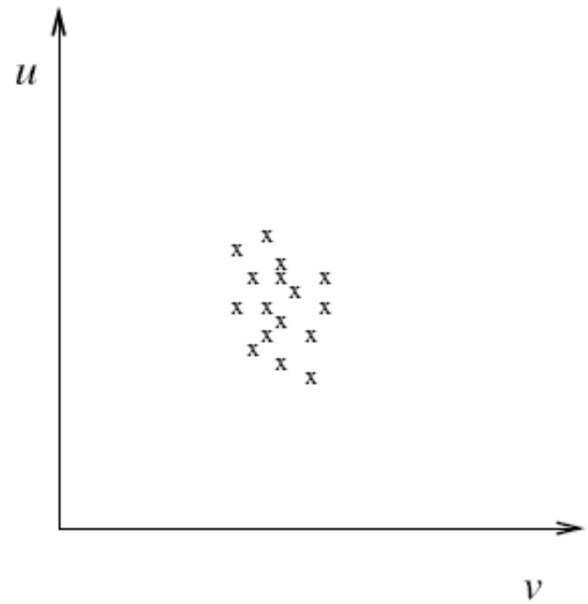
Occlusion



Multiple motions within a finite region.

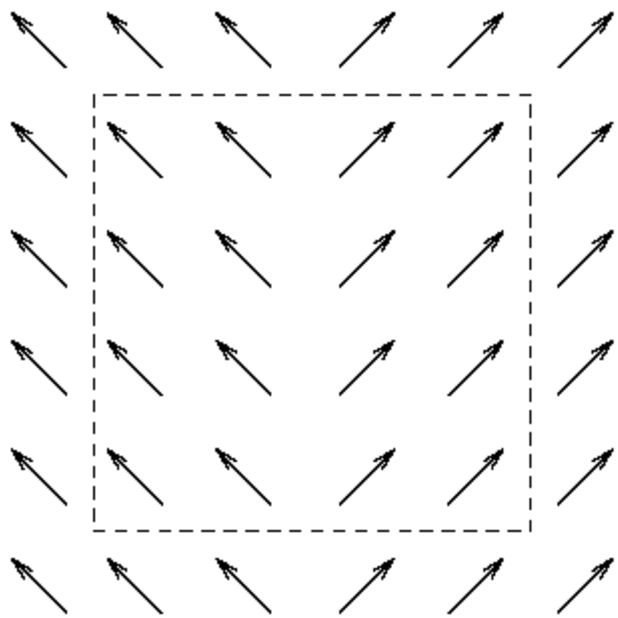
Coherent Motion

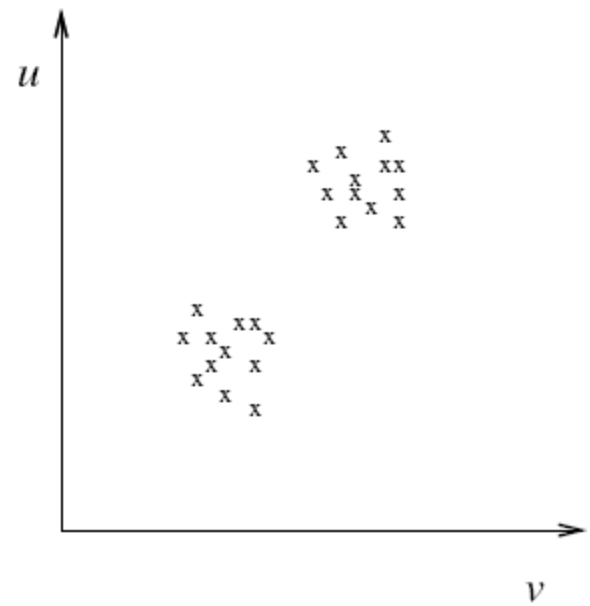




Possibly Gaussian.

