Hands On: Multimedia Methods for Large Scale Video Analysis (Lecture)

Dr. Gerald Friedland,
fractor@icsi.berkeley.edu

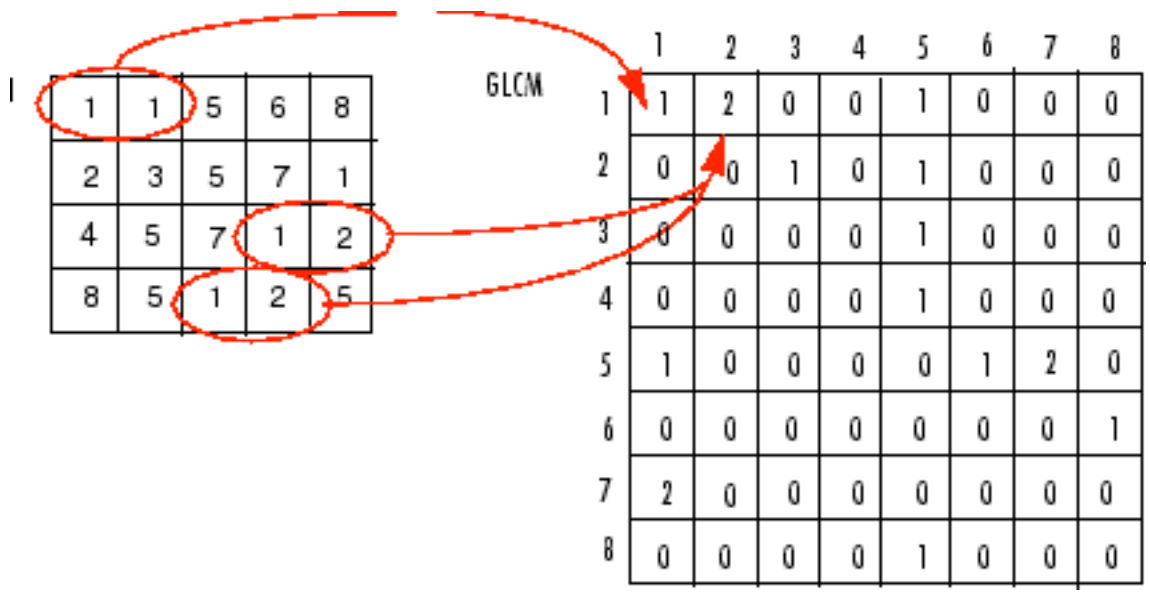
Today

- Response to Questions
- More on Visual Analysis:
  - More on Features
  - Outlook on Systems

FAQ

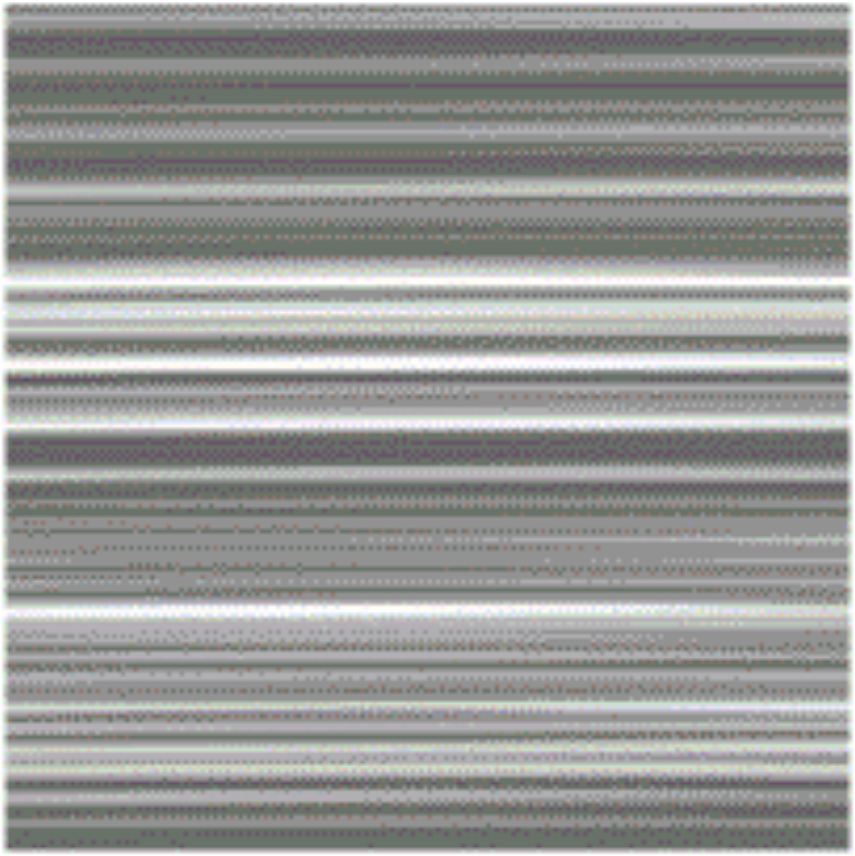
- Co-occurrence Matrix: What's the Intuition?

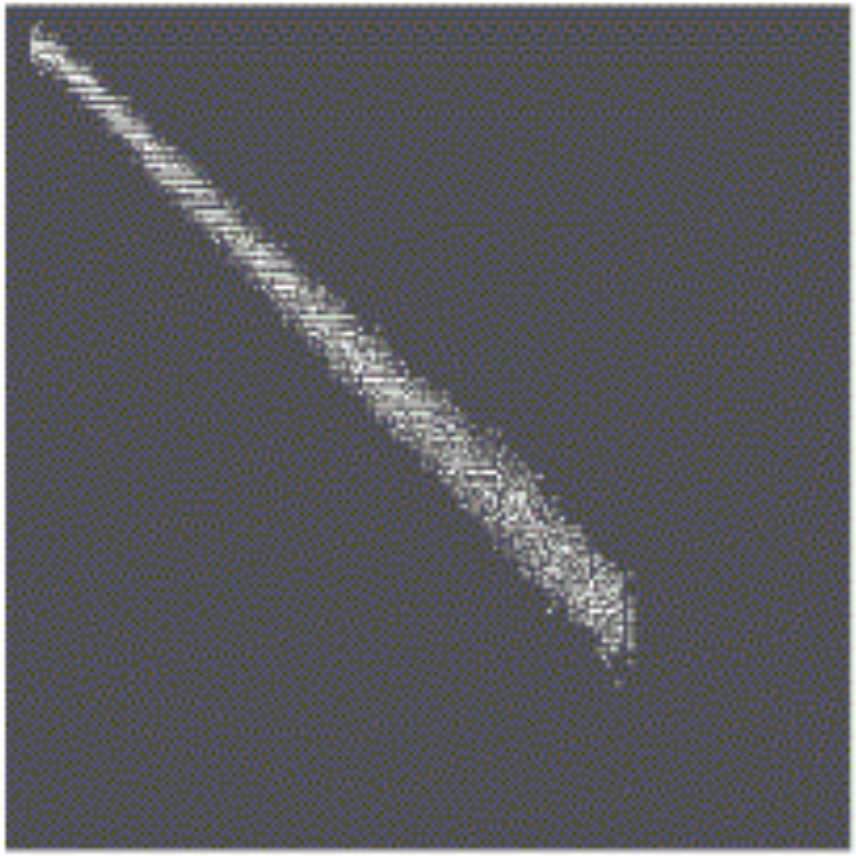
$$C_{\Delta x,\Delta y}(i,j) = \sum_{p=1}^{n}\sum_{q=1}^{m}\begin{cases}1, & \text{if } I(p,q)=i \text{ and } I(p+\Delta x, q+\Delta y)=j \\ 0, & \text{otherwise}\end{cases}$$



Co-Occurrence Matrix: Example

A.





