Introduction to Computer Vision

Michael J. Black

Object Recognition

News

- Last class on Wednesday.
- Attendance will be taken.
- Course evaluation(s).
- Reflections on vision.

Object categorization: the statistical viewpoint

$$p(zebra \mid image)$$

vs.

$$p(no\ zebra \mid image)$$

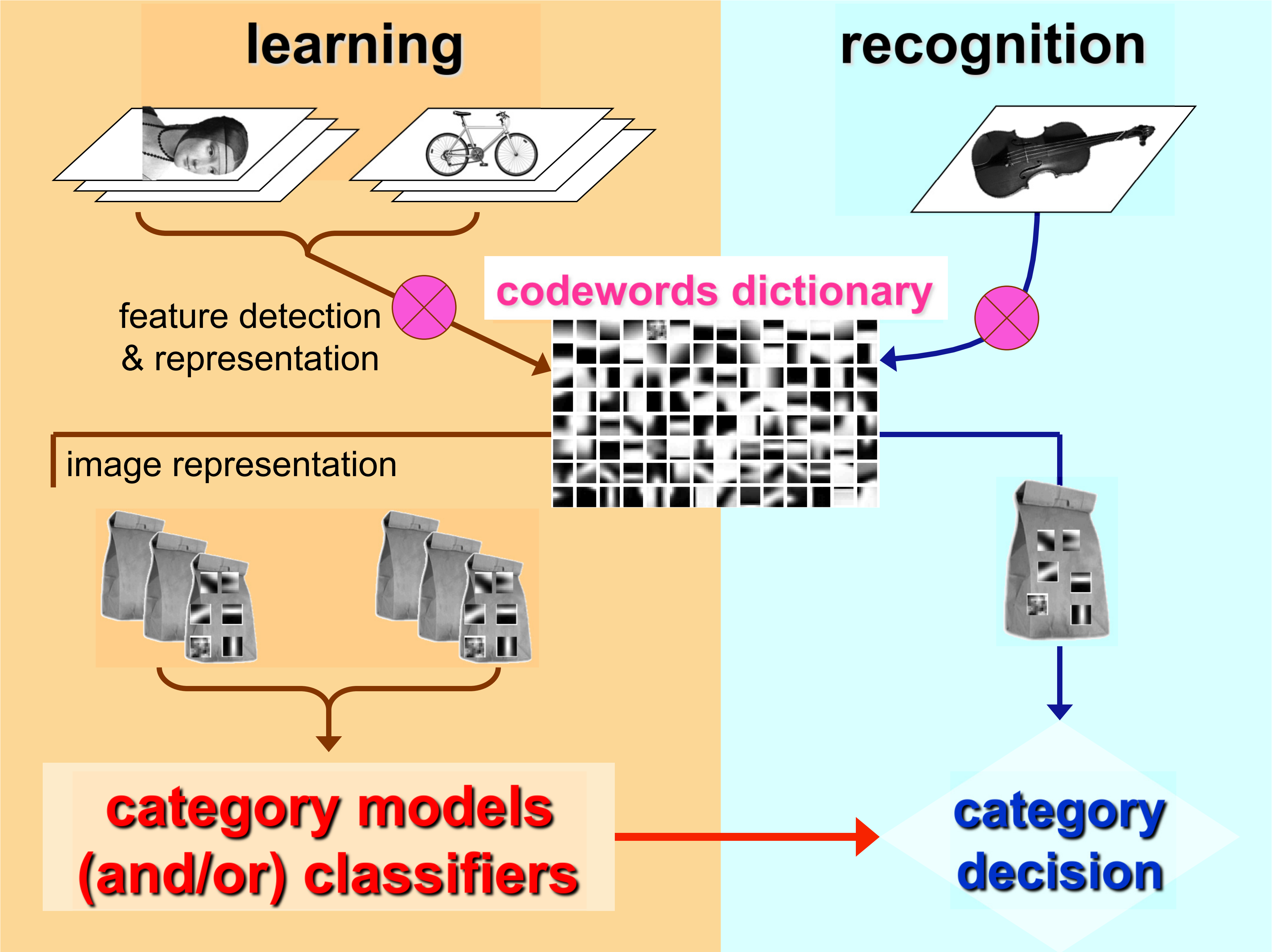
- Bayes rule:

$$\underbrace{\frac{p(zebra \mid image)}{p(no\ zebra \mid image)}}_{\text{posterior ratio}} = \underbrace{\frac{p(image \mid zebra)}{p(image \mid no\ zebra)}}_{\text{likelihood ratio}} \cdot \underbrace{\frac{p(zebra)}{p(no\ zebra)}}_{\text{prior ratio}}$$

Three main issues

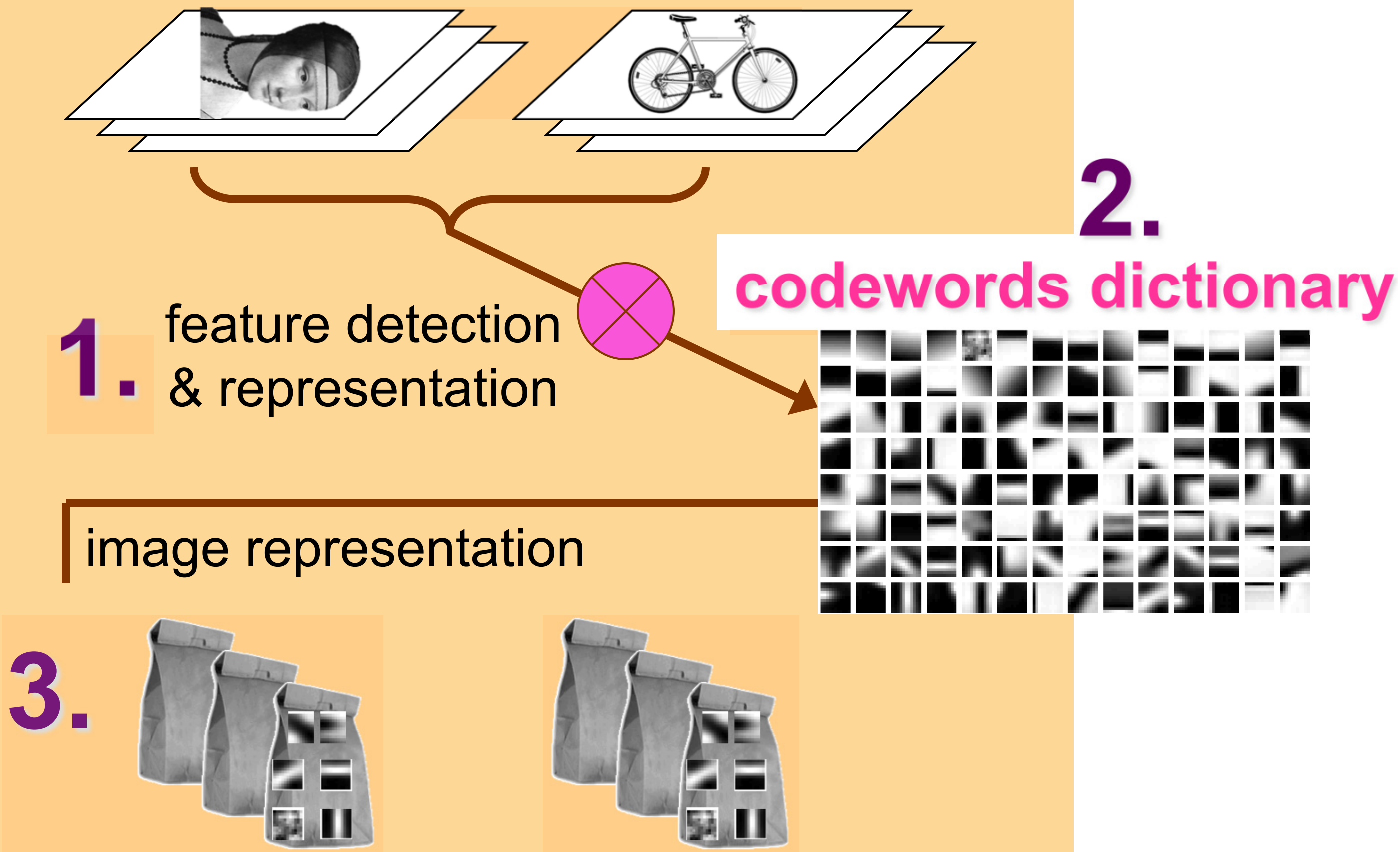
- Representation
  - How to represent an object category
- Learning
  - How to form the classifier, given training data
- Recognition
  - How the classifier is to be used on novel data

Object → **Bag of ‘words’**



learning
recognition
feature detection
& representation
codewords dictionary
image representation
category models
(and/or) classifiers
category
decision

Representation

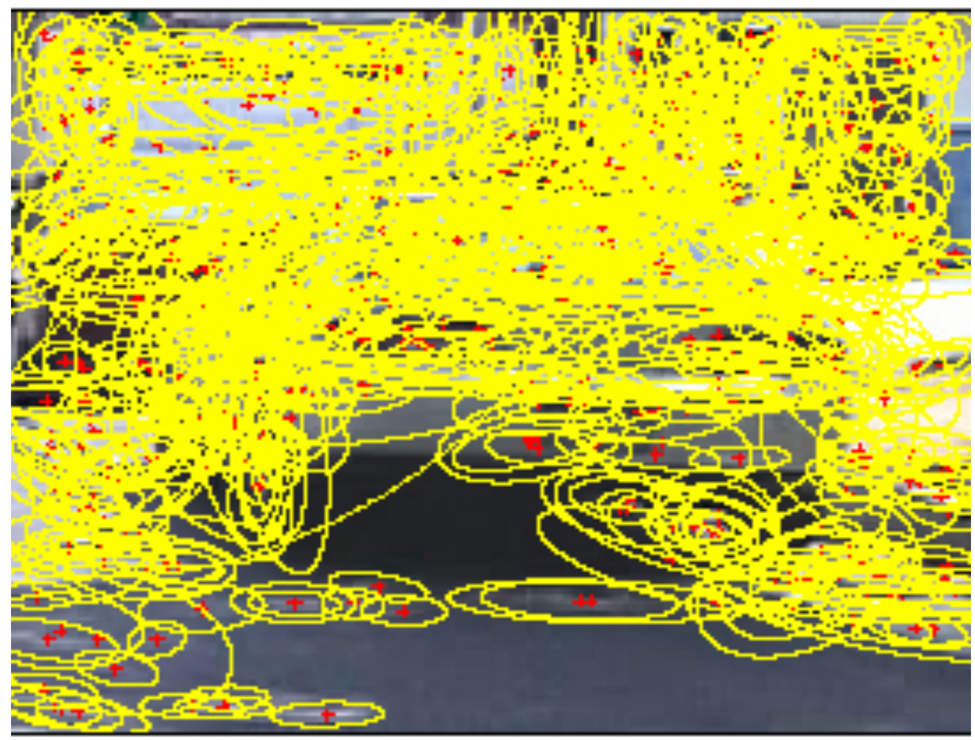


1.Feature detection and representation

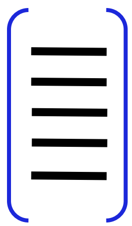
- Regular grid
  - Vogel & Schiele, 2003
  - Fei-Fei & Perona, 2005

1.Feature detection and representation

- Regular grid
  - Vogel & Schiele, 2003
  - Fei-Fei & Perona, 2005
- Interest point detector
  - Csurka, et al. 2004
  - Fei-Fei & Perona, 2005
  - Sivic, et al. 2005

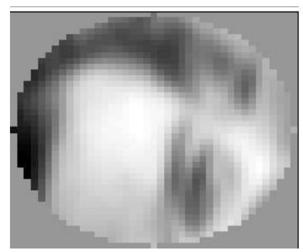


1.Feature detection and representation

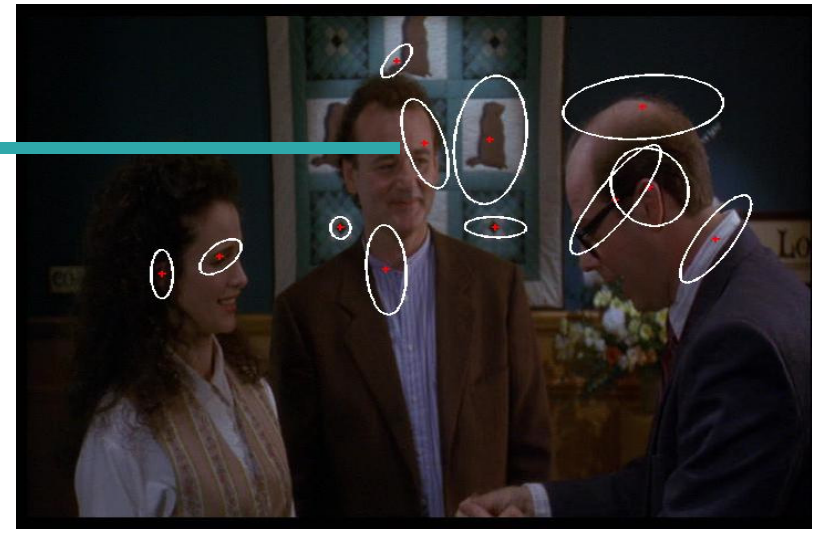


Compute SIFT descriptor

[Lowe'99]



Normalize patch



Detect patches

[Mikojaczyk and Schmid '02]

[Mata, Chum, Urban & Pajdla, '02]

[Sivic & Zisserman, '03]

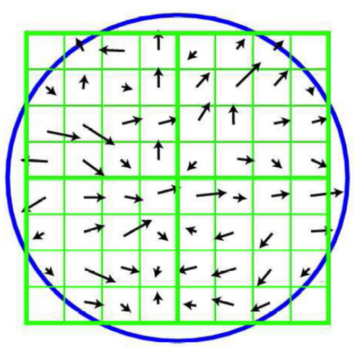
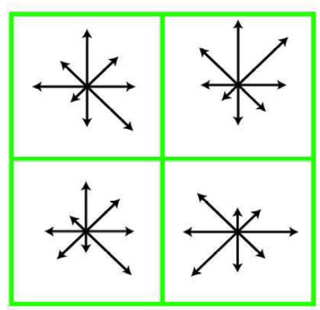


