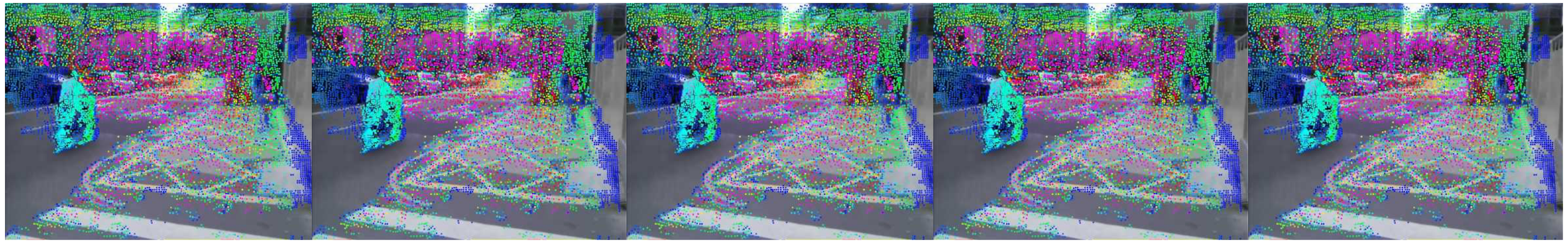
CPSC 425: Computer Vision



Lecture 20: Optical Flow (cont.)

Menu for Today (**March 19, 2019**)

Topics:

— Optical Flow (cont)
— Classification
— Naive Bayes Classifier
— Bayes' Risk

Redings:

— **Today's** Lecture: Forsyth & Ponce (2nd ed.) 15.1, 15.2
— **Next** Lecture: Forsyth & Ponce (2nd ed.) 16.1.3, 16.1.4, 16.1.9

Reminders:

— **Assignment 4**: Local Invariant Features and RANSAC due **today**

— **Assignment 5**: Scene Recognition with Bag of Words out **soon**

Today's "**fun**" Example: Visual Microphone

Follow-up work to previous lecture's example of **Eulerian** video magnification

Lecture 19: Re-cap

Optical flow is the apparent motion of brightness patterns in the image

Applications

— image and video stabilization in digital cameras, camcorders

— motion-compensated video compression schemes such as MPEG

— image registration for medical imaging, remote sensing

— action recognition

— motion segmentation

Lecture 19: Re-cap

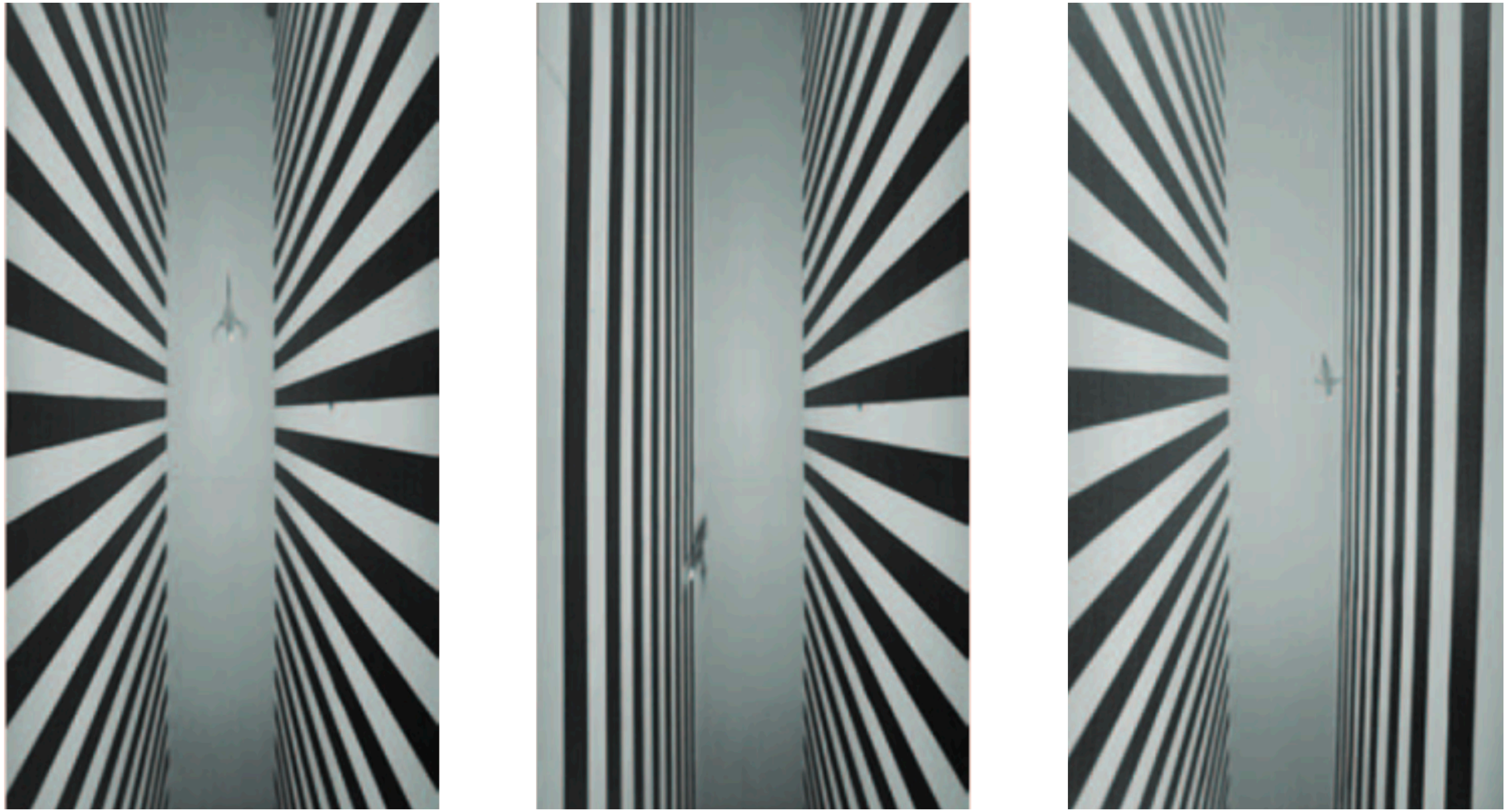
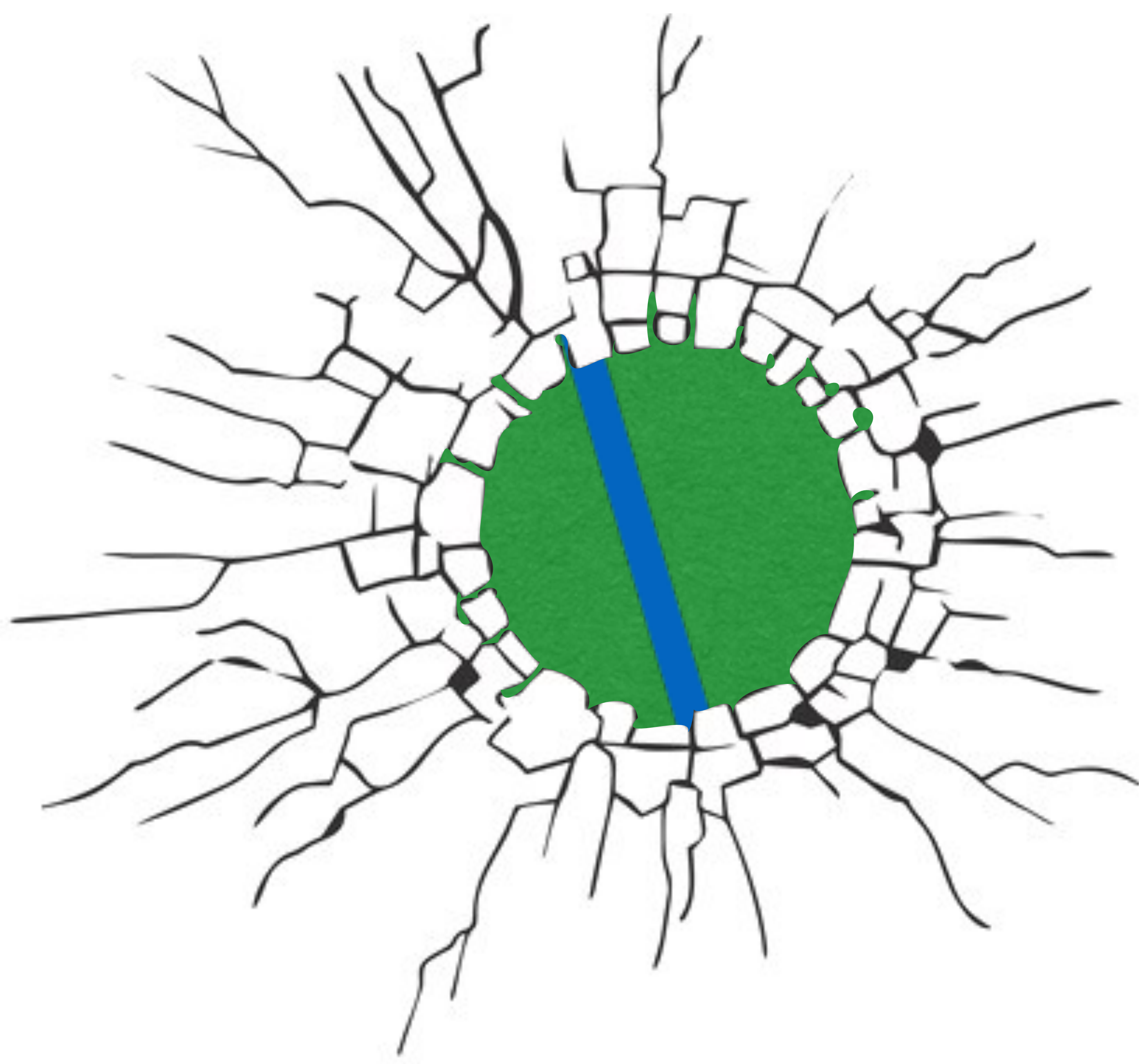


Figure credit: M. Srinivasan

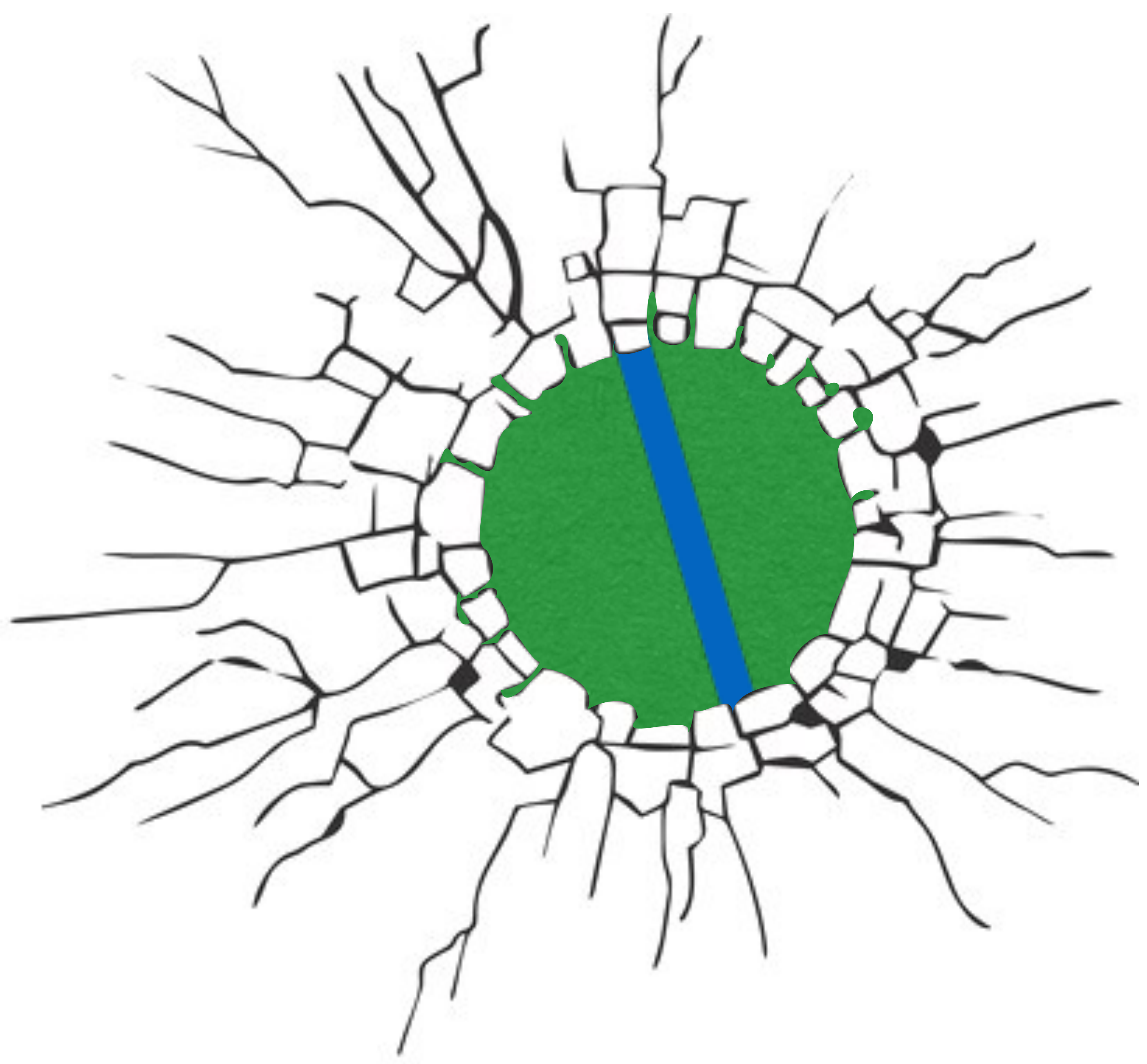
Aperture Problem



In which direction is the line moving?

