

EE795: Computer Vision and Intelligent Systems

Spring 2013

TTh 17:30-18:45 FDH 204

Lecture 18

130404

Outline

- Object Recognition Intro (Chapter 14)

Slides from Steve Seitz, Washington

- Excellent References

- <http://people.csail.mit.edu/torralba/shortCourseRLOC/index.html>
- <http://web.eecs.umich.edu/~silvio/teaching/lectures/lecture19.pdf>

Object Recognition Introduction



The "Margaret Thatcher Illusion", by Peter Thompson

Object Recognition Introduction



The "Margaret Thatcher Illusion", by Peter Thompson

What do we mean by “object recognition”?

Next 15 slides adapted from Li, Fergus, & Torralba’s excellent [short course](#) on category and object recognition



Verification: is that a lamp?



Detection: are there people?



Identification: is that Potala Palace?



Object categorization



mountain

tree

building

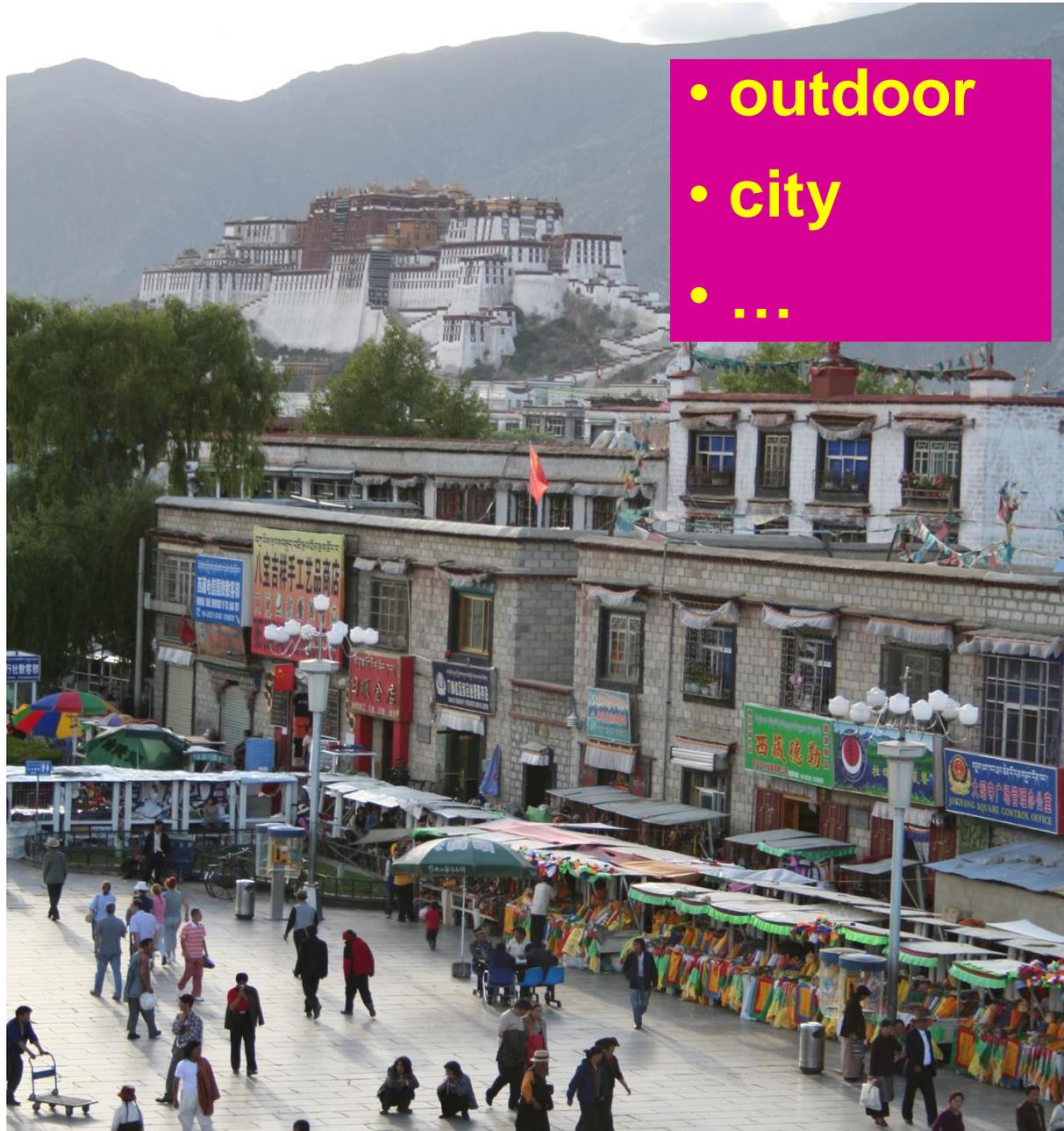
banner

street lamp

vendor

people

Scene and context categorization



- outdoor
- city
- ...

