Lecture 16:
Object recognition:
Part-based generative models

Professor Fei-Fei Li

Stanford Vision Lab

What we will learn today?

- Introduction
- Constellation model
  - Weakly supervised training
  - One-shot learning
- (Problem Set 4 (Q1))

Challenges: intra-class variation






Usual Challenges:

Variability due to:

- View point
- Illumination
- Occlusions

Basic issues

- **Representation**
  - 2D Bag of Words (BoW) models;
  - Part-based models;
  - Multi-view models;
- **Learning**
  - Generative & Discriminative BoW models
  - Generative models
  - Probabilistic Hough voting
- **Recognition**
  - Classification with BoW
  - Classification with Part-based models

Basic issues

- **Representation**
  - 2D Bag of Words (BoW) models;
  - Part-based models;
  - Multi-view models;
- **Learning**
  - Generative & Discriminative BoW models
  - Generative models
  - Probabilistic Hough voting
- **Recognition**
  - Classification with BoW
  - Classification with Part-based models

Basic issues

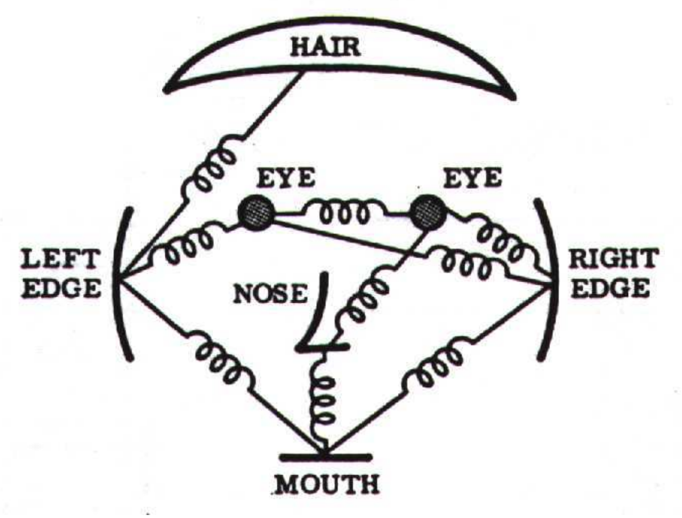
- **Representation**
  - 2D Bag of Words (BoW) models;
  - Part-based models;
  - Multi-view models (Lecture #19);
- **Learning**
  - Generative & Discriminative BoW models
  - Generative models
  - Probabilistic Hough voting
- **Recognition**
  - Classification with BoW
  - Classification with Part-based models

Problem with bag-of-words

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

Model: Parts and Structure

Parts and Structure Literature



- Fischler & Elschlager 1973

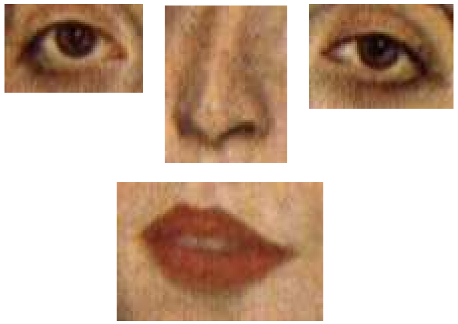
- Yuille '91
- Brunelli & Poggio '93
- Lades, v.d. Malsburg et al. '93
- Cootes, Lanitis, Taylor et al. '95
- Amit & Geman '95, '99
- et al. Perona '95, '96, '98, '00, '03
- Huttenlocher et al. '00
- Agarwal & Roth '02

etc…

The Constellation Model

T. Leung →	Representation	Shape statistics – F&G '95 Affine invariant shape – CVPR '98
M. Burl →	Detection	CVPR '96 ECCV '98
M. Weber M. Welling →	Unsupervised Learning	ECCV '00 Multiple views - F&G '00 Discovering categories - CVPR '00
R. Fergus →	Joint shape & appearance learning Generic feature detectors	CVPR '03 Polluted datasets - ECCV '04
L. Fei-Fei →	One-Shot Learning Incremental learning	ICCV '03 CVPR '04

Deformations



A

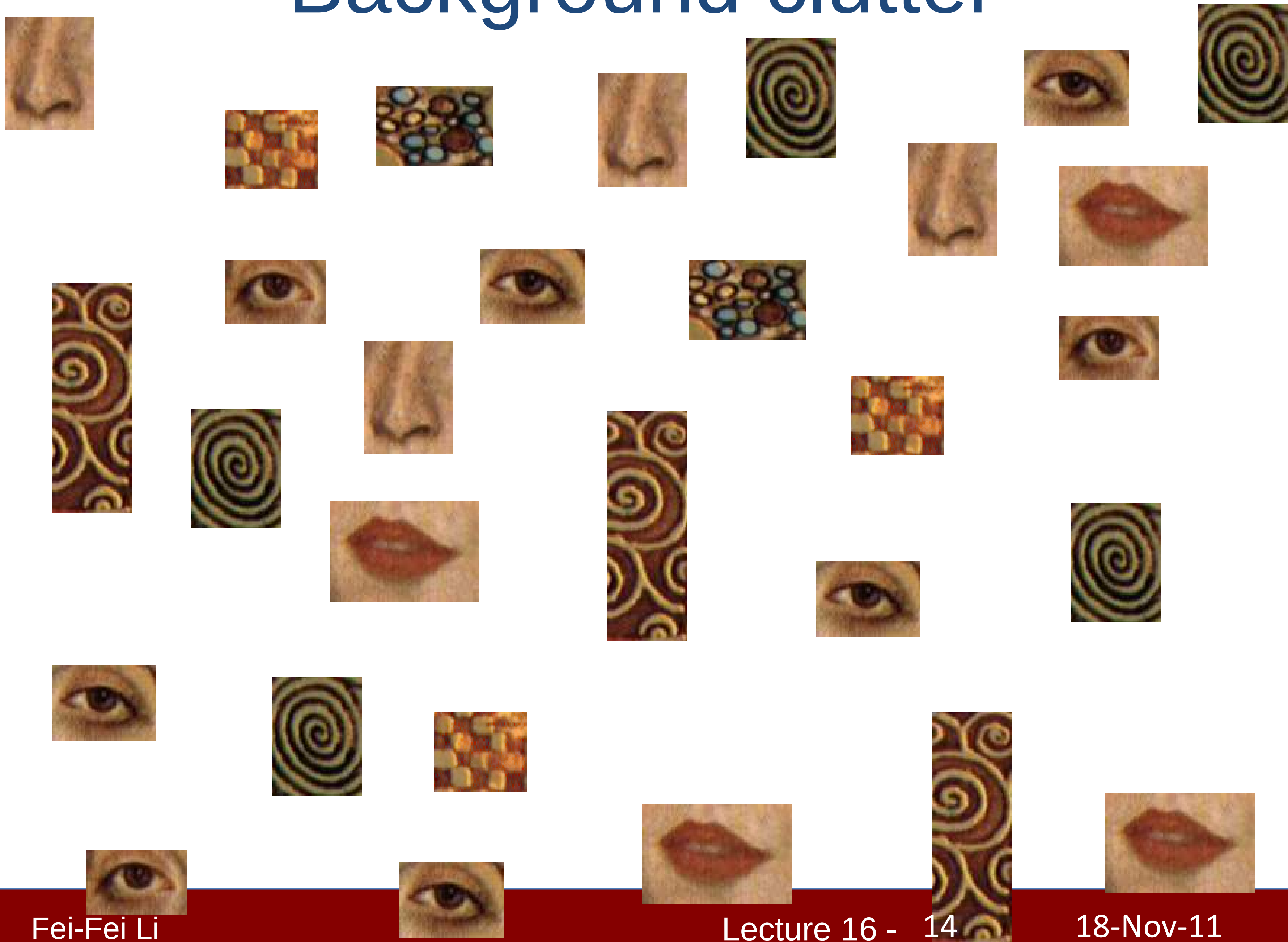
B

C

D

Presence / Absence of Features

Background clutter

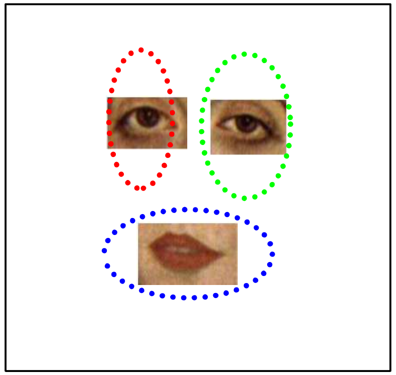


Generative probabilistic model

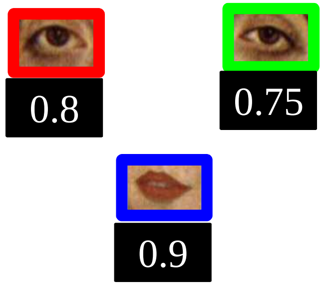
Foreground model

Clutter model

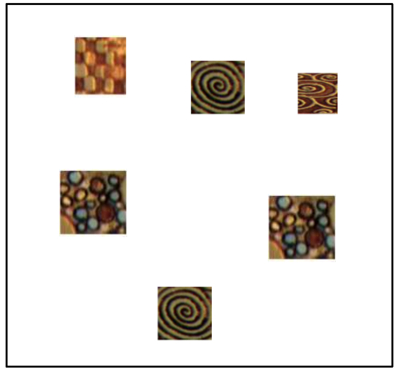
Gaussian shape pdf



Prob. of detection



Uniform shape pdf



\# detections

$p_{\text{Poisson}}(N_1|\lambda_1)$

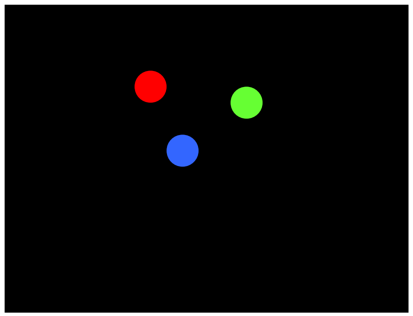
$p_{\text{Poisson}}(N_2|\lambda_2)$

$p_{\text{Poisson}}(N_3|\lambda_3)$

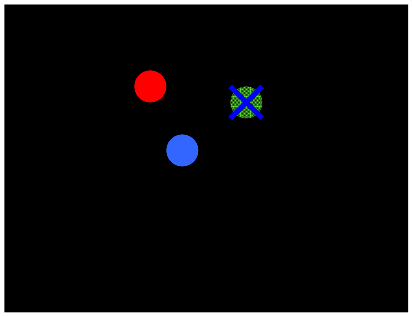
Assumptions: (a) Clutter independent of foreground detections
(b) Clutter detections independent of each other

Example

1. Object Part Positions

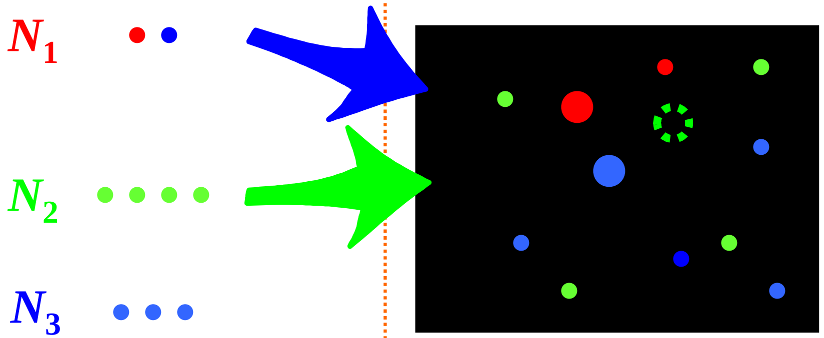


2. Part Absence

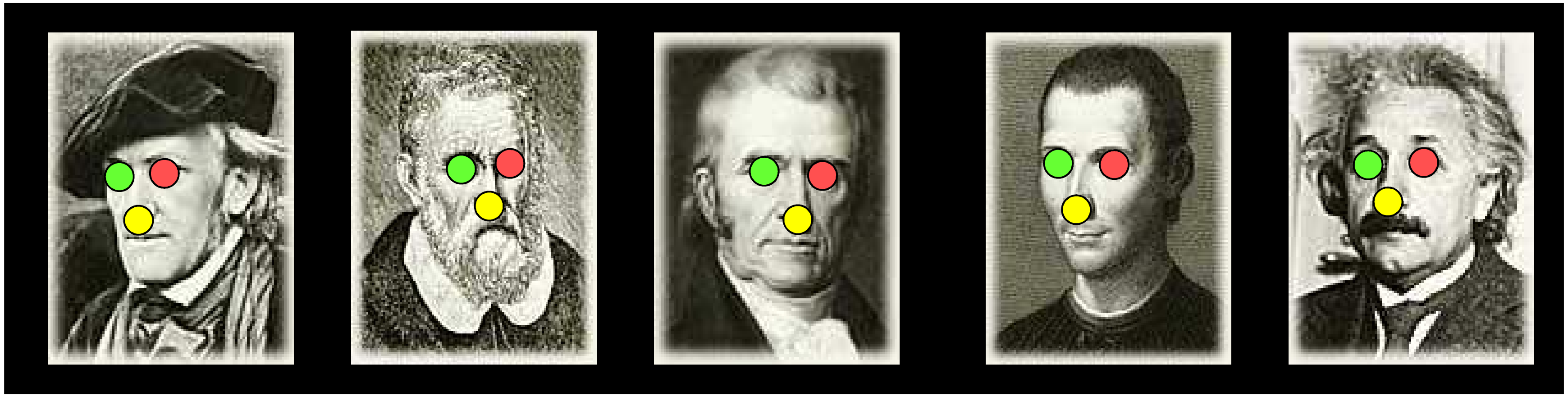


3a. N false detect

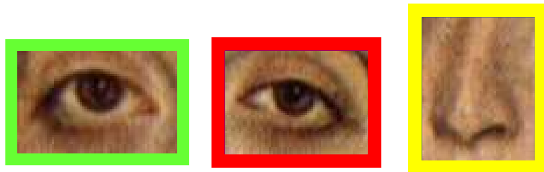
3b. Position f. detect



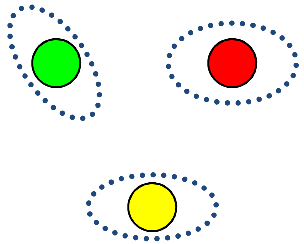
Learning Models `Manually'



- Obtain set of training images
- Choose parts

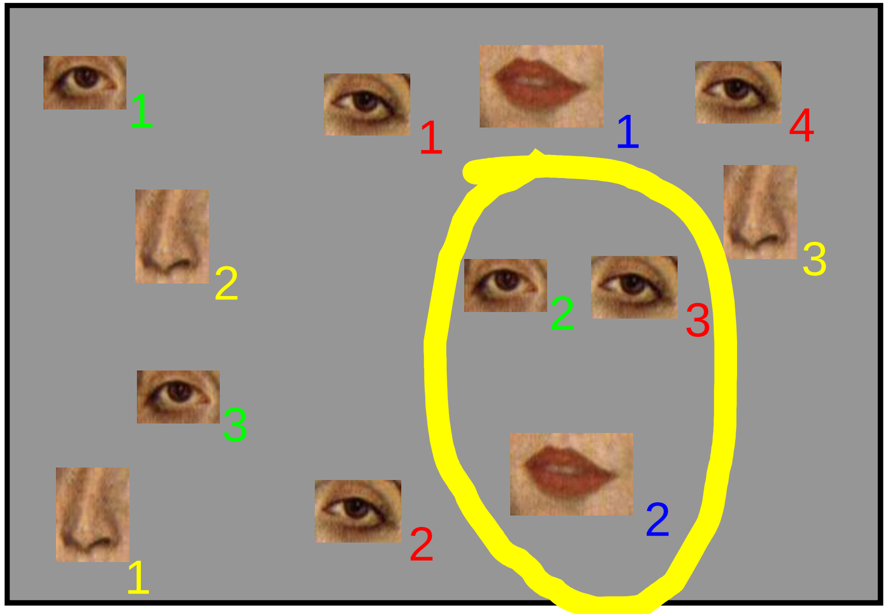


- Label parts by hand, train detectors
- Learn model from labeled parts



Recognition

1. Run part detectors exhaustively over image



$$h = \begin{pmatrix} 0 \ldots N_1 \\ 0 \ldots N_2 \\ 0 \ldots N_3 \\ 0 \ldots N_4 \end{pmatrix} \qquad \text{e.g. } h = \begin{pmatrix} 2 \\ 3 \\ 0 \\ 2 \end{pmatrix}$$

2. Try different combinations of detections in model
   - Allow detections to be missing (occlusion)
3. Pick hypothesis which maximizes:

$$\frac{p(Data \mid Object, Hyp)}{p(Data \mid Clutter, Hyp)}$$

4. If ratio is above threshold then, instance detected

So far…..