FAQ

- How do you Equalize/Gaussianize?

- Easy Answer: Use Kernel
- Sophisticated Answer: Google Research Papers...

FAQ

- If you hold the Nintendo 3DS upside down, will the 3D effect still work?

See yourself...

More on Features

- Finishing up Basic Features
- More Advanced Features:
  - Color Signatures
  - SIFT
  - GIST

Simple Visual Features

- Edge features:

```
gray_image = rgb2gray(my_image);
double_i = im2double(gray_image);
subplot(221),imshow(gray_image);
edge_image1 =
    edge(double_i,'sobel');
subplot(222),imshow(edge_image1)
edge_image2 =
    edge(double_i,'prewitt');
subplot(223),imshow(edge_image2)
edge_image3 =
    edge(double_i,'canny');
subplot(224),imshow(edge_image3);
```

Advanced Visual Features

- Shape features
  - A good shape representation feature for an object should be invariant to translation, rotation, and scaling
  - The use of shape features for image retrieval has been limited to special applications where objects or regions are readily available
  - Shape description can be categorized into either boundary-based or region-based method

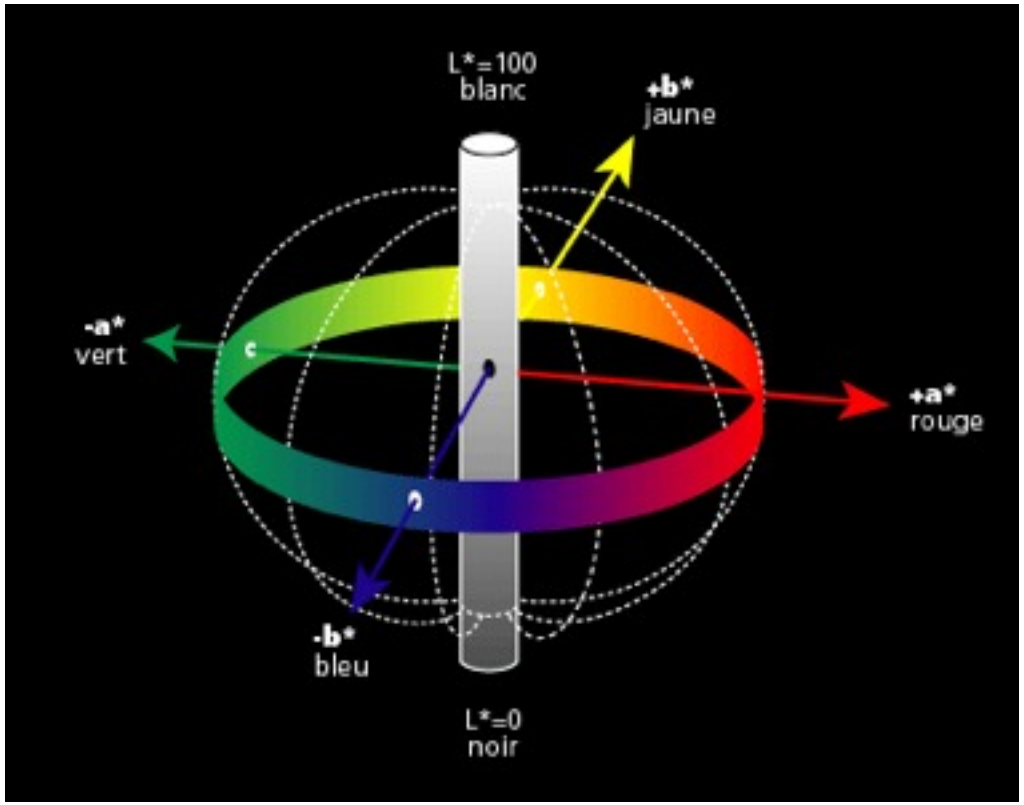
Color Signatures for Retrieval

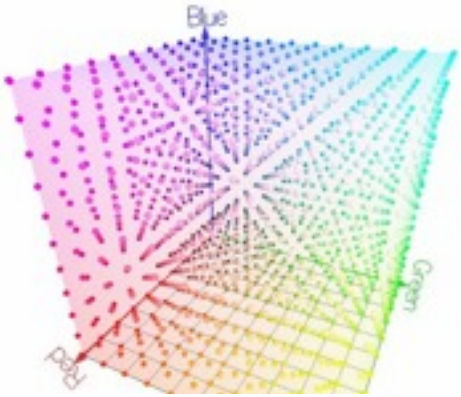
Image retrieval:

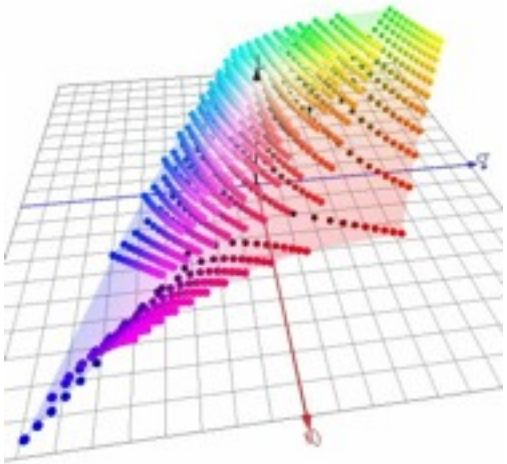
Y. Rubner, C. Tomasi, and L. J. Guibas: The Earth Mover's Distance as a Metric for Image Retrieval. Int. Journal of Computer Vision, 40(2):99–121, 2000.

Color Signatures

CIELAB space:







Create Color Signatures

Find representative colors by equal-size clustering (modified KD-Tree clustering [Rubner et. al.]):

Stage 1:

In each recursion look at one of L, a, and b axis.

Create two equal-sized subintervals.

Repeat until interval is smaller than a threshold.

Calculate the centroids for all clusters.

