9.913 Pattern Recognition for Vision

Class I - Overview

Instructors: B. Heisele, Y. Ivanov, T. Poggio

TOC

- Administrivia
- Problems of Computer Vision and Pattern Recognition
- Overview of classes
- Quick review of Matlab

Administrivia

- Instructors:
  - Bernd Heisele
  - Yuri Ivanov
  - Tomaso Poggio
- Meet
  - Th 10-12
- Credits: 10H
- Assignments:
  - Small weekly
  - Paper presentation/discussion
  - Final project

Syllabus

Sep 9 - Overview, Introduction	*Y&B*
Sep 16 - Vision - Image formation and processing	*Y*
Sep 23 - Vision – Feature extraction I	*B*
Sep 30 - PR/Vis - Feature Extraction II/Bayesian decisions	*B&Y*
Oct 7 - PR - Density estimation	*Y papers*
Oct 14 - PR – Clasification	*B*
Oct 21 - Biological Object Recognition	*T*
Oct 28 - PR - Clustering	*Y&B proj*
Nov 4 - Paper Discussion	*All*
Nov 11 - App I - Object Detection/Recognition	*B*
Nov 18 - App II - Morphable models	*T&B*
Nov 25 - No class - Thanksgiving day	
Dec 2 - App III - Tracking	*C&Y*
Dec 9 - App IV - Gesture and Action Recognition	*Y*
Dec 16 - Project presentation	*All*

Course Materials

- Books
  - Duda, Hart and Stork, *Pattern Recognition*
  - Optional - Mallot, *Computational Vision: Information Processing in Perception and Visual Behavior*
  - *Suggested further reading*
    - *Vision: Forsyth, Ponce, "Computer Vision: a Modern Approach"*
    - *Machine Learning: Hastie, Tibshirani, Friedman, "The Elements of Statistical Learning"*
- Slides, links
- Papers, notes, tutorials
- Office hours
  - No set hours – e-mail, or call

Problems of Computer Vision

- Shape from Shading
- Stereo
- Structure from Motion
- Tracking
- Object Detection/Recognition
- Activity Detection/Recognition
- … many more….

Machine Vision - Applications

- Automatic quality control
- Robotics
- Perceptual interfaces – human/machine interactions
- Surveillance
- …

Images removed due to copyright considerations.

Computer Vision – Shape-from-Shading

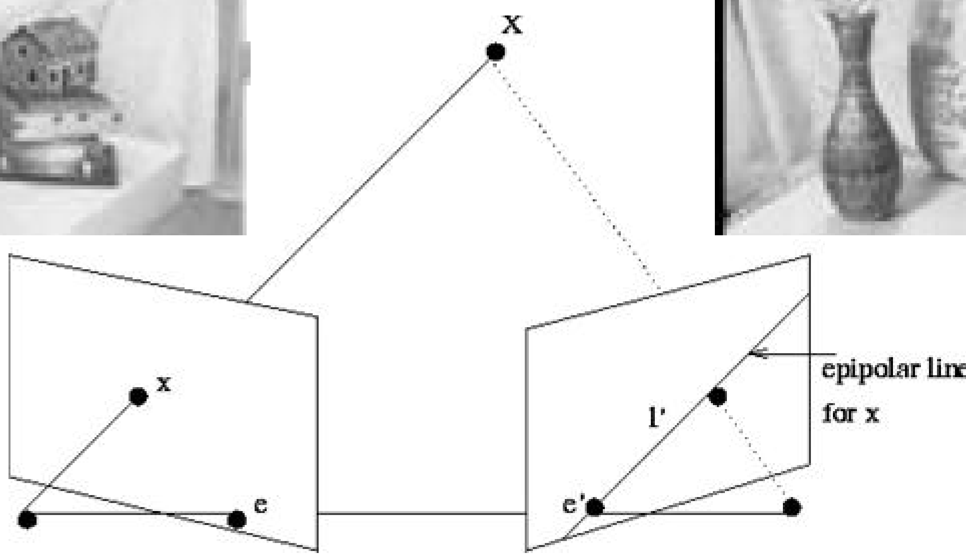
Find 3D surface parameters from a single image based on shading

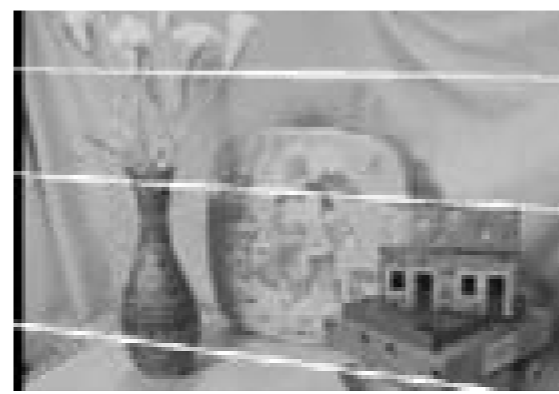
Images removed due to copyright considerations.

Images: NASA

Computer Vision - Stereo

Reconstruct 3D surface from 2 2D images taken simultaneously





Images: A. Zisserman

Images by A. Zisserman and P. Beardsley. Used with permission.

Reconstruct 3D surface from 2D images taken from a moving camera

Images ———

Structure ———

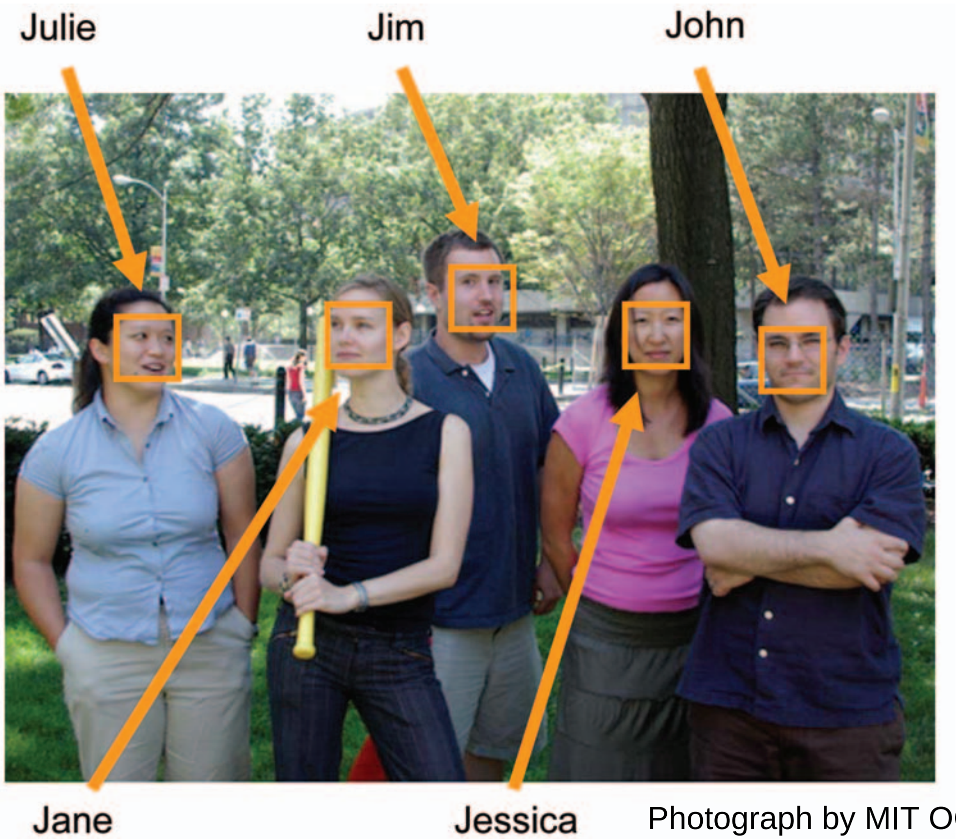
Texture ———

Textured model ———

Photos removed due to copyright considerations. Please see Figure 9 in: Azarbayejani, and Pentland. "Recursive Estimation of Motion, Structure, and Focal Length." IEEE Transactions on Pattern Analysis and Machine Intelligence 7, no. 6 (June 1995): 562-575.

Find objects in the image, determine what they are

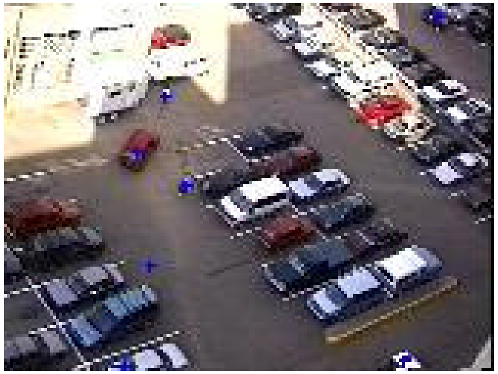
Eg: Face detection and recognition:



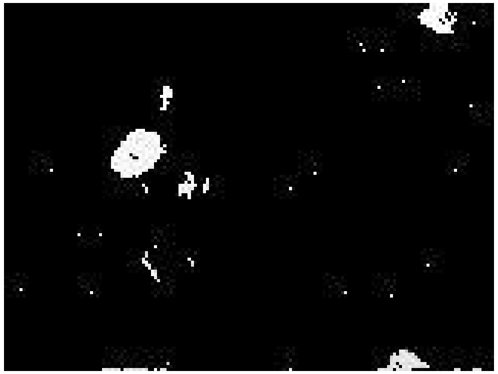
Photograph by MIT OCW.

