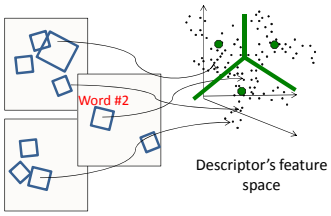

## Visual words

- Map high-dimensional descriptors to tokens/words by quantizing the feature space

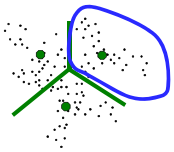



- Quantize via clustering, let cluster centers be the prototype "words"
- Determine which word to assign to each new image region by finding the closest cluster center.

2 Kristen Grauman

## Visual words

- Example: each group of patches belongs to the same visual word



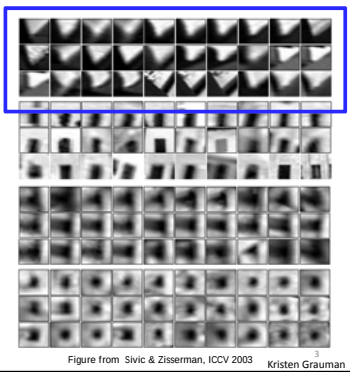


Figure from Sivic & Zisserman, ICCV 2003

3 Kristen Grauman

## Inverted file index


| Word # | Image # |
|---|---|
| 1 | 3 |
| 2 | |
| … | |
| 7 | 1, 2 |
| 8 | 3 |
| 9 | |
| 10 | |
| … | |
| 91 | 2 |

…

- Database images are loaded into the index mapping words to image numbers

4
Kristen Grauman

## Inverted file index

New query image

| Word # | Image # |
|---|---|
| 1 | 3 |
| 2 | |
| … | |
| 7 | 1, 2 |
| 8 | 3 |
| 9 | |
| 10 | |
| … | |
| 91 | 2 |

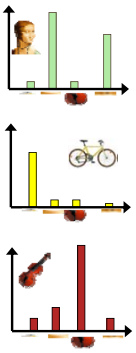
…

- New query image is mapped to indices of database images that share a word.

5
Kristen Grauman

## Bags of visual words

- Summarize entire image based on its distribution (histogram) of word occurrences.
- Analogous to bag of words representation commonly used for documents.

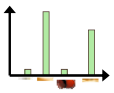


6

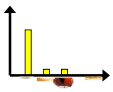
## Comparing bags of words

- Rank frames by normalized scalar product between their (possibly weighted) occurrence counts---*nearest neighbor* search for similar images.

[1 8 1 4]



[5 1 1 0]



$\vec{d}_j$

$\vec{q}$

$$sim(d_j, q) = \frac{\langle d_j, q \rangle}{\|d_j\| \|q\|}$$

$$= \frac{\sum_{i=1}^{V} d_j(i) * q(i)}{\sqrt{\sum_{i=1}^{V} d_j(i)^2} * \sqrt{\sum_{i=1}^{V} q(i)^2}}$$

for vocabulary of *V* words

7

Kristen Grauman

## Application: Large-Scale Retrieval

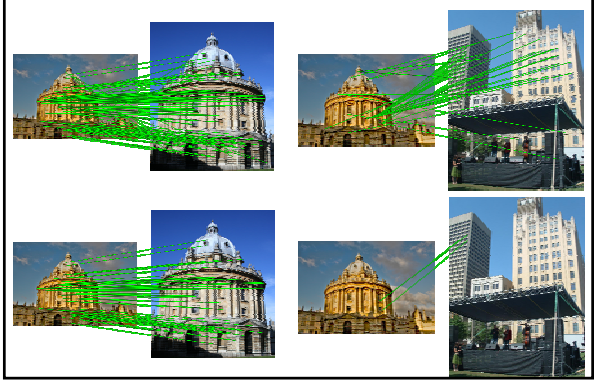
Query

Results from 5k Flickr images (demo available for 100k set)

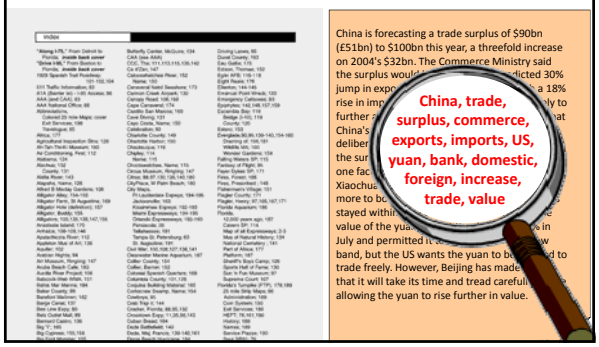
8

[Philbin CVPR'07]