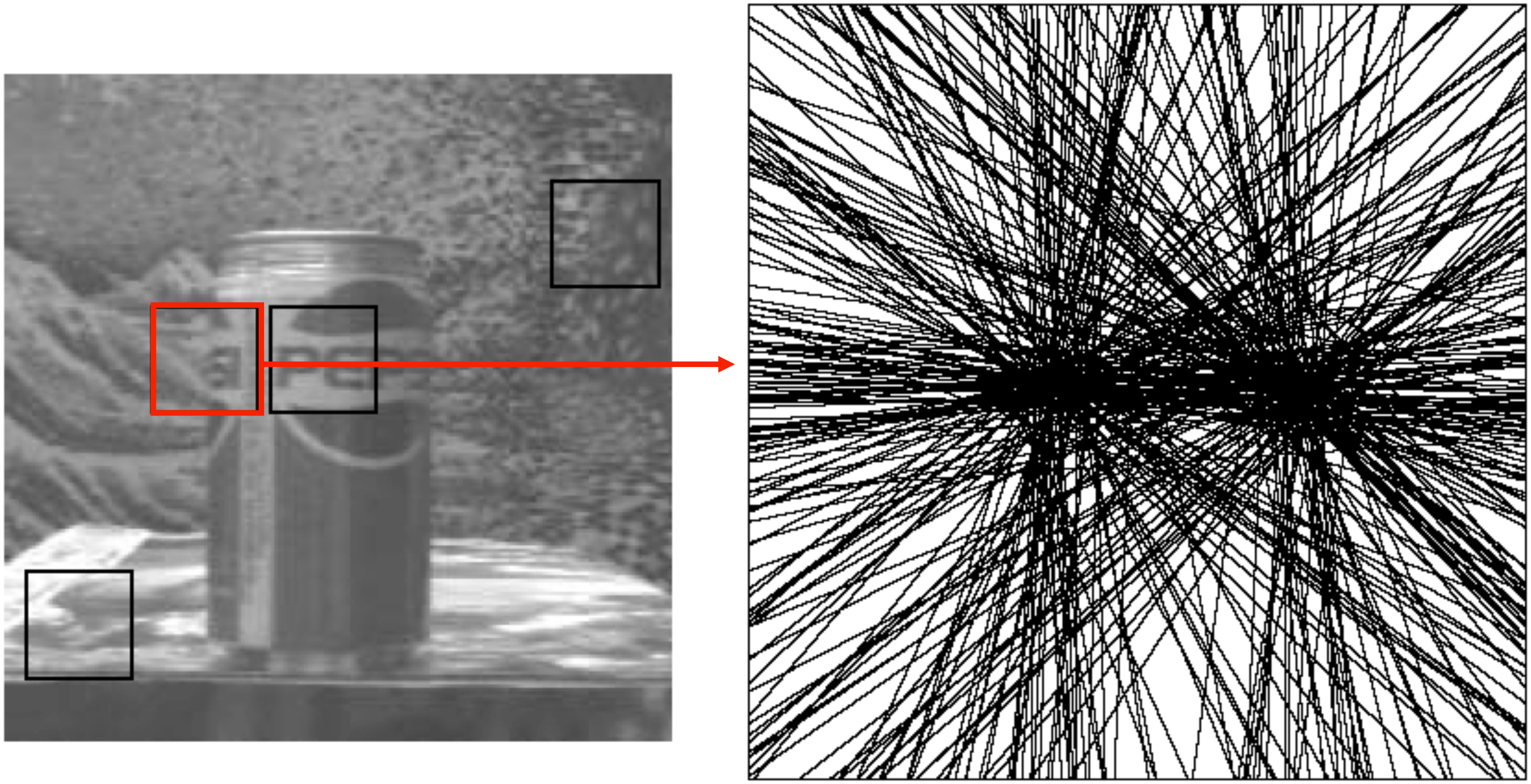
Multiple Motions