- Representation
  - Joint model of part locations
  - Ability to deal with background clutter and occlusions
- Learning
  - Manual construction of part detectors
  - Estimate parameters of shape density
- Recognition
  - Run part detectors over image
  - Try combinations of features in model
  - Use efficient search techniques to make fast

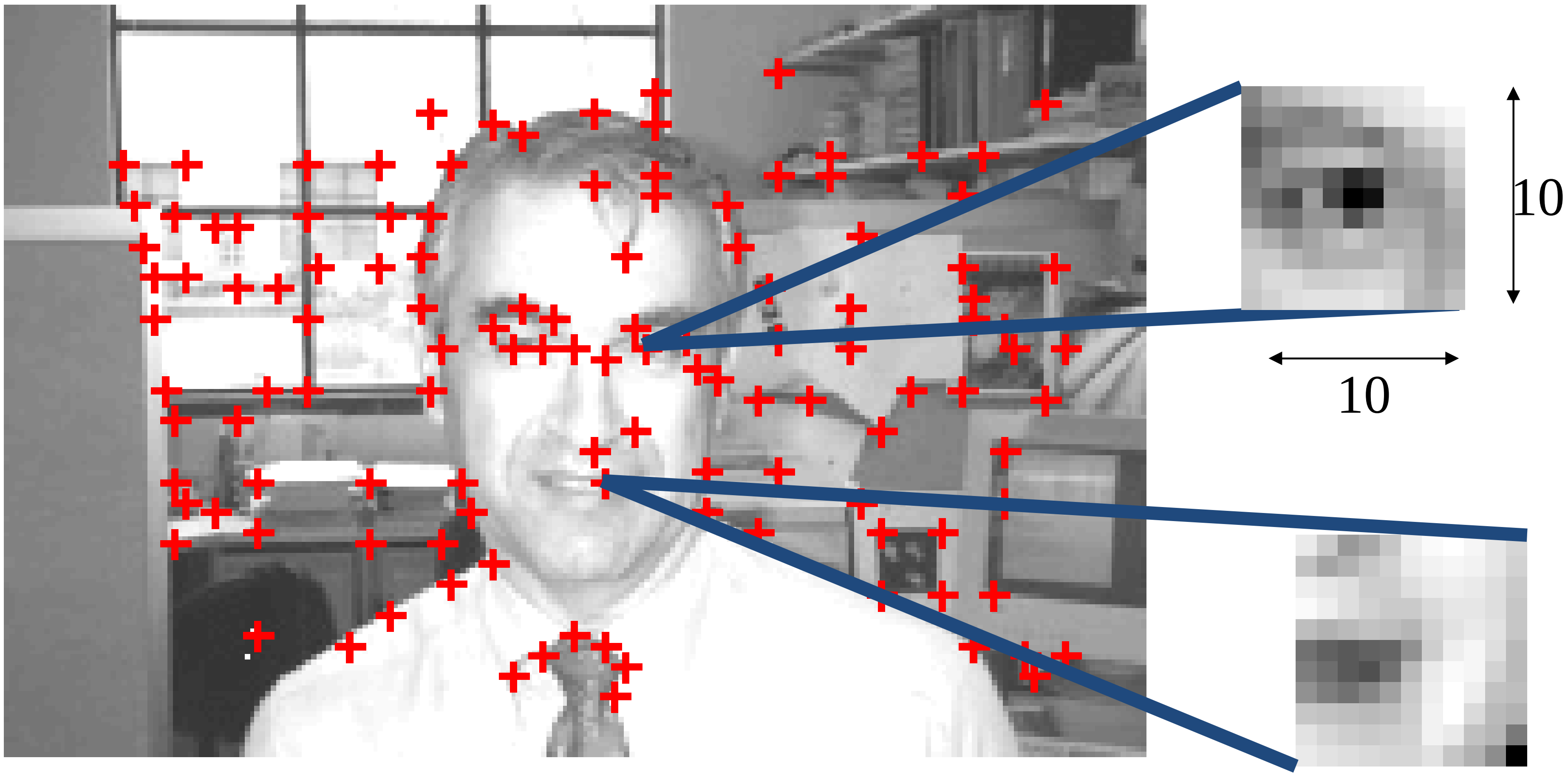
Unsupervised Learning

Weber & Welling et. al.

(Semi) Unsupervised learning

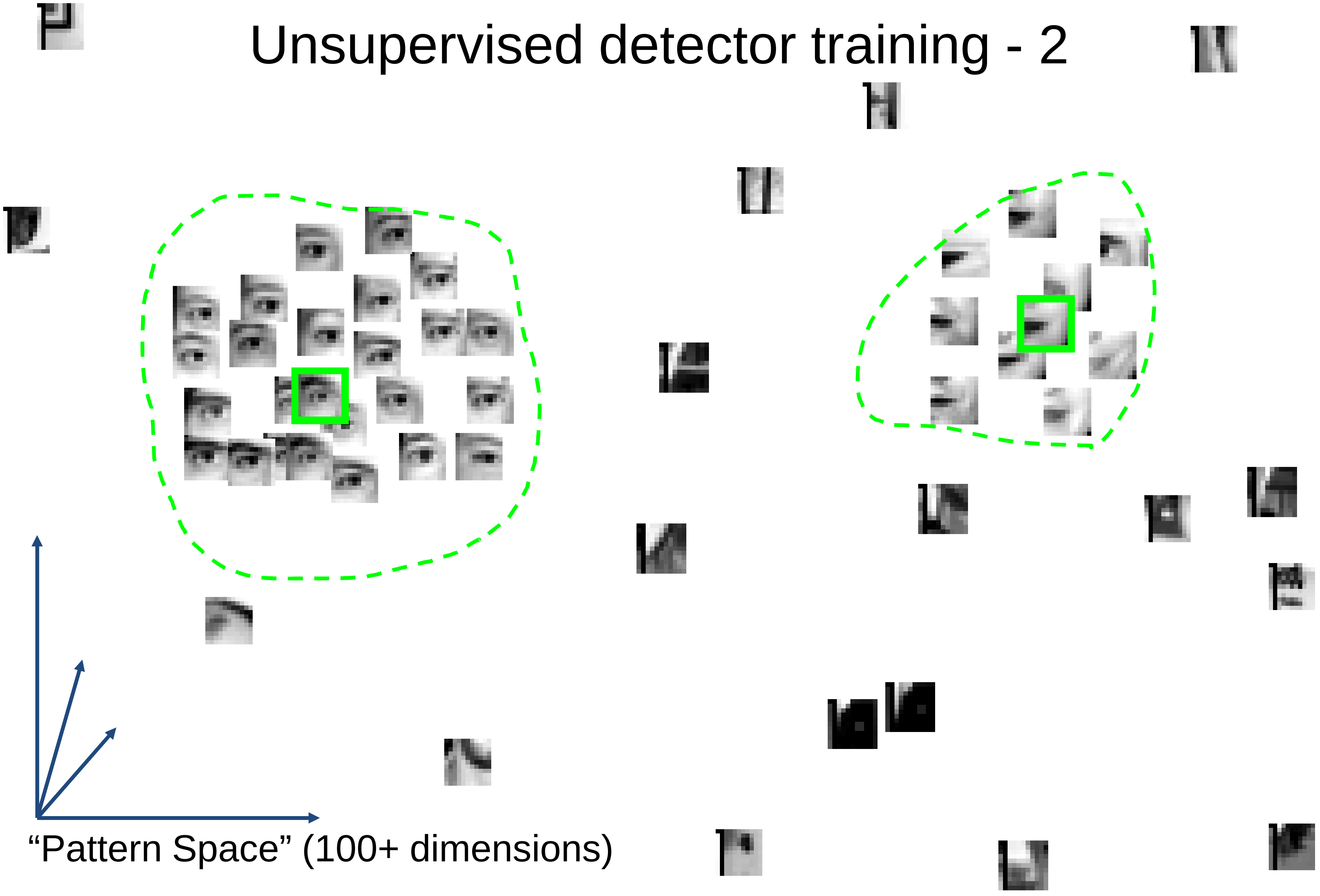
•Know if image contains object or not
•But no segmentation of object or manual selection of features

Unsupervised detector training - 1



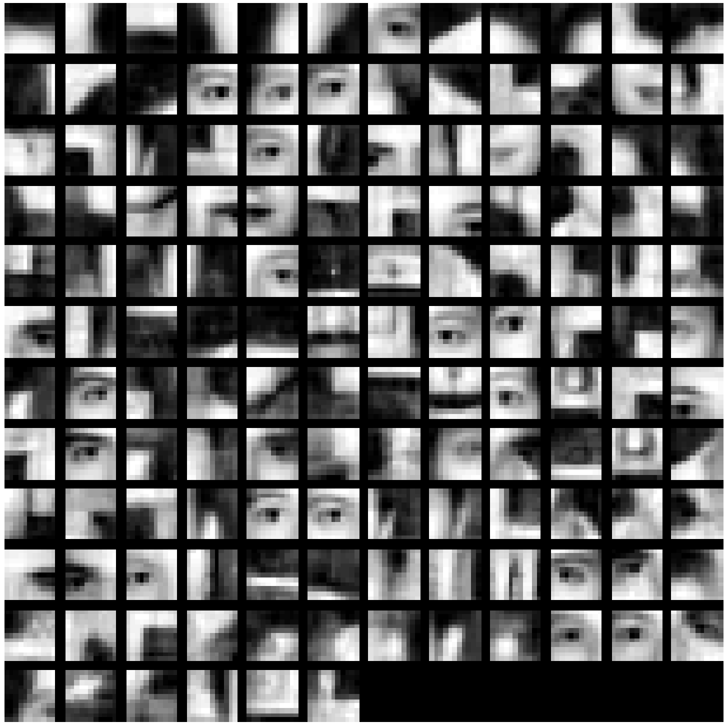
- Highly textured neighborhoods are selected automatically
- produces 100-1000 patterns per image

Unsupervised detector training - 2



Unsupervised detector training - 3

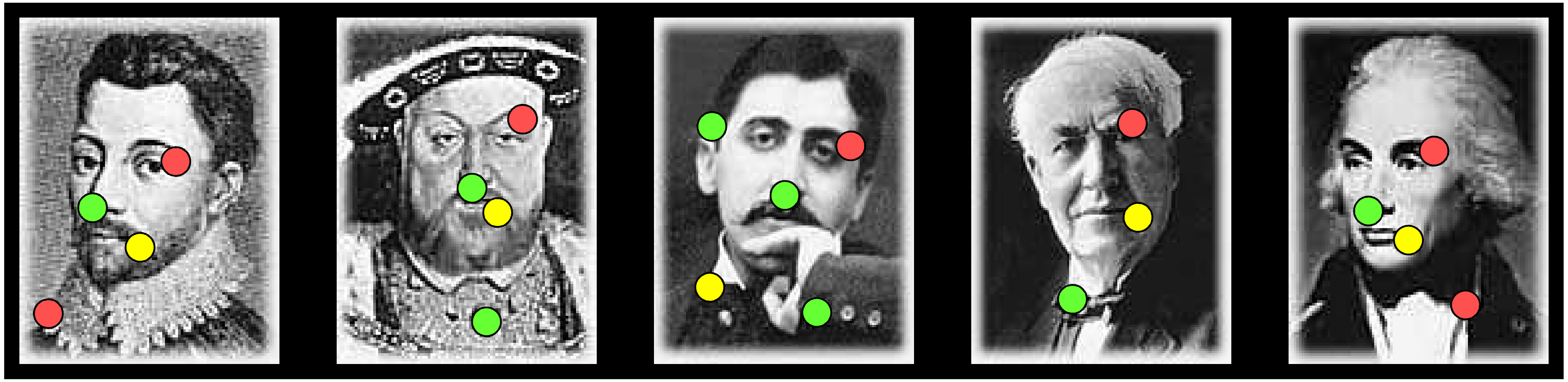
100-1000 images



~100 detectors

Learning

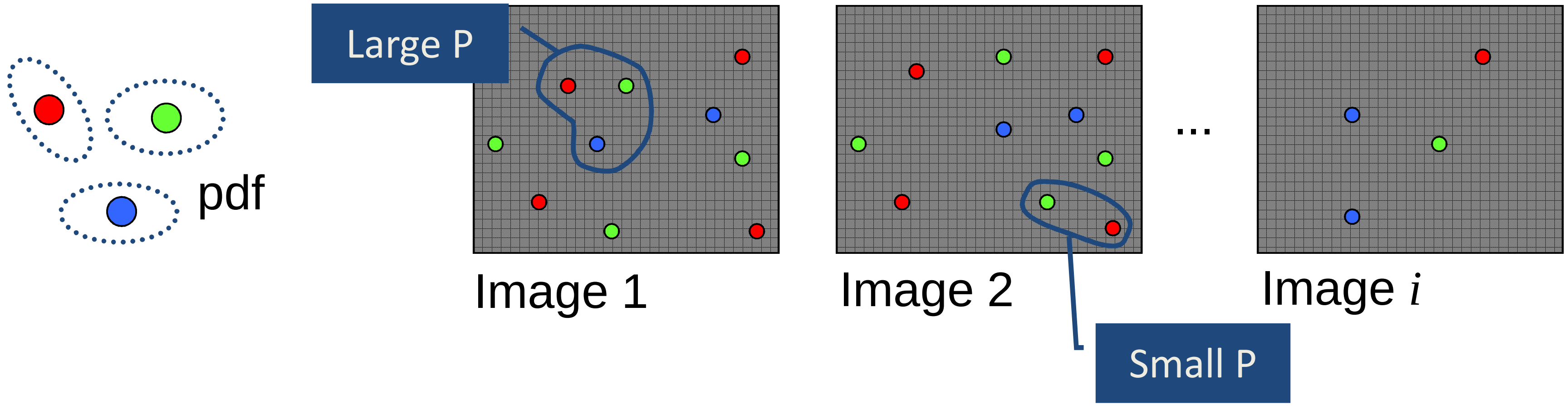
- Take training images. Pick set of detectors. Apply detectors.
- Task: Estimation of model parameters
- Chicken and Egg type problem, since we initially know neither:
  - Model parameters
  - Assignment of regions to foreground / background
- Let the assignments be a hidden variable and use EM algorithm to learn them and the model parameters



ML using EM

1. Current estimate

2. Assign probabilities to constellations

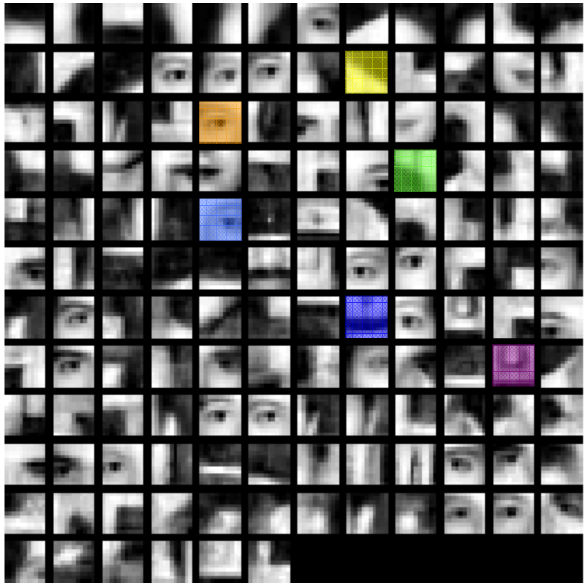


3. Use probabilities as weights to re-estimate parameters. Example: μ

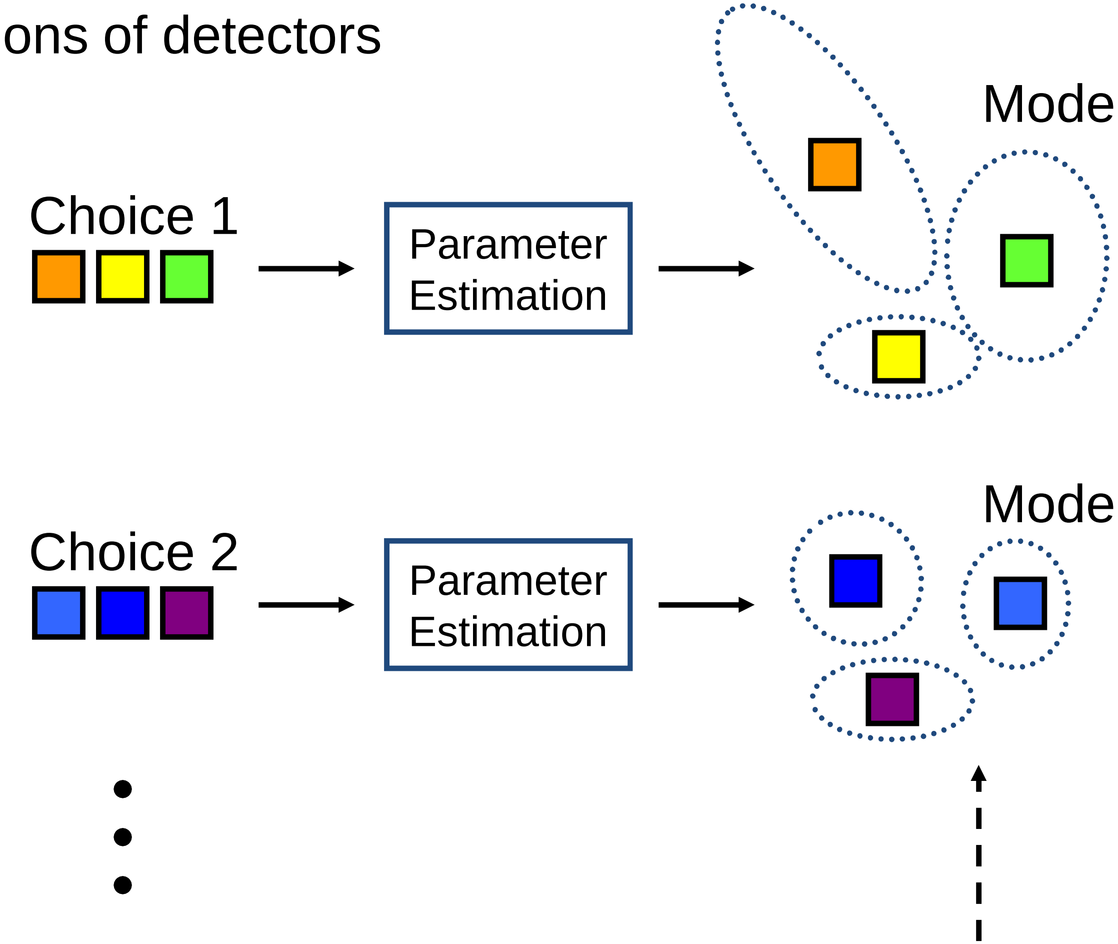
new estimate of μ

Detector Selection

- Try out different combinations of detectors (Greedy search)



Detectors (≈100)



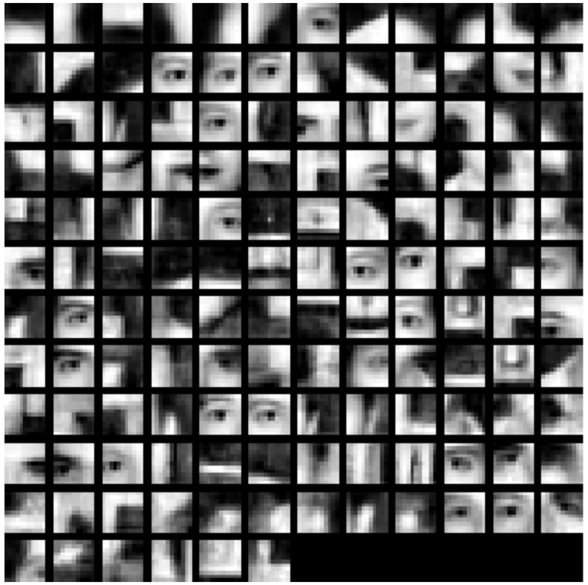
Predict / measure model performance (validation set or directly from model)