## Spatial Verification: RANSAC



## What else can we borrow from text retrieval?




## *tf-idf* weighting

- **T**erm **f**requency – **i**nverse **d**ocument **f**requency
- Describe frame by frequency of each word within it, downweight words that appear often in the database
- (Standard weighting for text retrieval)

Number of occurrences of word i in document d

Number of words in document d

$$t_i = \frac{n_{id}}{n_d} \log \frac{N}{n_i}$$

Total number of documents in database

Number of documents word i occurs in, in whole database

11

Kristen Grauman

## Query expansion

Query: ***golf green***

Results:

- How can the grass on the ***greens*** at a ***golf*** course be so perfect?
- For example, a skilled ***golf***er expects to reach the ***green*** on a par-four hole in **...**
- Manufactures and sells synthetic ***golf*** putting ***green***s and mats.

Irrelevant result can cause a `topic drift':

- Volkswagen ***Golf***, 1999, ***Green***, 2000cc, petrol, manual, , hatchback, 94000miles, 2.0 GTi, 2 Registered Keepers, HPI Checked, Air-Conditioning, Front and Rear Parking Sensors, ABS, Alarm, Alloy

12

Slide credit: Ondrej Chum

## Query expansion


Chum, Philbin, Sivic, Isard, Zisserman: Total Recall…, ICCV 2007

Slide credit: Ondrej Chum

13

## Recognition via local-feature based alignment

**Pros**:

- Effective when we are able to find reliable features within clutter
- Great results for matching specific instances

**Cons**:

- Spatial verification as post-processing – not seamless, expensive for large-scale problems
- Not suited for generic category recognition

14

Kristen Grauman

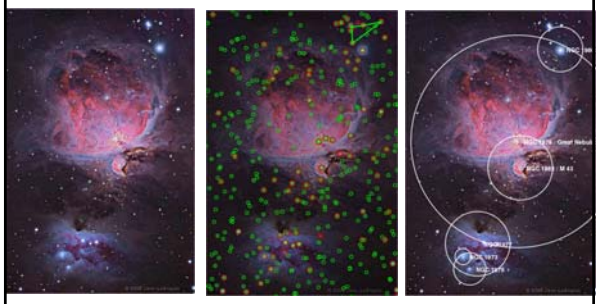
## Making the Sky Searchable: Fast Geometric Hashing for Automated Astrometry

Sam Roweis, Dustin Lang & Keir Mierle
University of Toronto

David Hogg & Michael Blanton
New York University

15

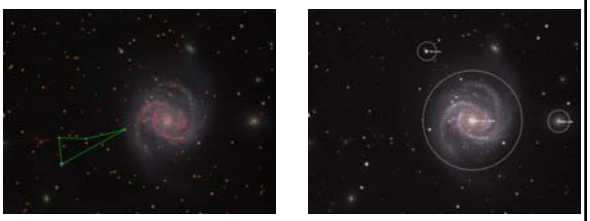
## Example



A shot of the Great Nebula, by Jerry Lodriguss (c.2006), from astropix.com
http://astrometry.net/gallery.html

16

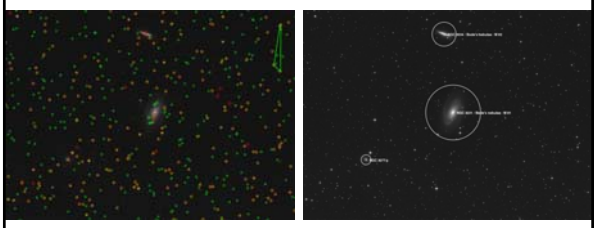
## Example



An amateur shot of M100, by Filippo Ciferri (c.2007) from flickr.com
http://astrometry.net/gallery.html

17

## Example



A beautiful image of Bode's nebula (c.2007) by Peter Bresseler, from starlightfriend.de
http://astrometry.net/gallery.html

18

## Today

- Generic object recognition

19

## What does recognition involve?

20

Source: Fei-Fei Li, Rob Fergus, Antonio Torralba.

## Verification: is that a lamp?

21

Source: Fei-Fei Li, Rob Fergus, Antonio Torralba.

Detection: are there people?

Source: Fei-Fei Li, Rob Fergus, Antonio Torralba.

