Object class recognition using unsupervised scale-invariant learning

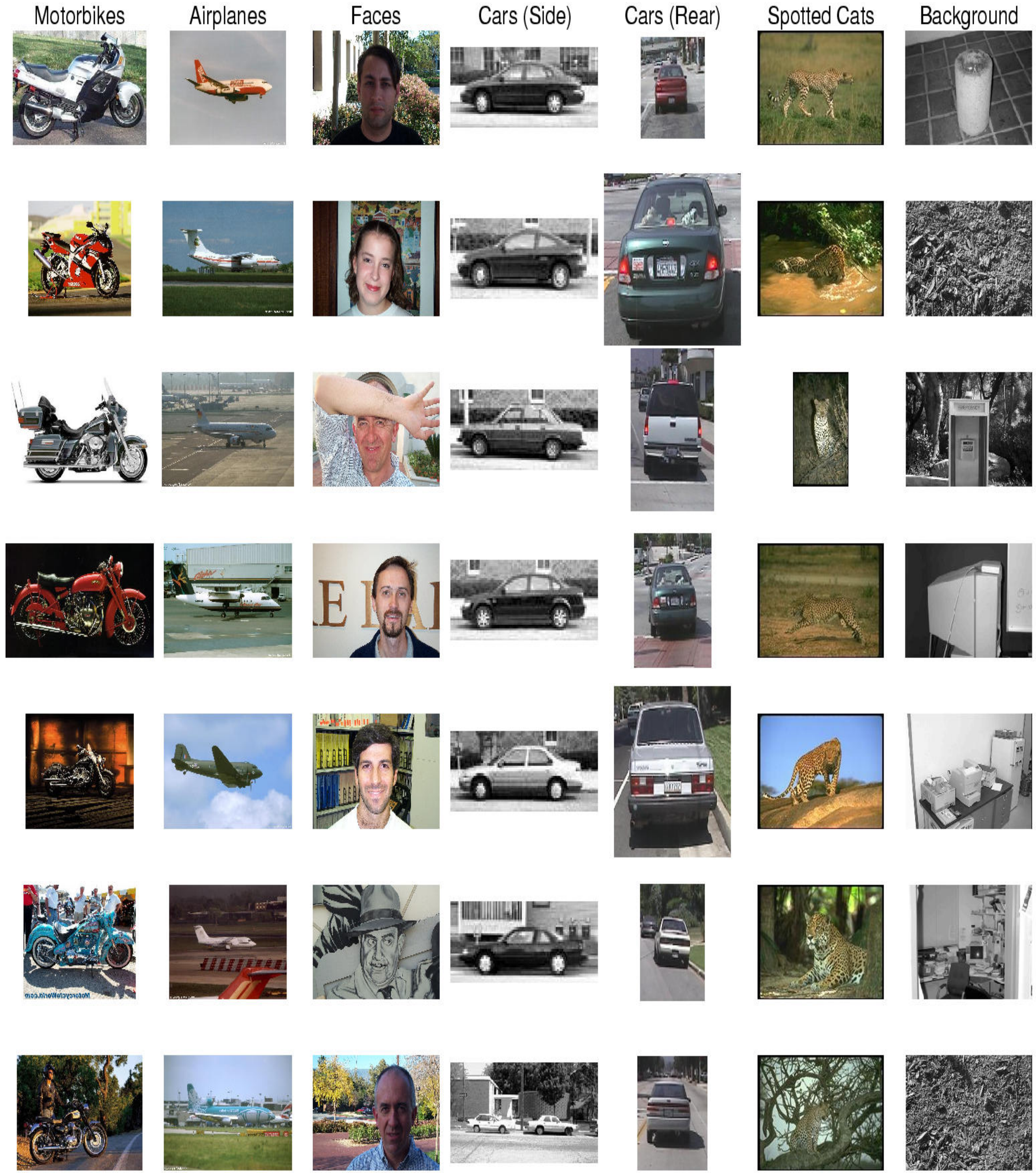
Rob Fergus
Pietro Perona
Andrew Zisserman