Frontal Views of Faces

- 200 Images (100 training, 100 testing)
- 30 people, different for training and testing

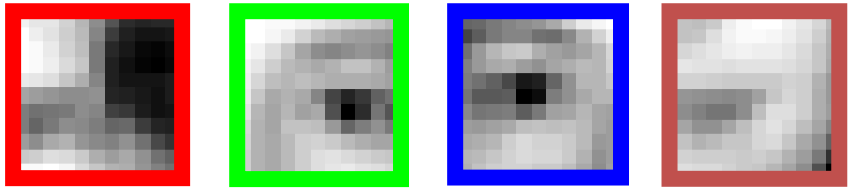
Learned face model

Pre-selected Parts

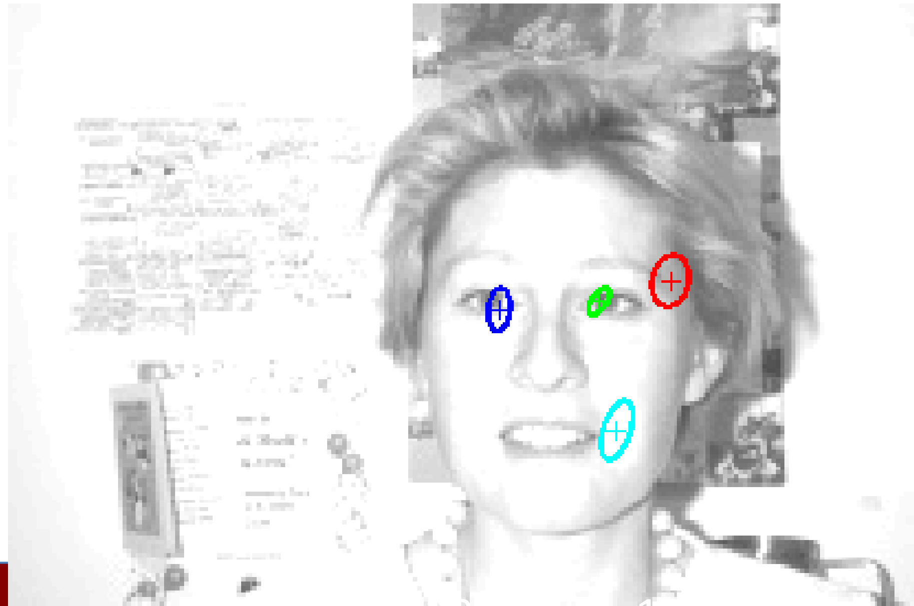


Test Error: 6% (4 Parts)

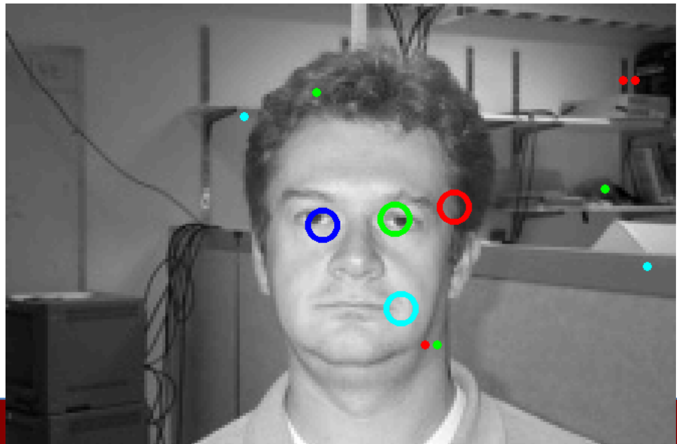
Parts in Model



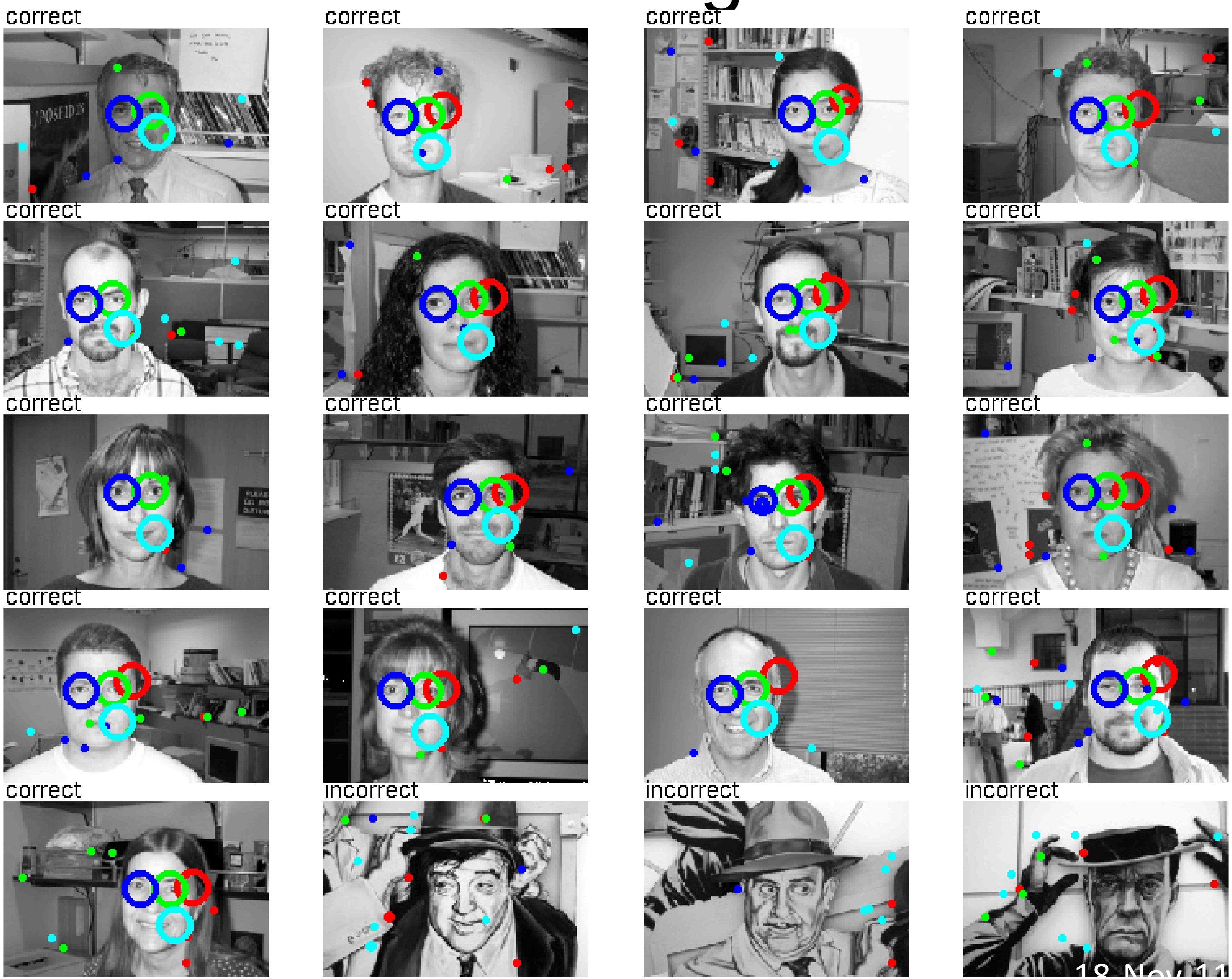
Model Foreground pdf



Sample Detection

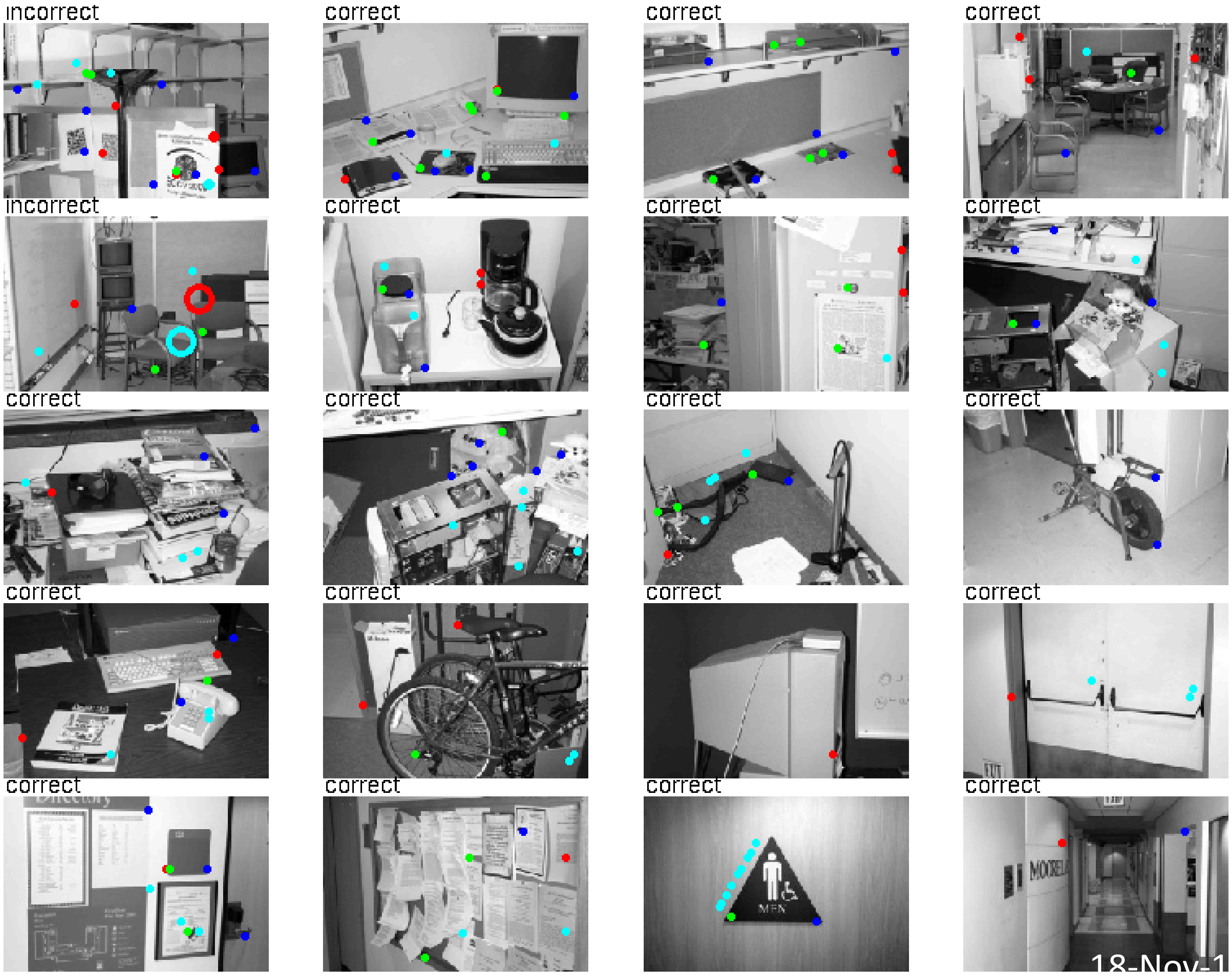


Face images



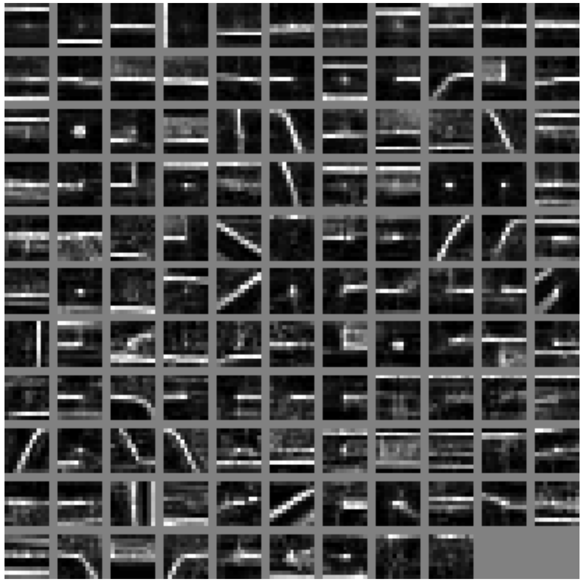
correct
correct
correct
correct
correct
correct
correct
correct
correct
correct
correct
correct
correct
correct
correct
correct
correct
incorrect
incorrect
incorrect

Background images



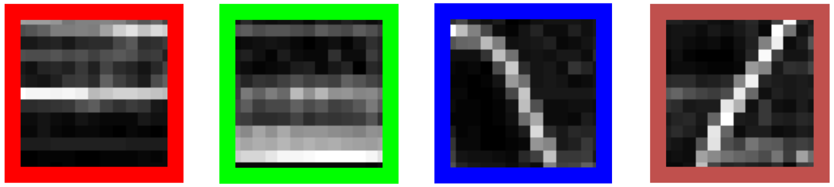
Car from Rear

Preselected Parts

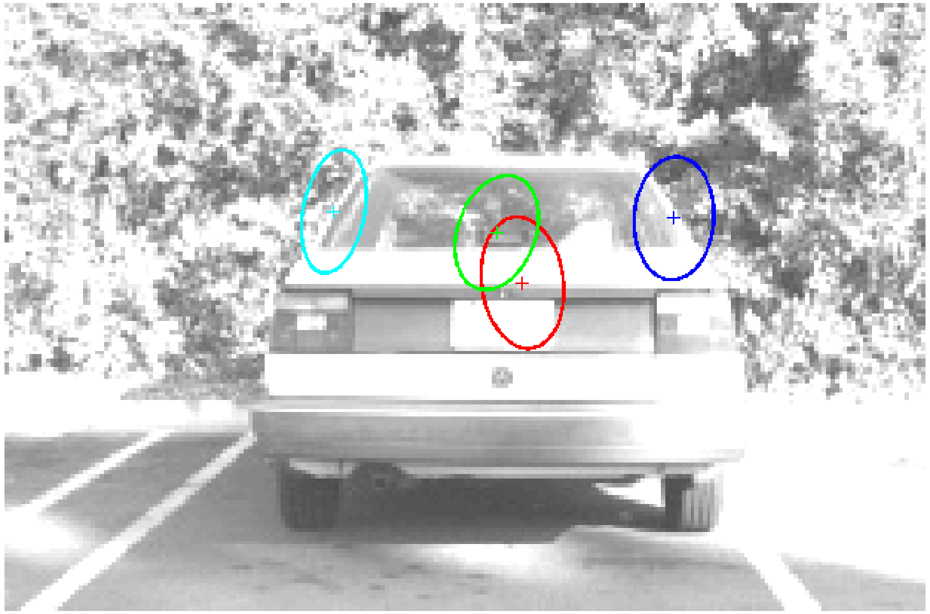


Test Error: 13% (5 Parts)