Stage 2:

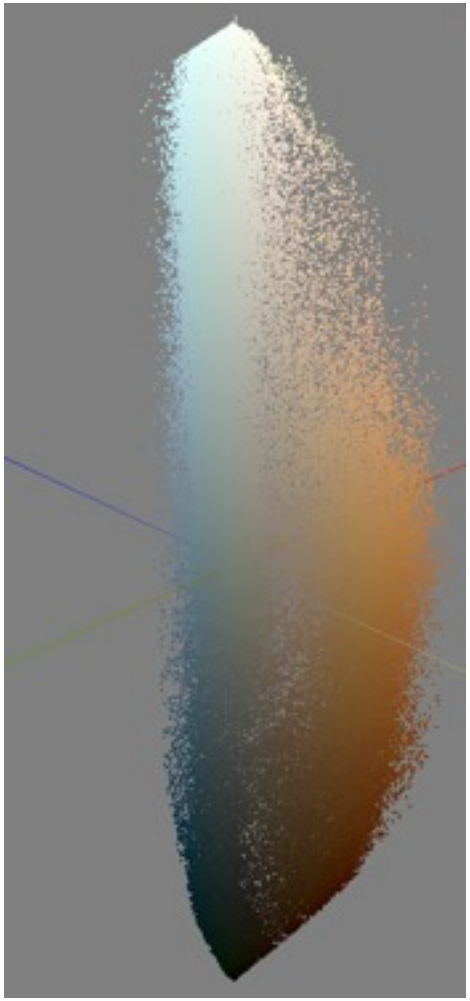
Repeat Step 1 using the cluster centroids.

Eliminate all clusters that represent less points in the image than a threshold.

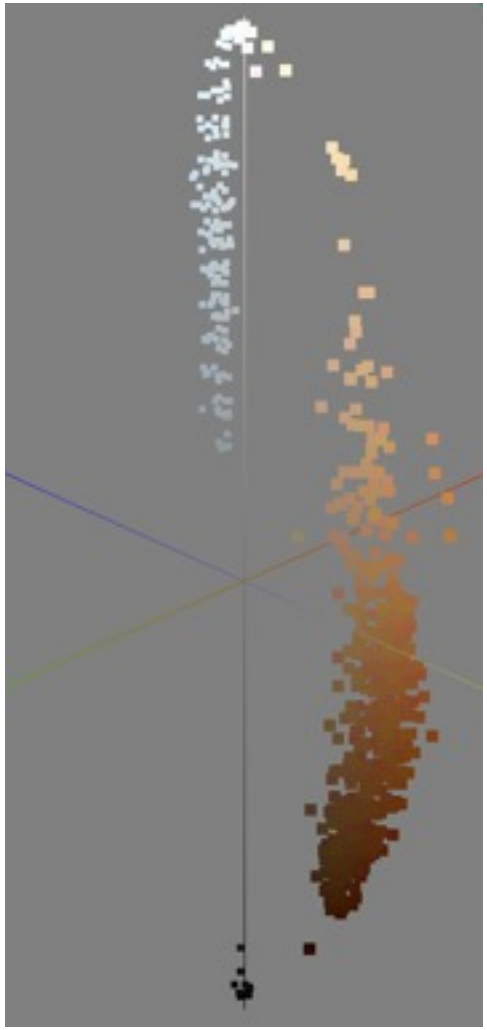
Create Color Signatures

Example:

→



→



Scale Invariant Feature Transform (SIFT)

- Used for Object Recognition:

  Lowe, David G. 'Distinctive Image Features from Scale Invariant Features', International Journal of Computer Vision, Vol. 60, No. 2, 2004, pp. 91-110

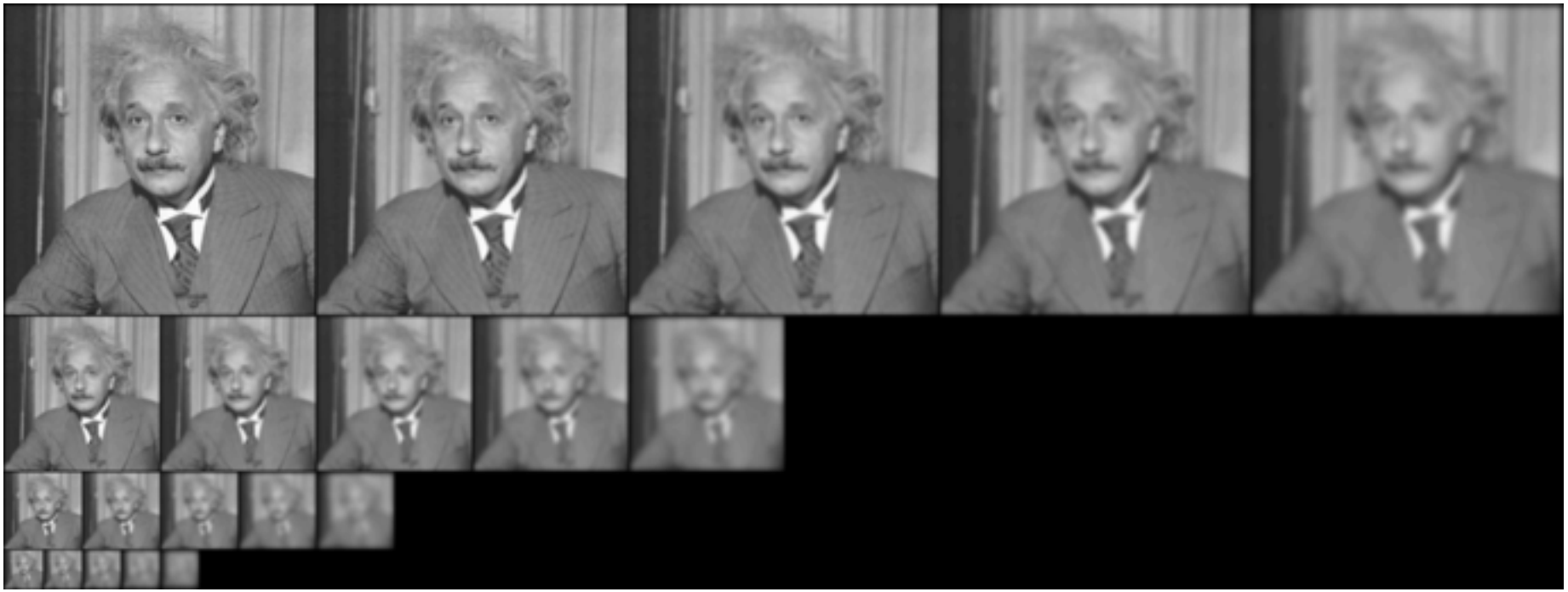
- Mostly invariant to:
  - Scale
  - Rotation
  - Illumination
  - Viewpoint

Scale Invariant Feature Transform (SIFT)

Steps:

- Scaling + blurring
- Finding difference of Gaussians (keypoints)
- Remove “bad keypoints”
- Assign orientation to keypoints
- Rotation and scale normalization

