Columbia University
In the City of New York

EE4830 Digital Image Processing
Lecture 11

Object Recognition

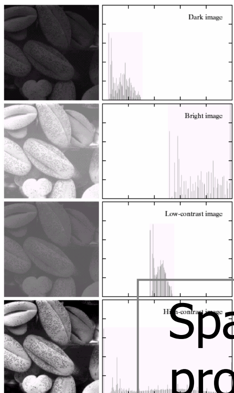
April 16, 2007

Lexing Xie
xlx at ee.columbia.edu

Announcements

- EXP#3 extended to Wed by 10am
- Problem Set #6 assigned
  - One analytical question, one practical
  - Due next Monday 04/23

Roadmap to Date



Spatial Domain processing and enhancement

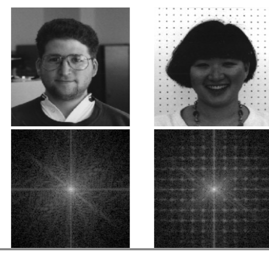


Image Transform and Filtering

Morphological Processing

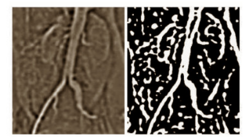


Image Segmentation

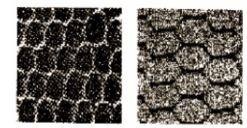


Image Descriptors

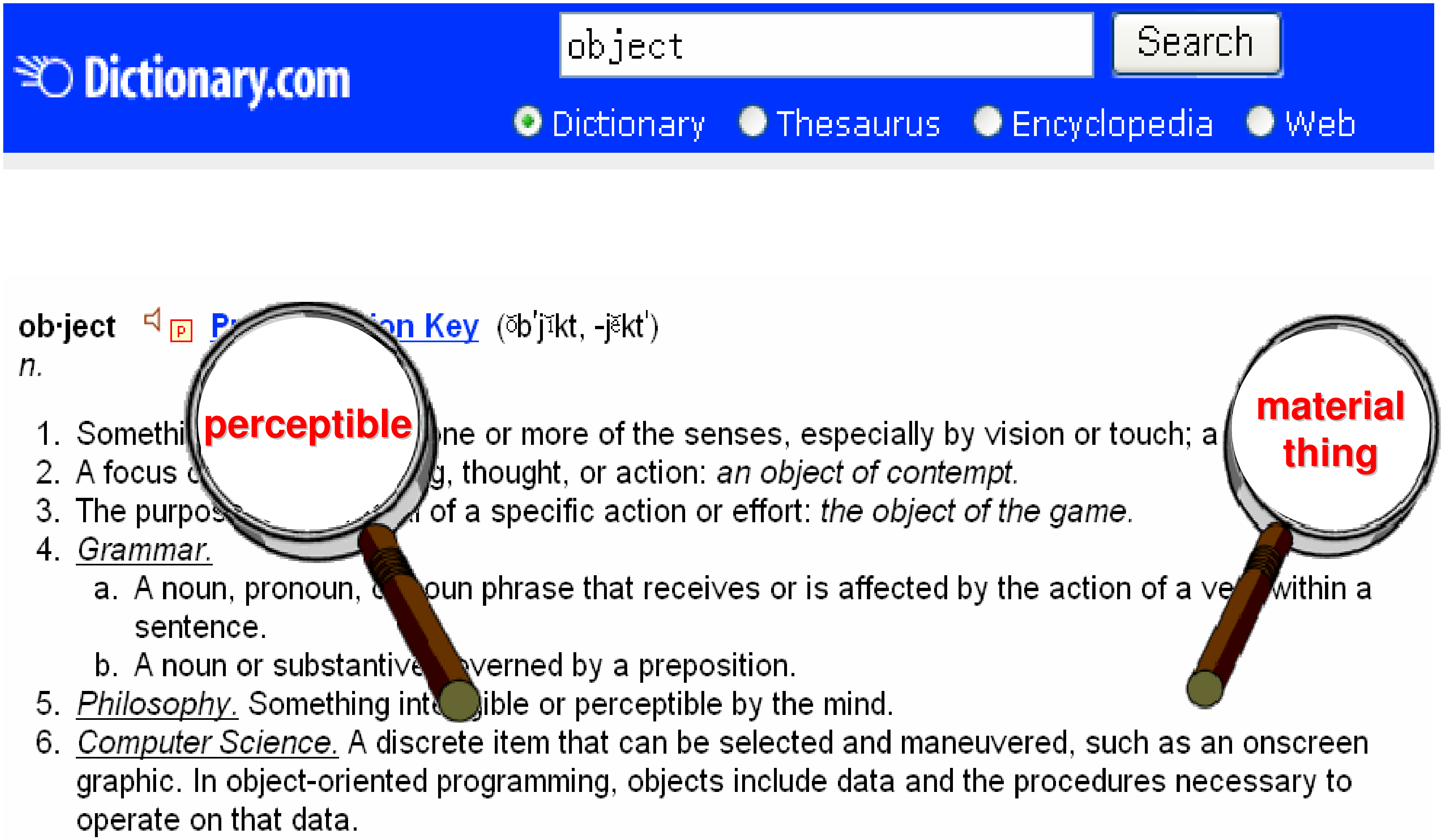
Applicatoions:
Object recognition, Image Compression
Indexing and Retrieval, Reconstruction

Lecture Outline

Problem: object recognition from images.

- What and why
- Object recognition as pattern classification
- General object recognition systems

What is Object Recognition?



Courtesy of http://people.csail.mit.edu/torralba/iccv2005/

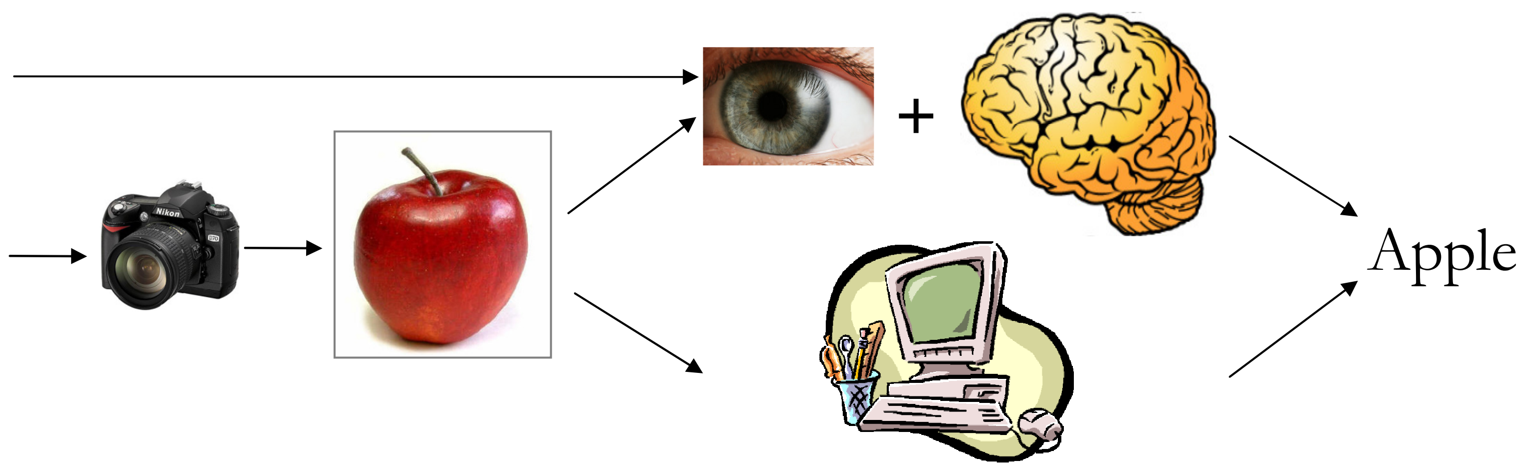
What is Object Recognition?

Sensory data → Descriptions

Color, texture, shape, motion, size, weight, smell, touch, sound, …

"toy", "stuffed Pooh", "a frontal, close-up shot of stuffed Pooh", "ToysRus #345812", …

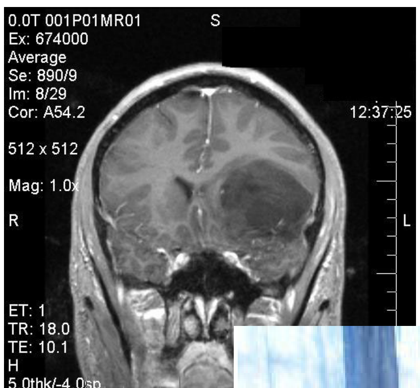
One of the fundamental problems of computer vision:

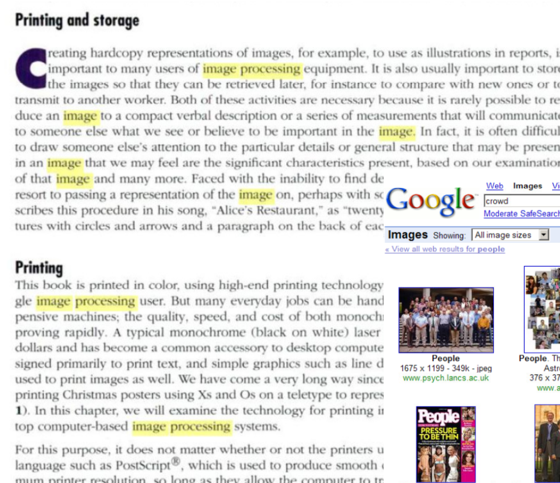


Why?