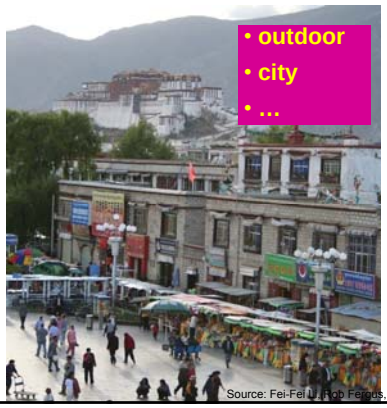
Identification: is that Potala Palace?

Source: Fei-Fei Li, Rob Fergus, Antonio Torralba.

Object categorization


Source: Fei-Fei Li, Rob Fergus, Antonio Torralba.

## Scene and context categorization




Source: Fei-Fei Li, Rob Fergus, Antonio Torralba.

25

## *Instance-level* recognition problem

John's car

26

## *Generic* categorization problem

27

## Object Categorization

- Task Description
  - "Given a small number of training images of a category, recognize a-priori unknown instances of that category and assign the correct category label."
- Which categories are feasible visually?

"Fido" → German shepherd → dog → animal → living being

Visual Object Recognition Tutorial

K. Grauman, B. Leibe

28

## Visual Object Categories

- Basic Level Categories in human categorization [Rosch 76, Lakoff 87]
  - The highest level at which category members have similar perceived shape
  - The highest level at which a single mental image reflects the entire category
  - The level at which human subjects are usually fastest at identifying category members
  - The first level named and understood by children
  - The highest level at which a person uses similar motor actions for interaction with category members

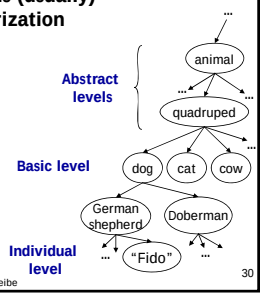
Visual Object Recognition Tutorial

K. Grauman, B. Leibe

29

## Visual Object Categories

- Basic-level categories in humans seem to be defined predominantly visually.
- There is evidence that humans (usually) start with basic-level categorization *before* doing identification.

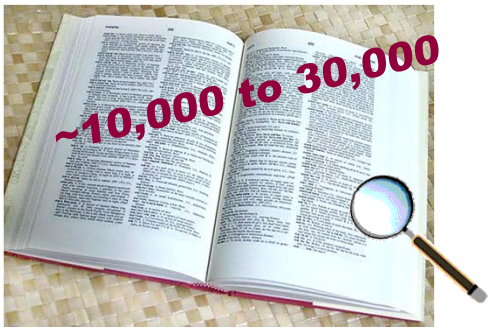



Visual Object Recognition Tutorial

K. Grauman, B. Leibe

30

How many object categories are there?




Source: Fei-Fei Li, Rob Fergus, Antonio Torralba.

Biederman 1987

31


32

## Other Types of Categories

- **Functional Categories**
  - **e.g. chairs =** *"something you can sit on"*

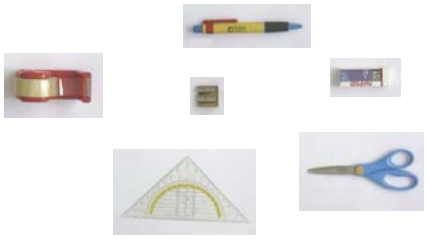
K. Grauman, B. Leibe

Visual Object Recognition Tutorial

33

## Other Types of Categories

- **Ad-hoc categories**
  - **e.g.** *"something you can find in an office environment"*



K. Grauman, B. Leibe

Visual Object Recognition Tutorial

34

## Why recognition?

– Recognition a fundamental part of perception
  - e.g., robots, autonomous agents

– Organize and give access to visual content
  - Connect to information
  - Detect trends and themes

35

## Posing visual queries

Yeh et al., MIT


Belhumeur et al.

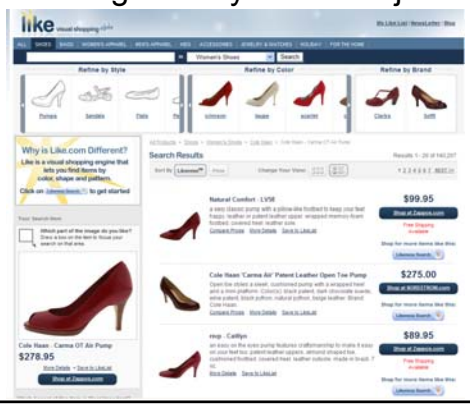

Kooaba, Bay & Quack et al.