Image Credit: Ioannis (Yannis) Gkioulekas (CMU)

Aperture Problem



In which direction is the line moving?

Image Credit: Ioannis (Yannis) Gkioulekas (CMU)

Aperture Problem

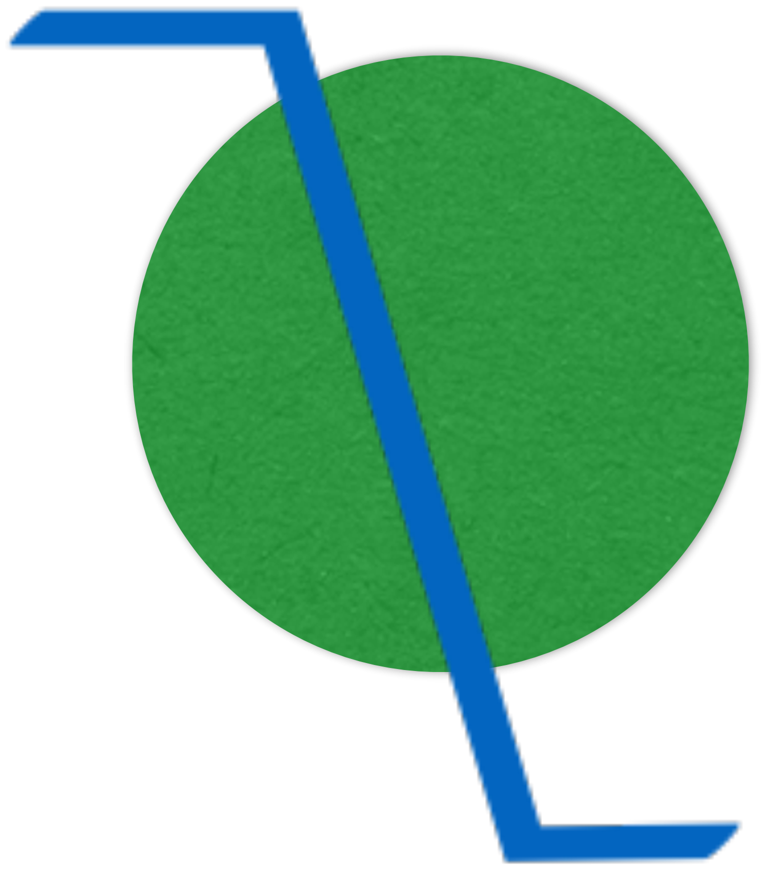


Image Credit: Ioannis (Yannis) Gkioulekas (CMU)

Aperture Problem

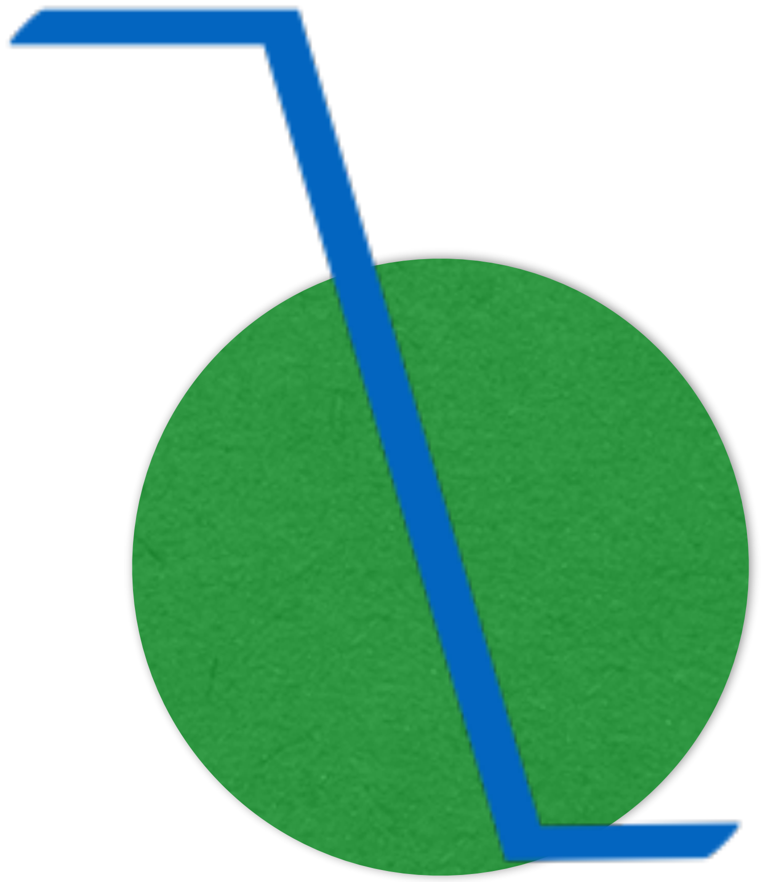


Image Credit: Ioannis (Yannis) Gkioulekas (CMU)

Aperture Problem

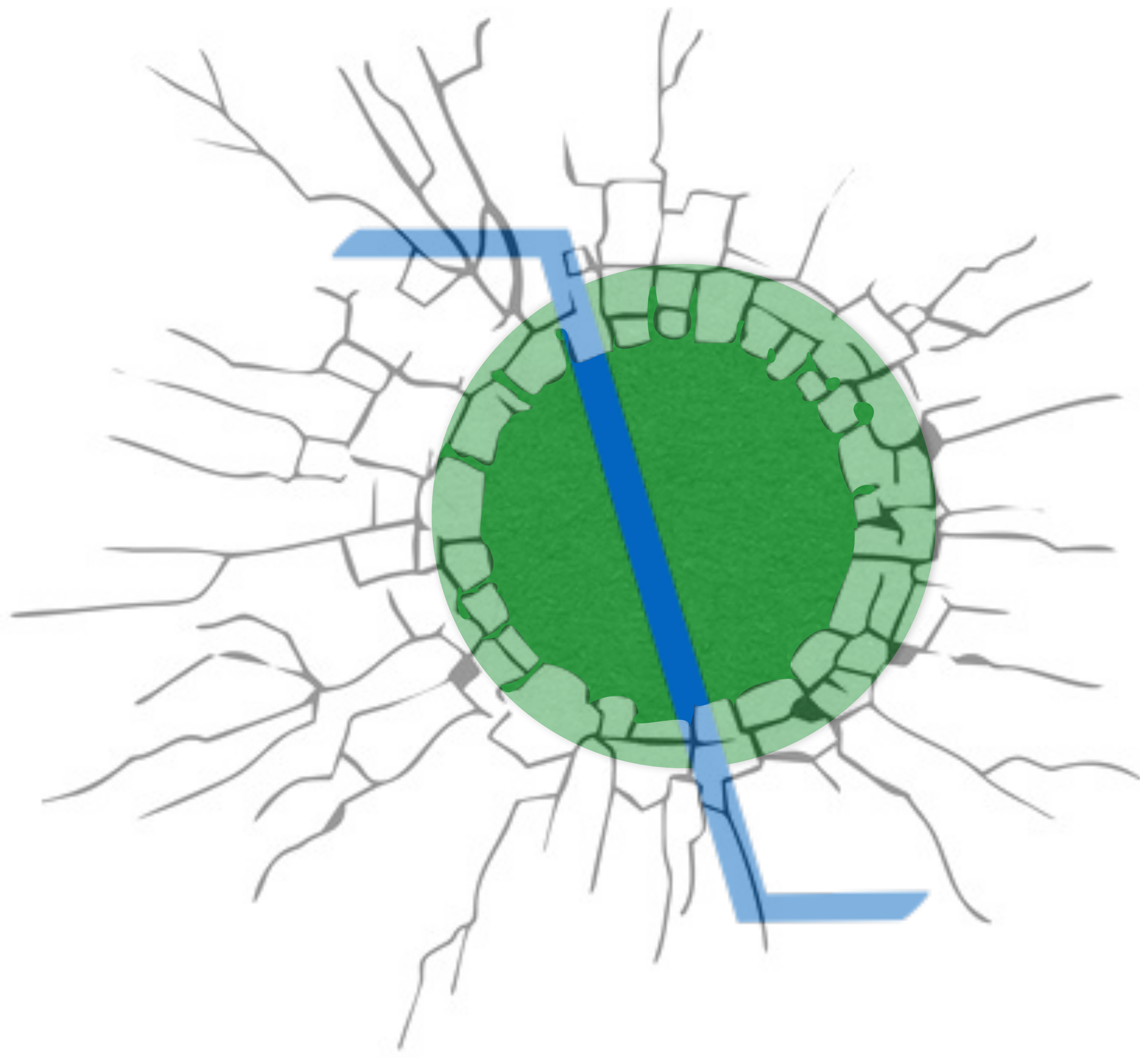


Image Credit: Ioannis (Yannis) Gkioulekas (CMU)