- Science
  - How do we recognize objects?
- Practice
  - Robot navigation
  - Medical diagnosis
  - Security
  - Industrial inspection and automation
  - Human-computer interface
  - Information retrieval
  - …

Applications of Object Recognition





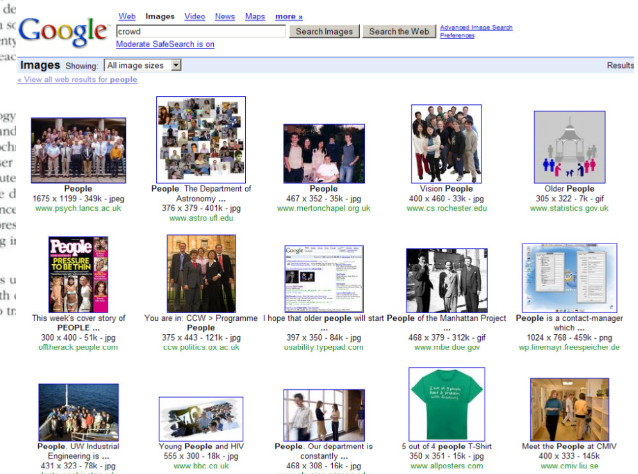
Printing and storage

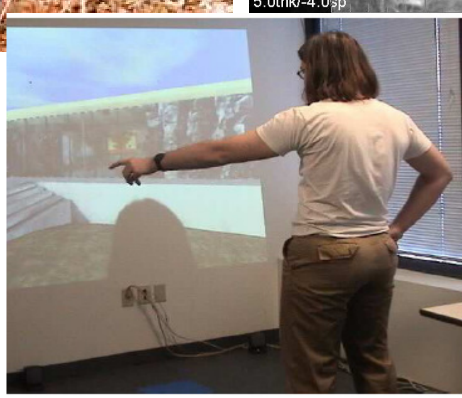
Creating hardcopy representations of images, for example, to use as illustrations in reports, is important to many users of image processing equipment. It is also usually important to store the images so that they can be retrieved later, for instance to compare with new ones or to transmit to another worker. Both of these activities are necessary because it is rarely possible to reduce an image to a compact verbal description or a series of measurements that will communicate to someone else what we see or believe to be important in the image. In fact, it is often difficult to draw someone else's attention to the particular details or general structure that may be present in an image that we may feel are the significant characteristics present, based on our examination of that image and many more. Faced with the inability to find de
resort to passing a representation of the image on, perhaps with sc
scribes this procedure in his song, "Alice's Restaurant," as "twenty
tures with circles and arrows and a paragraph on the back of eac

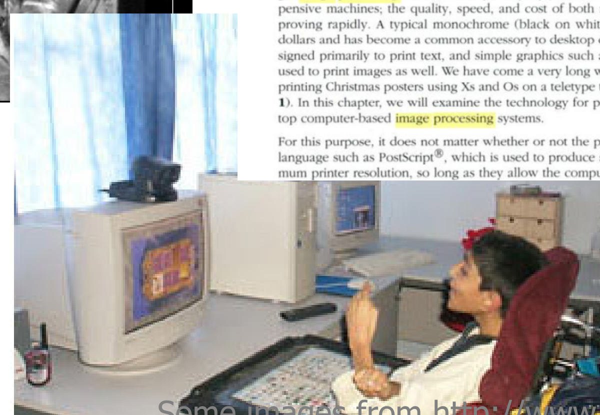
Printing

This book is printed in color, using high-end printing technology
gle image processing user. But many everyday jobs can be hand
pensive machines; the quality, speed, and cost of both monoch
proving rapidly. A typical monochrome (black on white) laser
dollars and has become a common accessory to desktop compute
signed primarily to print text, and simple graphics such as line d
used to print images as well. We have come a very long way since
printing Christmas posters using Xs and Os on a teletype to repres
1). In this chapter, we will examine the technology for printing i
top computer-based image processing systems.

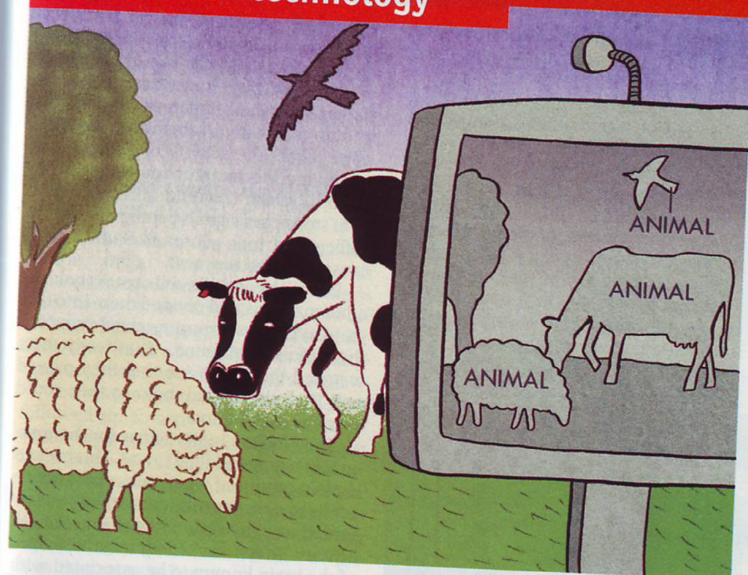
For this purpose, it does not matter whether or not the printers u
language such as PostScript®, which is used to produce smooth
mum printer resolution, so long as they allow the computer to tr







Some images from http://www.cs.utexas.edu/~grauman/research/research.html



Also in this section

Tech.view, our online column on personal technology, appears on Economist.com on Fridays. The column can be viewed at **www.economist.com/techview**

Computer vision

Easy on the eyes

A computer can now recognise classes of things as accurately as a person can

NEVER underestimate a computer. Never overestimate one either. For many years Garry Kasparov, a world chess champion, said that a computer would never beat him (or, indeed, any other human in his position). In May 1997 he had to eat his words. Deep Blue, an invention of IBM, did just that.

such as a rectangle or even a human face, and it can make a reasonable fist of the task. Ask it to find "animals" among photographs of dragonflies, trees, sharks, cars and monkeys, and it falls over. Indeed a monkey—or even a human baby—would leave it in the dust.

ised into numerous areas. Experiments on monkeys, in which researchers have recorded what excites individual nerve cells in each of these areas, give strong hints about how it works.

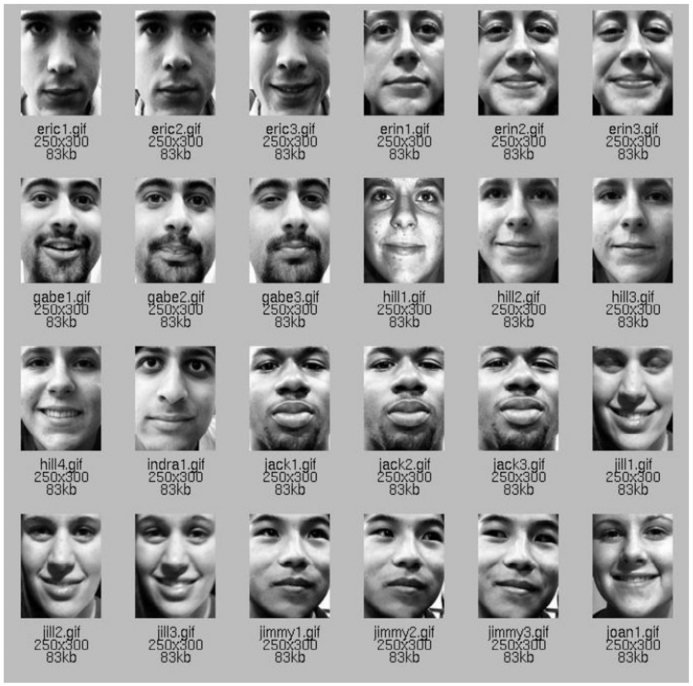
The pathway is hierarchical. Signals from the retina flow to the most basic processing area first; the cells in that area fire up others in the next area; and so on. Those in the first area are fussy. They react to edges or bars in particular orientations. By combining their signals, however, cells in the second area can respond to corners or bars in any orientation. And so the system builds up. Cells in the final area can recognise general things, animals included.