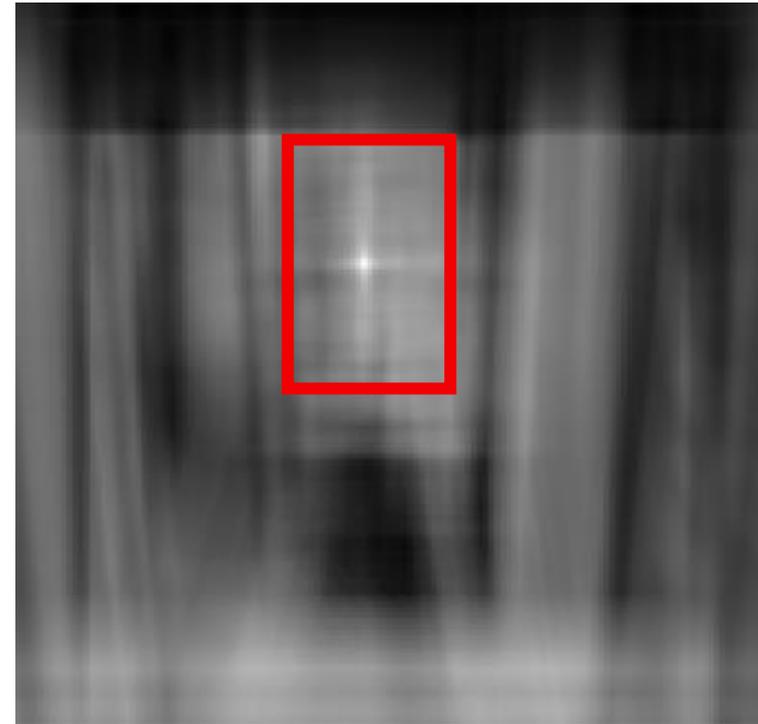
Object recognition

Is it really so hard?

Find the chair in this image



Output of normalized correlation



This is a chair

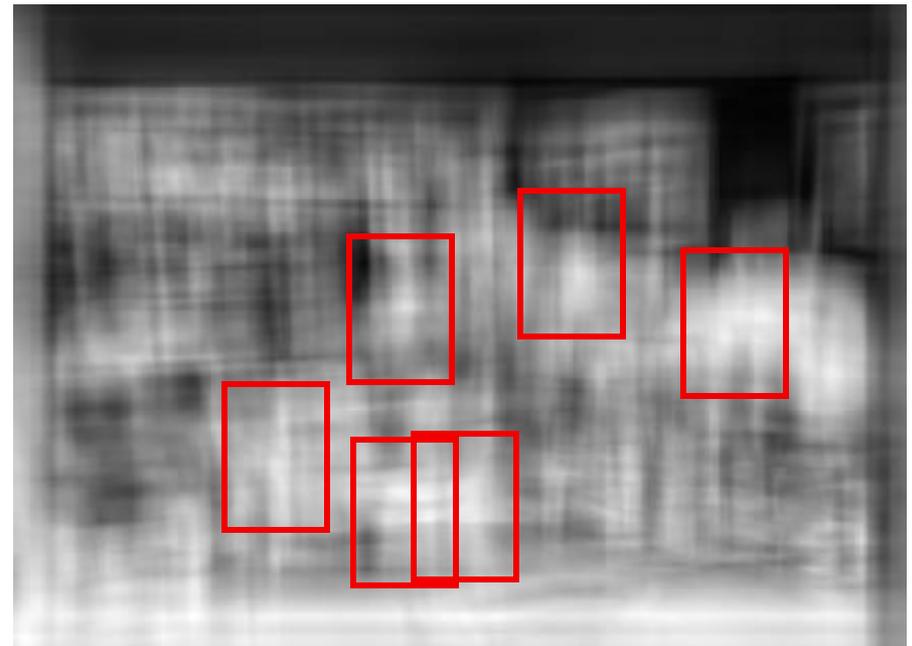
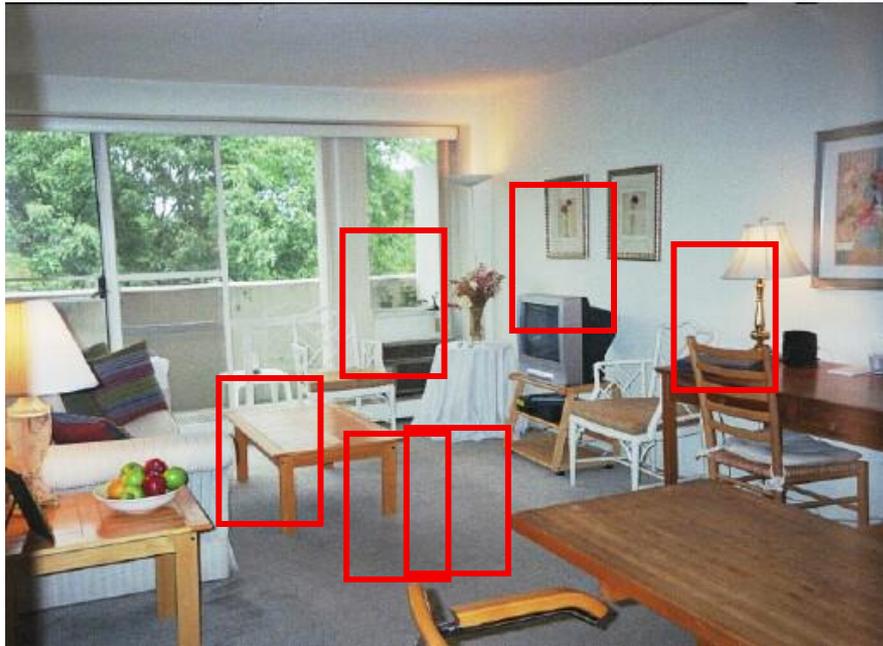




Object recognition

Is it really so hard?

Find the chair in this image



Pretty much garbage
Simple template matching is not going to make it



Object recognition

Is it really so hard?

Find the chair in this image



A “popular method is that of template matching, by point to point correlation of a model pattern with the image pattern. These techniques are inadequate for three-dimensional scene analysis for many reasons, such as occlusion, changes in viewing angle, and articulation of parts.” Nivatia & Binford, 1977.

Why not use SIFT matching for everything?