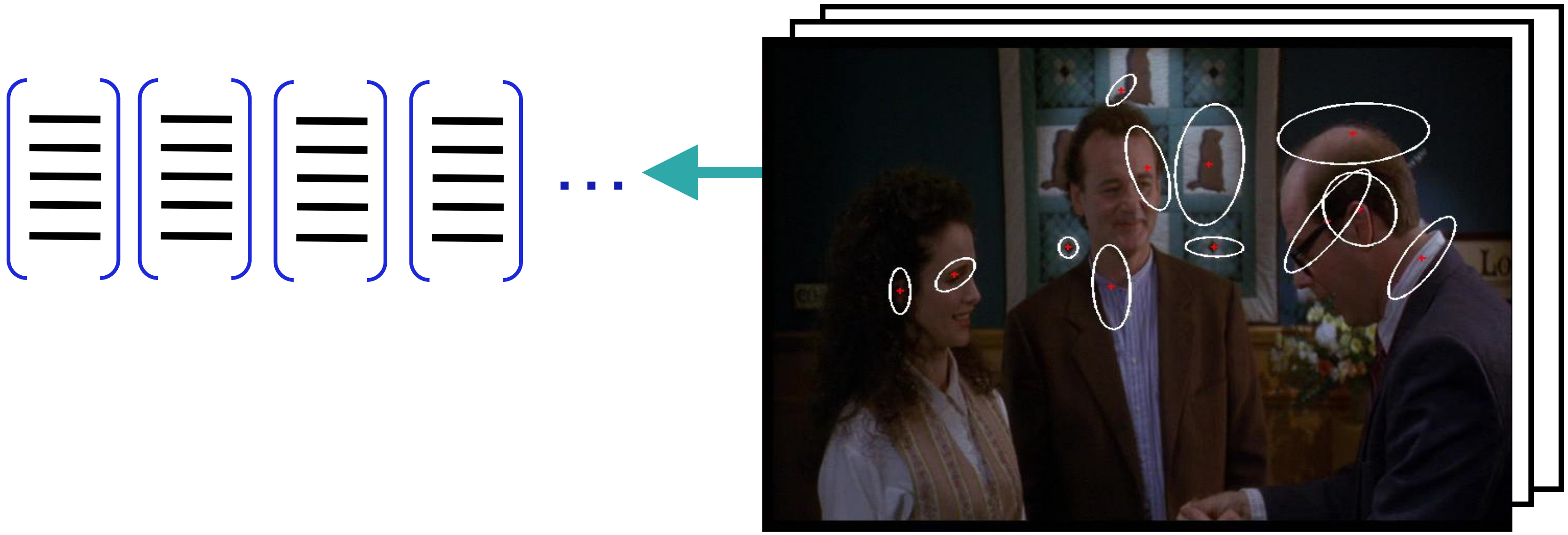
Image gradients



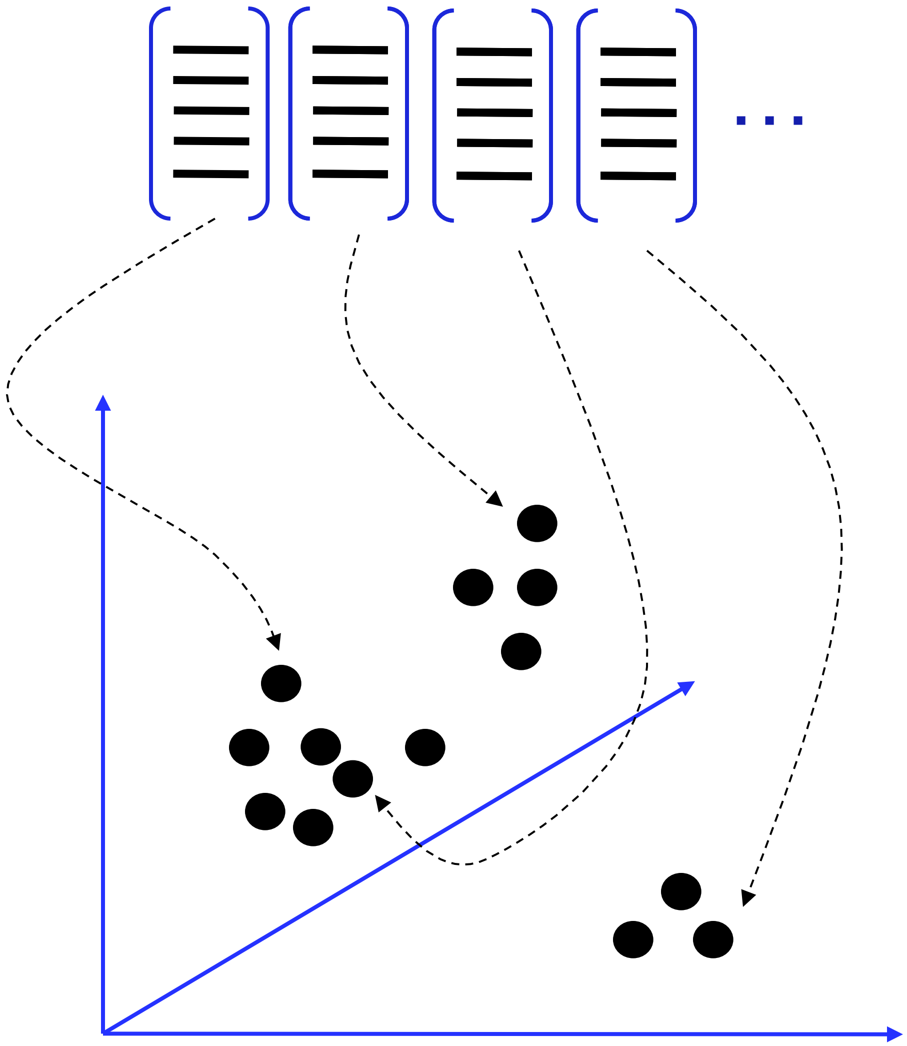
Keypoint descriptor

Slide credit: Josef Sivic

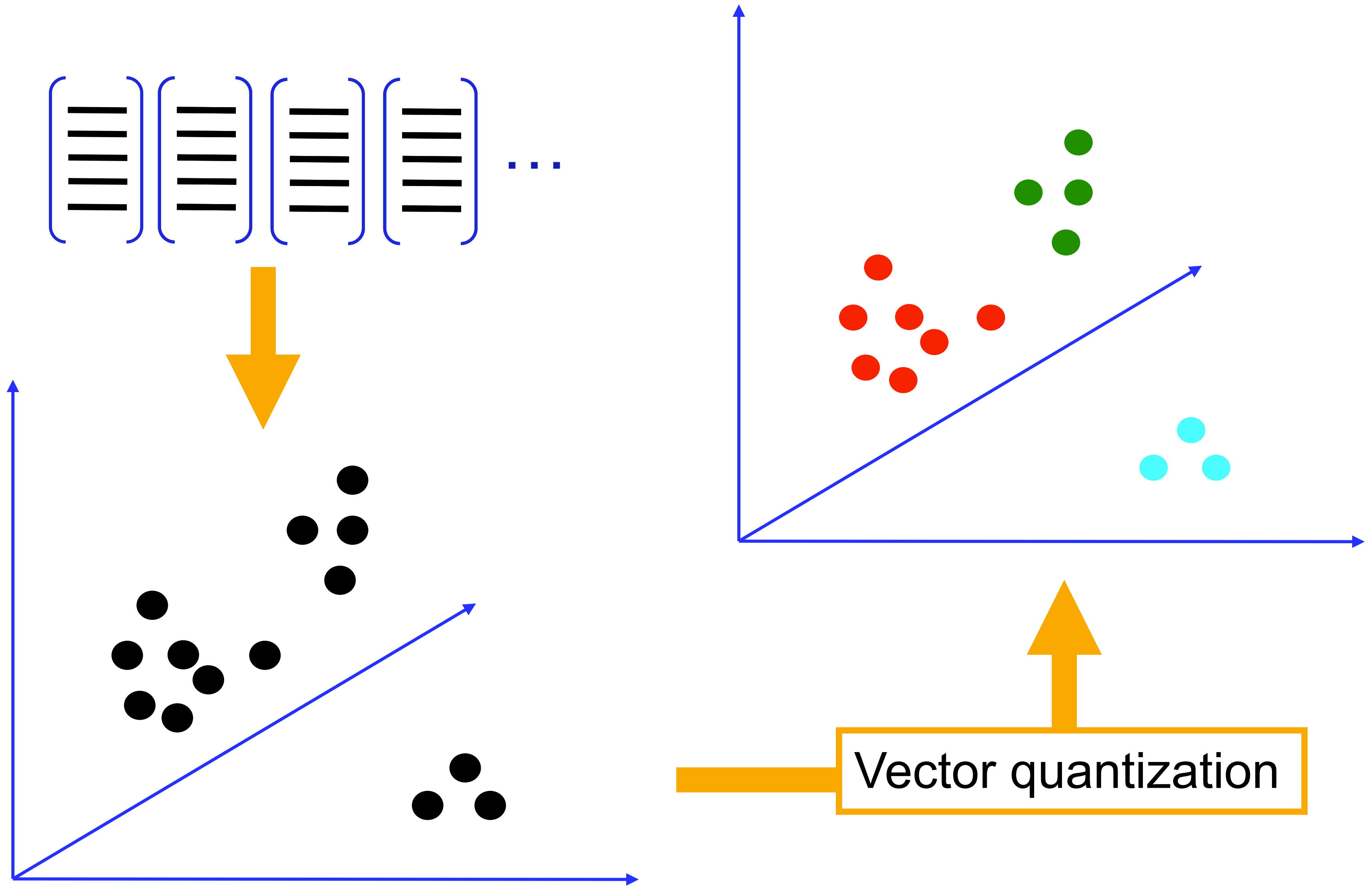
1.Feature detection and representation



2. Codewords dictionary formation

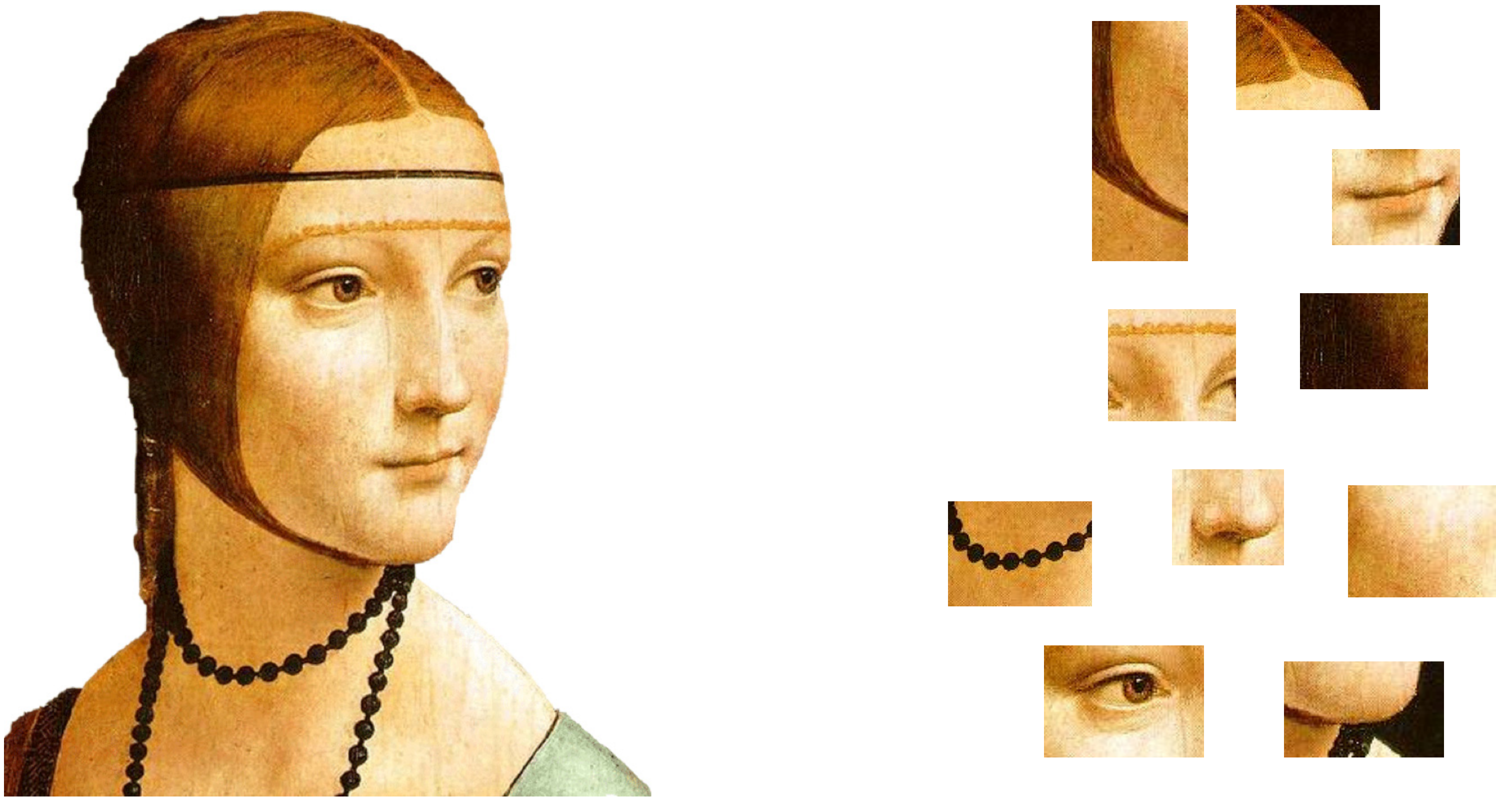


2. Codewords dictionary formation

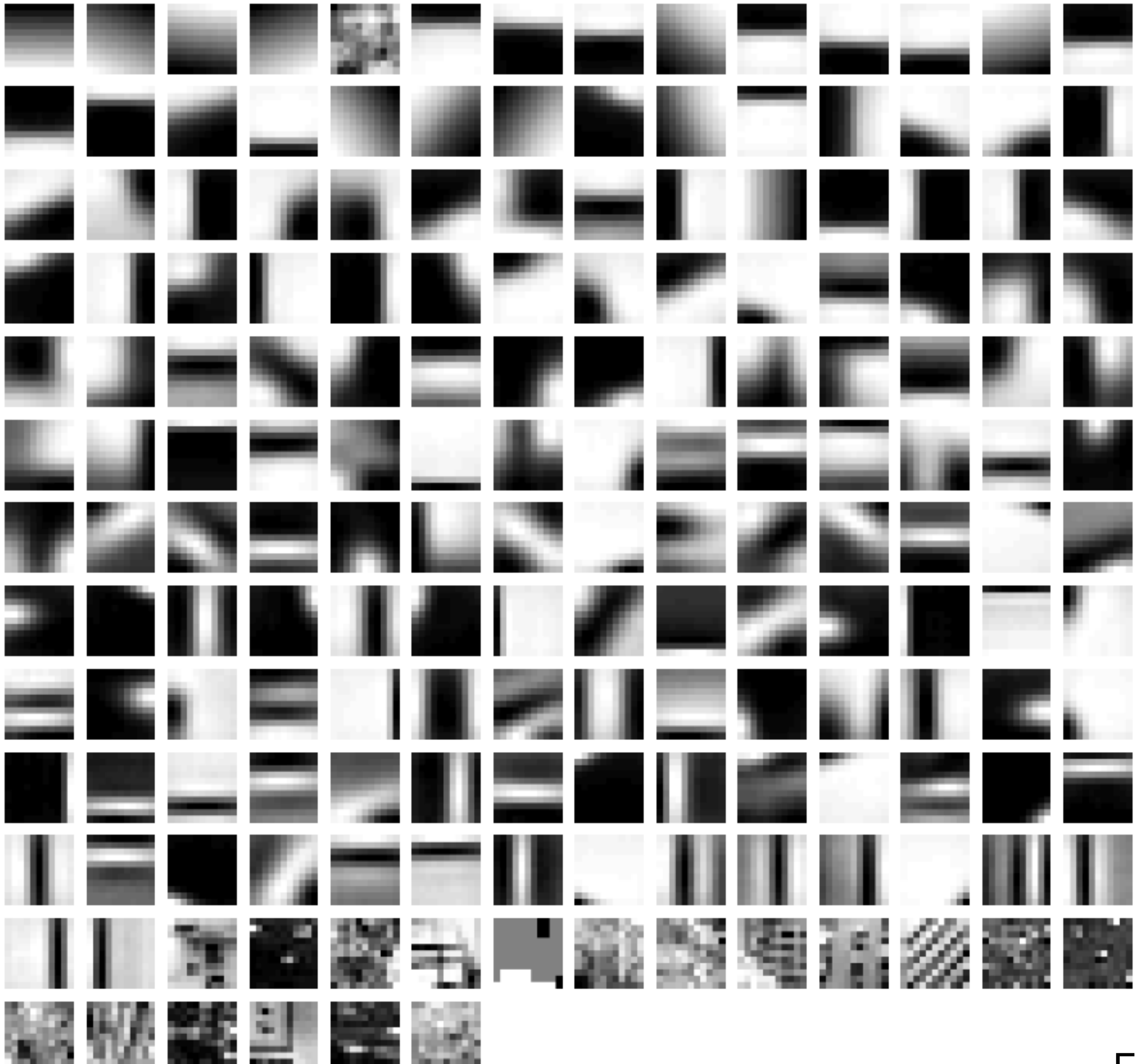


Slide credit: Josef Sivic

1.Feature detection and representation

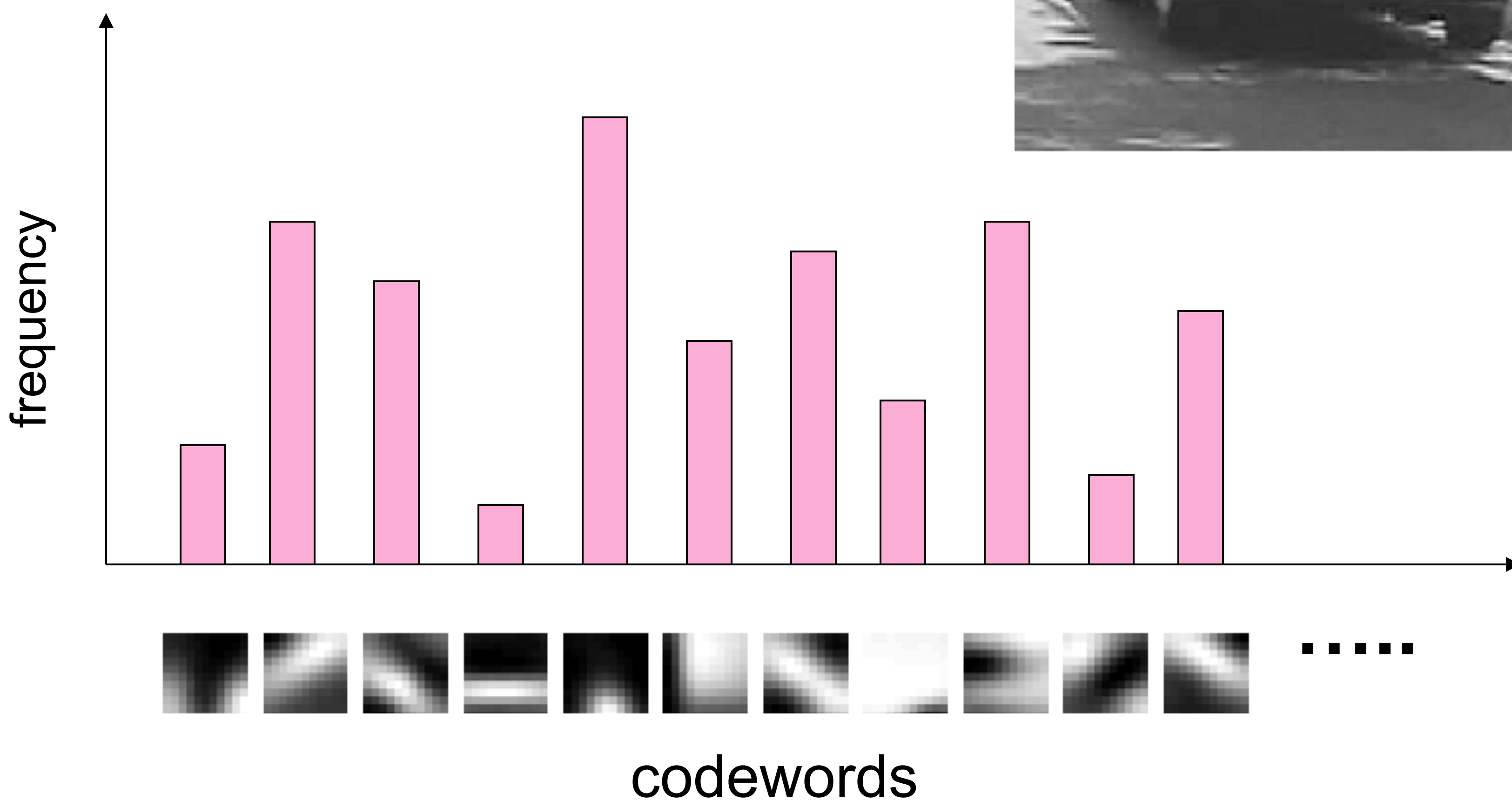


2. Codewords dictionary formation

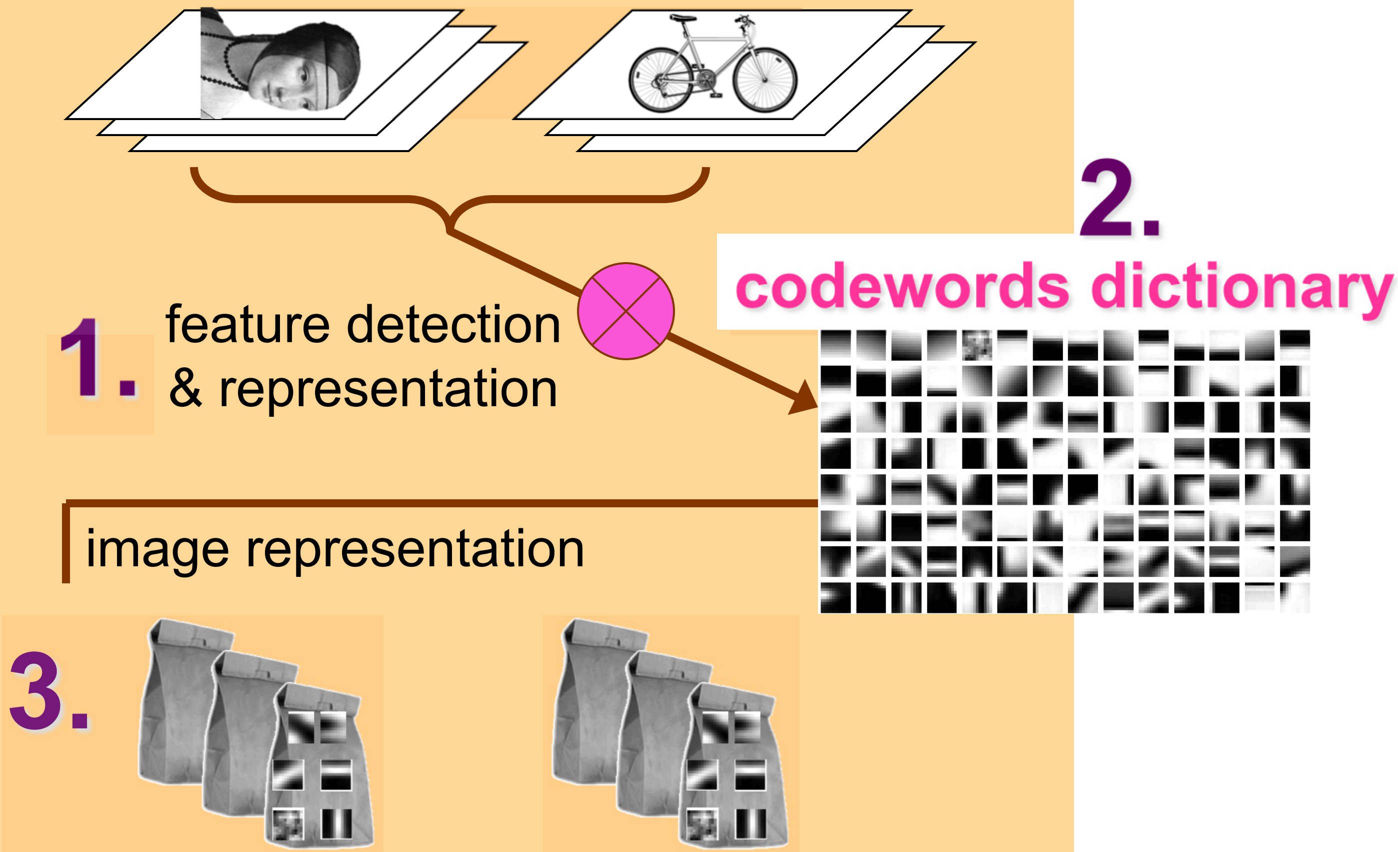


Fei-Fei et al. 2005

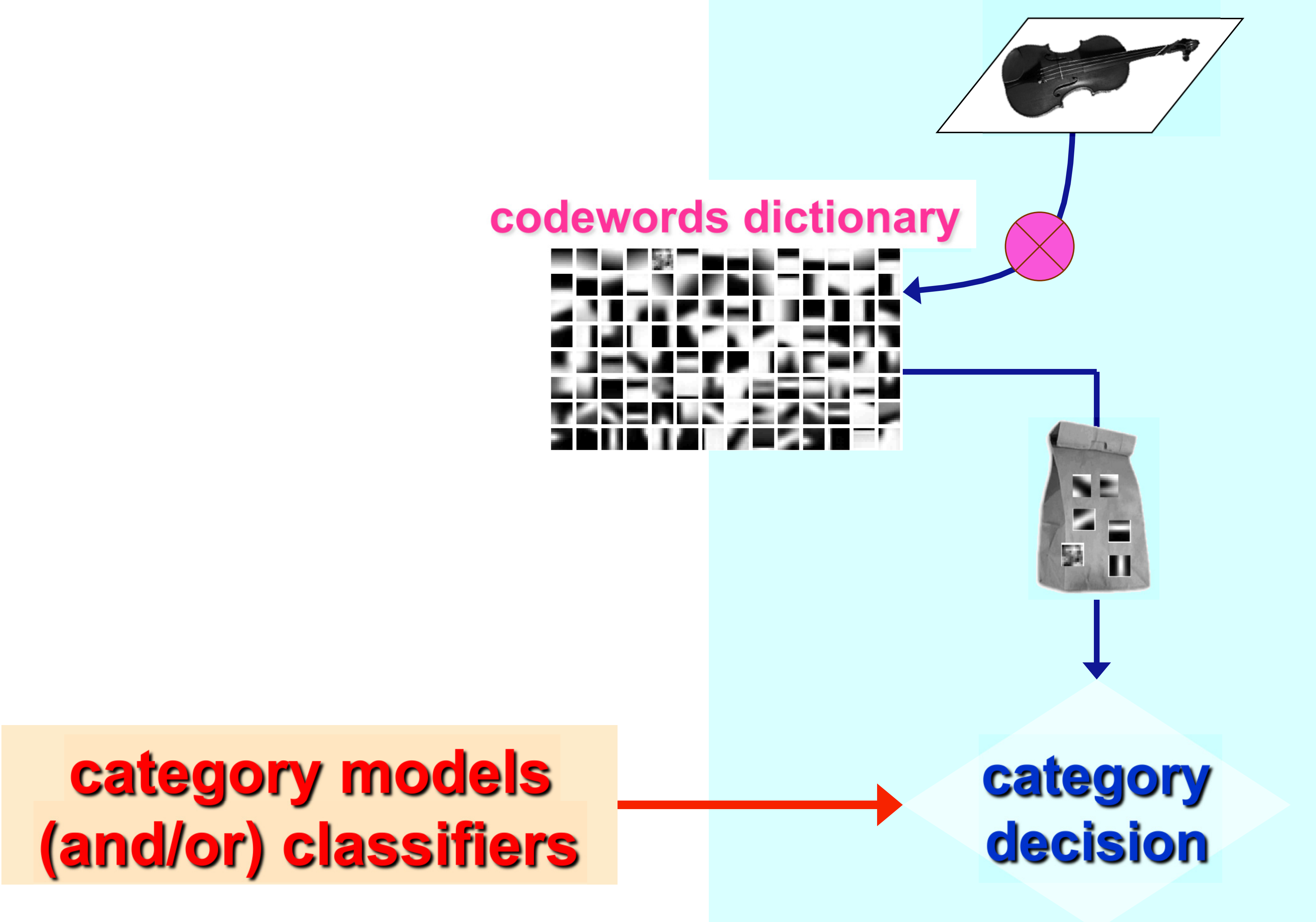
3. Image representation



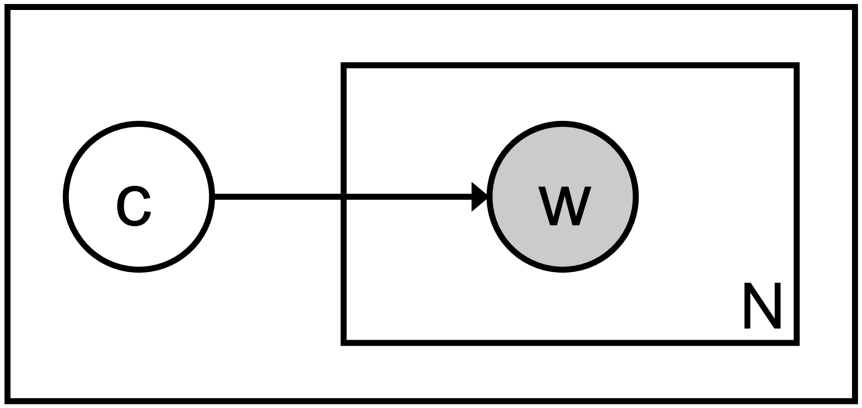
Representation

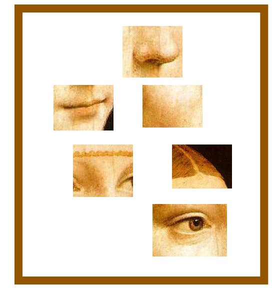


Learning and Recognition



Naïve Bayes model