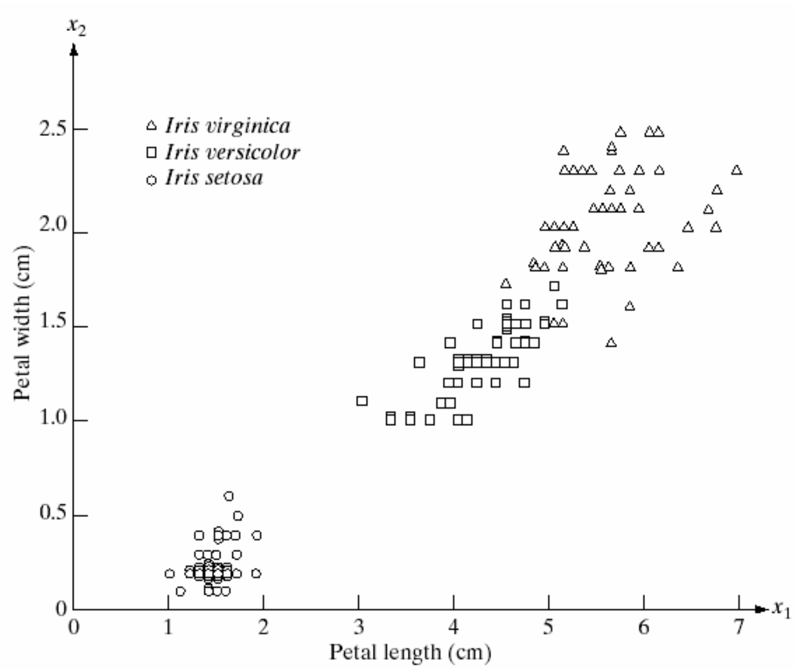
The Economists, April 7th, 2007

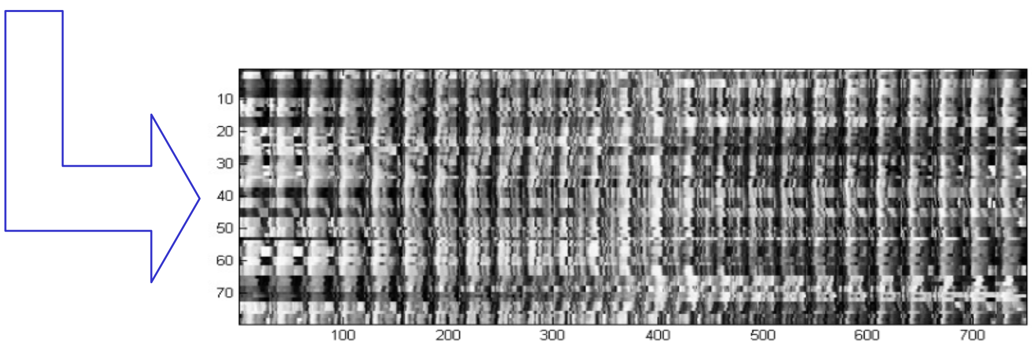
Lecture Outline

- Object recognition: what and why
- Object recognition as pattern classification
  - Distance-based classifiers
  - Neural networks
  - Bayes classifiers
  - Object recognition in practice
- General object recognition systems
- Summary

Objects as Vectors …







Minimum Distance Classifier

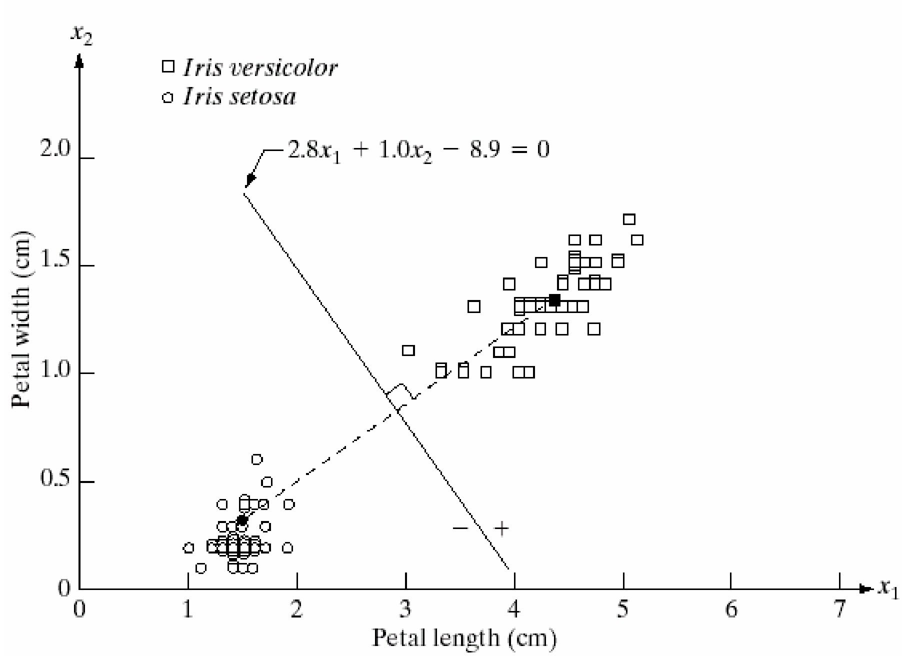
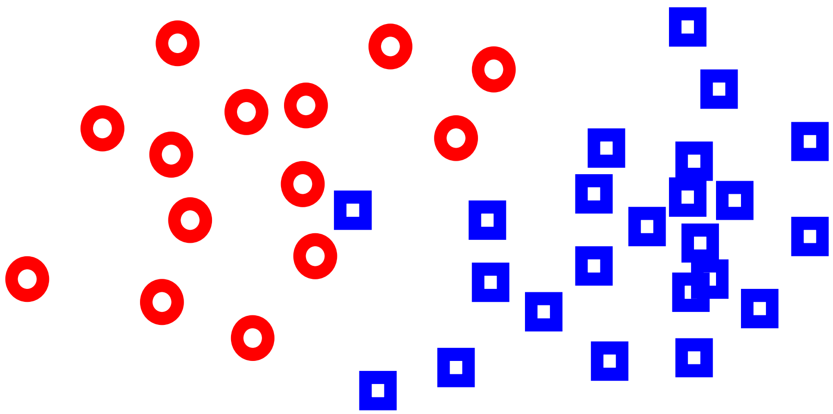
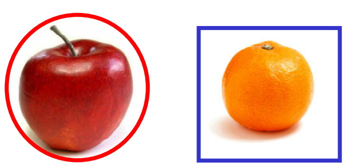


FIGURE 12.6 Decision boundary of minimum distance classifier for the classes of *Iris versicolor* and *Iris setosa*. The dark dot and square are the means.

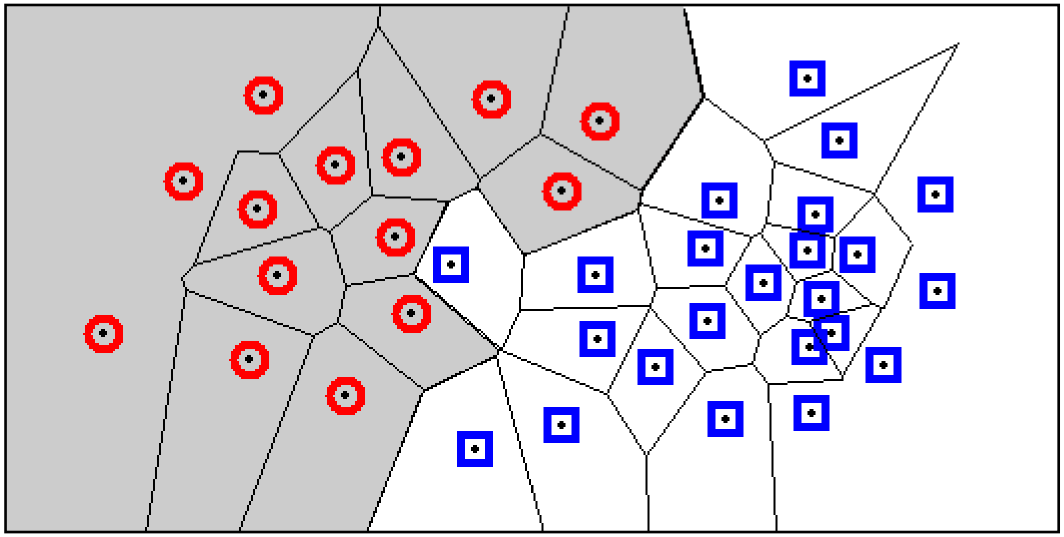
$$\hat{\omega} = \arg\min_j d(x, m_j), \;\; j = 1, 2$$