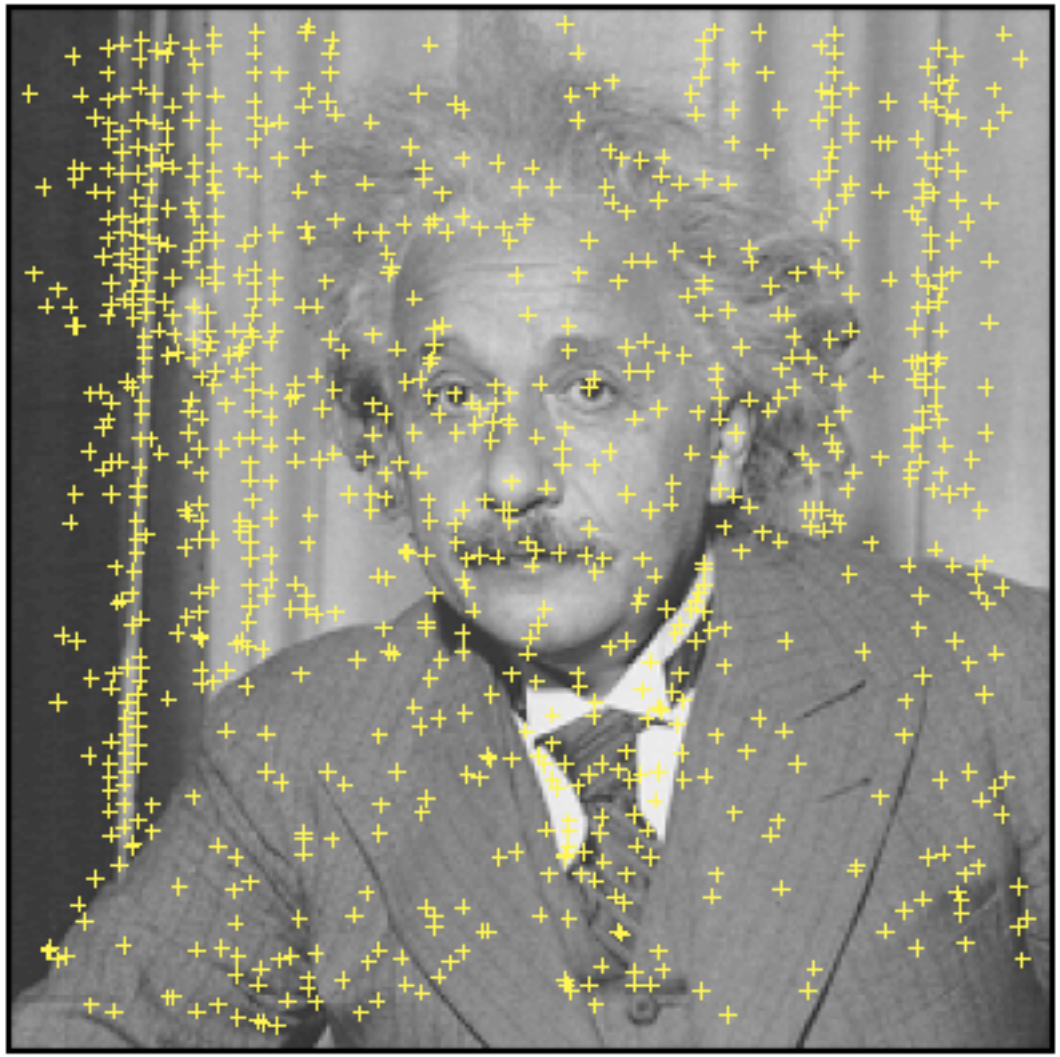
SIFT: Scaling, Blurring



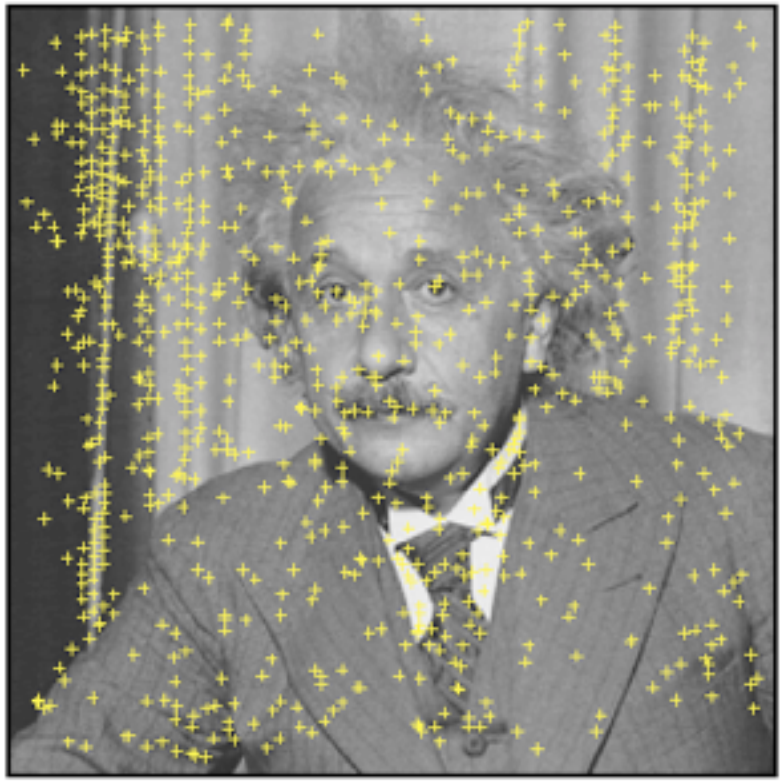
SIFT: Difference of Gaussians

SIFT: Difference of Gaussians

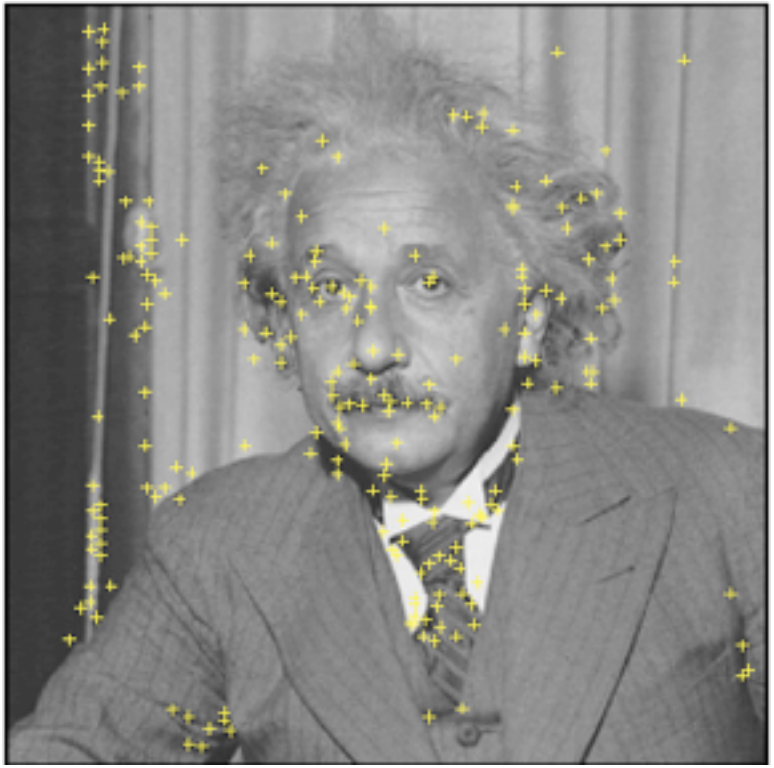
SIFT: Keypoints by Finding Maxima



SIFT: Removing "Bad" Keypoints

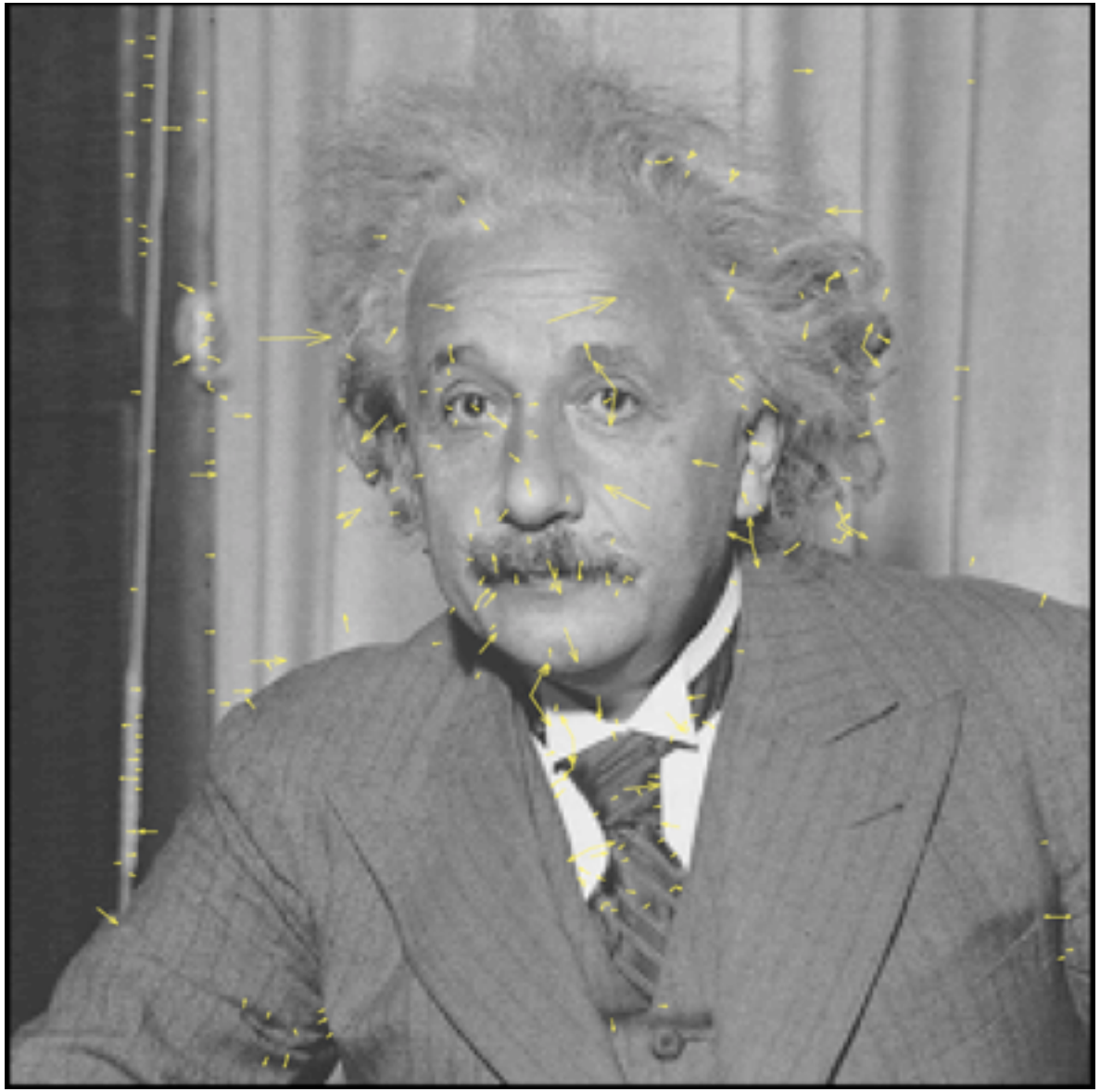


Removing low contrast