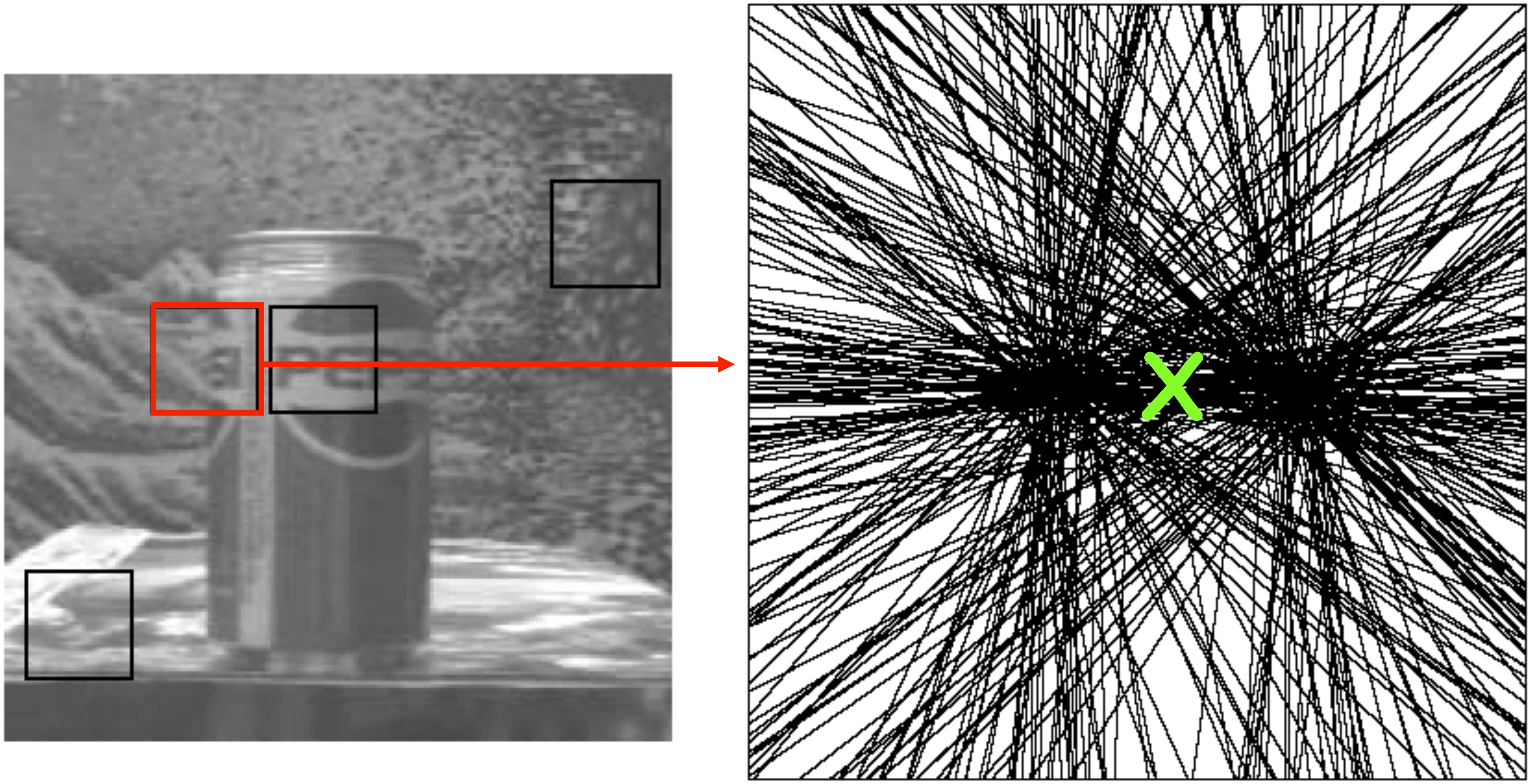
Definitely not Gaussian.