Nearest Neighbor Classifier





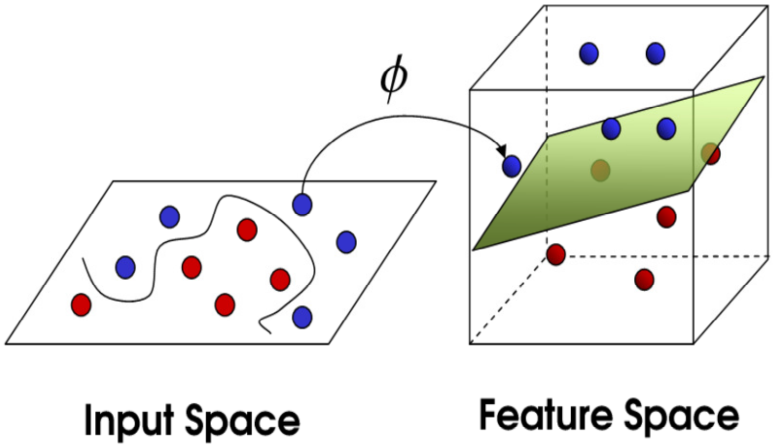
Nearest Neighbor Classifier



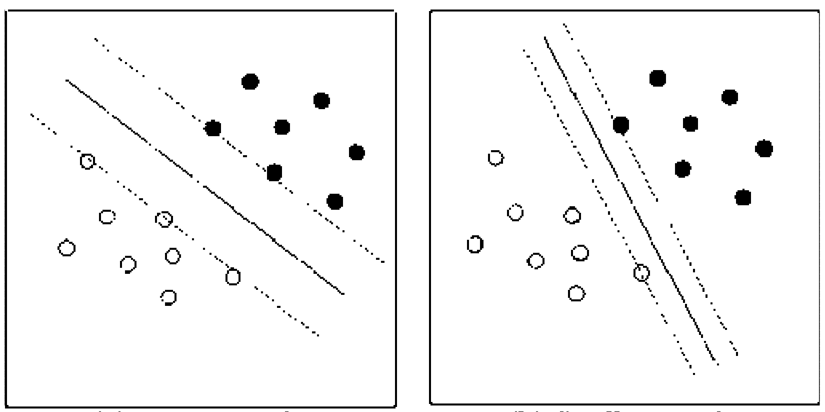
Support Vector Machines

- Two key ideas:
  - Class boundary can be linear in a higher-dimensional space, e.g.,

$$\Phi\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix}$$



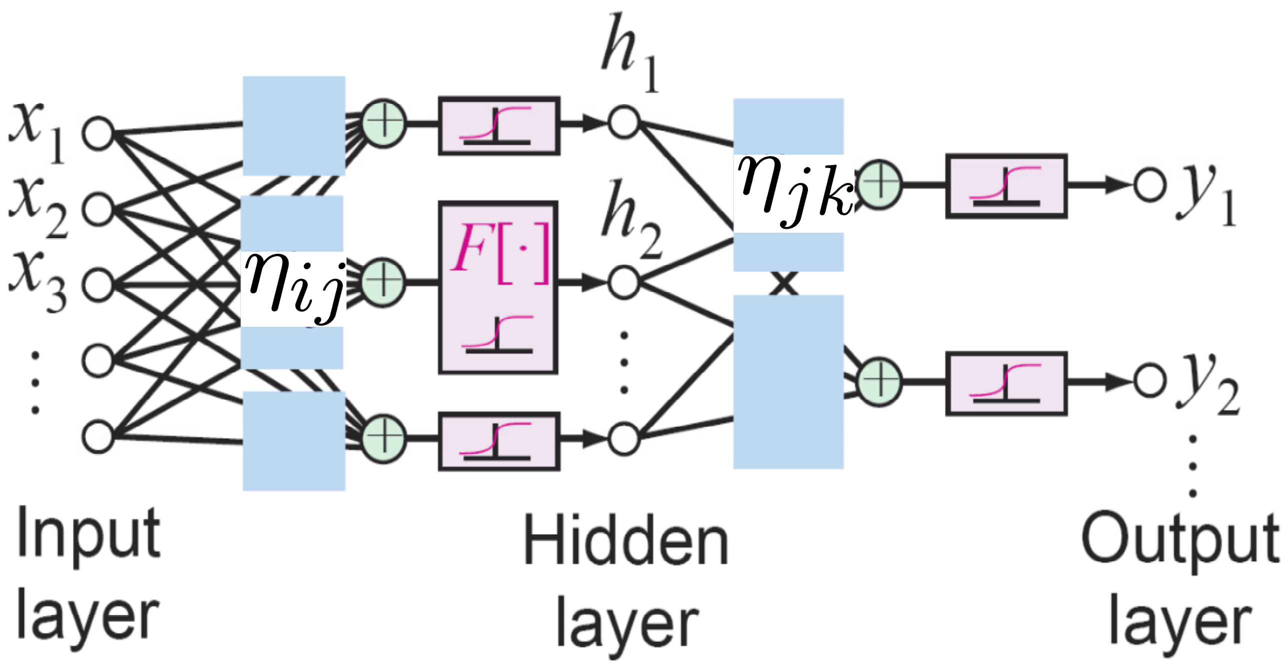
  - The "best" separating hyperplane has the largest margin.



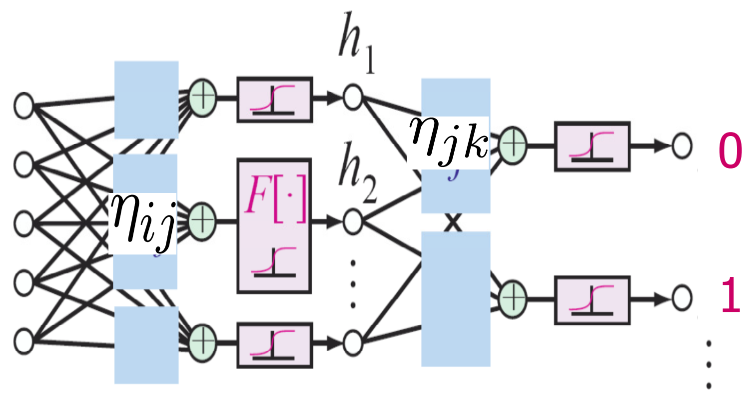
(a) Larger margin (b) Smaller margin

Neural Networks

$$y_k = F[\sum_j \eta_{jk} \cdot F[\sum_j \eta_{ij} x_i]]$$



Digit Recognition with Neural Net



- LeCun et al, 1992, 1998, …

http://yann.lecun.com/exdb/mnist/

40 PCA + quadratic classifier	none	3.3	LeCun et al. 1998
1000 RBF + linear classifier	none	3.6	LeCun et al. 1998
K-NN, Tangent Distance	subsampling to 16x16 pixels	1.1	LeCun et al. 1998
SVM, Gaussian Kernel	none	1.4	
SVM deg 4 polynomial	deskewing	1.1	LeCun et al. 1998
Reduced Set SVM deg 5 polynomial	deskewing	1.0	LeCun et al. 1998
Virtual SVM deg-9 poly [distortions]	none	0.8	LeCun et al. 1998
Virtual SVM, deg-9 poly, 1-pixel jittered	none	0.68	DeCoste and Scholkopf, MLJ 2002
Virtual SVM, deg-9 poly, 1-pixel jittered	deskewing	0.68	DeCoste and Scholkopf, MLJ 2002
Virtual SVM, deg-9 poly, 2-pixel jittered	deskewing	0.56	DeCoste and Scholkopf, MLJ 2002
2-layer NN, 300 hidden units, mean square error	none	4.7	LeCun et al. 1998
2-layer NN, 300 HU, MSE, [distortions]	none	3.6	LeCun et al. 1998
2-layer NN, 300 HU	deskewing	1.6	LeCun et al. 1998
2-layer NN, 1000 hidden units	none	4.5	LeCun et al. 1998
2-layer NN, 1000 HU, [distortions]	none	3.8	LeCun et al. 1998
3-layer NN, 300+100 hidden units	none	3.05	LeCun et al. 1998
3-layer NN, 300+100 HU [distortions]	none	2.5	LeCun et al. 1998
3-layer NN, 500+150 hidden units	none	2.95	LeCun et al. 1998
3-layer NN, 500+150 HU [distortions]	none	2.45	LeCun et al. 1998
3-layer NN, 500+300 HU, softmax, cross entropy, weight decay	none	1.53	Hinton, unpublished, 2005
2-layer NN, 800 HU, Cross-Entropy Loss	none	1.6	Simard et al., ICDAR 2003
2-layer NN, 800 HU, cross-entropy [affine distortions]	none		
2-layer NN, 800 HU, MSE [elastic distortions]	none	0.9	Simard et al., ICDAR 2003
2-layer NN, 800 HU, cross-entropy [elastic distortions]	none	0.7	Simard et al., ICDAR 2003

Bayes Classifier

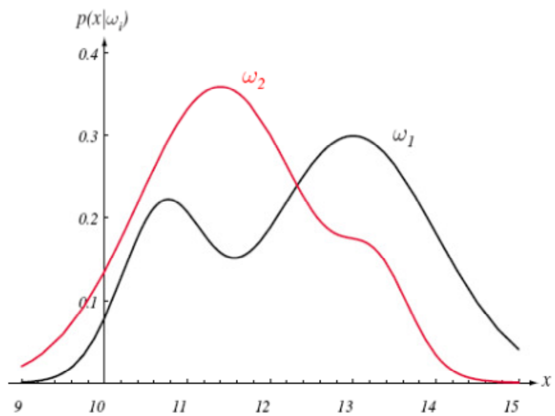


FIGURE 2.1. Hypothetical class-conditional probability density functions show t probability density of measuring a particular feature value x given the pattern is category ω_i. If x represents the lightness of a fish, the two curves might describe t difference in lightness of populations of two types of fish. Density functions are norm ized, and thus the area under each curve is 1.0. From: Richard O. Duda, Peter E. Ha and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & So Inc.