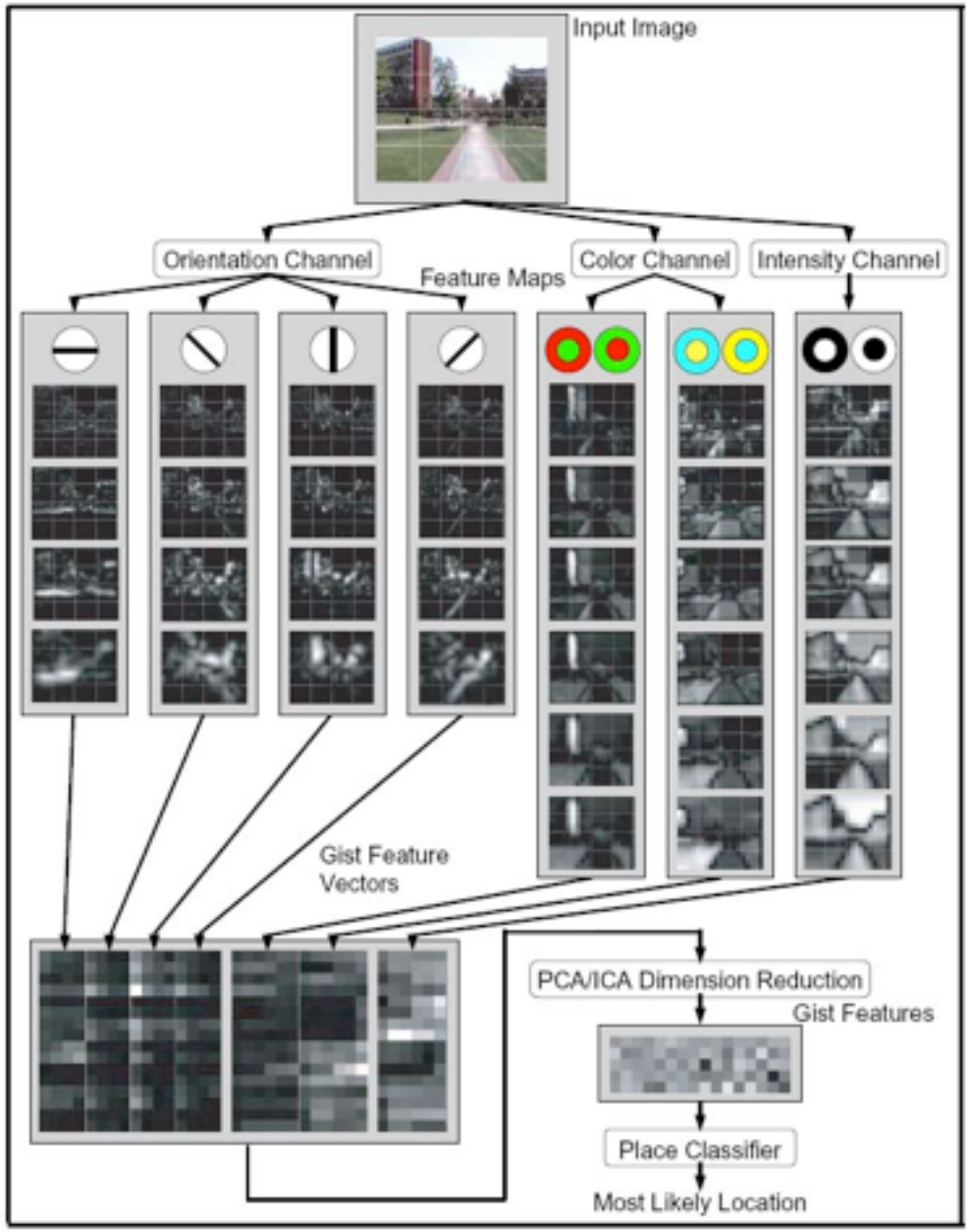


Removing edge responses

SIFT: Normalizing



GIST



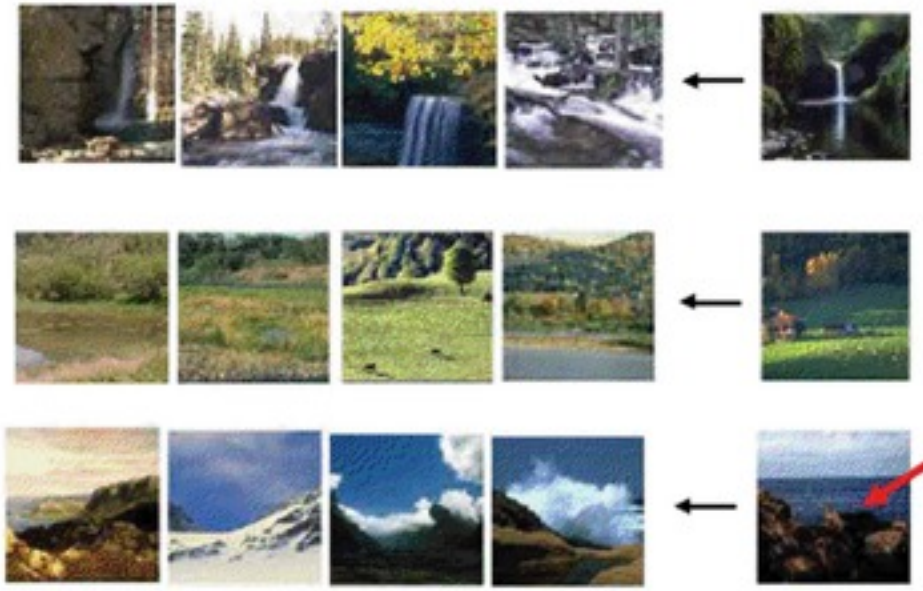
See: A. Oliva, and A. Torralba: Building the Gist of a Scene: The Role of Global Image Features in Recognition, Progress in Brain Research Volume: 155, Issue: 1, Publisher: Elsevier, Pages: 23-36, 2006

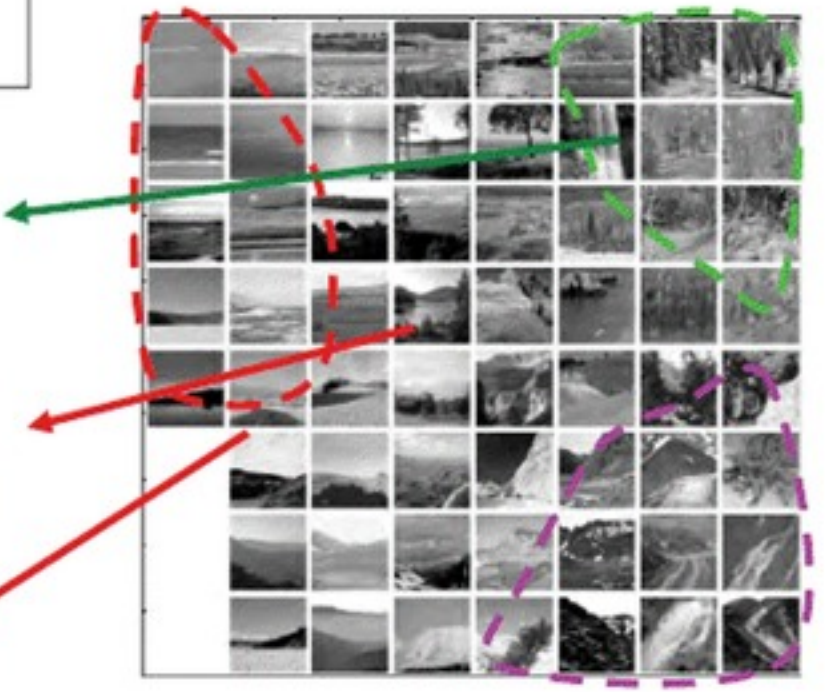
Categorization of Natural Scenes

Confusion Matrix (in % using Layout template) :
Classification of prototypical scenes (400 / category)

	Coast	Countryside	Forest	Mountain
Coast	88.6	8.9	1.2	1.3
Countryside	9.8	85.2	3.7	1.3
Forest	0.4	3.6	91.5	4.5
Mountain	0.4	4.6	3.8	91.2

Local organization: correct for 92 % images
(4 similar images on 7 K-NN)