Aperture Problem

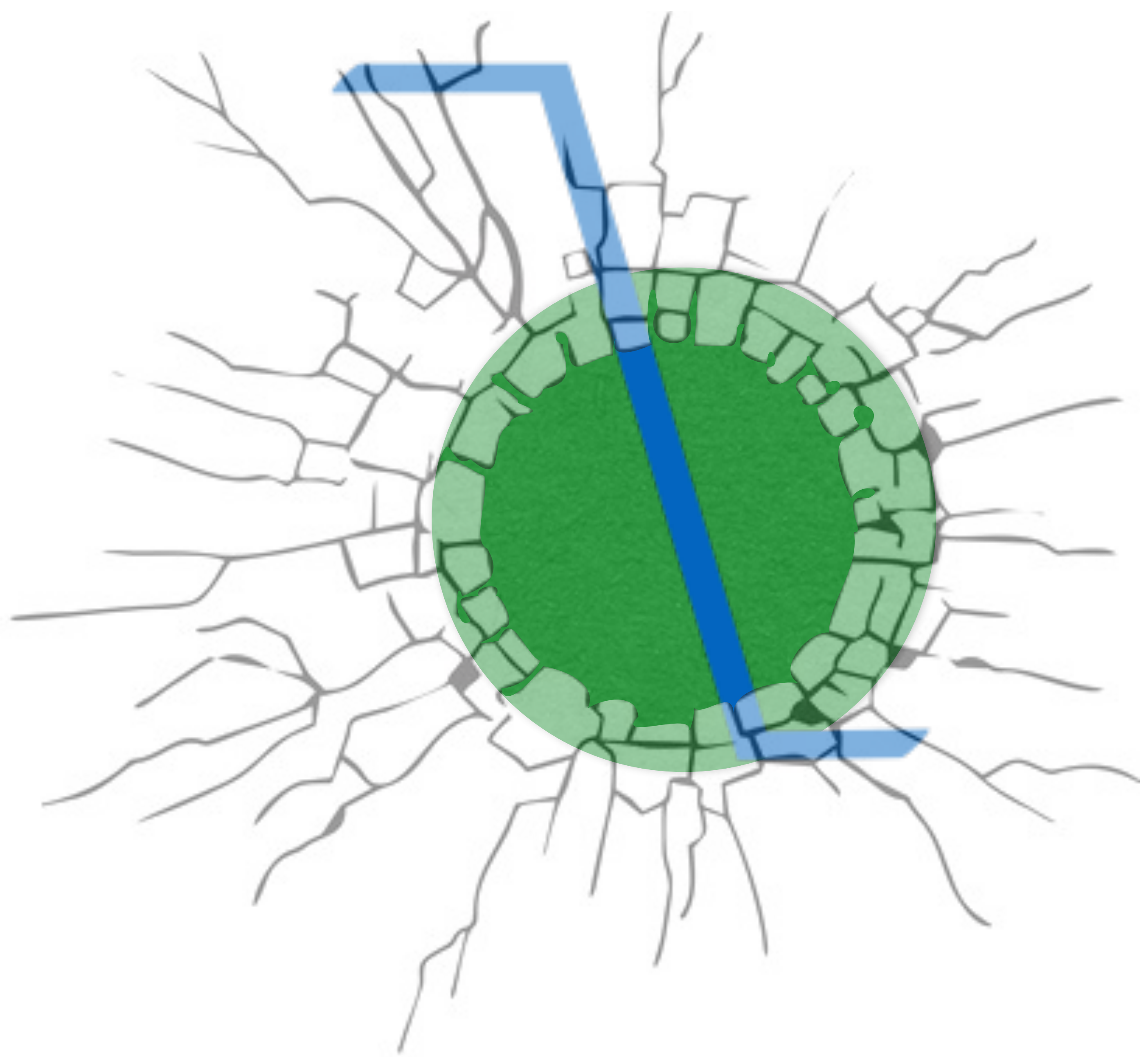
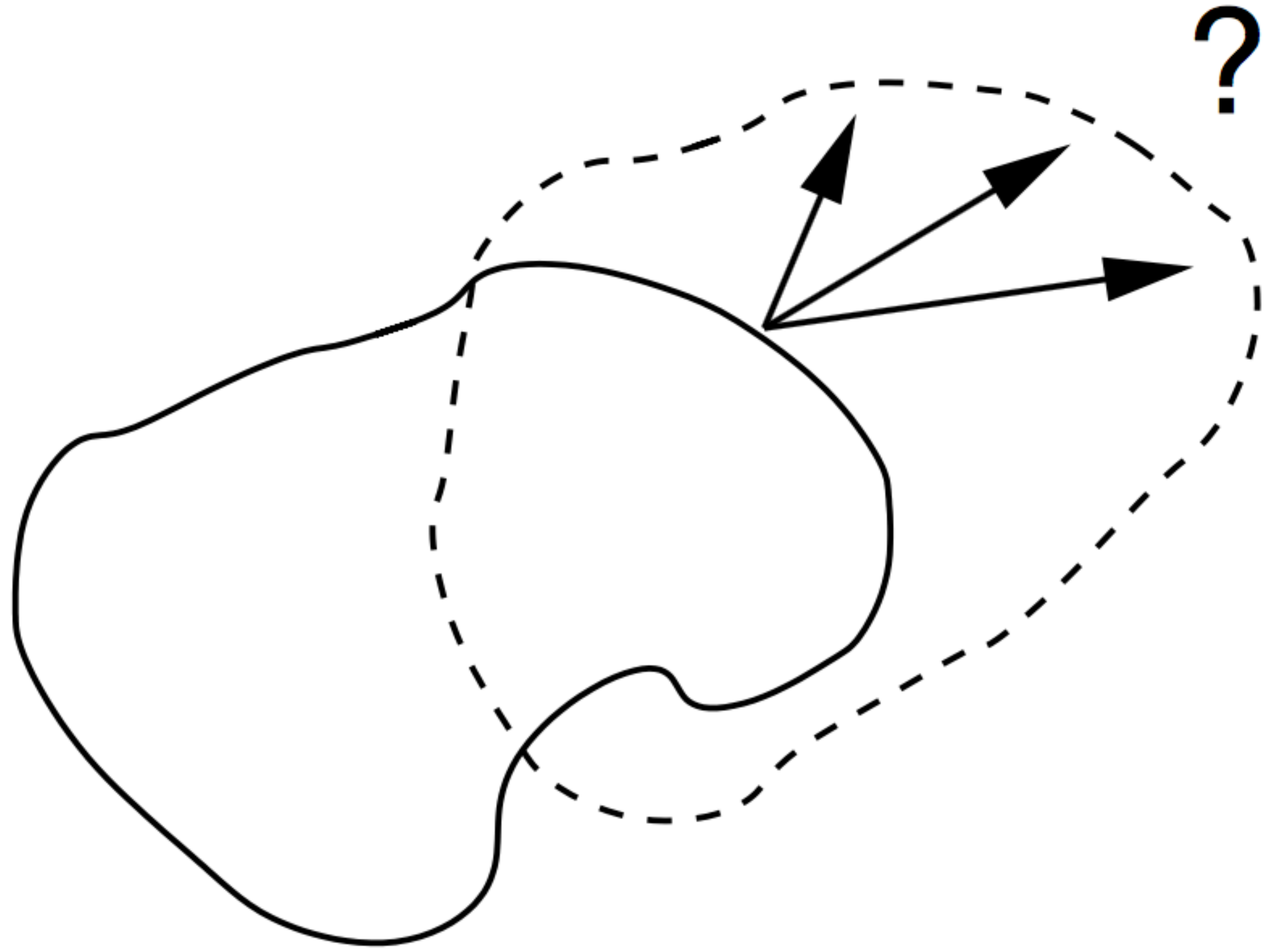


Image Credit: Ioannis (Yannis) Gkioulekas (CMU)

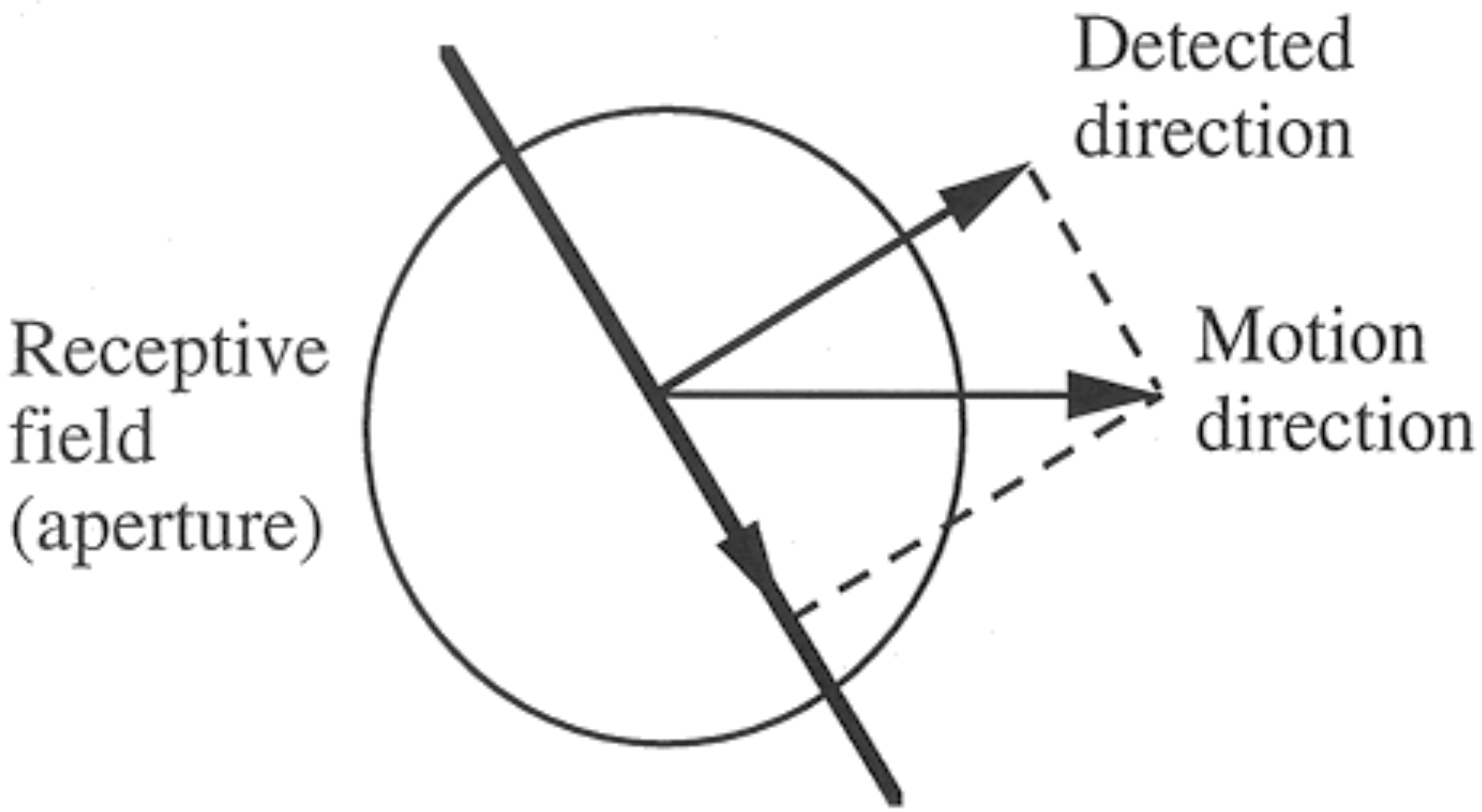
Aperture Problem



— Without distinct features to track, the true visual motion is ambiguous

— Locally, one can compute only the component of the visual motion in the direction perpendicular to the contour

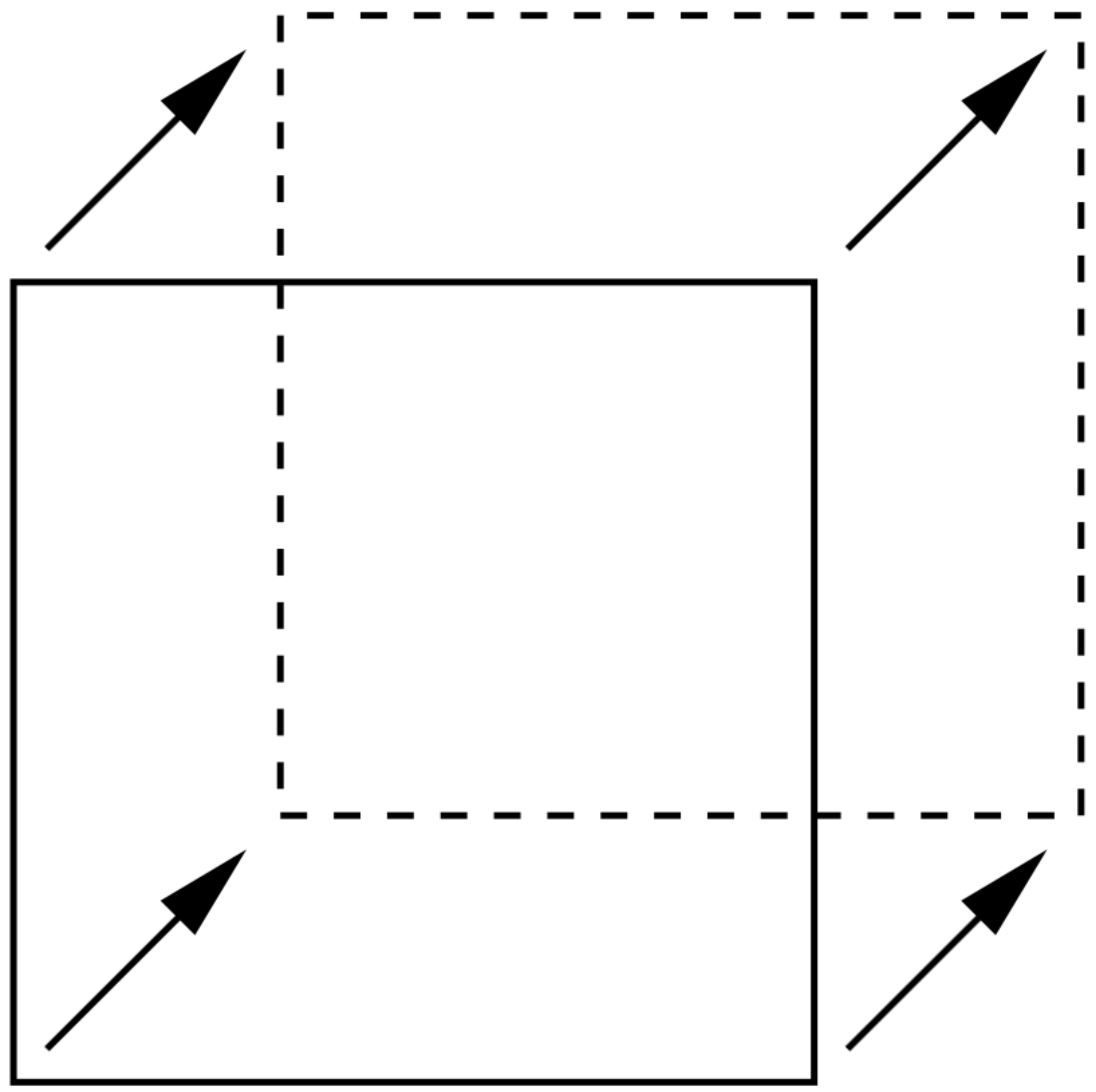
Aperture Problem



— Without distinct features to track, the true visual motion is ambiguous

— Locally, one can compute only the component of the visual motion in the direction perpendicular to the contour

Visual **Motion**



Visual motion is determined when there are distinct features to track, provided:

— the features can be detected and localized accurately; and

— the features can be correctly matched over time

Motion as **Matching**

Representation	Result is. . .
Point/feature based	(very) sparse
Contour based	(relatively) sparse
(Differential) gradient based	dense

Optical Flow **Constraint Equation**

Consider image intensity also to be a function of time, t. We write

$$I(x, y, t)$$

Optical Flow **Constraint Equation**

Consider image intensity also to be a function of time, t. We write

$$I(x, y, t)$$

Applying the **chain rule for differentiation**, we obtain

$$\frac{dI(x, y, t)}{dt} = I_x \frac{dx}{dt} + I_y \frac{dy}{dt} + I_t$$

where subscripts denote partial differentiation

Optical Flow **Constraint Equation**

Consider image intensity also to be a function of time, t. We write

$$I(x, y, t)$$

Applying the **chain rule for differentiation**, we obtain

$$\frac{dI(x, y, t)}{dt} = I_x \frac{dx}{dt} + I_y \frac{dy}{dt} + I_t$$

where subscripts denote partial differentiation

Define $u = \dfrac{dx}{dt}$ and $v = \dfrac{dy}{dt}$. Then $[u, v]$ is the 2-D motion and the space of all such u and v is the **2-D velocity space**