- Works well for object *instances*



- Not great for generic object *categories*



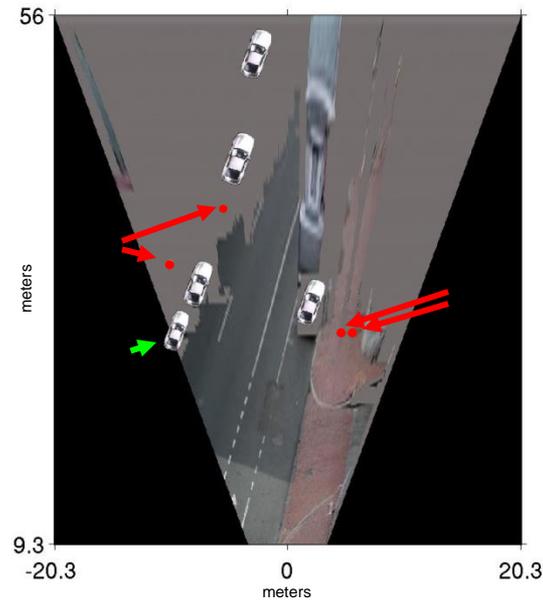
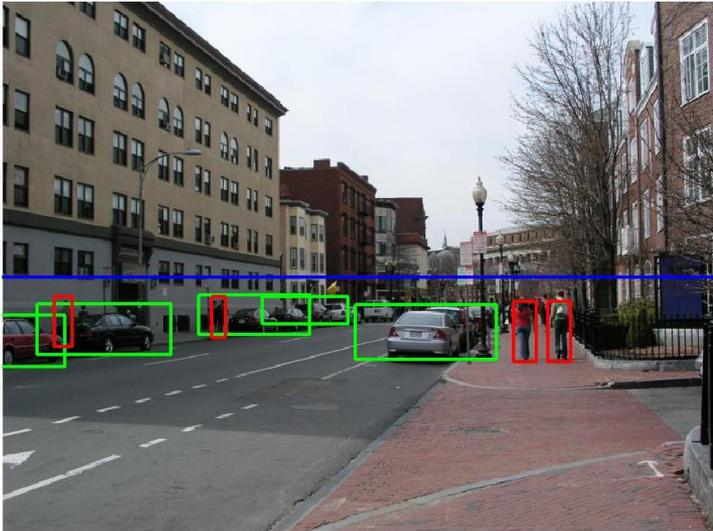
Applications: Computational photography



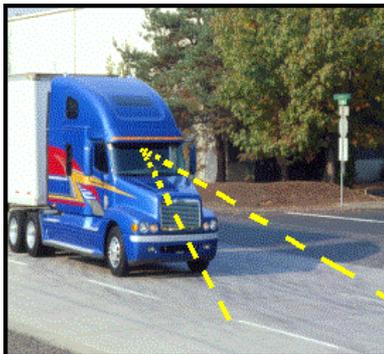
[Face priority AE] When a bright part of the face is too bright

Applications: Assisted driving

Pedestrian and car detection



Lane detection



- Collision warning systems with adaptive cruise control,
- Lane departure warning systems,
- Rear object detection systems,

Applications: image search



Search images

Places

- [London](#)
- [New York](#)
- [Egypt](#)
- [Forbidden City](#)

Celebrities

- [Michael Jordan](#)
- [Angelina Jolie](#)
- [Halle Berry](#)
- [Seth Rogan](#)
- [Rihanna](#)

Art

- [impressionism](#)
- [Keith Haring](#)
- [cubism](#)
- [Salvador Dali](#)
- [pointillism](#)

Shopping

- [evening gown](#)
- [necklace](#)
- [shoes](#)

Refine your image search with visual similarity

Similar Images allows you to search for images using pictures rather than words. Click the "[Similar images](#)" link under an image to find other images that look like it. Try a search of your own or click on an example below.

paris



[Similar images](#)



[Similar images](#)



[Similar images](#)



[Similar images](#)

temple



[Similar images](#)



[Similar images](#)



[Similar images](#)

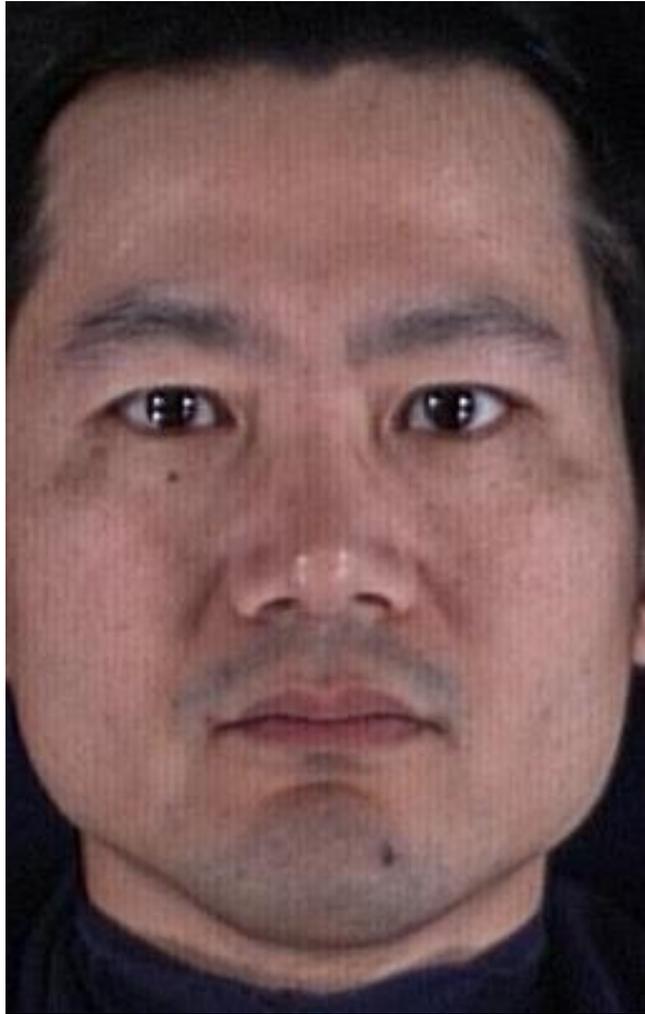


[Similar images](#)