36

## Autonomous agents able to detect objects

37

## Finding visually similar objects




38

## Discovering visual patterns

**Objects** Sivic & Zisserman

**Categories** Lee & Grauman

**Actions** Wang et al.

39

Kristen Grauman

## Auto-annotation


Gammeter et al.

T. Berg et al.

40

Kristen Grauman

## Challenges: robustness

**Illumination**

**Object pose**

**Clutter**

**Occlusions**

**Intra-class appearance**

**Viewpoint**

41

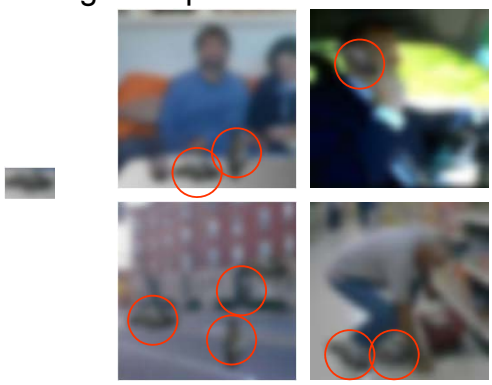
Kristen Grauman

## Challenges: robustness

Realistic scenes are crowded, cluttered, have overlapping objects.

42

## Challenges: importance of context



43

slide credit: Fei-Fei, Fergus & Torralba

## Challenges: importance of context

44

## Challenges: complexity

flickr
10 billion images

facebook
250 billion images

imgur
1 billion images served daily

YouTube
100 hours uploaded per minute

photobucket
10 billion images

cisco

**Almost 90% of web traffic is visual!**

45

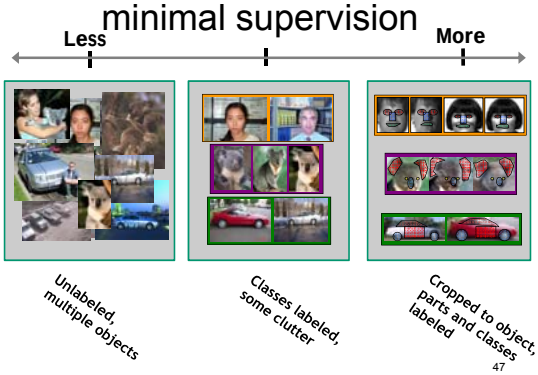
## Challenges: complexity

- Thousands to millions of pixels in an image
- 30+ degrees of freedom in the pose of articulated objects (humans)
- About half of the cerebral cortex in primates is devoted to processing visual information [Felleman and van Essen 1991]

46

Kristen Grauman

## Challenges: learning with minimal supervision




47

Kristen Grauman

## What works today

- Reading license plates, zip codes, checks

48

Source: Lana Lazebnik

## What works today

- Reading license plates, zip codes, checks
- Fingerprint recognition

49

Source: Lana Lazebnik

## What works today

- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection



50

Source: Lana Lazebnik

## What works today

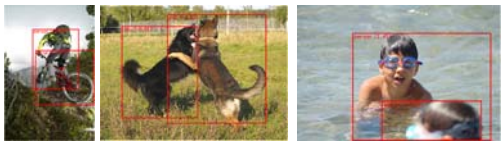
- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection
- Recognition of flat textured objects (CD covers, book covers, etc.)

51

Source: Lana Lazebnik

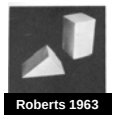
## What works today

- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection
- Recognition of flat textured objects (CD covers, book covers, etc.)
- Recognition of generic categories(*)!