$$c^* = \arg\max_c \; p(c \mid w) \propto p(c)\, p(w \mid c) = p(c) \prod_{n=1}^{N} p(w_n \mid c)$$

Object class decision

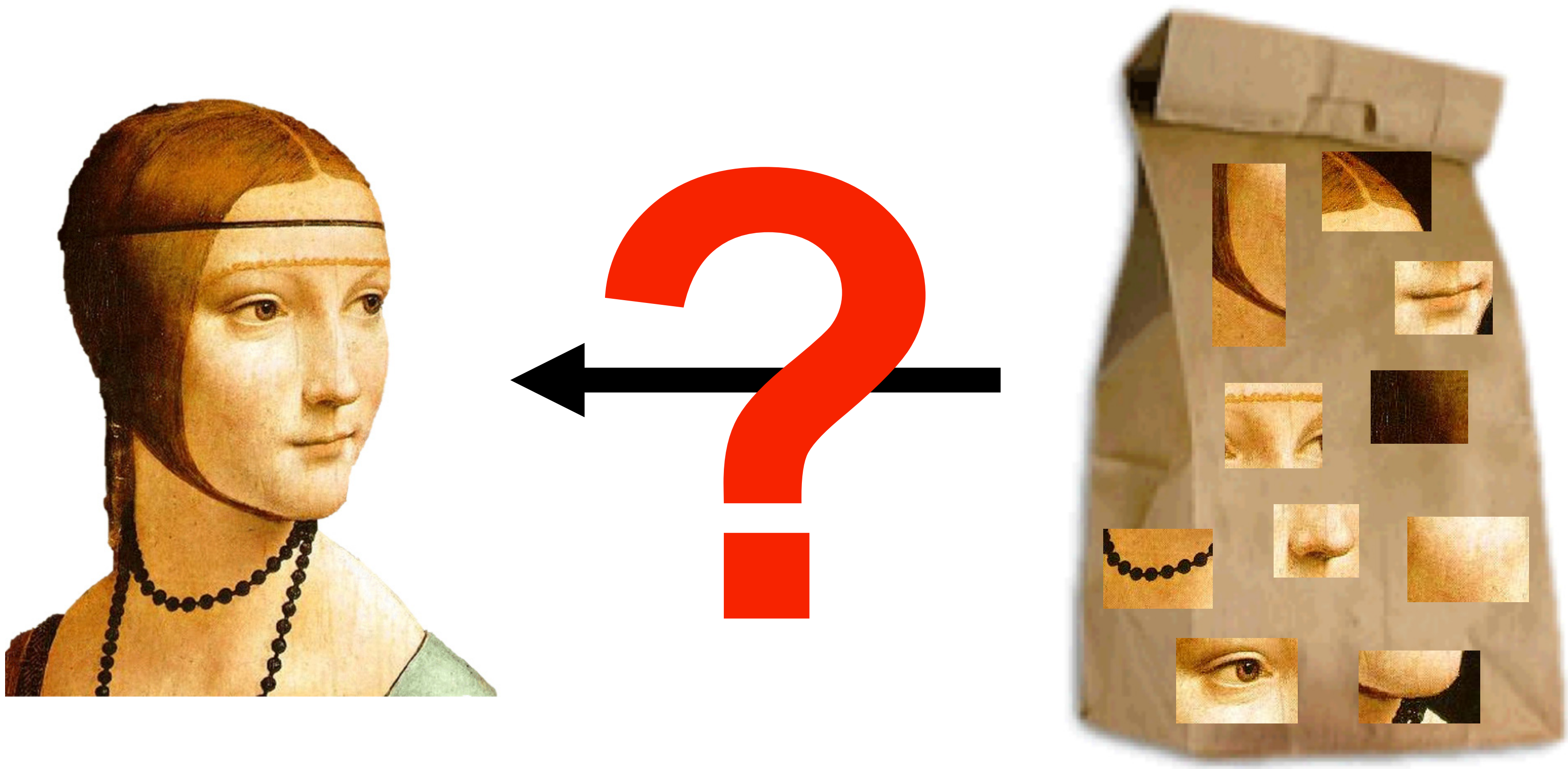
Prior prob. of the object classes

Image likelihood given the class

c: category of the image
w: patch in an image
N patches

Csurka et al. 2004

What about spatial info?



Problem with bag-of-words

- All have equal probability for bag-of-words methods
- Location information is important

Model: Parts and Structure

Representation

- Object as set of parts
  - Generative representation
- Model:
  - Relative locations between parts
  - Appearance of part
- Issues:
  - How to model location
  - How to represent appearance
  - Sparse or dense (pixels or regions)
  - How to handle occlusion/clutter

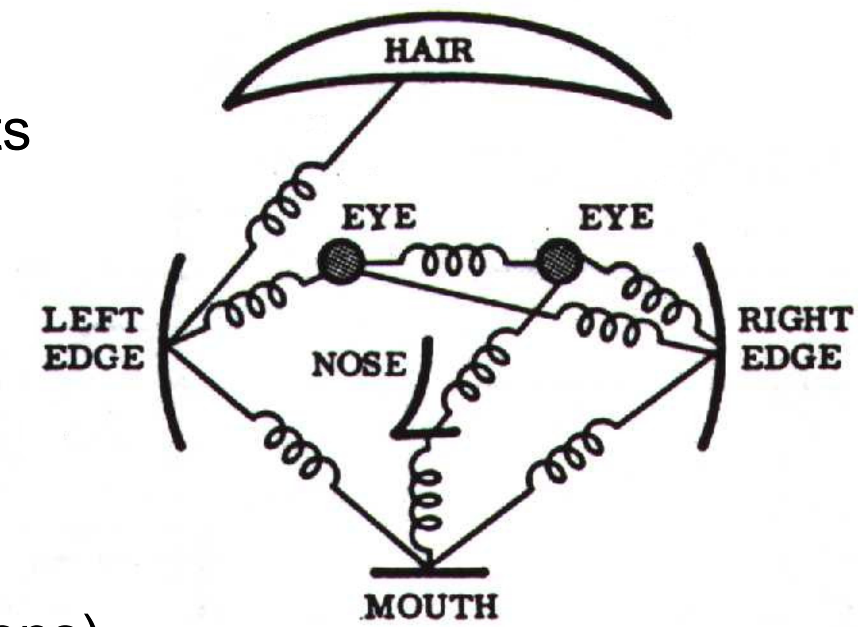
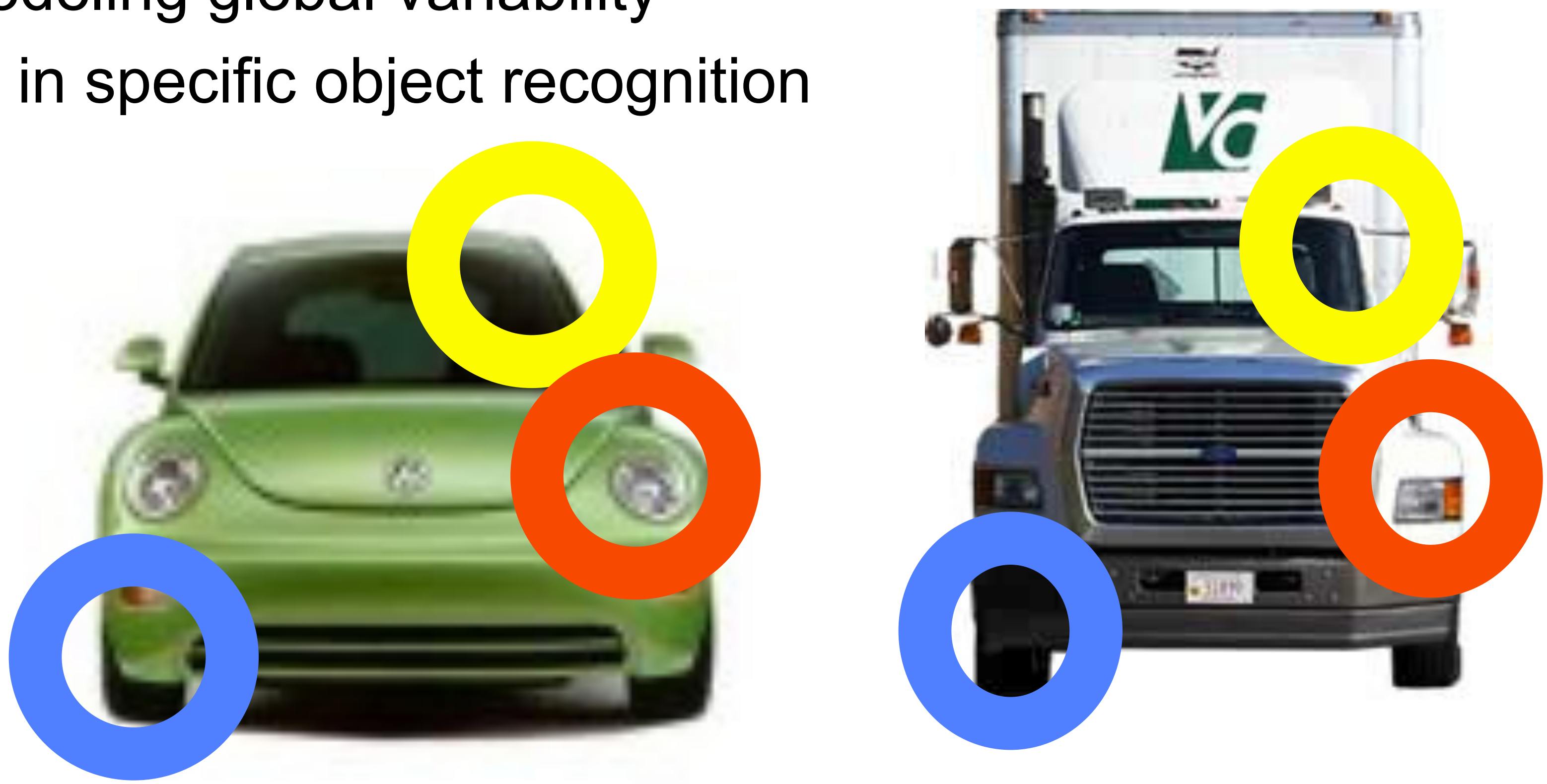