Multiple Motions



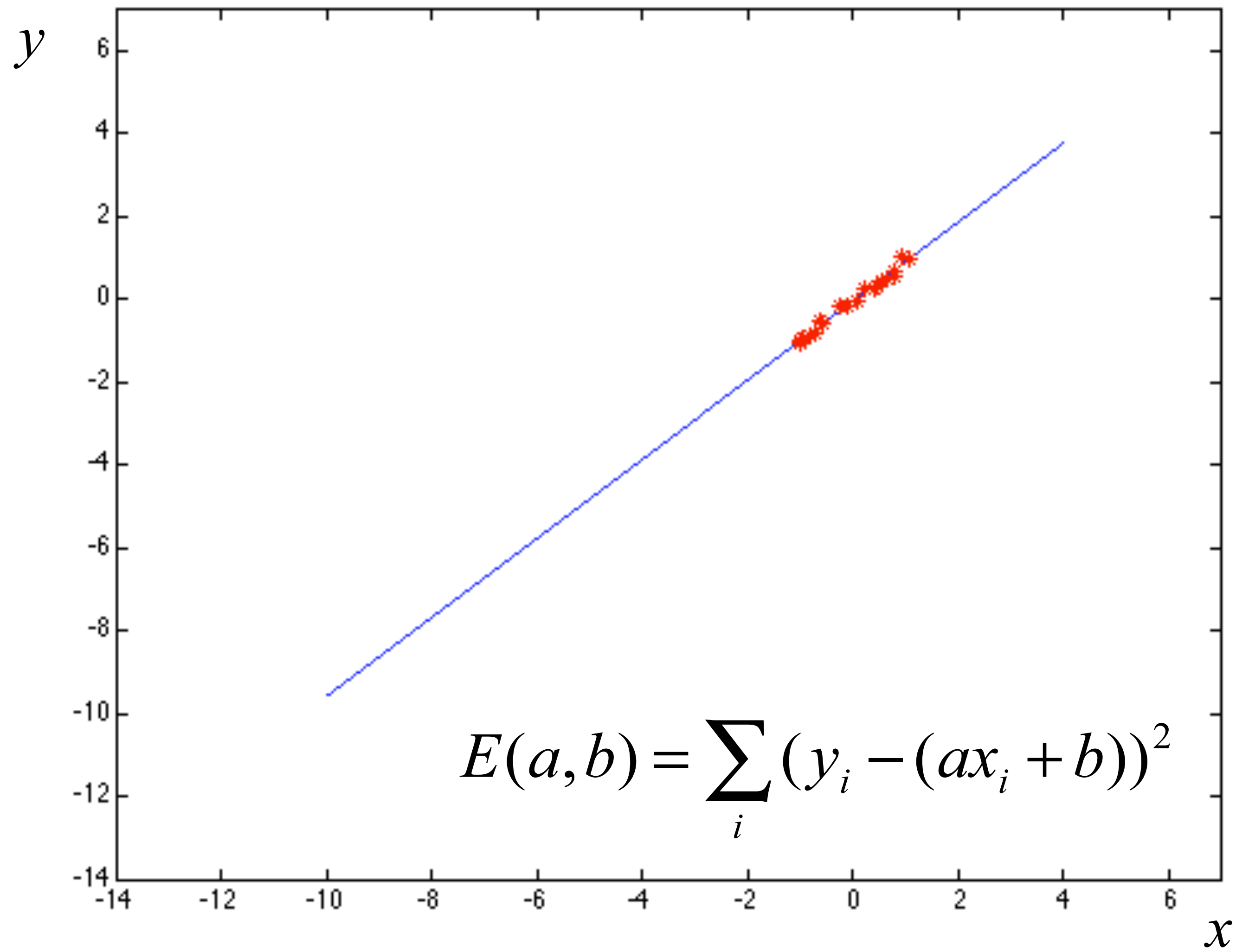
What is the “best” fitting translational motion?