Parts in Model



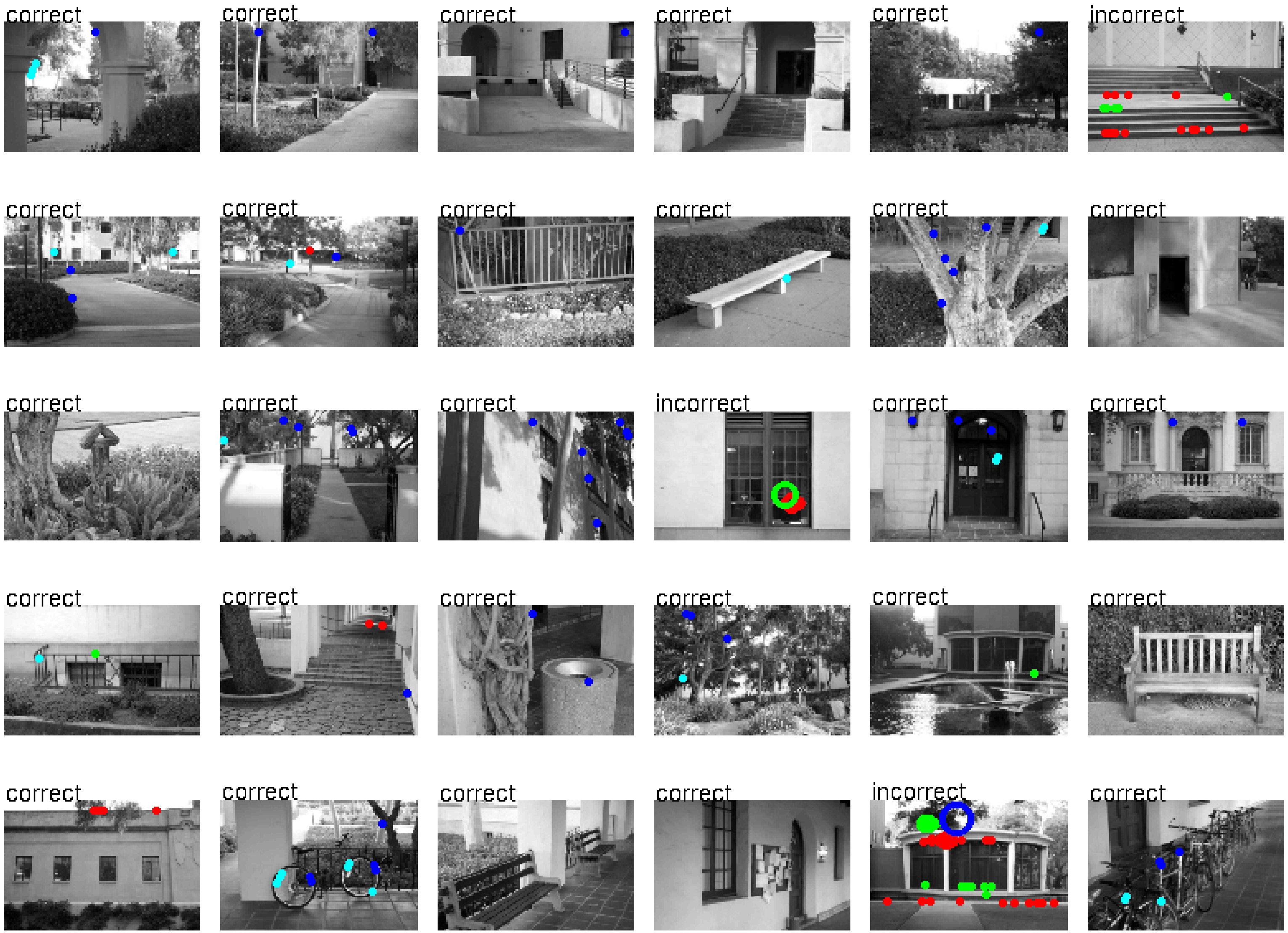
Model Foreground pdf



Sample Detection

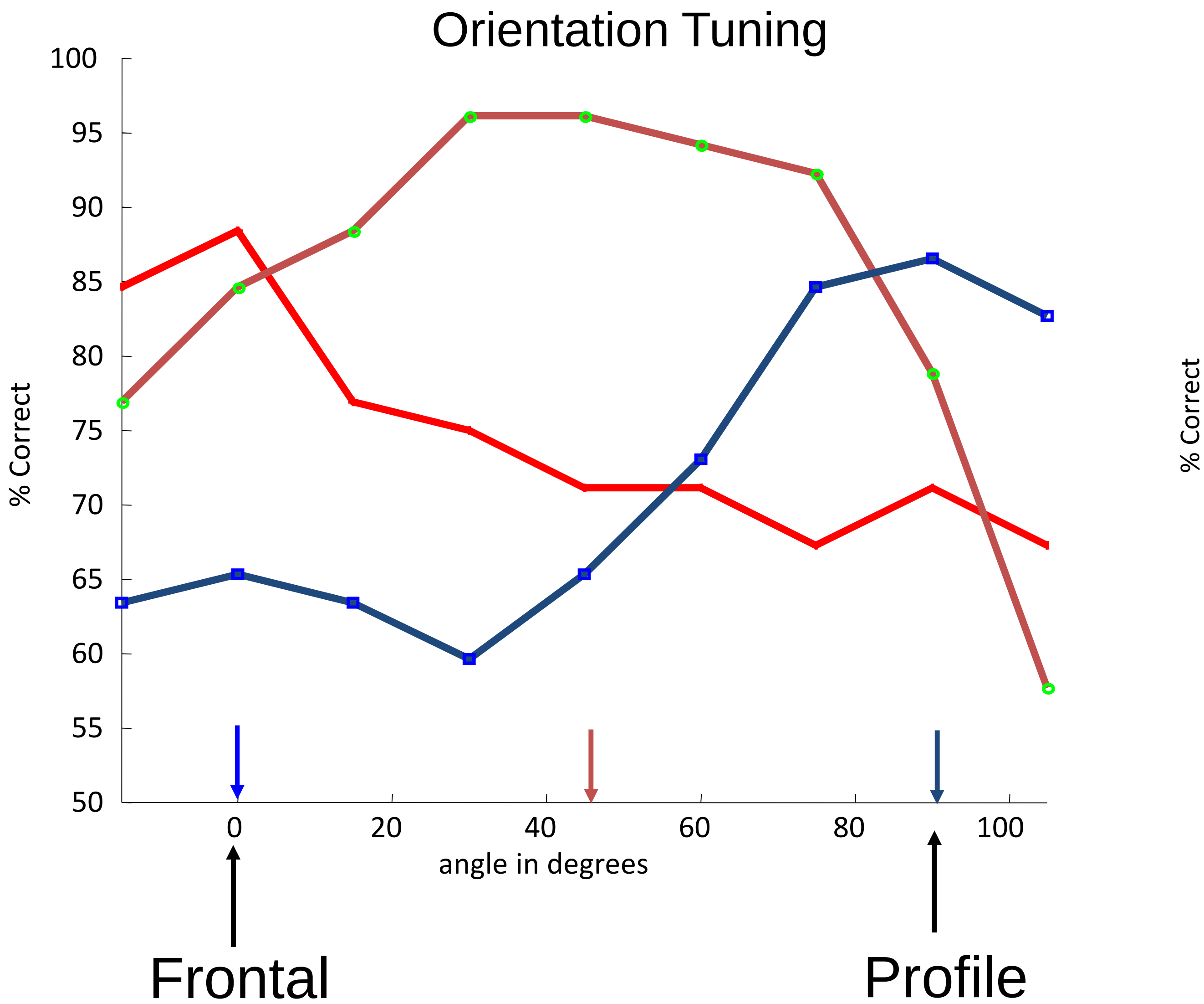
Detections of Cars

Background Images



3D Object recognition – Multiple mixture components

3D Orientation Tuning



So far (2)…..

- Representation
  - Multiple mixture components for different viewpoints
- Learning
  - Now semi-unsupervised
  - Automatic construction and selection of part detectors
  - Estimation of parameters using EM
- Recognition
  - As before

- Issues:

  -Learning is slow (many combinations of detectors)

  -Appearance learnt first, then shape

Issues

- Speed of learning
  - Slow (many combinations of detectors)
- Appearance learnt first, then shape
  - Difficult to learn part that has stable location but variable appearance
  - Each detector is used as a cross-correlation filter, giving a hard definition of the part's appearance

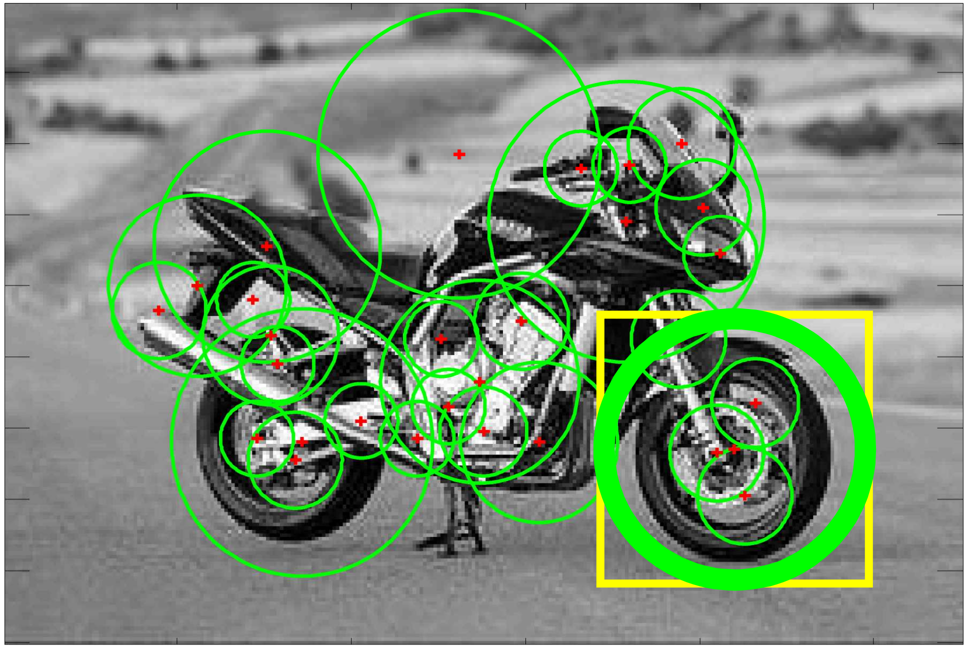
- Would like a fully probabilistic representation of the object

Object categorization

Fergus et. al.

CVPR ’03, IJCV ‘06

Detection & Representation of regions



- Find regions within image
- Use salient region operator (Kadir & Brady 01)

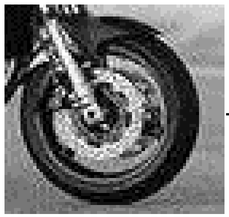
Location

(x,y) coords. of region centre

Scale

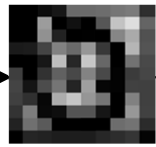
Radius of region (pixels)

Appearance



Normalize

11x11 patch



Projection onto PCA basis

$$\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{15} \end{pmatrix}$$

Gives representation of appearance in low-dimensional vector space

Motorbikes example

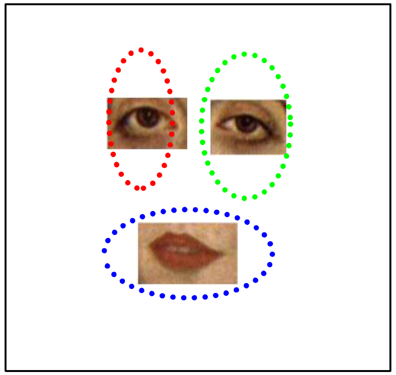
- Kadir & Brady saliency region detector

Generative probabilistic model (2)

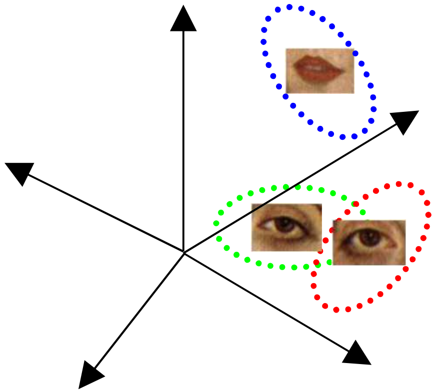
Foreground model

based on Burl, Weber et al. [ECCV '98, '00]