Figure from [Fischler & Elschlager 73]

Sparse representation

+ Computationally tractable (10^5 pixels → 10^1 -- 10^2 parts)
+ Generative representation of class
+ Avoid modeling global variability
+ Success in specific object recognition

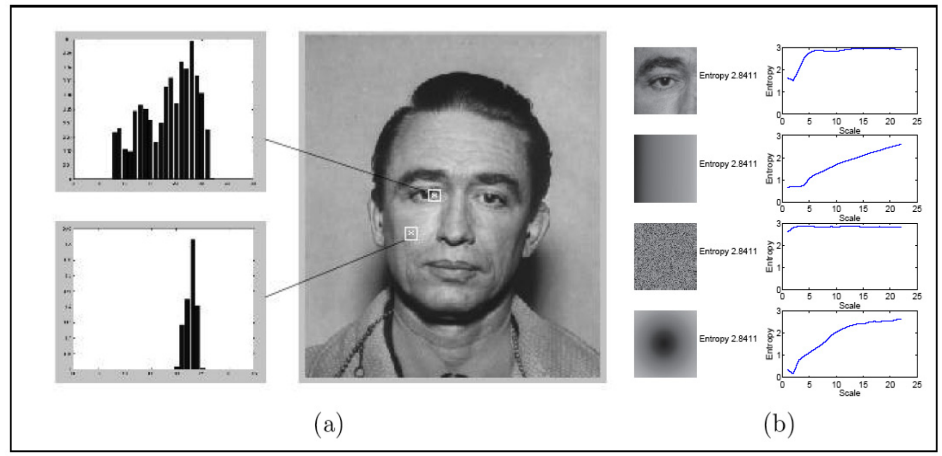


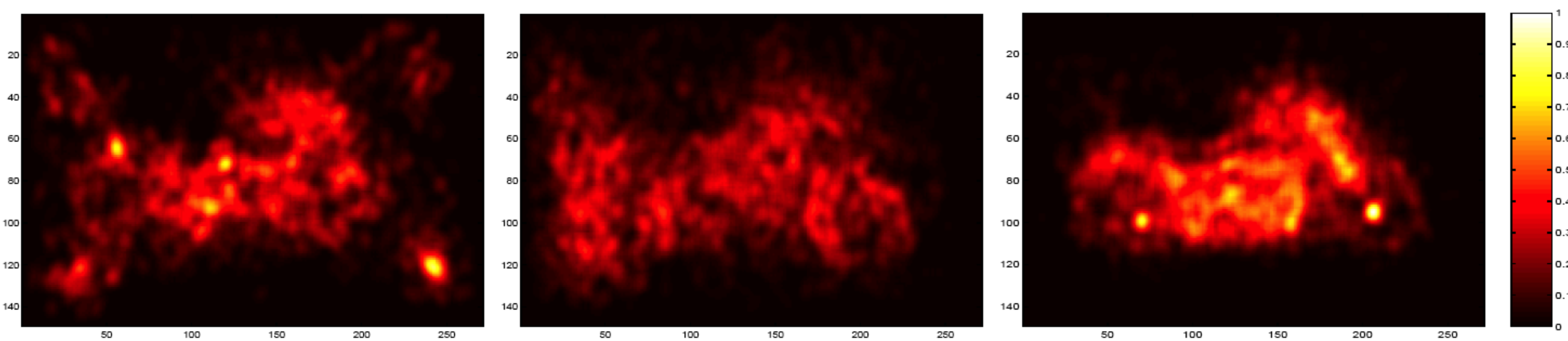
- Throw away most image information
- Parts need to be distinctive to separate from other classes

Region operators

- Local maxima of interest operator function

- Can give scale/ orientation invariance



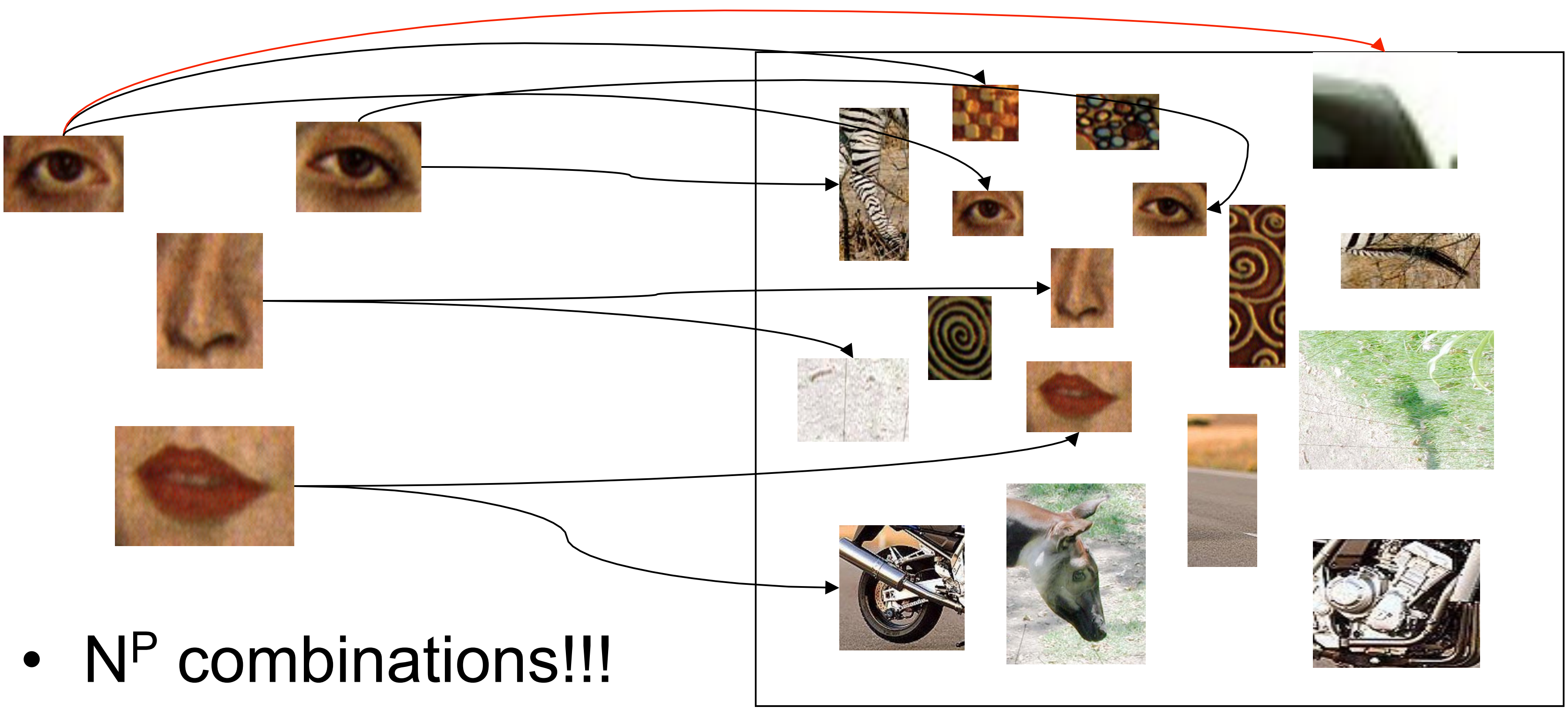


MultiScale Harris Difference-of-Gaussian Saliency

Figures from [Kadir, Zisserman and Brady 04]

The correspondence problem

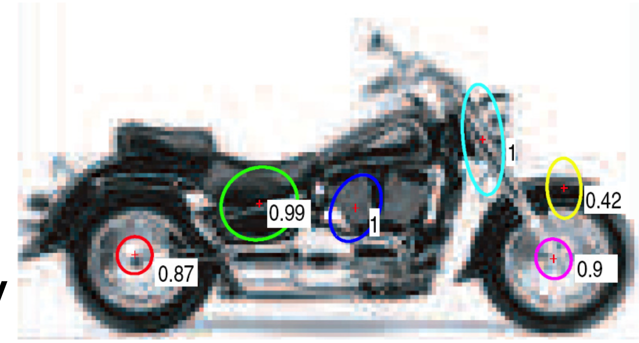
- Model with P parts
- Image with N possible assignments for each part
- Consider mapping to be 1-1



- N^P combinations!!!

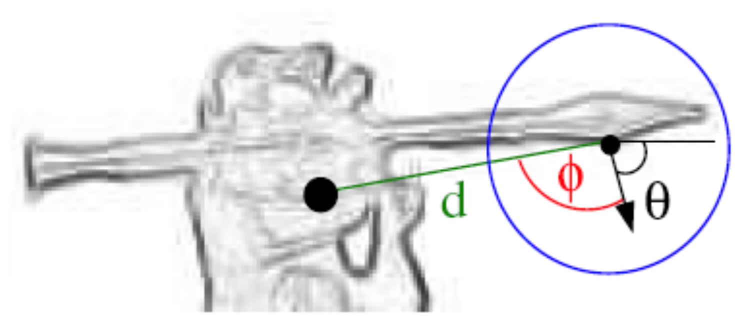
Explicit shape model

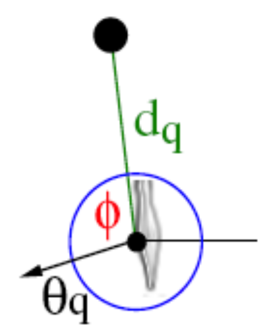
- Cartesian
  - E.g. Gaussian distribution
  - Parameters of model, mean and cov
  - Independence corresponds to zeros in cov
  - Burl et al. '96, Weber et al. '00, Fergus et al. '03



$$\mu = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ y_1 \\ y_2 \\ y_3 \end{pmatrix} \quad \Sigma = \begin{pmatrix} x_1x_1 & x_1x_2 & x_1x_3 & x_1y_1 & x_1y_2 & x_1y_3 \\ x_2x_1 & x_2x_2 & x_2x_3 & x_2y_1 & x_2y_2 & x_2y_3 \\ x_3x_1 & x_3x_2 & x_3x_3 & x_3y_1 & x_3y_2 & x_3y_3 \\ y_1x_1 & y_1x_2 & y_1x_3 & y_1y_1 & y_1y_2 & y_1y_3 \\ y_2x_1 & y_2x_2 & y_2x_3 & y_2y_1 & y_2y_2 & y_2y_3 \\ y_3x_1 & y_3x_2 & y_3x_3 & y_3y_1 & y_3y_2 & y_3y_3 \end{pmatrix}$$

- Polar
  - Convenient for invariance to rotation





Mikolajczyk et al., CVPR '06

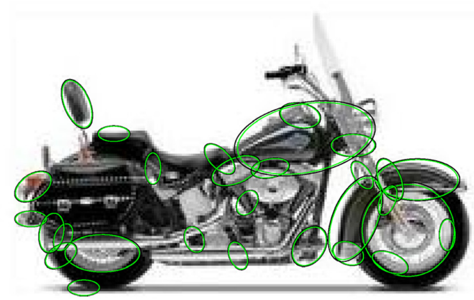
Representation of appearance

- Needs to handle intra-class variation
  - Task is no longer matching of descriptors
  - Implicit variation (VQ to get discrete appearance)
  - Explicit model of appearance (e.g. Gaussians in SIFT space)

- Dependency structure
  - Often assume each part's appearance is independent
  - Common to assume independence with location

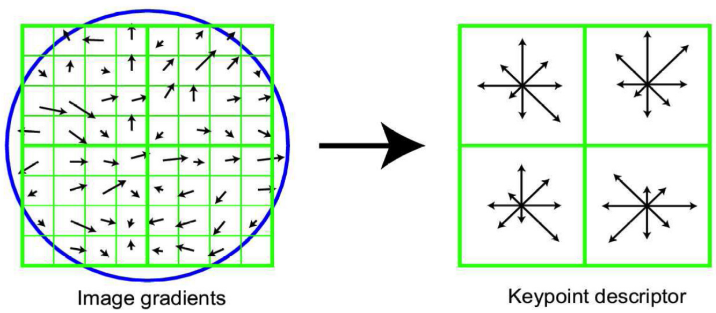
Representation of appearance

- Invariance needs to match that of shape model
- Insensitive to small shifts in translation/scale
  - Compensate for jitter of features
  - e.g. SIFT
- Illumination invariance
  - Normalize out

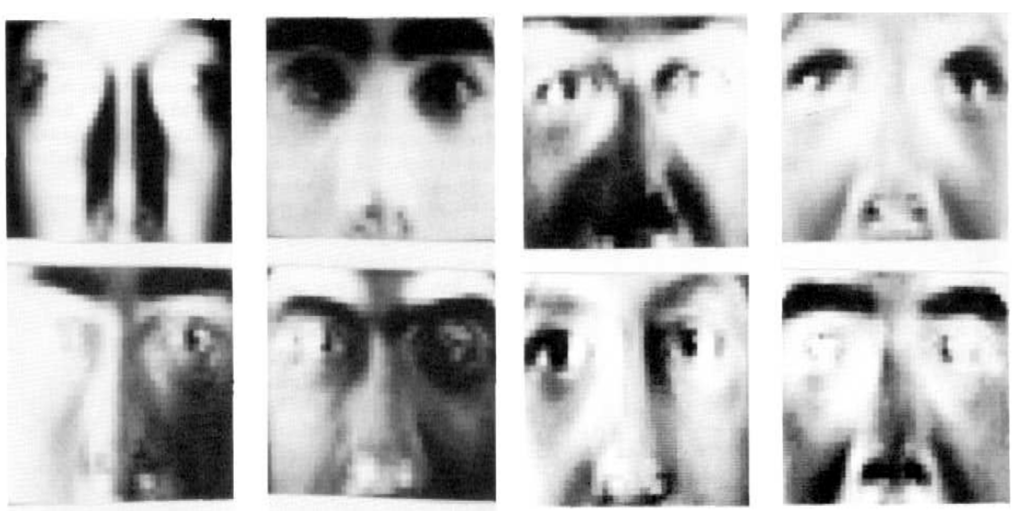


Appearance representation

- SIFT



- PCA



- Decision trees

[Lepetit and Fua CVPR 2005]

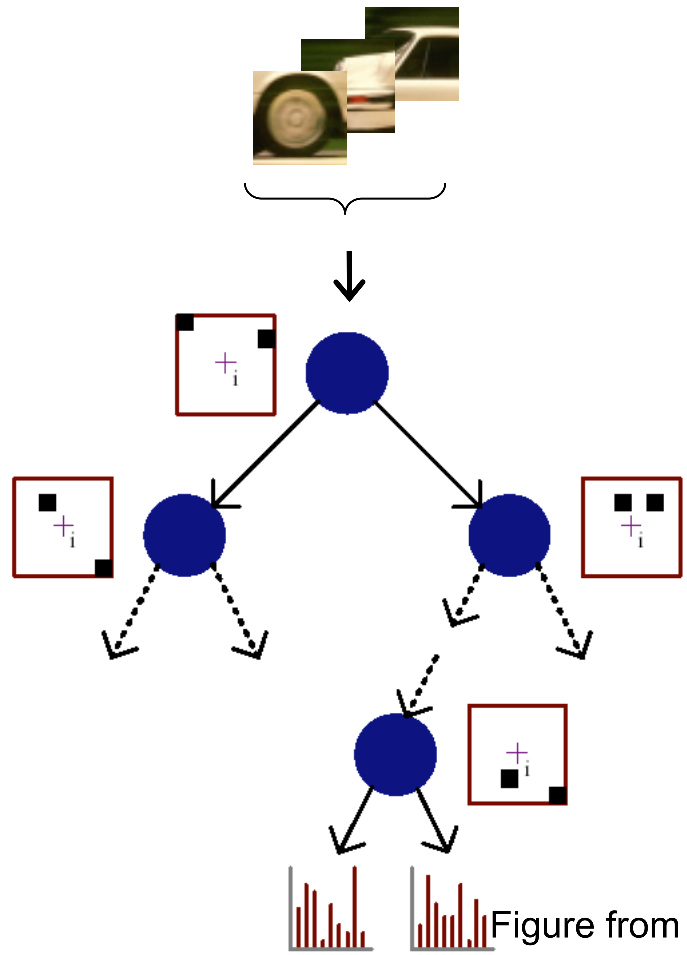
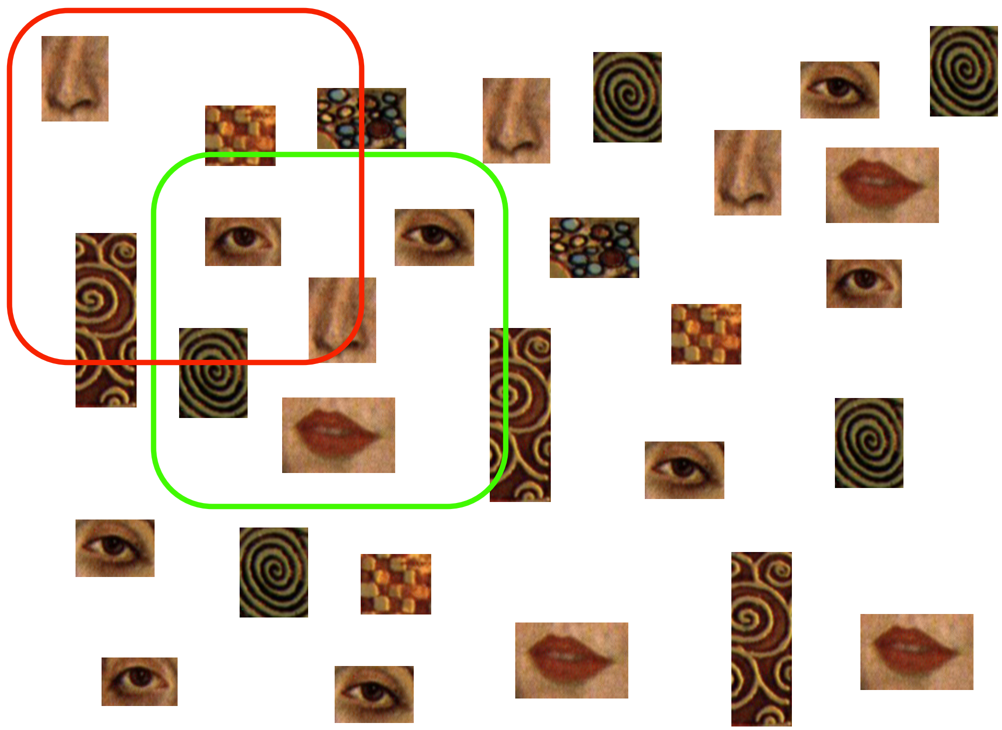