Multiple Motions

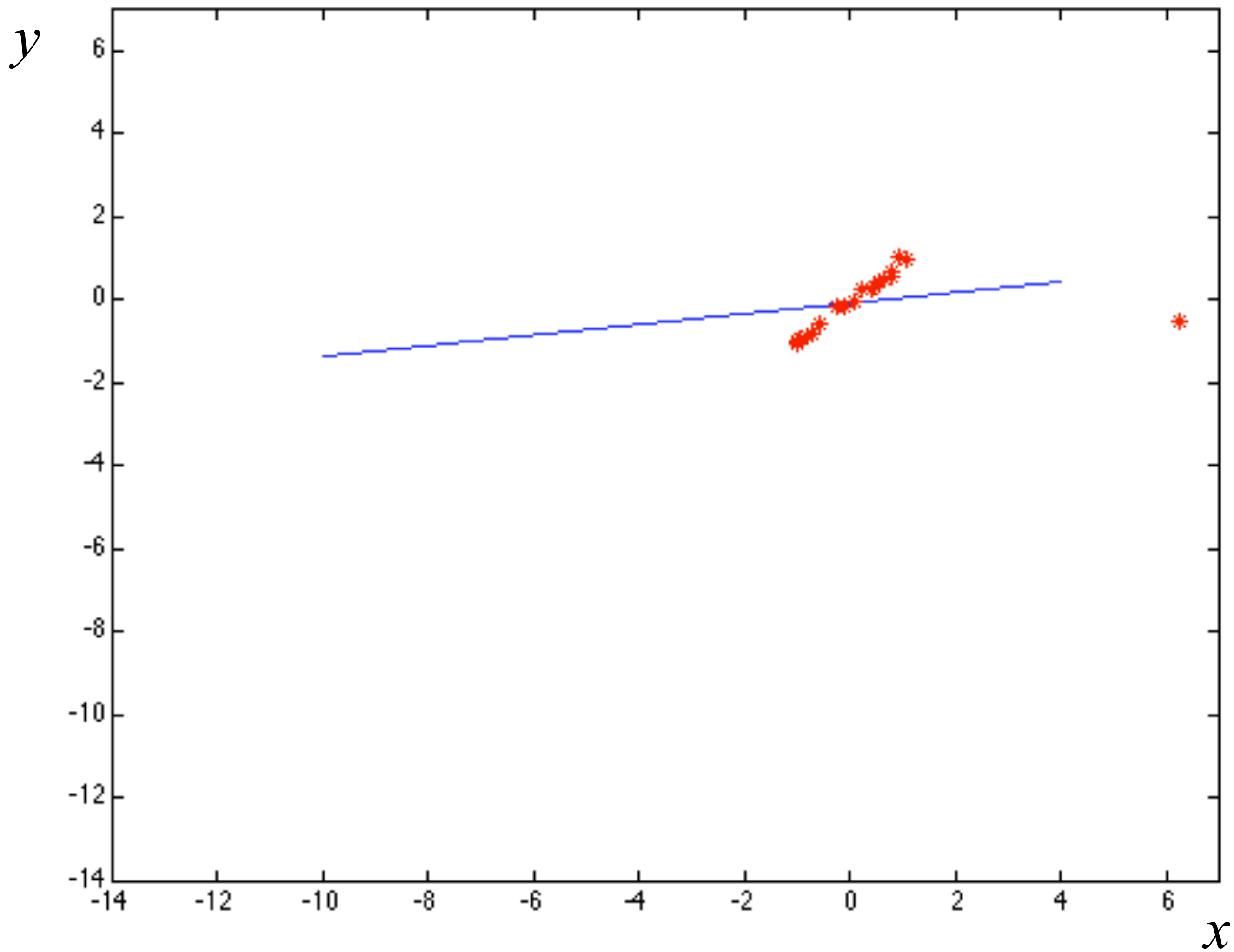


