# Lecture 22

## Object recognition III

# Class experiment

# Class experiment

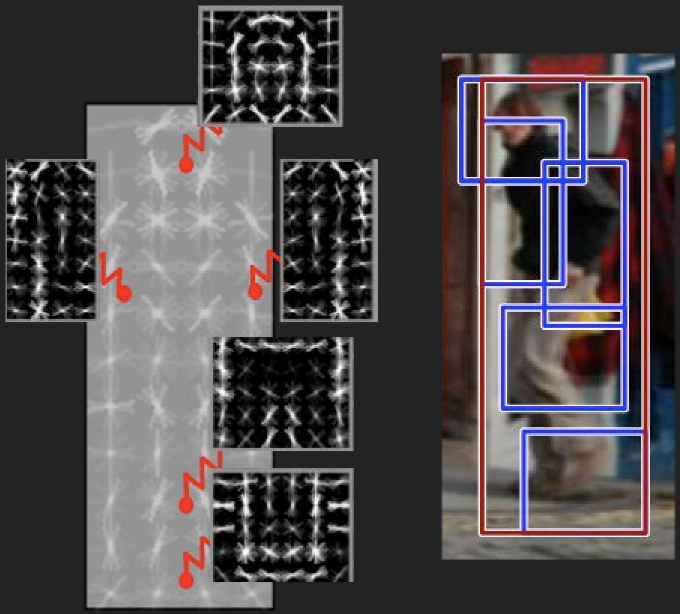
**Experiment 1:** draw a horse (the entire body, not just the head) in a white piece of paper.

Do not look at your neighbor! You already know how a horse looks like… no need to cheat.

# Class experiment

**Experiment 2:** draw a horse (the entire body, not just the head) but this time chose a viewpoint as weird as possible.

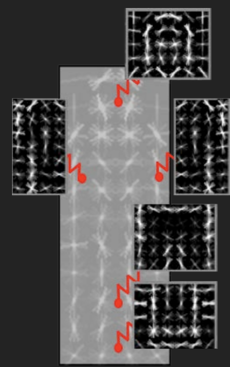
# Deformable part models

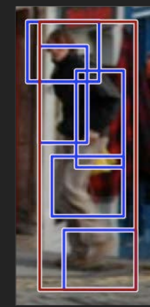


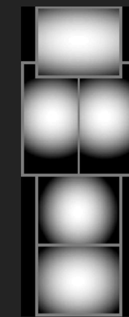
Model encodes local appearance + pairwise geometry

Source: Deva Ramanan

# Scoring function







$$\text{score}(x,z) = \sum_i w_i\, \phi(x, z_i) + \sum_{i,j} w_{ij}\, \Psi(z_i, z_j)$$

$x$ = image
$z_i = (x_i, y_i)$
$z = \{z_1, z_2 \ldots\}$

part template scores

spring deformation model

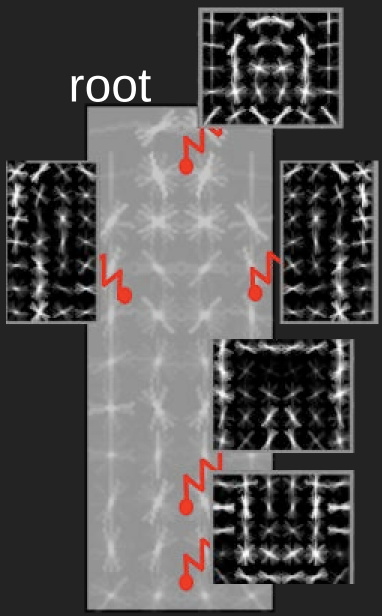
Score is linear in local templates $w_i$ and spring parameters $w_{ij}$

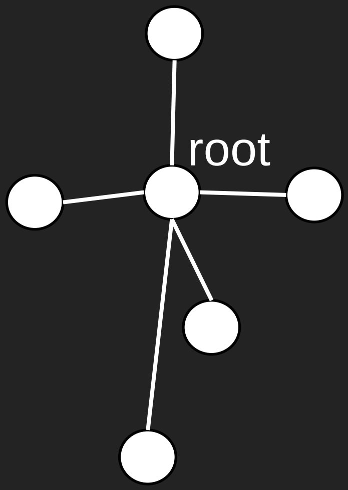
$$\text{score}(x,z) = w \cdot \Phi(x, z)$$

Source: Deva Ramanan

# Inference: $\max_z$ score(x,z)

Felzenszwalb & Huttenlocher 05




Star model: the location of the root filter is the anchor point
Given the root location, all part locations are independent

# Classification

$f_w(x)=w \cdot \Phi(x)$

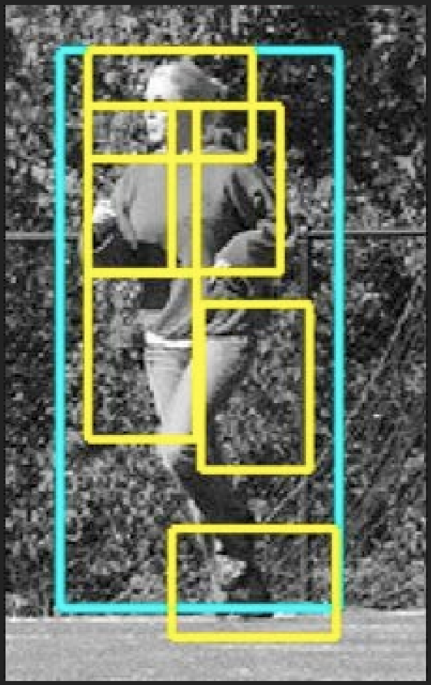
$f_w(x)>0$

Source: Deva Ramanan

# Latent-variable classification

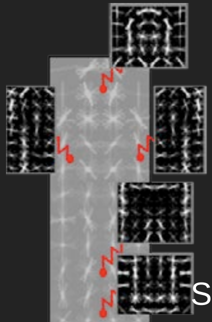
$f_w(x)=w \cdot \Phi(x)$

$f_w(x)>0$



$$f_w(x)=\max_z S(x,z)$$

$$=\max_z w \cdot \Phi(x, z)$$

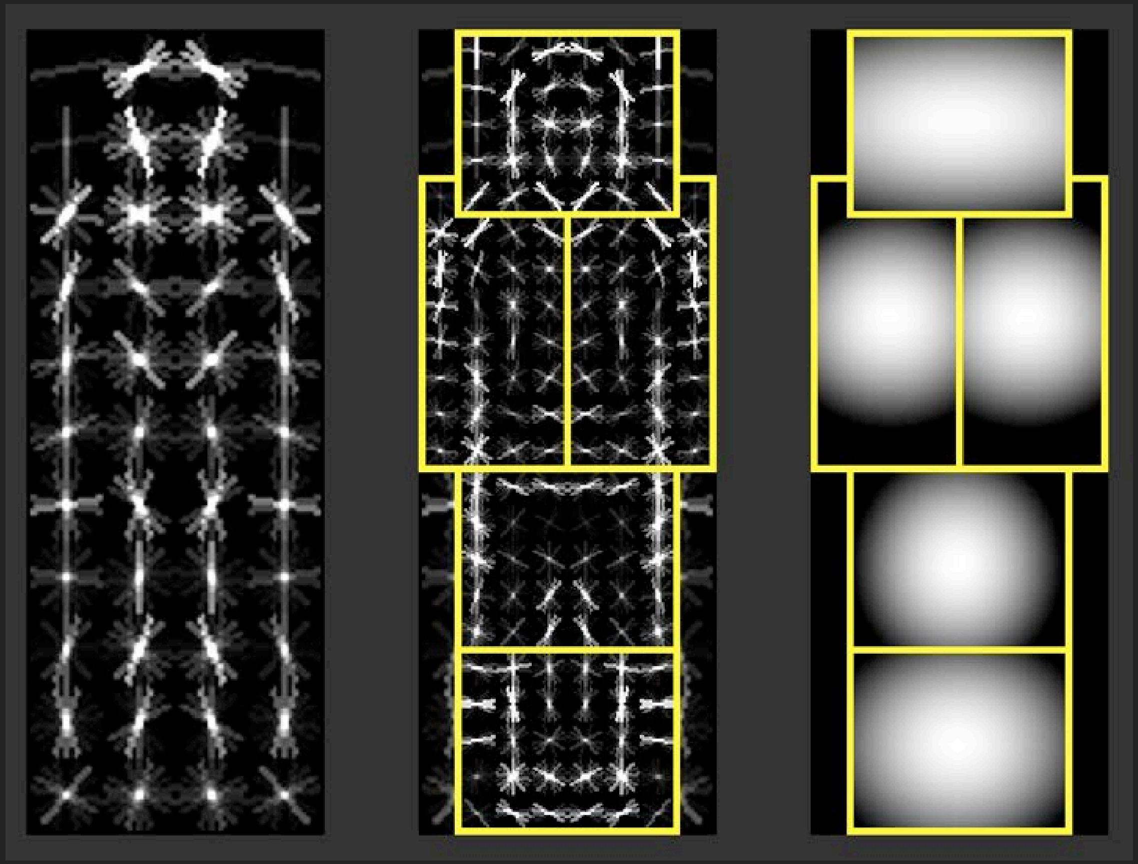


Source: Deva Ramanan

# Initialization

Learn root filter with SVM

Initialize part filters to regions in root filter with lots of energy



Source: Deva Ramanan

# Coordinate descent

1) Given positive part locations, learn w with a convex program

$$w = \operatorname*{argmin}_{w} L(w) \quad \text{with fixed} \quad \{z_n : n \in \text{pos}\}$$

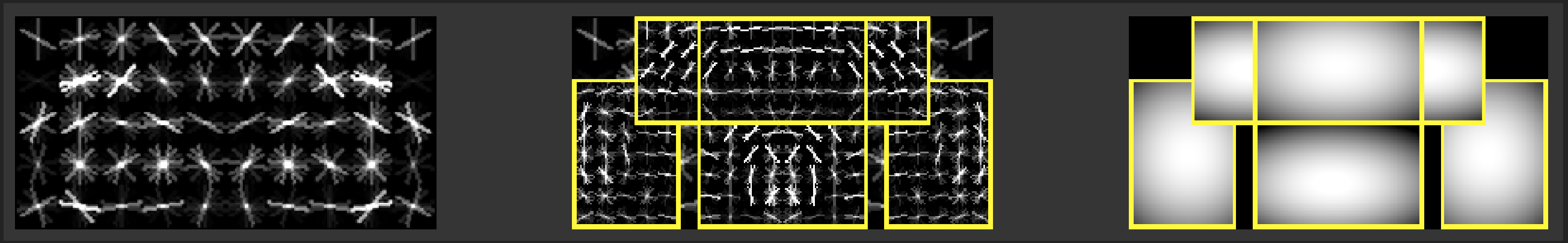
2) Given w, estimate part locations on positives

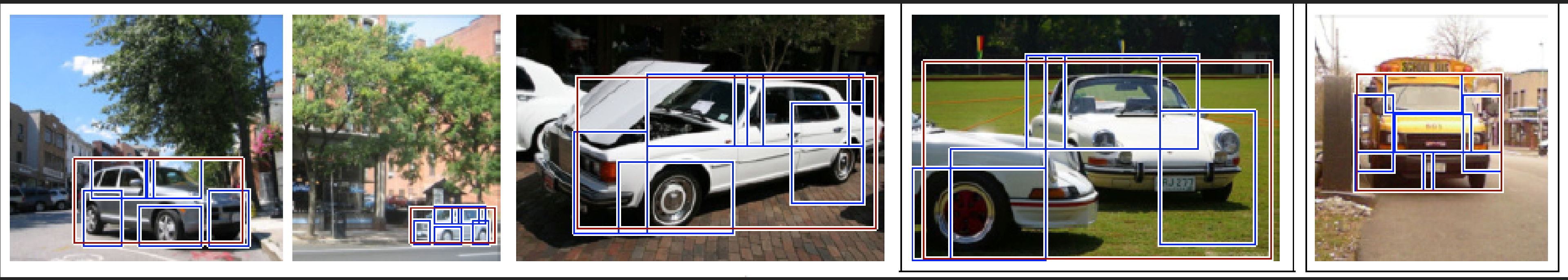
$$z_n = \operatorname*{argmax}_{z} w \cdot \Phi(x_n, z) \quad \forall n \in \text{pos}$$

The above steps perform coordinate descent on a joint loss

Source: Deva Ramanan

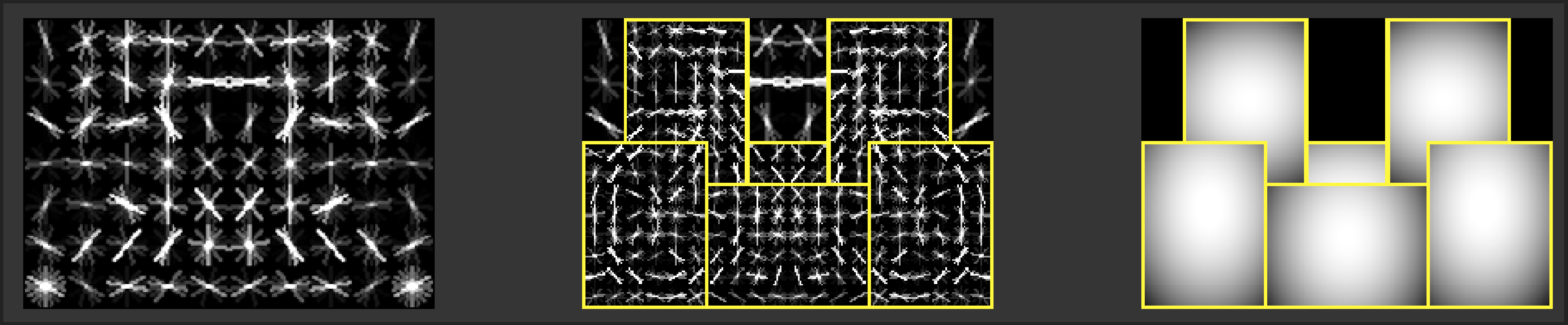
# Example models

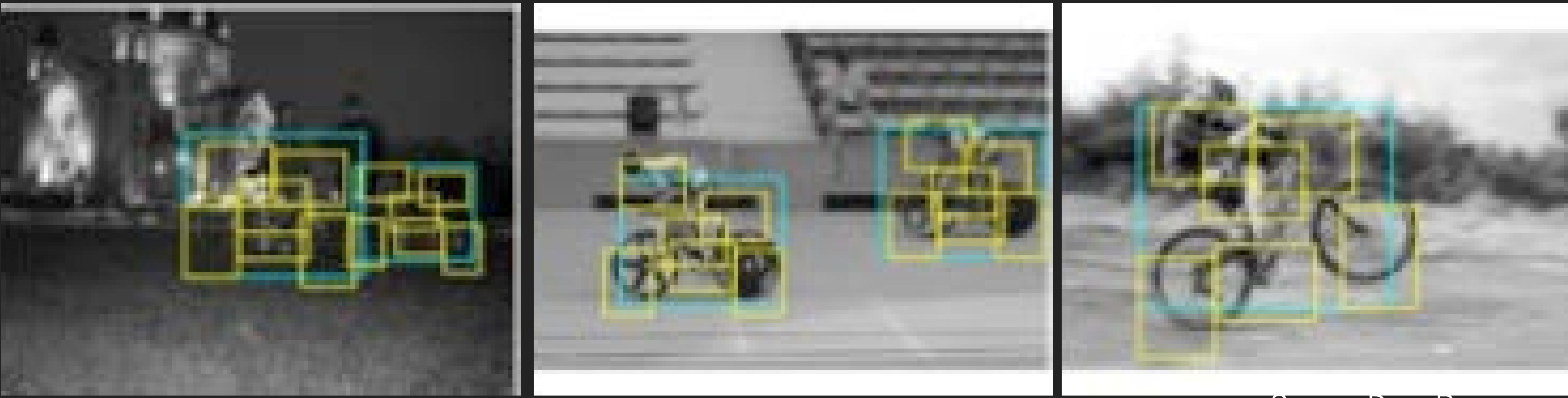




Source: Deva Ramanan

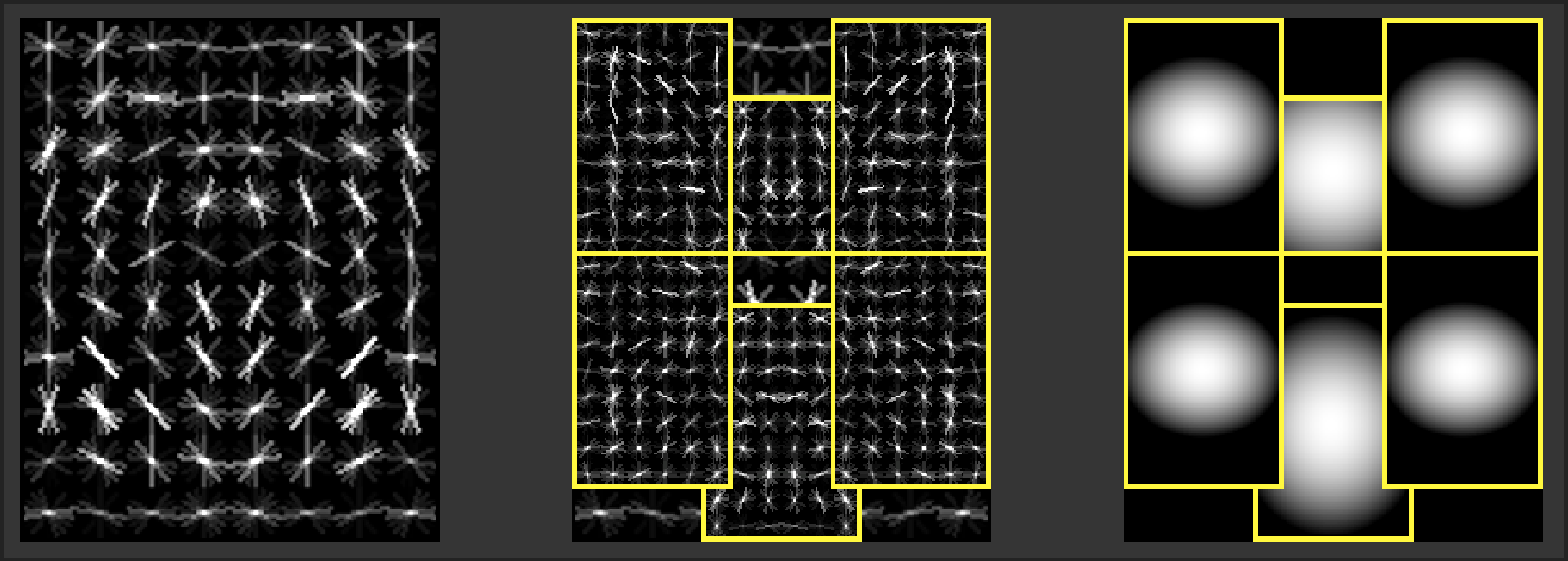
# Example models



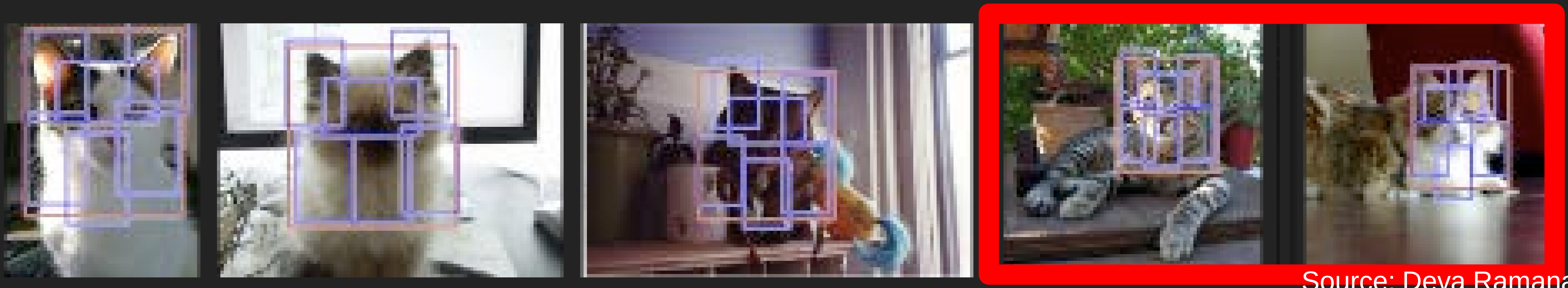


Source: Deva Ramanan

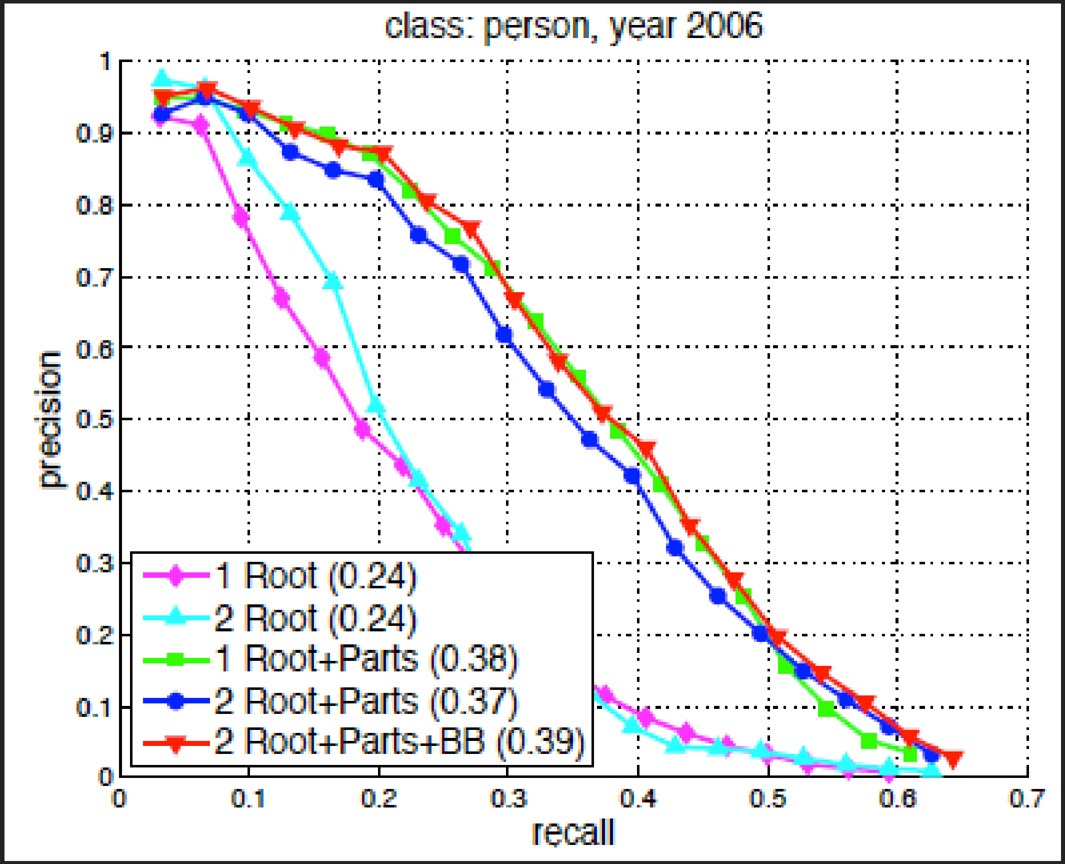
# Example models



False positive due to imprecise bounding box



Source: Deva Ramanan

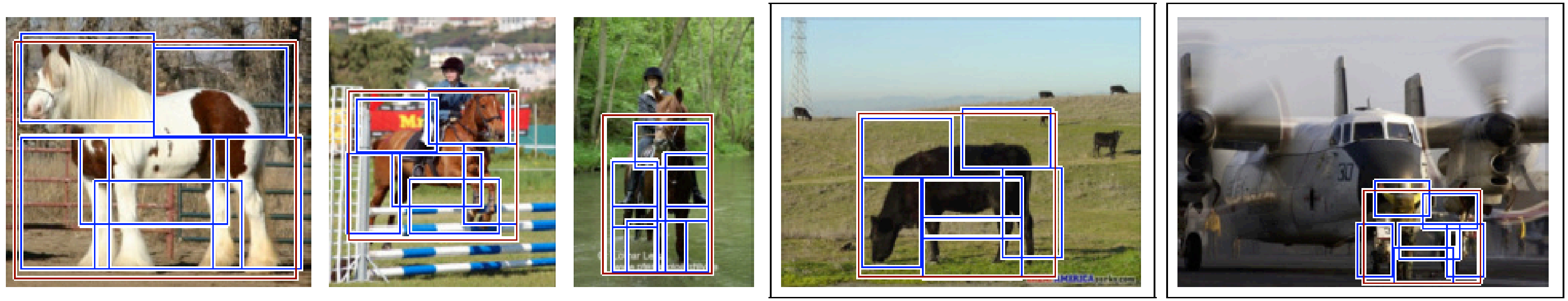



Other tricks:

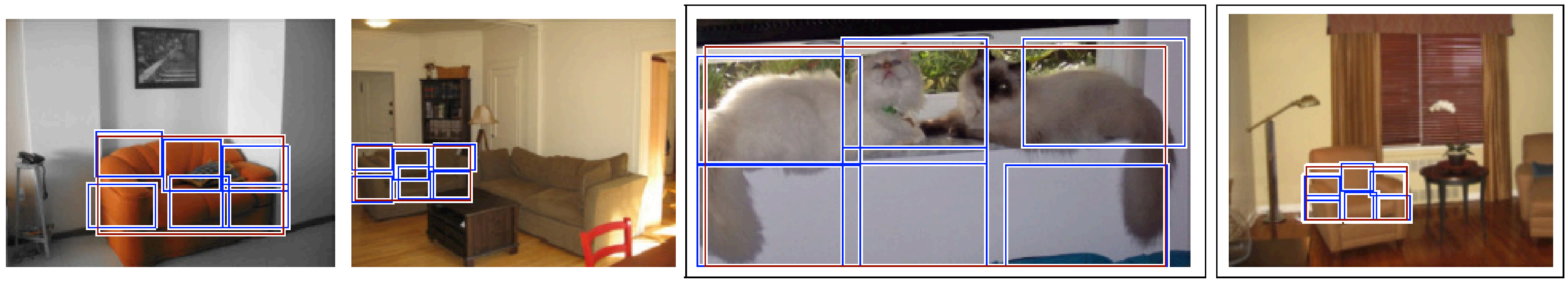
•Mining hard negative examples

•Noisy annotations

horse

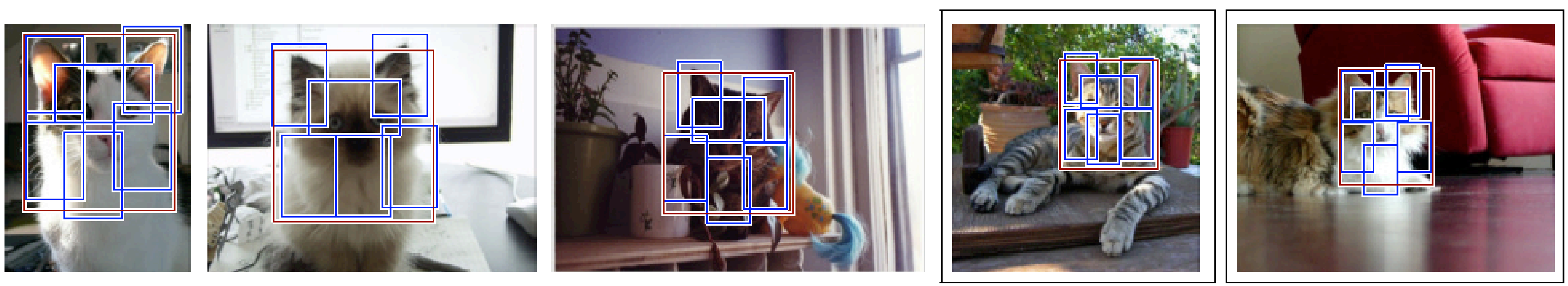


sofa



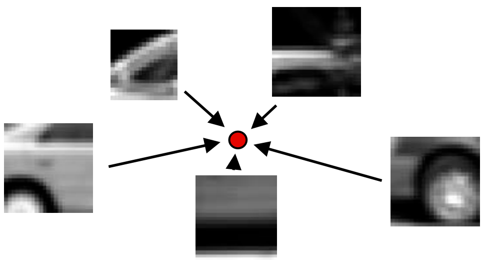
bottle


cat



# Structure models

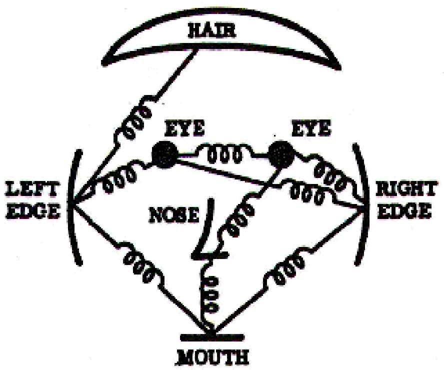
**Voting models**



- Many parts (>100)

Lecture 20

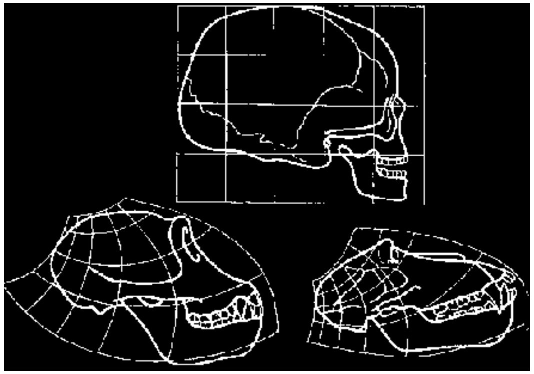
**Constellation models**




- Few parts (~6)