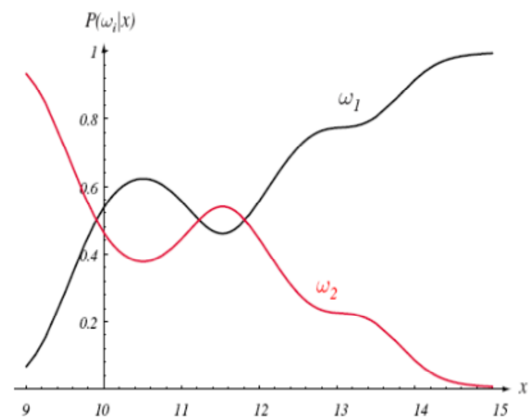


FIGURE 2.2. Posterior probabilities for the particular priors $P(\omega_1) = 2/3$ and $P(\omega_2) = 1/3$ for the class-conditional probability densities shown in Fig. 2.1. Thus in this case, given that a pattern is measured to have feature value $x = 14$, the probability it is in category ω_2 is roughly 0.08, and that it is in ω_1 is 0.92. At every x, the posteriors sum to 1.0. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

$$\begin{aligned} p(\omega_1|x) &= p(\omega_1)\frac{p(x|\omega_1)}{p(x)} \\ &= p(\omega_1)\frac{p(x|\omega_1)}{p(\omega_1)p(x|\omega_1) + p(\omega_2)p(x|\omega_2)} \end{aligned}$$

$$\begin{aligned} p(\omega_2|x) &= p(\omega_2)\frac{p(x|\omega_2)}{p(x)} \\ &= \ldots \end{aligned}$$

Bayes Classifier for Gaussian Classes

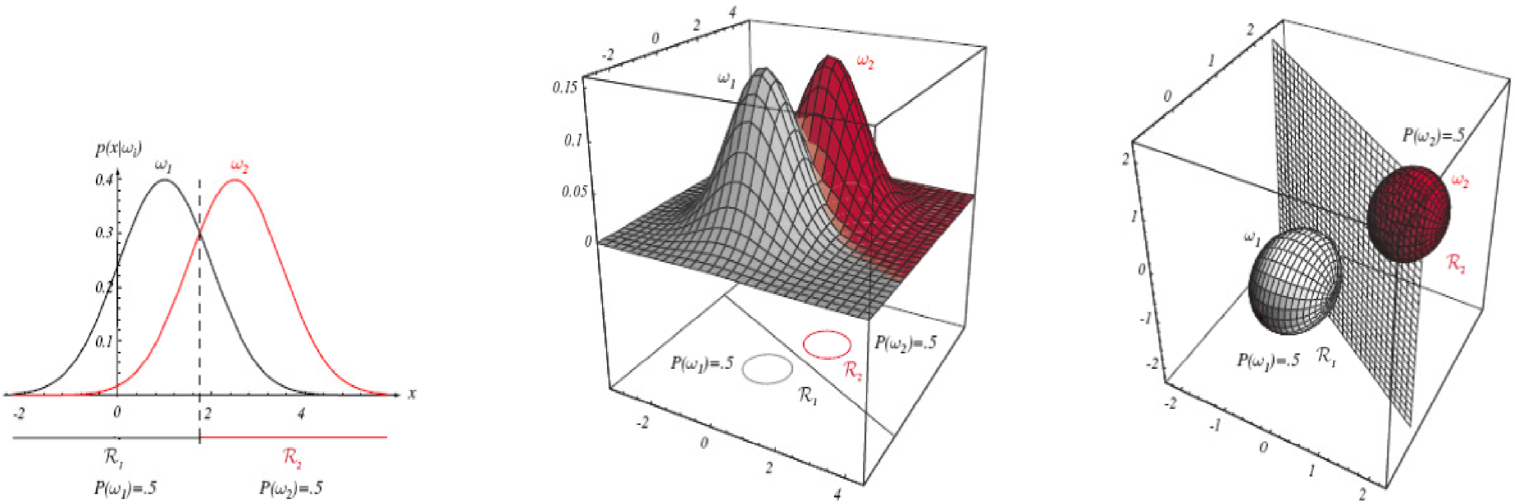


FIGURE 2.10. If the covariance matrices for two distributions are equal and proportional to the identity matrix, then the distributions are spherical in d dimensions, and the boundary is a generalized hyperplane of $d-1$ dimensions, perpendicular to the line separating the means. In these one-, two-, and three-dimensional examples, we indicate $p(\mathbf{x}|\omega_i)$ and the boundaries for the case $P(\omega_1) = P(\omega_2)$. In the three-dimensional case, the grid plane separates $\mathcal{R}_1$ from $\mathcal{R}_2$. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