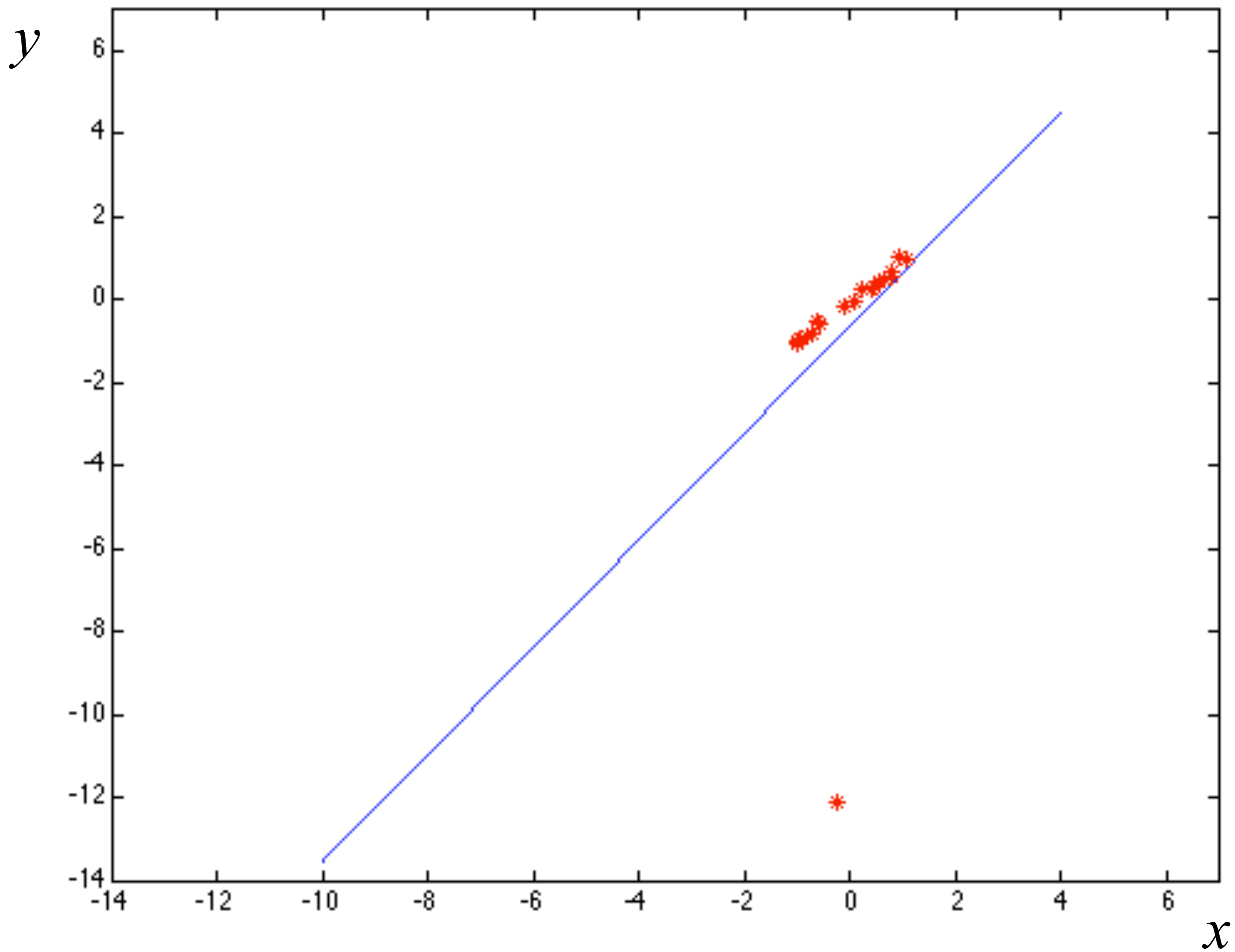
Least squares fit.