Optical Flow **Constraint Equation**

Consider image intensity also to be a function of time, t. We write

$$I(x, y, t)$$

Applying the **chain rule for differentiation**, we obtain

$$\frac{dI(x,y,t)}{dt} = I_x \frac{dx}{dt} + I_y \frac{dy}{dt} + I_t$$

where subscripts denote partial differentiation

Define $u = \frac{dx}{dt}$ and $v = \frac{dy}{dt}$. Then $[u, v]$ is the 2-D motion and the space of all such u and v is the **2-D velocity space**

Suppose $\frac{dI(x,y,t)}{dt} = 0$. Then we obtain the (classic) **optical flow constraint equation**

$$I_x u + I_y v + I_t = 0$$

Optical Flow **Constraint Equation**

Consider image intensity also to be a function of time, t. We write

$$I(x, y, t)$$

Applying the **chain rule for differentiation**, we obtain

$$\frac{dI(x,y,t)}{dt} = I_x \frac{dx}{dt} + I_y \frac{dy}{dt} + I_t$$

where subscripts denote partial differentiation

Define $u = \frac{dx}{dt}$ and $v = \frac{dy}{dt}$. Then $[u, v]$ is the 2-D motion and the space of all such u and v is the **2-D velocity space**

Suppose $\boxed{\frac{dI(x,y,t)}{dt} = 0}$. Then we obtain the (classic) **optical flow constraint equation**

$$I_x u + I_y v + I_t = 0$$

Optical Flow **Constraint Equation**

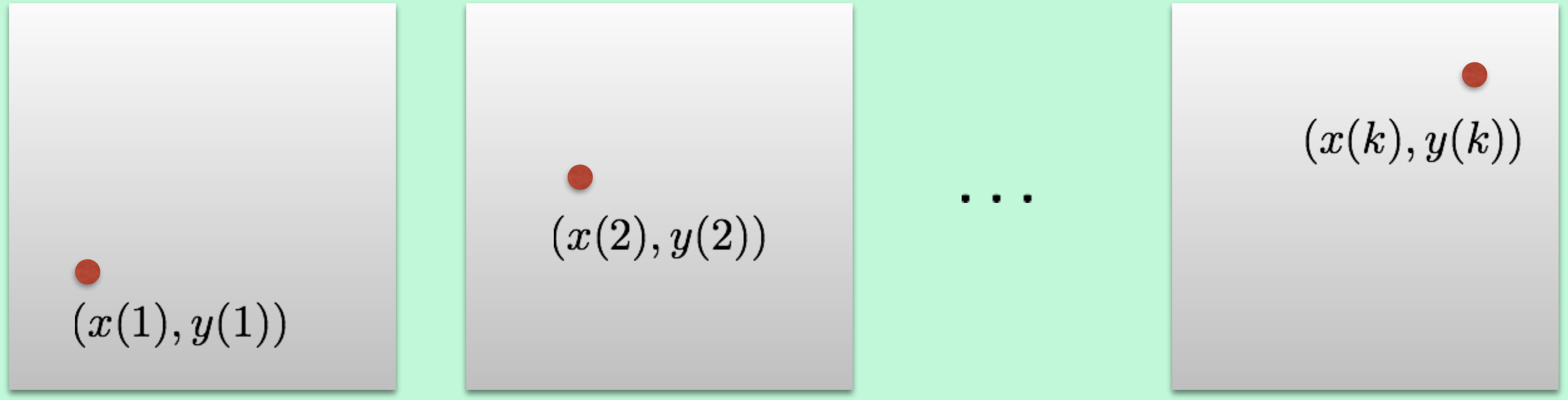
What does this mean, and why is it reasonable?

Suppose $\frac{dI(x,y,t)}{dt} = 0$. Then we obtain the (classic) **optical flow constraint equation**

$$I_x u + I_y v + I_t = 0$$

Optical Flow **Constraint Equation**

Scene point moving through image sequence



What does this mean, and why is it reasonable?

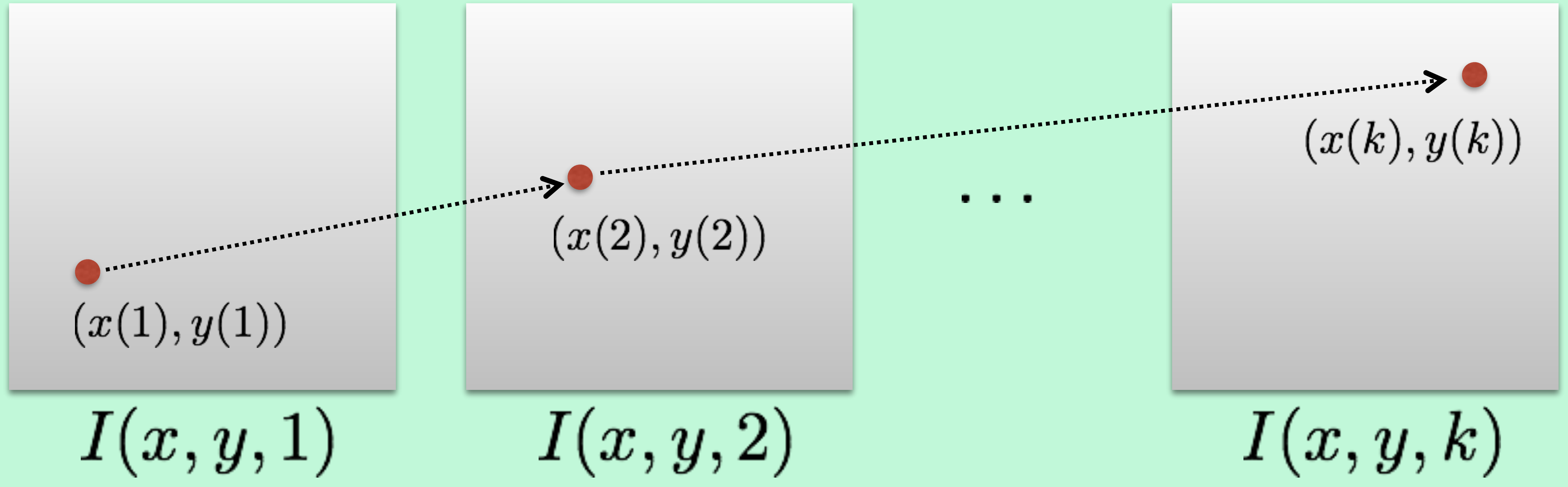
Suppose $\frac{dI(x,y,t)}{dt} = 0$. Then we obtain the (classic) **optical flow constraint equation**

$$I_x u + I_y v + I_t = 0$$

Image Credit: Ioannis (Yannis) Gkioulekas (CMU)

Optical Flow **Constraint Equation**

Scene point moving through image sequence



What does this mean, and why is it reasonable?

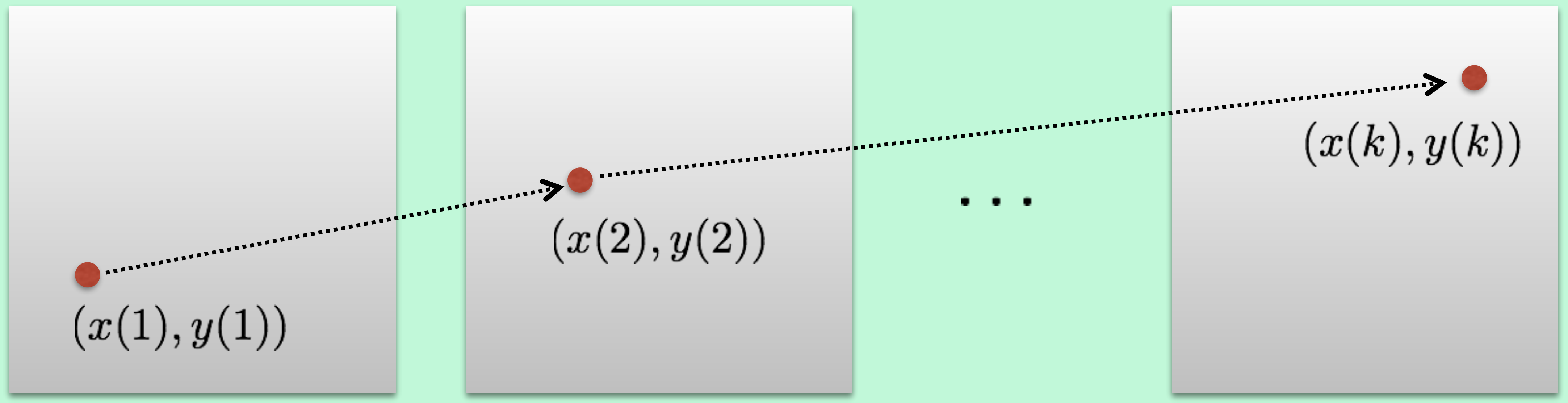
Suppose $\frac{dI(x,y,t)}{dt} = 0$. Then we obtain the (classic) **optical flow constraint equation**

$$I_x u + I_y v + I_t = 0$$

Image Credit: Ioannis (Yannis) Gkioulekas (CMU)

Optical Flow **Constraint Equation**

Brightness Constancy Assumption: Brightness of the point remains the same



$I(x, y, 1)$ $\quad$ $I(x, y, 2)$ $\quad \cdots \quad$ $I(x, y, k)$

$$I(x(t), y(t), t) = C$$

constant

What does this mean, and why is it reasonable?

Suppose $\dfrac{dI(x,y,t)}{dt} = 0$. Then we obtain the (classic) **optical flow constraint equation**

$$I_x u + I_y v + I_t = 0$$

Image Credit: Ioannis (Yannis) Gkioulekas (CMU)

Aside: Derivation of Optical Flow Constraint

$$I(x + u\delta t, y + v\delta t, t + \delta t) = I(x, y, t)$$

For small space-time step, brightness of a point is the same

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Aside: Derivation of Optical Flow Constraint

$$I(x + u\delta t, y + v\delta t, t + \delta t) = I(x, y, t)$$

For small space-time step, brightness of a point is the same

Insight:
If the time step is really small,
we can *linearize* the intensity function

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Aside: Derivation of Optical Flow Constraint

$$I(x + u\delta t, y + v\delta t, t + \delta t) = I(x, y, t)$$

Multivariable Taylor Series Expansion

(First order approximation, two variables)

$$f(x, y) \approx f(a, b) + f_x(a, b)(x - a) - f_y(a, b)(y - b)$$

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Aside: Derivation of Optical Flow Constraint

$$I(x + u\delta t, y + v\delta t, t + \delta t) = I(x, y, t)$$

Multivariable Taylor Series Expansion

(First order approximation, two variables)

$$f(x,y) \approx f(a,b) + f_x(a,b)(x-a) - f_y(a,b)(y-b)$$

$$I(x,y,t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t = I(x,y,t)$$ assuming small motion

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Aside: Derivation of Optical Flow Constraint

$$I(x + u\delta t, y + v\delta t, t + \delta t) = I(x, y, t)$$

Multivariable Taylor Series Expansion

(First order approximation, two variables)

$$f(x,y) \approx f(a,b) + f_x(a,b)(x-a) - f_y(a,b)(y-b)$$

partial derivative

$$I(x,y,t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t = I(x,y,t)$$

fixed point

assuming small motion

cancel terms

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Aside: Derivation of Optical Flow Constraint

$$I(x + u\delta t, y + v\delta t, t + \delta t) = I(x, y, t)$$

Multivariable Taylor Series Expansion

(First order approximation, two variables)

$$f(x, y) \approx f(a, b) + f_x(a, b)(x - a) - f_y(a, b)(y - b)$$

$$I(x, y, t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t = I(x, y, t)$$ assuming small motion

$$\frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t = 0$$ cancel terms

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Aside: Derivation of Optical Flow Constraint

$$I(x + u\delta t, y + v\delta t, t + \delta t) = I(x, y, t)$$

Multivariable Taylor Series Expansion

(First order approximation, two variables)

$$f(x, y) \approx f(a, b) + f_x(a, b)(x - a) - f_y(a, b)(y - b)$$

$$I(x, y, t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t = I(x, y, t)$$

assuming small motion

$$\frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t = 0$$

divide by δt

take limit $\delta t \to 0$

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Aside: Derivation of Optical Flow Constraint

$$I(x + u\delta t, y + v\delta t, t + \delta t) = I(x, y, t)$$

Multivariable Taylor Series Expansion

(First order approximation, two variables)

$$f(x,y) \approx f(a,b) + f_x(a,b)(x-a) - f_y(a,b)(y-b)$$

$$I(x,y,t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t = I(x,y,t)$$ assuming small motion

$$\frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t = 0$$

divide by δt

take limit $\delta t \to 0$

$$\frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t} = 0$$ **Brightness Constancy Equation**