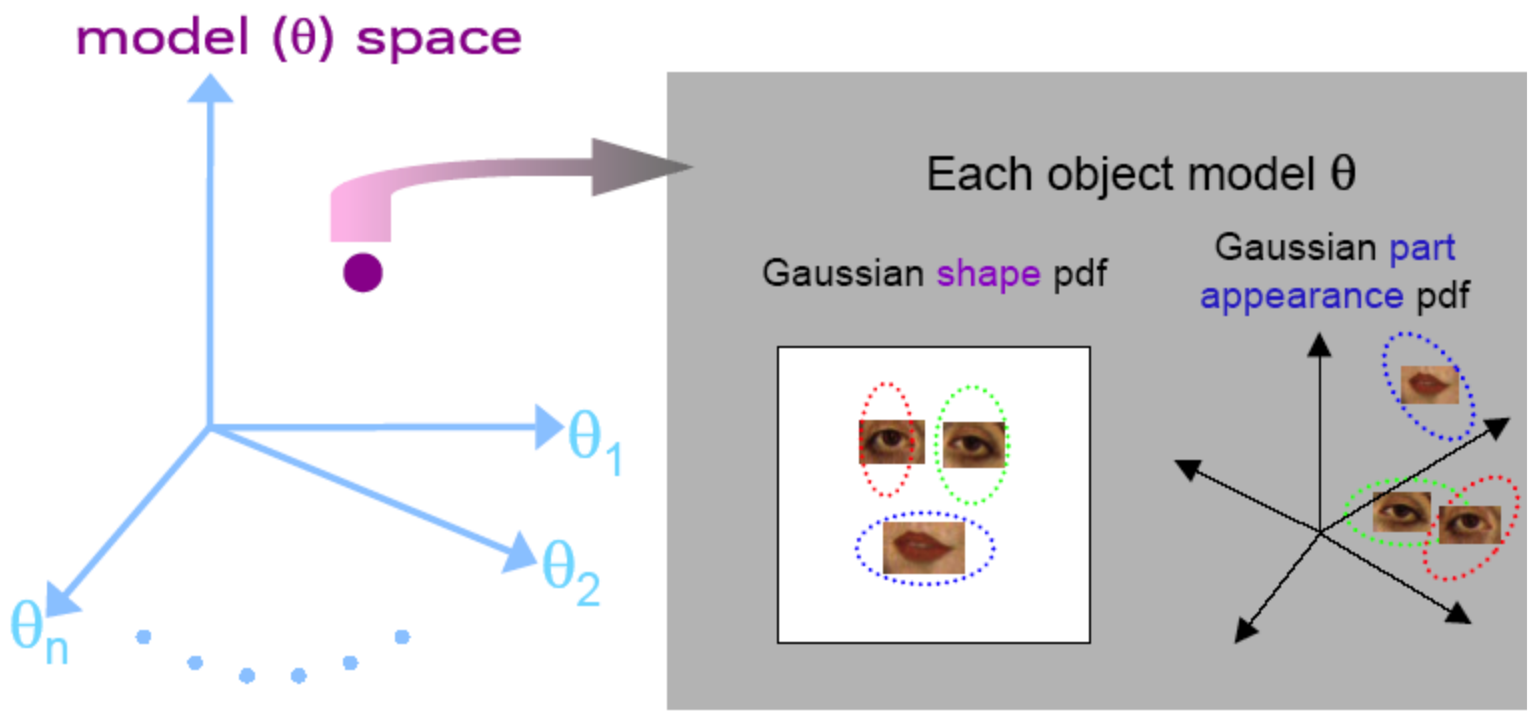
Figure from Winn & Shotton, CVPR '06

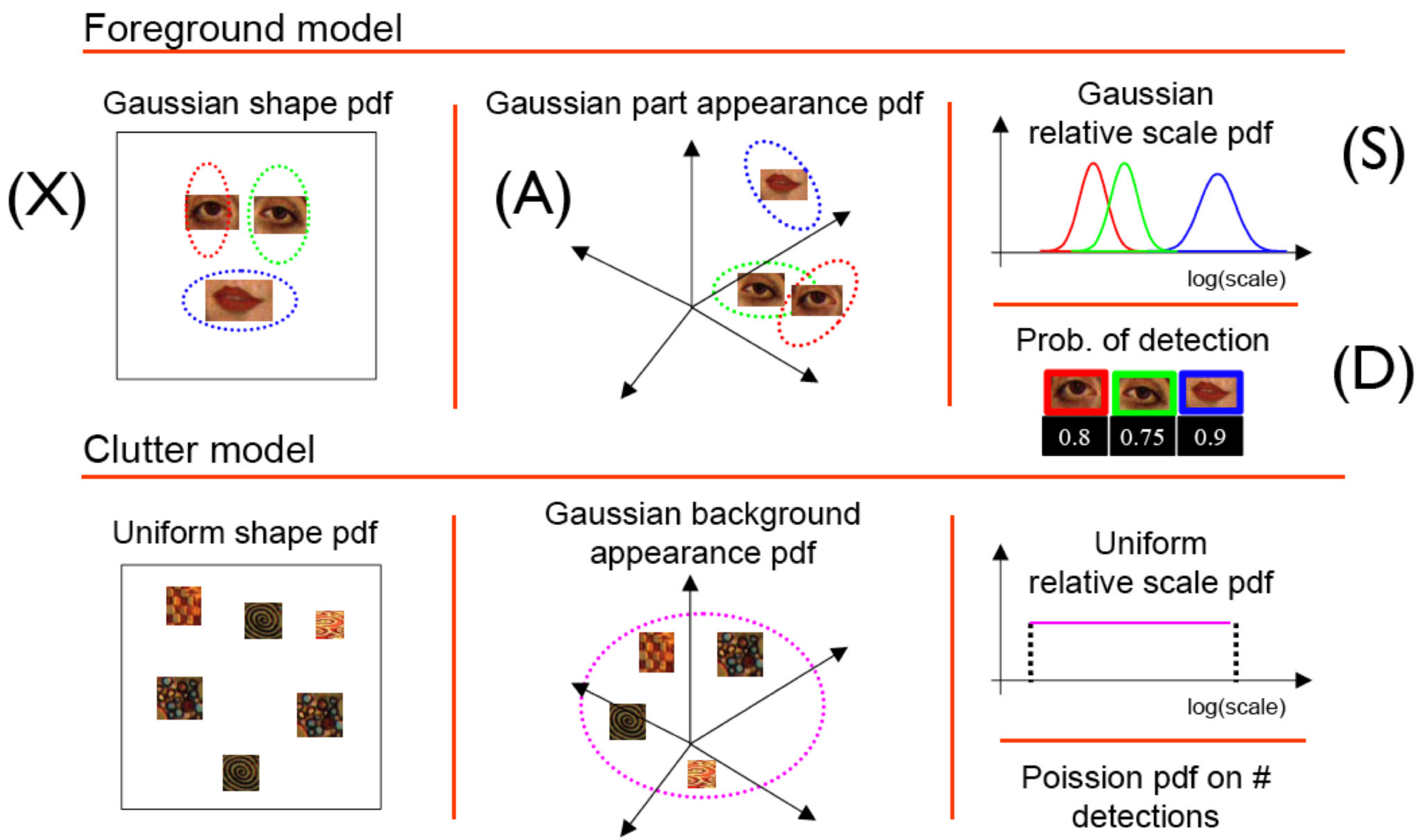
Background clutter

- Explicit model
  - Generative model for clutter as well as foreground object
- Use a sub-window
  - At correct position, no clutter is present



Representing Objects




Foreground model
Gaussian shape pdf
(X)
Gaussian part appearance pdf
(A)
Gaussian
relative scale pdf
log(scale)
(S)
Prob. of detection
0.8
0.75
0.9
(D)
Clutter model
Uniform shape pdf
Gaussian background
appearance pdf
Uniform
relative scale pdf
log(scale)
Poission pdf on #
detections

Demo (2)

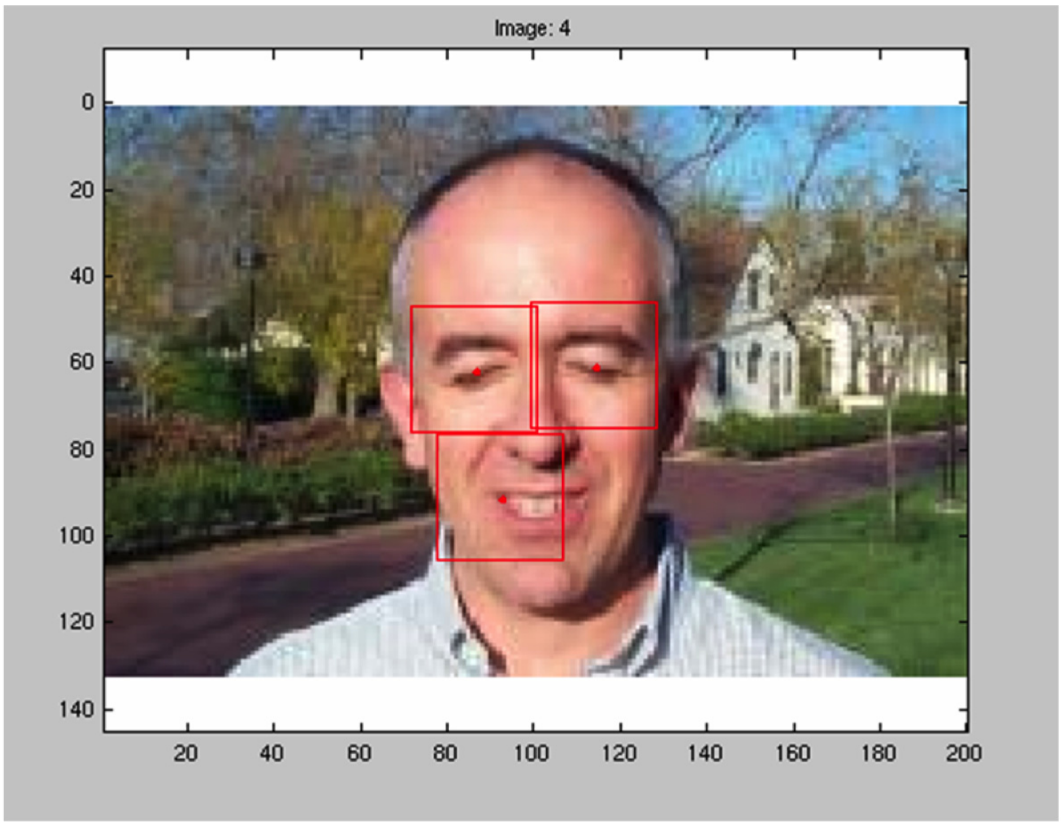
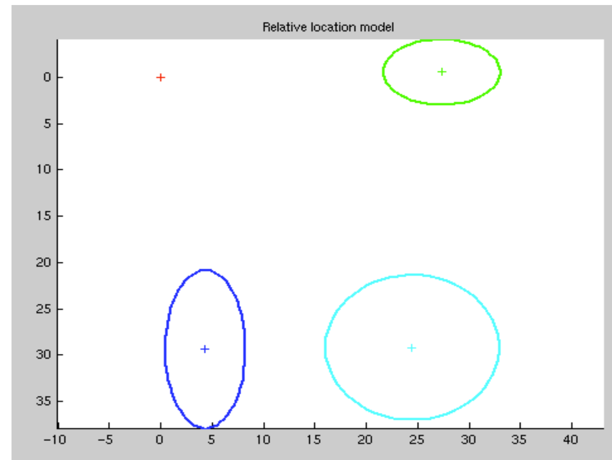
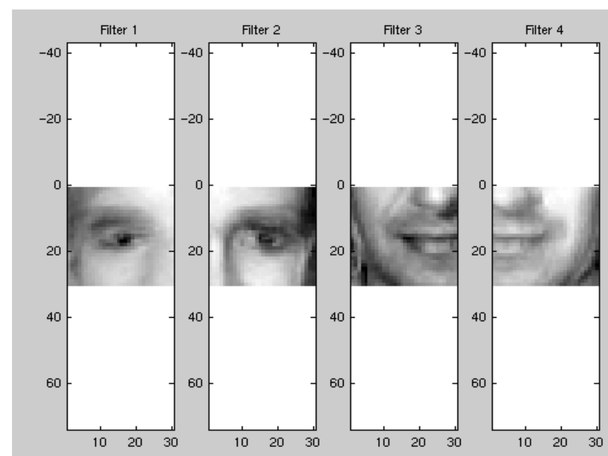
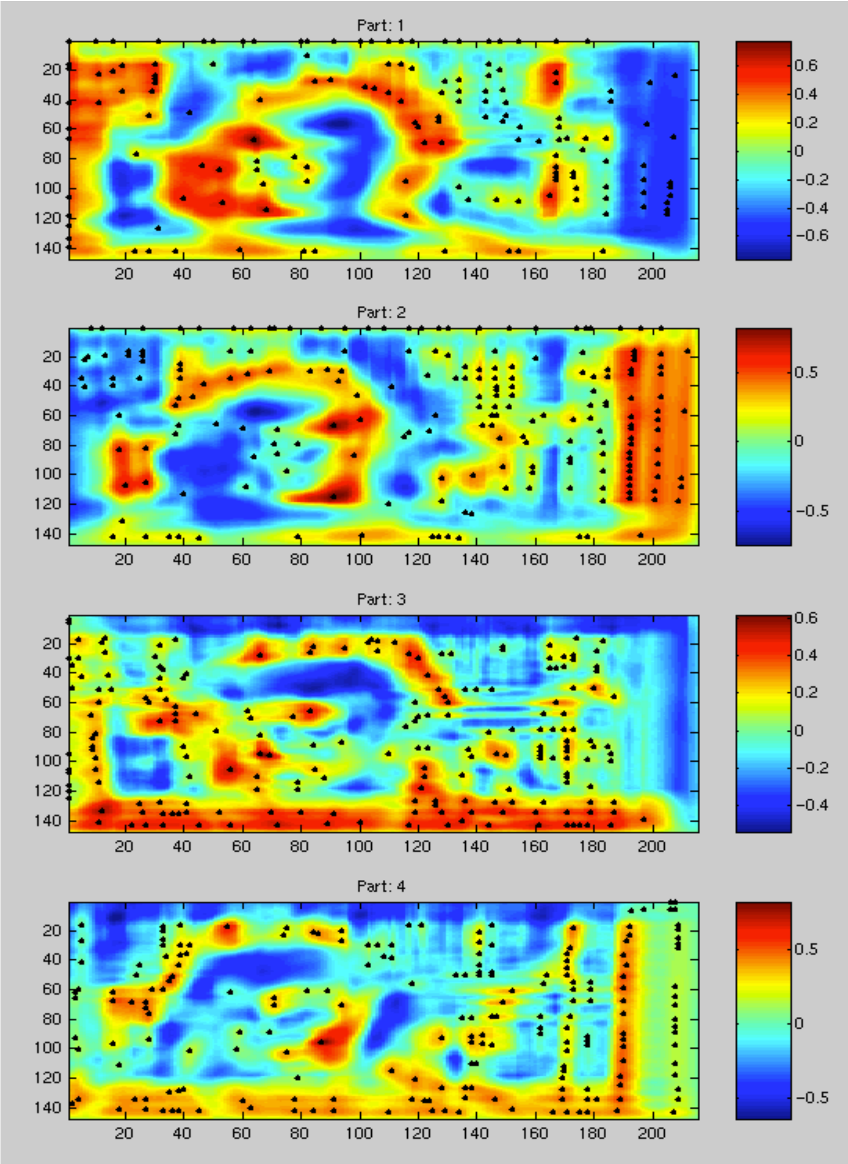


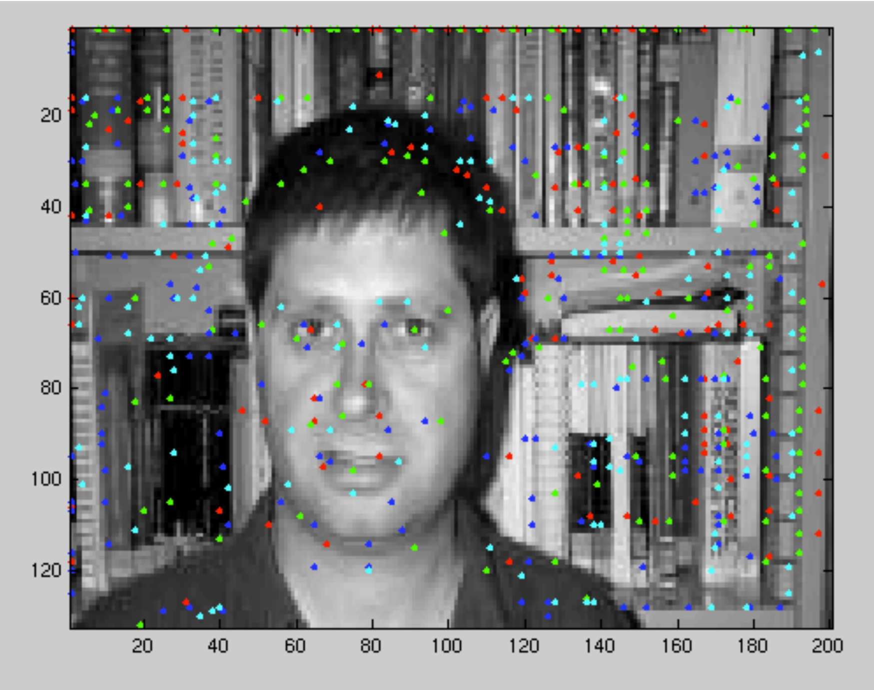
Image: 4
0
20
40
60
80
100
120
140
20 40 60 80 100 120 140 160 180 200