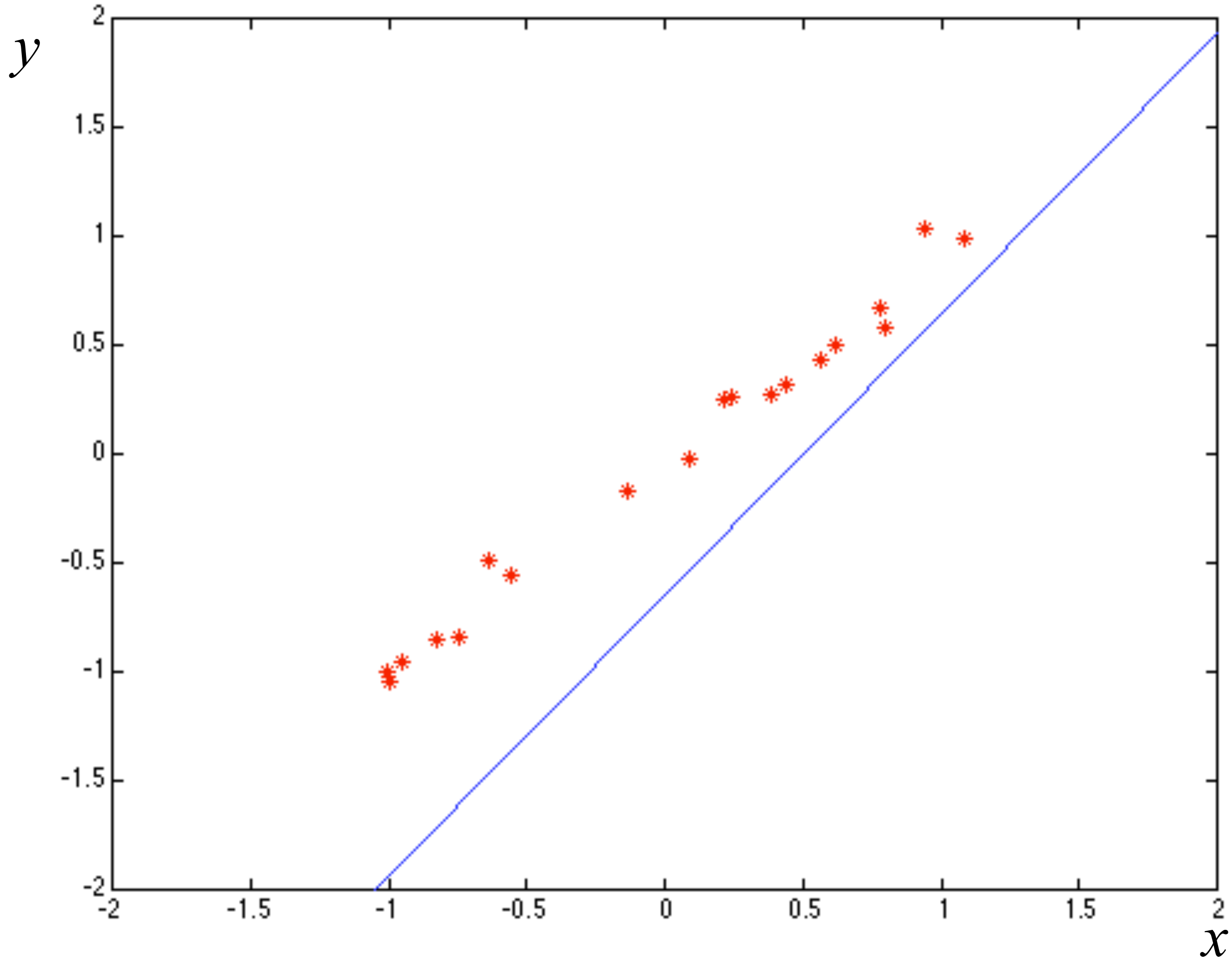
Simpler problem: fitting a line to data



$$E(a,b)=\sum_i (y_i-(ax_i+b))^2$$


y
6
4
2
0
-2
-4
-6
-8
-10
-12
-14
-14
-12
-10
-8
-6
-4
-2
0
2
4
6
x


y
6
4
2
0
-2
-4
-6
-8
-10
-12
-14
-14
-12
-10
-8
-6
-4
-2
0
2
4
6
x

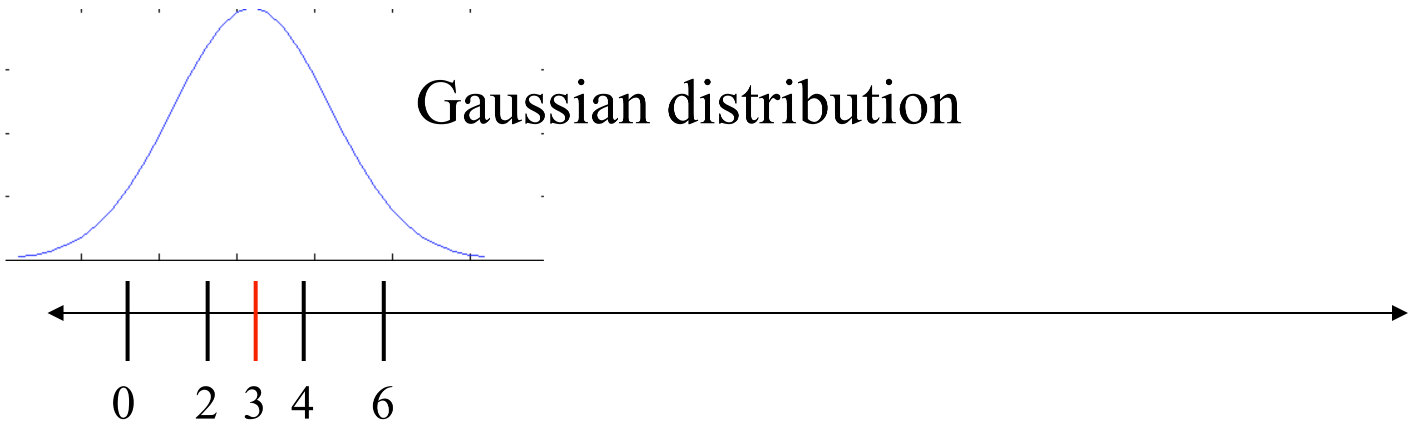

y
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
x

Robust Statistics

- Recover the best fit to the majority of the data.
- Detect and reject outliers.

History.

Estimating the mean



Mean is the optimal solution to:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} d_i$$