52

## Progress charted by datasets



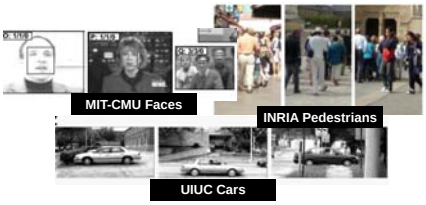
Roberts 1963

COIL

1963 ··· 1996

Slide credit: Kristen Grauman

## Progress charted by datasets



MIT-CMU Faces

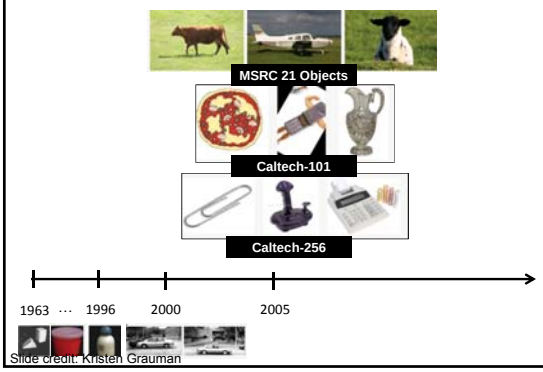
INRIA Pedestrians

UIUC Cars

1963 ··· 1996 2000

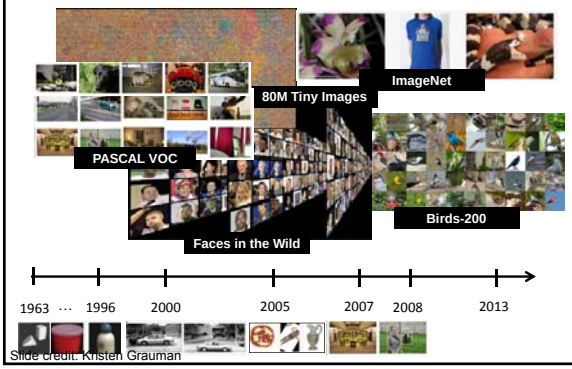
Slide credit: Kristen Grauman

## Progress charted by datasets




Slide credit: Kristen Grauman

## Progress charted by datasets




Slide credit: Kristen Grauman

## Evolution of methods

- Hand-crafted models
- 3D geometry
- Hypothesize and align

- Hand-crafted features
- Learned models
- Data-driven

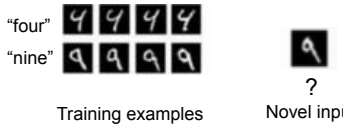
- "End-to-end" learning of features and models*,**

Kristen Grauman

* Labeled data availability
** Architecture design decisions, parameters.

## Supervised classification

- Given a collection of *labeled* examples, come up with a function that will predict the labels of new examples.




- How good is the function we come up with to do the classification?
- Depends on
  - Mistakes made
  - Cost associated with the mistakes

58

Kristen Grauman

## Supervised classification

- Given a collection of *labeled* examples, come up with a function that will predict the labels of new examples.
- Consider the two-class (binary) decision problem
  - L(4→9): Loss of classifying a 4 as a 9
  - L(9→4): Loss of classifying a 9 as a 4
- **Risk** of a classifier *s* is expected loss:

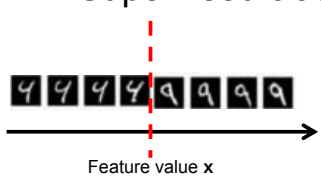
$$R(s) = \Pr(4 \to 9 \mid \text{using } s) L(4 \to 9) + \Pr(9 \to 4 \mid \text{using } s) L(9 \to 4)$$

- We want to choose a classifier so as to minimize this total risk

59

Kristen Grauman

## Supervised classification




Optimal classifier will minimize total risk.

At decision boundary, either choice of label yields same expected loss.

If we choose class "four" for point x at boundary, expected loss is:

$$= P(\text{class is } 9 \mid \mathbf{x})\, L(9 \to 4) + P(\text{class is } 4 \mid \mathbf{x}) L(4 \to 4)$$

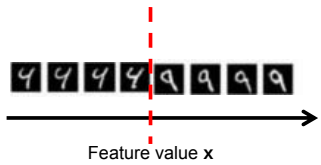
If we choose class "nine" for point x at boundary, expected loss is:

$$= P(\text{class is } 4 \mid \mathbf{x})\, L(4 \to 9)$$

60

Kristen Grauman

## Supervised classification




Optimal classifier will minimize total risk.

At decision boundary, either choice of label yields same expected loss.

So, best decision boundary is at point **x** where

$$P(\text{class is } 9 \mid \mathbf{x})\, L(9 \to 4) = P(\text{class is } 4 \mid \mathbf{x}) L(4 \to 9)$$

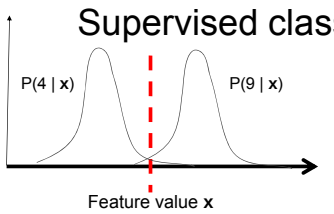
To classify a new point, choose class with lowest expected loss; i.e., choose "four" if

$$P(4 \mid \mathbf{x}) L(4 \to 9) > P(9 \mid \mathbf{x}) L(9 \to 4)$$

61

Kristen Grauman

## Supervised classification




Optimal classifier will minimize total risk.