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

How do we **compute** …

$$I_x u + I_y v + I_t = 0$$

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

How do we **compute** …

$$I_x u + I_y v + I_t = 0$$

$$I_x = \frac{\partial I}{\partial x} \quad I_y = \frac{\partial I}{\partial y}$$

spatial derivative

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

How do we **compute** …

$$I_x u + I_y v + I_t = 0$$

$$I_x = \frac{\partial I}{\partial x} \quad I_y = \frac{\partial I}{\partial y}$$

spatial derivative

Forward difference
Sobel filter
Scharr filter
…

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

How do we **compute** …

$$I_x u + I_y v + I_t = 0$$

$$I_x = \frac{\partial I}{\partial x} \quad I_y = \frac{\partial I}{\partial y}$$

spatial derivative

$$I_t = \frac{\partial I}{\partial t}$$

temporal derivative

Forward difference
Sobel filter
Scharr filter
…

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

How do we **compute** …

$$I_x u + I_y v + I_t = 0$$

$$I_x = \frac{\partial I}{\partial x} \quad I_y = \frac{\partial I}{\partial y}$$

spatial derivative

$$I_t = \frac{\partial I}{\partial t}$$

temporal derivative

Forward difference
Sobel filter
Scharr filter
…

Frame differencing

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Frame Differencing: Example

$t+1$

1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	10	10	10
1	1	10	10	10
1	1	10	10	10

-

t

1	1	1	1	1
1	1	1	1	1
1	10	10	10	10
1	10	10	10	10
1	10	10	10	10
1	10	10	10	10

=

$$I_t = \frac{\partial I}{\partial t}$$

0	0	0	0	0
0	0	0	0	0
0	-9	-9	-9	-9
0	-9	0	0	0
0	-9	0	0	0
0	-9	0	0	0

(example of a forward temporal difference)

t

1	1	1	1	1
1	1	1	1	1
1	10	10	10	10
1	10	10	10	10
1	10	10	10	10
1	10	10	10	10

$t+1$

1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	10	10	10
1	1	10	10	10
1	1	10	10	10

$$I_x = \frac{\partial I}{\partial x}$$

-	0	0	0	-
-	0	0	0	-
-	9	0	0	-
-	9	0	0	-
-	9	0	0	-
-	9	0	0	-

-1 0 1

$$I_y = \frac{\partial I}{\partial y}$$

-	-	-	-	-
0	0	0	0	0
0	9	9	9	9
0	0	0	0	0
0	0	0	0	0
-	-	-	-	-

-1
0
1

$$I_t = \frac{\partial I}{\partial t}$$

0	0	0	0	0
0	0	0	0	0
0	-9	-9	-9	-9
0	-9	0	0	0
0	-9	0	0	0
0	-9	0	0	0

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

How do we **compute** …

$$I_x u + I_y v + I_t = 0$$

$$I_x = \frac{\partial I}{\partial x} \quad I_y = \frac{\partial I}{\partial y}$$

spatial derivative

$$u = \frac{dx}{dt} \quad v = \frac{dy}{dt}$$

optical flow

$$I_t = \frac{\partial I}{\partial t}$$

temporal derivative

Forward difference
Sobel filter
Scharr filter
…

How do you compute this?

Frame differencing

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

How do we **compute** …

$$I_x u + I_y v + I_t = 0$$

$$I_x = \frac{\partial I}{\partial x} \quad I_y = \frac{\partial I}{\partial y}$$

spatial derivative

$$u = \frac{dx}{dt} \quad v = \frac{dy}{dt}$$

optical flow

$$I_t = \frac{\partial I}{\partial t}$$

temporal derivative

Forward difference
Sobel filter
Scharr filter
…

We need to solve for this!
(this is the unknown in the optical flow problem)

Frame differencing

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

How do we **compute** …

$$I_x u + I_y v + I_t = 0$$

$$I_x = \frac{\partial I}{\partial x} \quad I_y = \frac{\partial I}{\partial y}$$

spatial derivative

$$u = \frac{dx}{dt} \quad v = \frac{dy}{dt}$$

optical flow

$$I_t = \frac{\partial I}{\partial t}$$

temporal derivative

Forward difference
Sobel filter
Scharr filter
…

(u, v)

Solution lies on a line

Cannot be found uniquely
with a single constraint

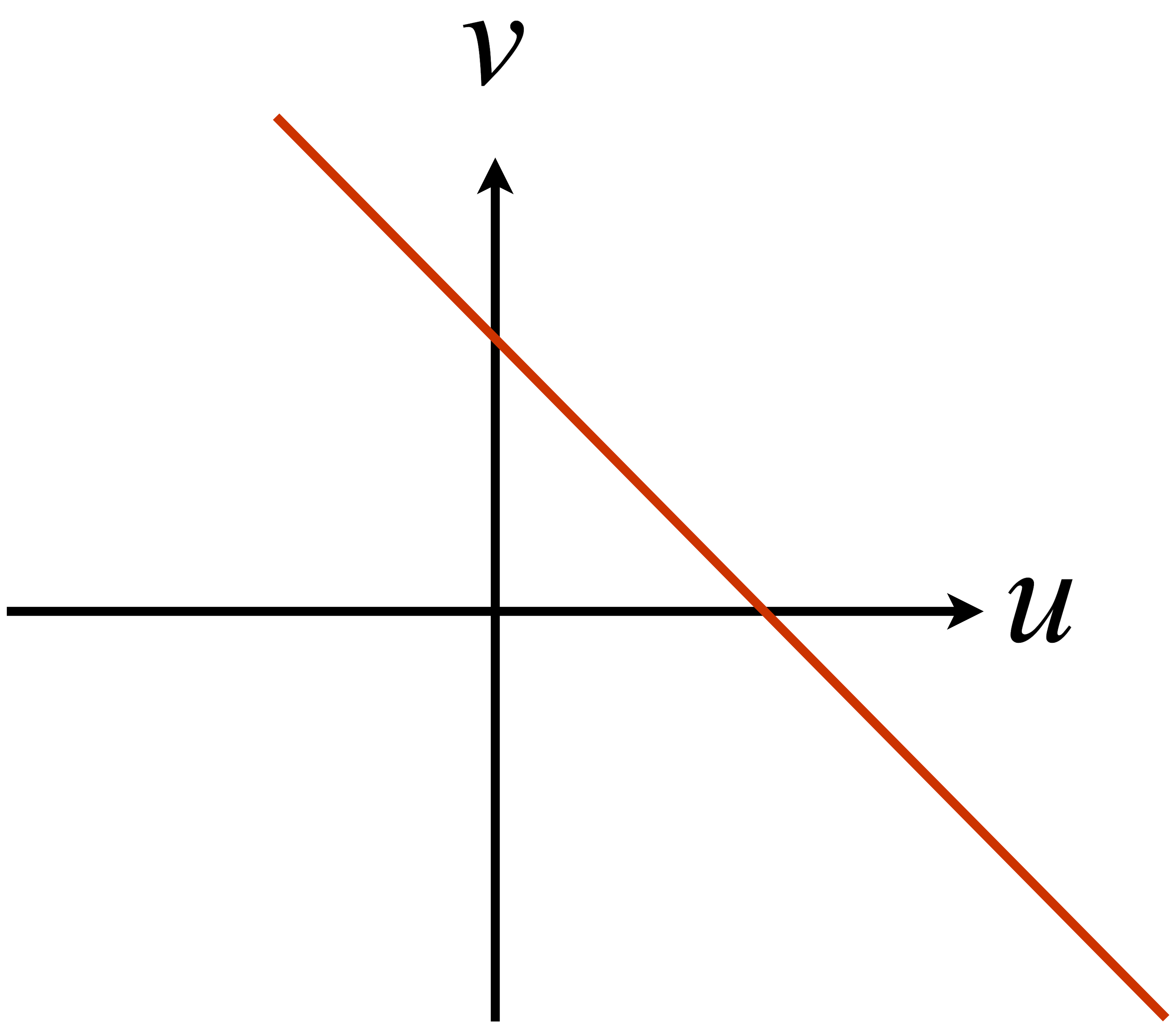
Frame differencing

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Optical Flow **Constraint Equation**

$$\boxed{I_x u + I_y v + I_t = 0}$$

many combinations of u and v will satisfy the equality



Equation determines a **straight line** in velocity space

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Lucas-Kanade

Observations:

1. The 2-D motion, $[u, v]$, at a given point, $[x, y]$, has two degrees-of-freedom

2. The partial derivatives, I_x, I_y, I_t, provide one constraint

3. The 2-D motion, $[u, v]$, cannot be determined locally from I_x, I_y, I_t alone

Lucas-Kanade

Observations:

1. The 2-D motion, $[u, v]$, at a given point, $[x, y]$, has two degrees-of-freedom

2. The partial derivatives, I_x, I_y, I_t, provide one constraint

3. The 2-D motion, $[u, v]$, cannot be determined locally from I_x, I_y, I_t alone

Lucas–Kanade Idea:

Obtain additional local constraint by computing the partial derivatives, I_x, I_y, I_t, in a window centered at the given $[x, y]$

Lucas-Kanade

Observations:

1. The 2-D motion, $[u, v]$, at a given point, $[x, y]$, has two degrees-of-freedom

2. The partial derivatives, I_x, I_y, I_t, provide one constraint

3. The 2-D motion, $[u, v]$, cannot be determined locally from I_x, I_y, I_t alone

Lucas–Kanade Idea:

Obtain additional local constraint by computing the partial derivatives, I_x, I_y, I_t, in a window centered at the given $[x, y]$

Constant Flow Assumption: nearby pixels will likely have same optical flow

Lucas-Kanade

Optical Flow Constraint Equation: $I_x u + I_y v + I_t = 0$

Suppose $[x_1, y_1] = [x, y]$ is the (original) center point in the window. Let $[x_2, y_2]$ be any other point in the window. This gives us two equations that we can write

$$I_{x_1} u + I_{y_1} v = -I_{t_1}$$
$$I_{x_2} u + I_{y_2} v = -I_{t_2}$$

and that can be solved locally for u and v as

$$\begin{bmatrix} u \\ v \end{bmatrix} = -\begin{bmatrix} I_{x_1} & I_{y_1} \\ I_{x_2} & I_{y_2} \end{bmatrix}^{-1} \begin{bmatrix} I_{t_1} \\ I_{t_2} \end{bmatrix}$$

provided that u and v are the same in both equations and provided that the required matrix inverse exists.

Lucas-Kanade

Optical Flow Constraint Equation: $I_x u + I_y v + I_t = 0$

Considering all n points in the window, one obtains

$$
\begin{gathered}
I_{x_1} u + I_{y_1} v = -I_{t_1} \\
I_{x_2} u + I_{y_2} v = -I_{t_2} \\
\vdots \\
I_{x_n} u + I_{y_n} v = -I_{t_n}
\end{gathered}
$$

which can be written as the matrix equation

$$\mathbf{A}\mathbf{v} = \mathbf{b}$$

where $\mathbf{v} = [u, v]^T, \ \mathbf{A} = \begin{bmatrix} I_{x_1} & I_{y_1} \\ I_{x_2} & I_{y_2} \\ \vdots & \vdots \\ I_{x_n} & I_{y_n} \end{bmatrix}$ and $\mathbf{b} = -\begin{bmatrix} I_{t_1} \\ I_{t_2} \\ \vdots \\ I_{t_n} \end{bmatrix}$

Lucas-Kanade

The standard least squares solution, $\bar{\mathbf{v}}$, to is

$$\bar{\mathbf{v}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

again provided that u and v are the same in all equations and provided that the rank of $\mathbf{A}^T\mathbf{A}$ is 2 (so that the required inverse exists)

Lucas-Kanade

Note that we can explicitly write down an expression for $\mathbf{A}^T\mathbf{A}$ as

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & I_y^2 \end{bmatrix}$$

which is identical to the matrix $\mathbf{C}$ that we saw in the context of Harris corner detection

Lucas-Kanade

Note that we can explicitly write down an expression for $\mathbf{A}^T\mathbf{A}$ as

$$\mathbf{A}^T\mathbf{A} = \left[\begin{array}{cc} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & I_y^2 \end{array} \right]$$

which is identical to the matrix $\mathbf{C}$ that we saw in the context of Harris corner detection

What does that mean?

Lucas-Kanade **Summary**

A dense method to compute motion, $[u, v]$ at every location in an image

Key Assumptions:

1. Motion is slow enough and smooth enough that differential methods apply (i.e., that the partial derivatives, I_x, I_y, I_t, are well-defined)

2. The optical flow constraint equation holds (i.e., $\frac{dI(x,y,t)}{dt} = 0$)

3. A window size is chosen so that motion, $[u, v]$, is constant in the window

4. A window size is chosen so that the rank of $\mathbf{A}^T\mathbf{A}$ is 2 for the window

Aside: Optical Flow Smoothness Constraint

Many methods trade off a 'departure from the optical flow constraint' cost with a 'departure from smoothness' cost.

The optimization objective to minimize becomes

$$E = \int\int (I_x u + I_y v + I_t)^2 + \lambda(|| \bigtriangledown u||^2 + || \bigtriangledown v||^2)$$

where λ is a weighing parameter.