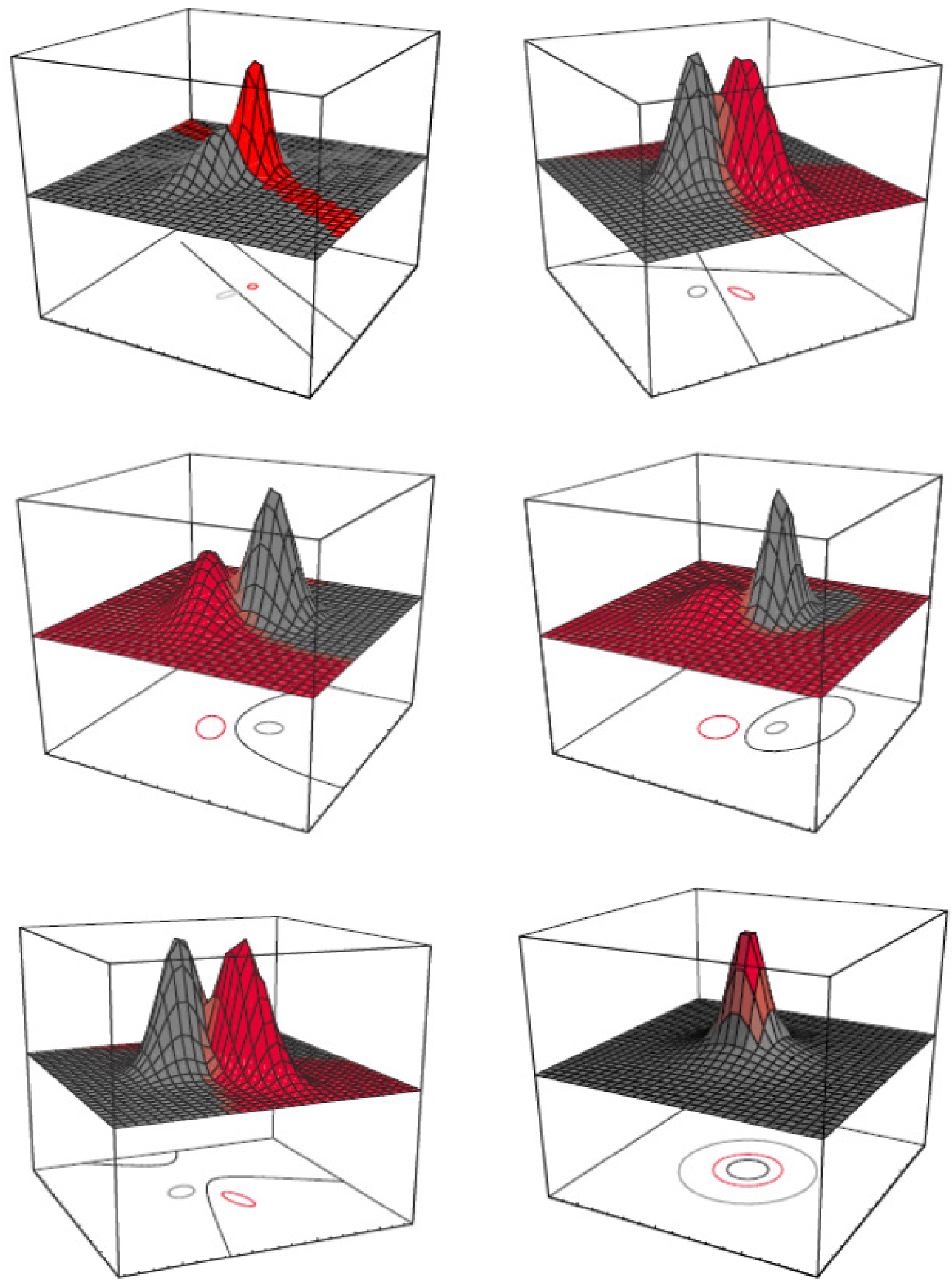
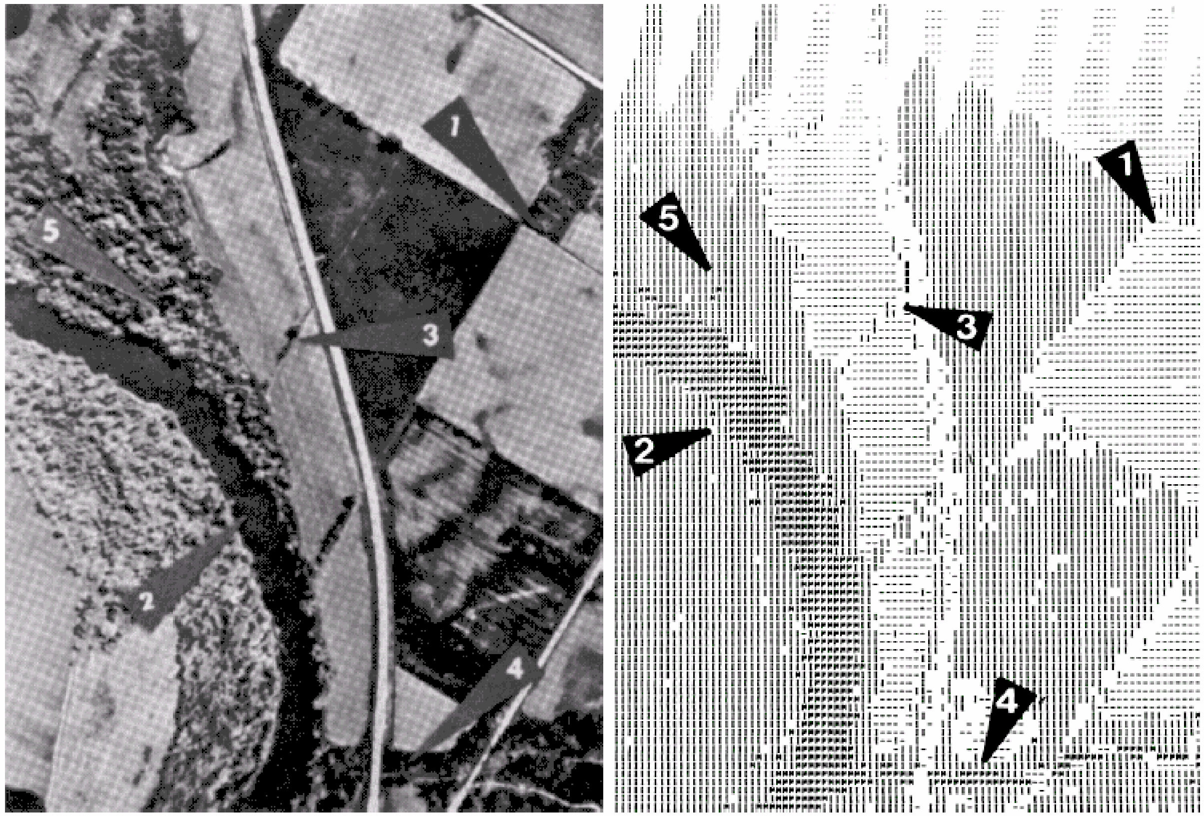


FIGURE 2.14. Arbitrary Gaussian distributions lead to Bayes decision boundaries that are general hyperquadrics. Conversely, given any hyperquadric, one can find two Gaus-



a b

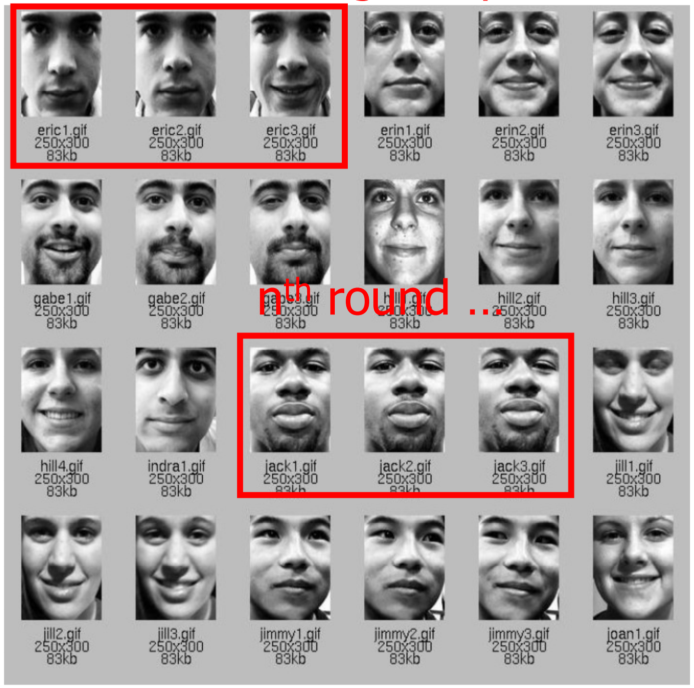
FIGURE 12.13 (a) Multispectral image. (b) Printout of machine classification results using a Bayes classifier. (Courtesy of the Laboratory for Applications of Remote Sensing, Purdue University.)

Homework: Classifying Faces …

- Goal: learn male/female face model from images
- Steps
  - Read images
  - Generate corresponding label (F/M +1/-1)
  - Train classifier (SVM/NN/ …)
- Estimate performance
  - Leave-one-out
    - "leave-one-person-out"
  - Report error rate:

$$err = \frac{\#\ \text{miss-classified faces}}{\text{total}\ \#\text{of faces}} \times 100\%$$

1st round testing samples



Object Recognition in Practice

- Commercial object recognition
  - Currently a $4 billion/year industry for inspection and assembly
  - Almost entirely based on template matching
- Upcoming applications
  - Mobile robots, toys, user interfaces
  - Location recognition
  - Digital camera panoramas, 3D scene modeling

This section courtesy of David Lowe, website and CVPR 2003 Tutorial

Industrial Applications

The Computer Vision Industry

David Lowe

This web page provides links to companies that develop products using computer vision. Computer vision (also often referred to as "machine vision" or "automated imaging") is the automated extraction of information from images. This differs from image processing, in which an image is processed to produce another image. This page covers only products based on computer or machine vision, and it does not cover image processing or any of the many suppliers of sensors or other equipment to the industry.

Companies are categorized under their principal application area, and then listed alphabetically. Companies are listed only if they have web pages giving information about their products. Please let me know of any links that are missing.

Automobile driver assistance

Iteris (Anaheim, California). Lane departure warning systems for trucks and cars that monitor position on the road. Used in over 10,000 trucks (2005). Also creates traffic monitoring systems.

MobilEye (Jerusalem, Israel). Vision systems that warn automobile drivers of danger, provide adaptive cruise control, and give driver assistance.

Smart Eye (Göteborg, Sweden). Systems to track eye and gaze position of a driver to detect drowsiness or inattention.

Automobile traffic management

Appian Technology (Bourne End, Buckinghamshire, UK). Systems for reading automobile license plates.

AutoVu (Montreal, Canada). Systems for reading automobile license plates.

Image Sensing Systems (St. Paul, Minnesota). Created the Autoscope system that uses roadside video cameras for real-time traffic management. Over 40,000 cameras are in use.

Film and Television

2D3 (Oxford, UK). Systems for tracking objects in video or film and solving for 3D motion to allow for precise augmentation with 3D computer graphics.

Hawkeye (Winchester, UK). Uses multiple cameras to precisely track tennis and cricket balls for sports refereeing and commentary.

Image Metrics (Manchester, England). A markerless tracking system for the human face that can be used to map detailed motion and facial expressions to synthetic characters.