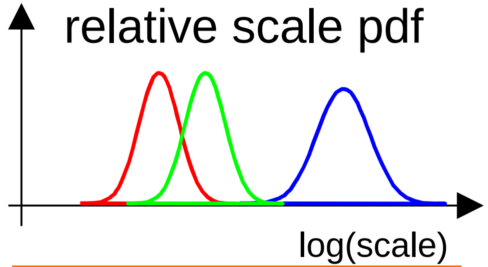
Gaussian shape pdf



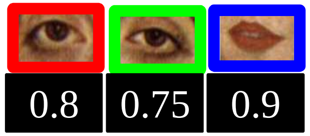
Gaussian part appearance pdf



Gaussian
relative scale pdf

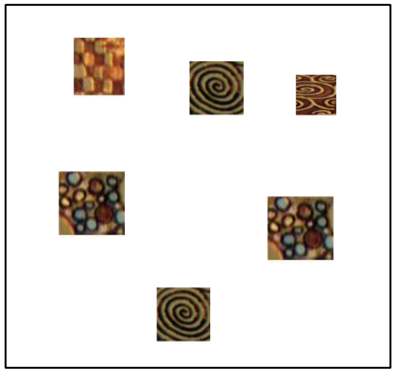


Prob. of detection

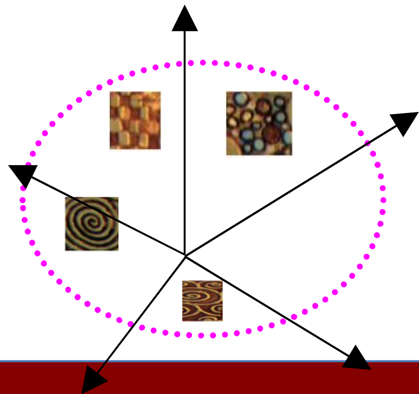


Clutter model

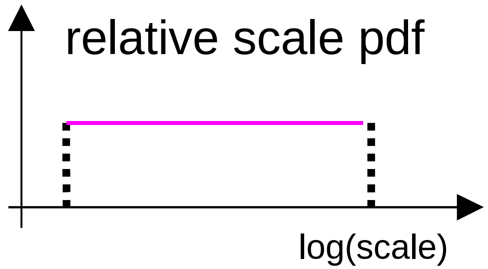
Uniform shape pdf



Gaussian background
appearance pdf



Uniform
relative scale pdf

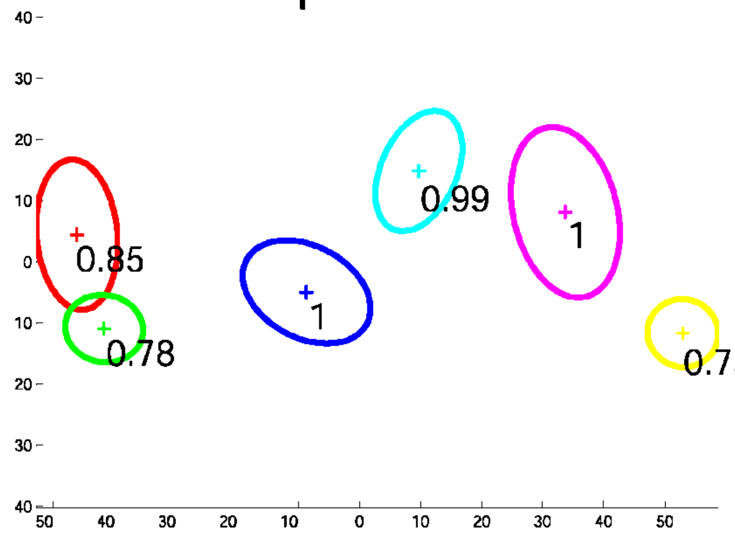


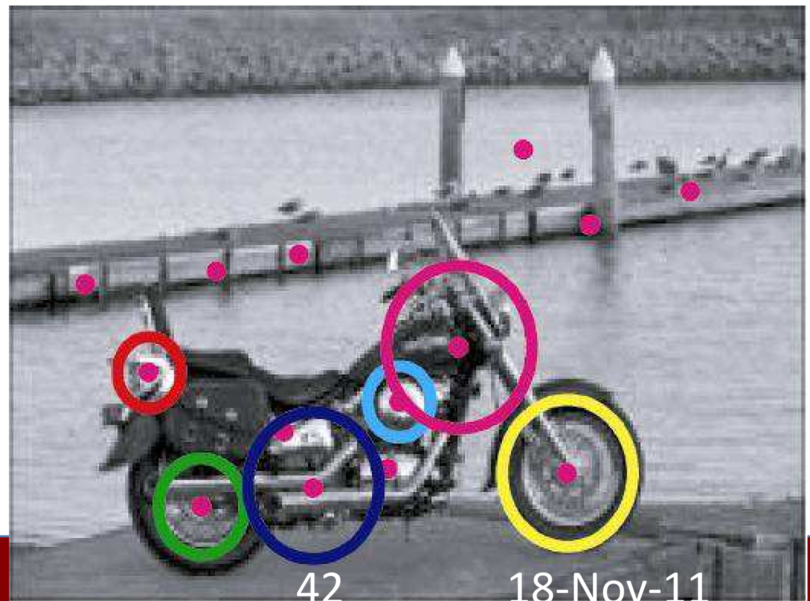
Poission pdf on #
detections

Motorbikes

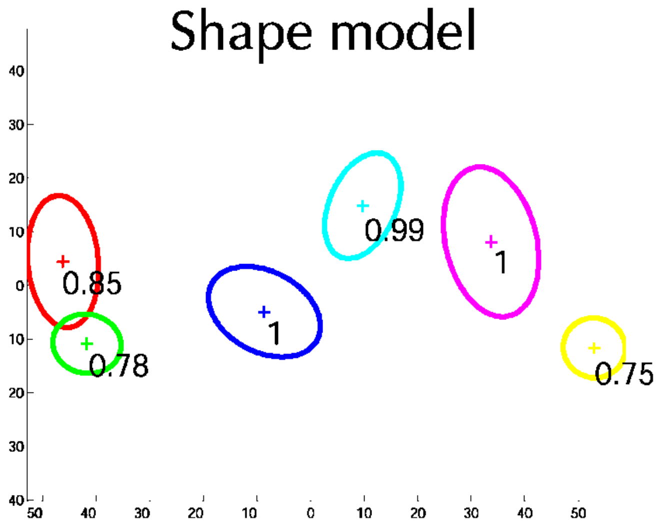
Samples from appearance model

Shape model



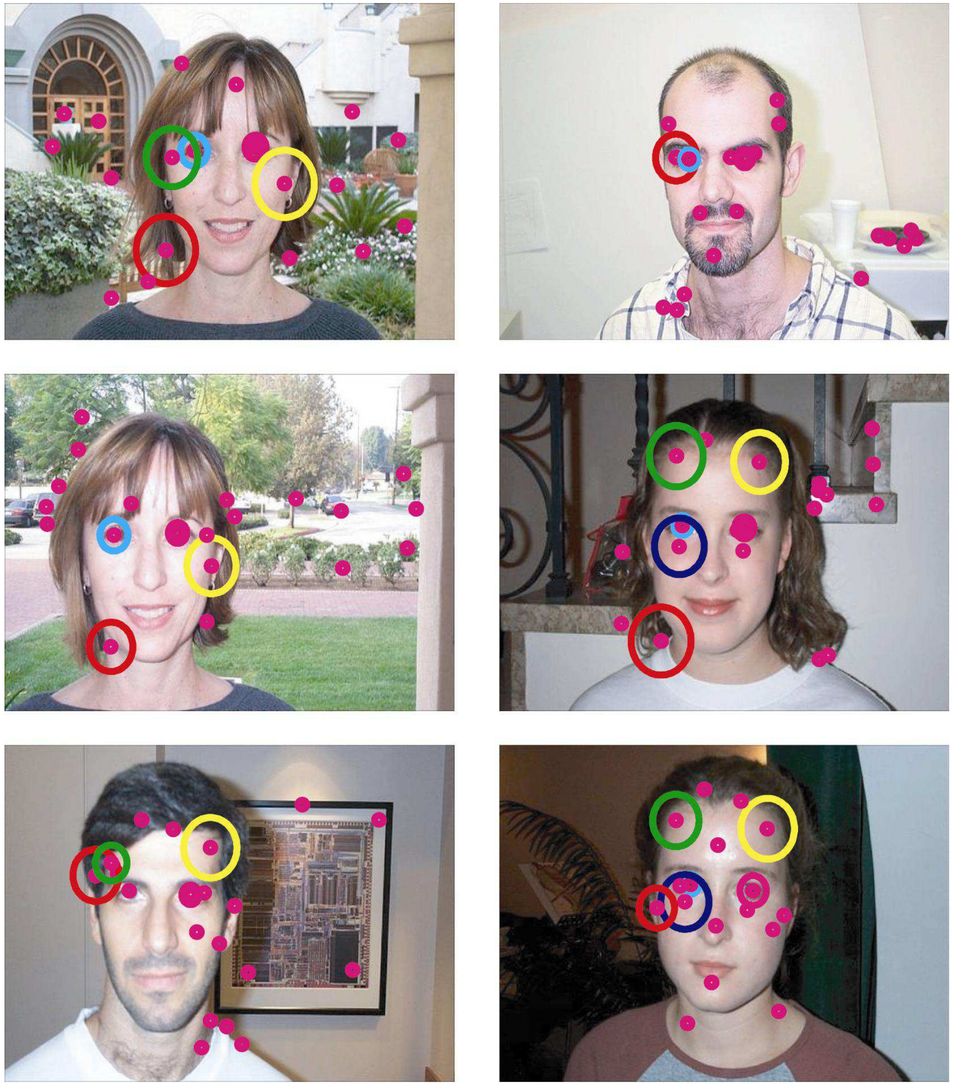


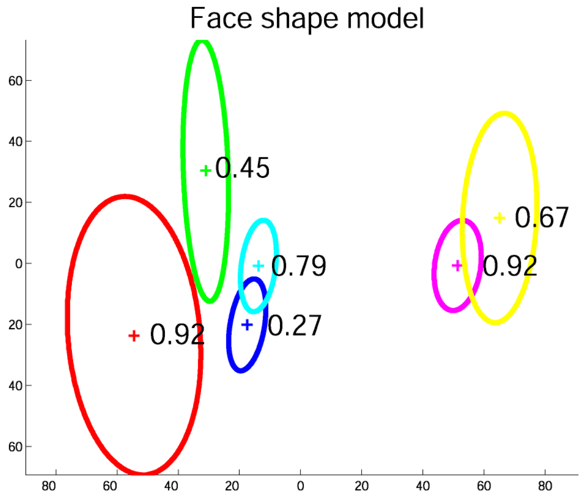
Recognized Motorbikes

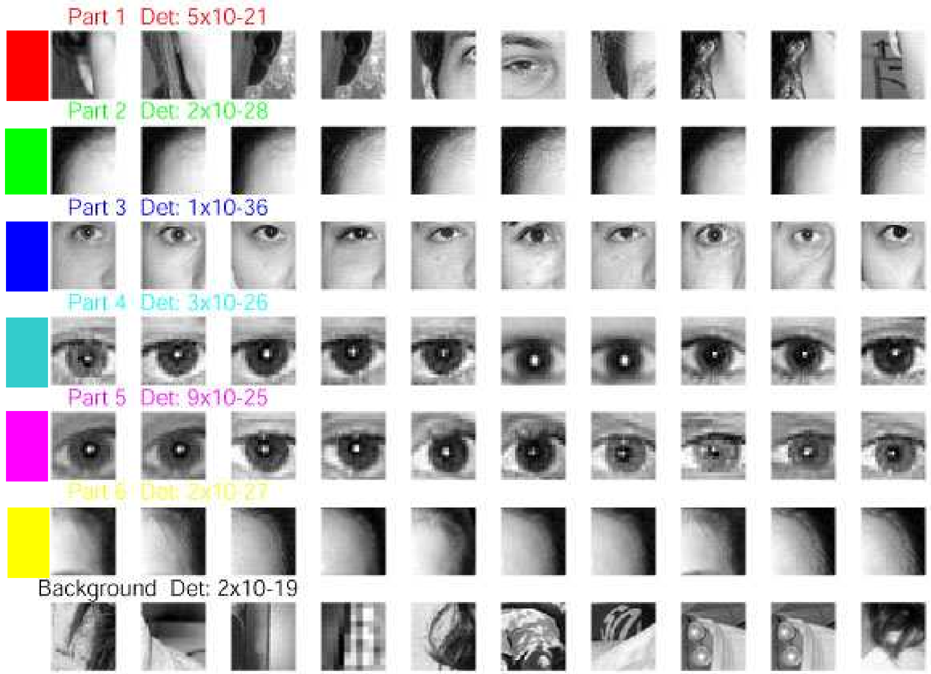


Background images evaluated with motorbike model

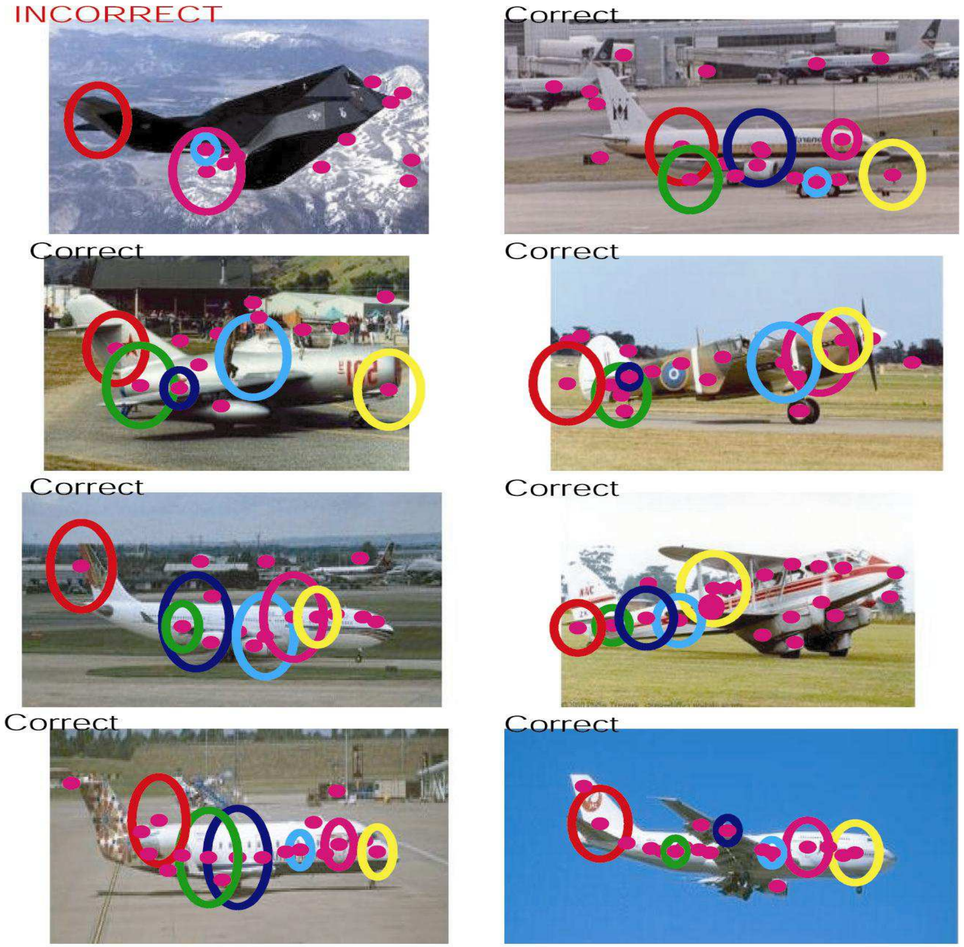
Frontal faces

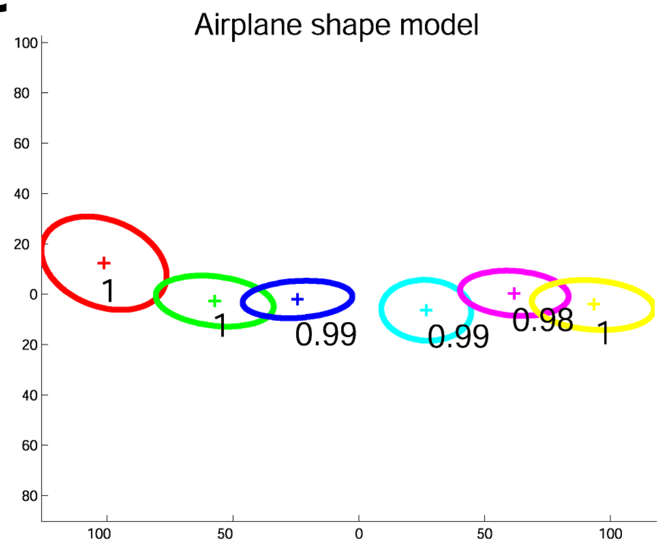


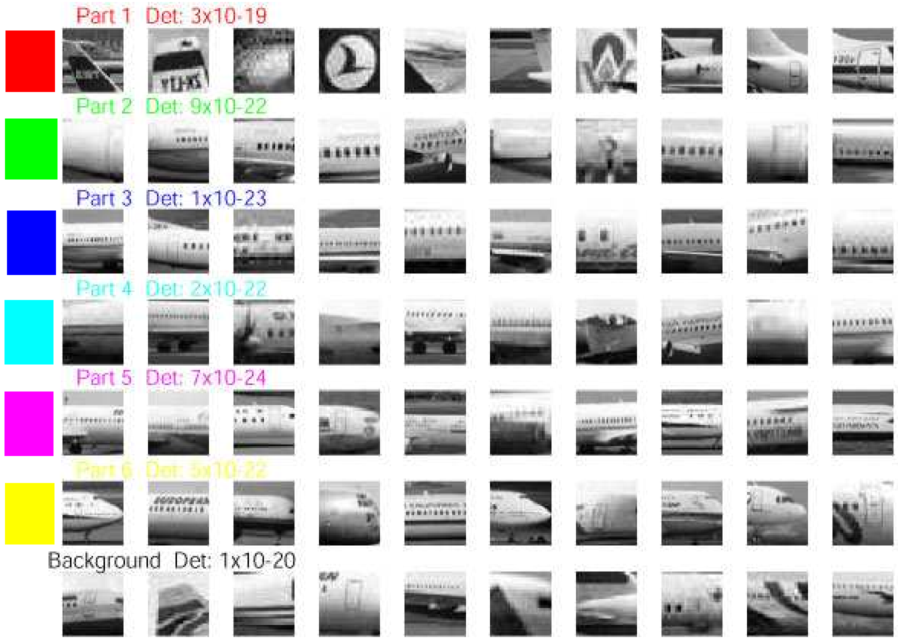




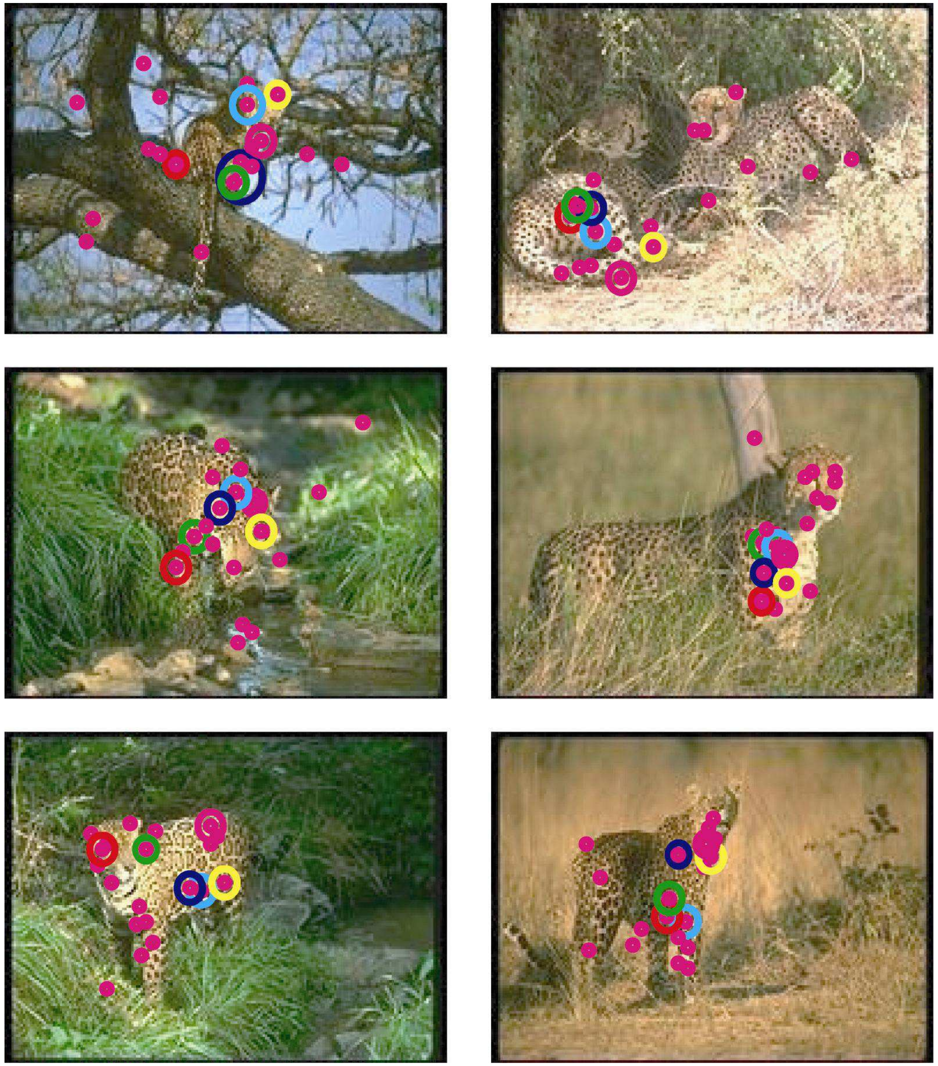
Airplanes

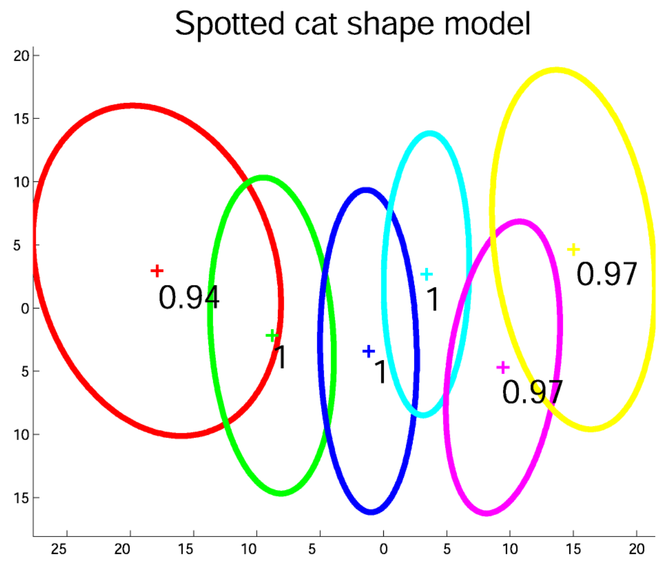

INCORRECT
Correct
Correct
Correct
Correct
Correct
Correct
Correct

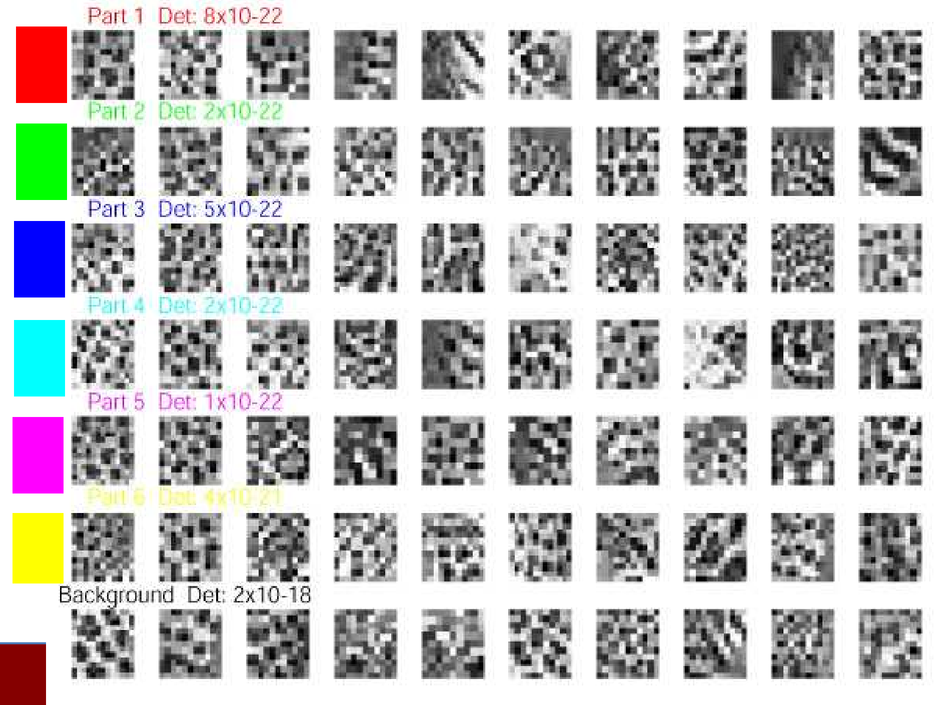

Airplane shape model
100
80
60
40
20
0
20
40
60
80
100
50
0
50
100
1
1
0.99
0.99
0.98
1


Part 1 Det: 3x10-19
Part 2 Det: 9x10-22
Part 3 Det: 1x10-23
Part 4 Det: 2x10-22
Part 5 Det: 7x10-24
Part 6 Det: 5x10-22
Background Det: 1x10-20

Spotted cats







Summary of results

Dataset	Fixed scale experiment	Scale invariant experiment
Motorbikes	7.5	6.7
Faces	4.6	4.6
Airplanes	9.8	7.0
Cars (Rear)	15.2	9.7
Spotted cats	10.0	10.0

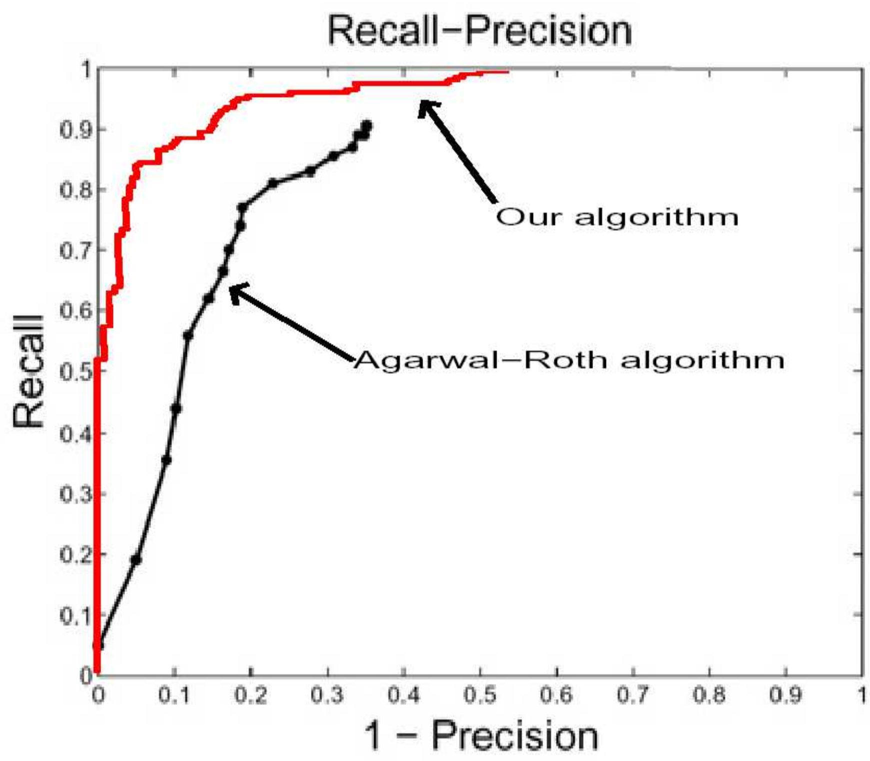
% equal error rate

Note: Within each series, same settings used for all datasets

Comparison to other methods

Dataset	Ours	Others	
Motorbikes	7.5	16.0	Weber et al. [ECCV '00]
Faces	4.6	6.0	Weber
Airplanes	9.8	32.0	Weber
Cars (Side)	11.5	21.0	Agarwal Roth [ECCV '02]

% equal error rate



Why this design?