Relative location model
0
5
10
15
20
25
30
35
−10 −5 0 5 10 15 20 25 30 35 40

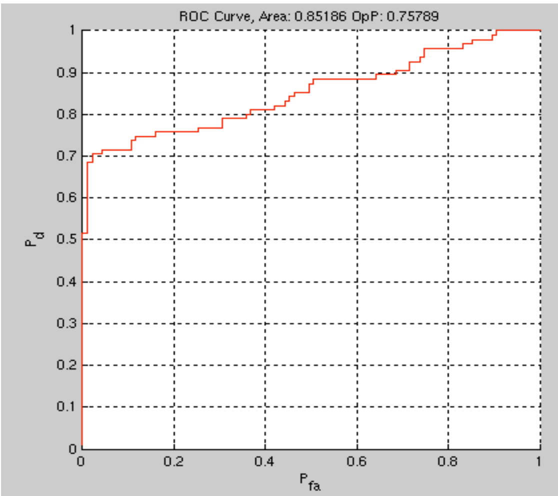

Filter 1
Filter 2
Filter 3
Filter 4
−40
−20
0
20
40
60
10 20 30

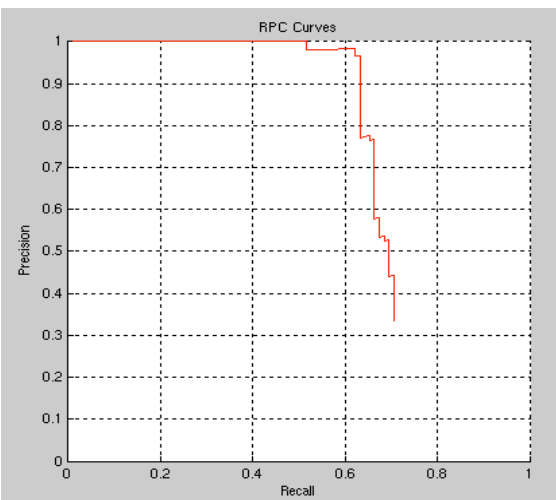
Demo (3)





Demo (4)


ROC Curve, Area: 0.85186 OpP: 0.75789
P_d
P_fa


RPC Curves
Precision
Recall

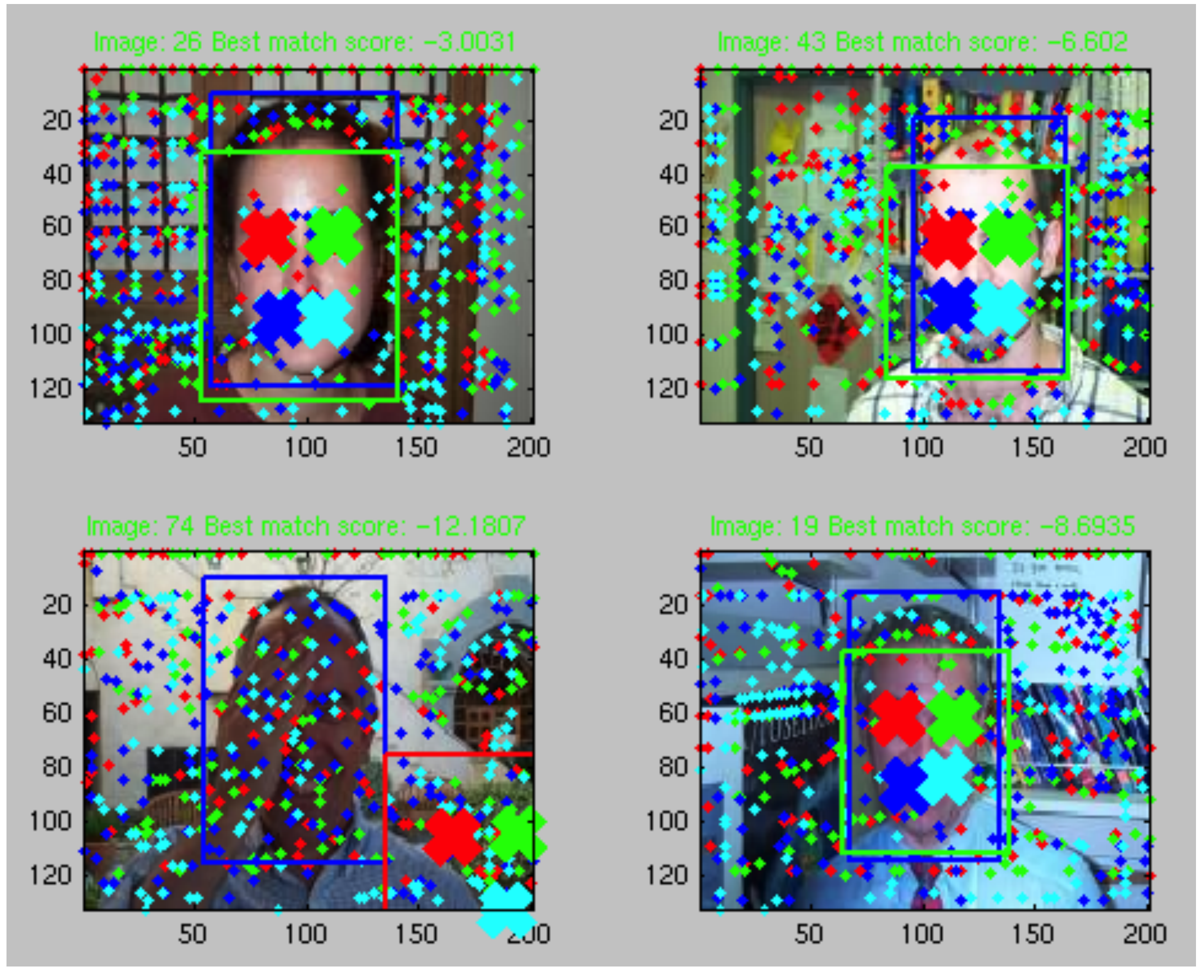
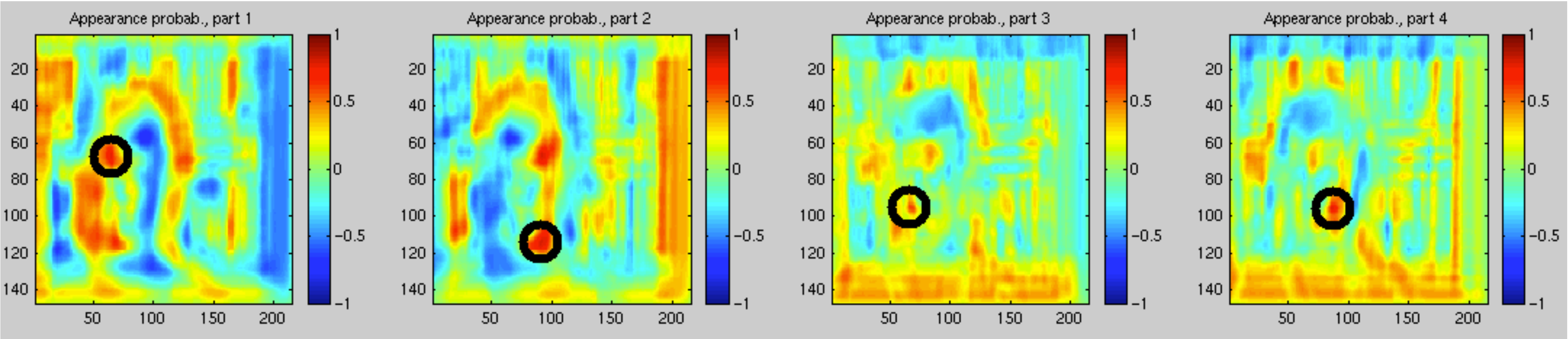
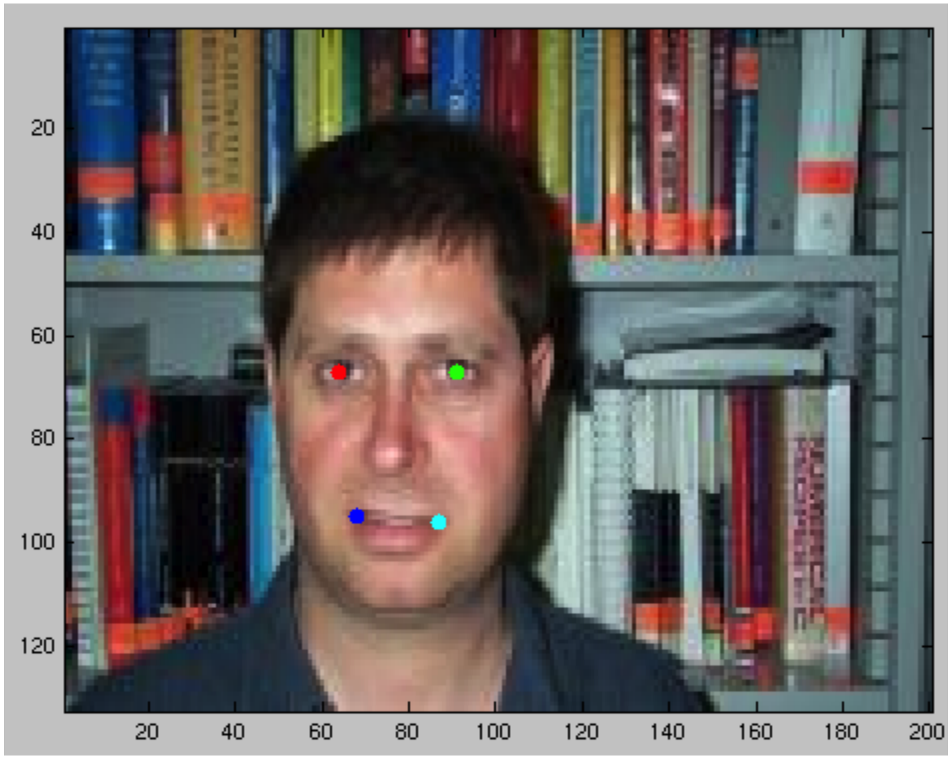
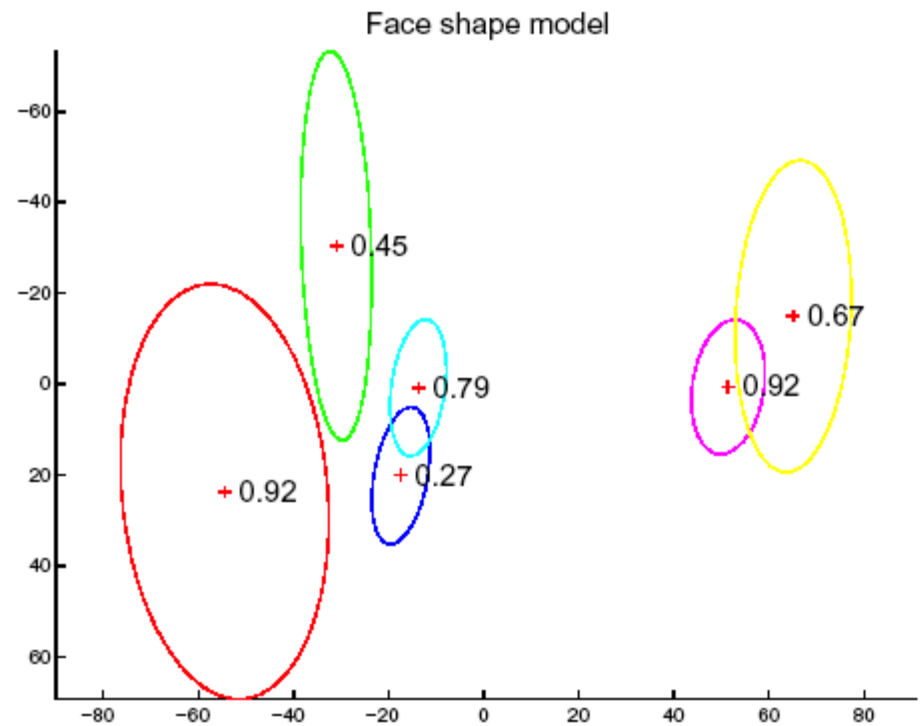


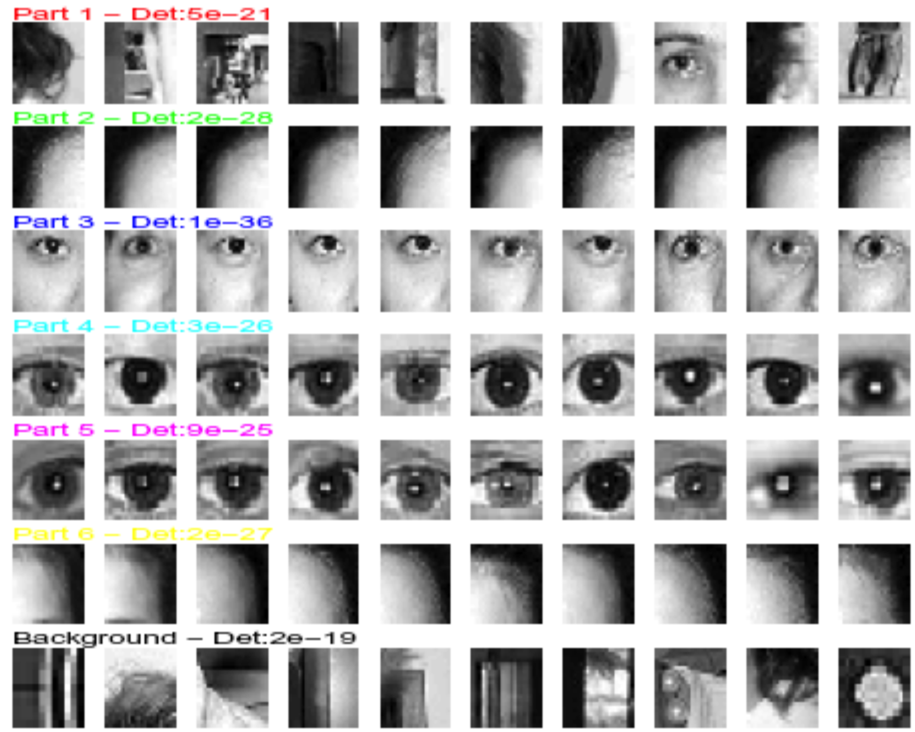
Image: 26 Best match score: -3.0031
Image: 43 Best match score: -6.602
Image: 74 Best match score: -12.1807
Image: 19 Best match score: -8.6935