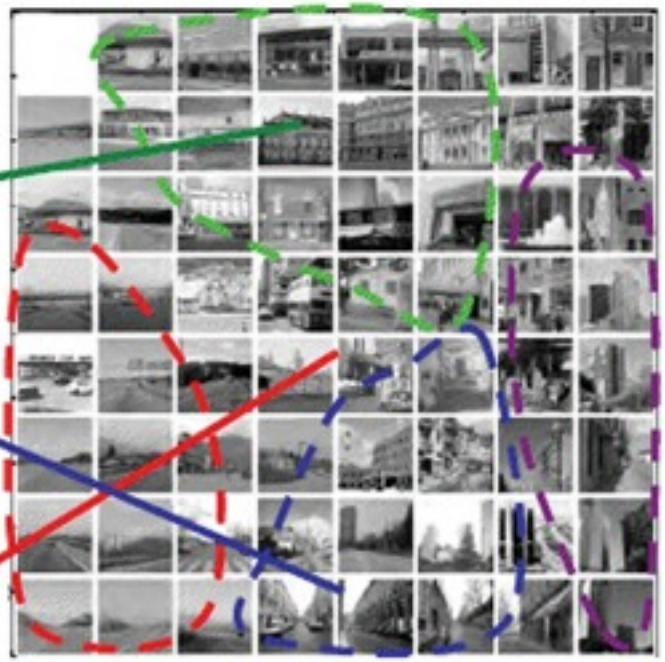
Slide Credit: Olivia

Categorization of Manmade Scenes

Confusion Matrix (in % using Layout template) :
Classification of prototypical scenes (400 / category)

	Highway	Street	City centre	tall building
Highway	91.6	4.8	2.7	0.9
Street	4.7	89.6	1.8	3.4
Centre	2.5	2.3	87.8	7.4
Tall Building	0.1	3.4	8.5	88

Local organization: correct for 86 % images
(4 similar images on 7 K-NN)



Slide Credit: Olivia

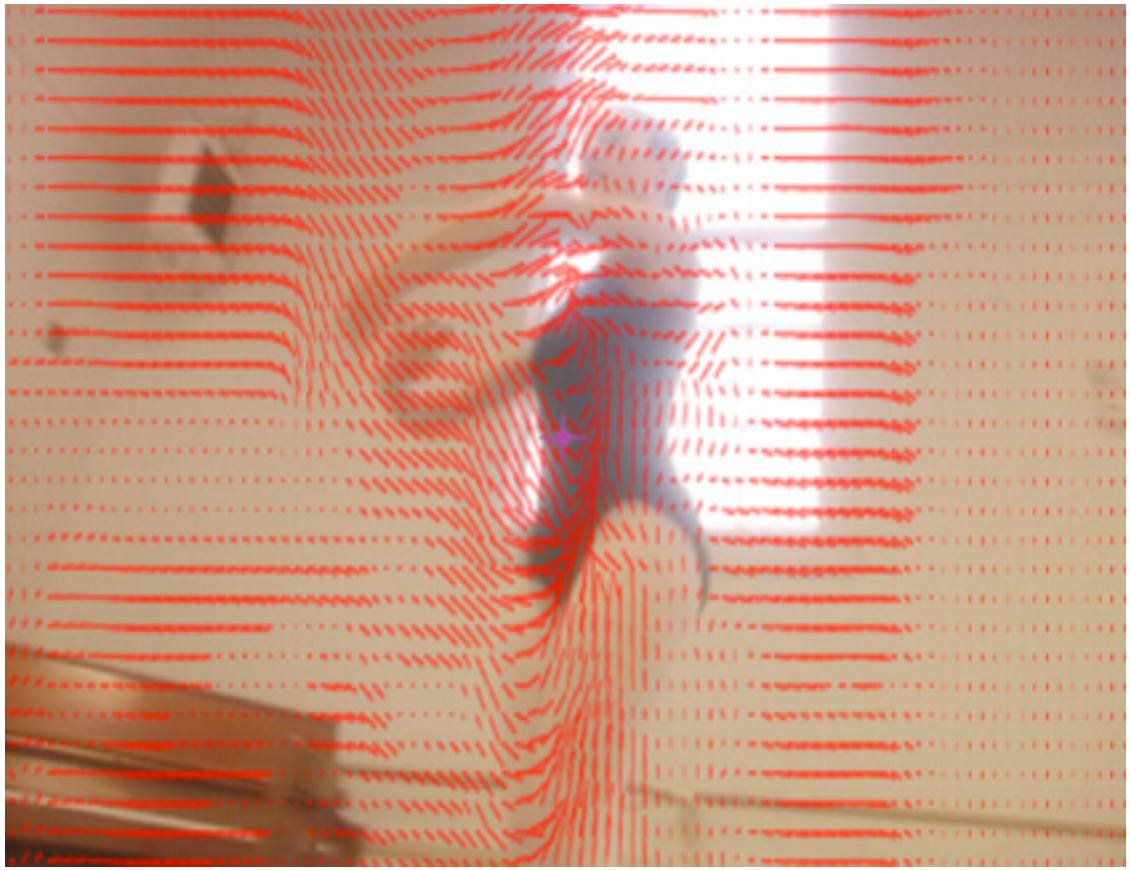
Temporal Video Features

Difference frame:

Other Features

- Histogram of Oriented Gradients (HoG)
- Speeded Up Robust Feature (SURF)
- Local Energy-based Shape Histogram (LESH)
- etc...

Temporal Video Features



Optical flow (compressed domain)

High-Level Visual Features

- Region detection: Skin color, textures
- Categories:
  indoor and outdoor, play and non-play, etc.
- Face detection:
  number of faces, location of face, etc.

High-Level Visual Features

- Objects, concepts, events, etc.
- Existence of an entity:
  e.g., trees
- Descriptive meaning
  e.g., sports

Traditional Approach

- Manual Feature Selection:
  - Choose a feature subset from the original feature set, which best represents the target semantic concepts
  - Having more features should surely result in more discriminating power, but adding irrelevant or distracting features often confuses system