Horn-Schunck Optical Flow

$$\min_{\boldsymbol{u},\boldsymbol{v}} \sum_{i,j} \left\{ \overbrace{E_s(i,j)}^{\text{smoothness}} + \overbrace{\underbrace{\lambda}_{\text{weight}} E_d(i,j)}^{\text{brightness constancy}} \right\}$$

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

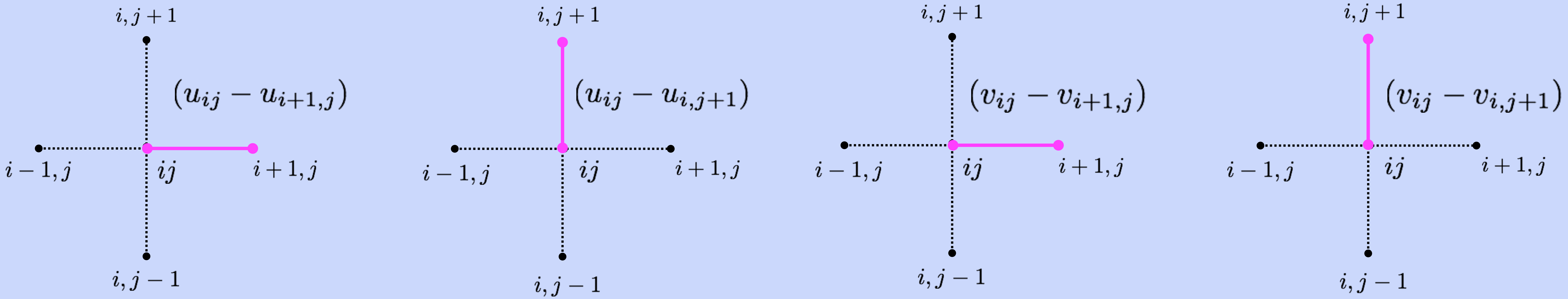
Horn-Schunck Optical Flow

Brightness constancy

$$E_d(i,j) = \left[I_x u_{ij} + I_y v_{ij} + I_t \right]^2$$

Smoothness

$$E_s(i,j) = \frac{1}{4}\left[(u_{ij} - u_{i+1,j})^2 + (u_{ij} - u_{i,j+1})^2 + (v_{ij} - v_{i+1,j})^2 + (v_{ij} - v_{i,j+1})^2 \right]$$



Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Summary

Motion, like binocular stereo, can be formulated as a matching problem. That is, given a scene point located at (x_0, y_0) in an image acquired at time t_0, what is its position, (x_1, y_1), in an image acquired at time t_1?

Assuming image intensity does not change as a consequence of motion, we obtain the (classic) **optical flow constraint equation**

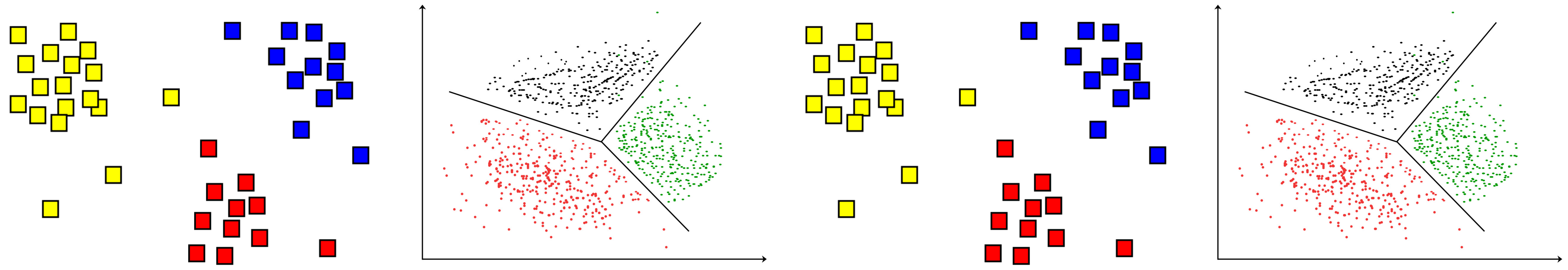
$$I_x u + I_y v + I_t = 0$$

where $[u, v]$, is the 2-D motion at a given point, $[x, y]$, and I_x, I_y, I_t are the partial derivatives of intensity with respect to x, y, and t

Lucas–Kanade is a dense method to compute the motion, $[u, v]$, at every location in an image

THE UNIVERSITY OF BRITISH COLUMBIA

CPSC 425: Computer Vision



Lecture 20: Classification

Classification

Problem:

Assign new observations into one of a fixed set of categories (classes)

Key Idea(s):

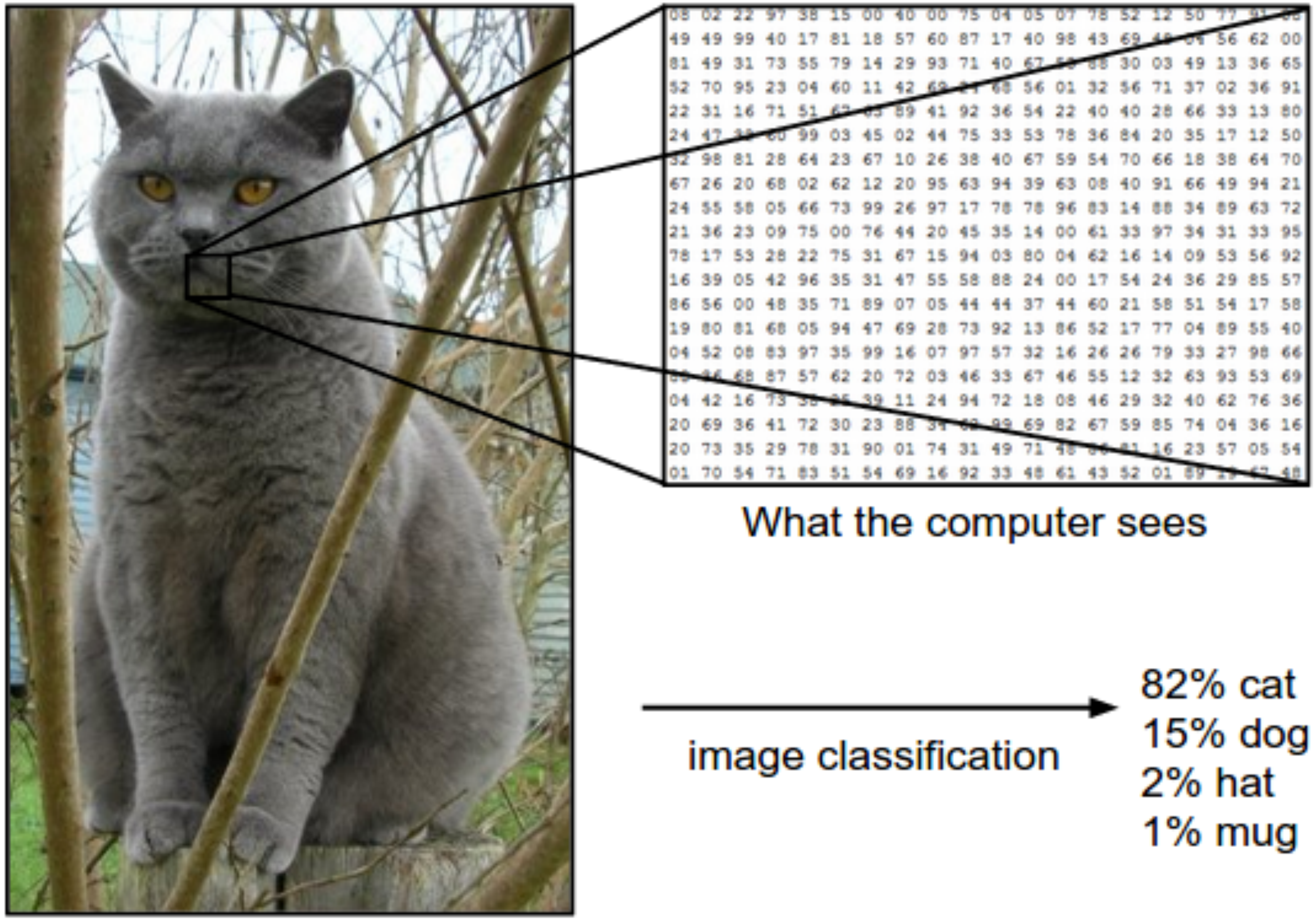
Build a model of data in a given category based on observations of instances in that category

Classification

(assume given set of discrete labels)
{dog, cat, truck, plane, ...}

→ cat

Classification



Classification

A **classifier** is a procedure that accepts as input a set of features and outputs a class **label**

Classifiers can be binary (face vs. not-face) or multi-class (cat, dog, horse, ...).

We build a classifier using a **training set** of labelled examples $\{(\mathbf{x}_i, y_i)\}$, where each $\mathbf{x}_i$ is a feature vector and each y_i is a class label.

Given a previously unseen observation, we use the classifier to predict its class label.

Classification

- Collect a database of images with labels
- Use ML to train an image classifier
- Evaluate the classifier on test images

Label

Feature vector computed from the image

Example 1: A Classification Problem

Categorize images of fish
— “Atlantic salmon” vs “Pacific salmon”

Use **features** such as length, width, lightness, fin shape & number, mouth position, etc.

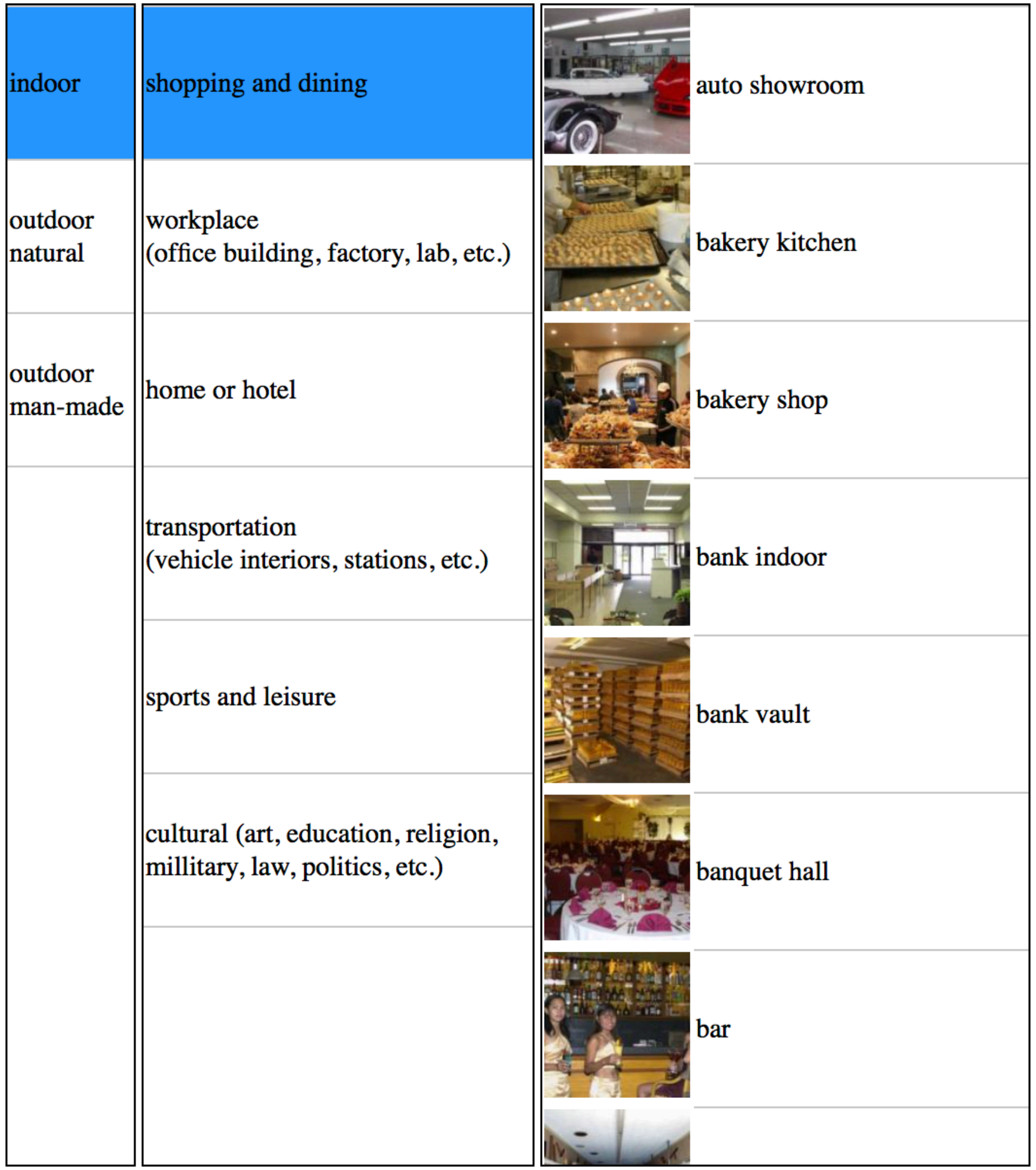
Given a previously unobserved image of a salmon, use the learned classifier to guess whether it is an Atlantic or Pacific salmon