Computer Vision - Tracking

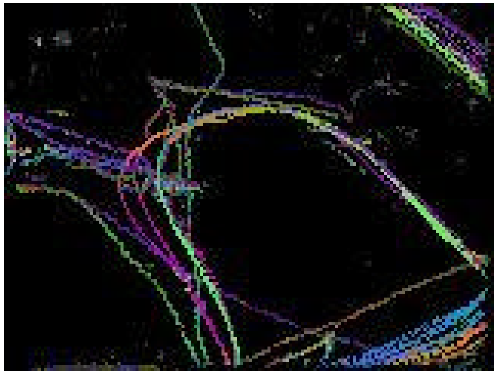
Determine objects' positions over multiple frames



Camera view



Connected components



Trajectories over time

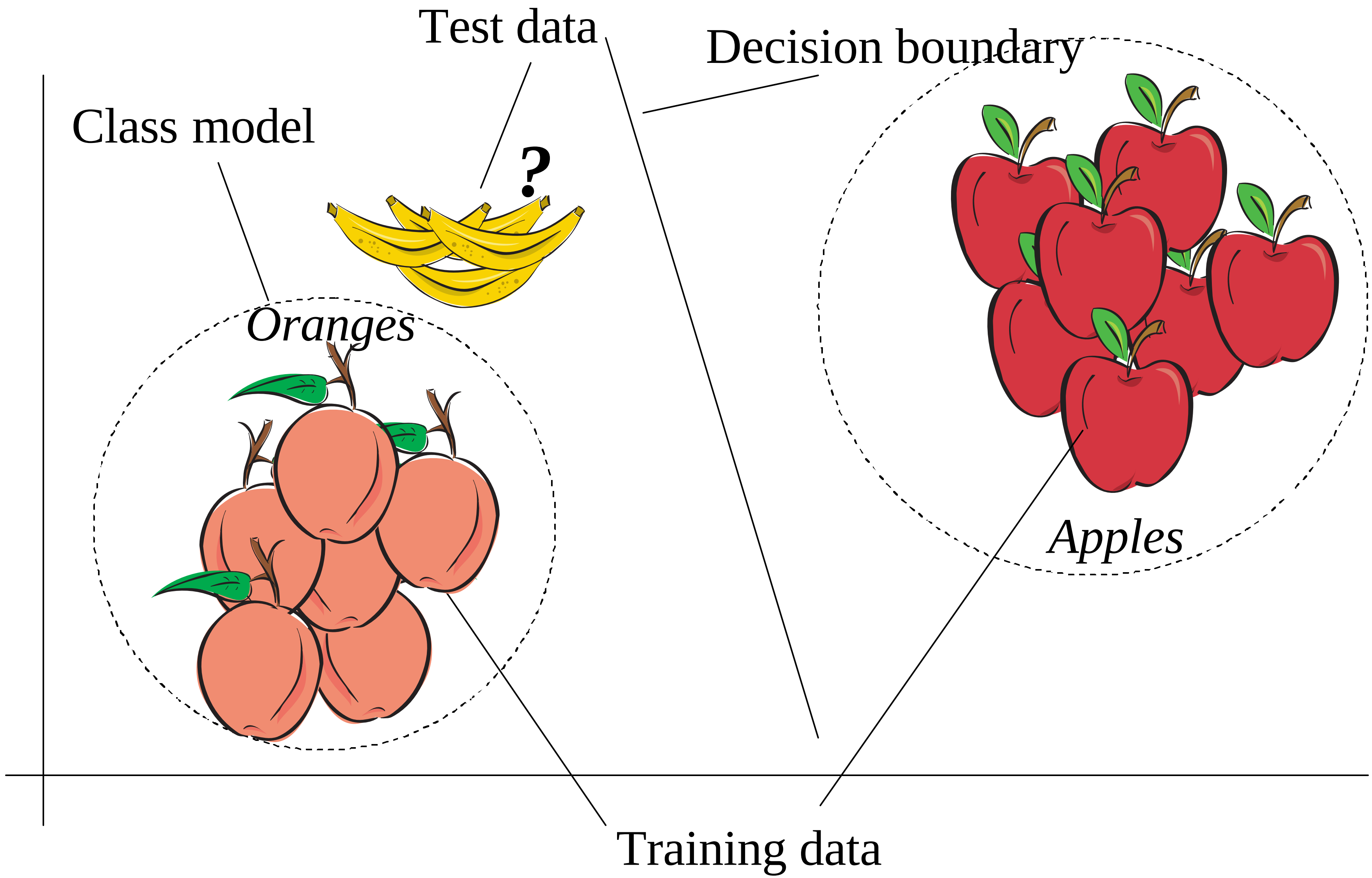
An object

Tracker – Chris Stauffer, Eric Grimson

“Big Four” Problems of Machine Learning

- Classification
- Density Estimation
- Clustering
- Regression

Classification



Images by MIT OCW.

Density Estimation

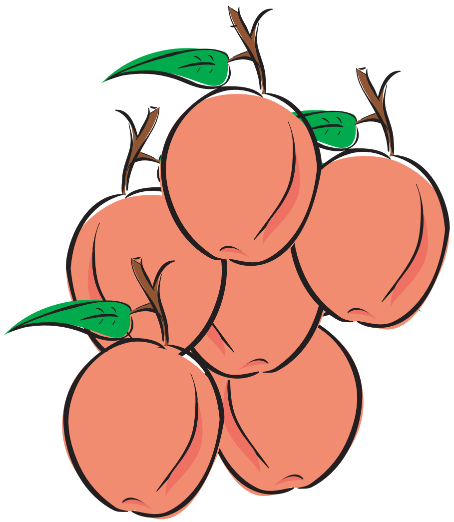
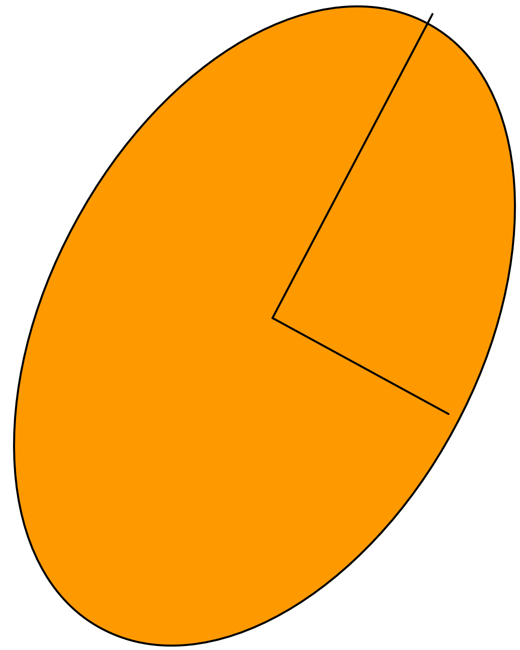


Image by MIT OCW.