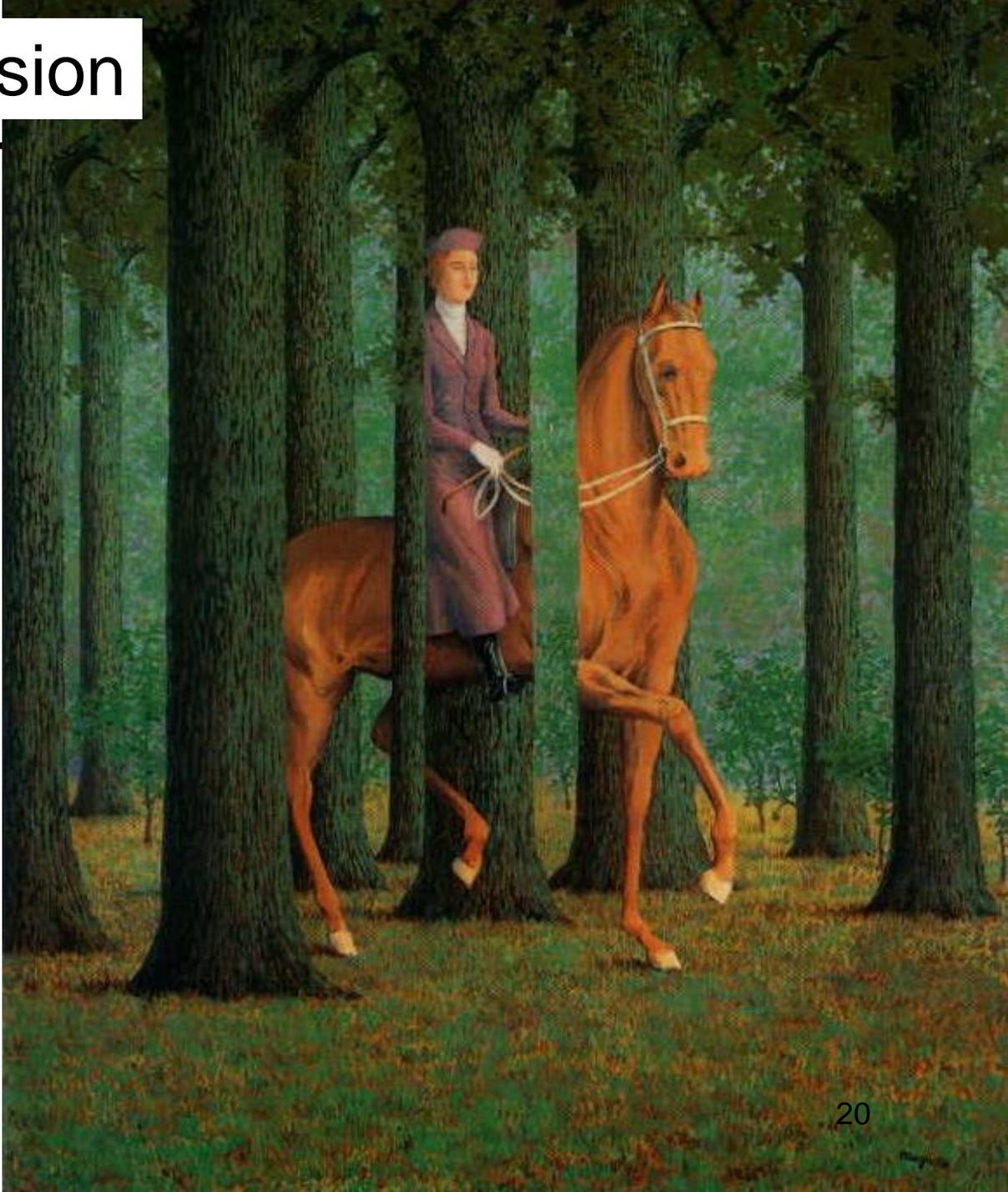
Challenges: viewpoint variation



Challenges: illumination variation



Challenges: occlusion

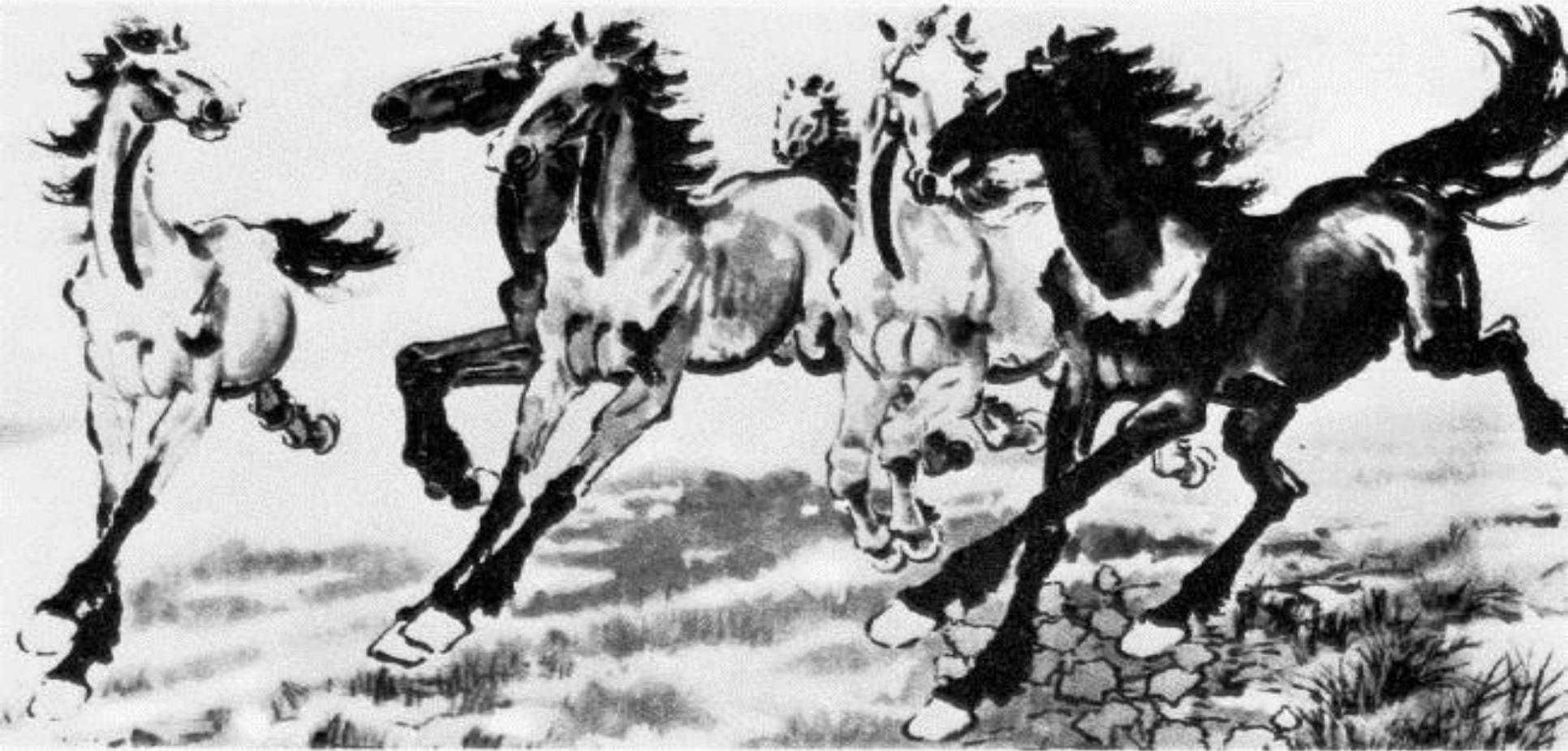


Magritte, 1957

Challenges: scale



Challenges: deformation

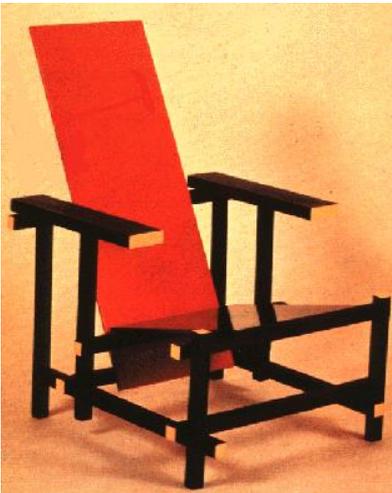


Challenges: background clutter



Klimt, 1913

Challenges: intra-class variation



Recognition problems

- What is it?
 - Object and scene recognition
- Who is it?
 - Identity recognition
- Where is it?
 - Object detection
- What are they doing?