Lecture 21

**Deformable models**

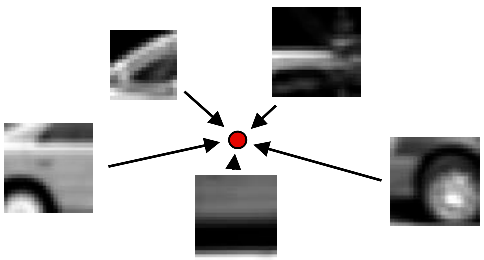


- No parts

Lecture 22

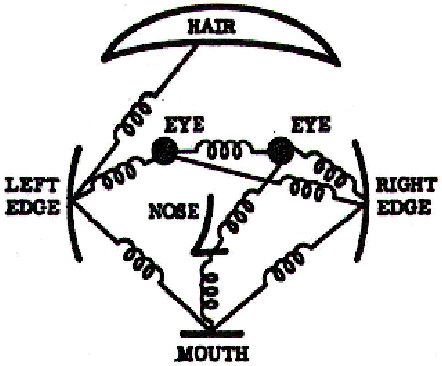
# Structure models

**Voting models**



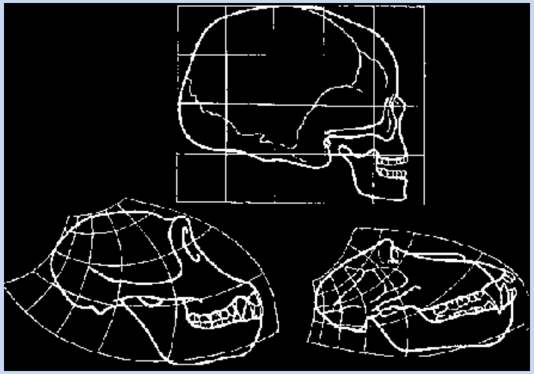
- Many parts (>100)

**Constellation models**




- Few parts (~6)

**Deformable models**

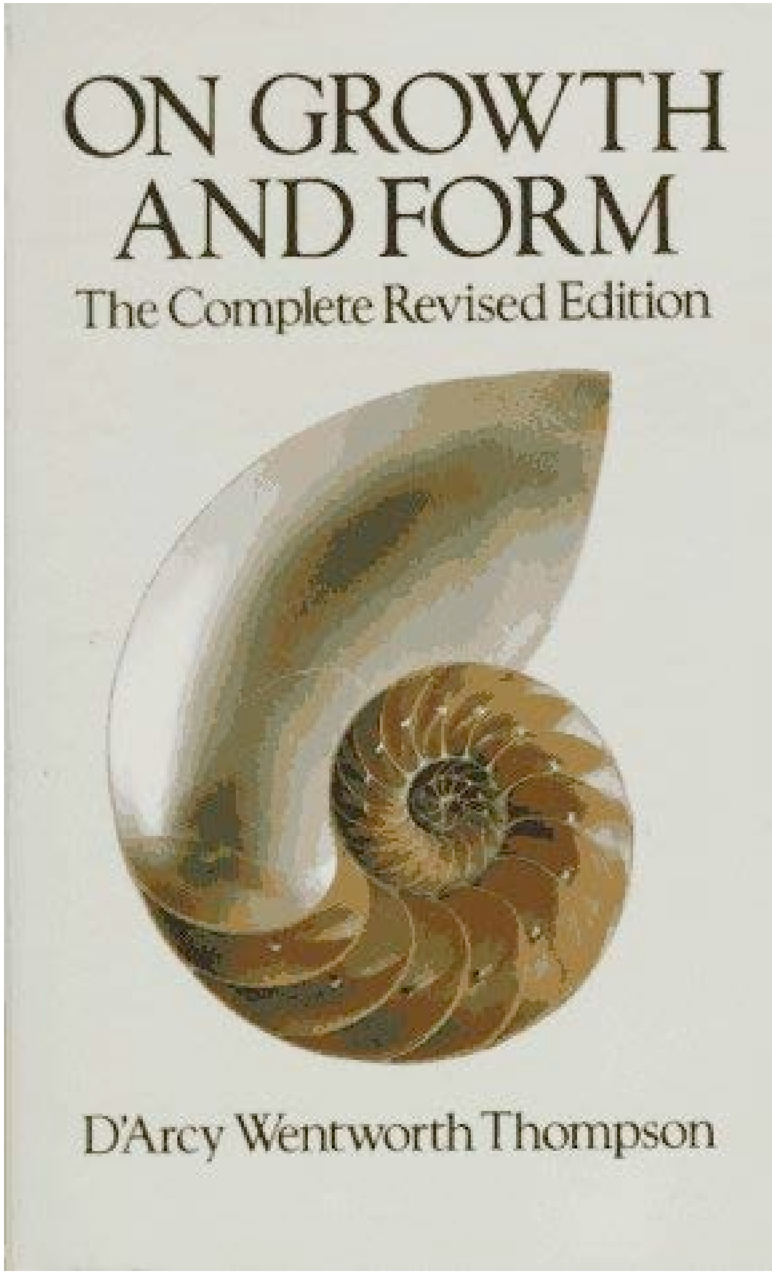


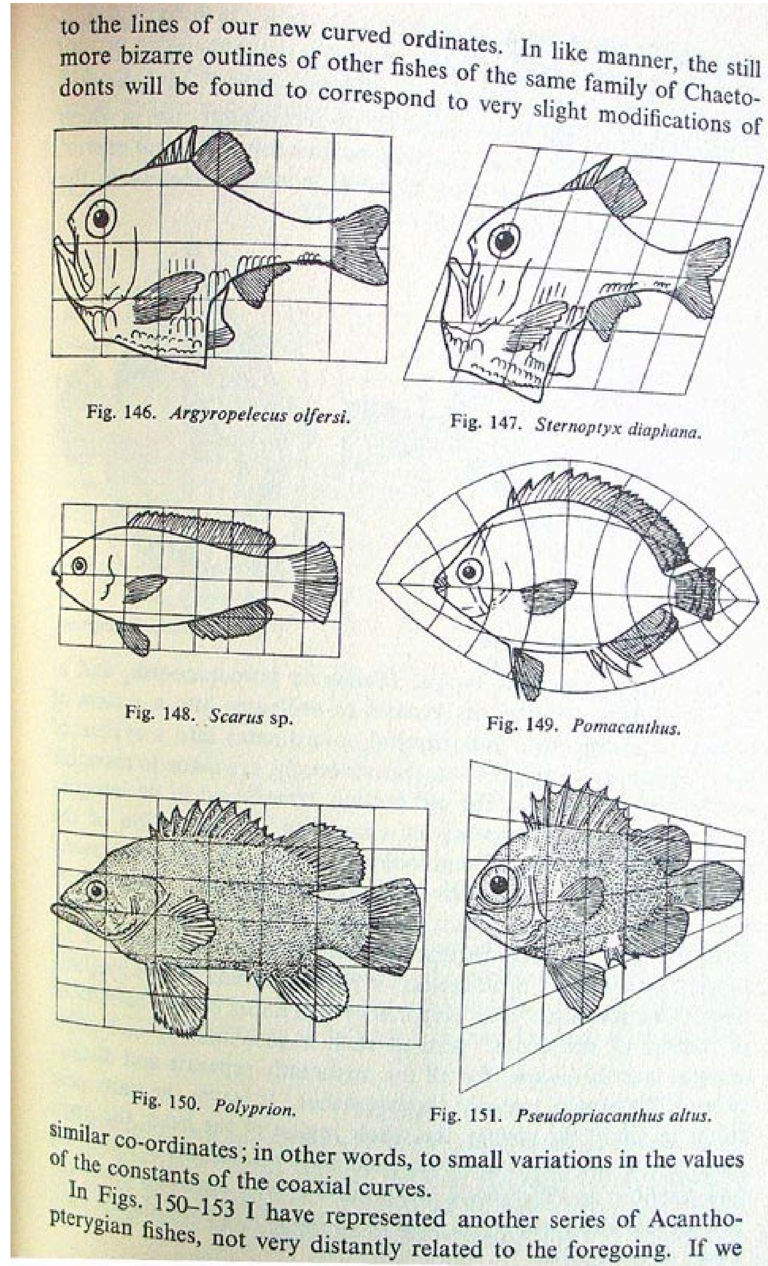
- No parts

Lecture 20

Lecture 21

Lecture 22



to the lines of our new curved ordinates. In like manner, the still more bizarre outlines of other fishes of the same family of Chaetodonts will be found to correspond to very slight modifications of

Fig. 146. *Argyropelecus olfersi.*

Fig. 147. *Sternoptyx diaphana.*

Fig. 148. *Scarus* sp.

Fig. 149. *Pomacanthus.*

Fig. 150. *Polyprion.*

Fig. 151. *Pseudopriacanthus altus.*

similar co-ordinates; in other words, to small variations in the values of the constants of the coaxial curves.

In Figs. 150–153 I have represented another series of Acanthopterygian fishes, not very distantly related to the foregoing. If we

From wikipedia: Perhaps the most famous part of the work is chapter XVII, "The Comparison of Related Forms," where Thompson explored the degree to which differences in the forms of related animals could be described by means of relatively simple mathematical transformations.

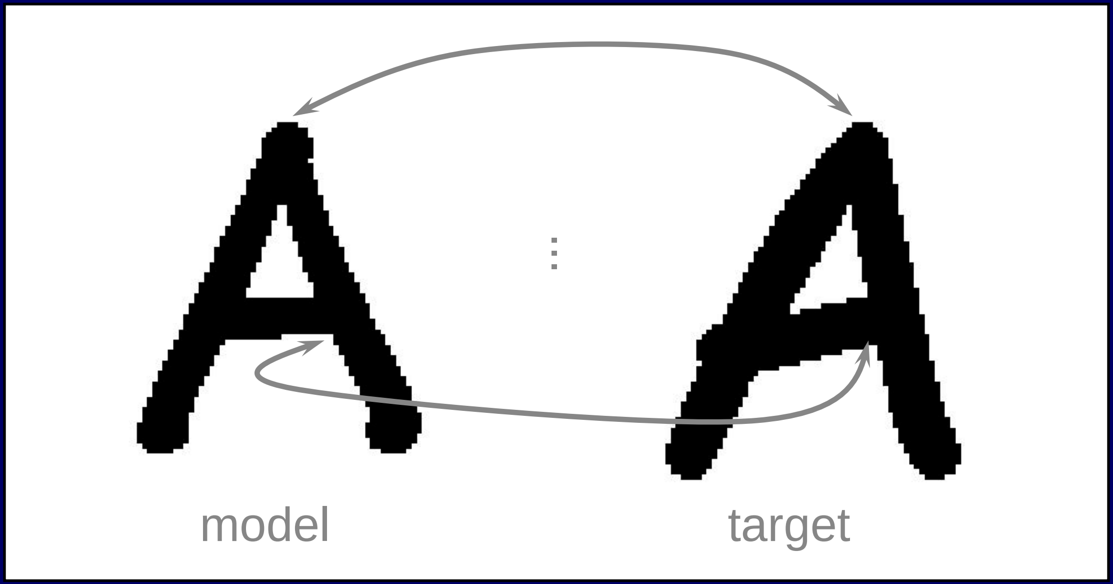
# Shape Matching and Object Recognition Using Shape Contexts

Serge Belongie, *Member, IEEE*, Jitendra Malik, *Member, IEEE*, and Jan Puzicha

**Abstract**—We present a novel approach to measuring similarity between shapes and exploit it for object recognition. In our framework, the measurement of similarity is preceded by 1) solving for correspondences between points on the two shapes, 2) using the correspondences to estimate an aligning transform. In order to solve the correspondence problem, we attach a descriptor, the *shape context*, to each point. The shape context at a reference point captures the distribution of the remaining points relative to it, thus offering a globally discriminative characterization. Corresponding points on two similar shapes will have similar shape contexts, enabling us to solve for correspondences as an optimal assignment problem. Given the point correspondences, we estimate the transformation that best aligns the two shapes; regularized thin-plate splines provide a flexible class of transformation maps for this purpose. The dissimilarity between the two shapes is computed as a sum of matching errors between corresponding points, together with a term measuring the magnitude of the aligning transform. We treat recognition in a nearest-neighbor classification framework as the problem of finding the stored prototype shape that is maximally similar to that in the image. Results are presented for silhouettes, trademarks, handwritten digits, and the COIL data set.

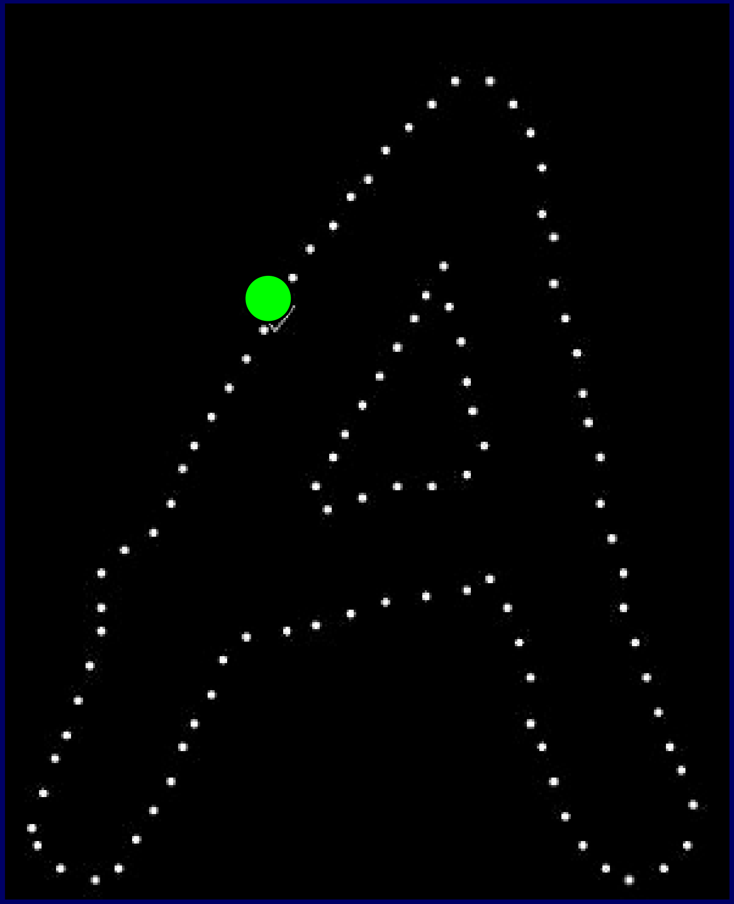
**Index Terms**—Shape, object recognition, digit recognition, correspondence problem, MPEG7, image registration, deformable templates.

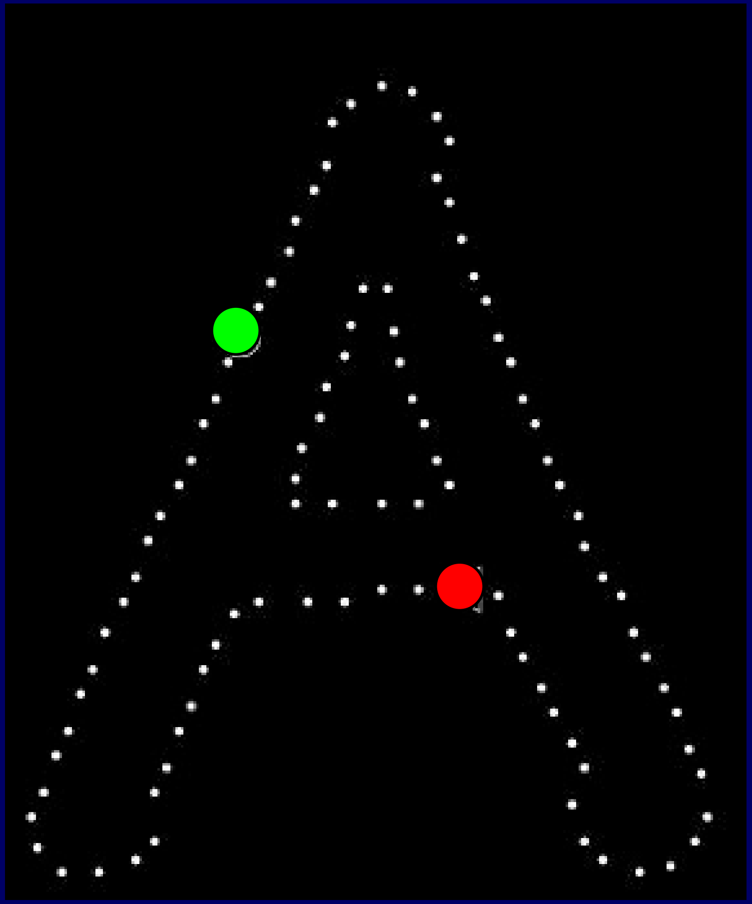
# Matching Framework



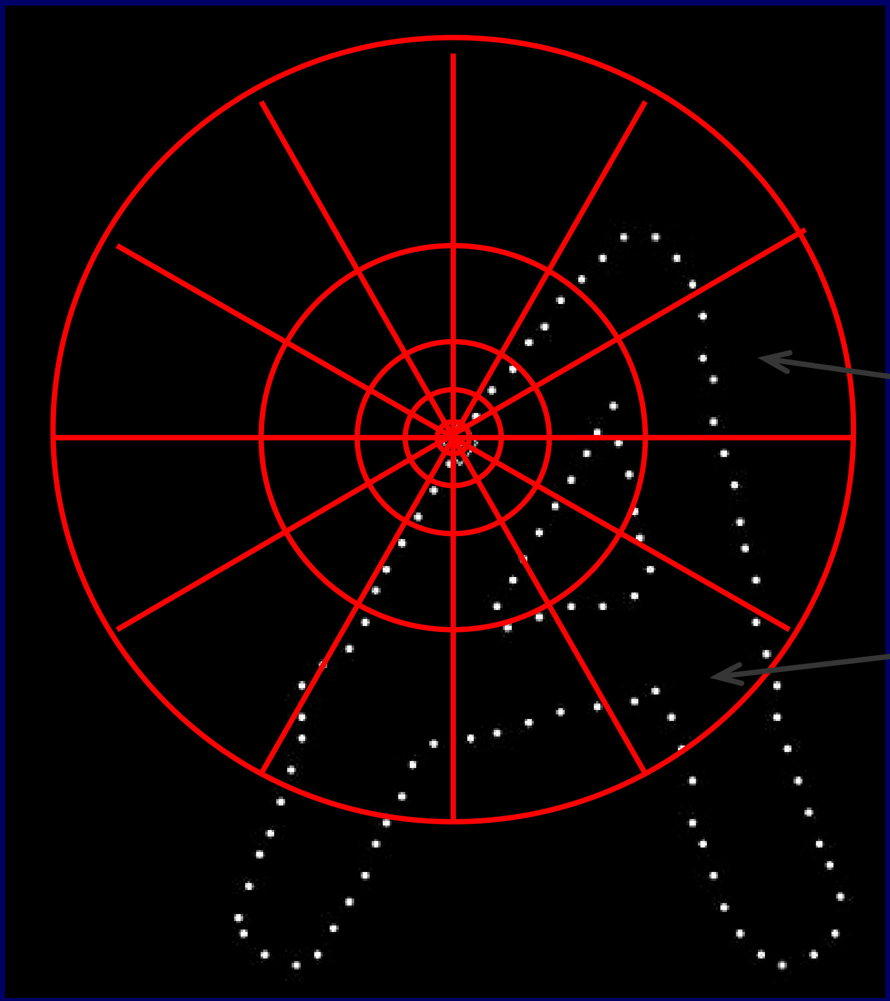

- Find correspondences between points on shape
- Fast pruning
- Estimate transformation & measure similarity

# Comparing Pointsets





# Shape Context



Count the number of points inside each bin, e.g.:

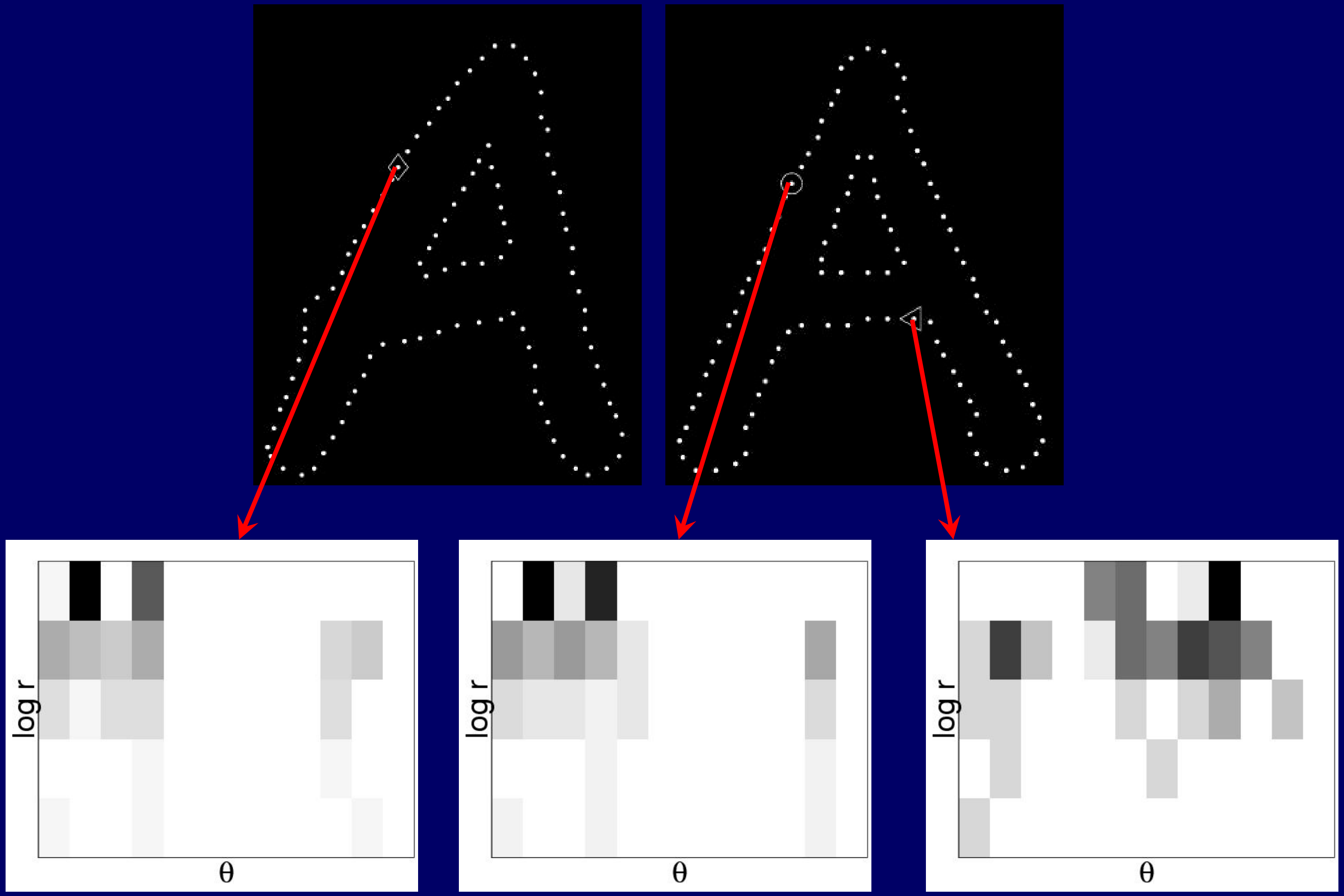
Count = 4

⋮

Count = 10

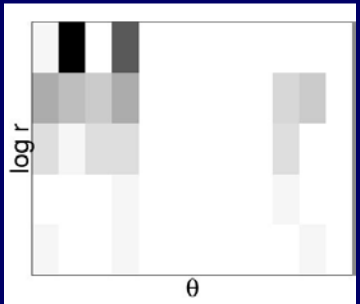
F Compact representation of distribution of points relative to each point

# Shape Context

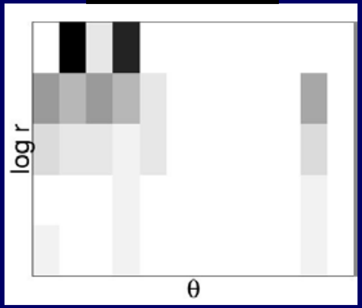


# Comparing Shape Contexts

$h_i(k)$

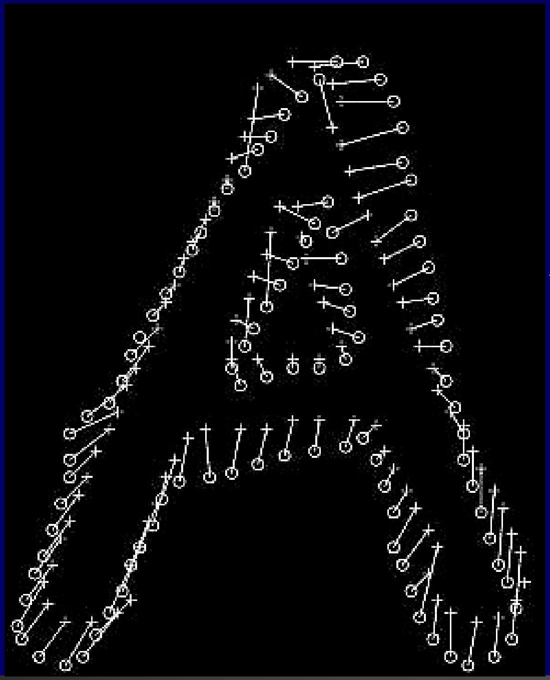



$h_j(k)$




Compute matching costs using Chi Squared distance:

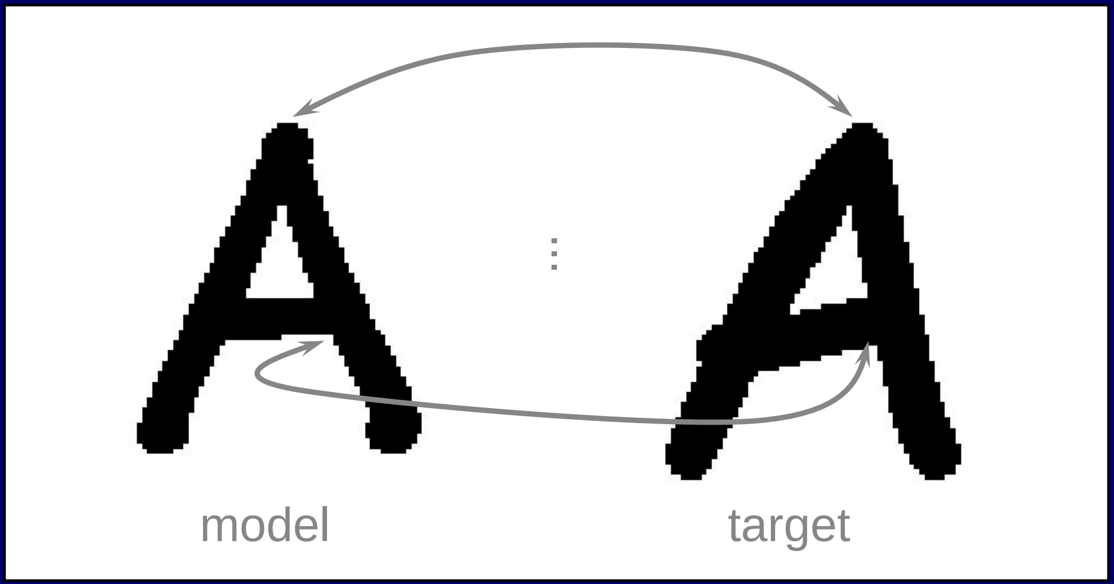
$$C_{ij} = \frac{1}{2} \sum_{k=1}^{K} \frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)}$$



Recover correspondences by solving linear assignment problem with costs $C_{ij}$

[Jonker & Volgenant 1987]