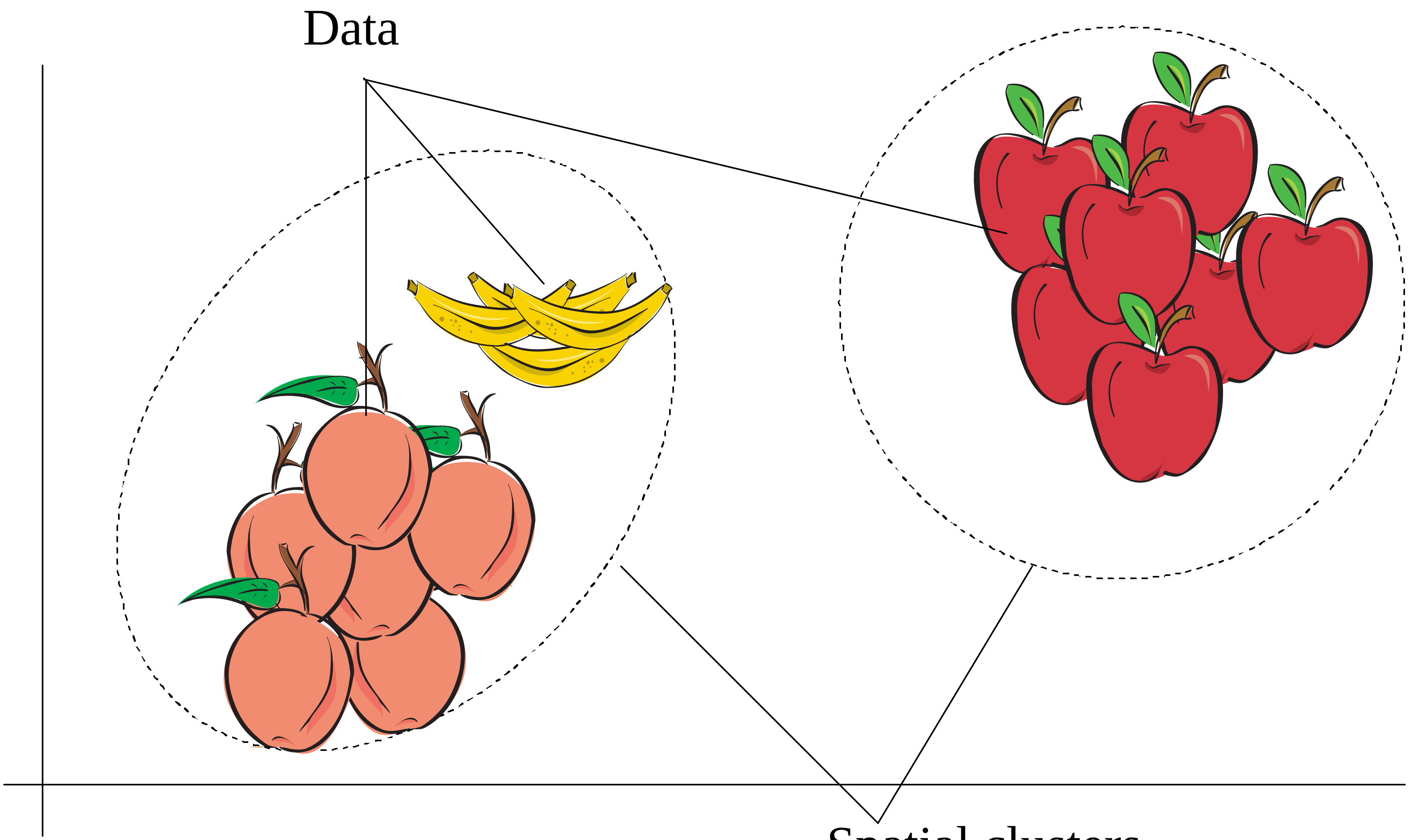


μ, Σ

Individual samples

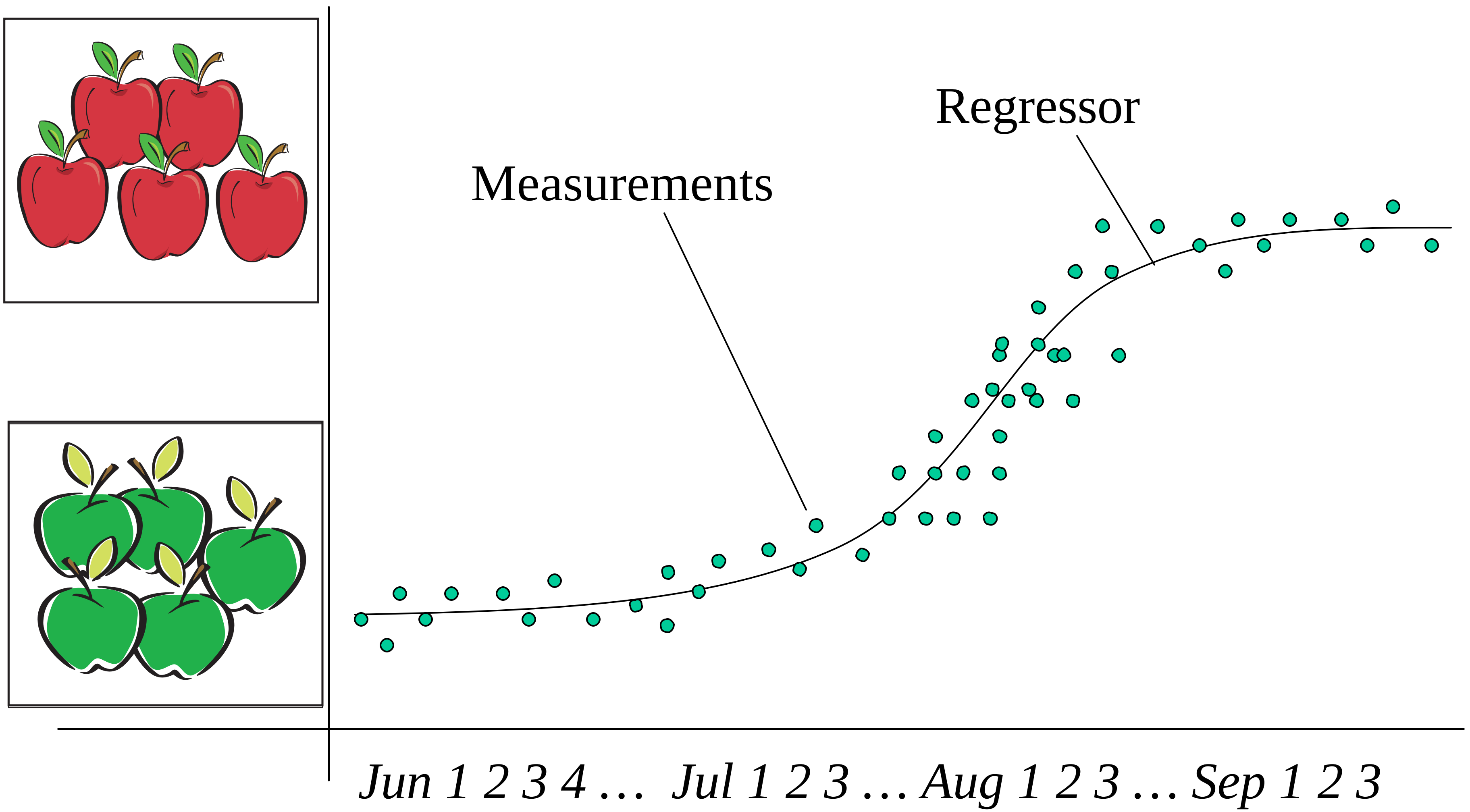
Generating density

Clustering



Images by MIT OCW.

Regression



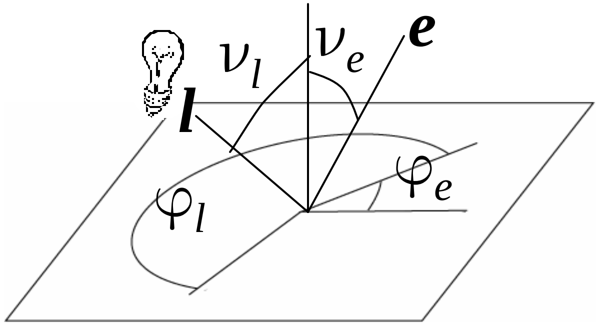
Images by MIT OCW.

Topics

- Image formation
- Color spaces
- Point operators
- Neighborhood operators
- Edge detection
- Motion

Vision - Fundamentals

- Image formation



- Source properties:
  - Type of source
  - Occlusions / shadows
  - Direction
- Surface properties:
  - Albedo (fraction of light reflected),
  - Shape,
  - Smoothness,
  - Orientation
- Imager:
  - Color properties
  - Distortion
  - Focal length
  - F-stop (how wide the iris)
  - Orientation
- Representation:
  - Bit depth
  - Color space

Vision - Color

- Light is perceived by two types of receptors in the human eye
  - Rods
  - Cones
- Color is perceived by 3 types of cones
  - “Red” (64%)
  - “Green” (34%)
  - “Blue” (2%)

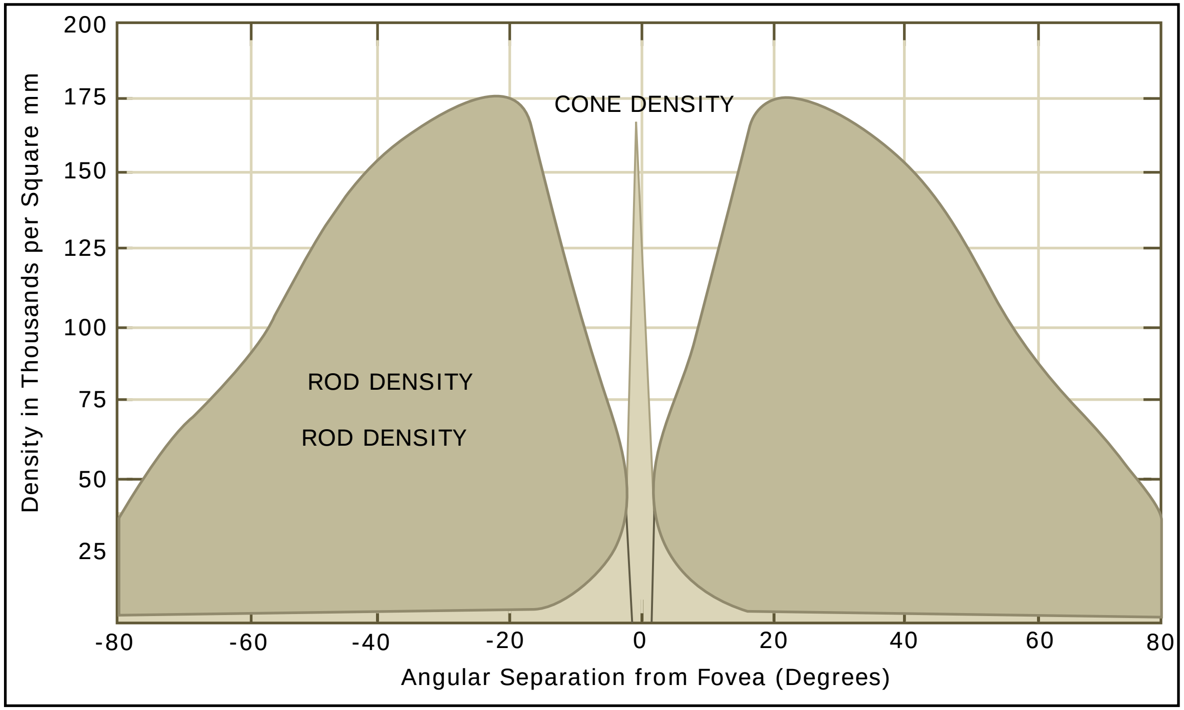


Figure by MIT OCW.

120,000,000 rods
6,000,000 cones

All Visible Colors (CIE 1931 Chromaticity diagram):

Figure removed due to copyright reasons. Please see: http://www.cie.co.at/cie/
and
http://www.ledtronics.com/datasheets/Pages/chromaticity/097b.htm

Point Operators – Image Arithmetic

Point operators query and affect only a single pixel at a time

Thresholding of a very famous person:

Original photo and accompanying processed photo both removed due to copyright reasons.