Figure credit: Duda & Hart

Example 2: Real Classification Problem

SUN Dataset

- 131K images
- 908 **scene** categories

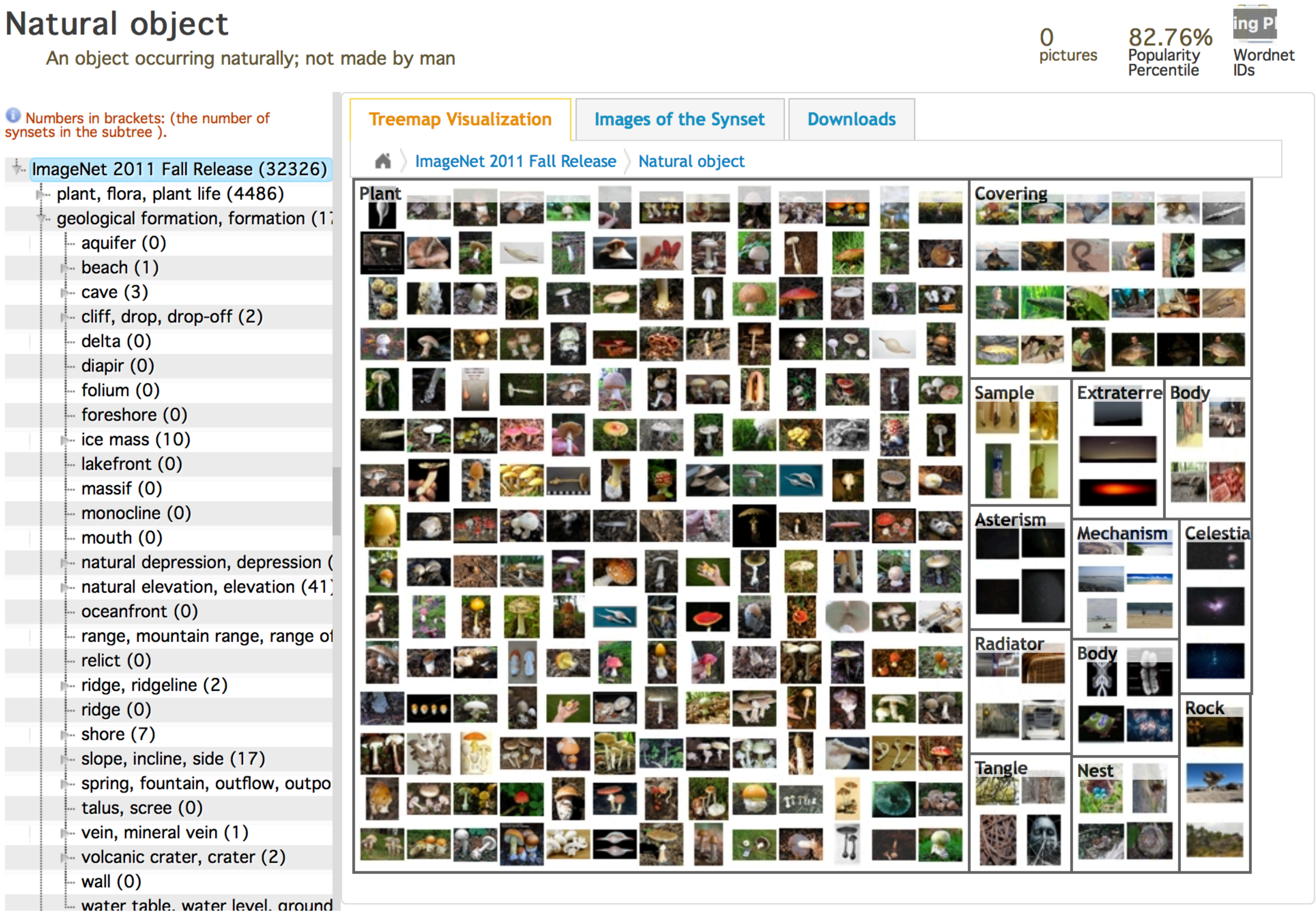


Example 3: Real Classification Problem

ImageNet Dataset

- 14 Million images

- 21K **object** categories



Bayes Rule (Review and Definitions)

Let c be the class label and let x be the measurement (i.e., evidence)

class–conditional probability (a.k.a. likelihood)

prior probability

$$\underbrace{P(c|x)}_{\text{posterior probability}} = \frac{P(x|c)\,p(c)}{P(x)}$$

unconditional probability (a.k.a. marginal likelihood)

Bayes Rule (Review and Definitions)

Let c be the **class label** and let x be the **measurement** (i.e., evidence)

Simple case:

— binary classification; i.e., $c \in \{1, 2\}$

— features are 1D; i.e., $x \in \mathbb{R}$

$$P(c|x) = \frac{P(x|c)p(c)}{P(x)}$$

General case:

— multi-class; i.e., $c \in \{1, ..., 1000\}$

— features are high-dimensional; i.e., $x \in \mathbb{R}^{2,000+}$

Example: Discrete Bayes Classifier

Assume we have two classes: $c_1 = \textbf{male}$ $c_2 = \textbf{female}$

We have a person who's gender we don't know, who's name is *drew*

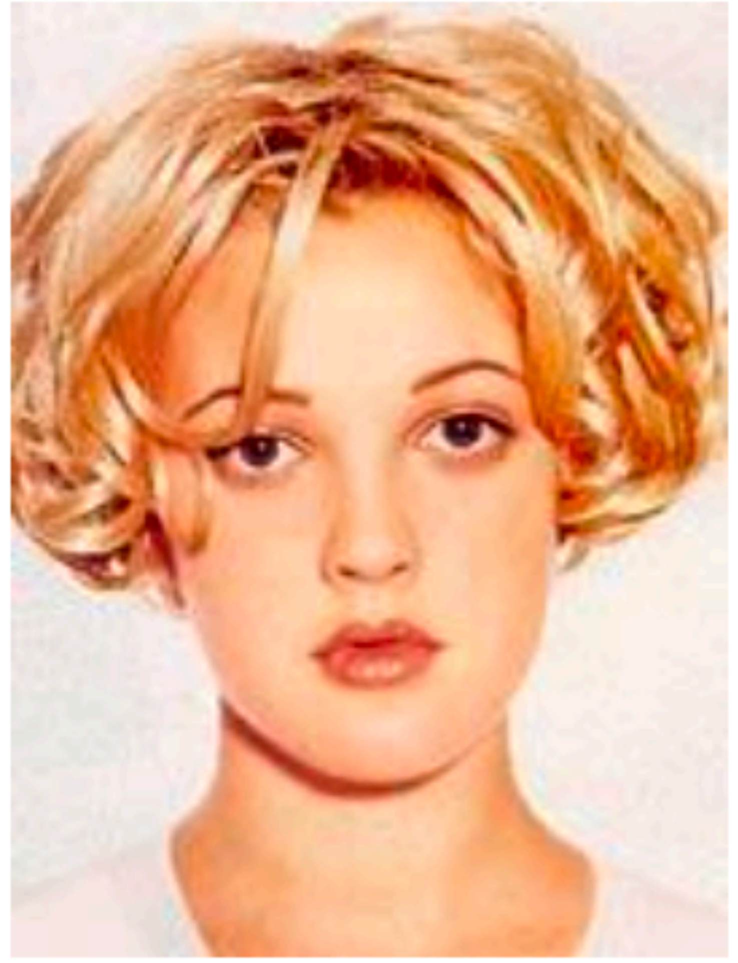
Example from: Eamonn Keogh

Example: Discrete Bayes Classifier

Assume we have two classes: $c_1 = \textbf{male}$ $c_2 = \textbf{female}$

We have a person who's gender we don't know, who's name is *drew*

Drew Carey



Drew Barrymore

Example from: Eamonn Keogh

Example: Discrete Bayes Classifier

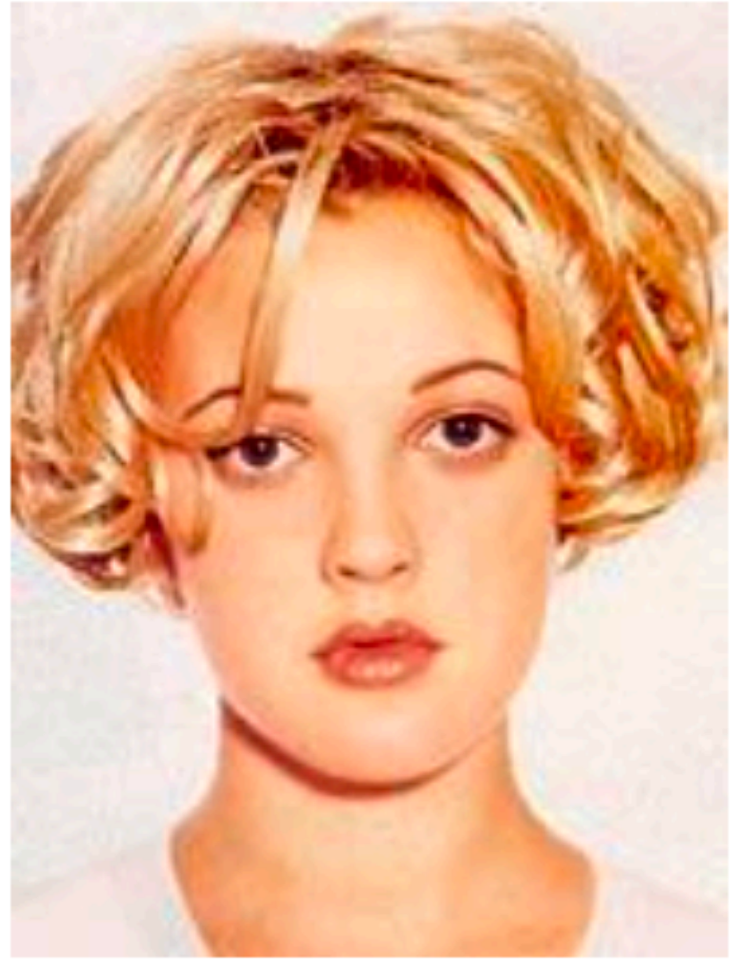
Assume we have two classes: $c_1 = \textbf{male}$ $c_2 = \textbf{female}$

We have a person who's gender we don't know, who's name is *drew*

Classifying drew as being male or female is equivalent to asking is it more probable that *drew* is male or female, i.e. which is greater $p(\textbf{male}|drew)$

$p(\textbf{female}|drew)$

Drew Carey



Drew Barrymore

Example from: Eamonn Keogh

Example: Discrete Bayes Classifier

Assume we have two classes: $c_1 = \textbf{male}$ $c_2 = \textbf{female}$

We have a person who's gender we don't know, who's name is *drew*

Classifying drew as being male or female is equivalent to asking is it more probable that *drew* is male or female, i.e. which is greater $p(\textbf{male}|drew)$

$p(\textbf{female}|drew)$

$$p(\textbf{male}|drew) = \frac{p(drew|\textbf{male})p(\textbf{male})}{p(drew)}$$

Example from: Eamonn Keogh

Example: Discrete Bayes Classifier

Name	Gender
Drew	Male
Claudia	Female
Drew	Female
Drew	Female
Alberto	Male
Karin	Female
Nina	Female
Sergio	Male

$$p(\mathbf{male}|drew) = \frac{p(drew|\mathbf{male})p(\mathbf{male})}{p(drew)}$$

Example from: Eamonn Keogh

Example: Discrete Bayes Classifier

$$p(\mathbf{male}) =$$

$$p(drew|\mathbf{male}) =$$

$$p(drew) =$$

$$p(\mathbf{male}|drew) = \frac{p(drew|\mathbf{male})p(\mathbf{male})}{p(drew)}$$

Name	Gender
Drew	Male
Claudia	Female
Drew	Female
Drew	Female
Alberto	Male
Karin	Female
Nina	Female
Sergio	Male

Example from: Eamonn Keogh

Example: Discrete Bayes Classifier

Name	Gender
Drew	Male
Claudia	Female
Drew	Female
Drew	Female
Alberto	Male
Karin	Female
Nina	Female
Sergio	Male

$$p(\mathbf{male}) = \frac{3}{8}$$

$$p(drew|\mathbf{male}) =$$

$$p(drew) =$$

$$p(\mathbf{male}|drew) = \frac{p(drew|\mathbf{male})p(\mathbf{male})}{p(drew)}$$

Example from: Eamonn Keogh

Example: Discrete Bayes Classifier

$$p(\mathbf{male}) = \frac{3}{8}$$

$$p(drew|\mathbf{male}) = \frac{1}{3}$$

$$p(drew) =$$

$$p(\mathbf{male}|drew) = \frac{p(drew|\mathbf{male})p(\mathbf{male})}{p(drew)}$$

Name	Gender
Drew	Male
Claudia	Female
Drew	Female
Drew	Female
Alberto	Male
Karin	Female
Nina	Female
Sergio	Male

Example from: Eamonn Keogh

Example: Discrete Bayes Classifier

Name	Gender
Drew	Male
Claudia	Female
Drew	Female
Drew	Female
Alberto	Male
Karin	Female
Nina	Female
Sergio	Male

$$p(\mathbf{male}) = \frac{3}{8}$$

$$p(drew|\mathbf{male}) = \frac{1}{3}$$

$$p(drew) = \frac{3}{8}$$

$$p(\mathbf{male}|drew) = \frac{p(drew|\mathbf{male})p(\mathbf{male})}{p(drew)}$$