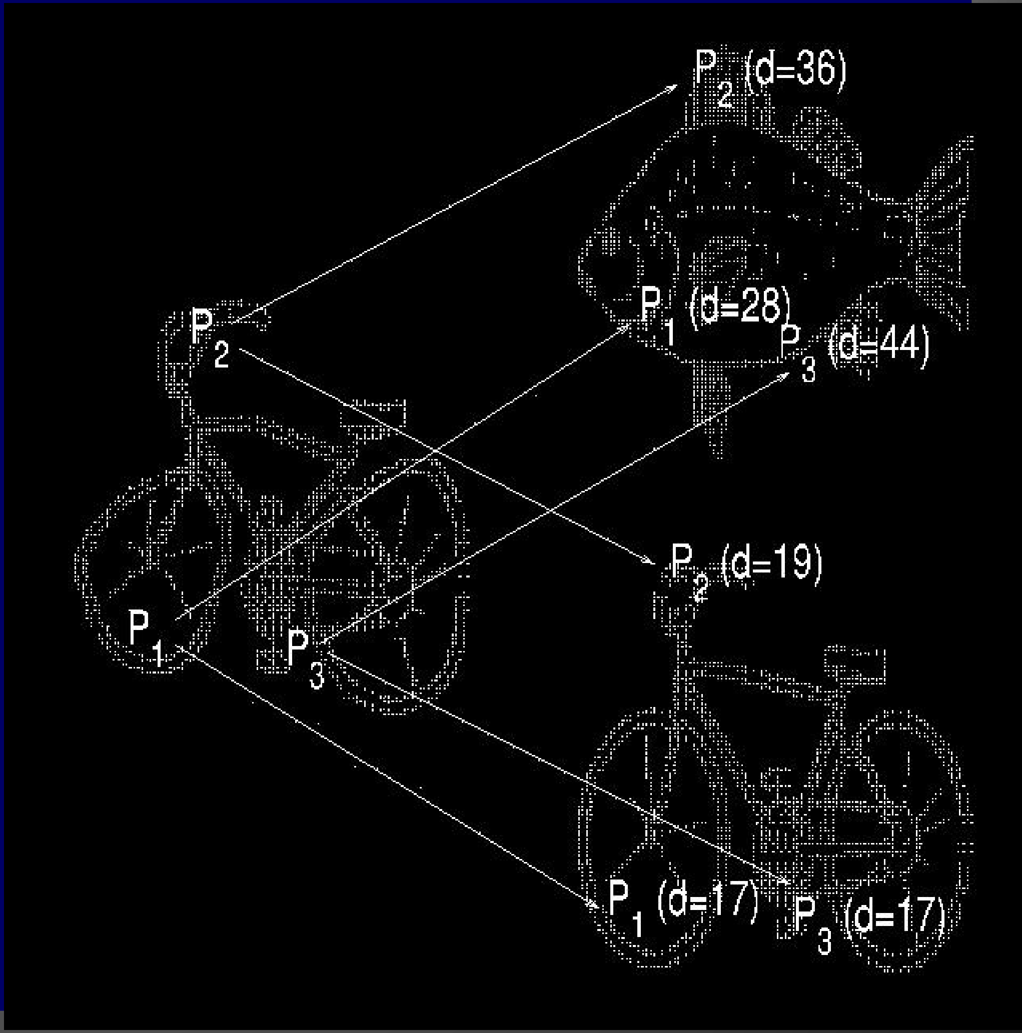
# Matching Framework



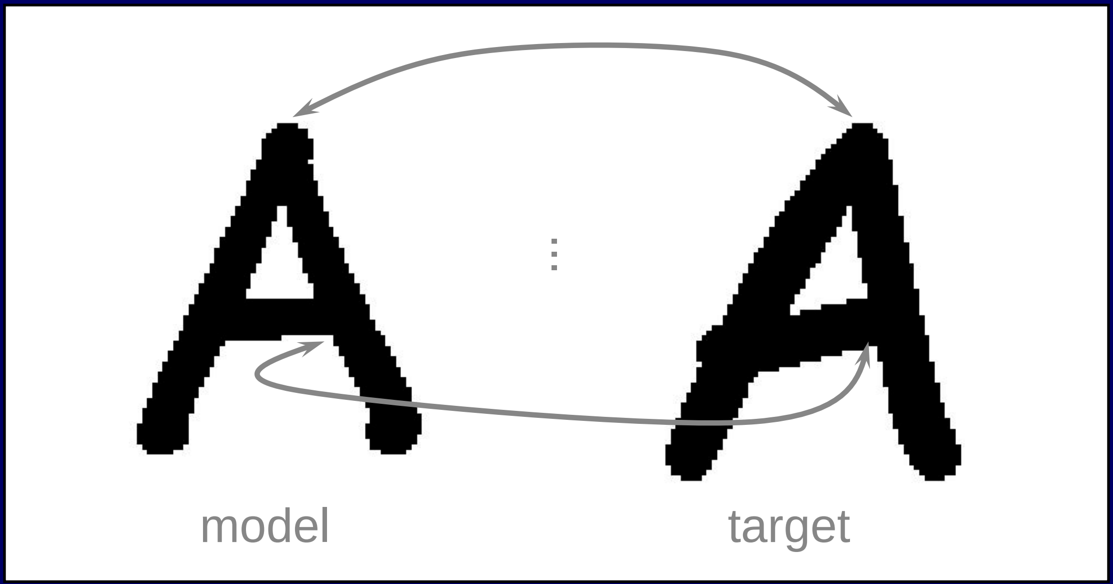

- Find correspondences between points on shape
- Fast pruning
- Estimate transformation & measure similarity

# Fast pruning

- Find best match for the shape context at  only a few random points and add up cost

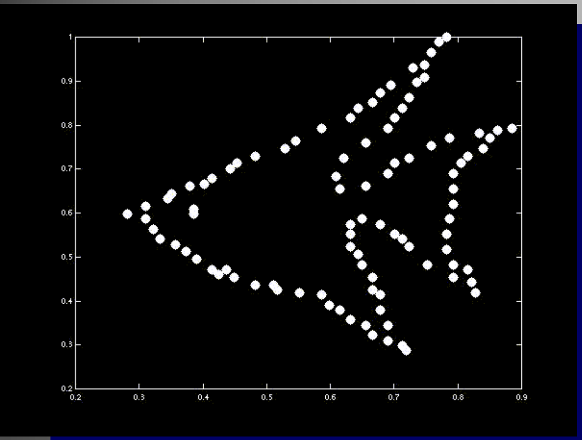


# Matching Framework

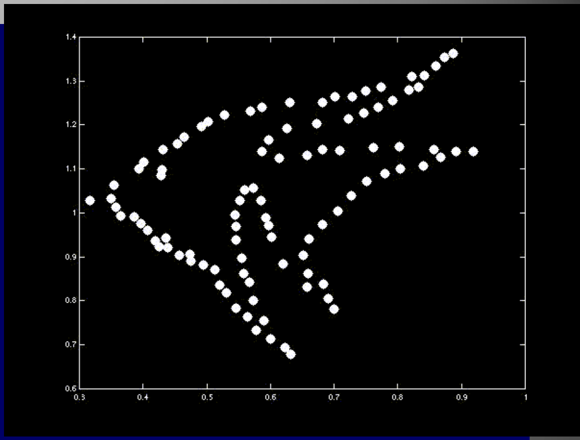



- Find correspondences between points on shape
- Fast pruning
- Estimate transformation & measure similarity

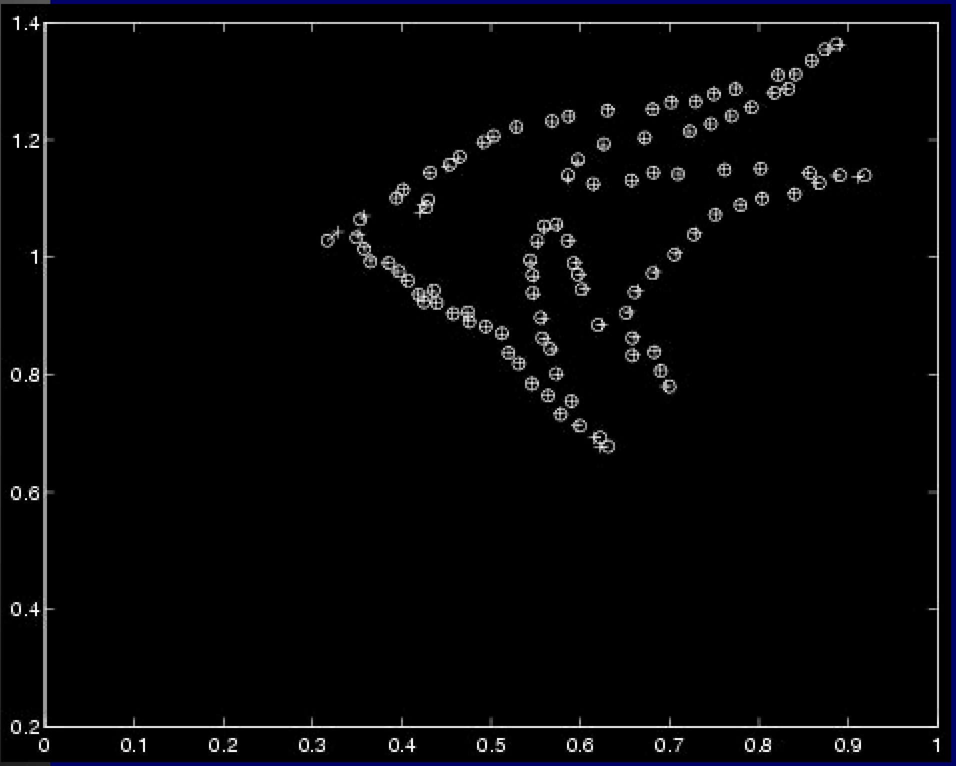
# Matching Example

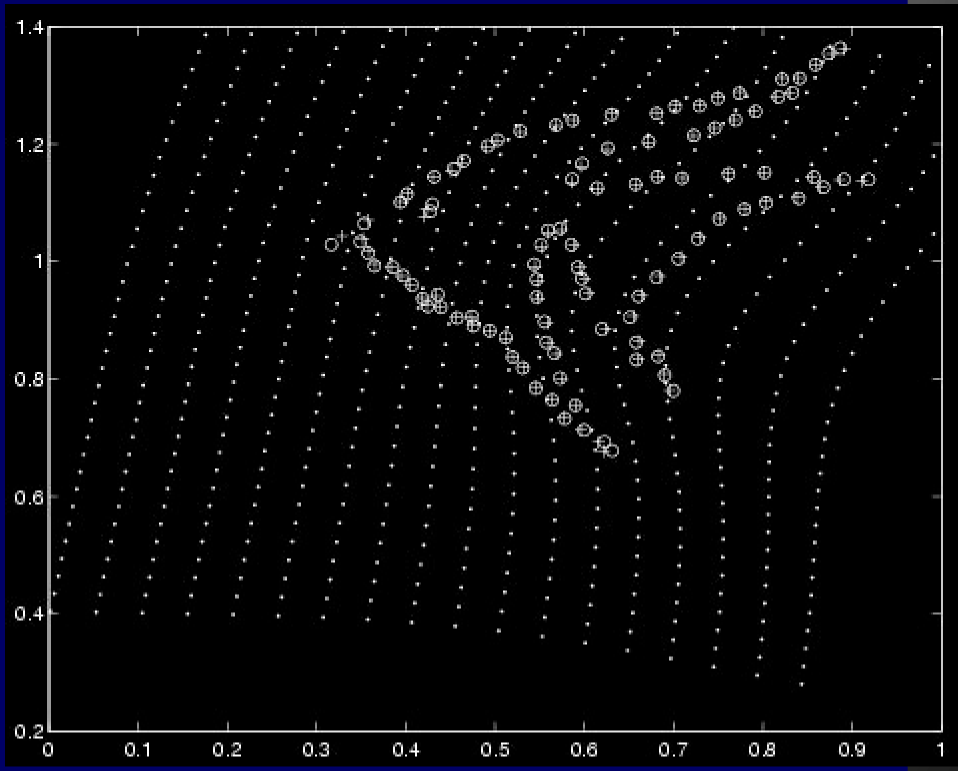


model

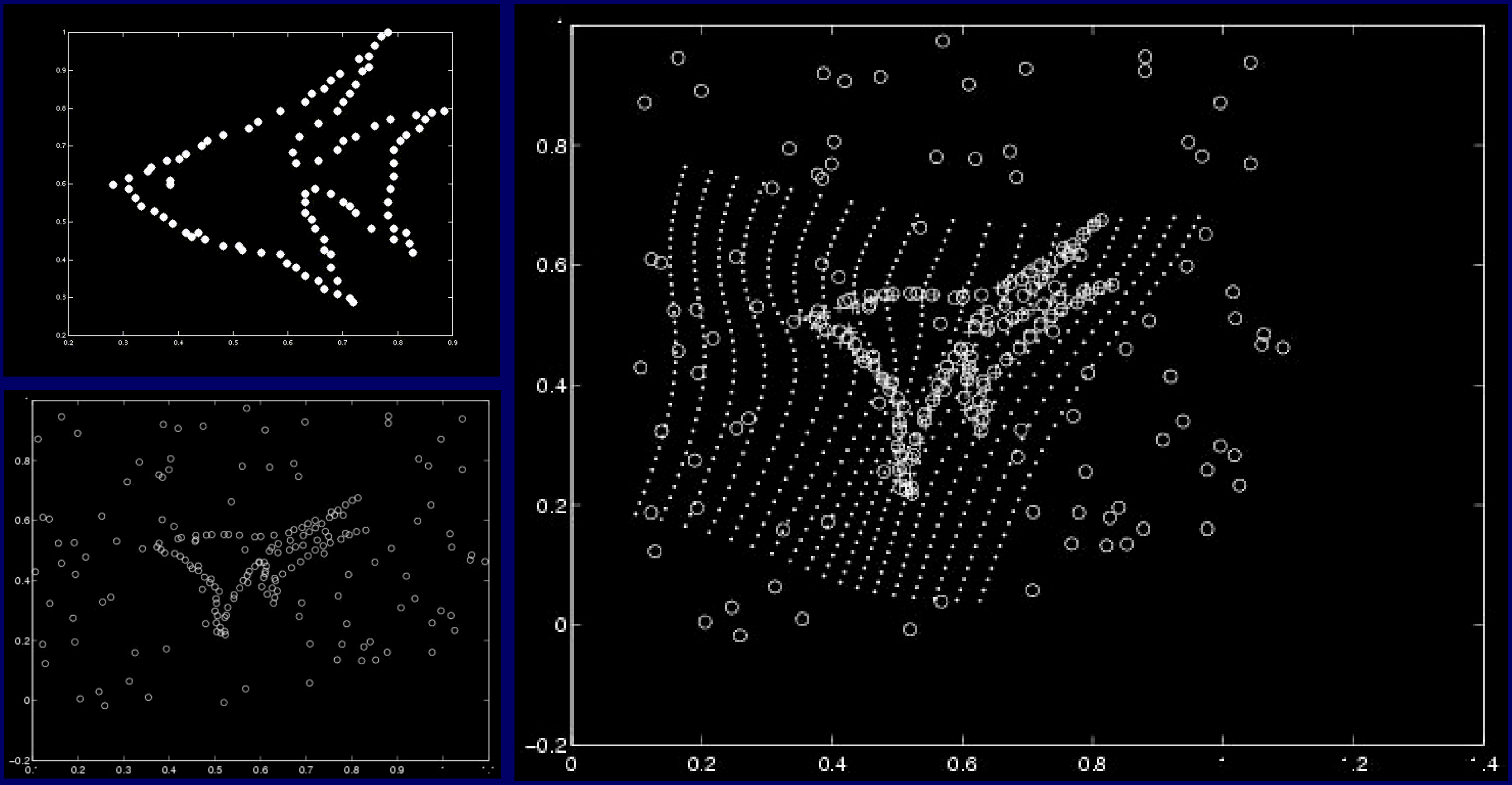


target



# Outlier Test Example



# The spaces of faces is not convex

The average of two faces is not another face

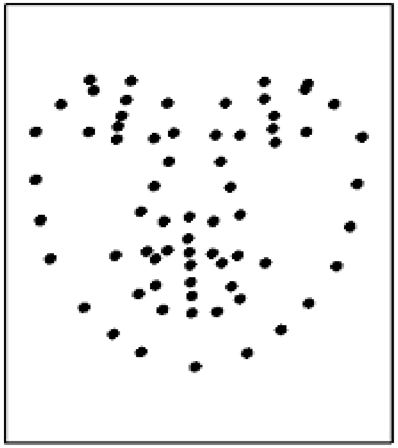
# The spaces of faces is not convex

The average of two faces is not another face

# A shape-texture face model

Model

"Shape"



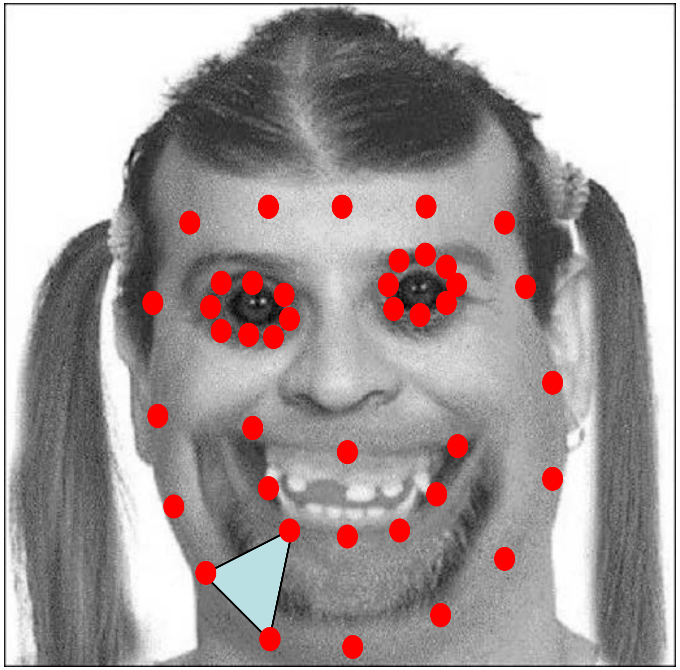
"texture"

Cootes, Edwards, and Taylor, [“Active Appearance Models”](), ECCV 1998

Slide: Dhruv Batra

# Image warping

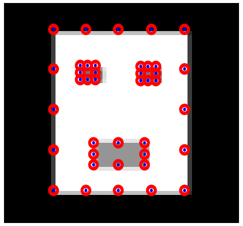
# Image warping

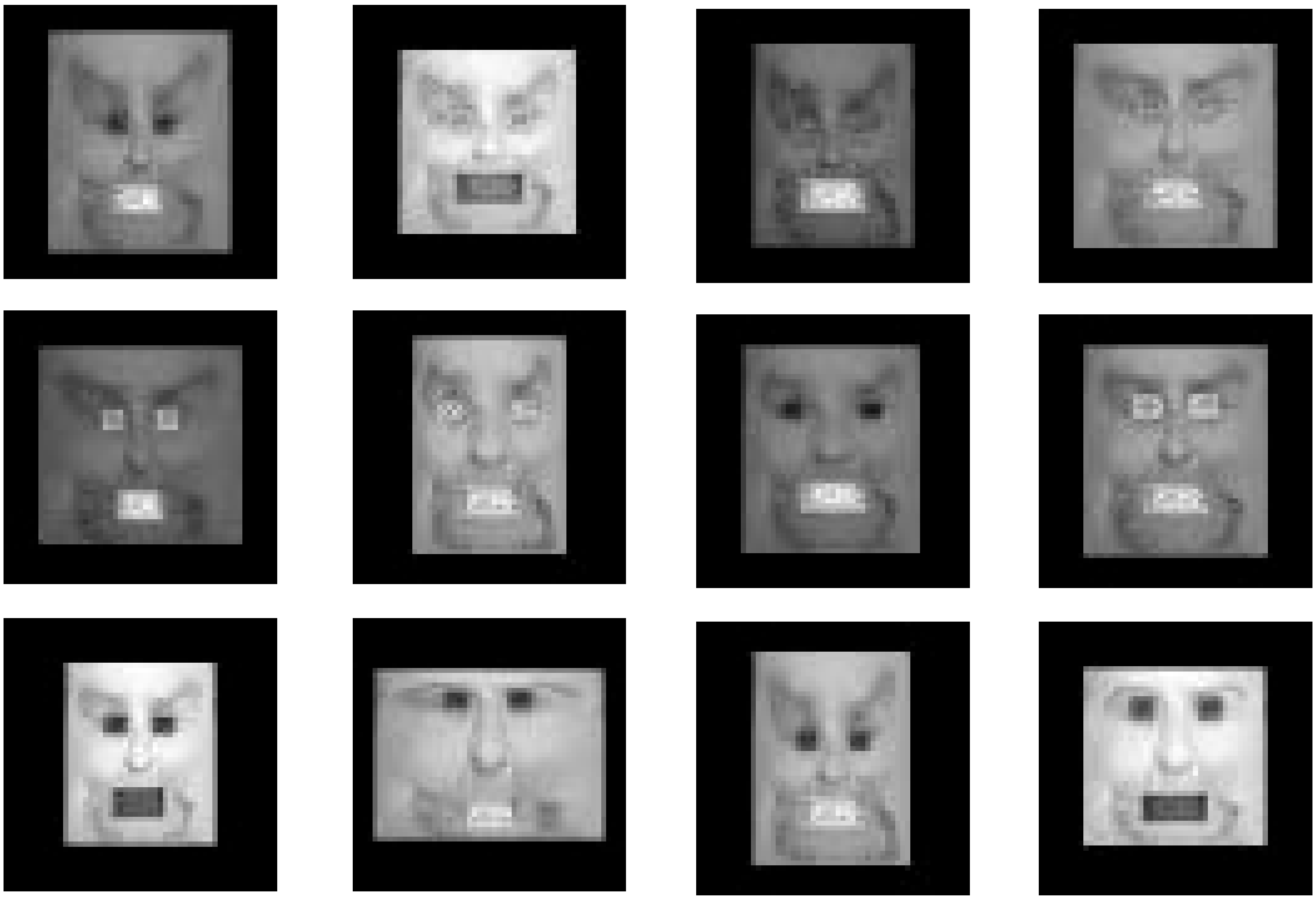


Original image

Background

# Face database



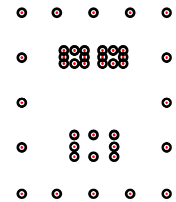


# Appearance Model (AppModel.m)

• Each image is represented as (1) a collection of correspondence points (shape) and (2) a texture image normalized in shape.

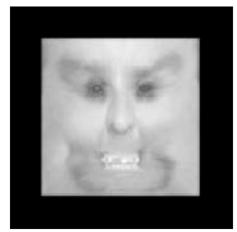
Original image

$$\begin{pmatrix} x1 \\ x2 \\ \ldots \\ xi \end{pmatrix} \longleftrightarrow$$



**1 - Shape information (texture free)**

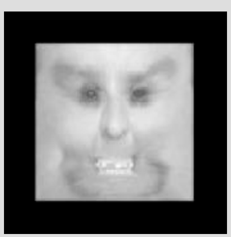
$$I \longleftrightarrow$$



**2 - Texture information (shape free)**

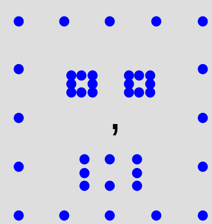
• 2 - Shape normalization is obtained by warping the image into the mean shape of the training database.

Shape free texture



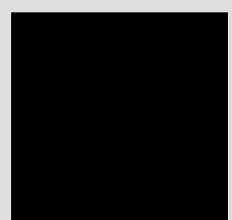
= ImageWarp (

Mean shape



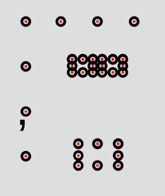
,

zeros

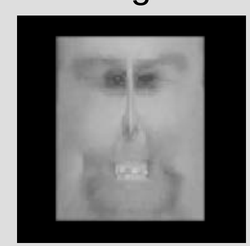


;

shape

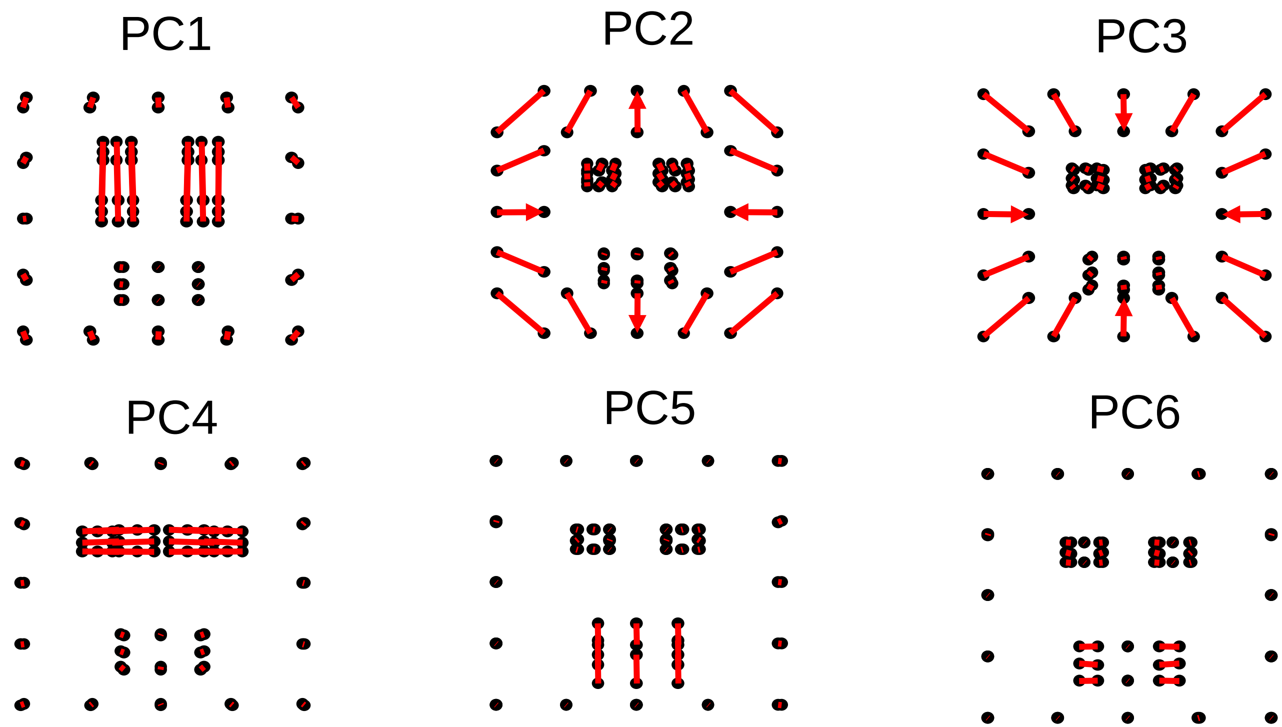


Original image

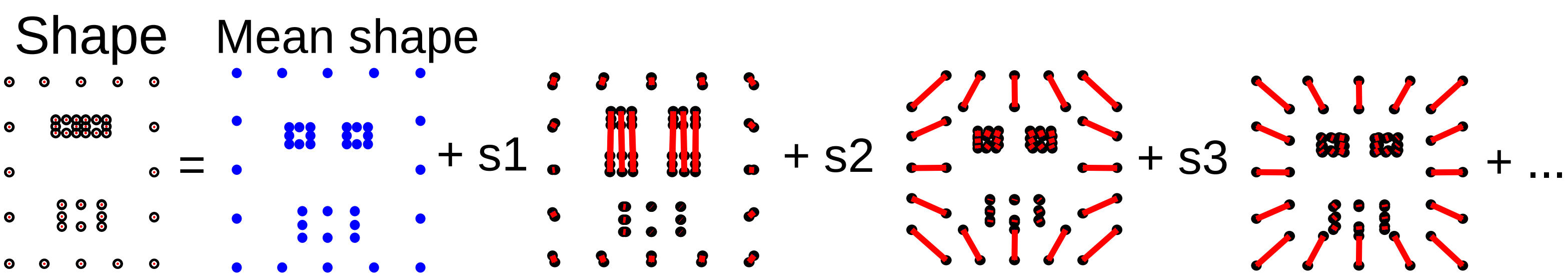


# Shape model

- **PCA of shape information for the training database:**




- **Each shape can be decomposed as:**

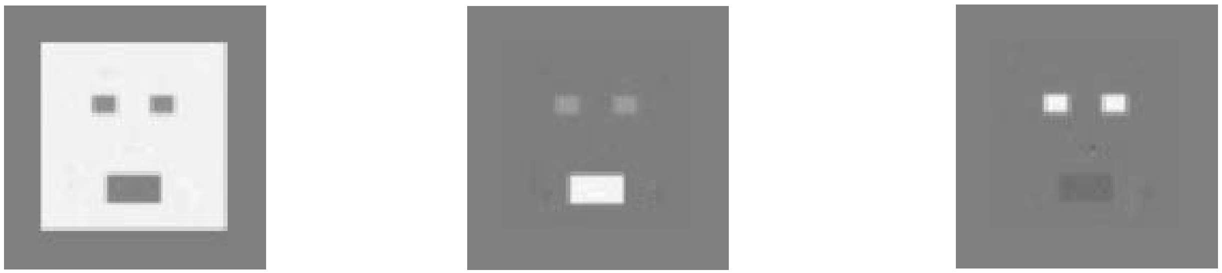


# Texture model

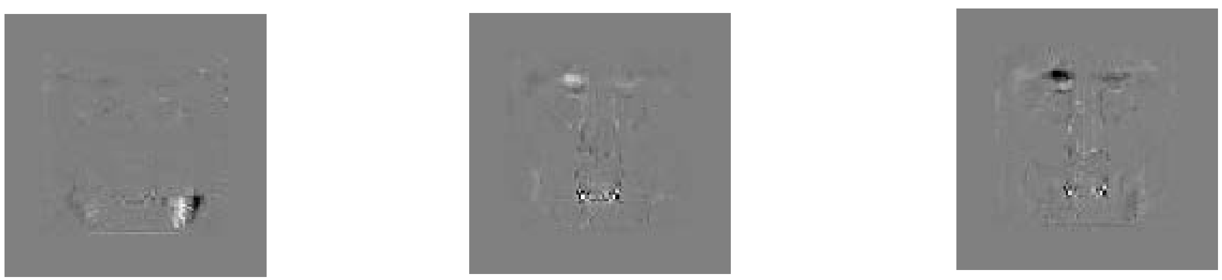
- **PCA of texture information for the training database:**