Topics

- Intro to Pattern Recognition and Machine Learning
- Bayes rule
- Normal density
- Minimum error rate classification
- Decision surfaces
- ROC curves and classifier performance

Bayes Rule:

Posterior *Likelihood* *Prior*

$$\boxed{P(w \mid x)} = \frac{\boxed{P(x \mid w)}\,\boxed{P(w)}}{\boxed{P(x)}}$$

Evidence

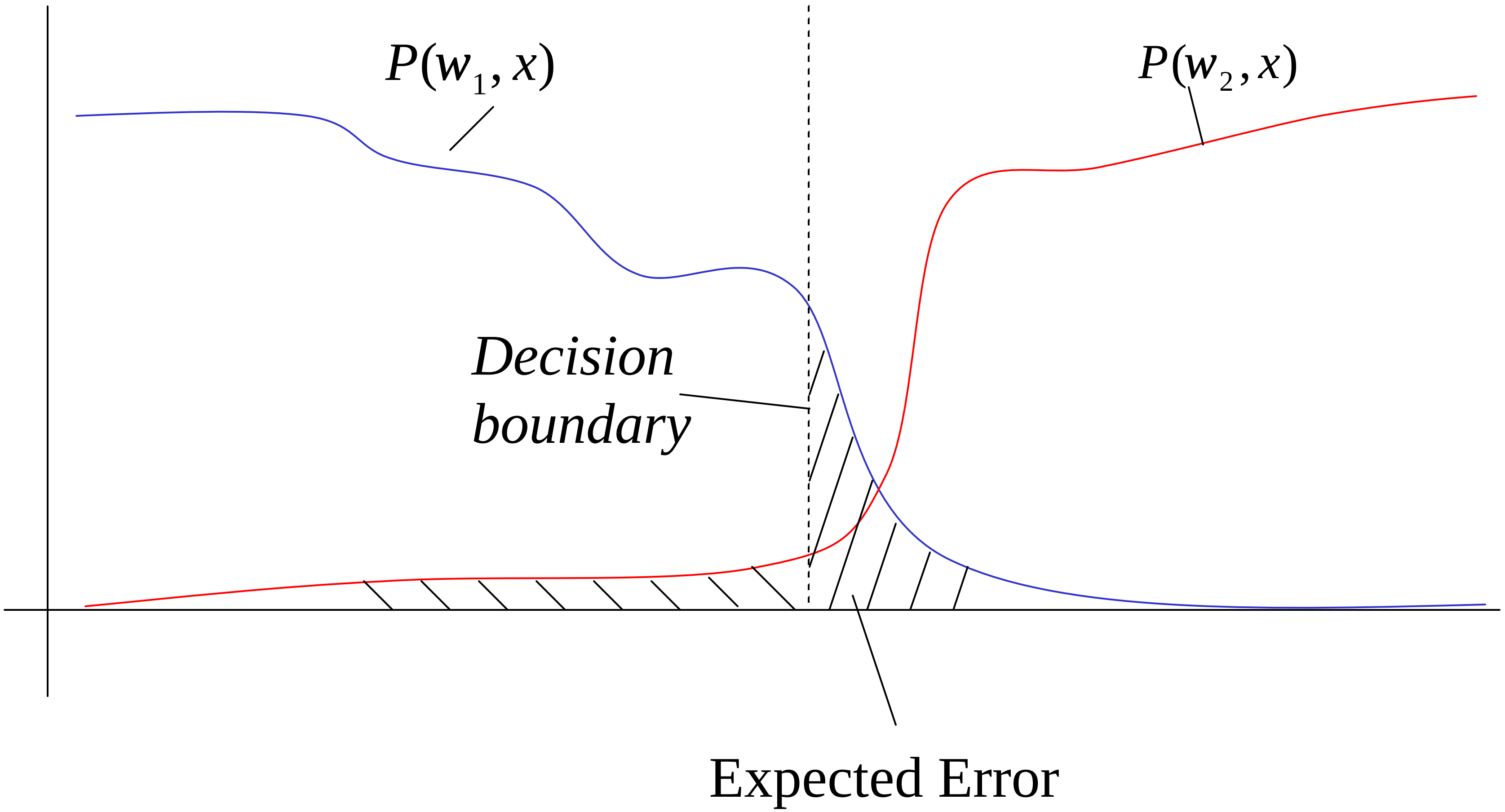
Image removed due to copyright considerations.

Rev. Thomas Bayes
(1702-1761)

Fundamental Law of Probability and Statistics

Expected Error

How do we compare two class models given the data?



How do we compare two class models given the data?

$$\frac{P(w_1 \mid x)}{P(w_2 \mid x)} <> 1 \quad \Rightarrow$$

$$\frac{P(x \mid w_1) P(w_1)}{P(x \mid w_2) P(w_2)} <> 1 \quad \Rightarrow$$

$$\frac{P(x \mid w_1)}{P(x \mid w_2)} <> \frac{P(w_2)}{P(w_1)}$$

Likelihood Ratio Test

Topics

- Parametric density estimation
    - ML parameter estimation
    - Bayesian parameter estimation
- Non-parametric
    - K-Nearest Neighbors
    - Parzen Windows

Density Estimation

Say we need to teach a robot about oranges

Main feature: color

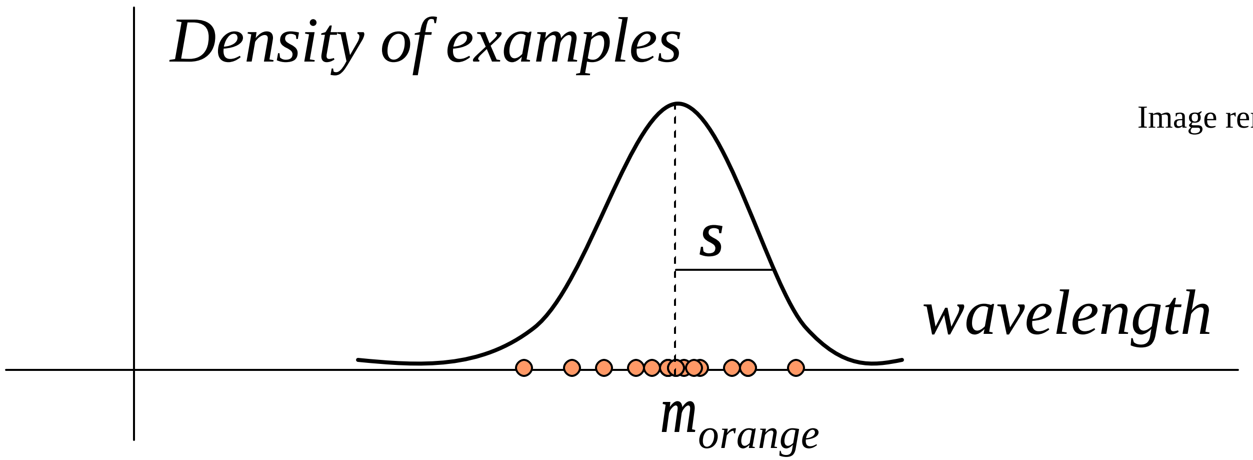


Image removed due to copyright considerations.

Assume a Gaussian: $P(w_i \mid m, s) = \dfrac{1}{\sqrt{2p}\, s} e^{-\frac{1}{2}\left(\frac{w_i - m}{s}\right)^2}$

Need μ and σ to maximize $P(w_1\ w_2\ w_3 \ldots w_N \mid m, s)$

Solution: factor, log, differentiate and set to 0

Density Estimation

Say we need to teach a robot about apples

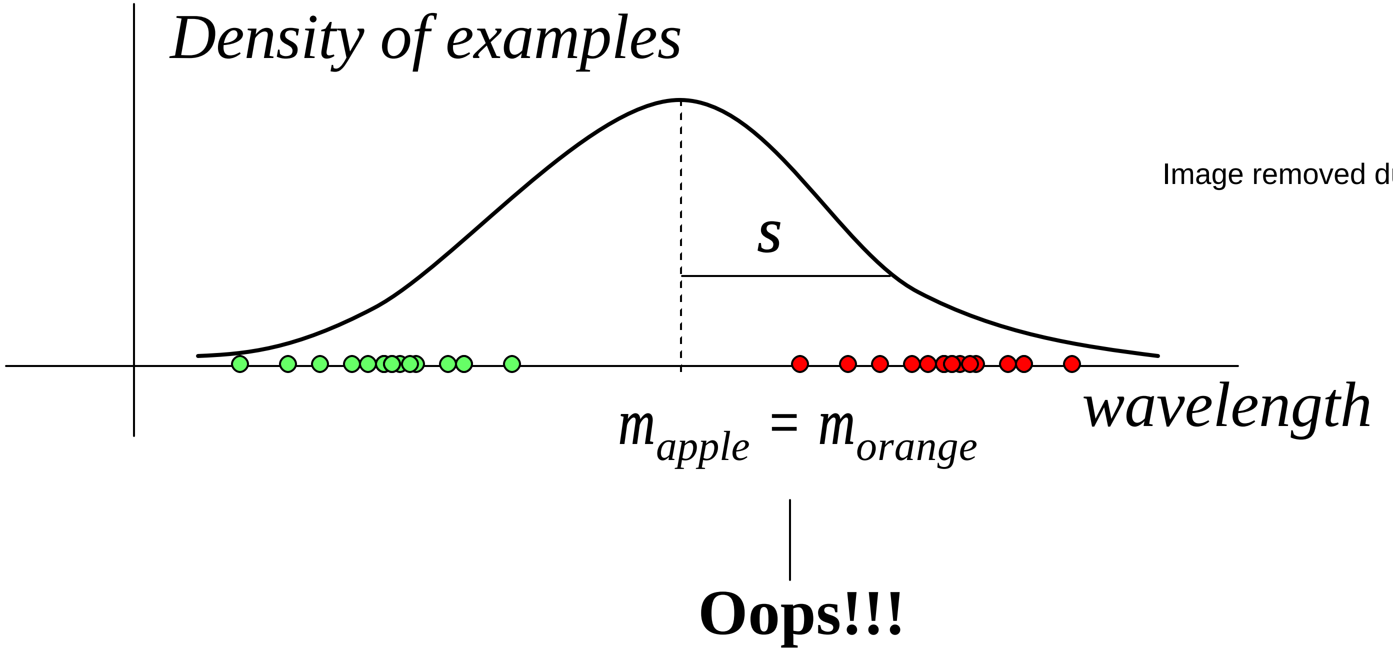


Image removed due to copyright considerations.