$$\min_{\mu} \sum_{i=1}^{N} \underbrace{(d_i - \mu)}_{\text{residual}}{}^2$$

Estimating the Mean

The mean maximizes this likelihood:

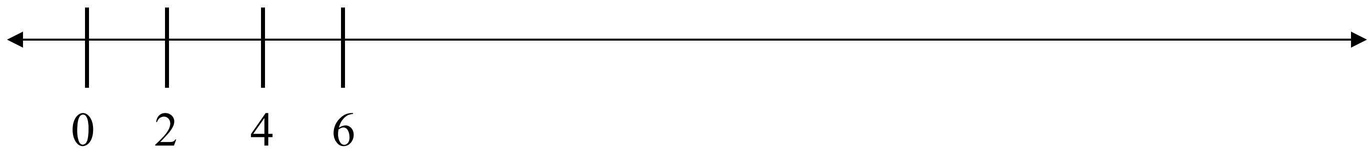
$$\max_{\mu} p(d_i \mid \mu) = \frac{1}{\sqrt{2\pi\sigma}} \prod_{i=1}^{N} \exp(-\frac{1}{2}(d_i - \mu)^2 / \sigma^2)$$

The negative log gives (with sigma=1):

$$\min_{\mu} \sum_{i=1}^{N} (d_i - \mu)^2$$

"least squares" estimate

Estimating the mean



Estimating the mean

What happens if we change just one measurement?

$$\mu' = \mu + \frac{\Delta}{N}$$

With a single “bad” data point I can move the mean arbitrarily far.