The PCA is done on the shape free images

PC1 PC2 PC3

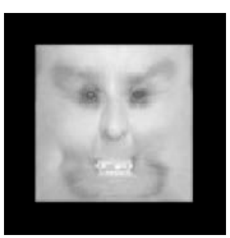


PC4 PC5 PC6



- **Each texture (shape free) can be decomposed as:**

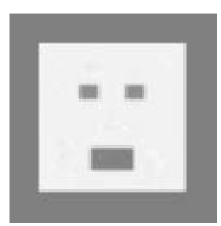
Shape free texture



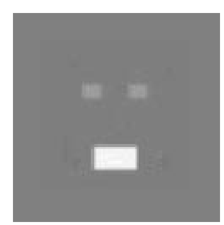
=

Mean texture

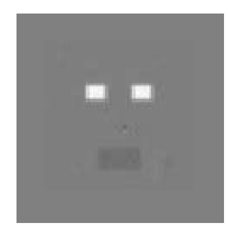
+ t1



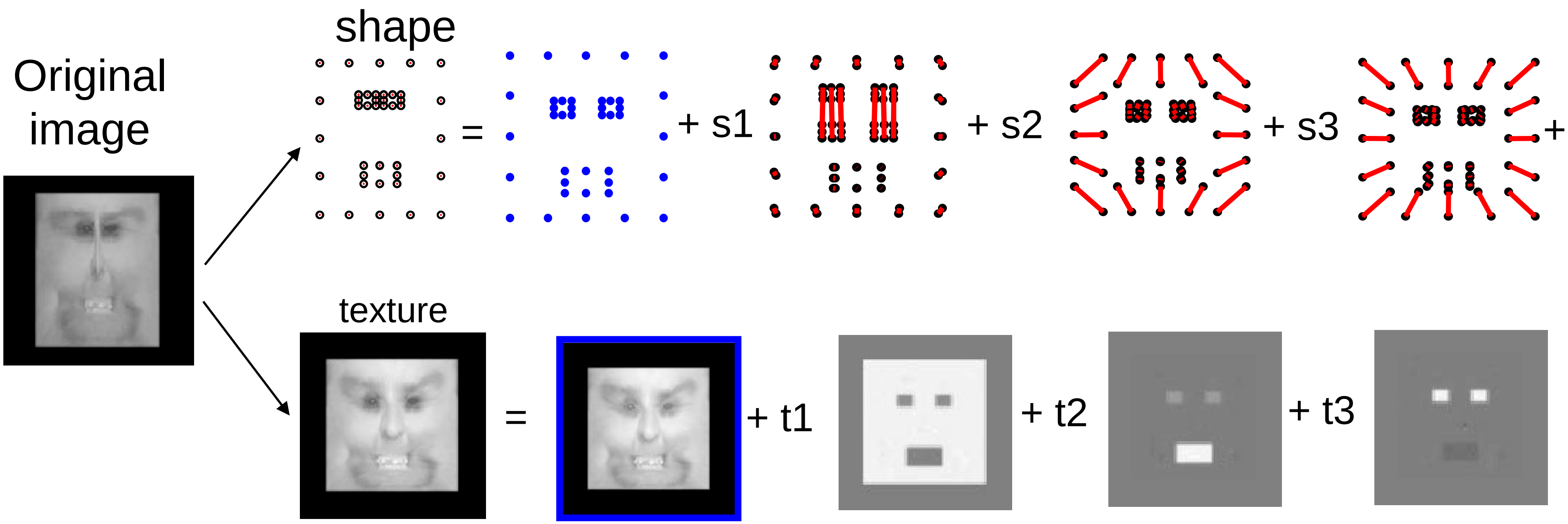
+ t2



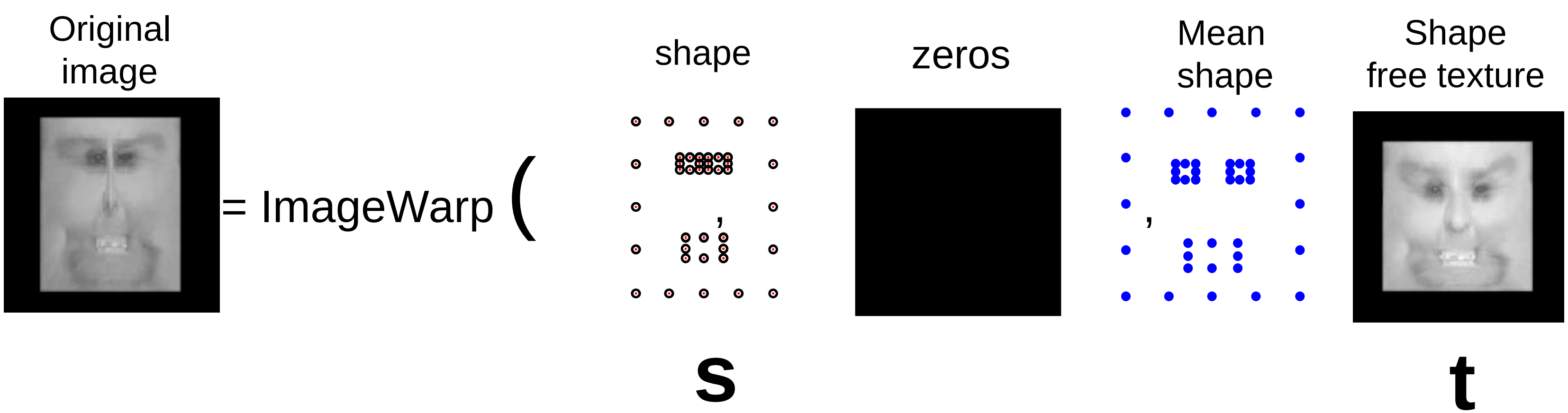
+ t3



# Summary of Appearance Model of one image



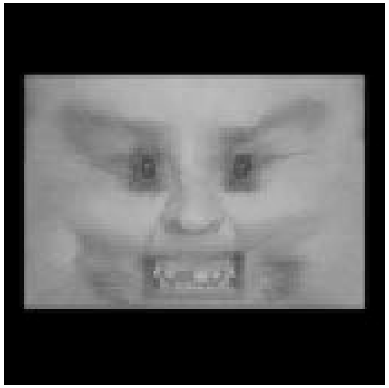

A set of model parameters encode shape and gray level variation learned from a training set



# Active Appearance Model Search

Given a new “face” the model has to build an appearance model (shape + texture) that reproduces the original image:

Novel face

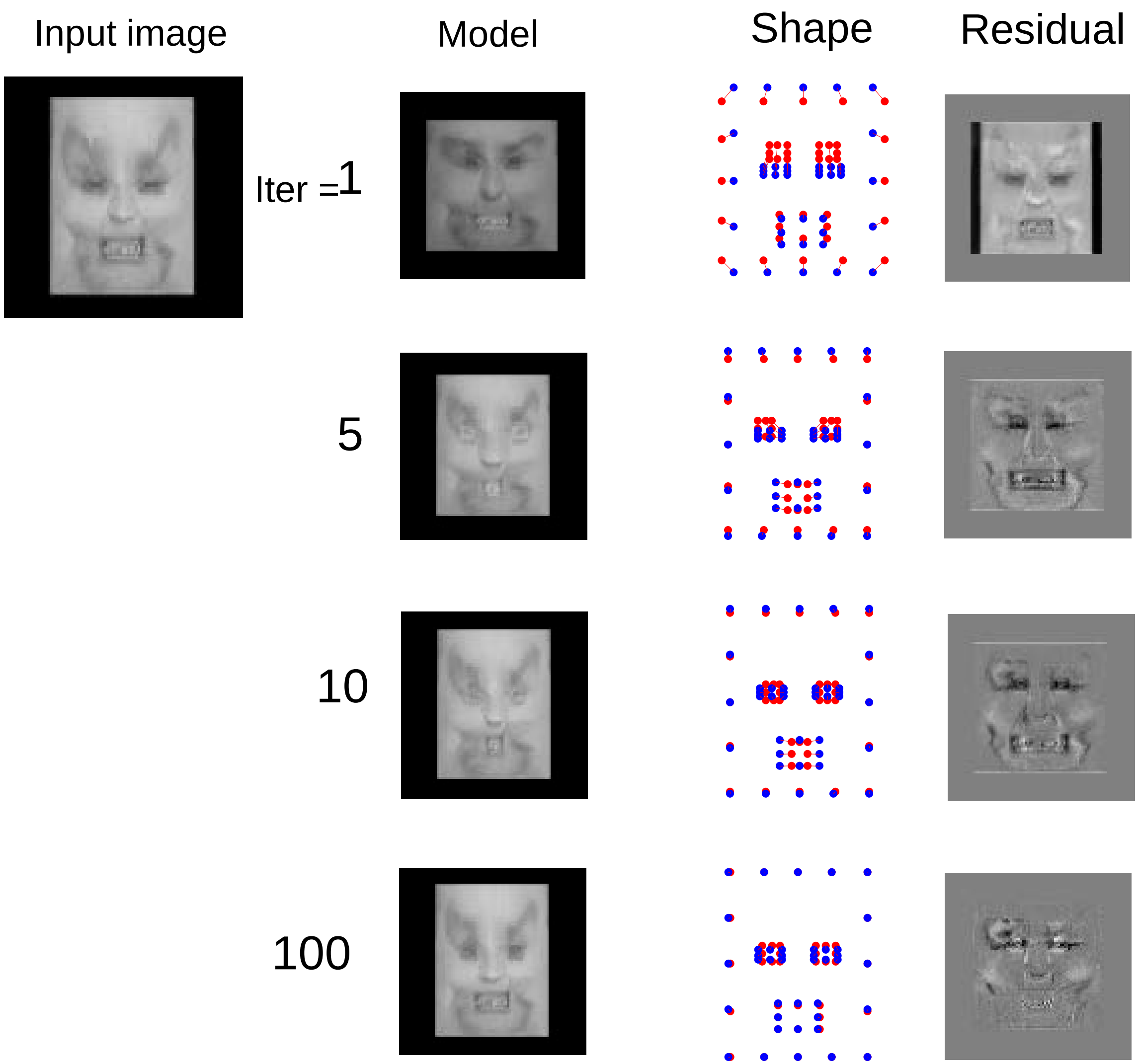


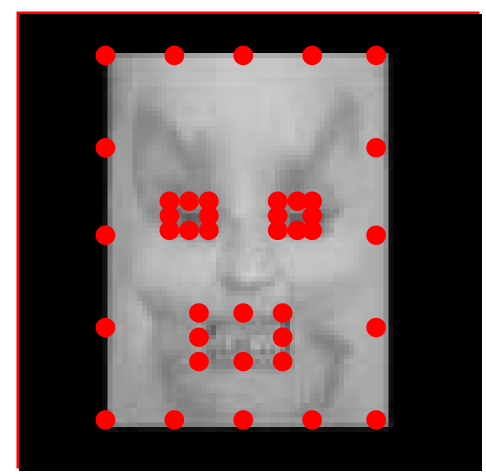
Shape = ?

Texture = ?

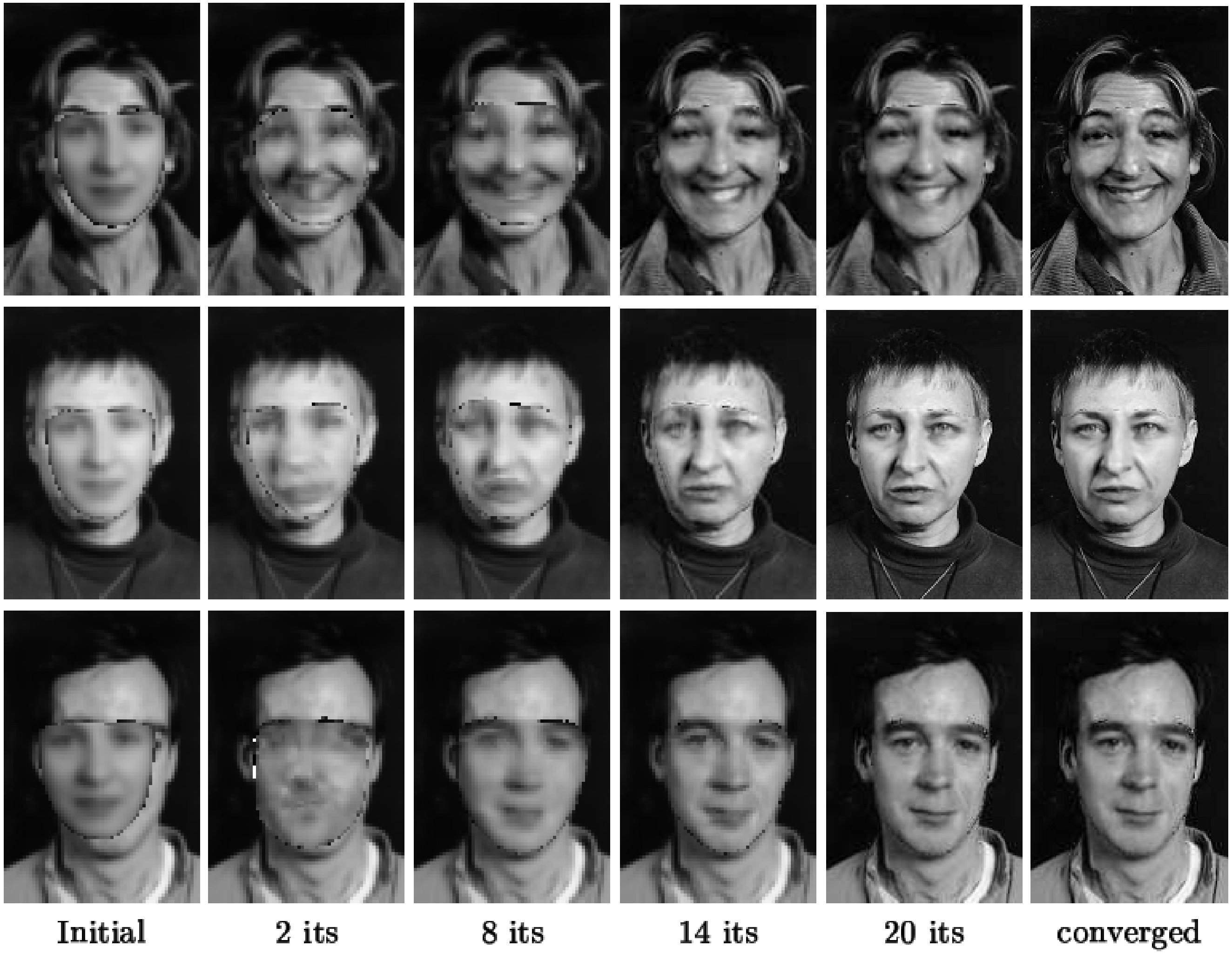
This is done in an iterative procedure that tries to minimize the reconstruction error.

# Results



# Active Appearance Model Search (Results)




Slide: Dhruv Batra

# Essence of the Idea: Recognition by Synthesis

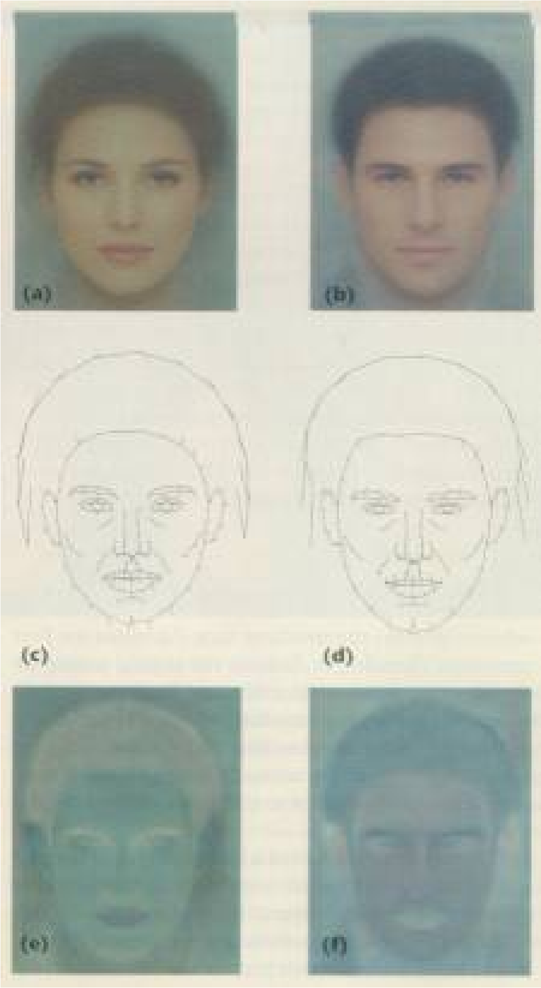
Explain a new example in terms of the model parameters

Initial    3 its    8 its    11 its    Converged    Original

Slide: Dhruv Batra

# Face Modeling

Compute *average* faces (color and shape)



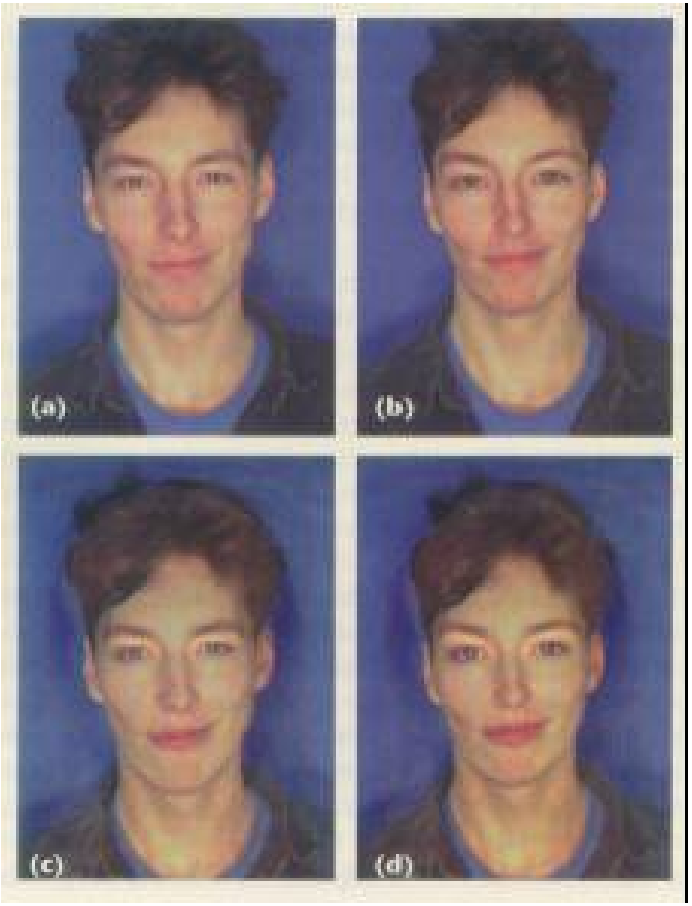

Compute *deviations* between male and female (vector and color differences)

# Changing gender

Deform shape and/or color of an input face in the direction of "more female"

original

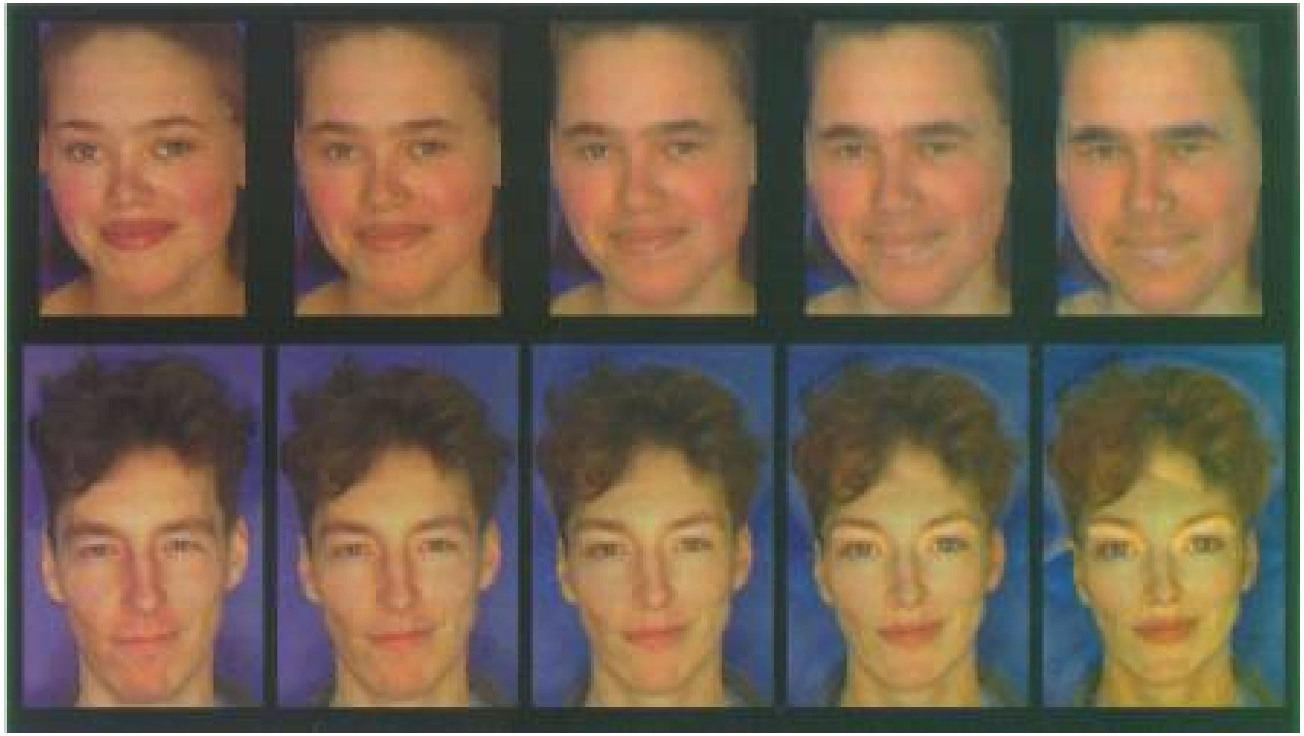
shape




color

both

# Enhancing gender



more same **original** androgynous more opposite

# Changing age

Face becomes "rounder" and "more textured" and "grayer"

original

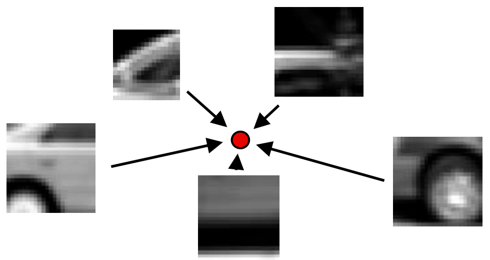
shape


color

both

# Structure models

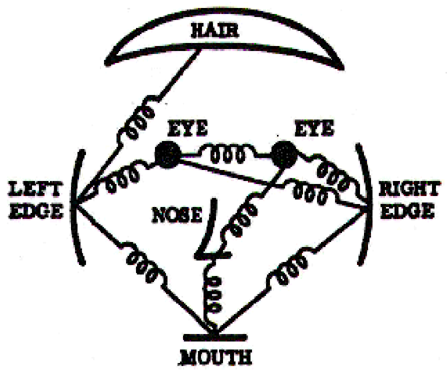
**Voting models**



- Many parts (>100)

Lecture 20

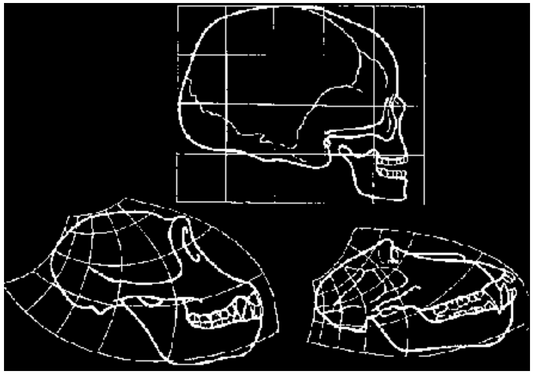
**Constellation models**




- Few parts (~6)

Lecture 21

**Deformable models**



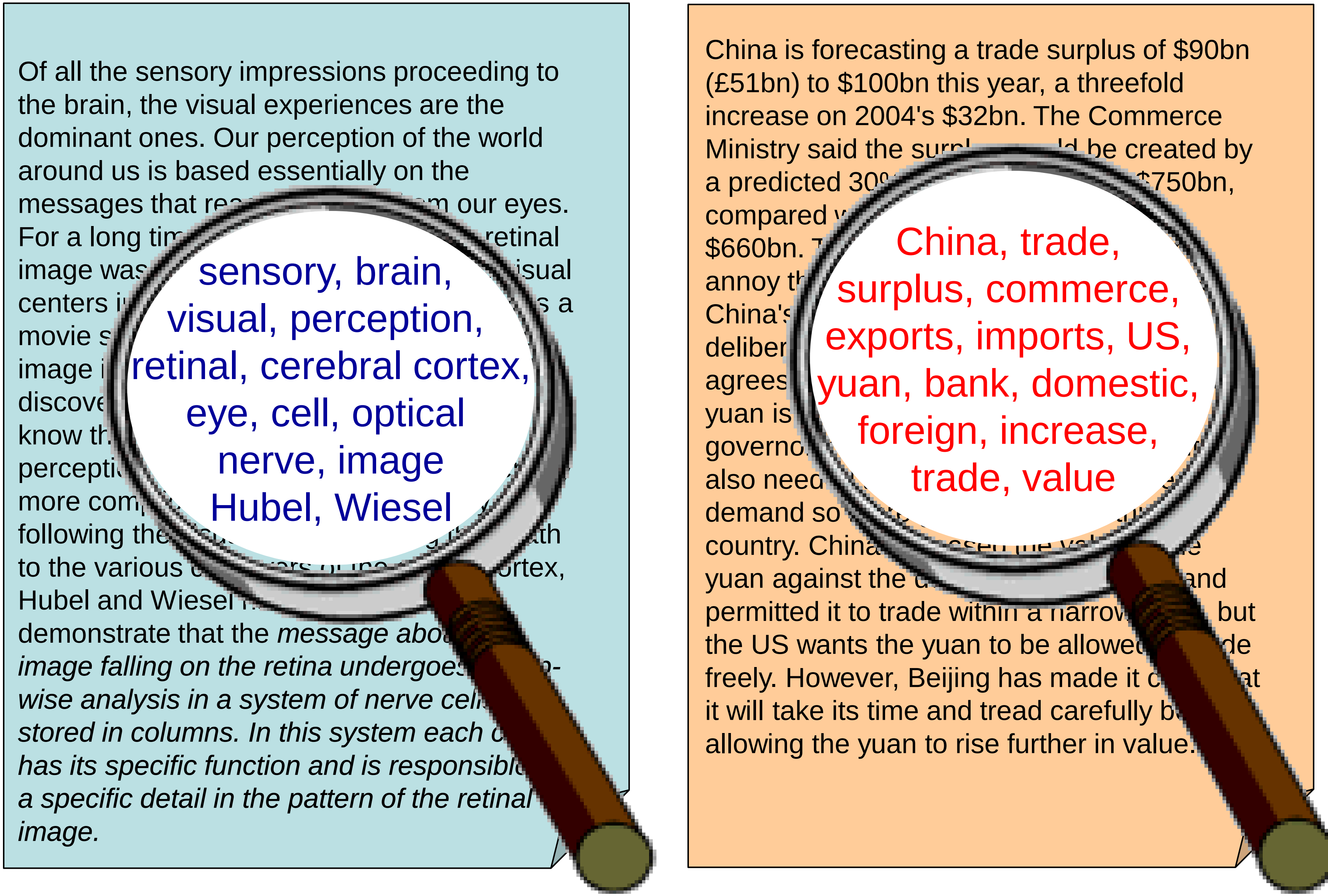
- No parts

Lecture 22

# Object → Bag of 'words'

Slide credit: Fei fei

# Analogy to documents




Slide credit: Fei fei

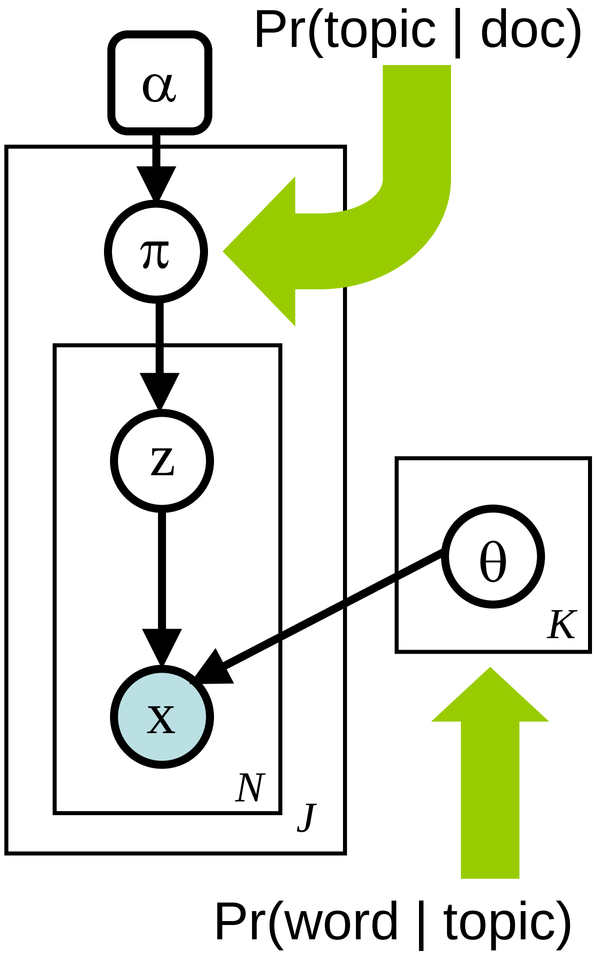
# Related works

- Early “bag of words” models: mostly texture recognition
  - Cula & Dana, 2001; Leung & Malik 2001; Mori, Belongie & Malik, 2001; Schmid 2001; Varma & Zisserman, 2002, 2003; Lazebnik, Schmid & Ponce, 2003;
- Hierarchical Bayesian models for documents (pLSA, LDA, etc.)
  - Hoffman 1999; Blei, Ng & Jordan, 2004; Teh, Jordan, Beal & Blei, 2004
- Object categorization
  - Csurka, Bray, Dance & Fan, 2004; Sivic, Russell, Efros, Freeman & Zisserman, 2005; Sudderth, Torralba, Freeman & Willsky, 2005;
- Natural scene categorization
  - Vogel & Schiele, 2004; Fei-Fei & Perona, 2005; Bosch, Zisserman & Munoz, 2006

Slide credit: Fei fei

# Hierarchical Topic Models

- Topic models typically use a *"bag of words"* approx.:
  - Learning topics allows transfer of information within a corpus of related documents
  - Mixing proportions capture the distinctive features of particular documents




Latent Dirichlet Allocation (LDA)

*Blei, Ng, & Jordan, JMLR 2003*

# Analogy: Discovering topics in text collections

Text document

The William Randolph Hearst Foundation will give $1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services," Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center's share will be $200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive $400,000 each. The Juilliard School, where music and the performing arts are taught, will get $250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual $100,000 donation, too.