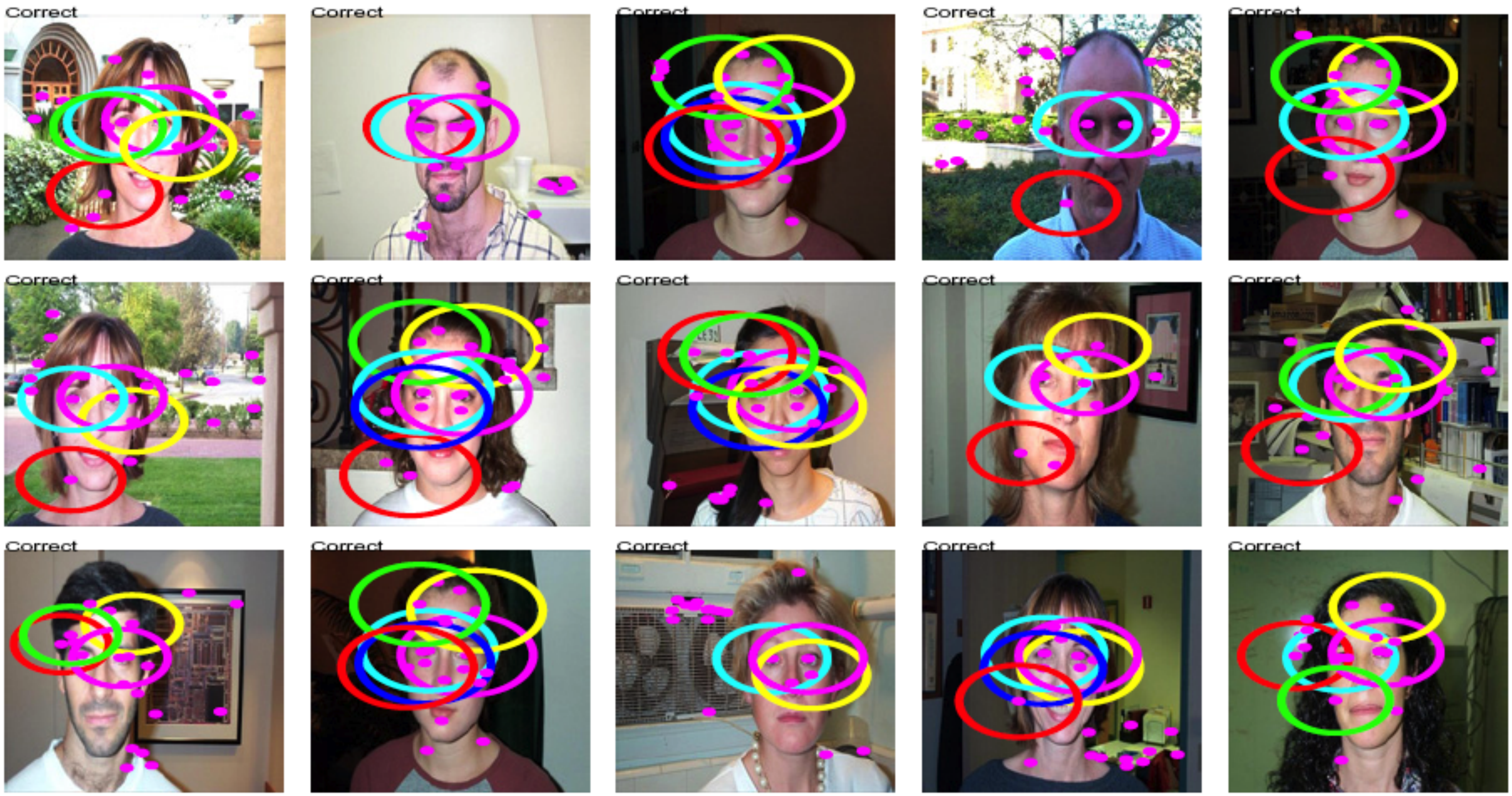
Demo: efficient methods

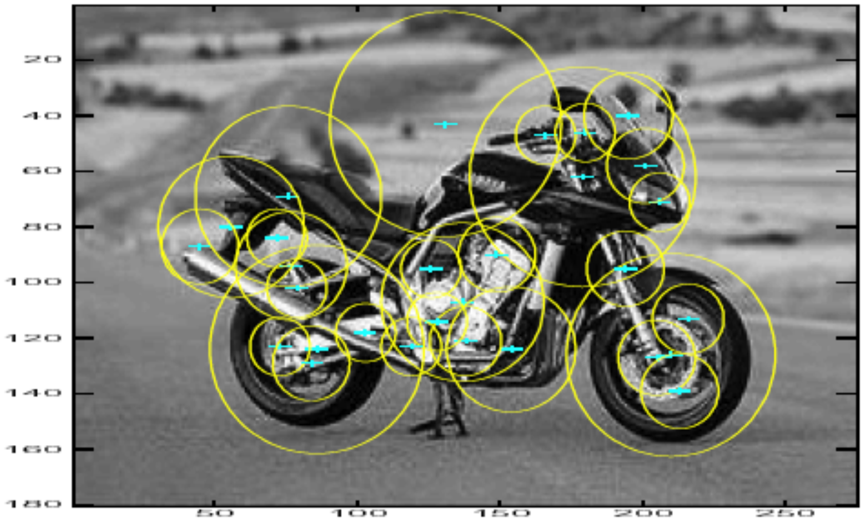


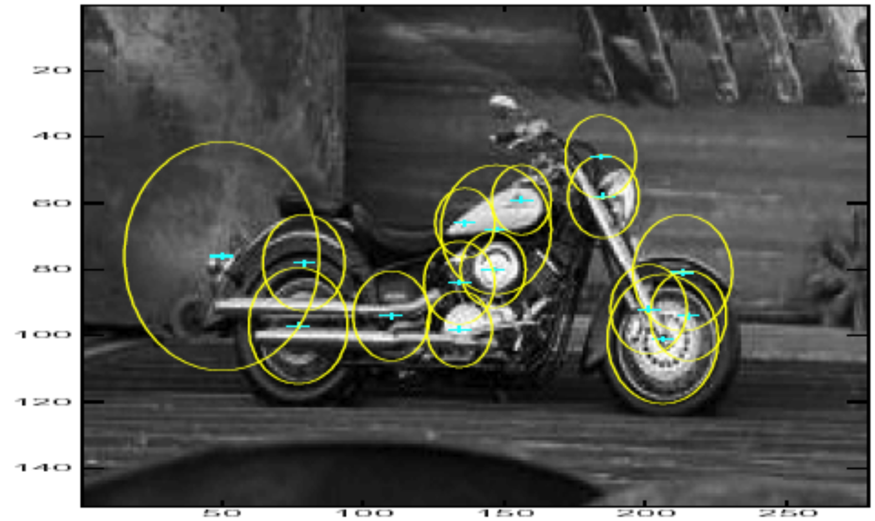


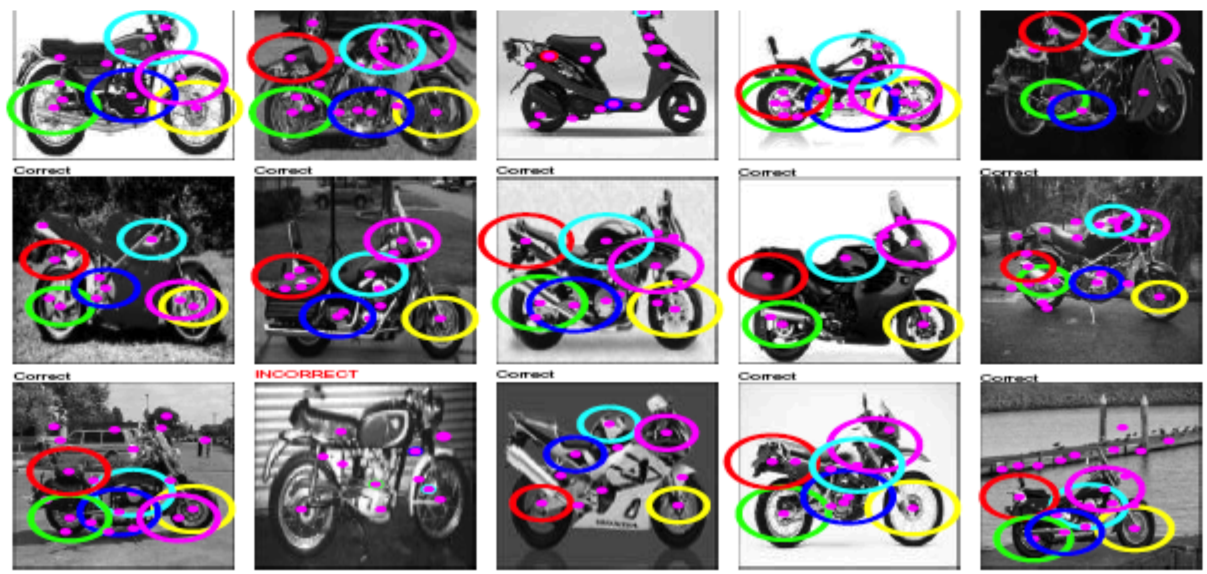

Face shape model
0.45
0.67
0.79
0.92
0.92
0.27
-60
-40
-20
0
20
40
60
-80
-60
-40
-20
0
20
40
60
80


Part 1 – Det:5e–21
Part 2 – Det:2e–28
Part 3 – Det:1e–36
Part 4 – Det:3e–26
Part 5 – Det:9e–25
Part 6 – Det:2e–27
Background – Det:2e–19


Correct
Correct
Correct
Correct
Correct
Correct
Correct
Correct
Correct
Correct
Correct
Correct
Correct
Correct
Correct


20
40
60
80
100
120
140
160
180
50
100
150
200
250


20
40
60
80
100
120
140
50
100
150
200
250


Correct
Correct
Correct
Correct
Correct
Correct
INCORRECT
Correct
Correct
Correct

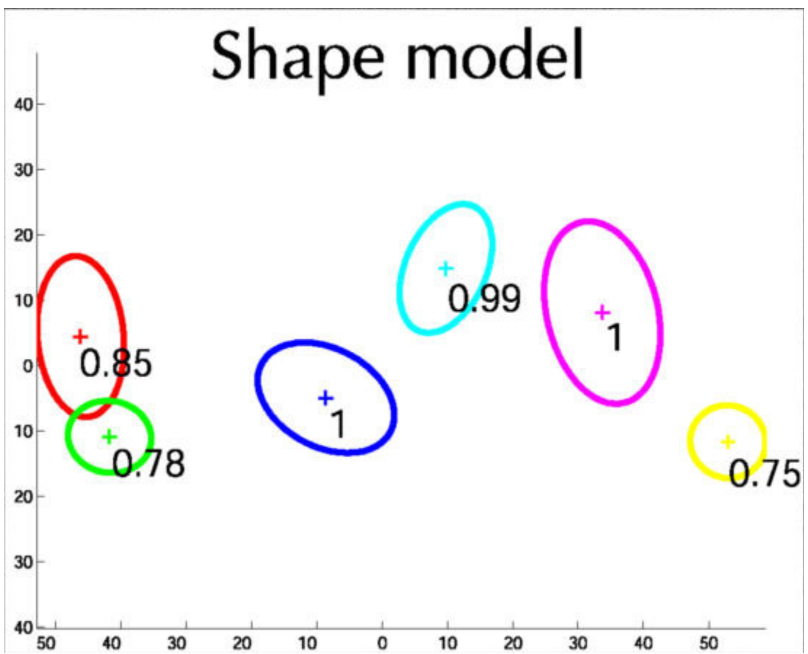
Learn parts from examples.

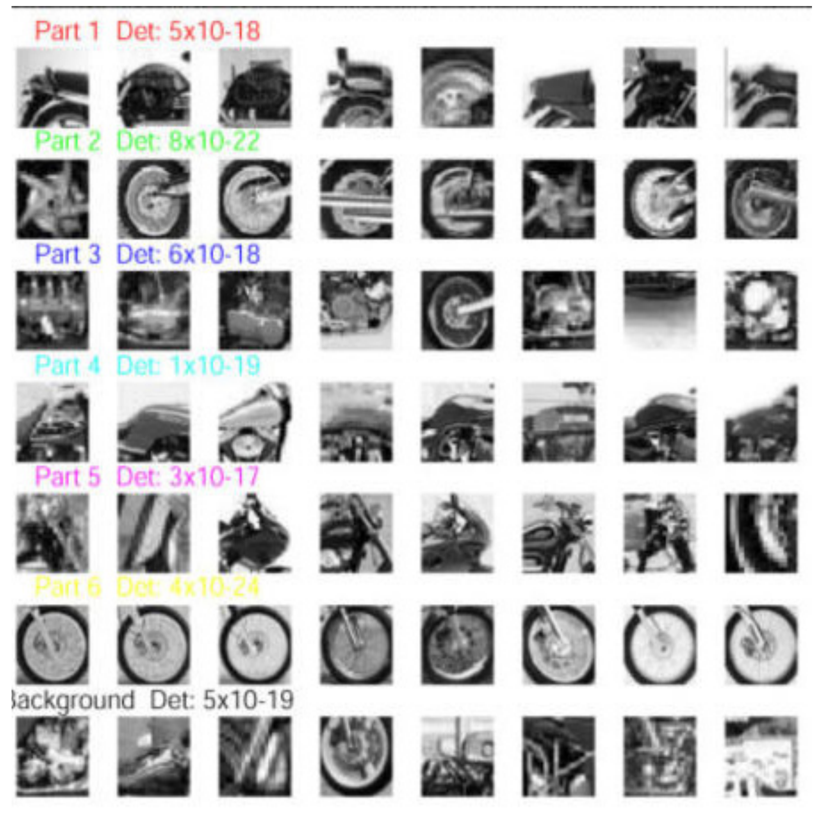
Find interesting points (structure tensor), find similar ones, use PCA to model them.

From: Rob Fergus http://www.robots.ox.ac.uk/%7Efergus/

Shape

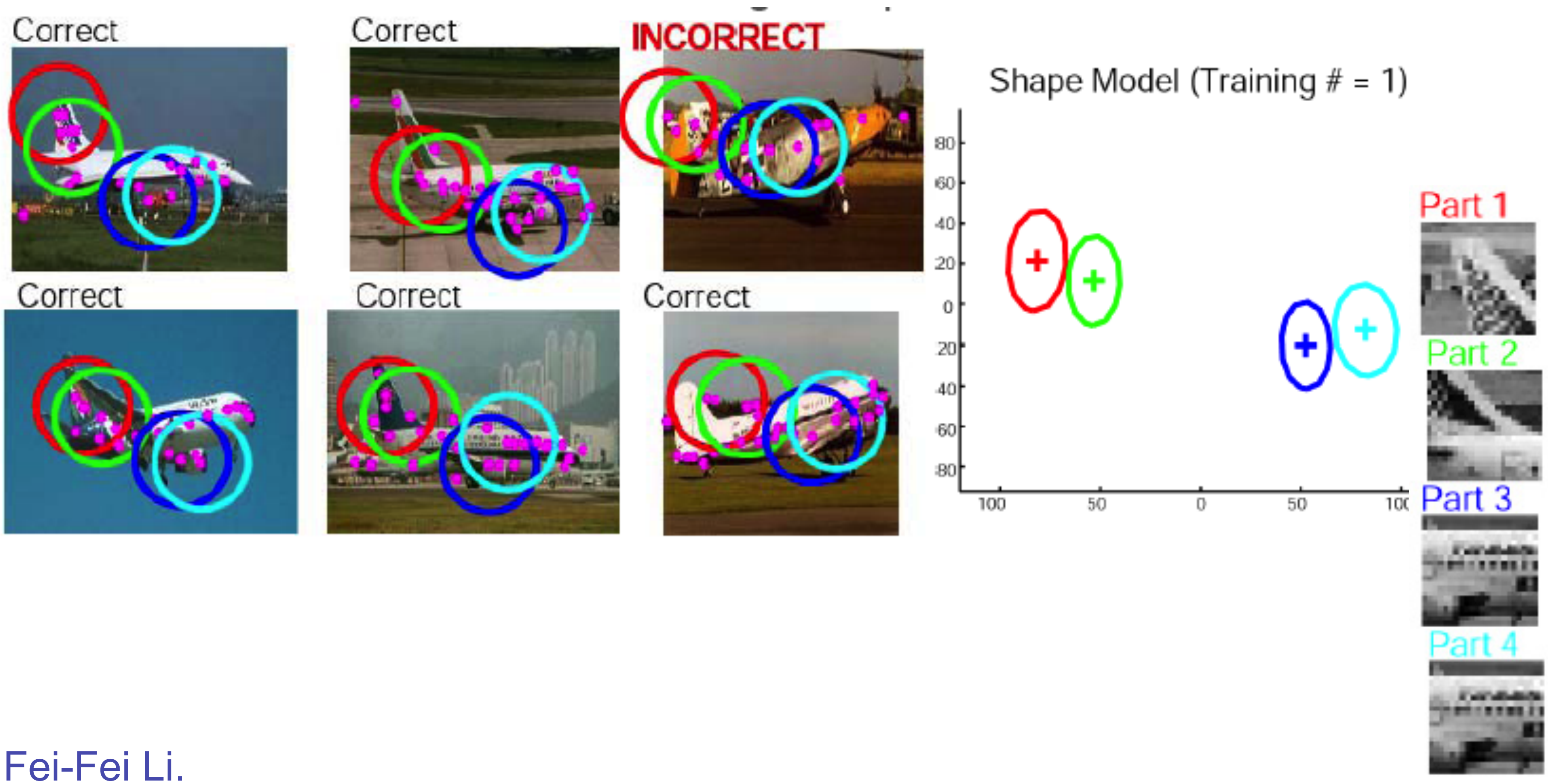
Given a “vocabulary” of parts, learn a model of their spatial relationships





From: Rob Fergus http://www.robots.ox.ac.uk/%7Efergus/

Recognizing Objects



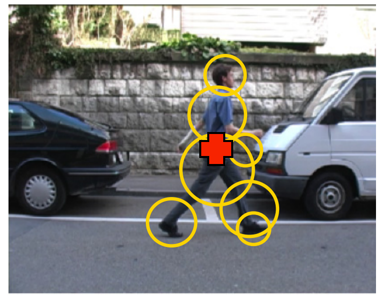
Fei-Fei Li.

Implicit shape model

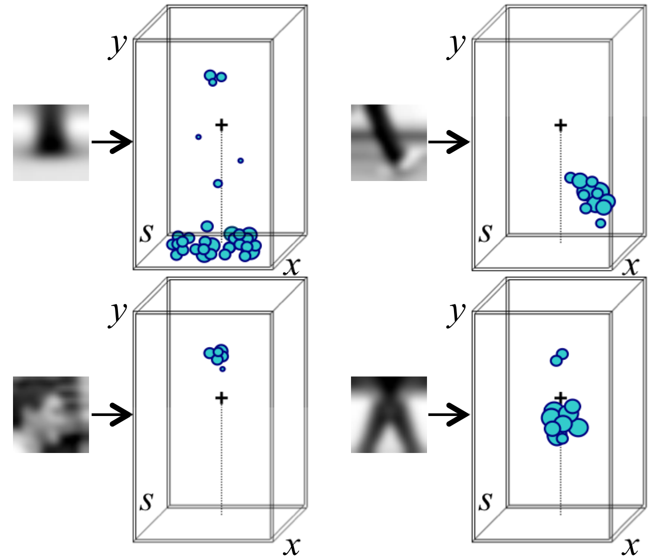
- Use Hough space voting to find object
- Leibe and Schiele ’03,’05

Learning

- Learn appearance codebook
  - Cluster over interest points on training images



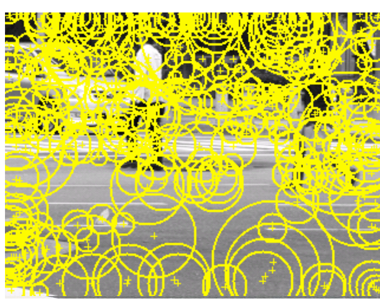
- Learn spatial distributions
  - Match codebook to training images
  - Record matching positions on object
  - Centroid is given



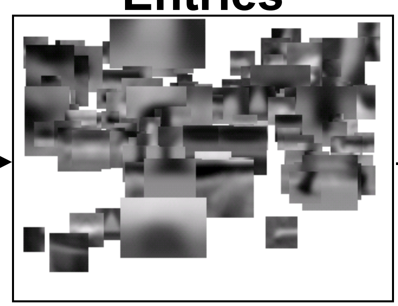
Spatial occurrence distributions

Recognition

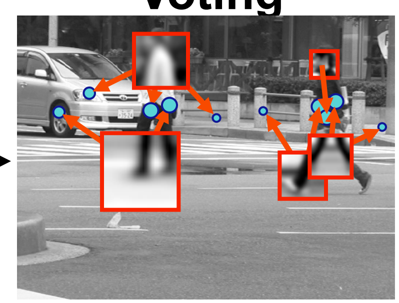
Interest Points



Matched Codebook Entries



Probabilistic Voting



~100 Things We've Learned

Pinhole camera
Perspective projection
Orthographic projection
Weak perspective
PCA
Eigenvalues
Eigenvectors
Inpainting
Markov random field
Particle filter
Image statistics
Continuation method
Graduated non-convexity
MAP estimate

Gaussian pyramid
Laplacian pyramid
Matlab
Linear filtering
Convolution
Gaussian
Gradient
Dimensionality reduction
Monte Carlo sampling
Convolution
Correlation

Projection
Finite differences
Steerable filter
Gradient magnitude
DoG
Template matching
Normalized correlation
SSD
Subspaces
Basis image
SVD
Eigenfaces
Histogram

~100 Things We've Learned

Random variable
Marginalize
Expectation
Statistical independence
Conditional independence
Joint probability
Conditional probability
Bayes Theorem
Likelihood
Prior
Classifier
Tracking
Regularization
Stereo

Posterior
Covariance
Structure tensor
Mahalanobis distance
Whitening
Denoising
Motion field
Optical flow
Taylor series
Brightness constancy
OFCE
Aperture problem
Outliers
Rectification
Epipole

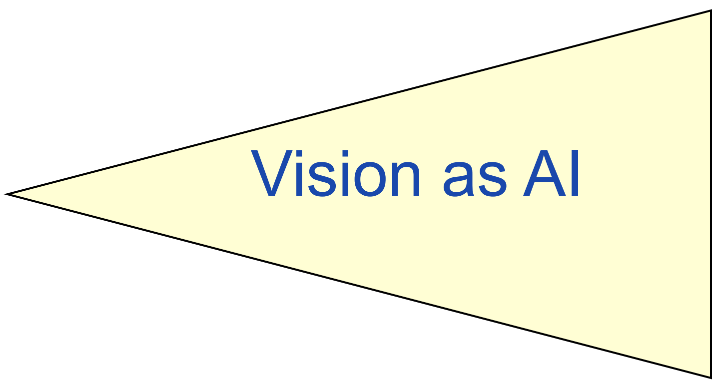
Affine
Least squares
Generative model
Warping
Interpolation
Super resolution
Occlusion
Robust statistics
Influence function
Breakdown point
Gradient descent
Annealing
Discontinuities
Binocular disparity

~100 Things We've Learned

So are we done?