Imagineer Systems (Guildford, UK). Computer vision software for the film and video industries.

http://www.cs.ubc.ca/spider/lowe/vision.html

APPIAN TECHNOLOGY PLC

HOME PRODUCTS APPLICATIONS COMPANY DISTRIBUTORS SERVICES INVESTORS NEWS CONTACT

RDI2CBW

LEADING EDGE SECURITY, SURVEILLANCE AND TRAFFIC MANAGEMENT PRODUCTS

Who we are...

Appian Technology PLC is the leading provider of high technology security, surveillance and traffic management products. We have unique in-house expertise backed by over 10 years experience in selling and supporting our products world-wide. Appian provide the world leading Talon ANPR/ALPR system, the Navigator Parking Guidance Information (PGI) system and the LaserCAM mobile digital speed enforcement camera. Our products provide accurate, cost-effective systems for the Security, Police, and Commercial markets. From counter terrorism to congestion charging, we have the solution...

Find us...

Appian Technology Plc [View Map]
Appian House
4 Wessex Road
Bourne End
Buckinghamshire Tel: +44 (1628) 554 750
SL8 5DT Fax: +44 (1628) 554 751
UK E-Mail: sales@appian-tech.com

OUR PRODUCTS

Automatic Number Plate Recognition (ANPR)

Talon is the world leading ANPR/ALPR solution. It employs highly accurate neural network processing technology. With over 10 years of experience in ANPR, the Talon system is proven in critical policing, counter terrorism and commercial deployments world-wide. more »

ANPR Products » ANPR Applications »

Parking Guidance Information (PGI)

Navigator is an accurate, reliable and cost effective PGI system for all types of parking facilities. Navigator is suitable for a range of parking guidance applications from car park access guidance, zone specific guidance to urban area traffic guidance.. more »

ALPR - Automatic License Plate Recogition | ANPR - Automatic Number Plate Recognition | Congestion Charging
MVI Systems | Road Tolling Systems | Parking Guidance Information
Traffic Solutions | Speed Solutions | Speed Enforcement | Appian Technology PLC

design by rawnet|limited

http://www.appian-tech.com/

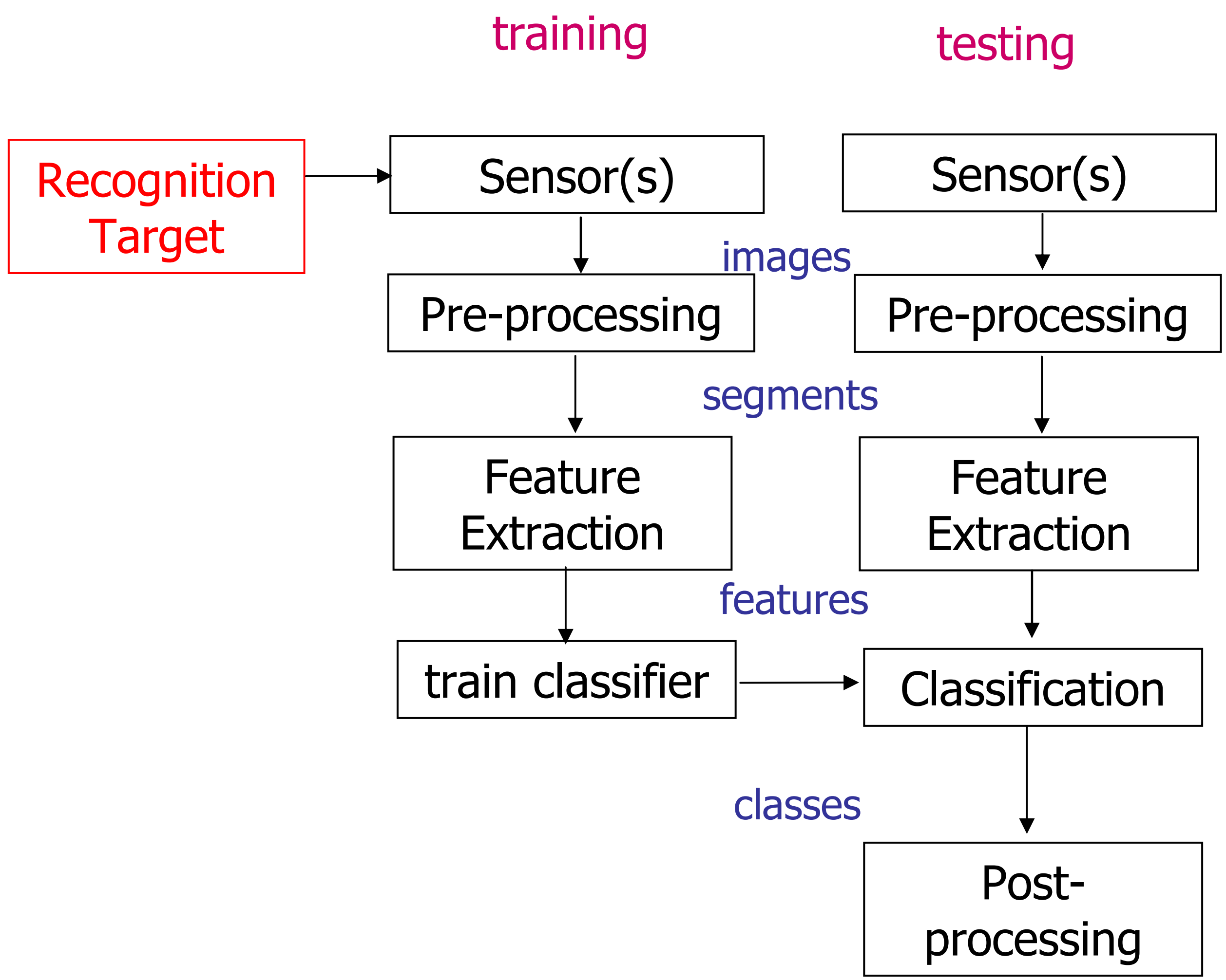
http://www.sportvision.com/

http://www.dipix.com/

Lecture Outline

- Object recognition: what and why
- Object recognition as pattern classification
- General object recognition systems
  - Another view of object recognition
  - Real-world challenges
  - Survey of state-of the art
- Summary

Object Recognition End-to-End



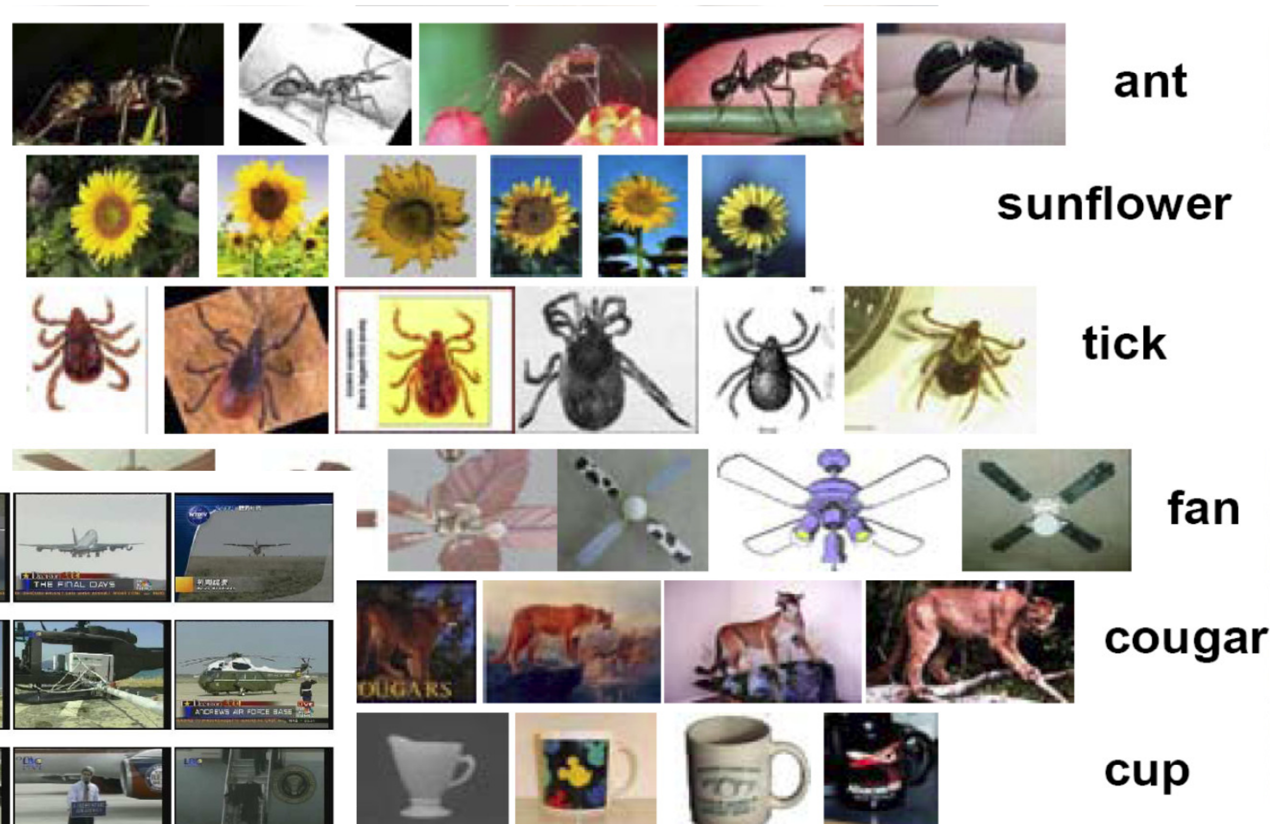
- window
- parts or patches
- salient points