Discovered topics

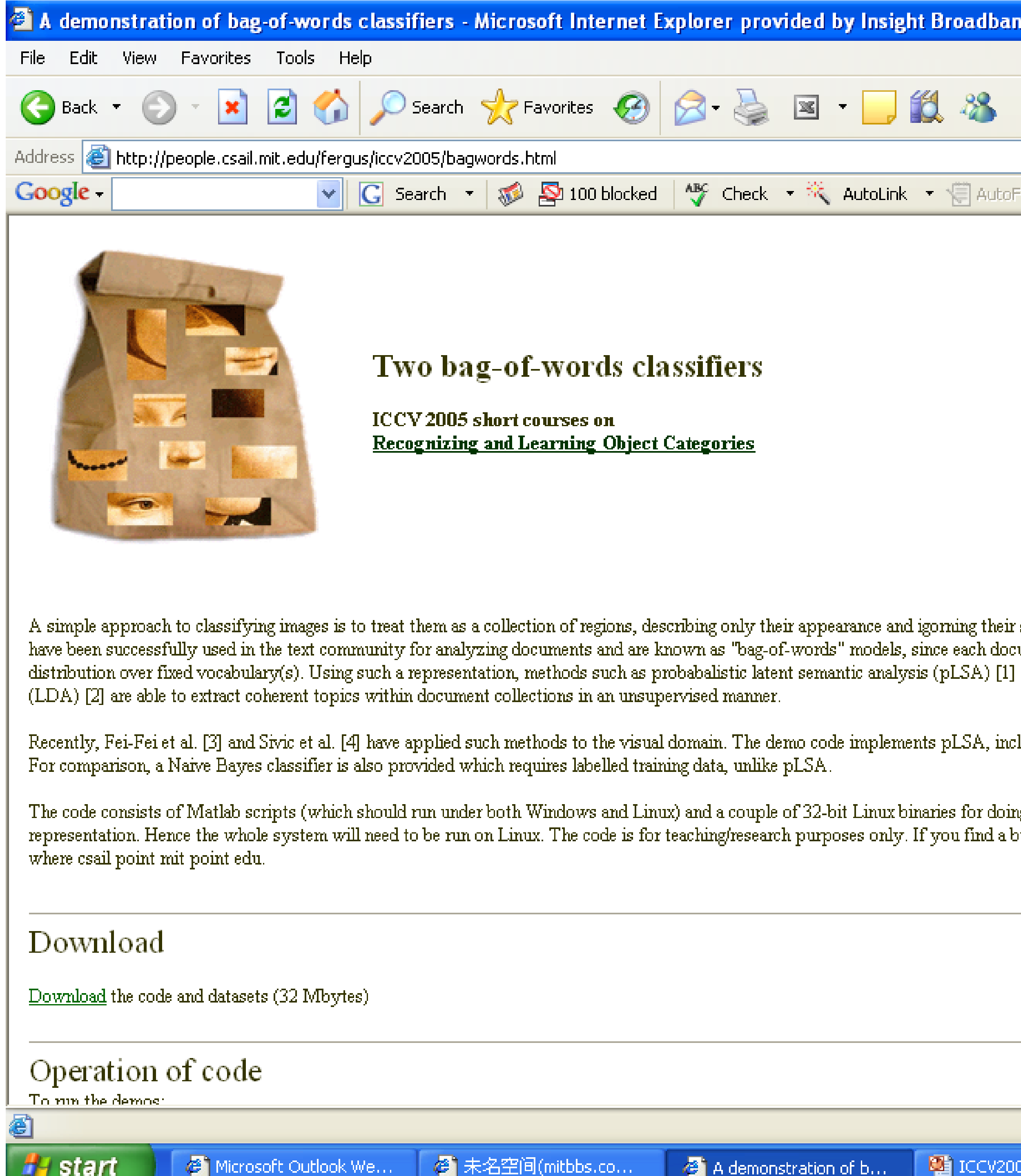
| "Arts" | "Budgets" | "Children" | "Education" |
|---|---|---|---|
| NEW | MILLION | CHILDREN | SCHOOL |
| FILM | TAX | WOMEN | STUDENTS |
| SHOW | PROGRAM | PEOPLE | SCHOOLS |
| MUSIC | BUDGET | CHILD | EDUCATION |
| MOVIE | BILLION | YEARS | TEACHERS |
| PLAY | FEDERAL | FAMILIES | HIGH |
| MUSICAL | YEAR | WORK | PUBLIC |
| BEST | SPENDING | PARENTS | TEACHER |
| ACTOR | NEW | SAYS | BENNETT |
| FIRST | STATE | FAMILY | MANIGAT |
| YORK | PLAN | WELFARE | NAMPHY |
| OPERA | MONEY | MEN | STATE |
| THEATER | PROGRAMS | PERCENT | PRESIDENT |
| ACTRESS | GOVERNMENT | CARE | ELEMENTARY |
| LOVE | CONGRESS | LIFE | HAITI |

Blei, et al. 2003

# Visual analogy

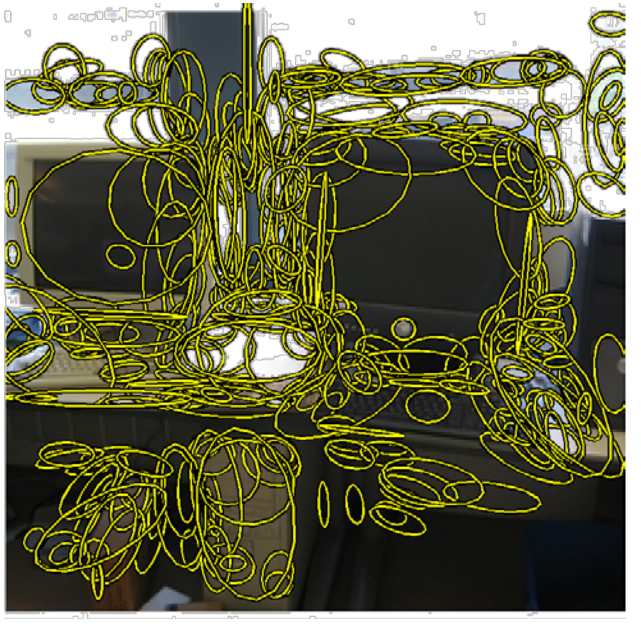
| | | |
|---|---|---|
| document | - | image |
| word | - | visual word |
| topics | - | objects |

# Demo


A demonstration of bag-of-words classifiers - Microsoft Internet Explorer provided by Insight Broadban
File Edit View Favorites Tools Help
Back
Search
Favorites
Address http://people.csail.mit.edu/fergus/iccv2005/bagwords.html
Google
Search
100 blocked
Check
AutoLink
Two bag-of-words classifiers
ICCV 2005 short courses on
Recognizing and Learning Object Categories
A simple approach to classifying images is to treat them as a collection of regions, describing only their appearance and igorning their
have been successfully used in the text community for analyzing documents and are known as "bag-of-words" models, since each docu
distribution over fixed vocabulary(s). Using such a representation, methods such as probabalistic latent semantic analysis (pLSA) [1]
(LDA) [2] are able to extract coherent topics within document collections in an unsupervised manner.
Recently, Fei-Fei et al. [3] and Sivic et al. [4] have applied such methods to the visual domain. The demo code implements pLSA, incl
For comparison, a Naive Bayes classifier is also provided which requires labelled training data, unlike pLSA.
The code consists of Matlab scripts (which should run under both Windows and Linux) and a couple of 32-bit Linux binaries for doin
representation. Hence the whole system will need to be run on Linux. The code is for teaching/research purposes only. If you find a b
where csail point mit point edu.
Download
Download the code and datasets (32 Mbytes)
Operation of code
start
Microsoft Outlook We...
未名空间(mitbbs.co...
A demonstration of b...

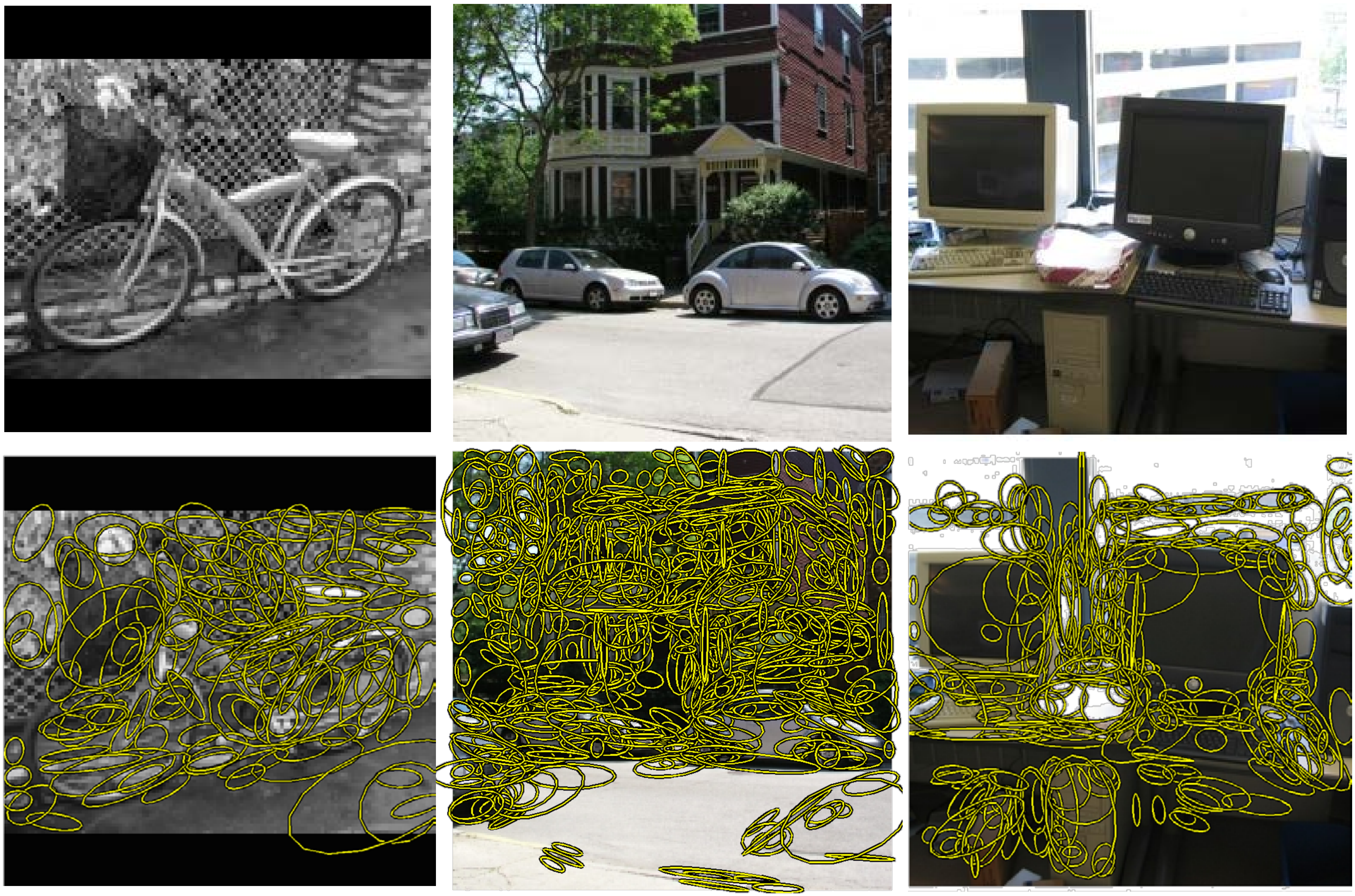
# From Images to Features

- Pixels are very sensitive to changes in lighting & pose
- Instead represent image as *affine covariant regions:*
  - Harris affine invariant regions (corners & edges)
  - Maximally stable extremal regions (segmentation)



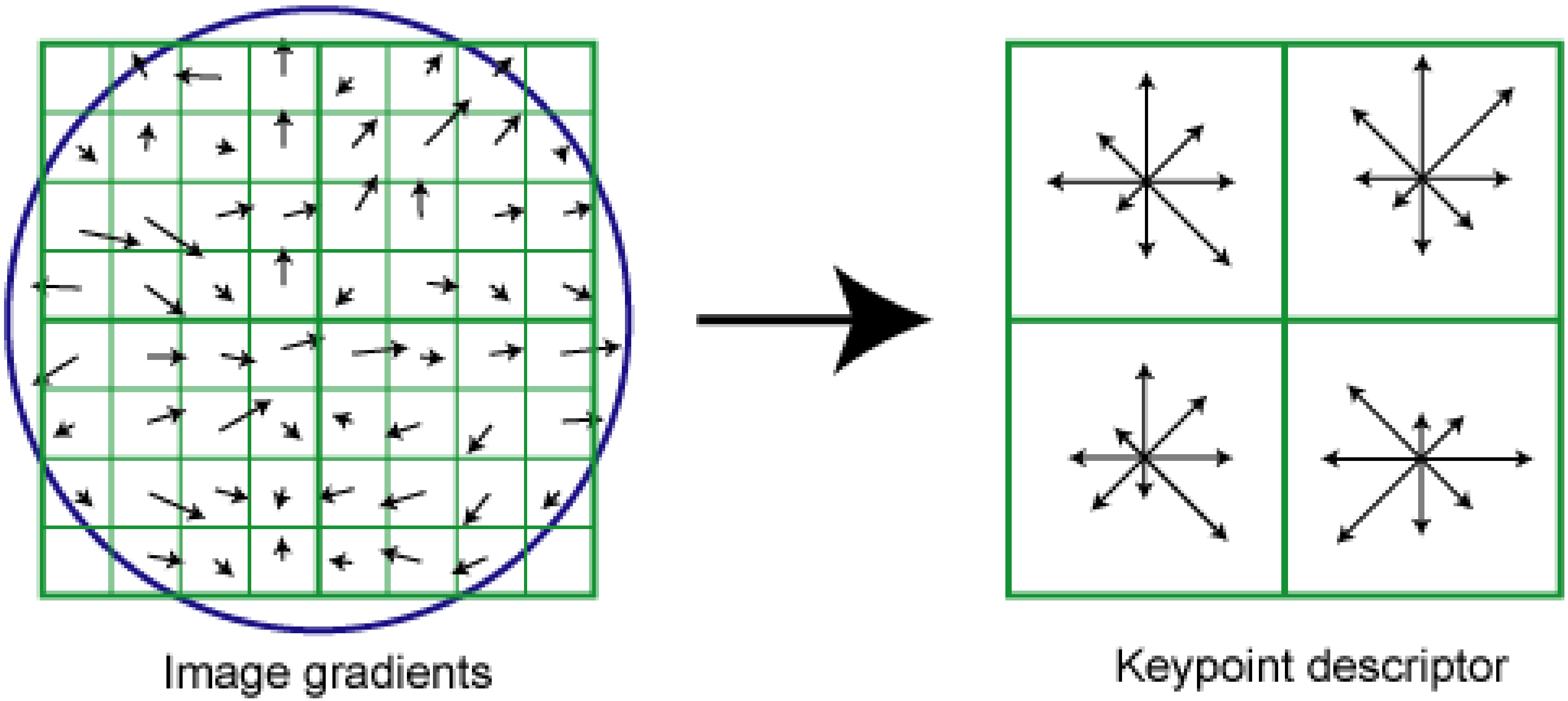
Software provided by
*Oxford Visual Geometry Group*

# Sample Detected Features



# Describing Feature Appearance

- **SIFT:** Scale Invariant Feature Transform
- Normalized histogram of orientation energy in each affinely adapted region (128-dim.)



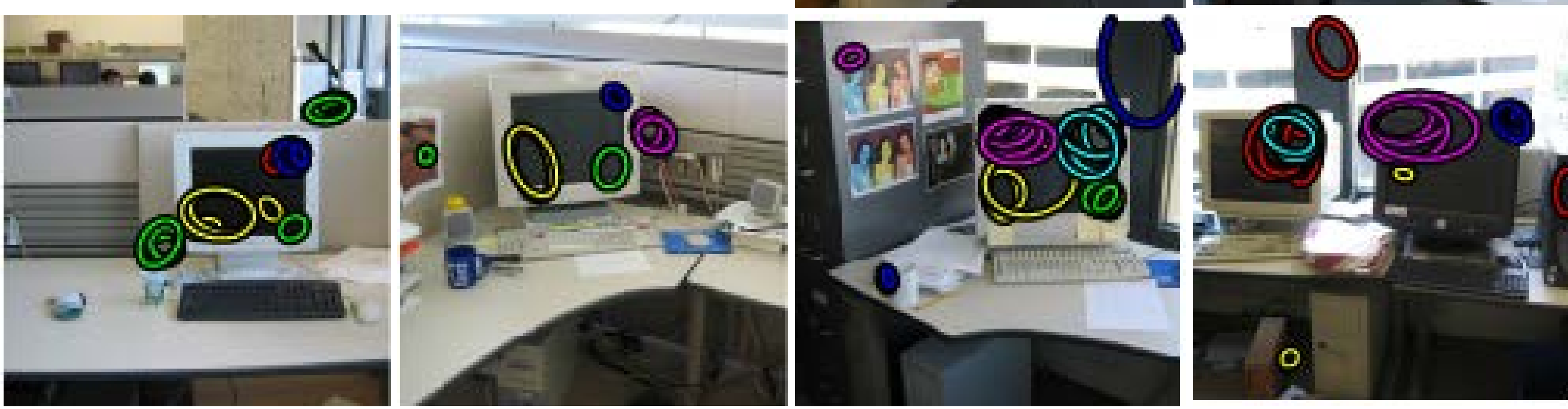

*D. Lowe, IJCV 2004*

# A Discrete Feature Vocabulary

- Using all training images, build a dictionary via K-means clustering (~1000 words)
- Map each SIFT descriptor to nearest word

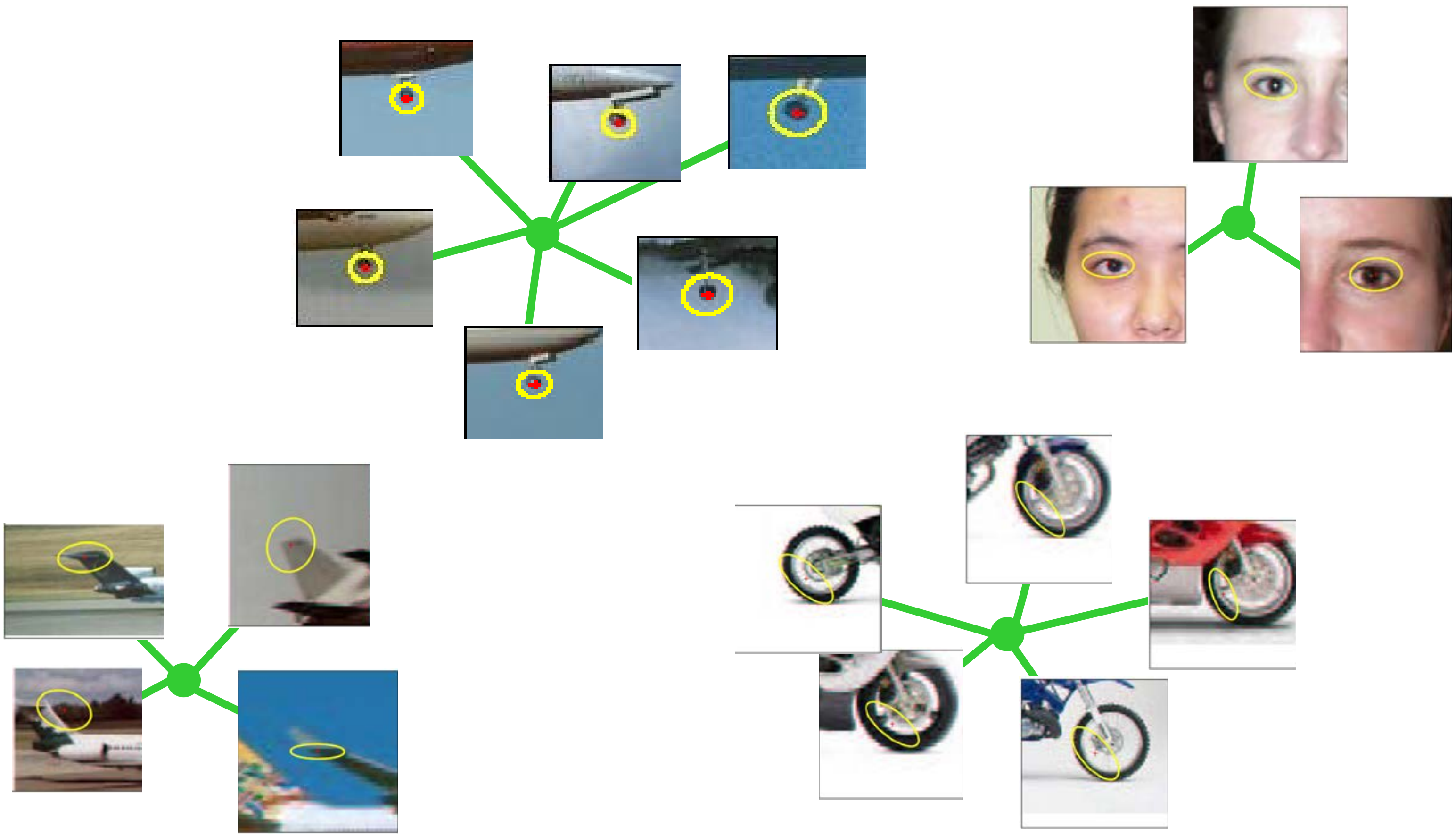
$w_{ji} \longrightarrow$ appearance of feature $i$ in image $j$

$y_{ji} \longrightarrow$ 2D position of feature $i$ in image $j$



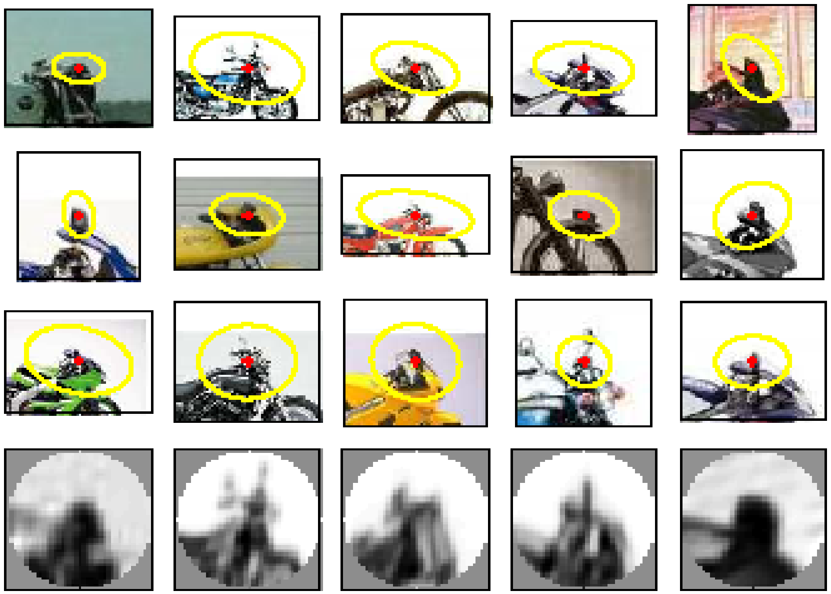
# Form dictionary

Build visual vocabulary by k-means clustering SIFT descriptors (K~2,000)

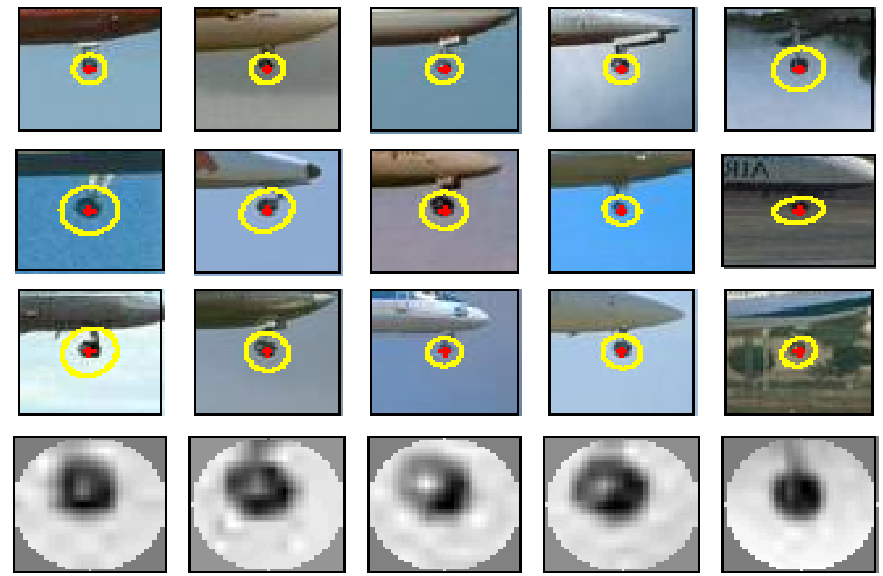


Slide credit: Bryan Russell & Josef Sivic

# Example regions assigned to the same dictionary cluster



Cluster 1

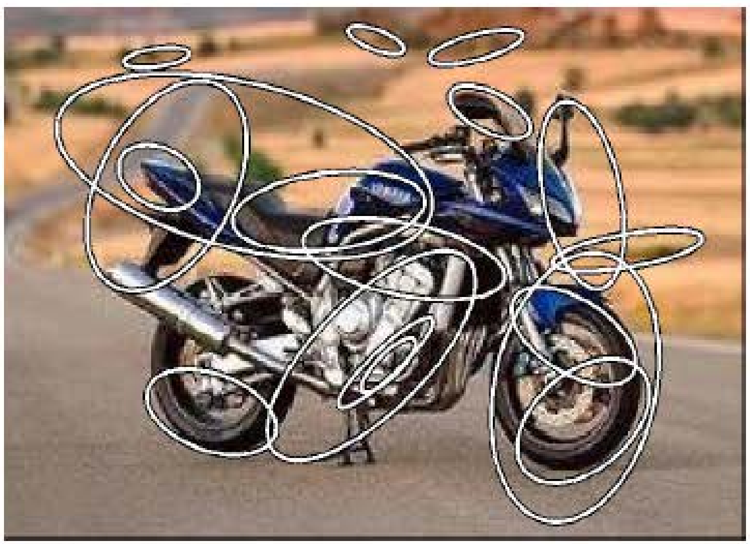


Cluster 2

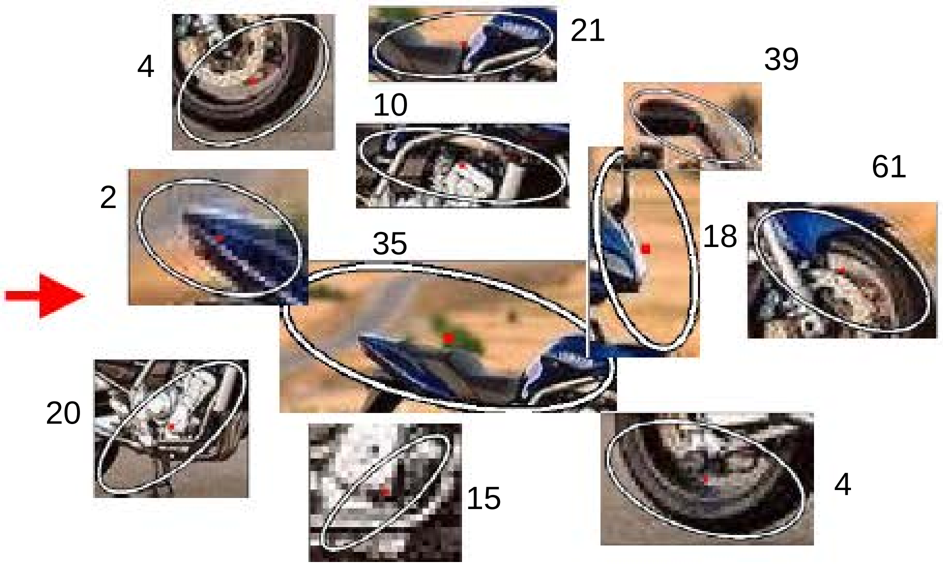
Slide credit: Bryan Russell & Josef Sivic

# Representing an image with visual words

Sivic & Zisserman '03



Interest regions




Visual words

Slide credit: Bryan Russell & Josef Sivic

# System overview

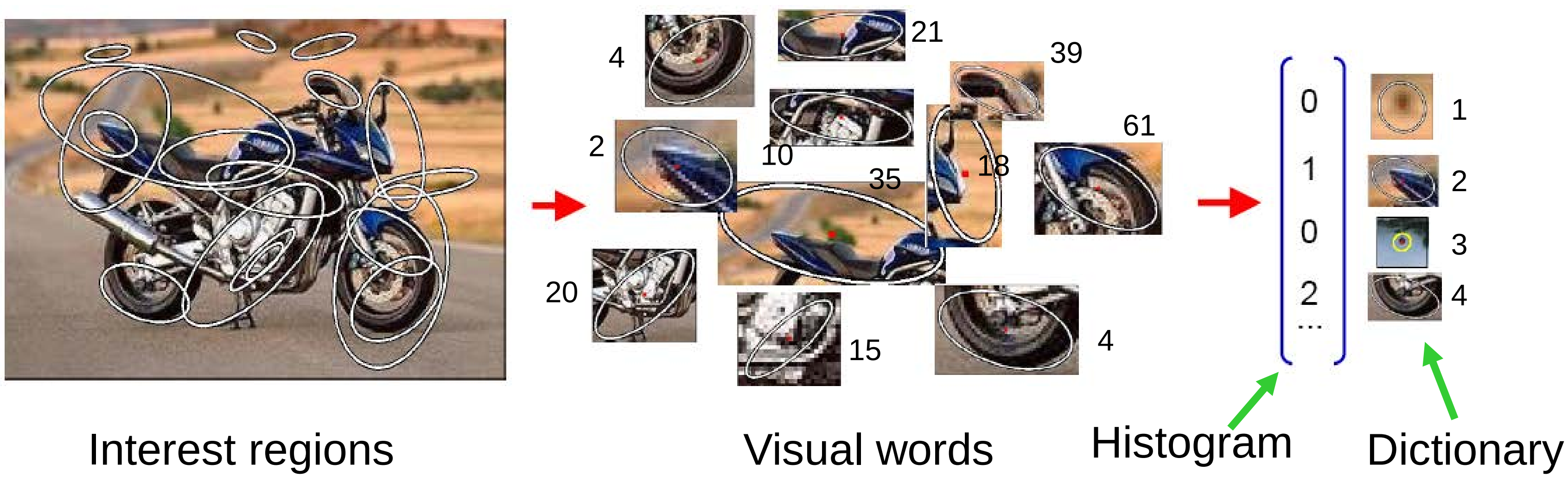
Input image

Compute visual words

Discover visual topics

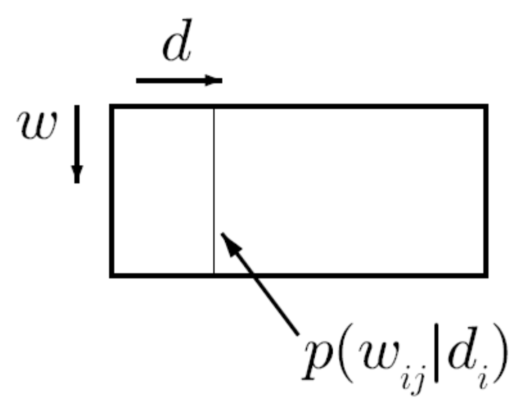
Slide credit: Bryan Russell & Josef Sivic

# Bag of words




Stack visual word histograms
as columns in matrix

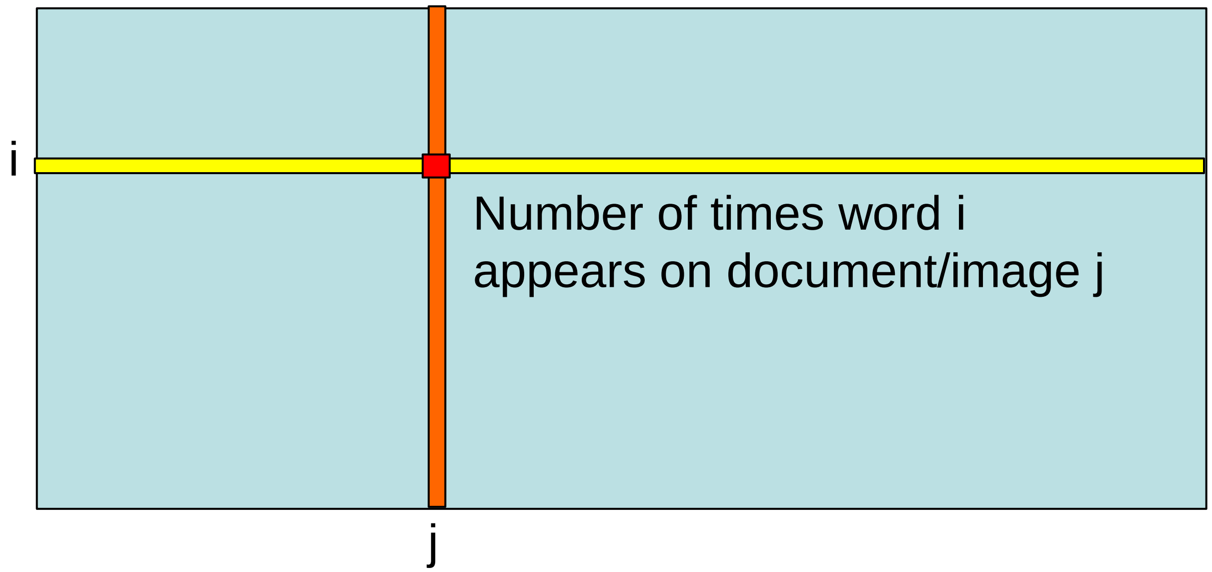
Throw away spatial information!