 - Activities
- All of these are **classification** problems
 - Choose one class from a list of possible candidates

What is recognition?

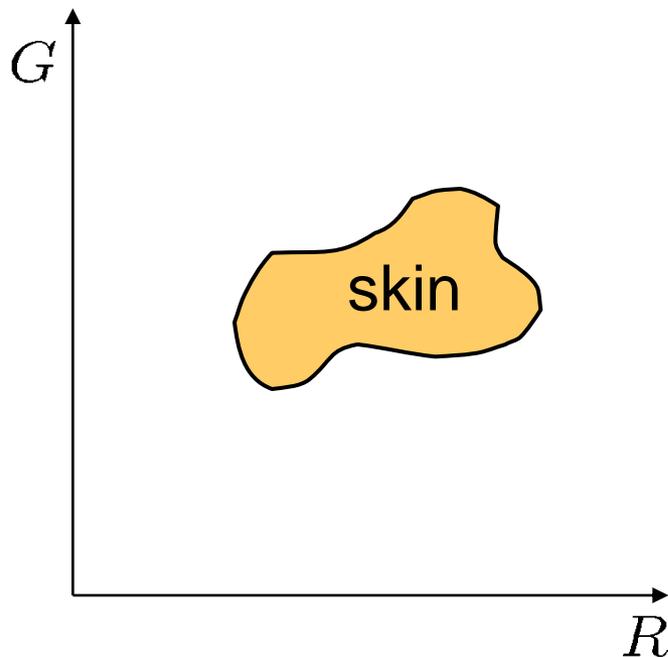
- A different taxonomy from [Csurka *et al.* 2006]:
- Recognition
 - Where is *this* particular object?
- Categorization
 - What *kind* of object(s) is(are) present?
- Content-based image retrieval
 - Find me something that looks similar
- Detection
 - Locate *all* instances of a given class

Face detection



How to tell if a face is present?