- Generic features seem to well in finding consistent parts of the object

- Some categories perform badly – different feature types needed

- Why PCA representation?
  - Tried ICA, FLD, Oriented filter responses etc.
  - But PCA worked best

- Fully probabilistic representation lets us use tools from machine learning community

S. Savarese, 2003

P. Bueget, 1562

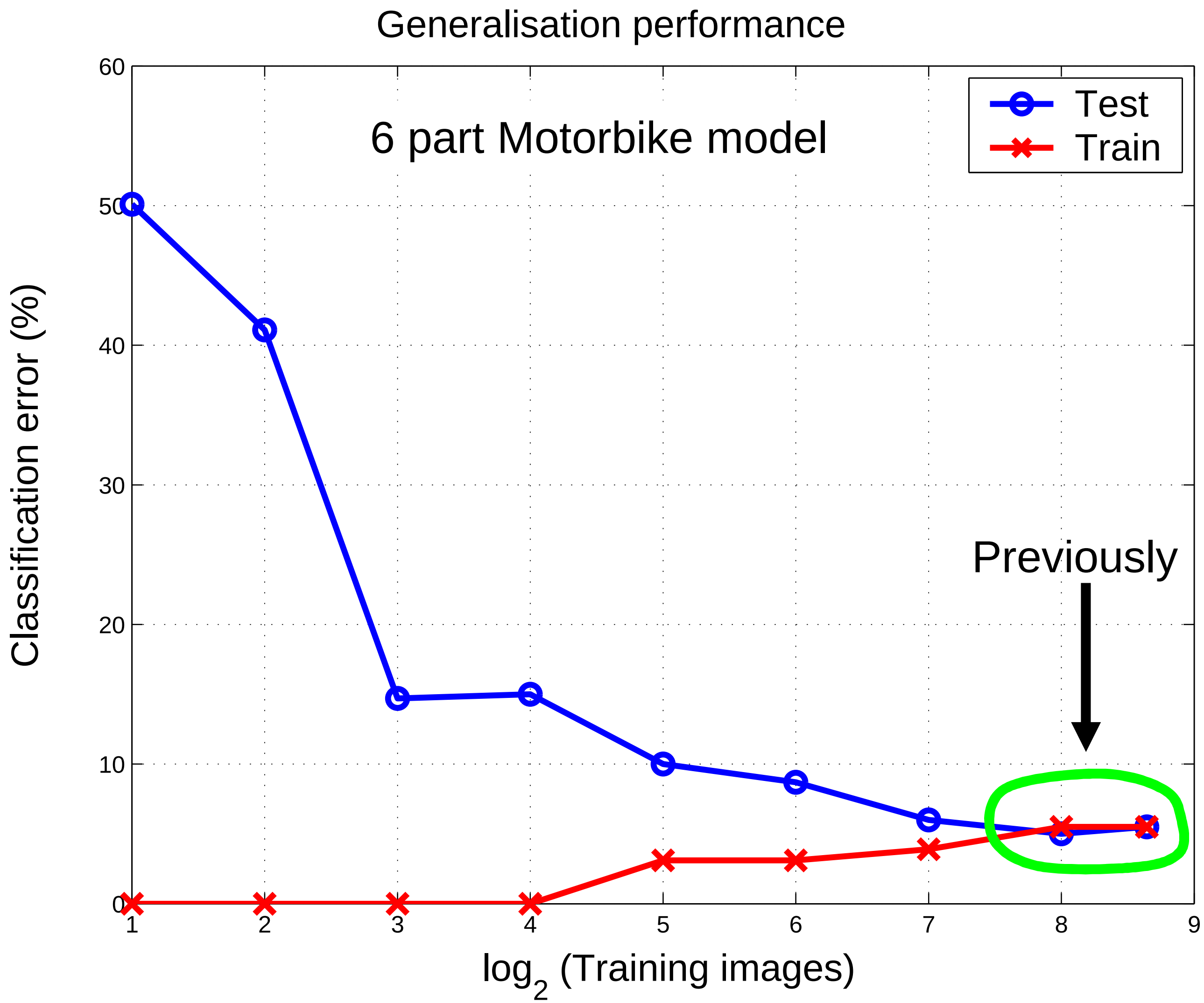
One-Shot learning

Fei-Fei et. al.

ICCV ’03, PAMI ‘06

Algorithm	Training Examples	Categories
Burl, et al. Weber, et al. Fergus, et al.	200 ~ 400	Faces, Motorbikes, Spotted cats, Airplanes, Cars
Viola et al.	~10,000	Faces
Schneiderman, et al.	~2,000	Faces, Cars
Rowley et al.	~500	Faces

Number of training examples



How do we do better than what statisticians have told us?

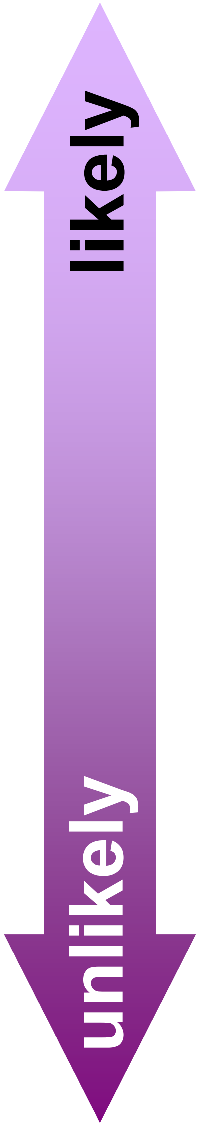
- Intuition 1: use Prior information
- Intuition 2: make best use of training information

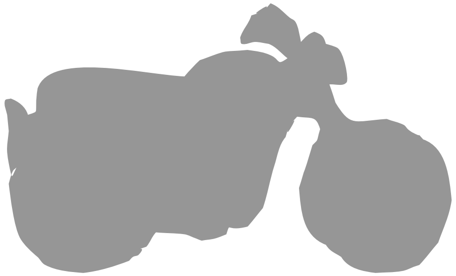
Prior knowledge: means

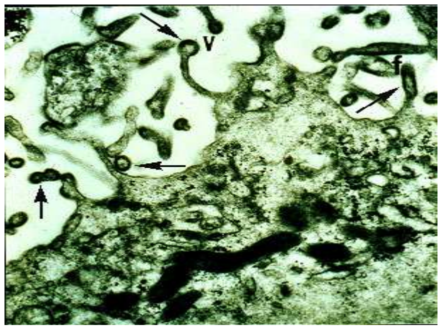
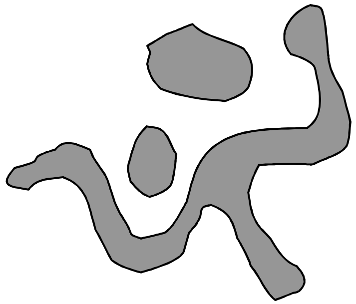
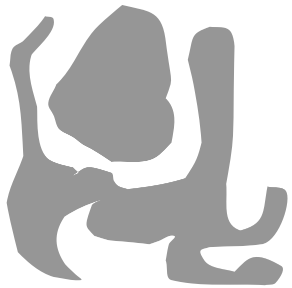
Appearance

Shape







Bayesian framework

$\boxed{\text{P(object | test, train)}}$ vs. P(clutter | test, train)

Bayes Rule

$$\boxed{p(\text{test} \mid \text{object}, \ \text{train})} \ p(\text{object})$$

Expansion by parametrization

$$\int p(\text{test} \mid \theta, \text{object}) \ \boxed{p(\theta \mid \text{object}, \text{train})} \ d\theta$$

Bayesian framework

$\boxed{\text{P(object | test, train)}}$ vs. P(clutter | test, train)

Bayes Rule

$$\boxed{p(\text{test} \mid \text{object, train})}\; p(\text{object})$$

Expansion by parametrization

$$\int p(\text{test} \mid \theta, \text{object}) \boxed{p(\theta \mid \text{object}, \text{train})}\, d\theta$$

Previous Work: $\boxed{\delta\left(\theta^{\mathrm{ML}}\right)}$

Bayesian framework

$\boxed{\text{P(object | test, train)}}$ vs. P(clutter | test, train)

$\xrightarrow{\text{Bayes Rule}}$

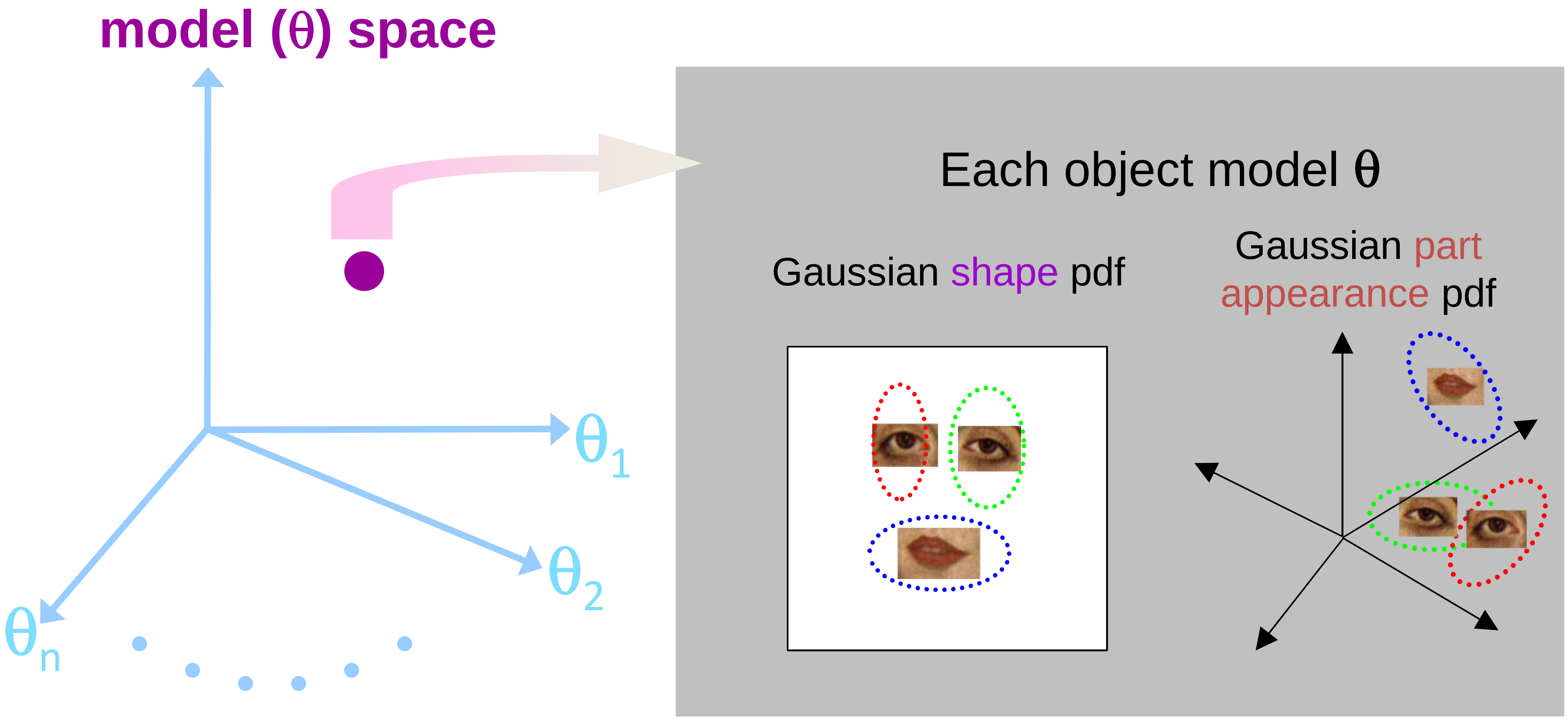
$$\boxed{p(\text{test} \mid \text{object, train})}\, p(\text{object})$$

$\xrightarrow{\textbf{Expansion by parametrization}}$

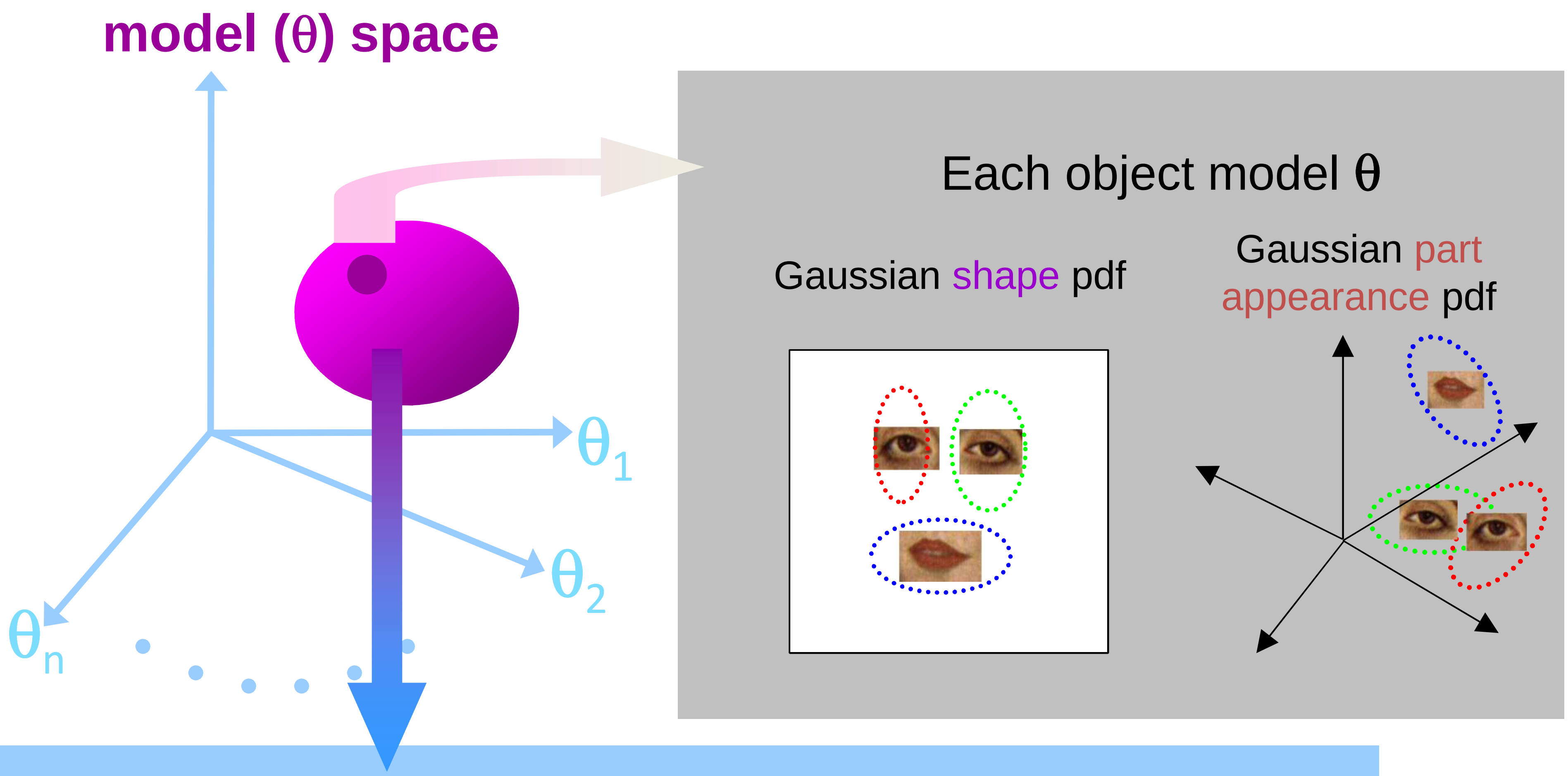
$$\int p(\text{test} \mid \theta, \text{object}) \boxed{p(\theta \mid \text{object}, \text{train})}\, d\theta$$

One-Shot learning: $\boxed{p(\text{train} \mid \theta, \text{object})\, p(\theta)}$

Model Structure



Model Structure



model distribution: p(θ)