Timeline

1975-1985

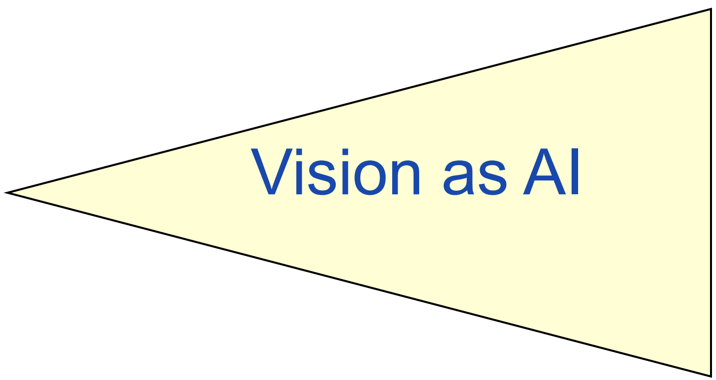


Early view (50's-60's): Minsky thought the vision sub-problem of AI could be solved by a single PhD student in a single summer. Done. Move on.

Lofty goals and early excitement.

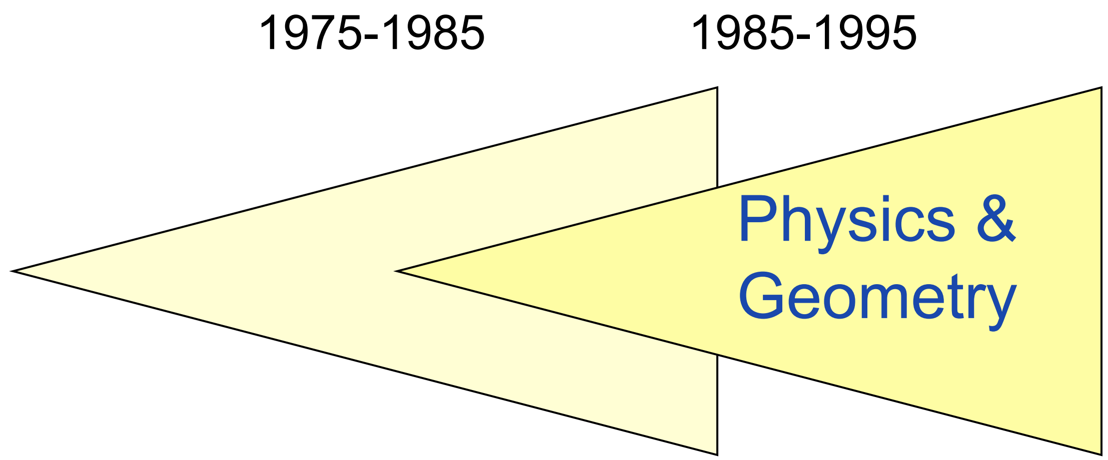
Timeline

1975-1985



Shattered dreams and early disappointment.

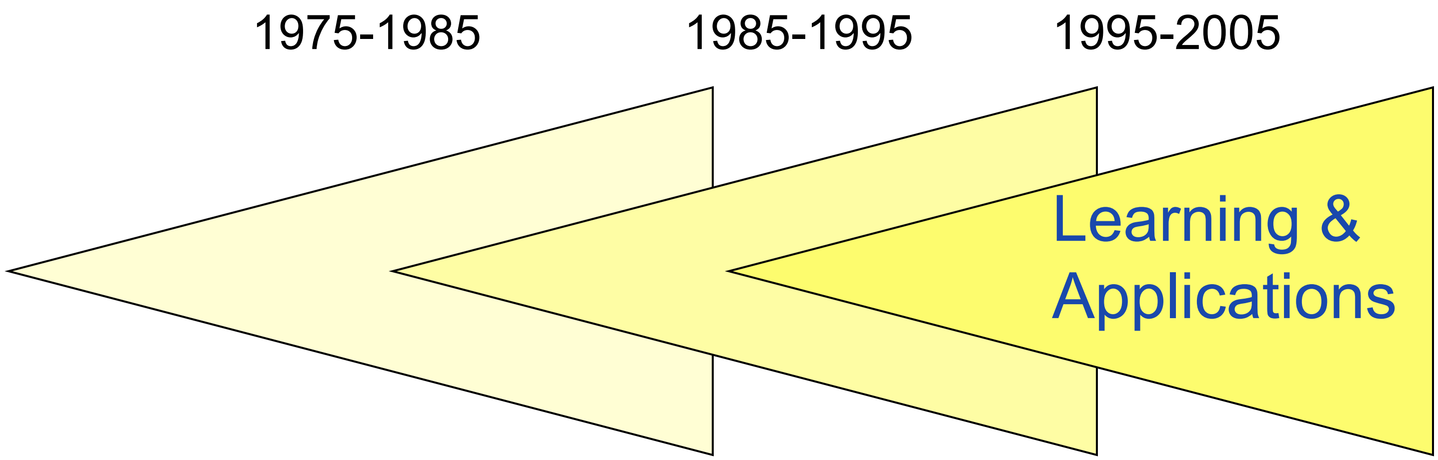
Timeline



Regroup, focus on the basics

* metric reconstruction, quantitative evaluation.
* optimization methods.

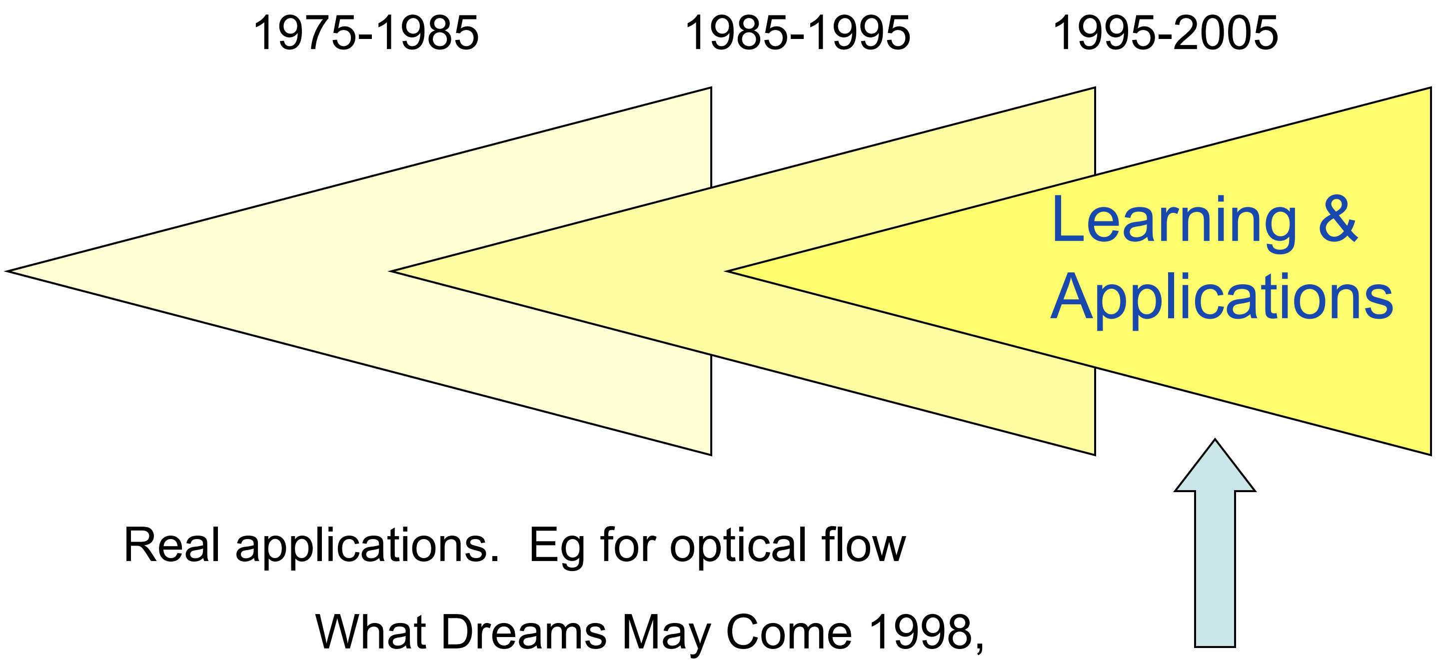
Timeline



Trends: big disks, digital cameras, Firewire, fast processors, desktop video.

Machine learning provides a new grounding.

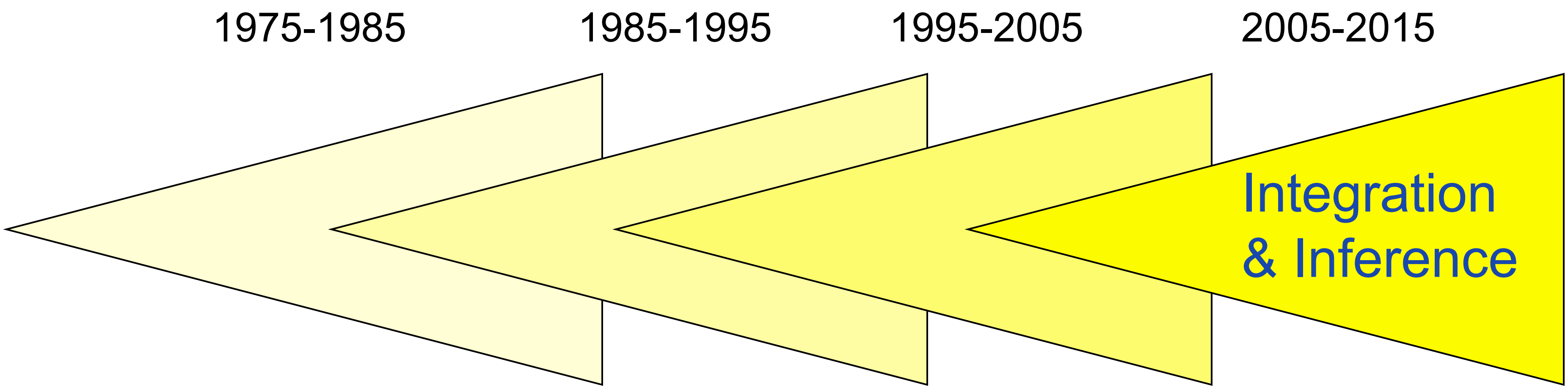
Timeline



Real applications. Eg for optical flow

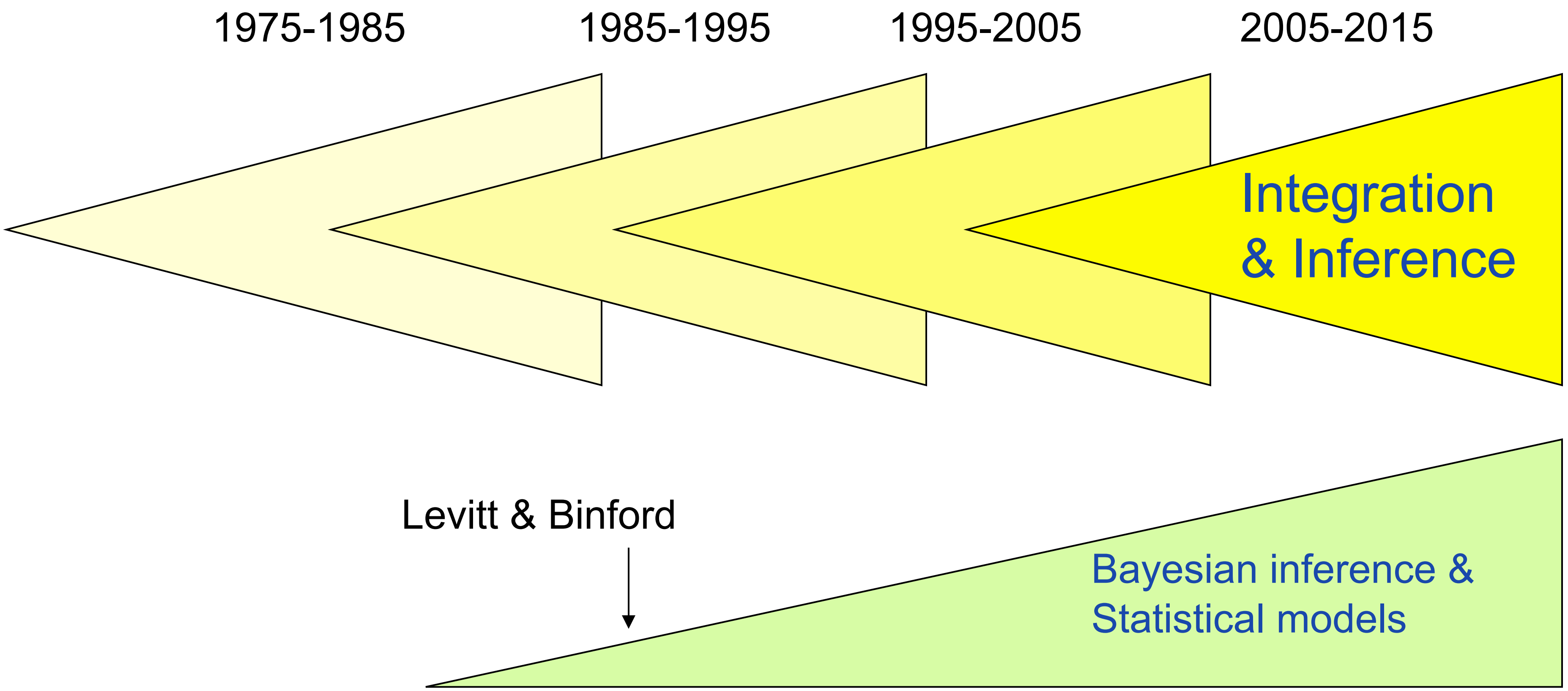
What Dreams May Come 1998,
The Matrix 1999

Timeline



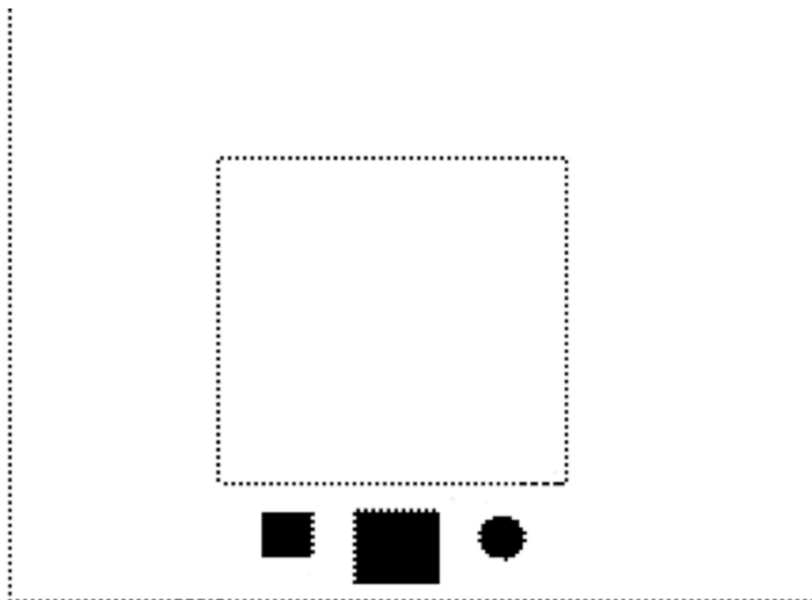
Return to some of the early goals with new tools.

Timeline


1975-1985
1985-1995
1995-2005
2005-2015
Integration
& Inference
Levitt & Binford
Bayesian inference &
Statistical models

What is still far off?

Motion interpretation.



<Play>

Heider&Simmel, 1944

* Here “vision” problem is trivial but explanation is hard.