Need a multi-modal distribution, or a *non-parametric* method

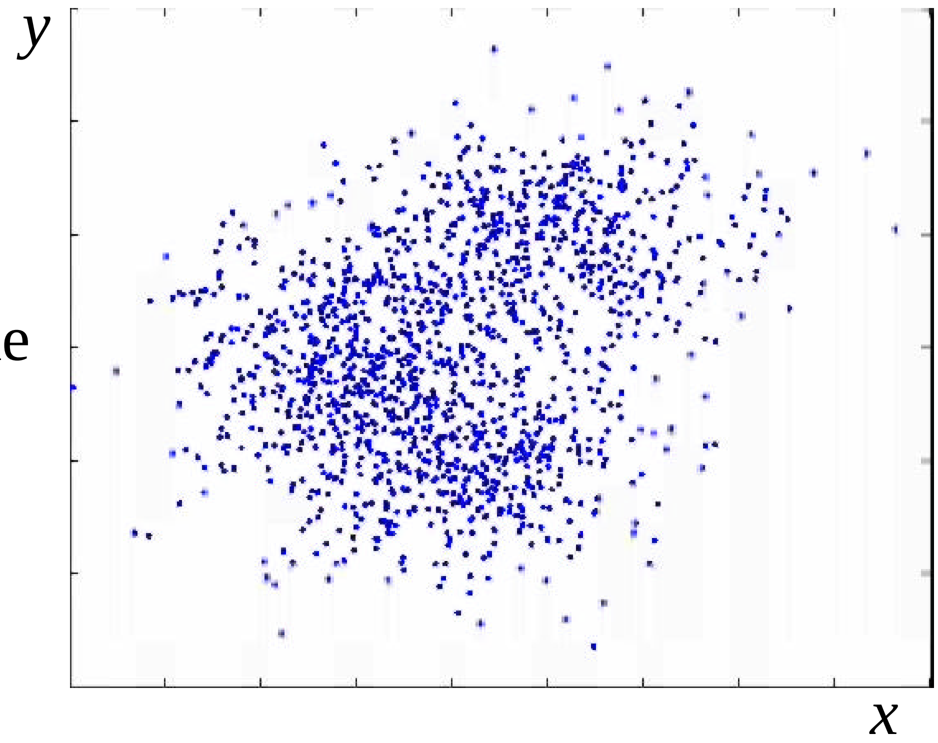
Topics

- K-Means algorithm
- EM algorithm
- Trees
- Clustering for tracking

Find tight groups in the data

K-Means algorithm

1. Randomly place *K* centers
2. Assign points to the closest one
3. Compute new centers (means)
4. Go to 2 until convergence



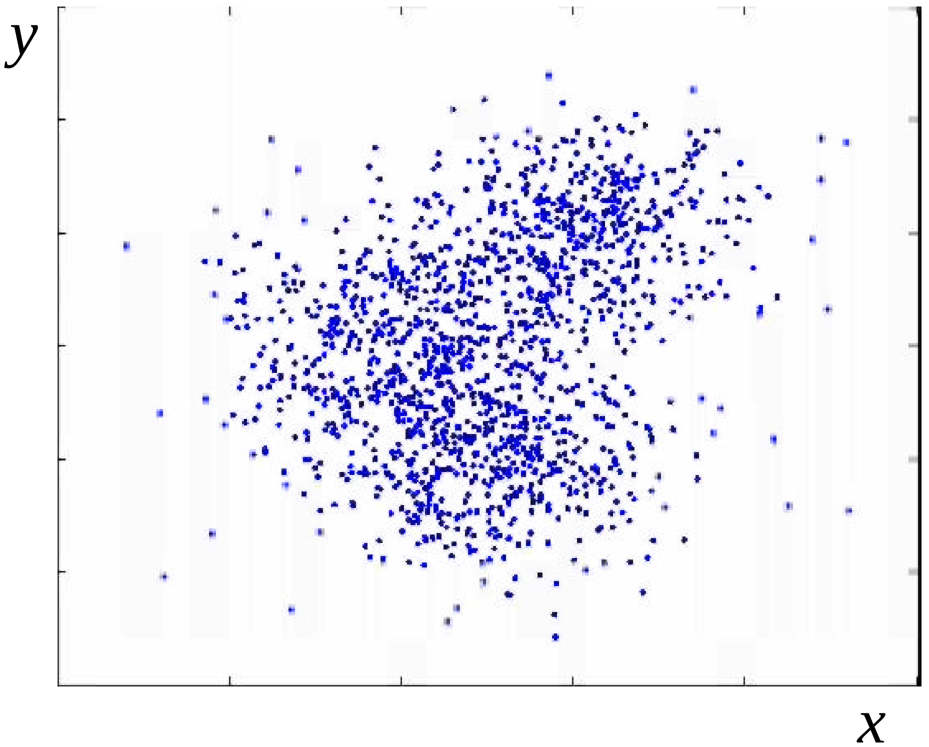
"Hard" label assignment

Clustering/Vector Quantization

Find tight groups in the data

EM algorithm

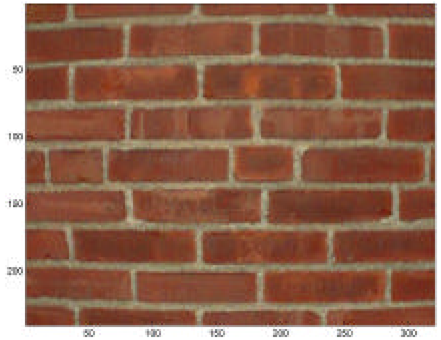
1. Randomly place *K* Gaussians
2. Compute the *posterior* for each point
3. Average data with respect to posterior of each class
4. Go to 2 until convergence



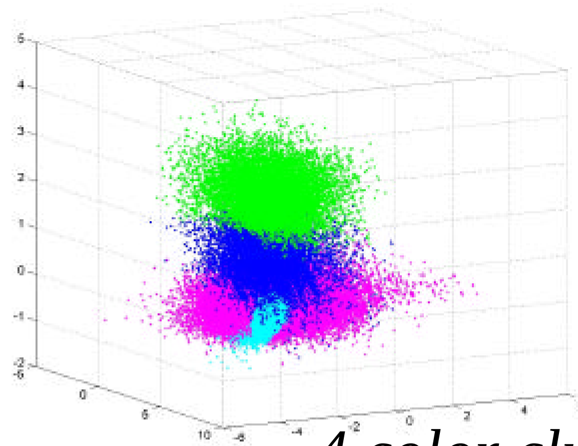
"Soft" label assignment

Clustering – Another Example

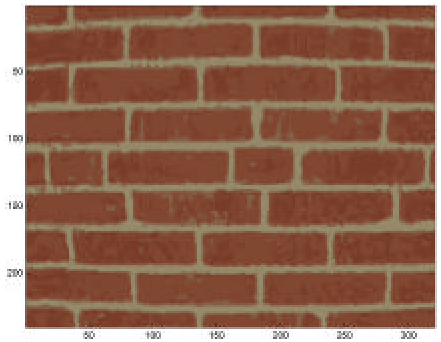
In the image we can cluster pixels in the RGB space



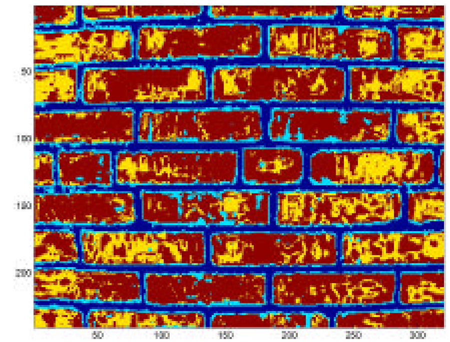
Source – 24 bit/pix



4 color clusters



Color-quantized – 2 bit/pix



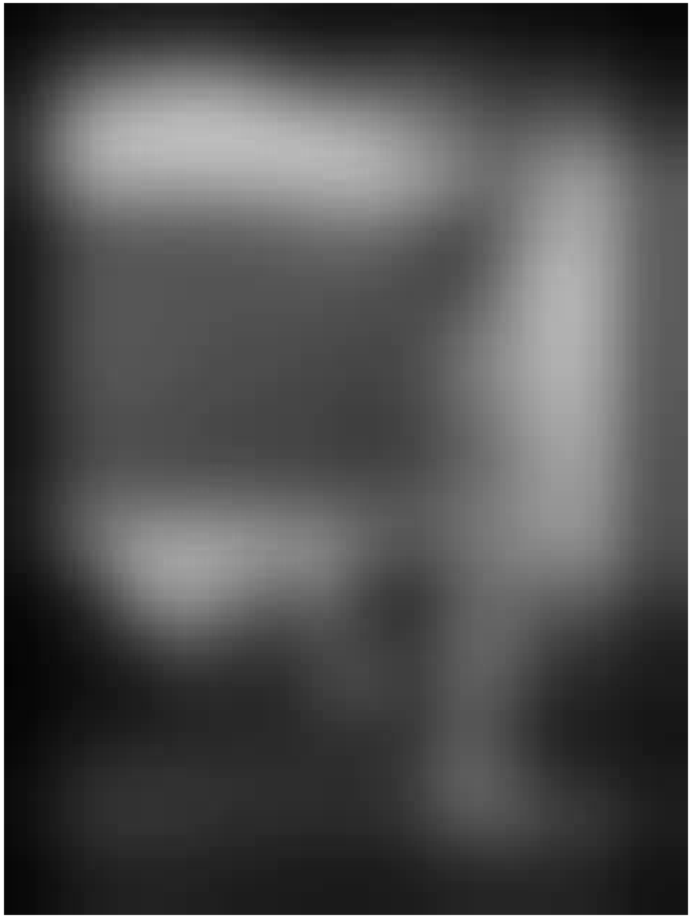
Cluster assignments

Topics

- Motion
- Gesture
- Surveillance

Application: Motion and Gesture Recognition

Is motion important for object recognition?