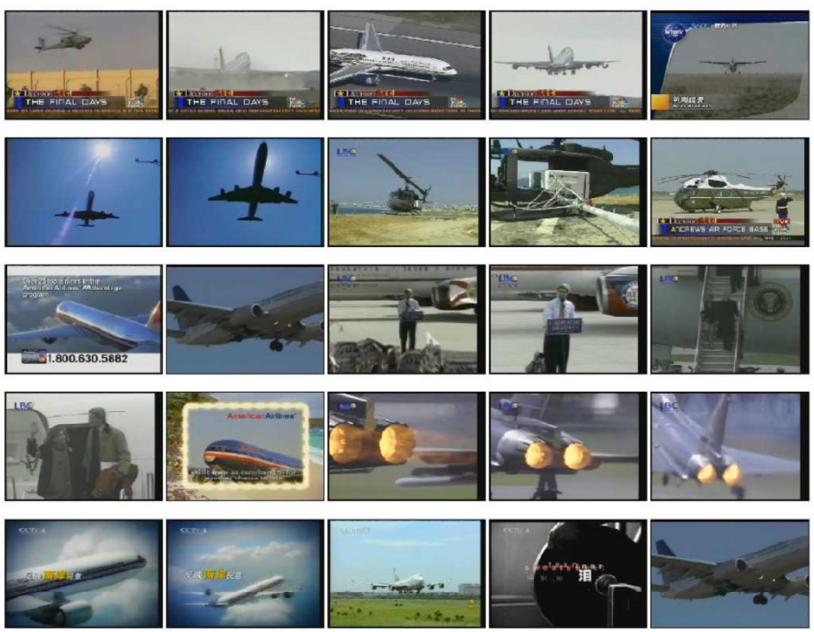
- pixel vector
- descriptors

- face? car?
- statue of liberty?

- spatial/temporal smoothing
- context, etc.

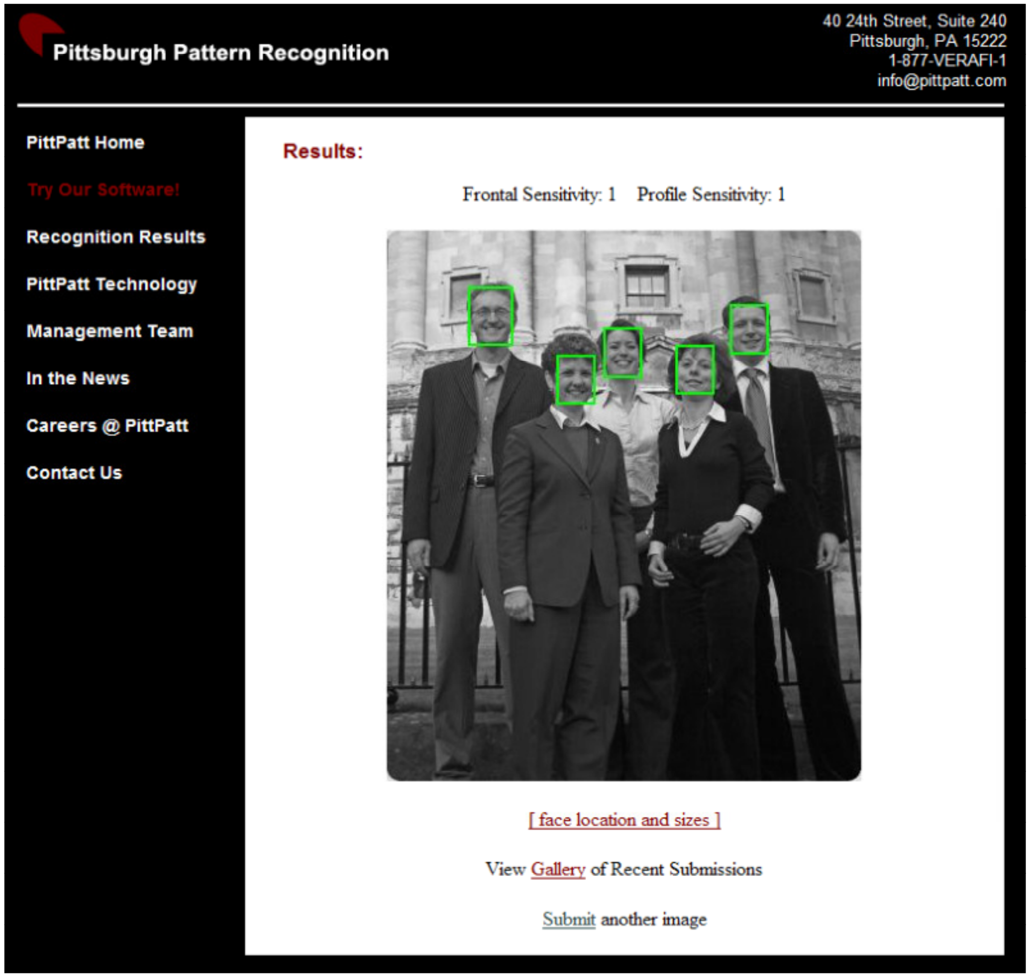
Object Category Recognition





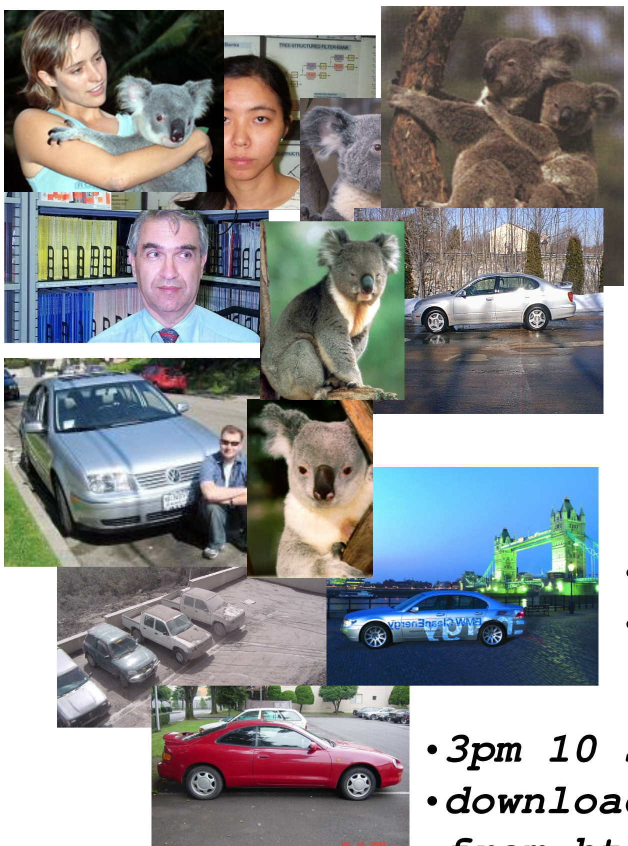
Demos

- Pittpatt http://demo.pittpatt.com/



It’s not just vision…

Integrate with mobile sensor information (GPS, time, nearby object or people), calendar, schedule…
Infer semantically rich meta-data labels from joint sources.



- **10am 7 Sep 05**
- **Australian park**
- **Jim, Jill nearby**

- **4pm 8 Sep 05**
- **Sydney**

- **8pm 10 Oct 05**
- **London**

- **3pm 10 Sep 05**
- **downloaded from http://...**

“two koalas seen on nat. park trip with Jim and Jill”

“Jill and koala on nat. park trip”

“John and his new car”

“office parking lot”

“car to consider purchasing”

Courtesy of Kristen Grauman http://www.cs.utexas.edu/~grauman/

Summary

- The object recognition problem
- Object recognition as pattern classification
- Object recognition grown up

- Readings: G&W 12.1-12.2
- Reference: Duda, Hart, Stork, "Pattern Classification", 2nd Ed.

- Next time: Image Compression

Other acknowledgements: Dan Ellis, EE6820 Slides; Duda, Hart& Stork, Pattern Classification 2nd Ed., David Claus and Christoph F. Eick: Nearest Neighbor Editing and Condensing Techniques.