Example: TrecVid MED 2010

Making a cake

Batting a run in

Assembling a shelter

Example: TrecVid MED 2010

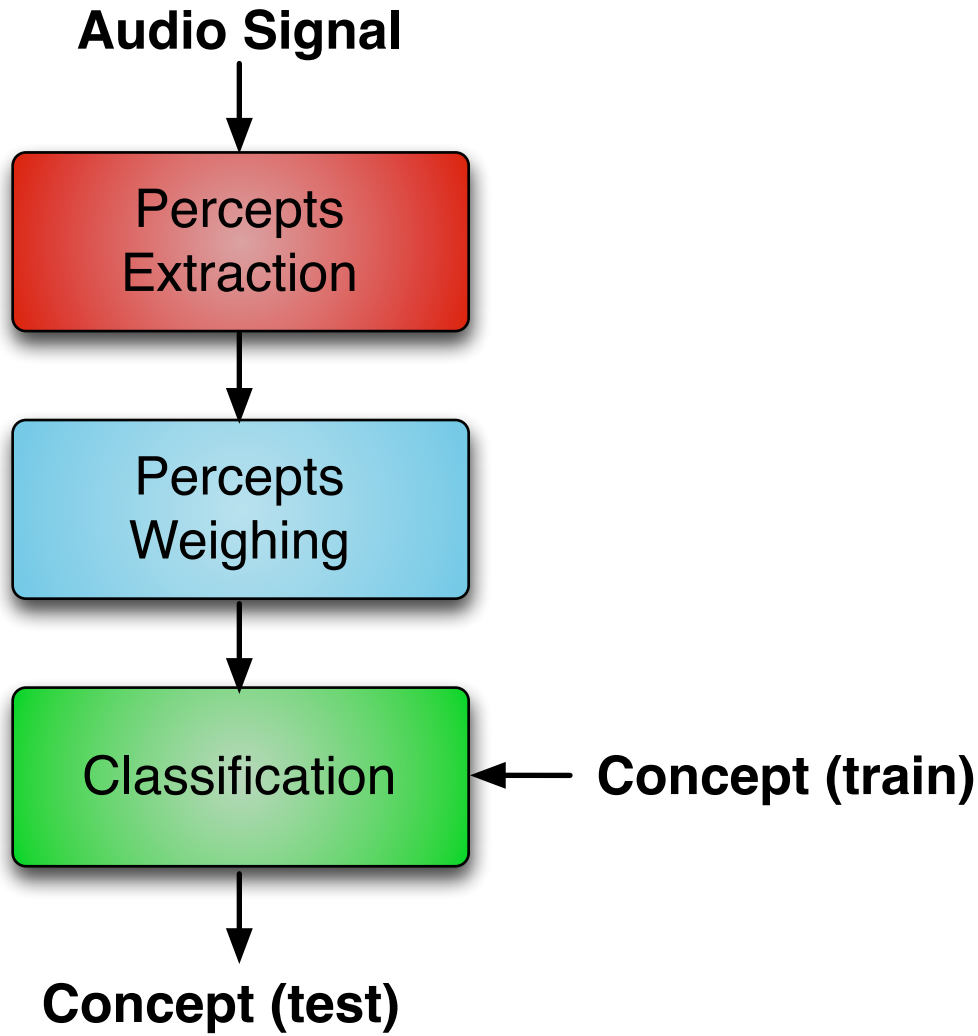
Human Action Concepts	*Scene Concepts*	*Audio Concepts*
▪ Person walking	▪ Indoor kitchen	▪ Outdoor rural
▪ Person running	▪ Outdoor with grass/trees visible	▪ Outdoor urban
▪ Person squatting	▪ Baseball field	▪ Indoor quiet
▪ Person standing up	▪ Crowd (a group of 3+ people)	▪ Indoor noisy
▪ Person making/assembling stuffs with hands (hands visible)	▪ Cakes (close-up view)	▪ Original audio
▪ Person batting baseball		▪ Dubbed audio
		▪ Speech comprehensible
		▪ Music
		▪ Cheering
		▪ Clapping

Yu-Gang Jiang, Xiaohong Zeng, Guangnan Ye, Subhabrata Bhattacharya, Dan Ellis, Mubarak Shah, Shih-Fu Chang: *Columbia-UCF TRECVID2010 Multimedia Event Detection: Combining Multiple Modalities, Contextual Concepts, and Temporal Matching*, Proceedings of TrecVid 2010, Gaithersburg, MD, December 2010.

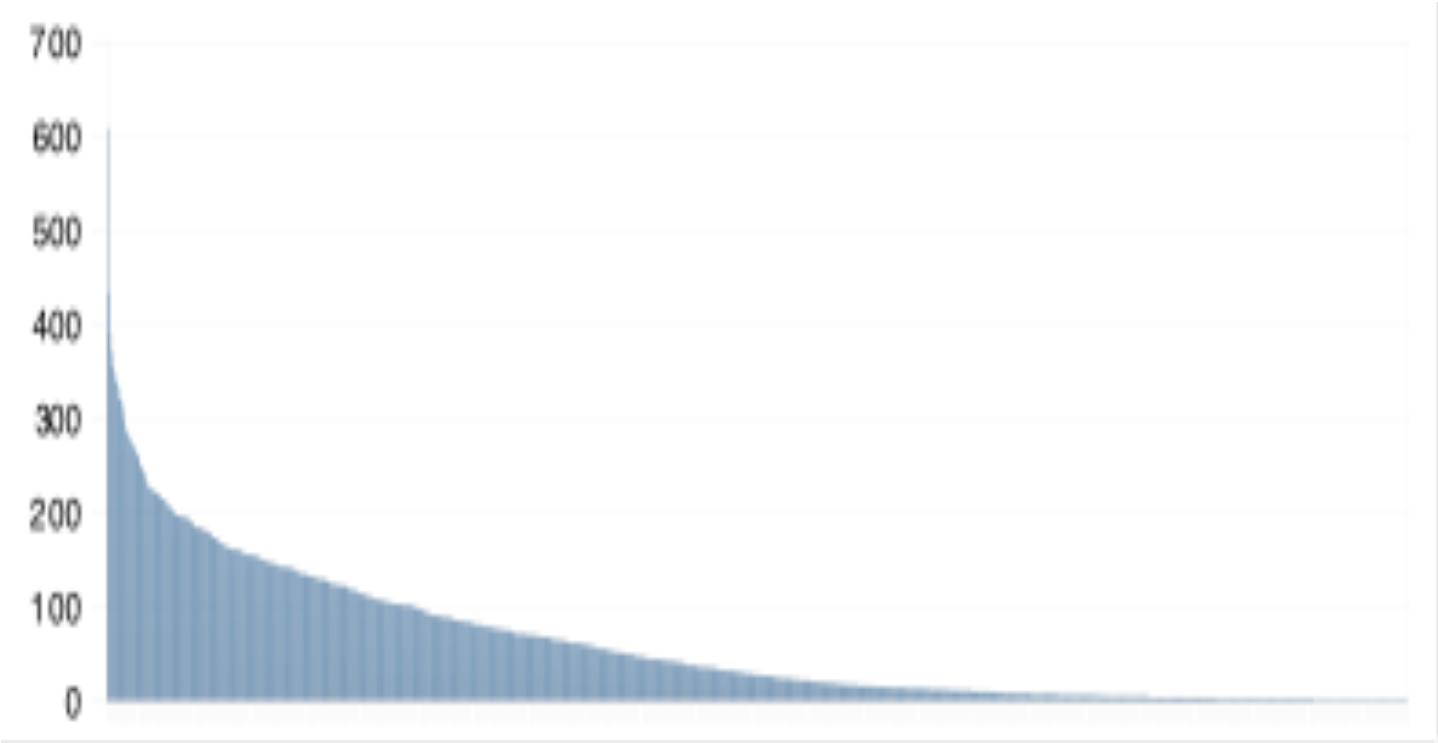
Newer Approaches

- Automatic Feature Selection:
  - Let the computer learn by example a subset from the original feature set, which best represents the target semantic concepts
  - Problem: Training set might not be representative enough

Automatic Feature Selection



Distribution of "Percepts"



Histogram of top-300 "words".

Near Zipfian Distribution!

Next Week (Project Meeting)

- More on Amazon
- Tools at ICSI

Next Week (Lecture)

- Multimedia Systems