Photos and figures from: Bobick, A., and J. Davis. "The Representation and Recognition of Action Using Temporal Templates." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, no. 3 (2002). Courtesy of IEEE, A. Bobick, and J. Davis. Copyright 2002 IEEE. Used with Permission.

Motion

One way to do it – turn it into object recognition:

Video:

Cumulative MEI:

Now shape can be matched

Problems: Direction, Segmentation, Only appropriate for determining gross body motion…

Photos and figures from: Bobick, A., and J. Davis. "The Representation and Recognition of Action Using Temporal Templates." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, no. 3 (2002). Courtesy of IEEE, A. Bobick, and J. Davis. Copyright 2002 IEEE. Used with Permission.

Linear Dynamic Systems:

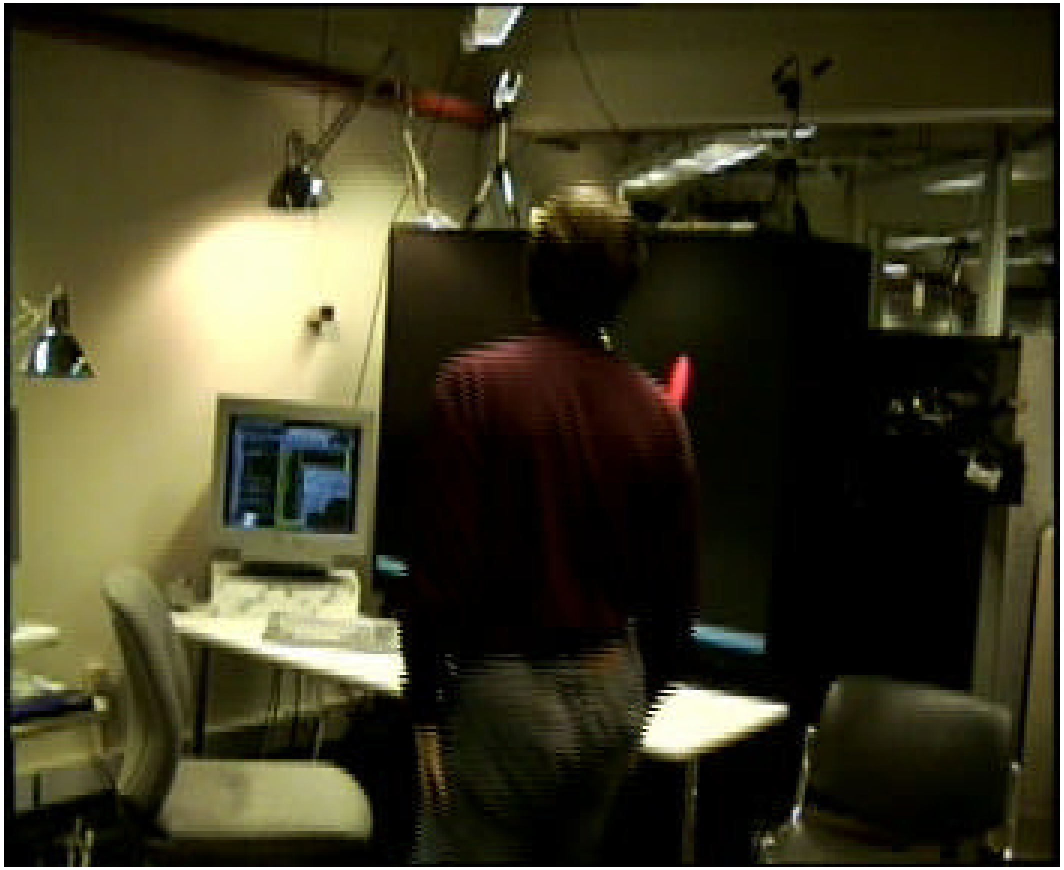
- Systems where the next **state** is a linear combination of the the previous state, a control signal, and noise.
- **Observations** are a linear combination of the current state and noise

$$\mathbf{x}_{t+1} = \mathbf{\Phi}_t \mathbf{x}_t + \mathbf{B}_t \mathbf{u}_t + \mathbf{L}_t \boldsymbol{\xi}_t$$
$$\mathbf{y}_{t+1} = \mathbf{H}_t \mathbf{x}_{t+1} + \boldsymbol{\theta}_t$$

Tracking

- Trackers combine system models $\{F, B, H\}$, with observations to estimate the state.
- ***Unimodal*** trackers, like Kalman Filters, maintain a single estimate of the state
- ***Miltimodal*** tracker frameworks combine multiple unimodal trackers to track multimodal distributions. (e.g.: Multiple Hypothesis Testing, Particle Filtering)

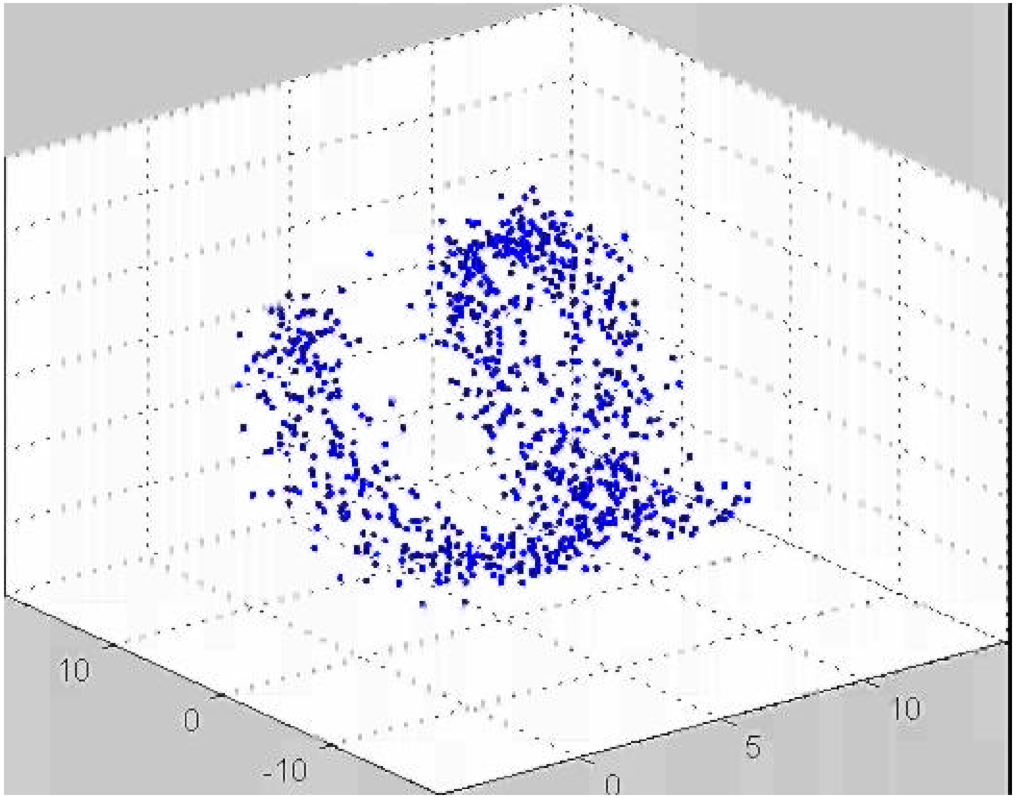
Tracking - Example



Courtesy of Chris Wren. Used with permission.

Tracking allows modeling gesture as time series:

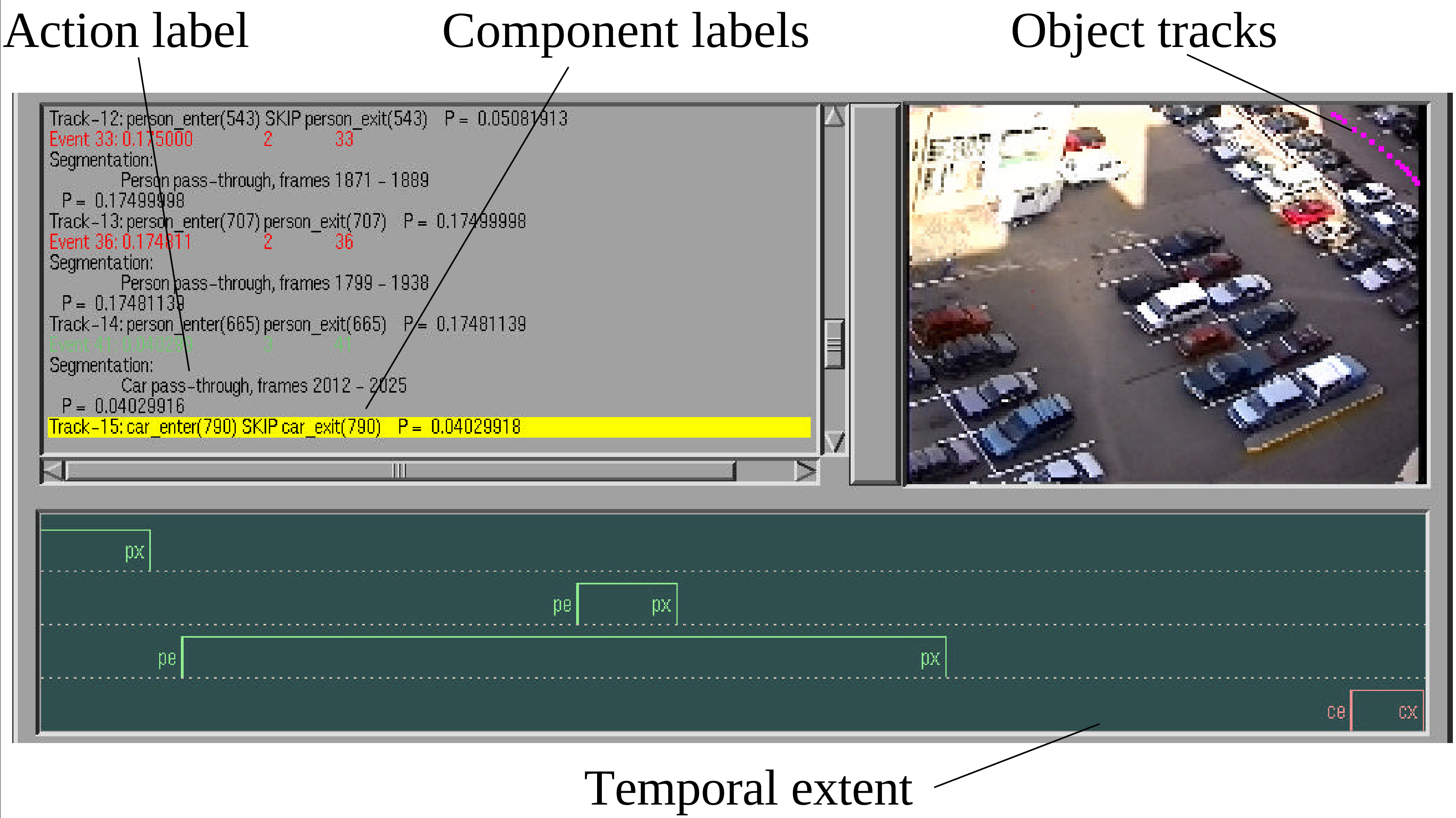
Hidden Markov Model (HMM):
- A mixture of simple densities
- With dynamic constraints



After this sequential density is estimated things are easy again

Applications IV - Surveillance

Parse: Car-Pass-Through



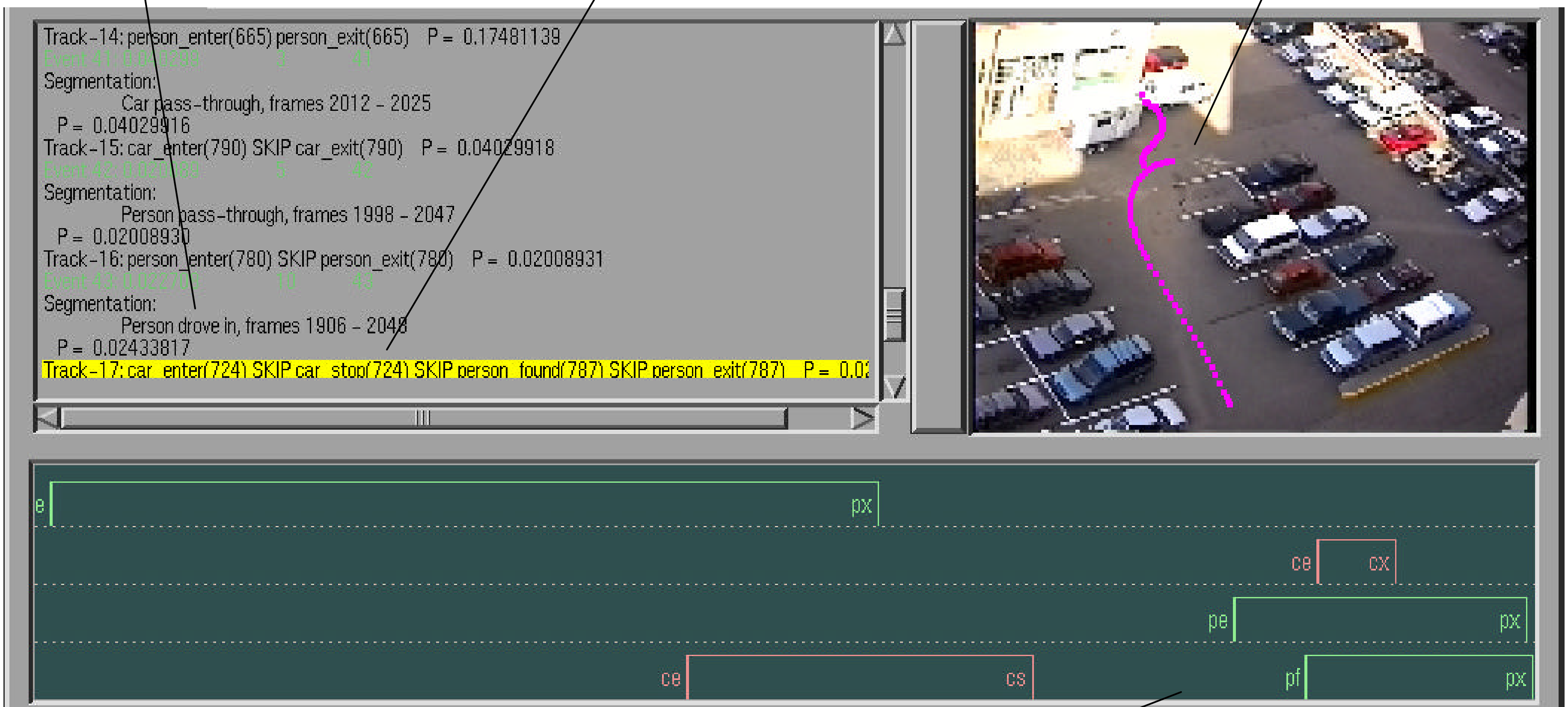
Figures from: Ivanov, Yuri, Chris Stauffer, Aaron Bobick, and W. E. L. Grimson. "Video Surveillance of Interactions." *IEEE Workshop on Visual Surveillance (ICCV 2001)* (1999). Courtesy of IEEE, Y. Ivanov, C. Stauffer, A. Bobick, W. E. L. Grimson. Used with Permission.

Parse: Drive-In

Action label

Component labels

Object tracks



Temporal extent

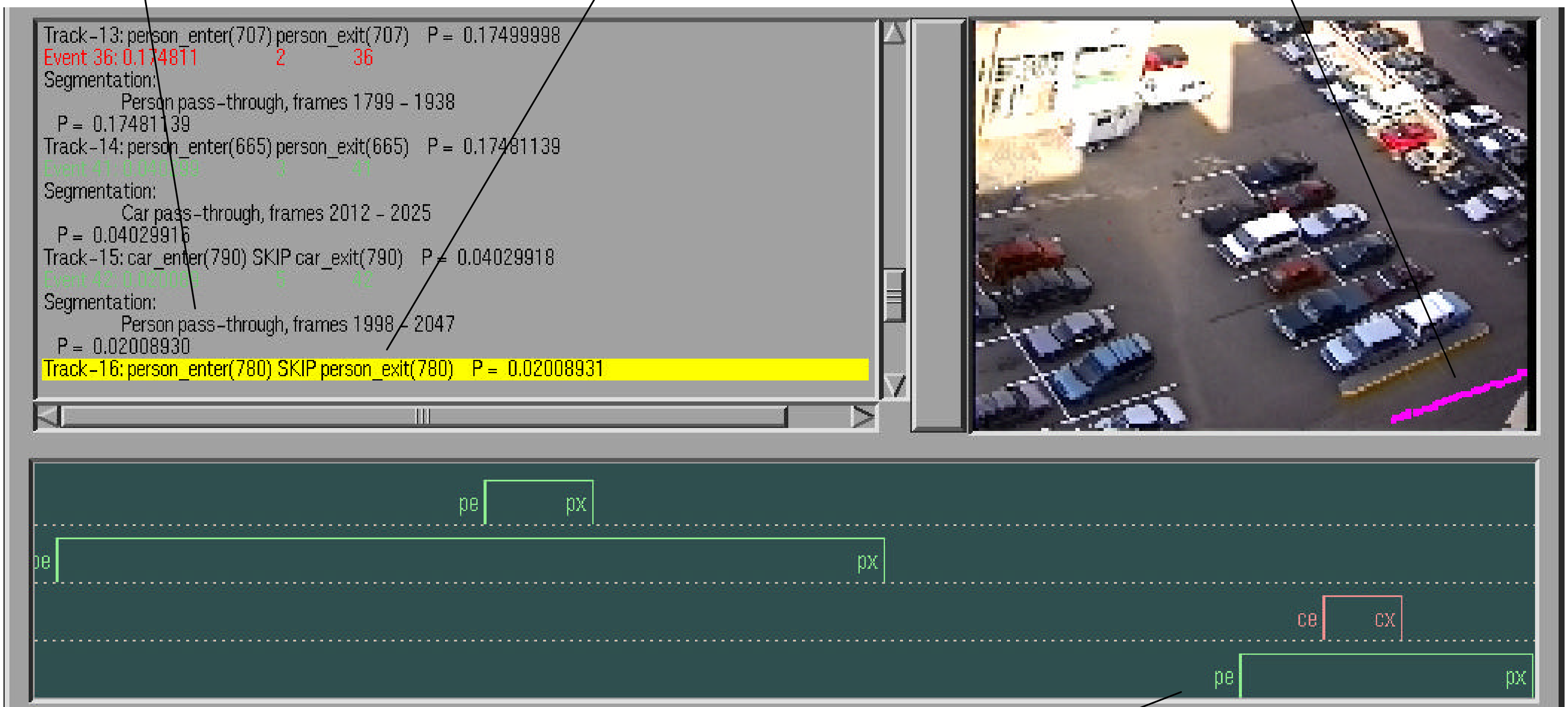
Figures from: Ivanov, Yuri, Chris Stauffer, Aaron Bobick, and W. E. L. Grimson. "Video Surveillance of Interactions." *IEEE Workshop on Visual Surveillance (ICCV 2001)* (1999). Courtesy of IEEE, Y. Ivanov, C. Stauffer, A. Bobick, W. E. L. Grimson. Copyright 1999 IEEE. Used with Permission.