One simple method: skin detection



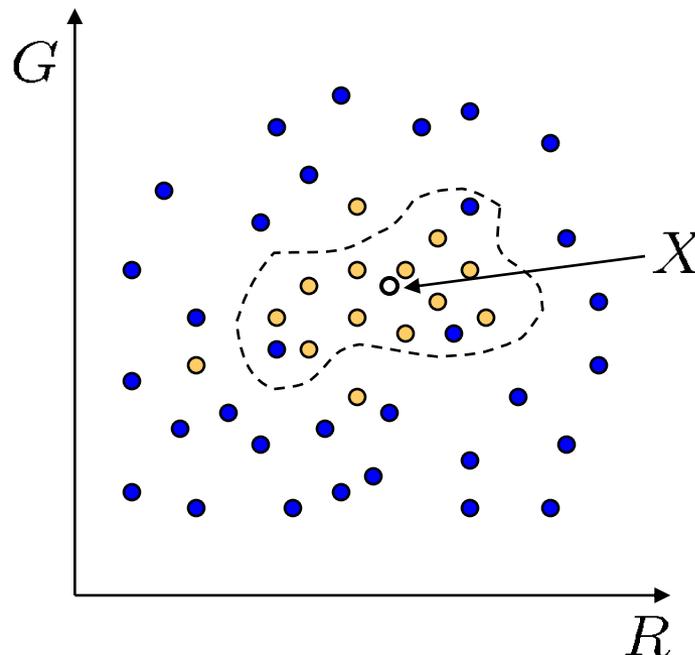
Skin pixels have a distinctive range of colors

- Corresponds to region(s) in RGB color space
 - for visualization, only R and G components are shown above

Skin classifier

- A pixel $X = (R, G, B)$ is skin if it is in the skin region
- But how to find this region?

Skin detection



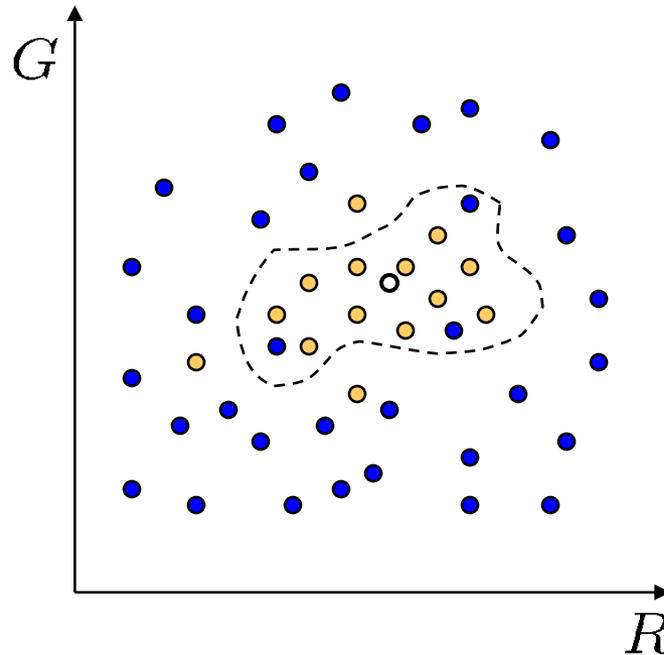
Learn the skin region from examples

- Manually label pixels in one or more “training images” as skin or not skin
- Plot the training data in RGB space
 - skin pixels shown in orange, non-skin pixels shown in blue
 - some skin pixels may be outside the region, non-skin pixels inside. Why?

Skin classifier

- Given $X = (R, G, B)$: how to determine if it is skin or not?

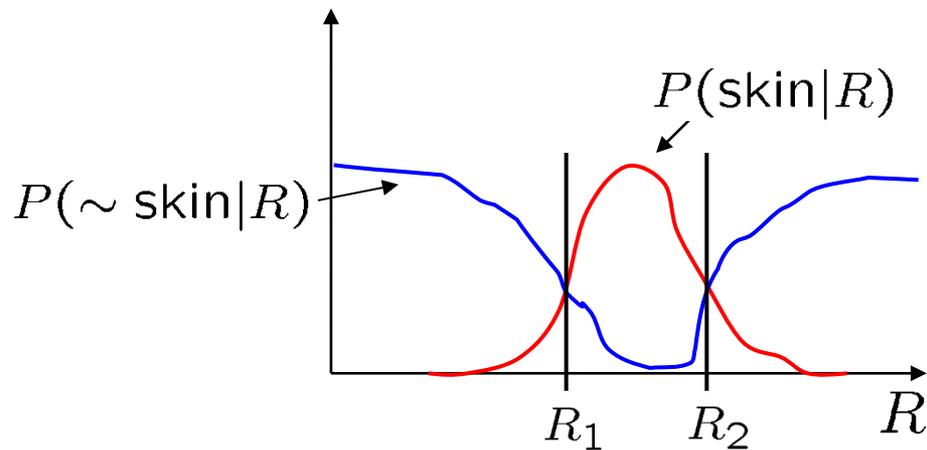
Skin classification techniques



Skin classifier

- Given $X = (R, G, B)$: how to determine if it is skin or not?
- Nearest neighbor
 - find labeled pixel closest to X
 - choose the label for that pixel
- Data modeling
 - fit a model (curve, surface, or volume) to each class
- Probabilistic data modeling
 - fit a probability model to each class

Probabilistic skin classification



Now we can model uncertainty

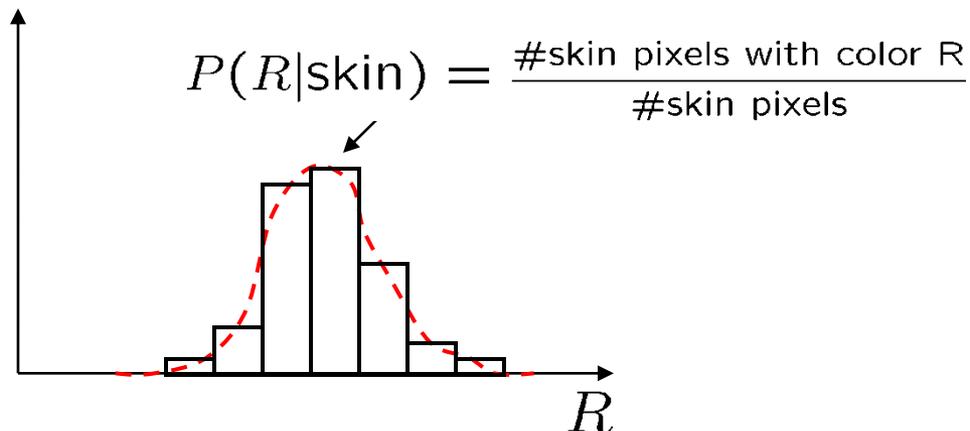
- Each pixel has a probability of being skin or not skin
 - $P(\sim \text{skin}|R) = 1 - P(\text{skin}|R)$

Skin classifier

- Given $X = (R,G,B)$: how to determine if it is skin or not?
- Choose interpretation of highest probability
 - set X to be a skin pixel if and only if $R_1 < X \leq R_2$

Where do we get $P(\text{skin}|R)$ and $P(\sim \text{skin}|R)$?