Slide credit: Bryan Russell & Josef Sivic

# Documents collection

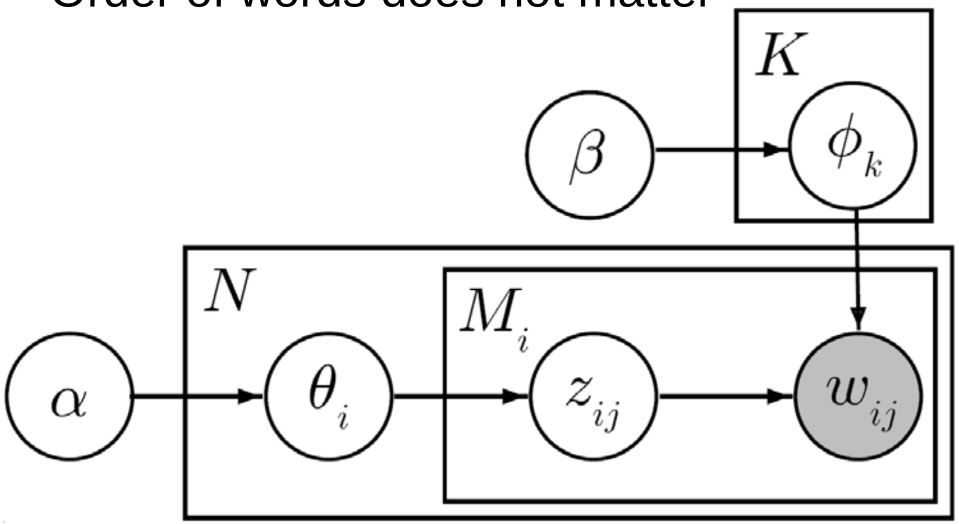
Co-ocurrence table:



# Latent Dirichlet Allocation (LDA)

Blei, et al. 2003

- LDA model assumes exchangeability
- Order of words does not matter




$w_{ij}$ - words

$z_{ij}$ - topic assignments

$\mu_i$ - topic mixing weights

$\Phi_k$ - word mixing weights

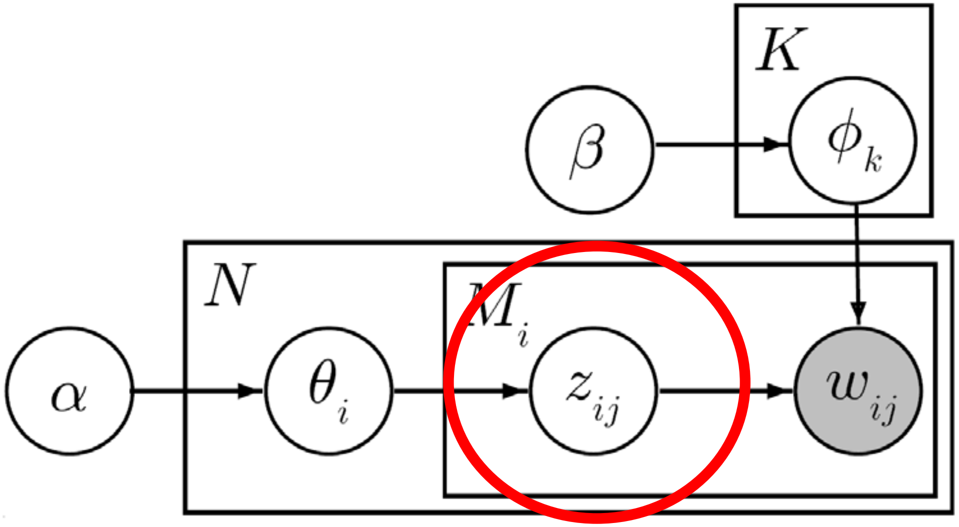
$$z_{ij}|\theta_i \sim \theta_i \qquad \theta_i|\alpha \sim Dirichlet(\alpha)$$

$$w_{ij}|z_{ij}=k,\phi \sim \phi_k \qquad \phi_k|\beta \sim Dirichlet(\beta)$$

$$p(w_{ij}) \propto \sum_{k=1}^{K} p(w_{ij}|z_{ij}=k,\phi_k)\ p(z_{ij}=k|\theta_i)$$

Slide credit: Bryan Russell & Josef Sivic

# Inference




$w_{ij}$ - words

$z_{ij}$ - topic assignments

$\mu_i$ - topic mixing weights

Á$_k$ - word mixing weights

Use Gibbs sampler to sample topic assignments

[Griffiths & Steyvers 2004]

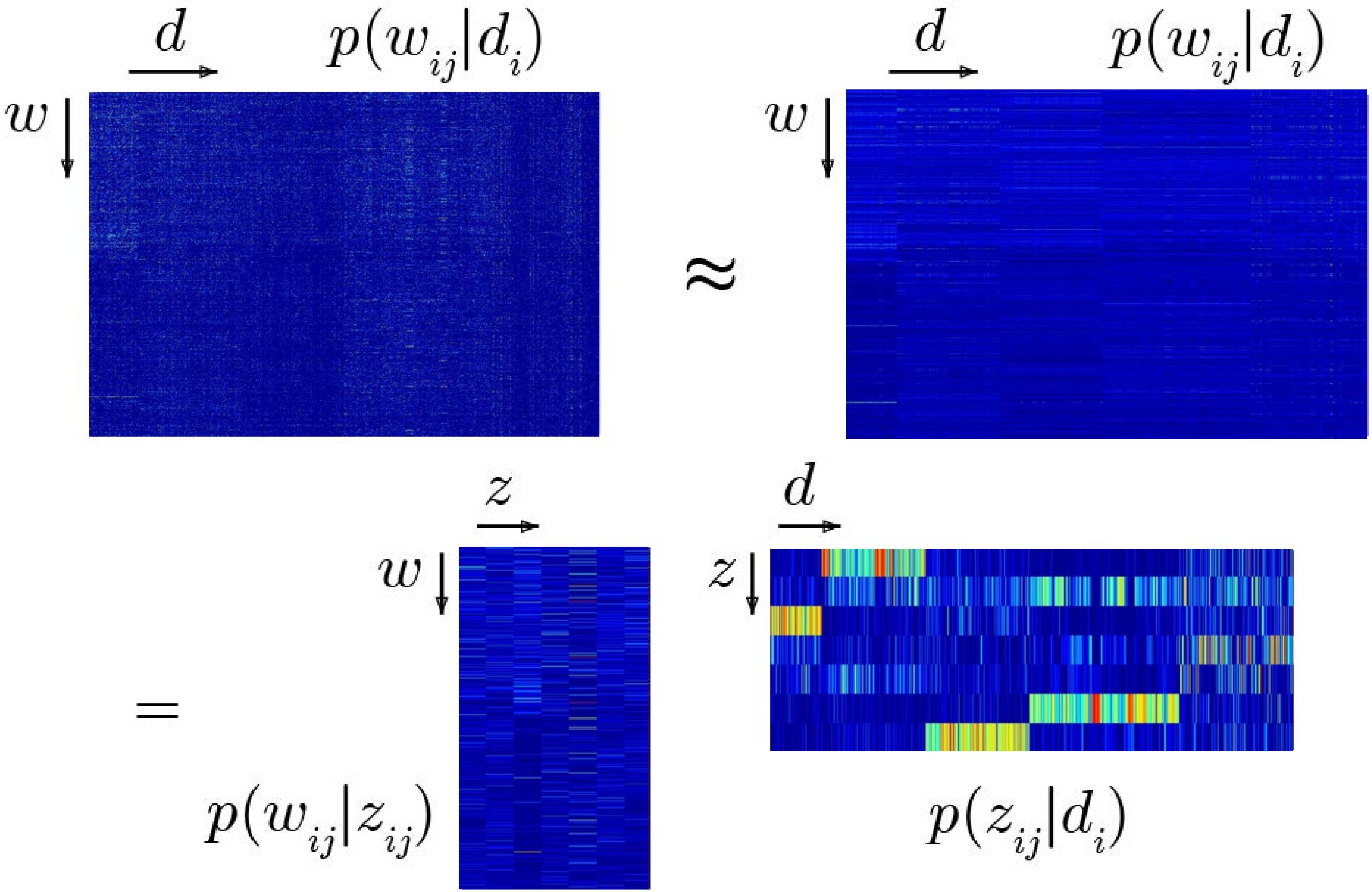
$$z_{ij} \sim p(z_{ij} = k | w_{ij} = v, w_{\setminus(ij)}, z_{\setminus(ij)}, \alpha, \beta)$$

- Only need to maintain counts of topic assignments
- Sampler typically converges in less than 50 iterations
- Run time is less than an hour

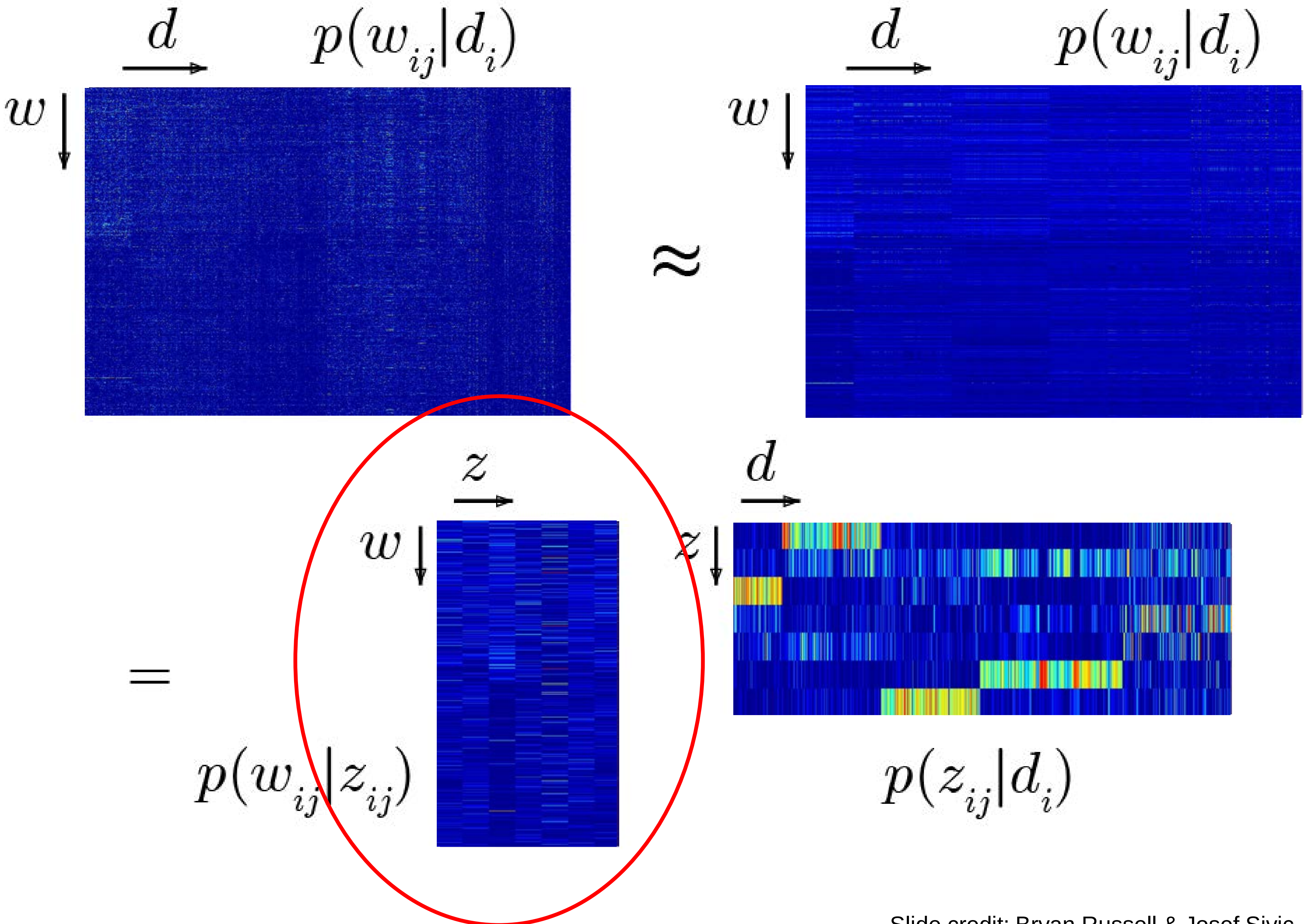
# Apply to Caltech 4 + background images

| | |
|---|---|
| Faces | 435 |
| Motorbikes | 800 |
| Airplanes | 800 |
| Cars (rear) | 1155 |
| Background | 900 |
| Total: | 4090 |

Slide credit: Bryan Russell & Josef Sivic




Slide credit: Bryan Russell & Josef Sivic



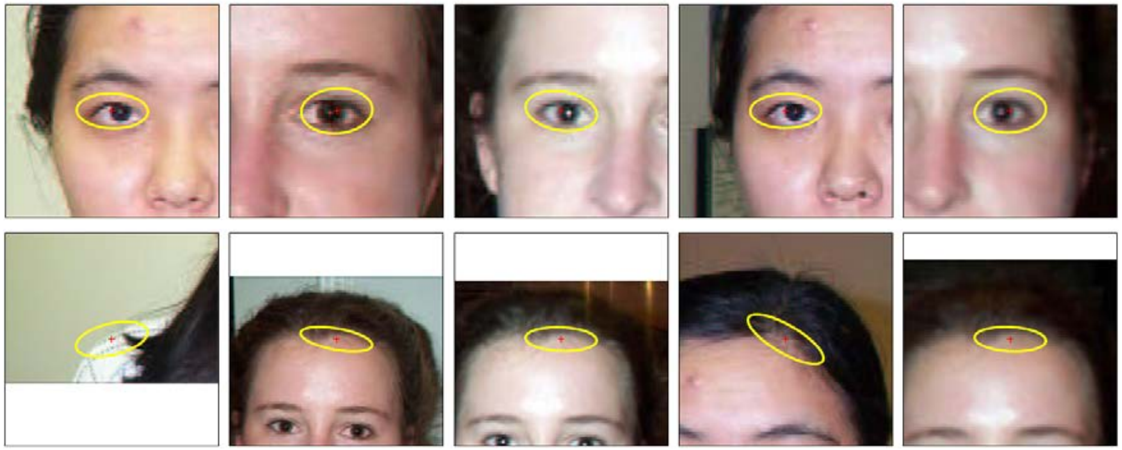

Slide credit: Bryan Russell & Josef Sivic

# Most likely words given topic

Topic 1

Word 1

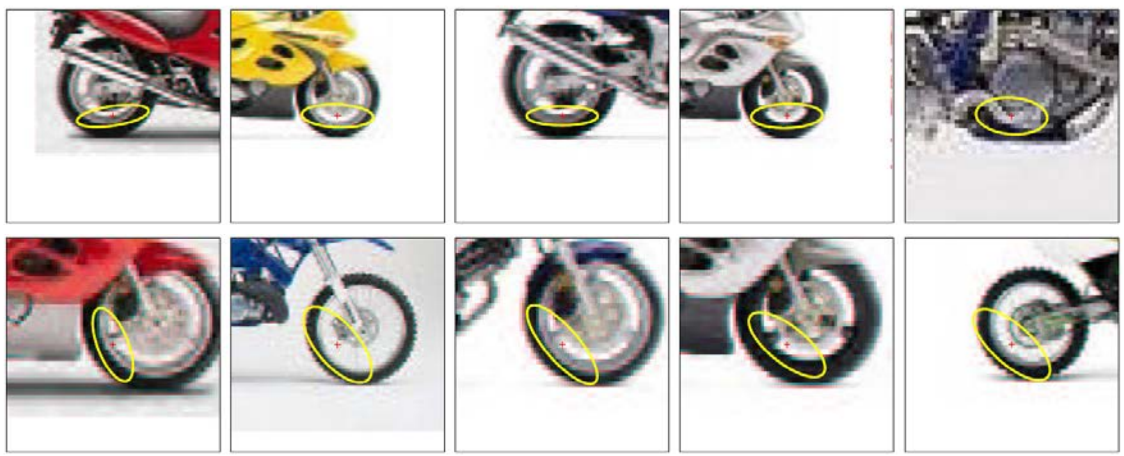
Word 2



Topic 2

Word 1

Word 2



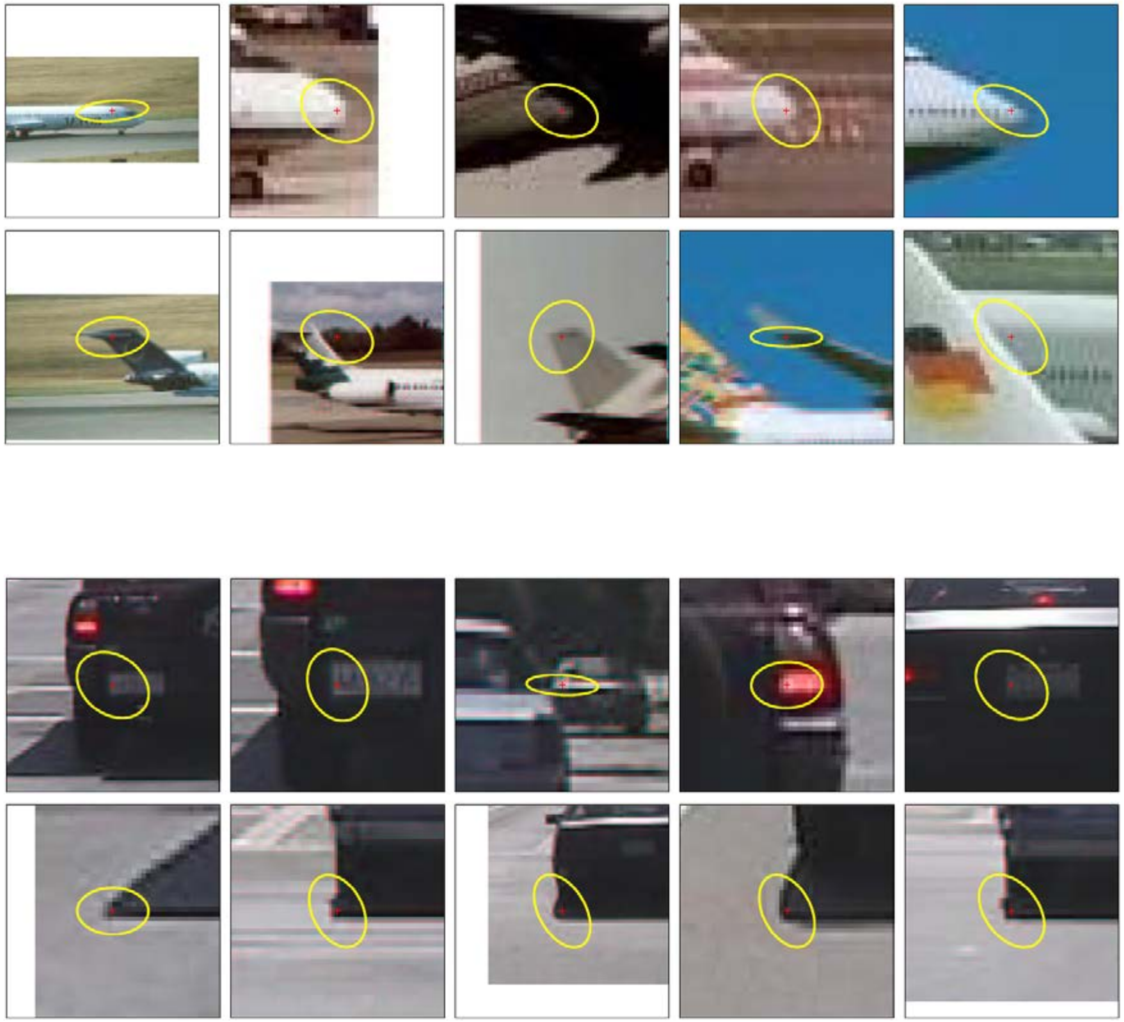
Slide credit: Bryan Russell & Josef Sivic

# Most likely words given topic

Topic 3

Word 1

Word 2

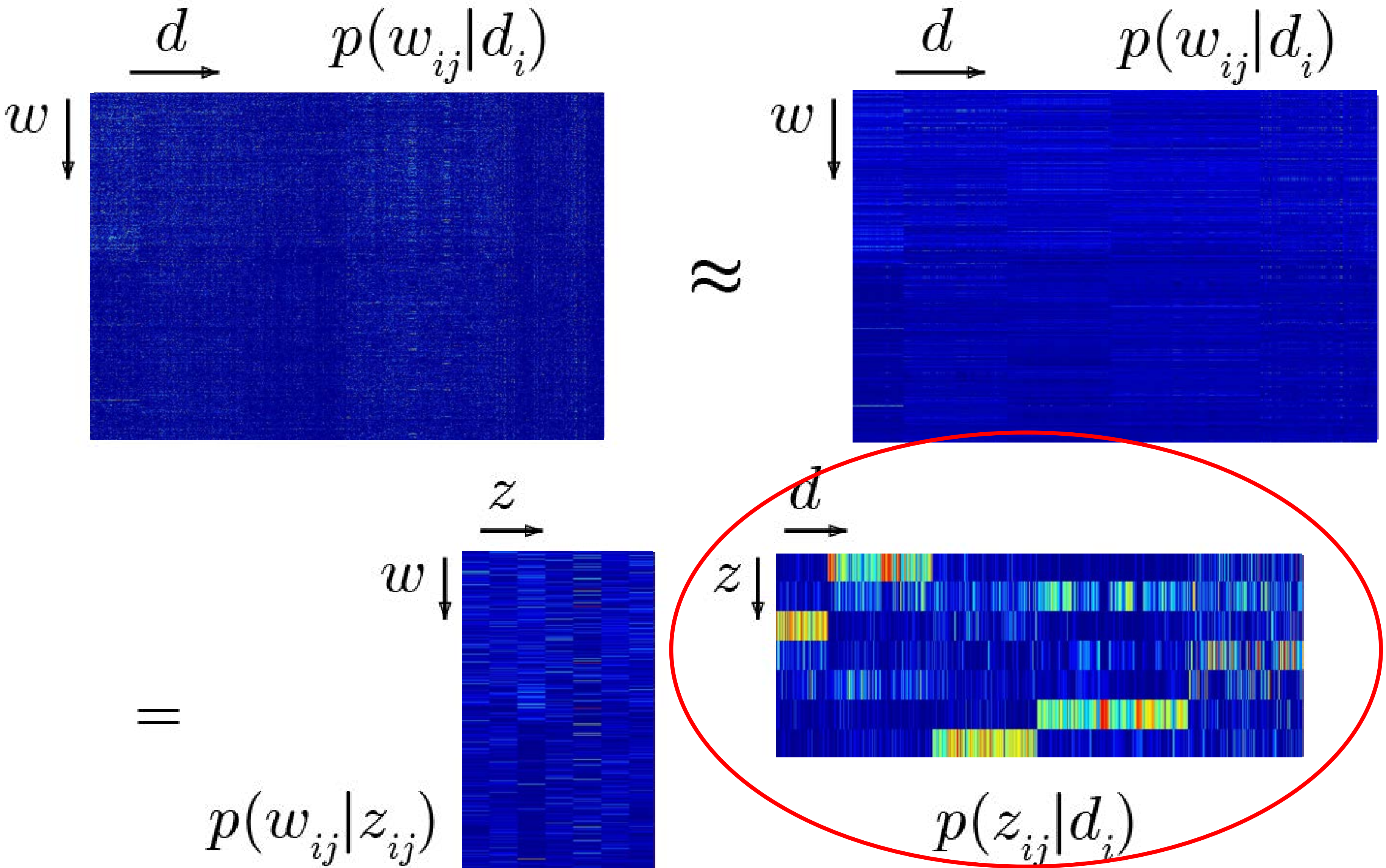


Topic 4

Word 1

Word 2

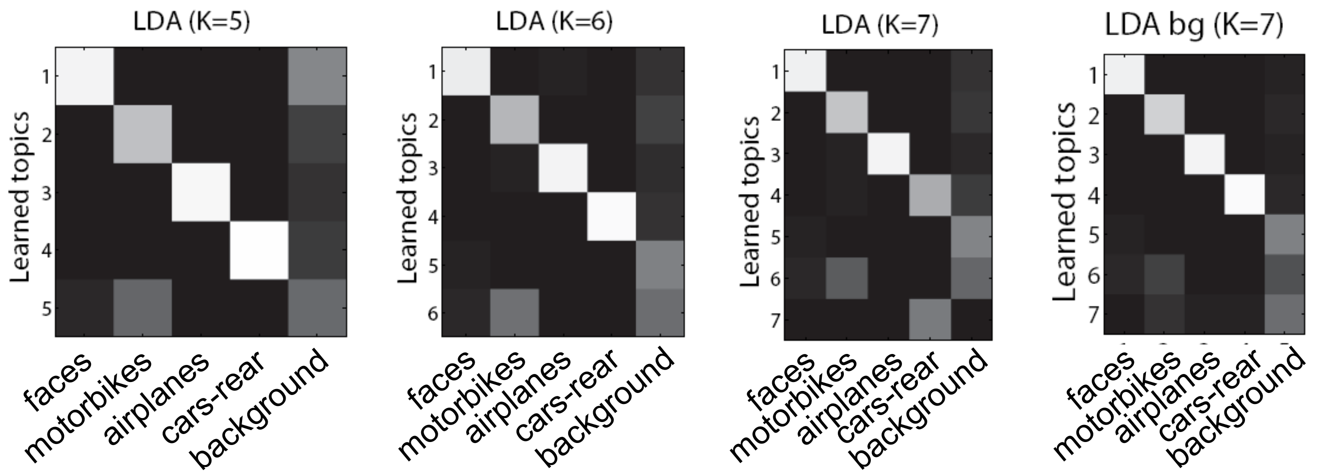
Slide credit: Bryan Russell & Josef Sivic




Slide credit: Bryan Russell & Josef Sivic

# Image clustering

Confusion matrices:

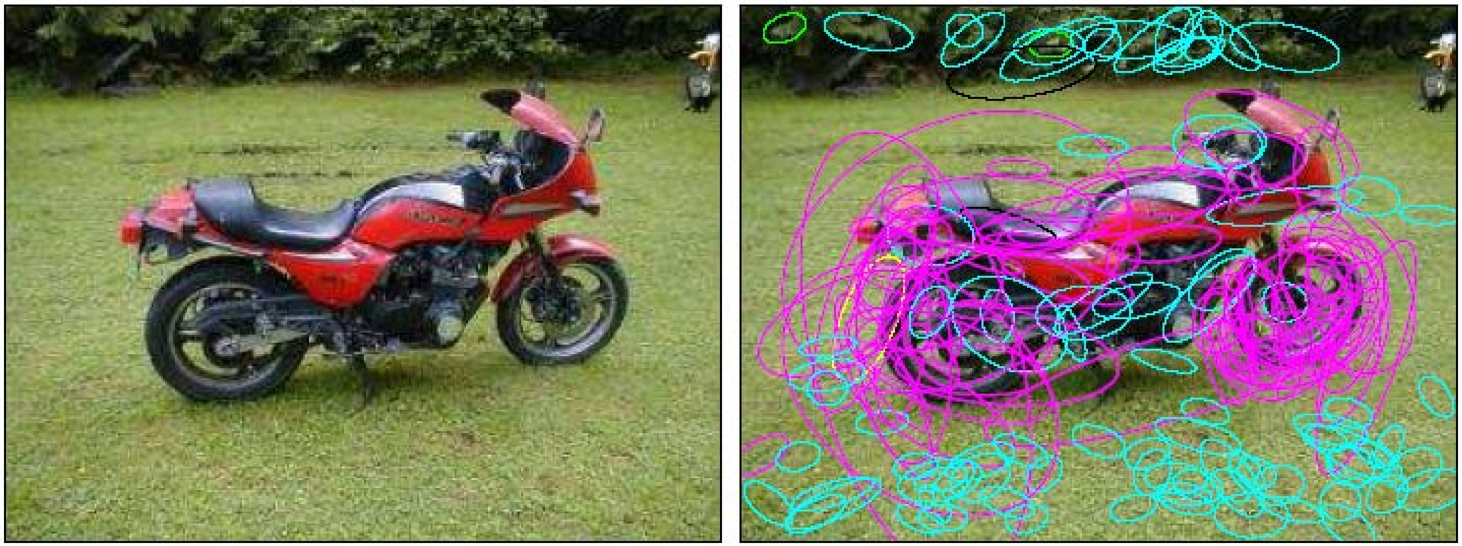



Average confusion:

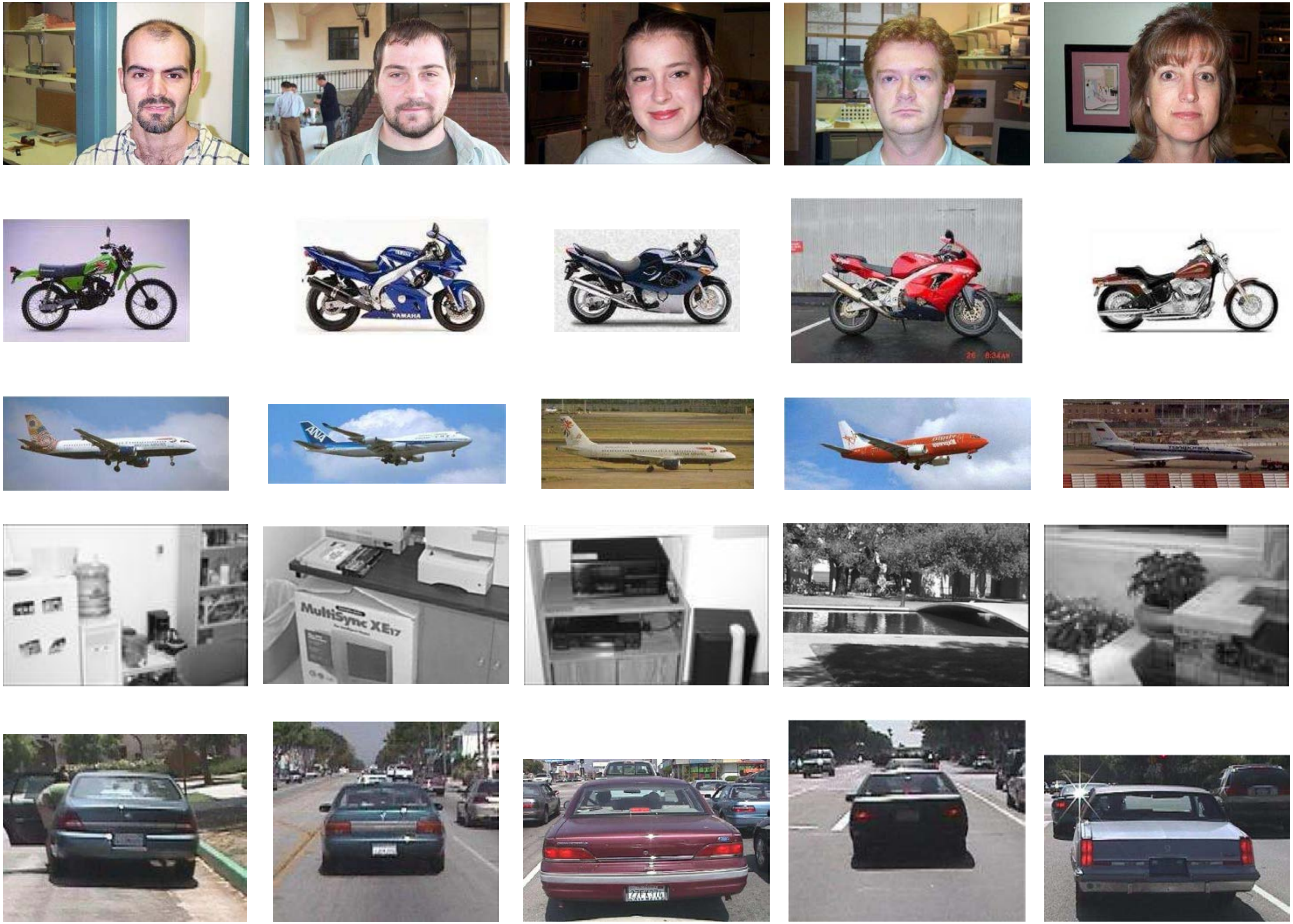
| Expt. | Categories | T | LDA % | LDA # | pLSA % | pLSA # | KM baseline % | KM baseline # |
|---|---|---|---|---|---|---|---|---|
| (1) | 4 | 4 | 97 | 86 | 98 | 70 | 72 | 908 |
| (2) | 4 + bg | 5 | 78 | 931 | 78 | 931 | 56 | 1820 |
| (2)* | 4 + bg | 6 | 84 | 656 | 76 | 1072 | – | – |
| (2)* | 4 + bg | 7 | 78 | 1007 | 83 | 768 | – | – |
| (2)* | 4 + bg-fxd | 7 | 90 | 330 | 93 | 238 | – | – |

Slide credit: Bryan Russell & Josef Sivic

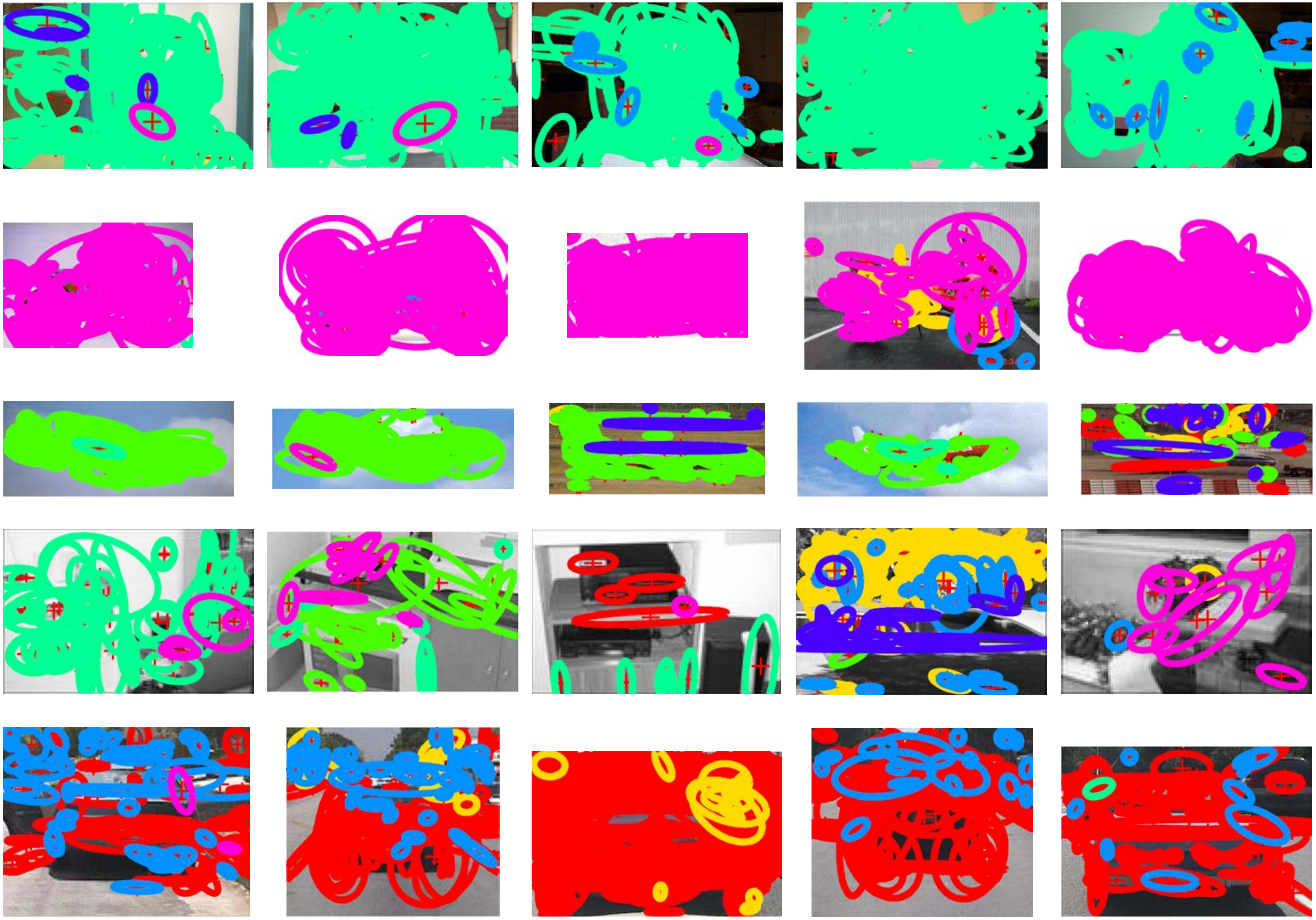
# Image as a mixture of topics (objects)



Slide credit: Bryan Russell & Josef Sivic




Slide credit: Bryan Russell & Josef Sivic



Slide credit: Bryan Russell & Josef Sivic

# Beyond single classes

- Multiclass
- Multiview
- Datasets

# Beyond single classes

- **Multiclass**
- Multiview
- Datasets

# Shared features

- Is learning the object class 1000 easier than learning the first?

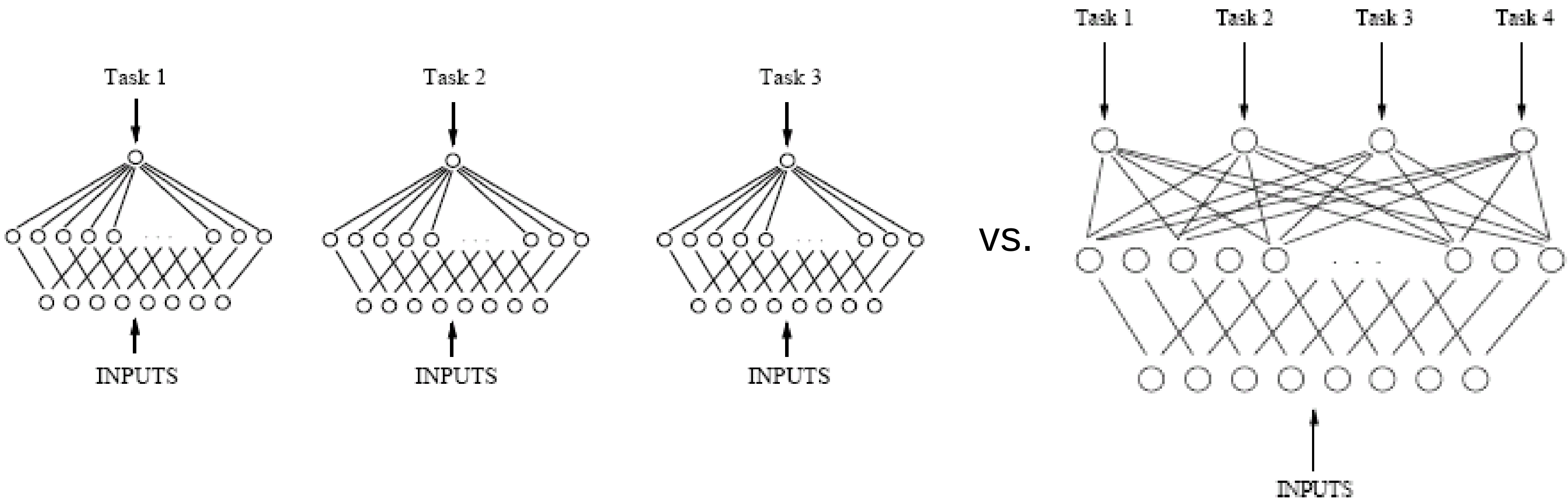
  … 

- Can we transfer knowledge from one object to another?
- Are the shared properties interesting by themselves?

# Multitask learning

**R. Caruana. Multitask Learning. ML 1997**

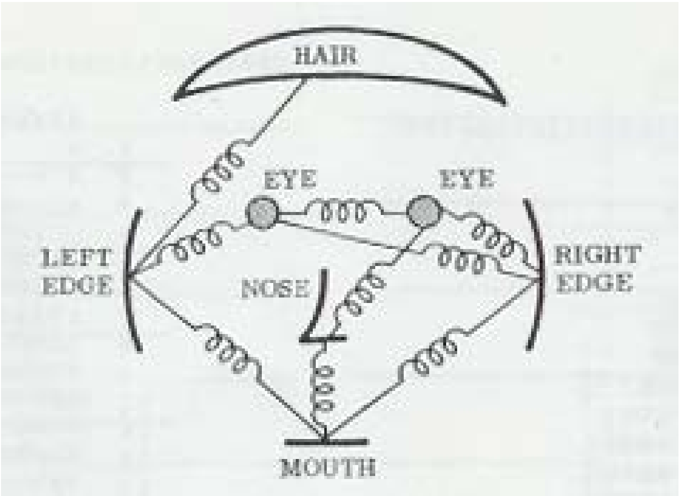
“MTL improves generalization by leveraging the domain-specific information contained in the training signals of *related* tasks. It does this by training tasks in parallel while using a shared representation”.



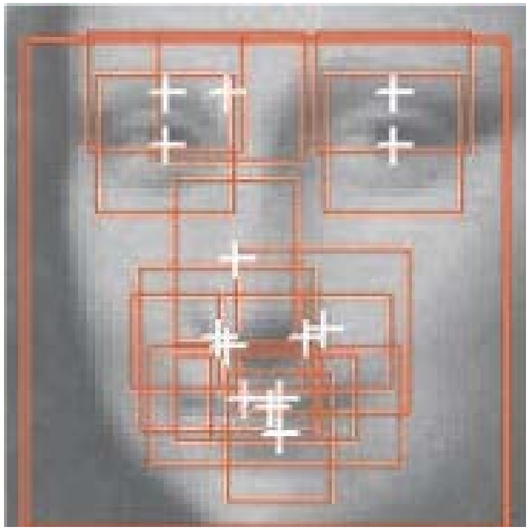

Sejnowski & Rosenberg 1986; Hinton 1986; Le Cun et al. 1989; Suddarth & Kergosien 1990; Pratt et al. 1991; Sharkey & Sharkey 1992; …

# Sharing in constellation models

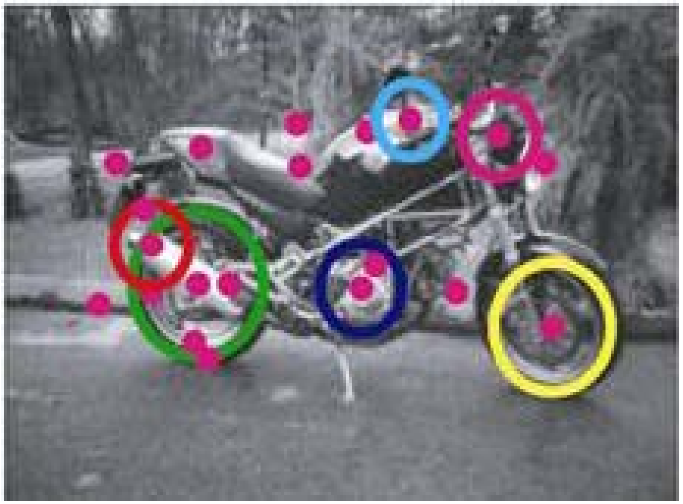
(next Wednesday)



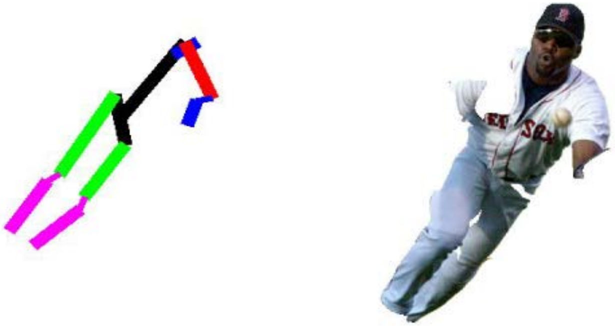

**Pictorial Structures**
*Fischler & Elschlager, IEEE Trans. Comp. 1973*



**SVM Detectors**
*Heisele, Poggio, et. al., NIPS 2001*



**Constellation Model**
*Fergus, Perona, & Zisserman, CVPR 2003*

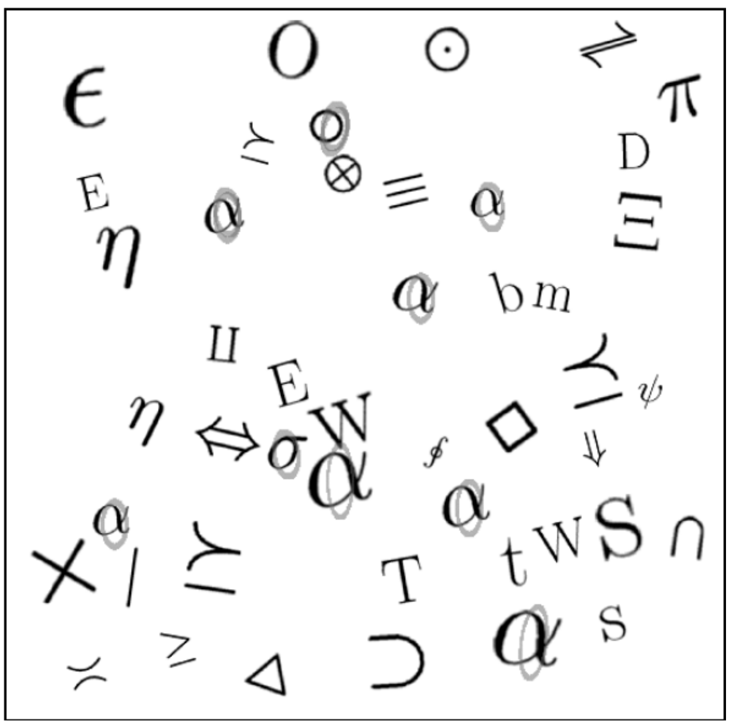


**Model-Guided Segmentation**
*Mori, Ren, Efros, & Malik, CVPR 2004*

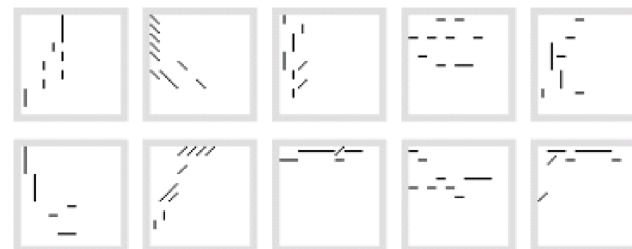
# Reusable Parts

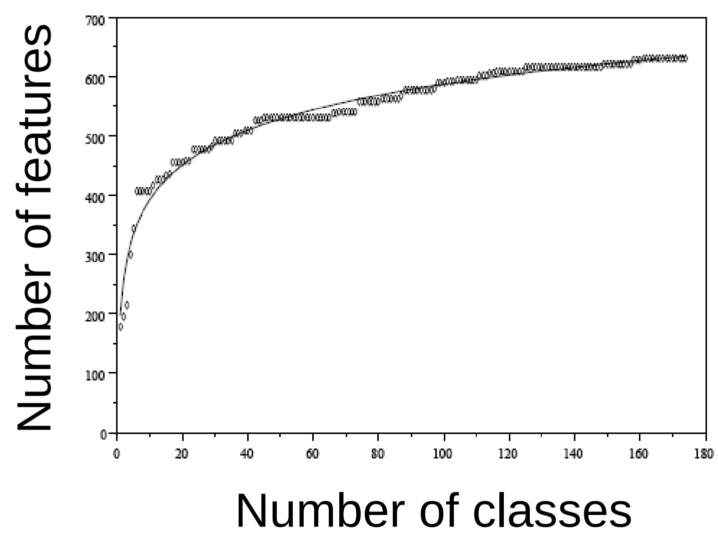
**Krempp, Geman, & Amit “Sequential Learning of Reusable Parts for Object Detection”. TR 2002**

Goal: Look for a vocabulary of edges that reduces the number of features.



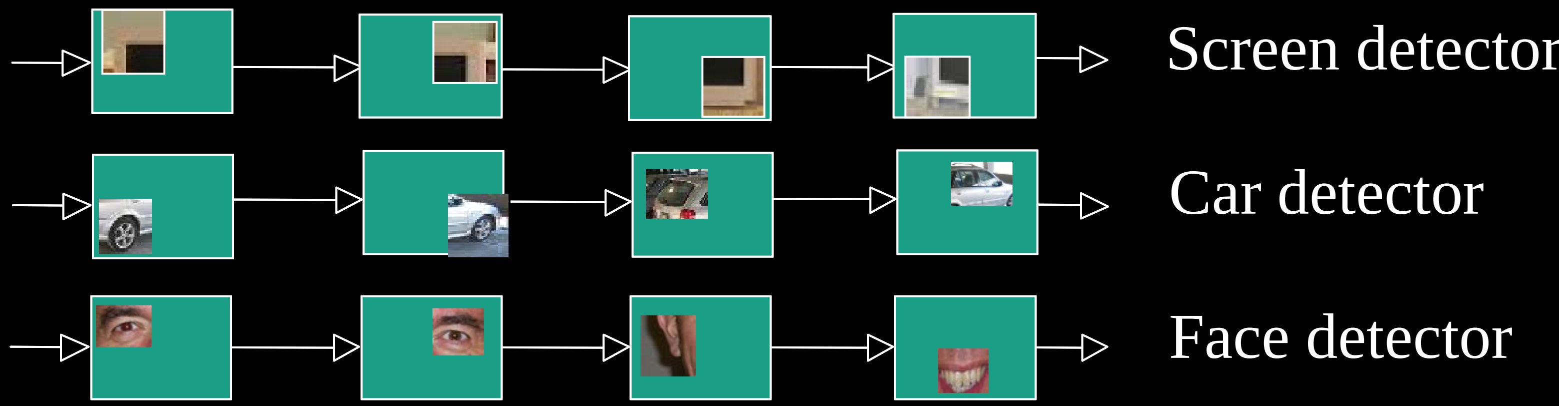
Examples of reused parts



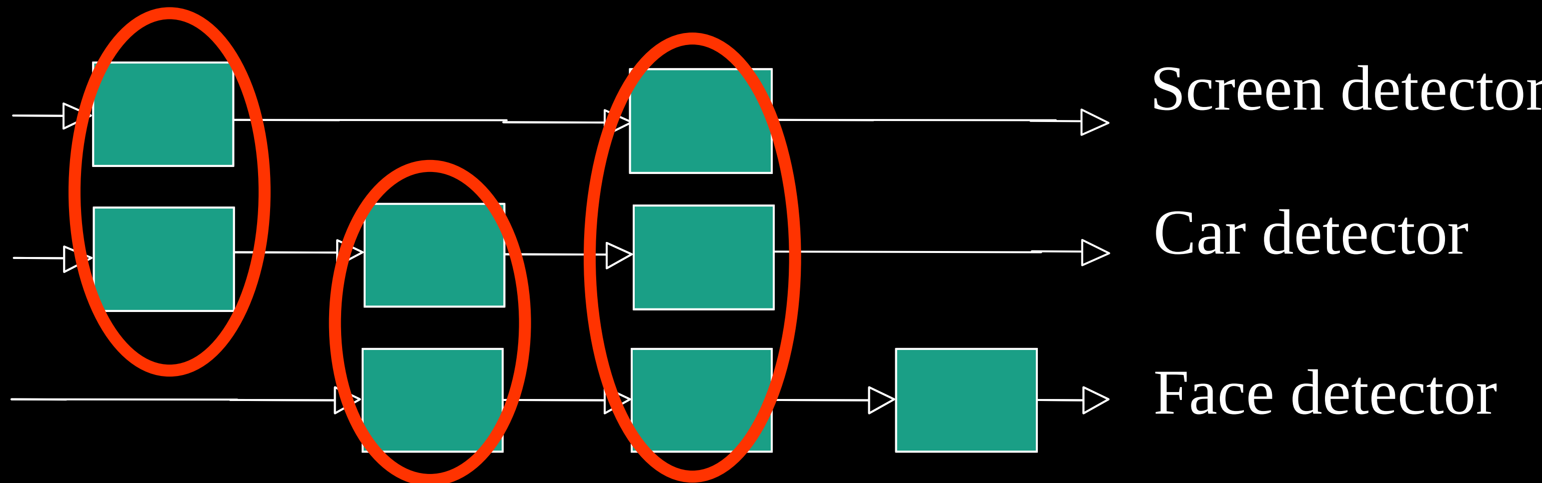


# Additive models and boosting

- Independent binary classifiers:



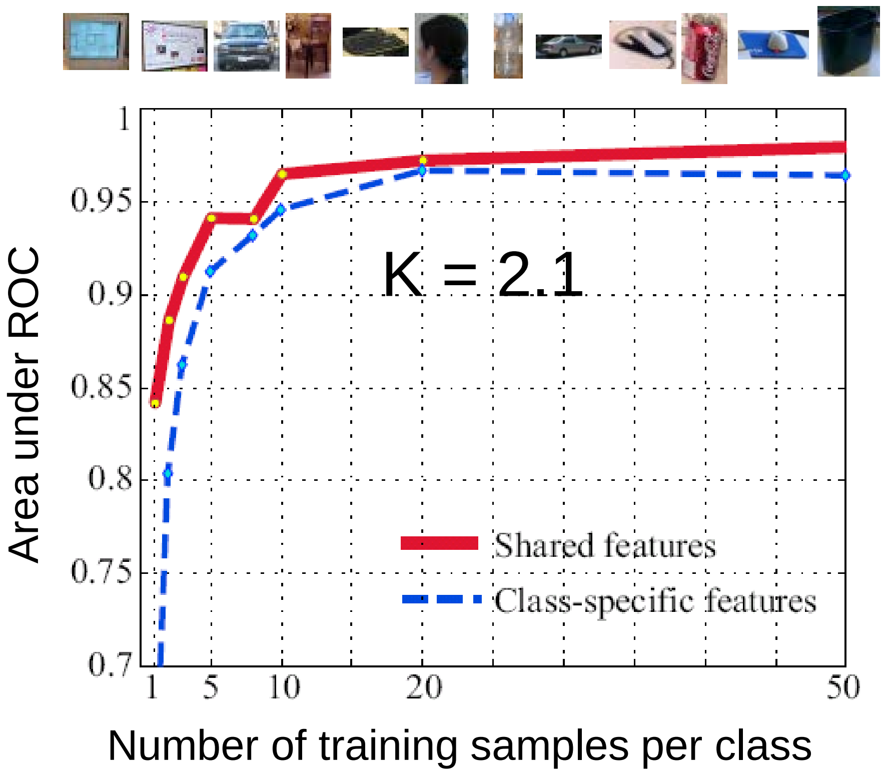

- Binary classifiers that share features:



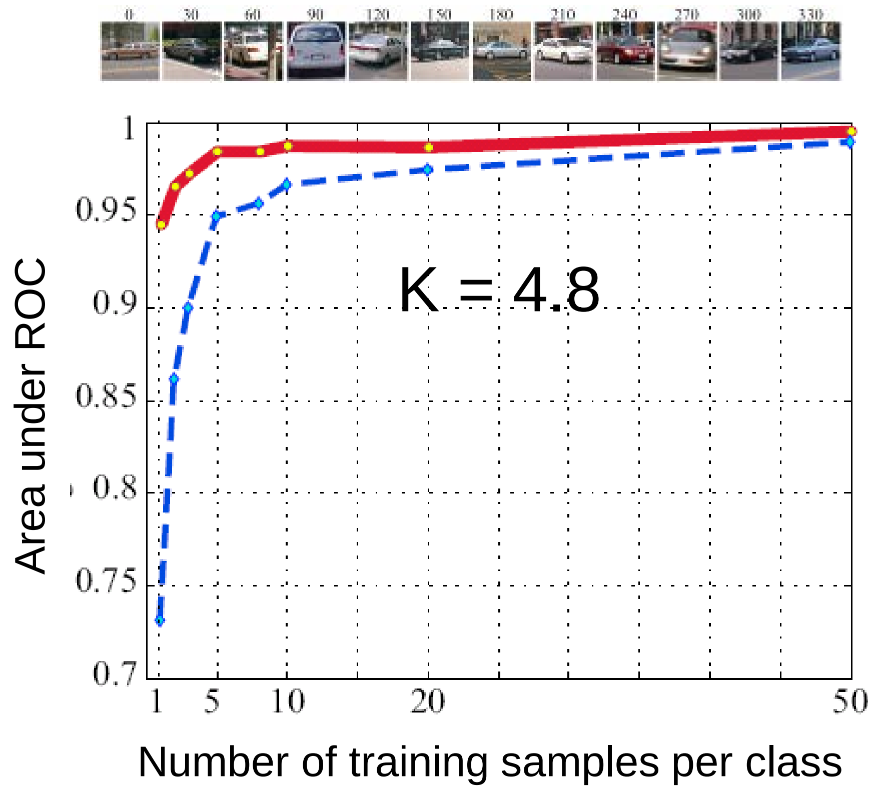

Torralba, Murphy, Freeman. CVPR 2004. PAMI 2007

# Generalization as a function of object similarities

12 unrelated object classes




12 viewpoints




Torralba, Murphy, Freeman. CVPR 2004. PAMI 2007

# Beyond single classes

- Multiclass
- **Multiview**
- Datasets

# 3D object categorization

Despite we can categorize all three pictures as being views of a horse, the three pictures do not look as being equally typical views of horses. And they do not seem to be recognizable with the same easiness.