Parse: Person-Pass-Through

Action label

Component labels

Object tracks



Temporal extent

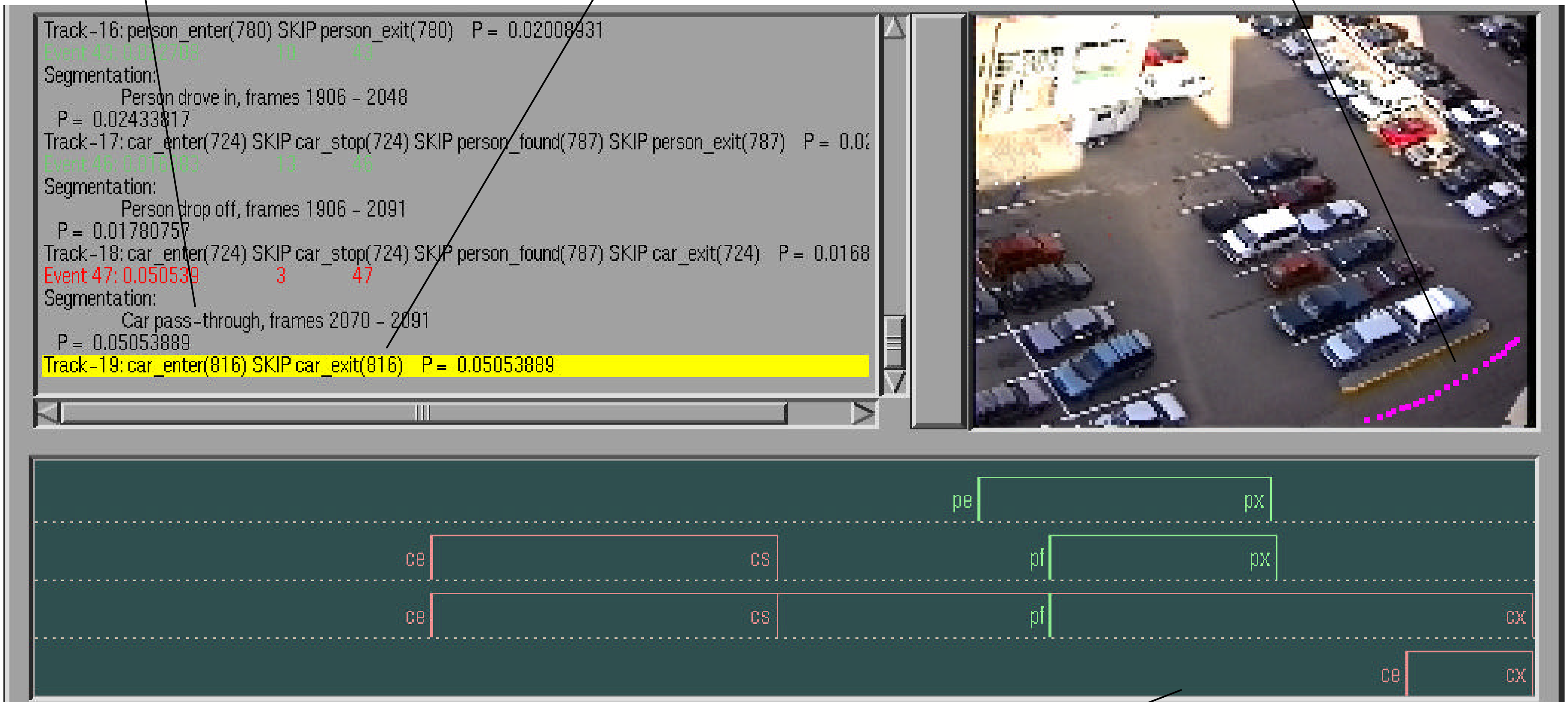
Figures from: Ivanov, Yuri, Chris Stauffer, Aaron Bobick, and W. E. L. Grimson. "Video Surveillance of Interactions." *IEEE Workshop on Visual Surveillance (ICCV 2001)* (1999). Courtesy of IEEE, Y. Ivanov, C. Stauffer, A. Bobick, W. E. L. Grimson. Copyright 1999 IEEE. Used with Permission.

Parse: Car-Pass-Through

Action label

Component labels

Object tracks



Temporal extent

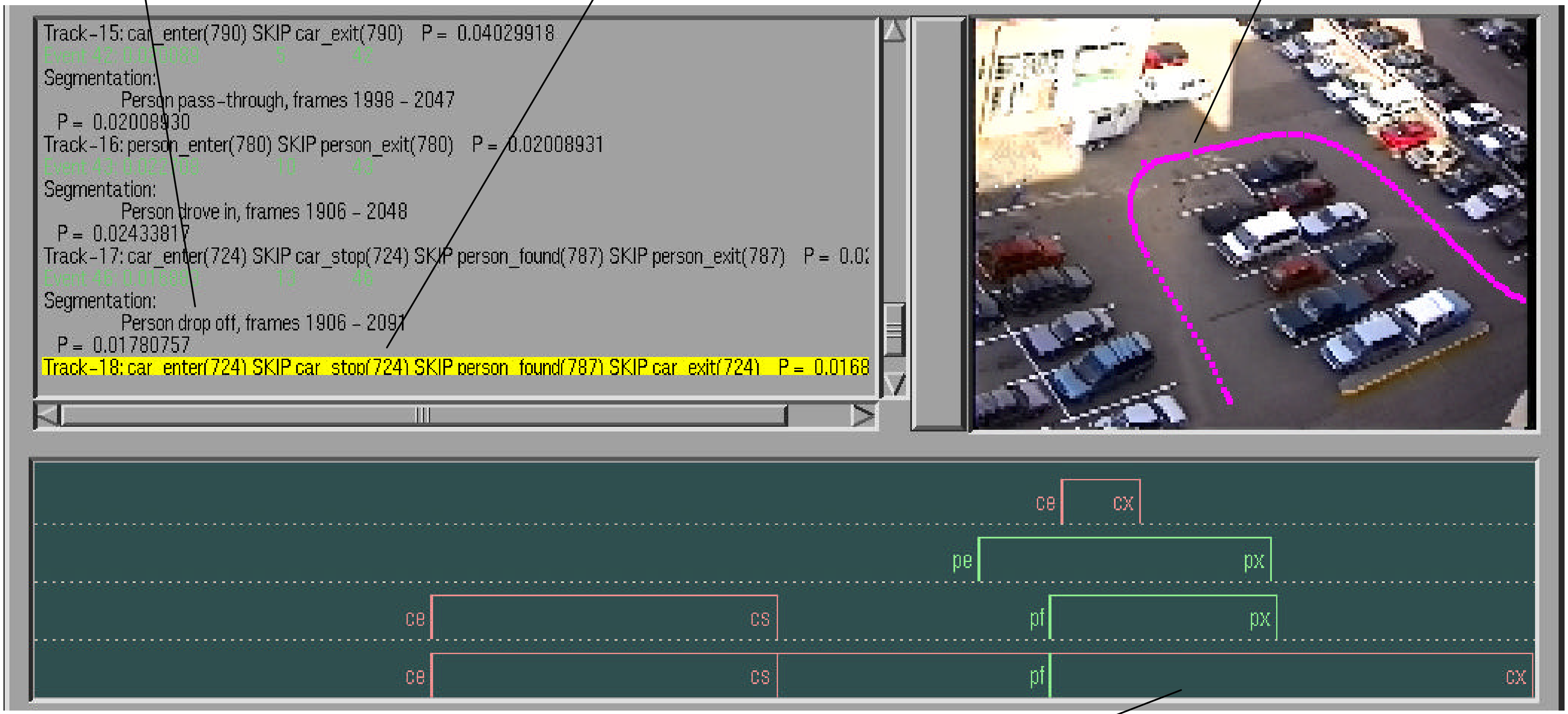
Figures from: Ivanov, Yuri, Chris Stauffer, Aaron Bobick, and W. E. L. Grimson. "Video Surveillance of Interactions." *IEEE Workshop on Visual Surveillance (ICCV 2001)* (1999). Courtesy of IEEE, Y. Ivanov, C. Stauffer, A. Bobick, W. E. L. Grimson. Copyright 1999 IEEE. Used with Permission.

Parse: Drop-Off

Action label

Component labels

Object tracks



Temporal extent

Figures from: Ivanov, Yuri, Chris Stauffer, Aaron Bobick, and W. E. L. Grimson. "Video Surveillance of Interactions." *IEEE Workshop on Visual Surveillance (ICCV 2001)* (1999). Courtesy of IEEE, Y. Ivanov, C. Stauffer, A. Bobick, W. E. L. Grimson. Copyright 1999 IEEE. Used with Permission.