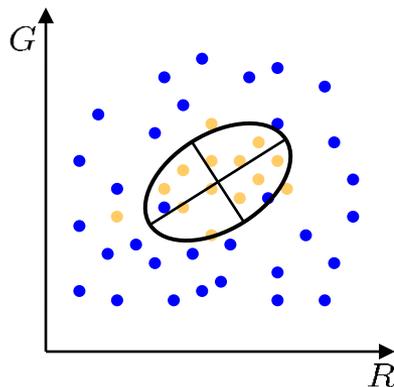
Learning conditional PDF's



We can calculate $P(R | \text{skin})$ from a set of training images

- It is simply a histogram over the pixels in the training images
 - each bin R_i contains the proportion of skin pixels with color R_i

This doesn't work as well in higher-dimensional spaces. Why not?

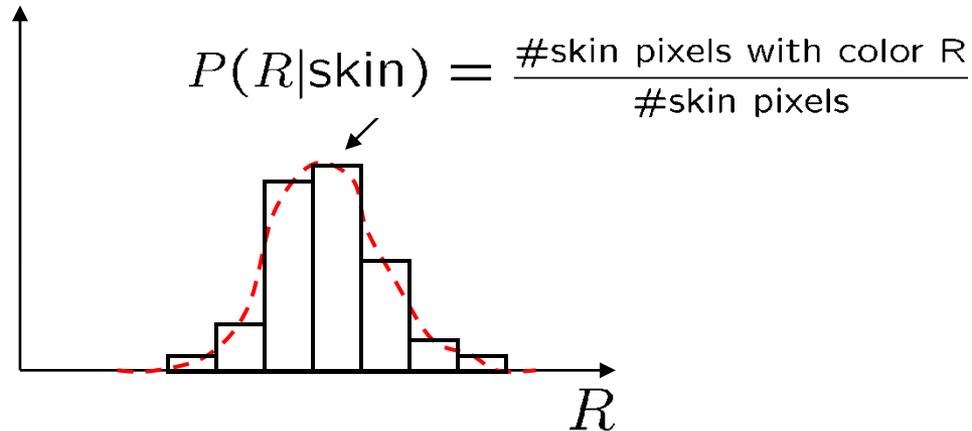


Approach: fit parametric PDF functions

- common choice is rotated Gaussian
 - center $c = \bar{X}$
 - covariance $\sum_X (X - \bar{X})(X - \bar{X})^T$

» orientation, size defined by eigenvecs, eigenvals

Learning conditional PDF's



We can calculate $P(R | \text{skin})$ from a set of training images

- It is simply a histogram over the pixels in the training images
 - each bin R_i contains the proportion of skin pixels with color R_i

But this isn't quite what we want

- Why not? How to determine if a pixel is skin?
- We want $P(\text{skin} | R)$ not $P(R | \text{skin})$
- How can we get it?

Bayes rule

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

In terms of our problem:

$$P(\text{skin}|R) = \frac{P(R|\text{skin}) P(\text{skin})}{P(R)}$$

what we measure
(likelihood) domain knowledge
(prior)

what we want
(posterior)

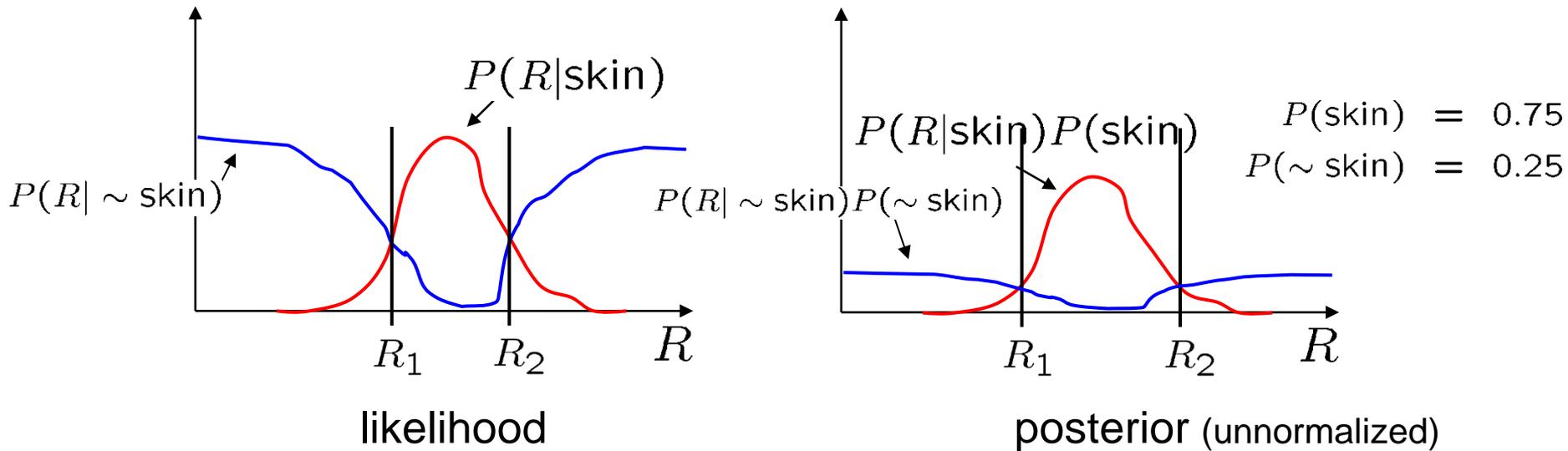
normalization term

$$P(R) = P(R|\text{skin})P(\text{skin}) + P(R|\sim \text{skin})P(\sim \text{skin})$$