- **conjugate distribution of p(train|θ,object)**

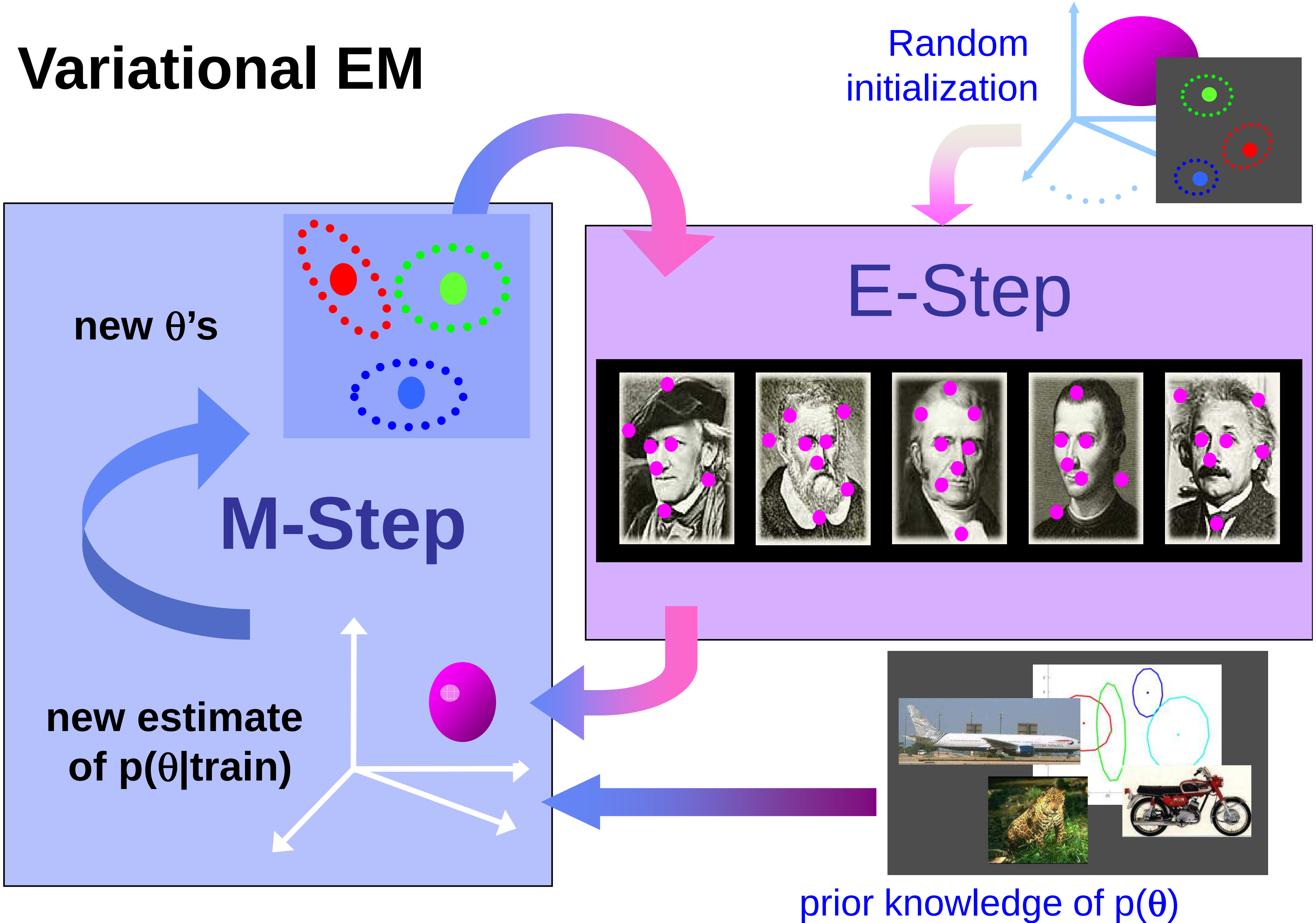
Learning Model Distribution

$$p(\theta|\text{object, train}) \propto p(\text{train}|\theta, \text{object})\, p(\theta)$$

- use Prior information
- Bayesian learning
  - marginalize over theta

❖ **Variational EM (Attias, Hinton, Minka, etc.)**

Variational EM



Experiments

Training:

1- 6 randomly

drawn images

Testing:

50 fg/ 50 bg images

object present/absent

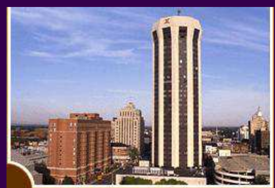
Datasets

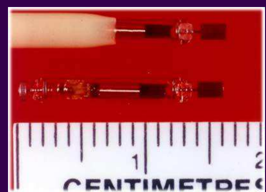
faces

airplanes

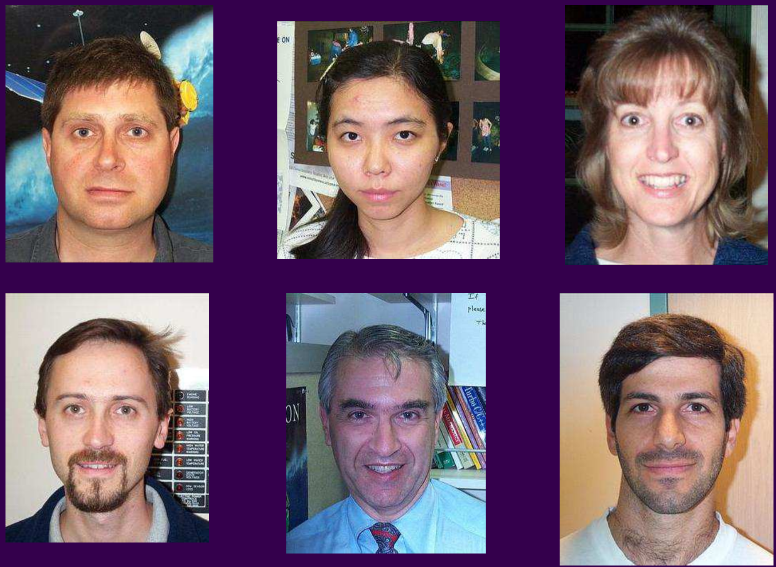
spotted cats

motorbikes





Faces



Motorbikes

Airplanes

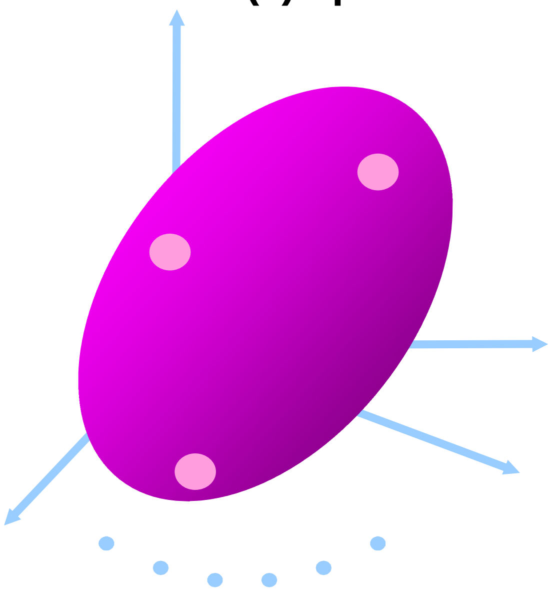
Spotted cats

Experiments: obtaining priors

airplanes

spotted cats

motorbikes



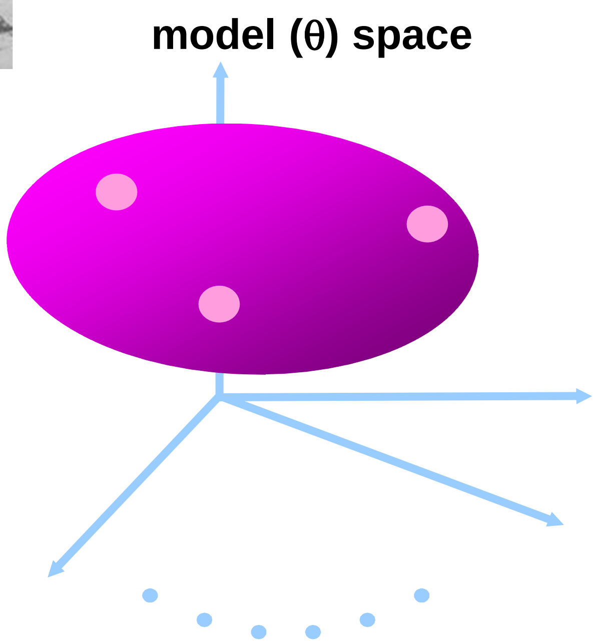
faces

Experiments: obtaining priors

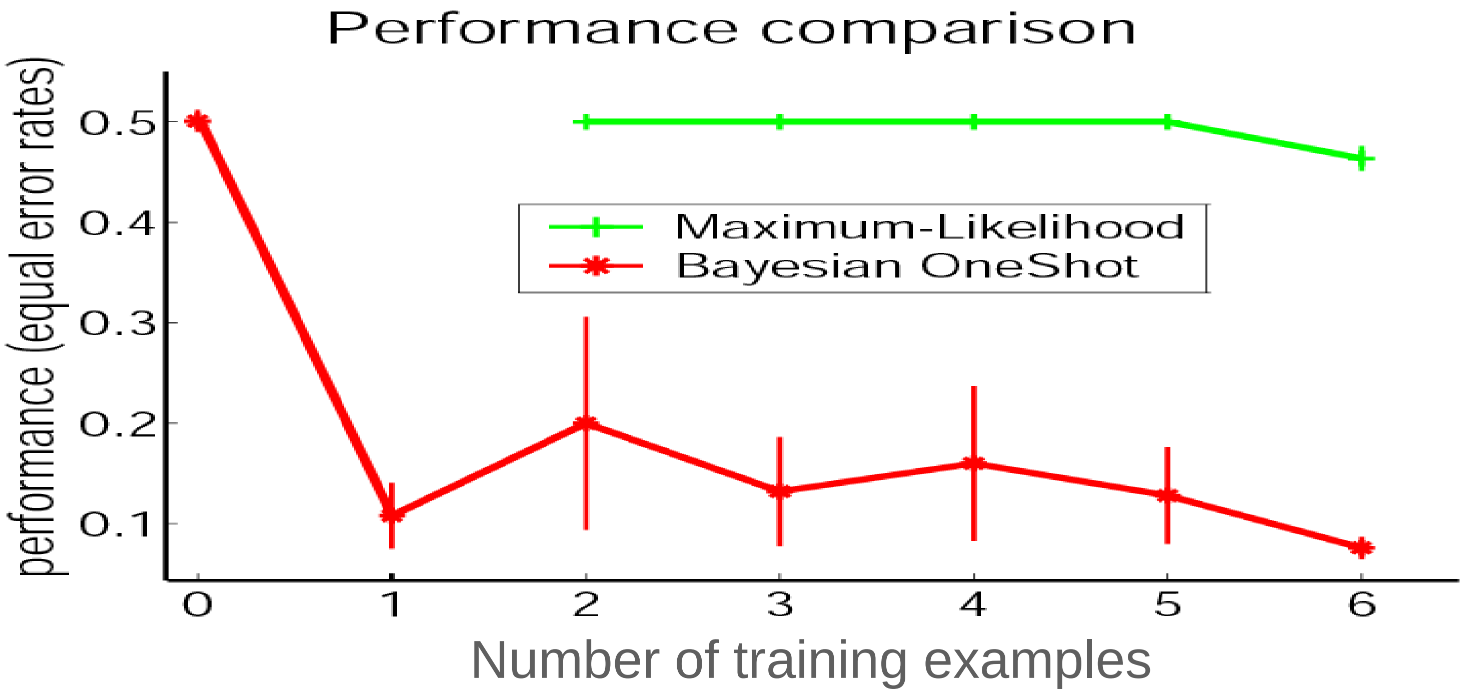
airplanes

faces

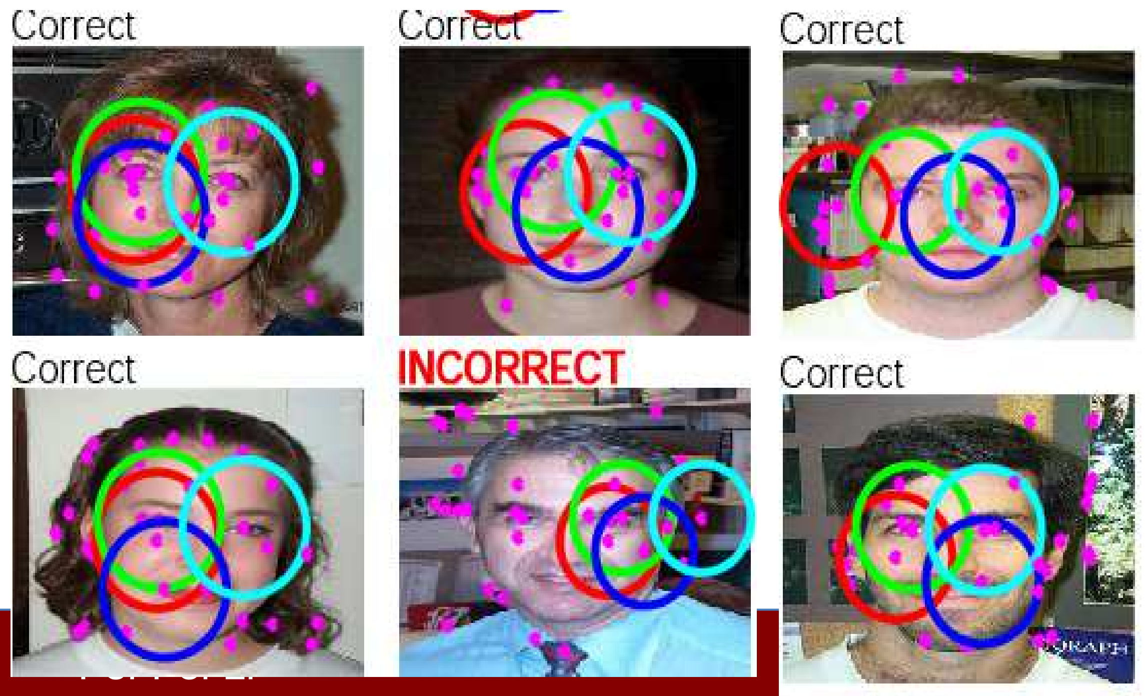
motorbikes

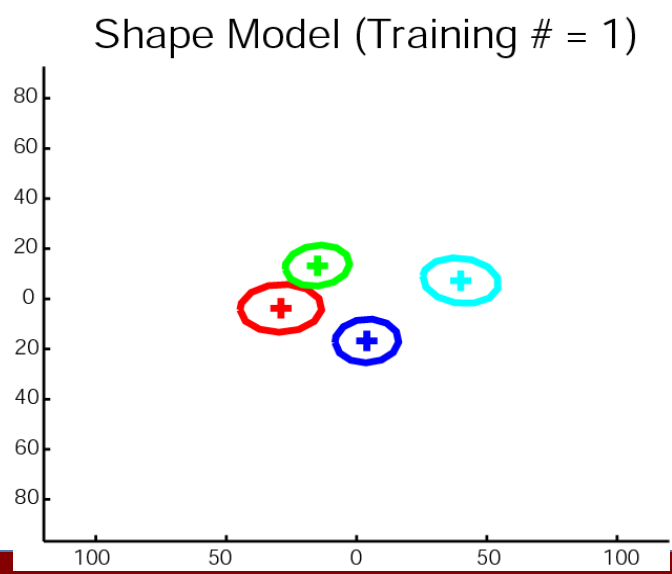


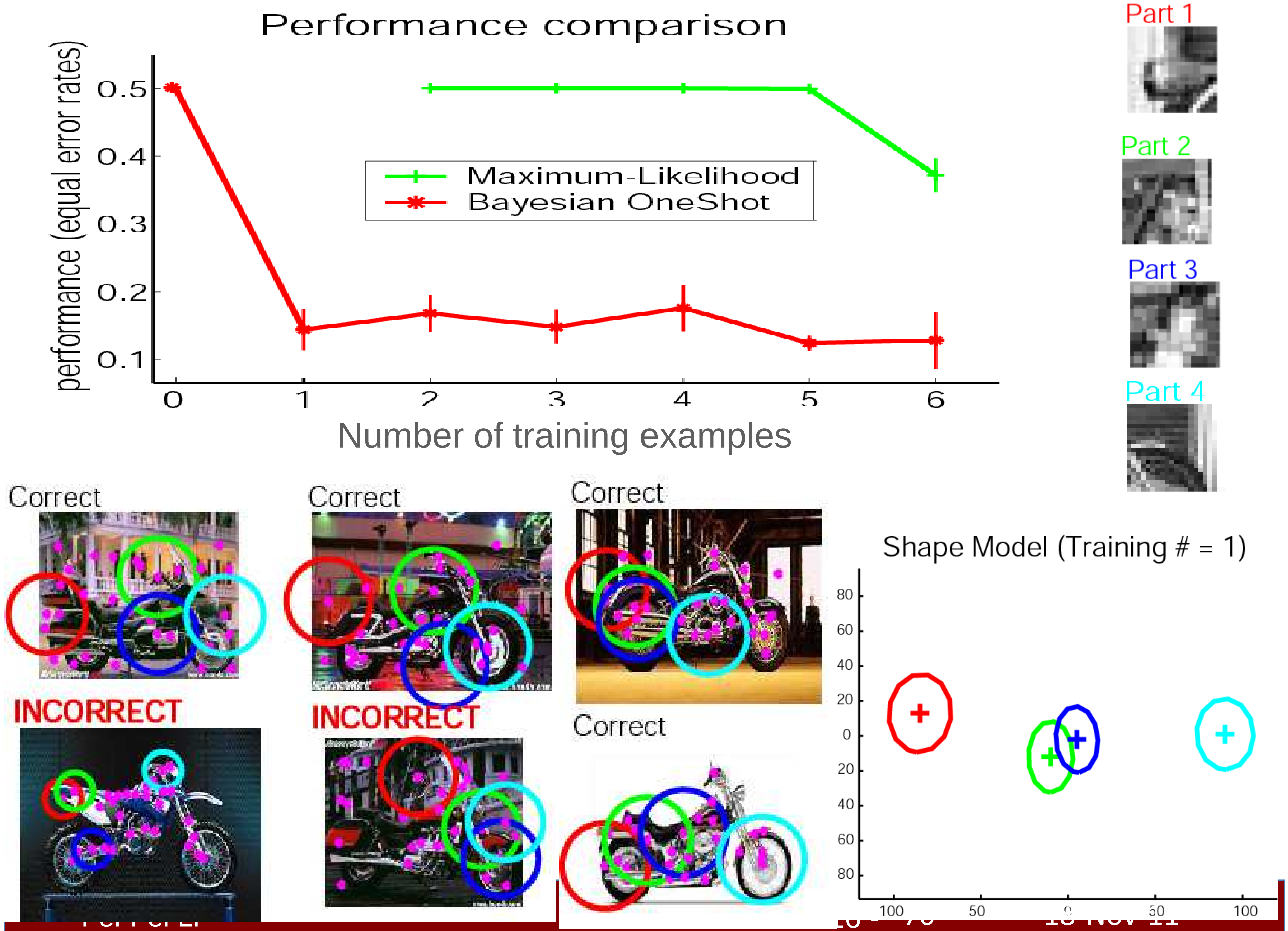
spotted cats

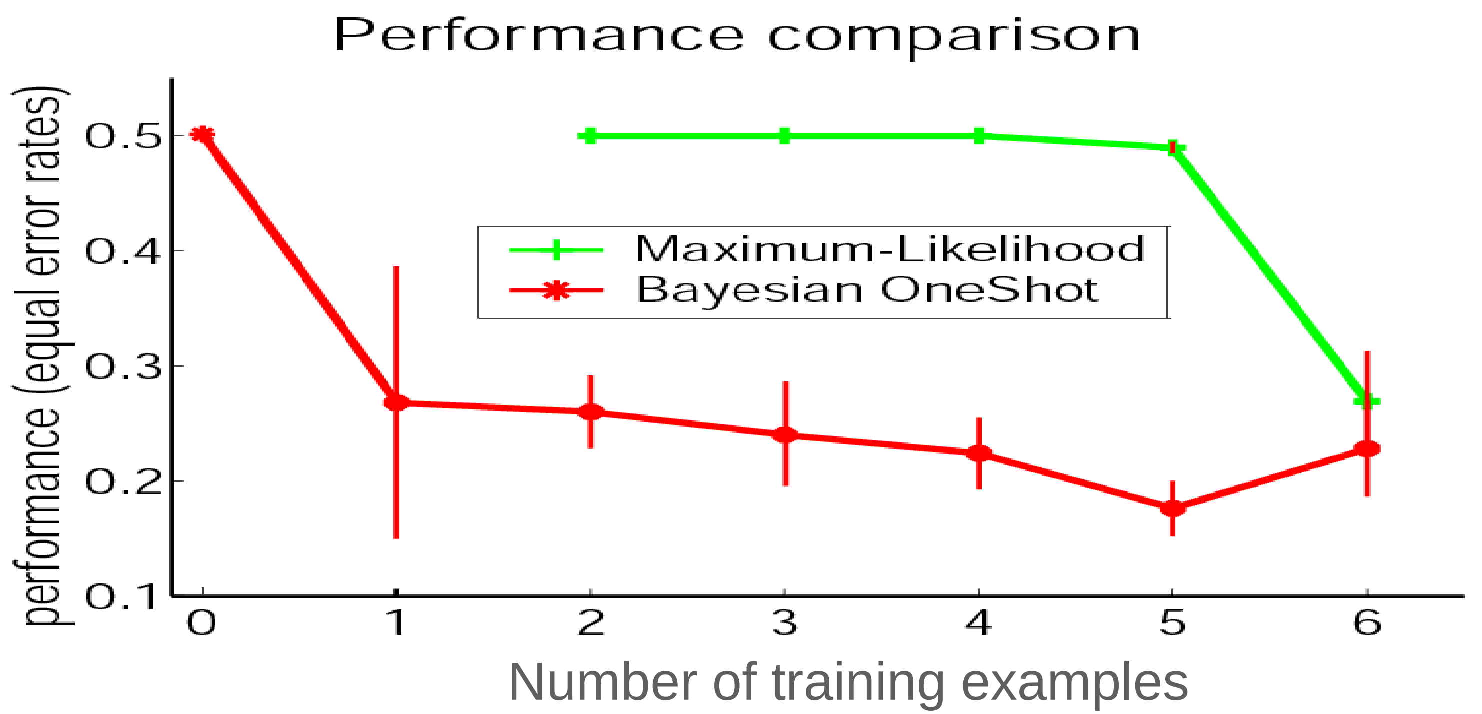

Performance comparison
performance (equal error rates)
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
Number of training examples
Maximum-Likelihood
Bayesian OneShot

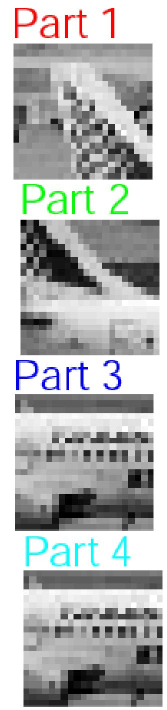

Part 1
Part 2
Part 3
Part 4

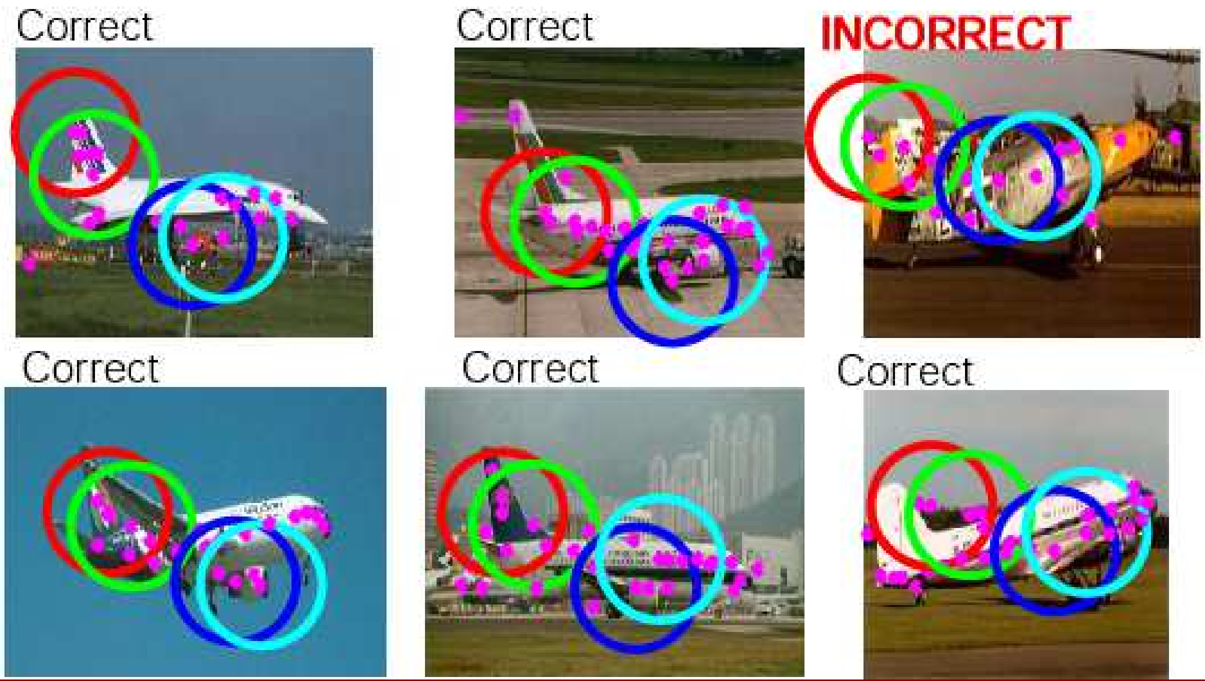

Correct
Correct
Correct
Correct
INCORRECT
Correct

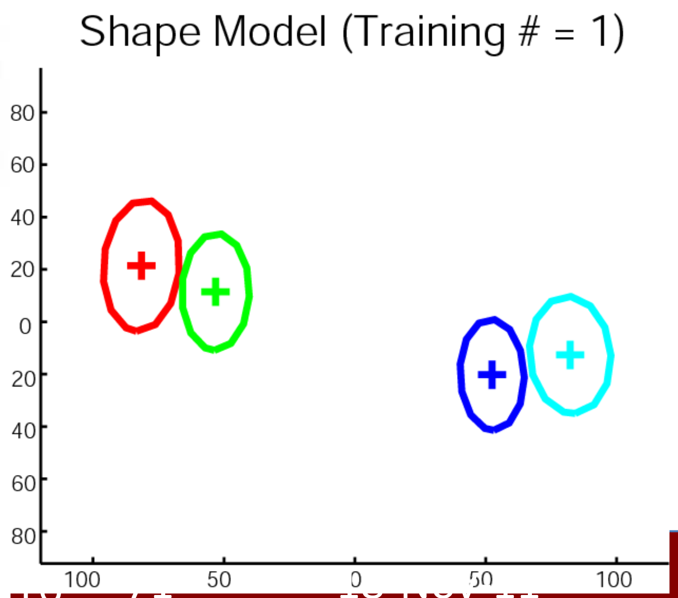

Shape Model (Training # = 1)
80
60
40
20
0
20
40
60
80
100
50
0
50
100

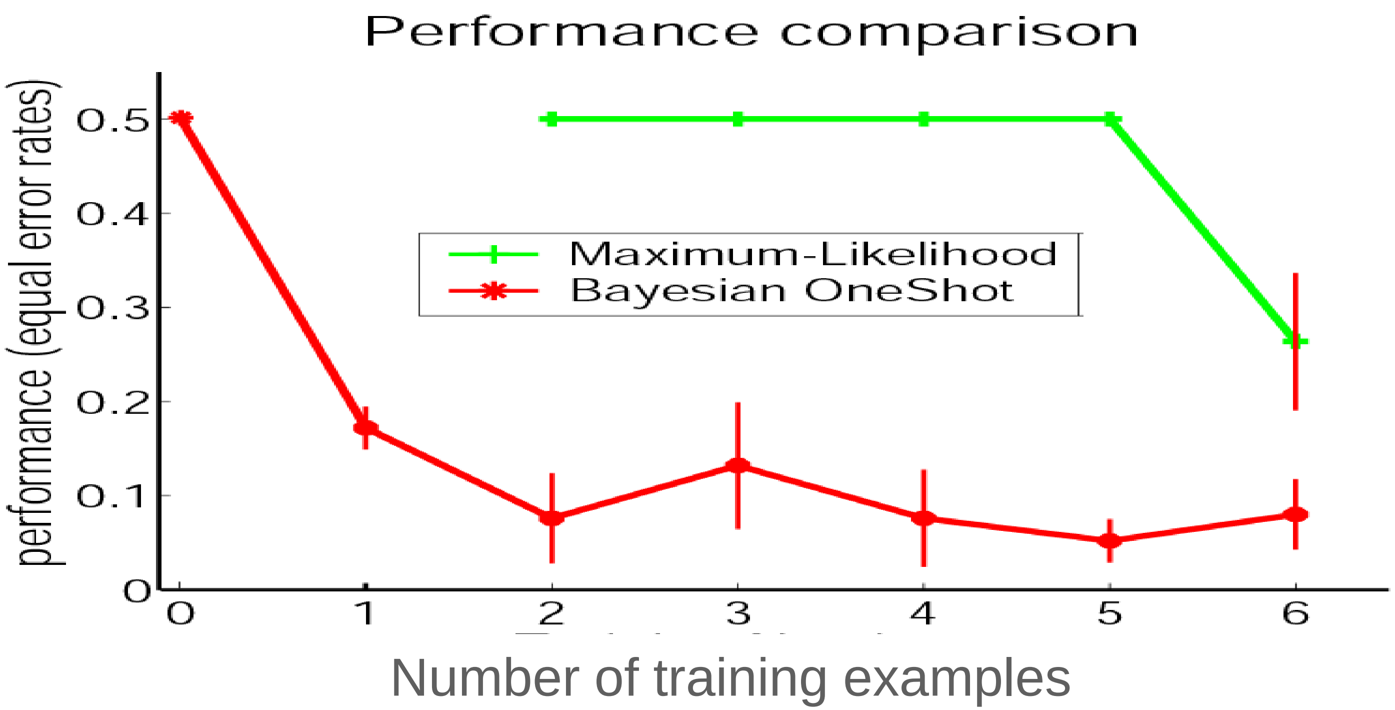

Performance comparison
performance (equal error rates)
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
Number of training examples
Maximum-Likelihood
Bayesian OneShot
Part 1
Part 2
Part 3
Part 4
Correct
Correct
Correct
INCORRECT
INCORRECT
Correct
Shape Model (Training # = 1)
80
60
40
20
0
20
40
60
80
100
50
50
100

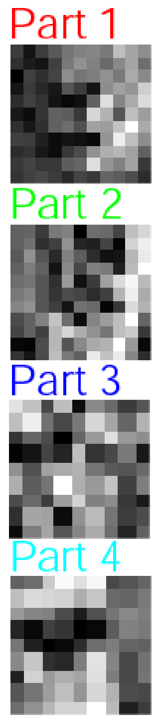

Performance comparison
performance (equal error rates)
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
Number of training examples
Maximum-Likelihood
Bayesian OneShot

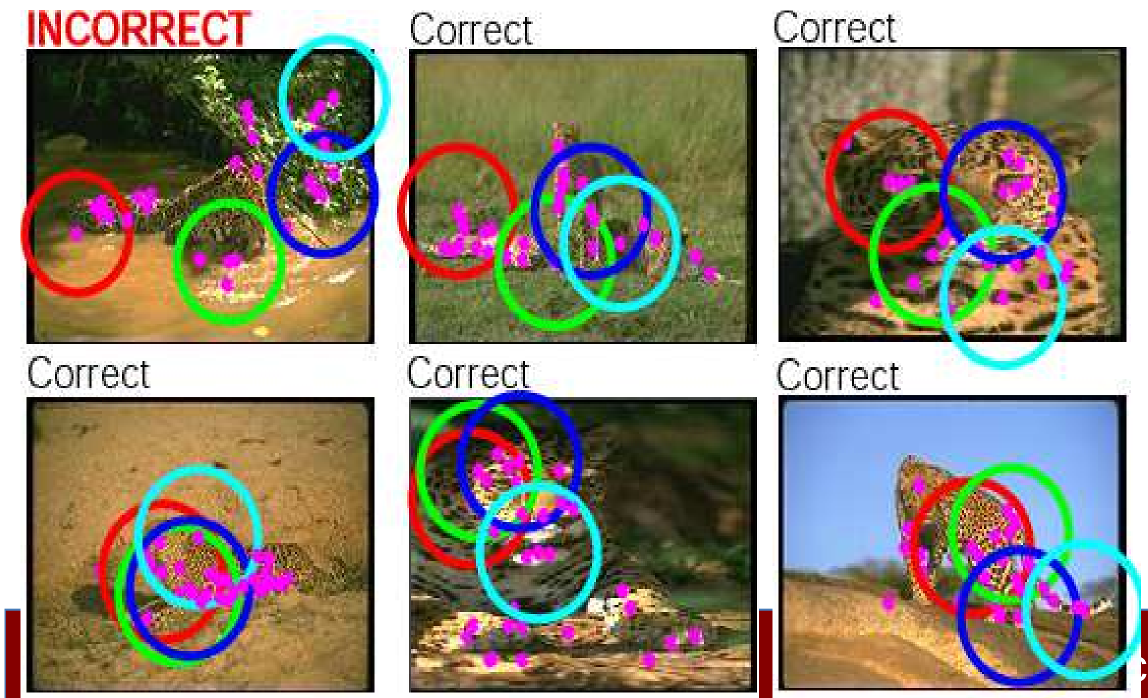

Part 1
Part 2
Part 3
Part 4

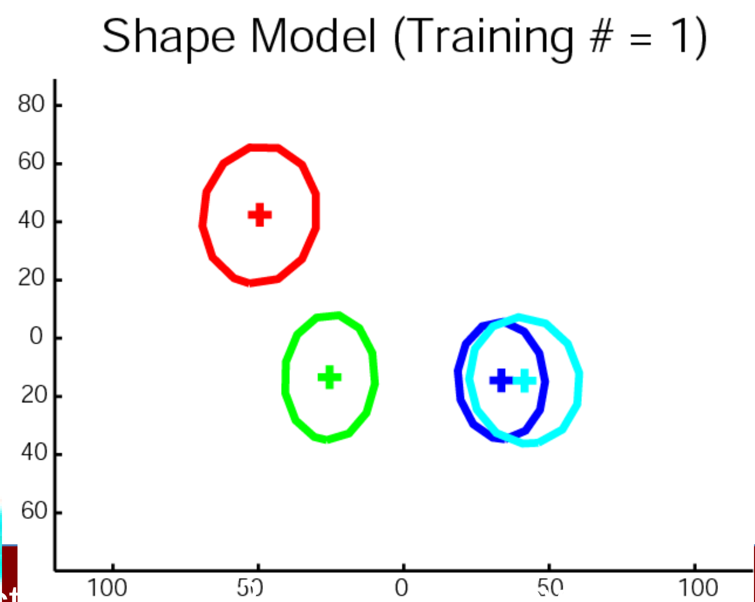

Correct
Correct
INCORRECT
Correct
Correct
Correct


Shape Model (Training # = 1)
80
60
40
20
0
20
40
60
80
100
50
0
50
100


Performance comparison
performance (equal error rates)
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6
Maximum-Likelihood
Bayesian OneShot
Number of training examples


Part 1
Part 2
Part 3
Part 4


INCORRECT
Correct
Correct
Correct
Correct
Correct


Shape Model (Training # = 1)
80
60
40
20
0
20
40
60
100 50 0 50 100

Algorithm	Training Examples	Categories	Results(error)
Burl, et al. Weber, et al. Fergus, et al.	200 ~ 400	Faces, Motorbikes, Spotted cats, Airplanes, Cars	5.6 - 10 %
Viola et al.	~10,000	Faces	7-21%
Schneiderman, et al.	~2,000	Faces, Cars	5.6 – 17%
Rowley et al.	~500	Faces	7.5 – 24.1%
Bayesian One-Shot	1 ~ 5	Faces, Motorbikes, Spotted cats, Airplanes	8 – 15 %

What we have learned today?

- Introduction
- Constellation model
  - Weakly supervised training
  - One-shot learning
- (Problem Set 4 (Q1))