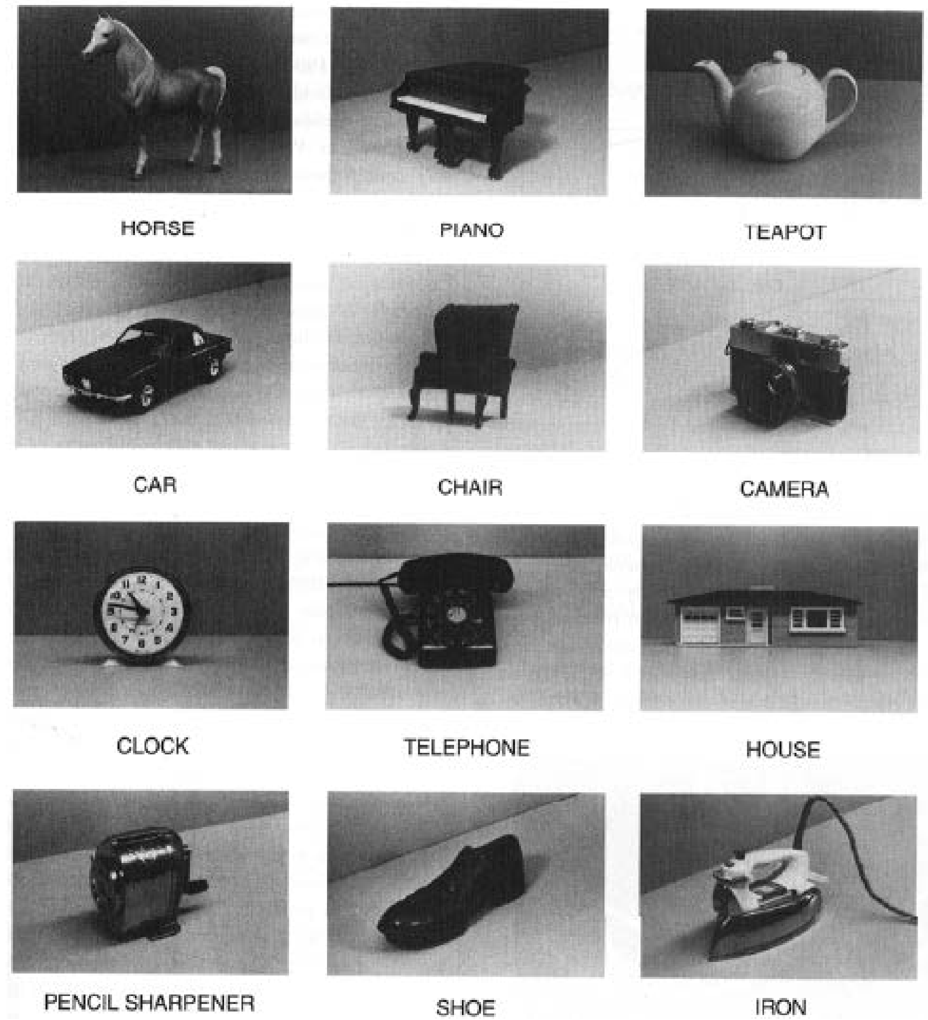

# Canonical Perspective

**Experiment** (Palmer, Rosch & Chase 81): participants are shown views of an object and are asked to rate “how much each one looked like the objects they depict” (scale; 1=very much like, 7=very unlike)

In a recognition task, reaction time correlated with the ratings.

Canonical views are recognized faster at the entry level.

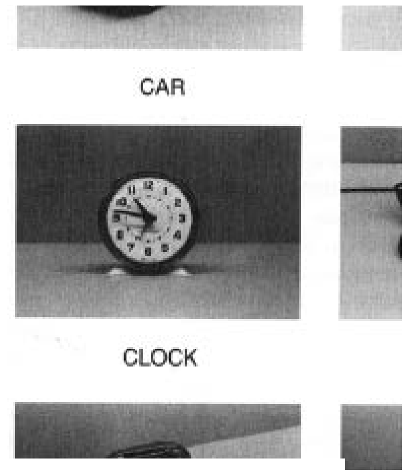
Examples of canonical perspective:

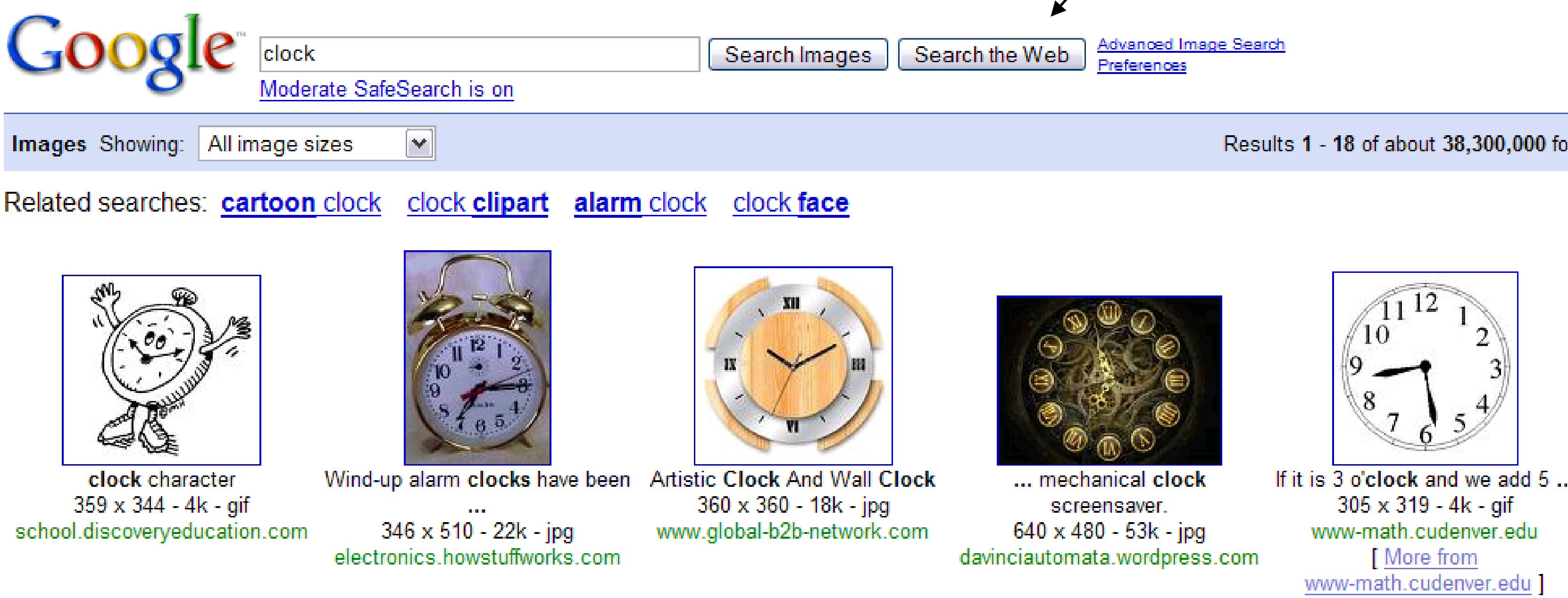



From *Vision Science*, Palmer

# Canonical Viewpoint

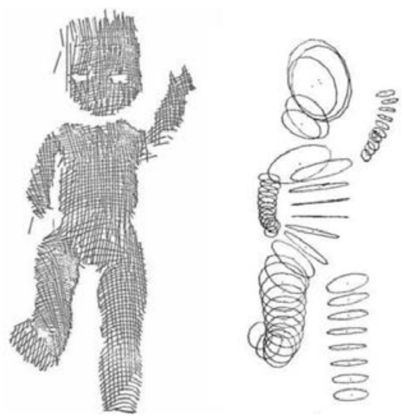
Clocks are preferred as purely frontal



# Object representations

**Explicit 3D models**: use volumetric representation. Have an explicit model of the 3D geometry of the object.



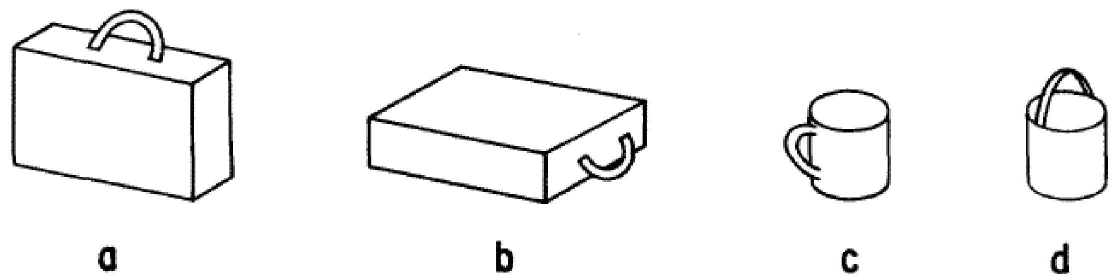



*Figure 3.* Different arrangements of the same components can produce different objects.

Appealing but hard to get it to work…

# Object representations

**Implicit 3D models**: matching the input 2D view to view-specific representations.

(b) For cars, classifiers are trained on 8 viewpoints

Not very appealing but somewhat easy to get it to work…

# Beyond single classes

- Multiclass
- Multiview
- **Datasets**

# The PASCAL Visual Object Classes

In 2007, the twenty object classes that have been selected are:

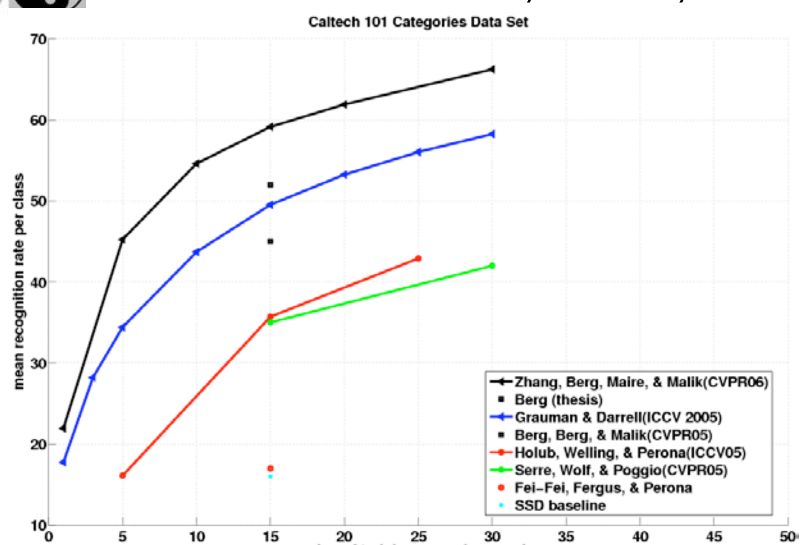
*Person:* person
*Animal:* bird, cat, cow, dog, horse, sheep
*Vehicle:* aeroplane, bicycle, boat, bus, car, motorbike, train
*Indoor:* bottle, chair, dining table, potted plant, sofa, tv/monitor

M. Everingham, Luc van Gool , C. Williams, J. Winn, A. Zisserman 2007

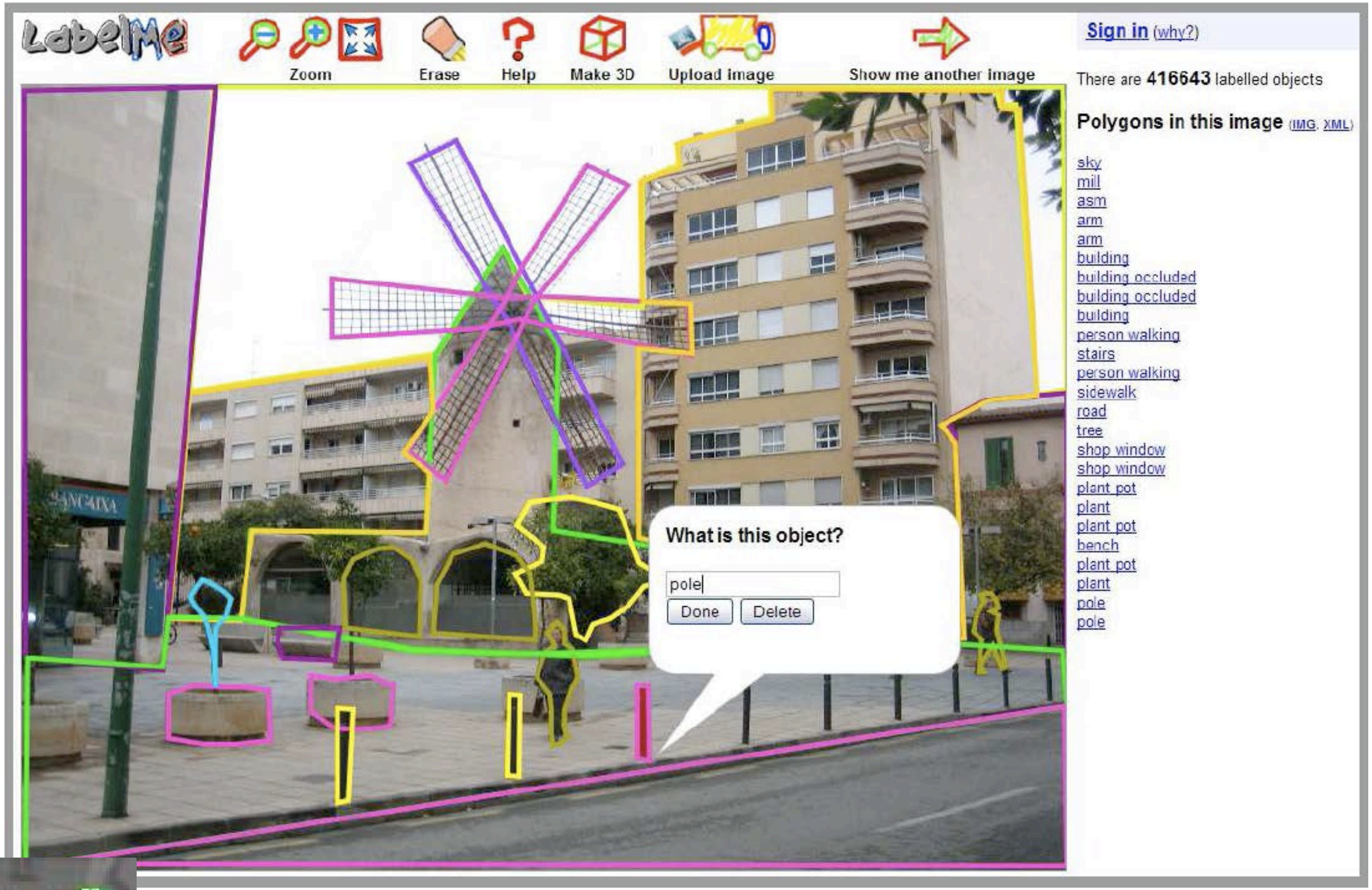
# Caltech 101 and 256


Fei-Fei, Fergus, Perona, 2004


Griffin, Holub, Perona, 2007



# LabelMe

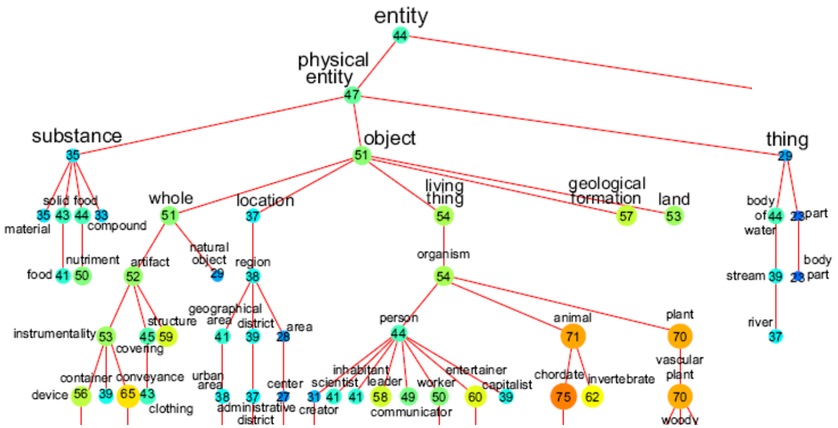



**Tool went online July 1st, 2005**
**530,000 object annotations collected**
Labelme.csail.mit.edu

B.C. Russell, A. Torralba, K.P. Murphy, W.T. Freeman, IJCV 2008

# 80.000.000 images

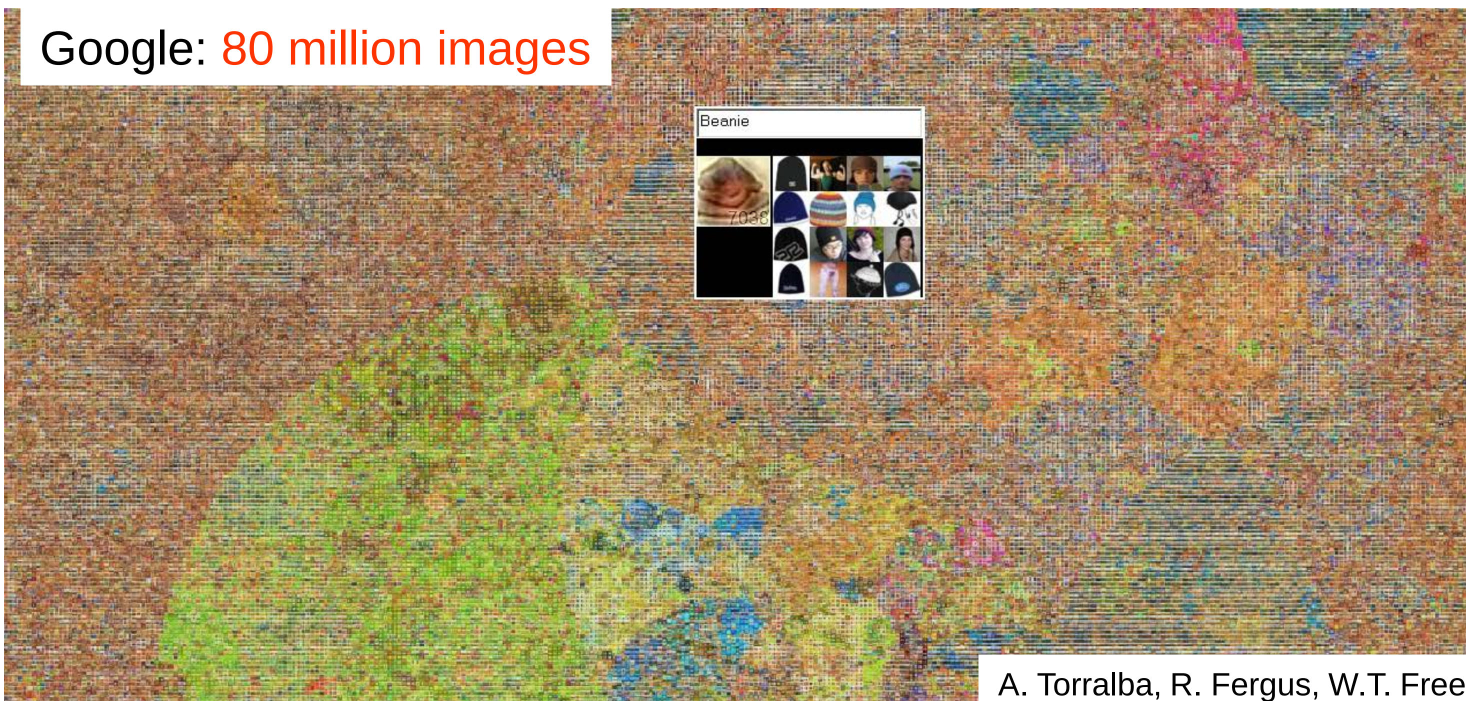
75.000 non-abstract nouns from WordNet




7 Online image search engines

And after 1 year downloading images

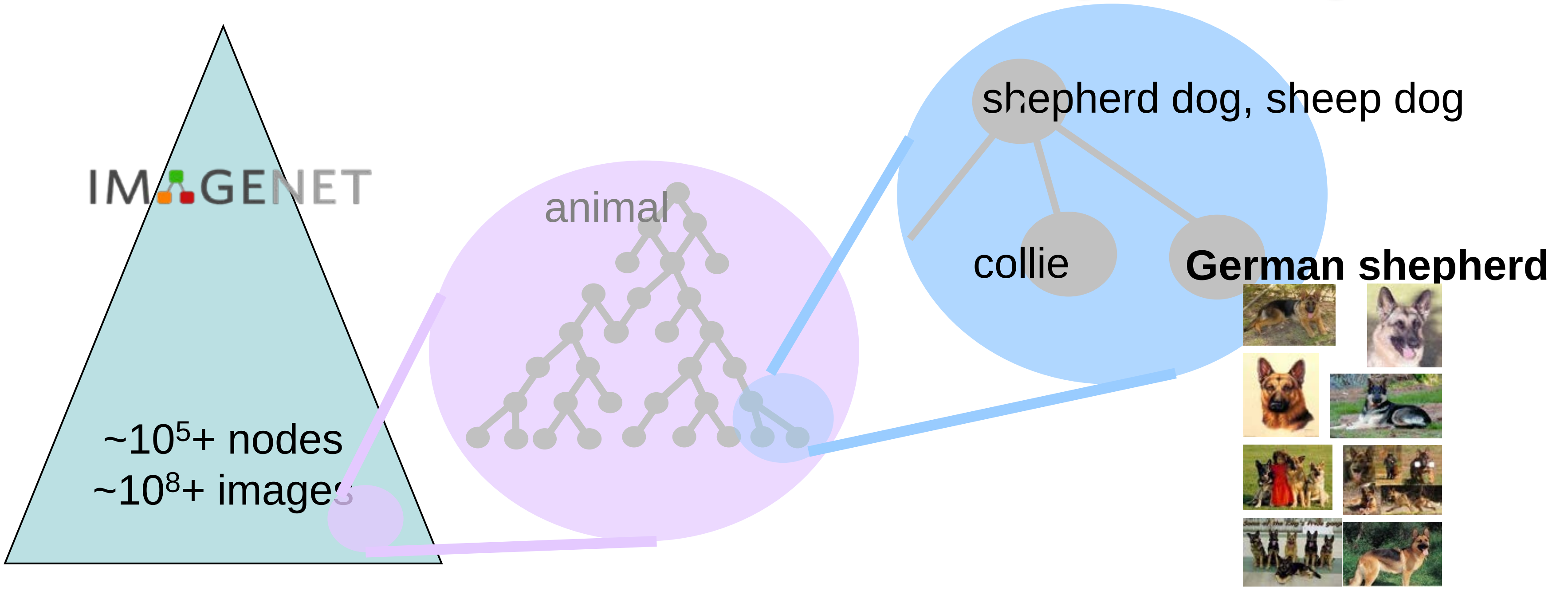
Google: 80 million images




A. Torralba, R. Fergus, W.T. Freeman. PAMI 2008

# IMAGENET

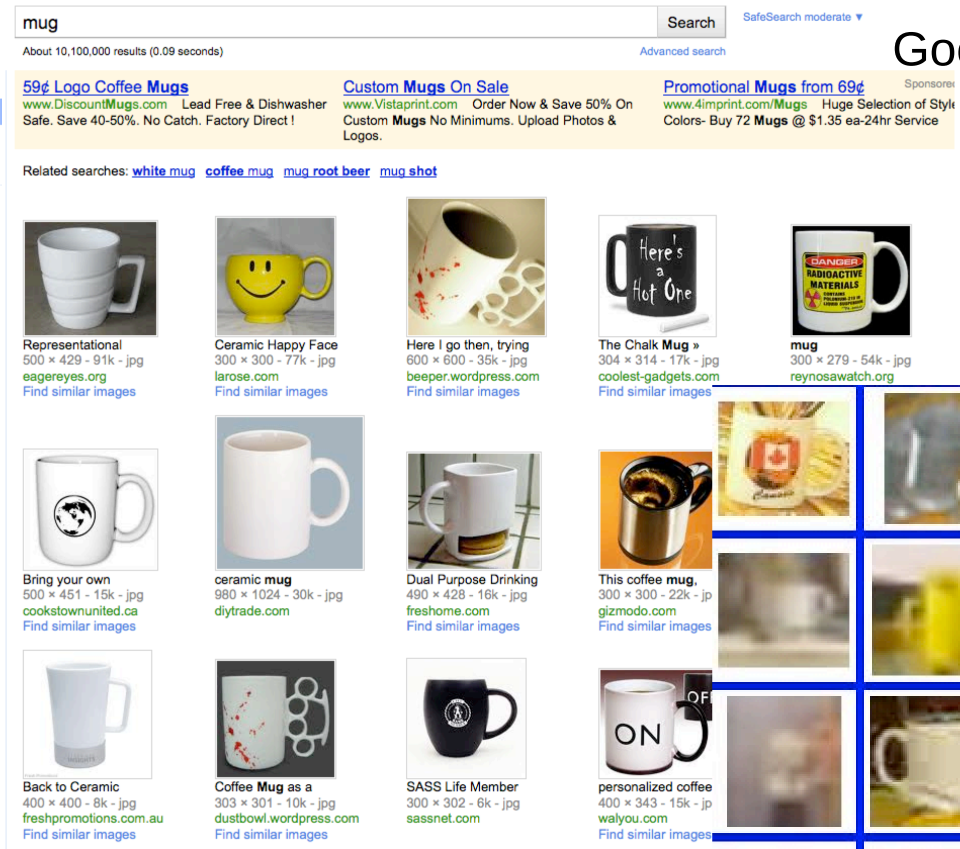
- An ontology of images based on WordNet
- ImageNet currently has
  - 13,000+ categories of visual concepts
  - 10 million human-cleaned images (~700im/categ)
  - 1/3+ is released online @ **www.image-net.org**




Deng, Dong, Socher, Li & Fei-Fei, CVPR 2009

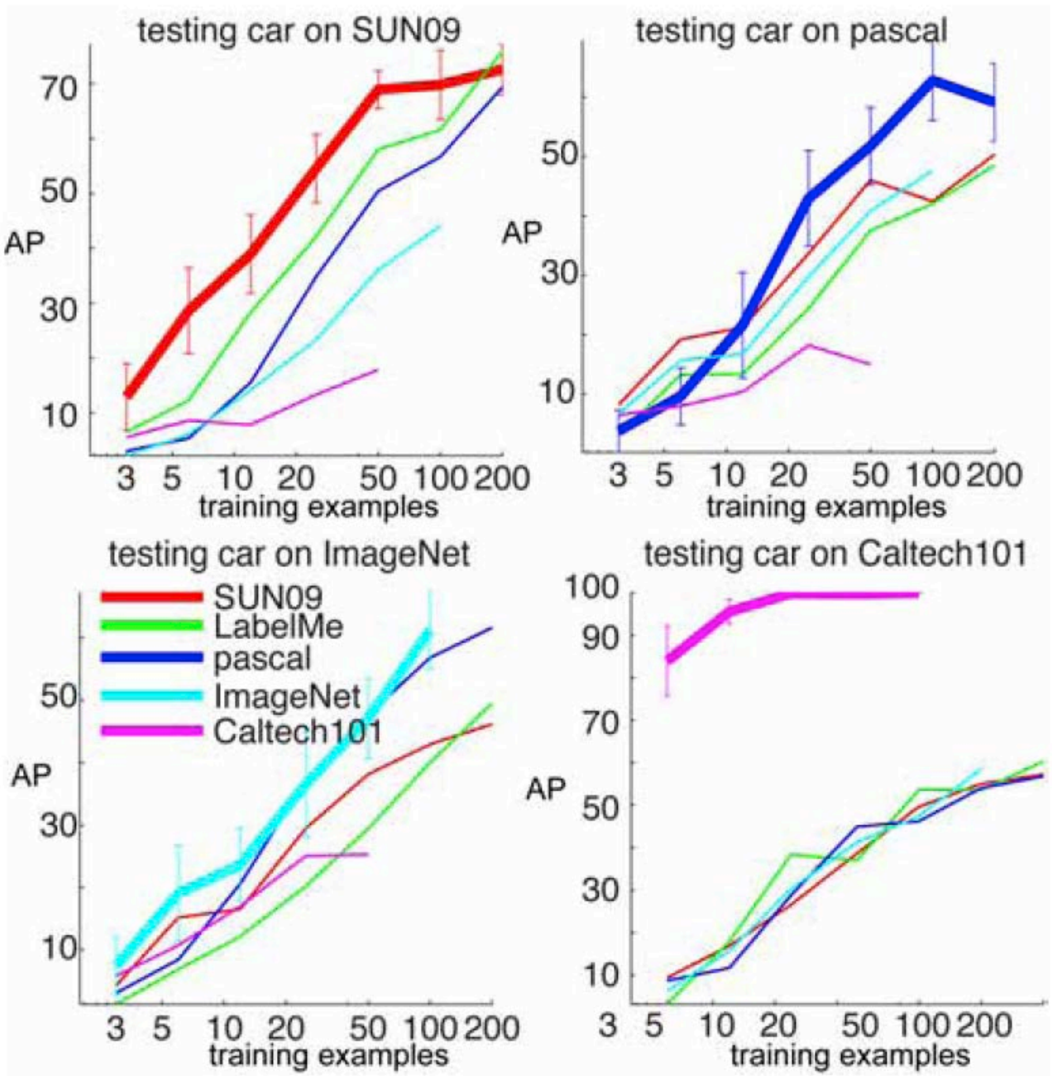
# Dataset biases

Google mugs




Mugs from LabelMe

# Dataset biases




Torralba, Efros. Unbiased Look at Dataset Bias. CVPR 2011