Example from: Eamonn Keogh

Example: Discrete Bayes Classifier

Name	Gender
Drew	Male
Claudia	Female
Drew	Female
Drew	Female
Alberto	Male
Karin	Female
Nina	Female
Sergio	Male

$$p(\mathbf{male}) = \frac{3}{8}$$

$$p(drew|\mathbf{male}) = \frac{1}{3}$$

$$\cancel{p(drew) = \frac{3}{8}}$$

$$p(\mathbf{male}|drew) = \frac{p(drew|\mathbf{male})p(\mathbf{male})}{\cancel{p(drew)}} = 0.125$$

Example from: Eamonn Keogh

Example: Discrete Bayes Classifier

Name	Gender
Drew	Male
Claudia	Female
Drew	Female
Drew	Female
Alberto	Male
Karin	Female
Nina	Female
Sergio	Male

$$p(\textbf{male}) = \frac{3}{8} \qquad p(\textbf{female}) = \frac{5}{8}$$

$$p(drew|\textbf{male}) = \frac{1}{3} \qquad p(drew|\textbf{female}) = \frac{2}{5}$$

$$\cancel{p(drew) = \frac{3}{8}}$$

$$p(\textbf{male}|drew) = \frac{p(drew|\textbf{male})p(\textbf{male})}{\cancel{p(drew)}} = 0.125$$

$$p(\textbf{female}|drew) = \frac{p(drew|\textbf{female})p(\textbf{female})}{\cancel{p(drew)}} = 0.25$$

Example from: Eamonn Keogh

Bayes Rule (Review and Definitions)

Let c be the **class label** and let x be the **measurement** (i.e., evidence)

Simple case:

— binary classification; i.e., $c \in \{1, 2\}$

— features are 1D; i.e., $x \in \mathbb{R}$

$$P(c|x) = \frac{P(x|c)p(c)}{P(x)}$$

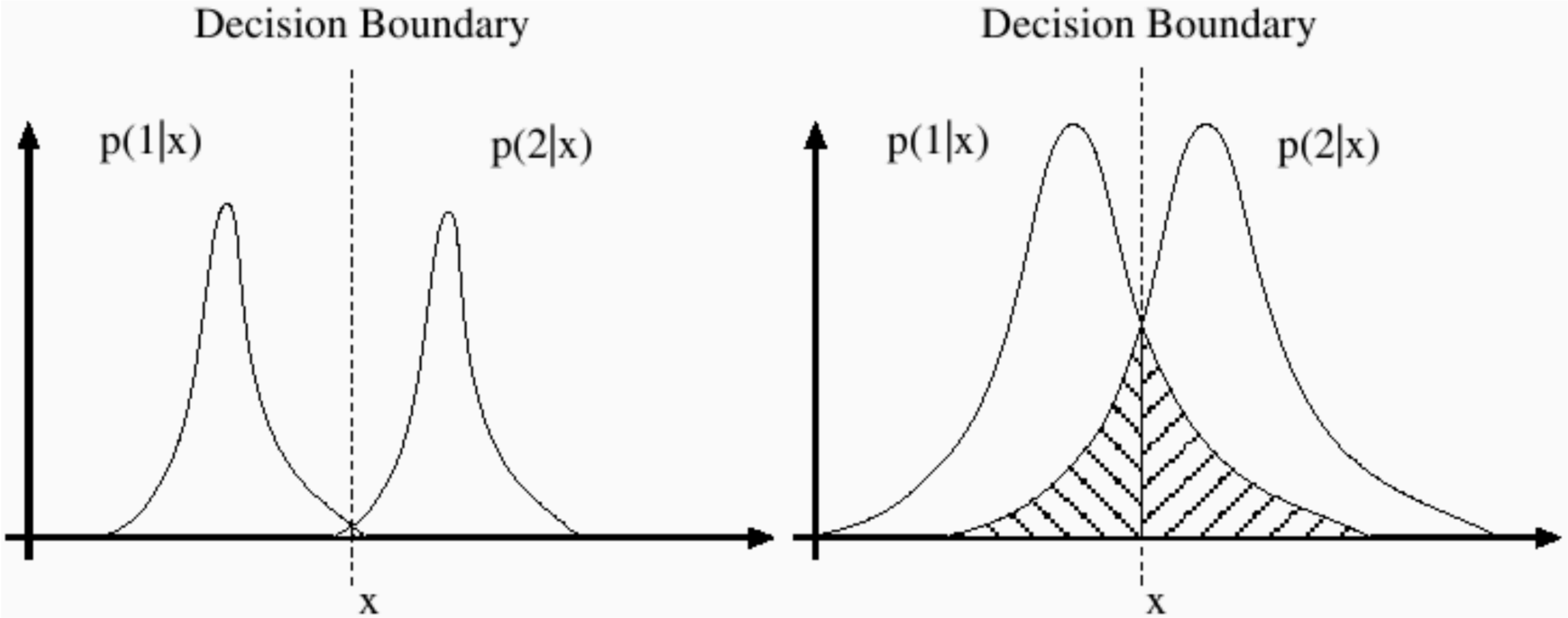
General case:

— multi-class; i.e., $c \in \{1, ..., 1000\}$

— features are high-dimensional; i.e., $x \in \mathbb{R}^{2,000+}$

Bayes’ Risk

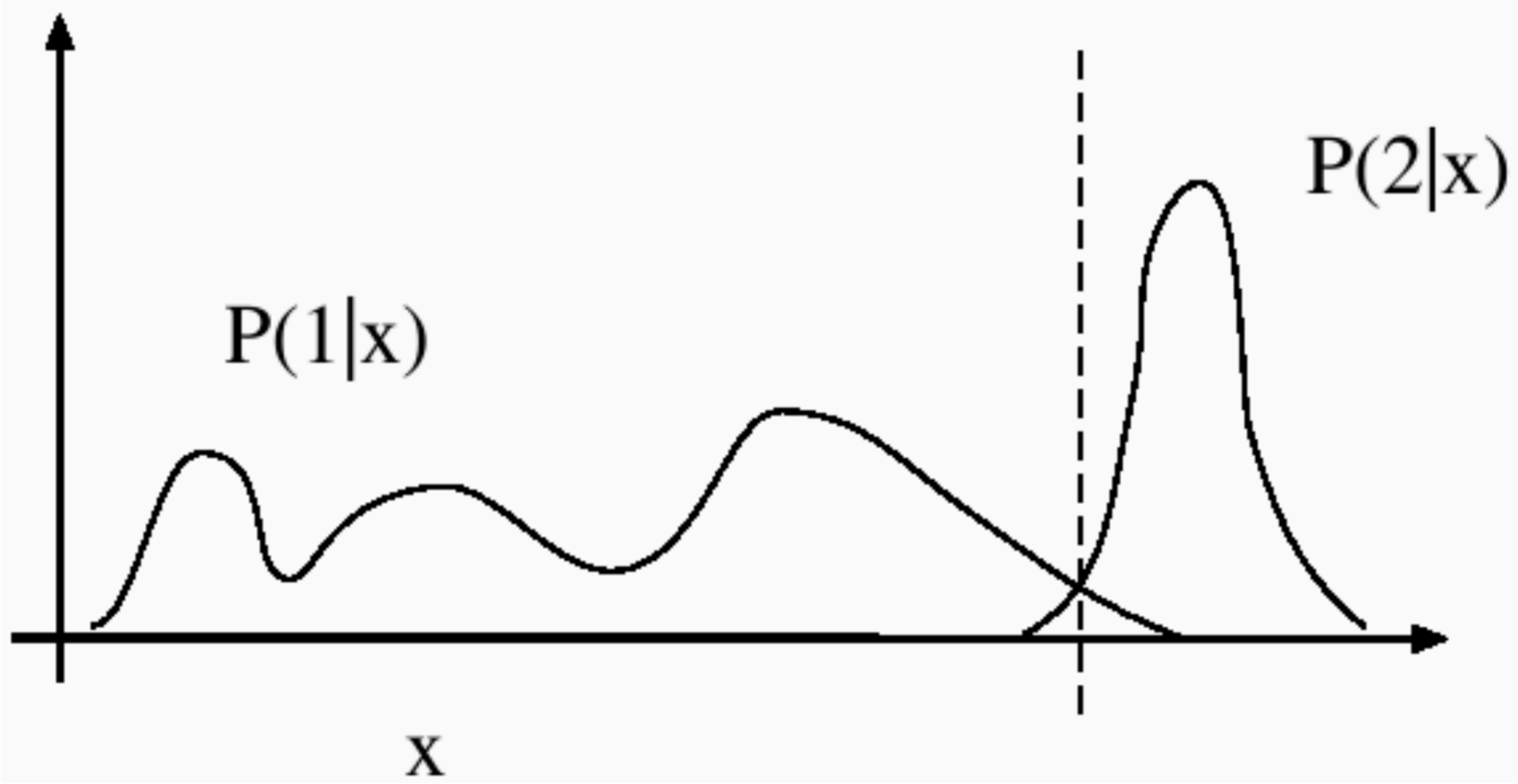
Some errors may be inevitable: the minimum risk (shaded area) is called the **Bayes’ risk**



Forsyth & Ponce (2nd ed.) Figure 15.1

Discriminative vs. **Generative**

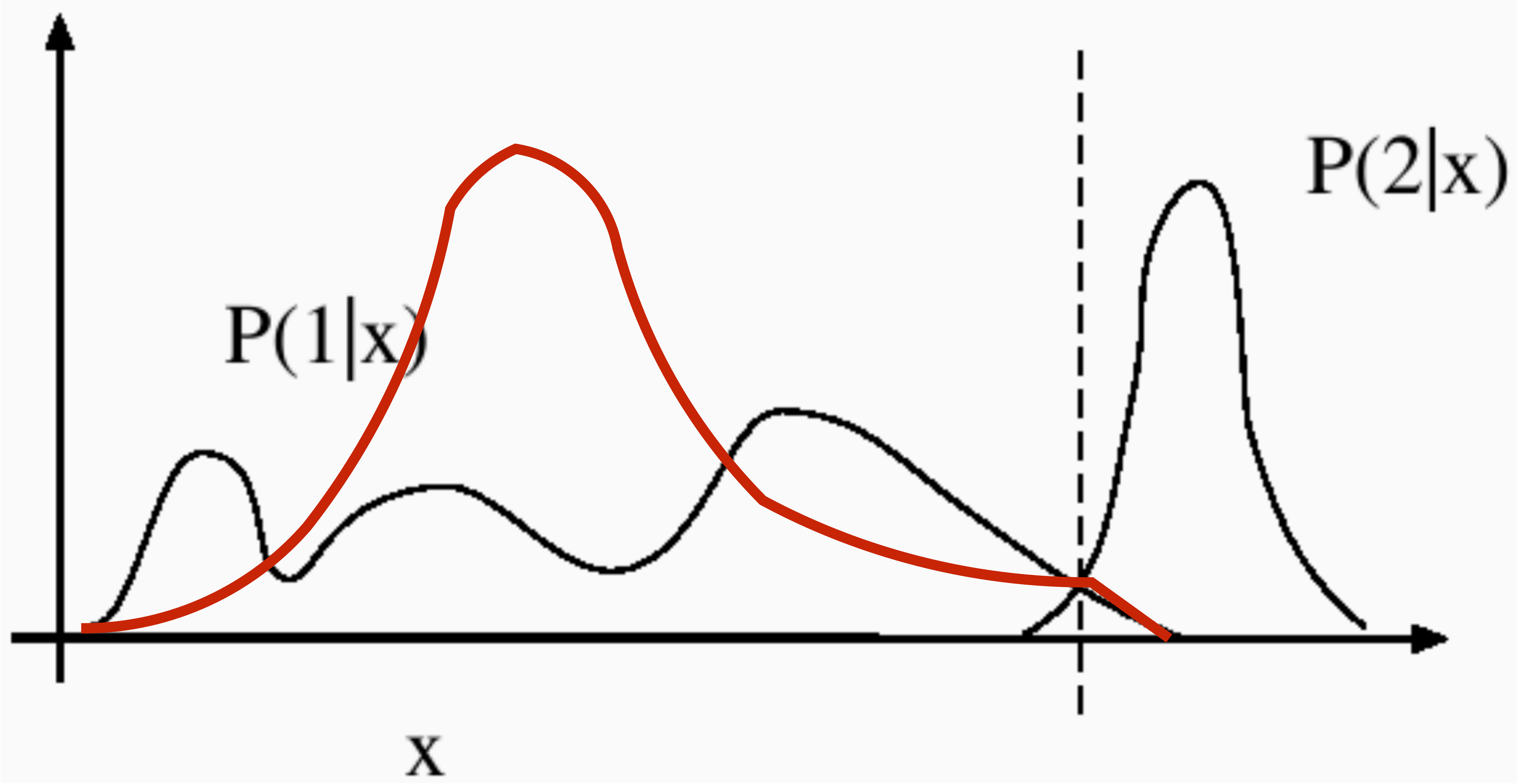
Finding a **decision boundary** is not the same as modeling a **conditional density** — while a normal density here is a poor fit to P(1|x), the quality of the classifier depends only on how well the boundary is positioned



Forsyth & Ponce (2nd ed.) Figure 15.5

Discriminative vs. **Generative**

Finding a **decision boundary** is not the same as modeling a **conditional density** — while a normal density here is a poor fit to P(1|x), the quality of the classifier depends only on how well the boundary is positioned



Forsyth & Ponce (2nd ed.) Figure 15.5