At decision boundary, either choice of label yields same expected loss.

So, best decision boundary is at point **x** where

$$P(\text{class is } 9 \mid \mathbf{x})\, L(9 \to 4) = P(\text{class is } 4 \mid \mathbf{x}) L(4 \to 9)$$

To classify a new point, choose class with lowest expected loss; i.e., choose "four" if

$$\boxed{P(4 \mid \mathbf{x})} L(4 \to 9) > \boxed{P(9 \mid \mathbf{x})} L(9 \to 4)$$

*How to evaluate these probabilities?*

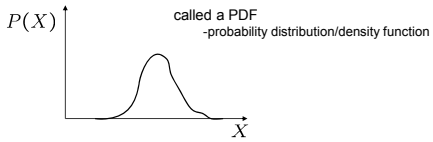
62

Kristen Grauman

## Probability

Basic probability

- X is a random variable
- P(X) is the probability that X achieves a certain value




- $0 \le P(X) \le 1$
- $\int_{-\infty}^{\infty} P(X)dX = 1$ or $\sum P(X) = 1$

  continuous X  discrete X

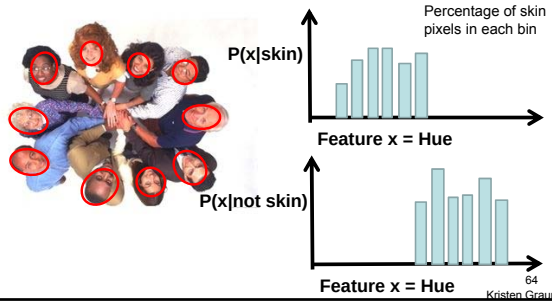
- Conditional probability: P(X | Y)
  - probability of X given that we already know Y

63

Source: Steve Seitz

## Example: learning skin colors

- We can represent a class-conditional density using a histogram (a "non-parametric" distribution)



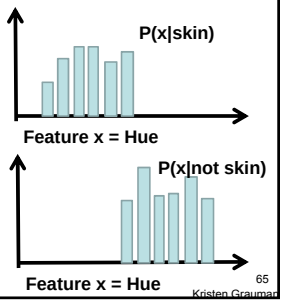

64
Kristen Grauman

## Example: learning skin colors

- We can represent a class-conditional density using a histogram (a "non-parametric" distribution)

Now we get a new image, and want to label each pixel as skin or non-skin.

*What's the probability we care about to do skin detection?*




65
Kristen Grauman

## Bayes rule

$$\underbrace{P(skin \mid x)}_{\text{posterior}} = \frac{\overbrace{P(x \mid skin)}^{\text{likelihood}}\overbrace{P(skin)}^{\text{prior}}}{P(x)}$$

*Where does the prior come from?*

*Why use a prior?*

66

## Example: classifying skin pixels

Now for every pixel in a new image, we can estimate probability that it is generated by skin.

Brighter pixels → higher probability of being skin

Classify pixels based on these probabilities

- if $p(\text{skin}|\boldsymbol{x}) > \theta$, classify as skin
- if $p(\text{skin}|\boldsymbol{x}) < \theta$, classify as not skin

67

## Example: classifying skin pixels

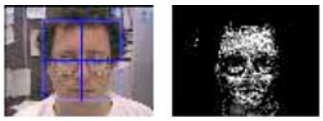


**Figure 6:** A video image and its flesh probability image

**Figure 7:** Orientation of the flesh probability distribution marked on the source video image

Gary Bradski, 1998

68

## Example: classifying skin pixels

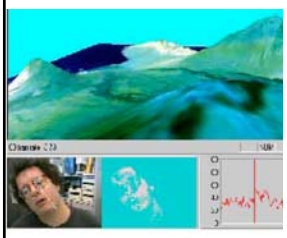


**Figure 13:** CAMSHIFT-based face tracker used to over a 3D graphic's model of Hawaii

**Figure 12:** CAMSHIFT-based face tracker used to play Quake 2 hands free by inserting control variables into the mouse queue

Using skin color-based face detection and pose estimation as a video-based interface

Gary Bradski, 1998

69

## Supervised classification

- Want to minimize the expected misclassification
- Two general strategies
  - Use the training data to build representative probability model; separately model class-conditional densities and priors (*generative*)
  - Directly construct a good decision boundary, model the posterior (*discriminative*)

70

## Coming up

Face detection

Categorization with local features and part-based models

Deep convolutional neural networks

71

## Questions?

See you Tuesday!

72