The prior: $P(\text{skin})$

- Could use domain knowledge
 - $P(\text{skin})$ may be larger if we know the image contains a person
 - for a portrait, $P(\text{skin})$ may be higher for pixels in the center
- Could learn the prior from the training set. How?
 - $P(\text{skin})$ may be proportion of skin pixels in training set

Bayesian estimation



Bayesian estimation = minimize probability of misclassification

- Goal is to choose the label (skin or \sim skin) that maximizes the posterior
 - this is called **Maximum A Posteriori (MAP) estimation**
- Suppose the prior is uniform: $P(\text{skin}) = P(\sim \text{skin}) = 0.5$
 - in this case $P(\text{skin}|R) = cP(R|\text{skin})$, $P(\sim \text{skin}|R) = cP(R|\sim \text{skin})$
 - maximizing the posterior is equivalent to maximizing the likelihood
 - » $P(\text{skin}|R) > P(\sim \text{skin}|R)$ if and only if $P(R|\text{skin}) > P(R|\sim \text{skin})$
 - this is called **Maximum Likelihood (ML) estimation**

Skin detection results

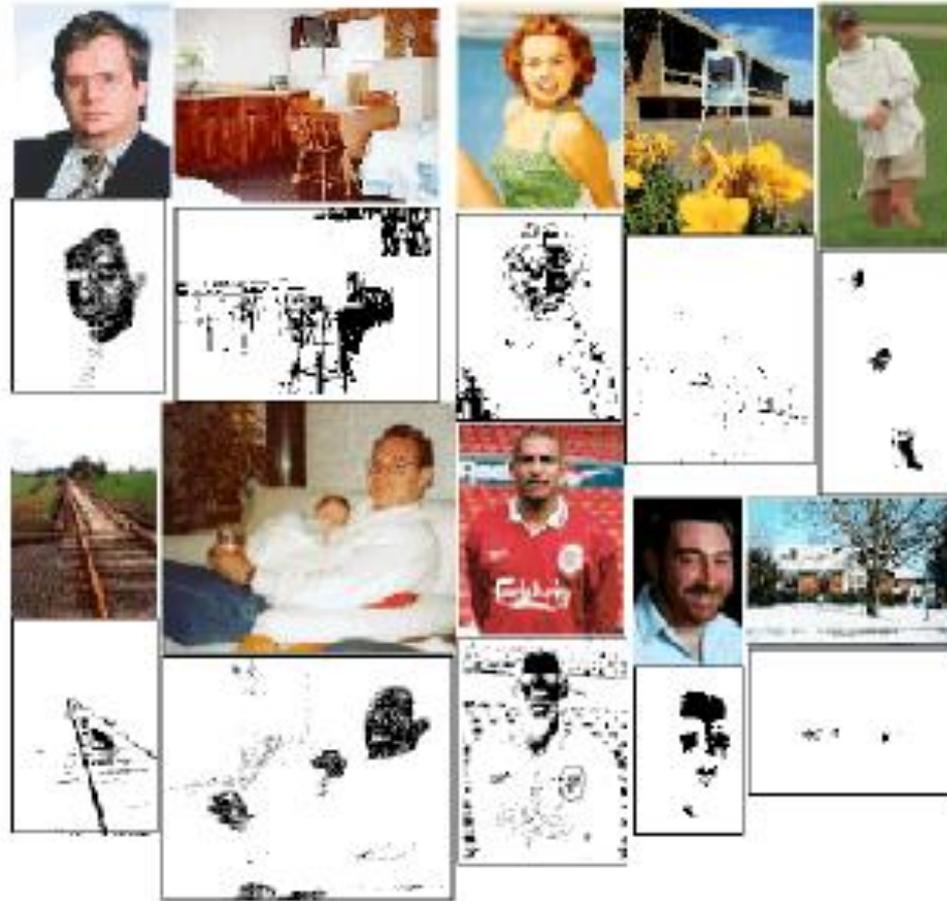
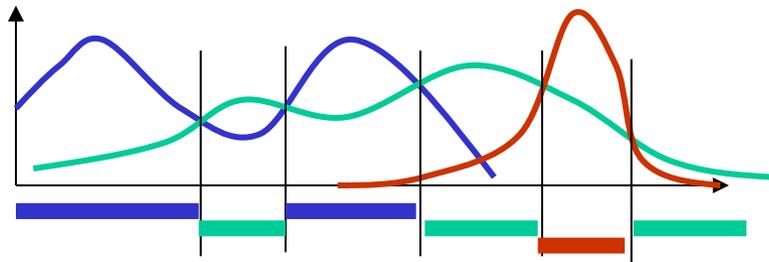


Figure 25.3. The figure shows a variety of images together with the output of the skin detector of Jones and Rehg applied to the image. Pixels marked black are skin pixels, and white are background. Notice that this process is relatively effective, and could certainly be used to focus attention on, say, faces and hands. *Figure from "Statistical color models with application to skin detection," M.J. Jones and J. Rehg, Proc. Computer Vision and Pattern Recognition, 1999 © 1999, IEEE*

General classification

This same procedure applies in more general circumstances

- More than two classes
- More than one dimension



Example: face detection

- Here, X is an image region
 - dimension = # pixels
 - each face can be thought of as a point in a high dimensional space