Oxford University
California Institute of Technology

Goal

- Recognition of object categories
- Unassisted learning

Some object categories

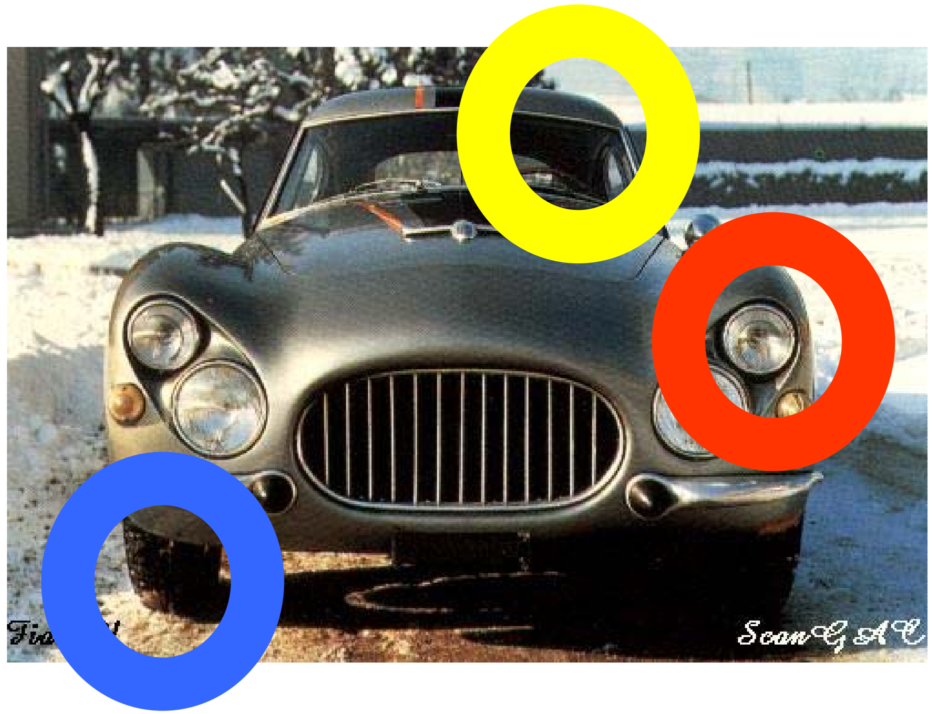


Learn from examples

Difficulties:

- Size variation
- Background clutter
- Occlusion
- Intra-class variation

Model: Constellation of Parts



Main issues:

- measuring the similarity of parts
- representing the configuration of parts

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
Perona et al. ‘95, ‘96, ’98, ’00
Agarwal & Roth ‘02

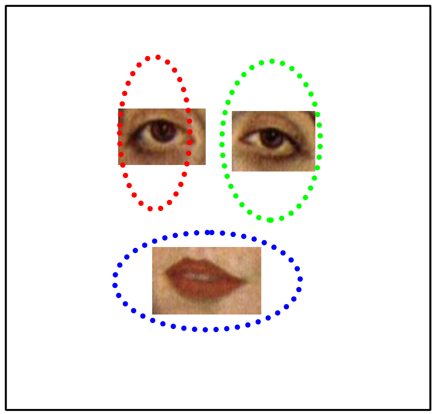
Overview of talk

- Representation
- Recognition
- Learning

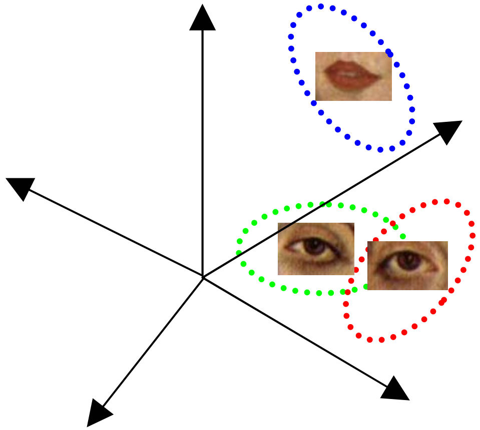
Generative probabilistic model

Foreground model

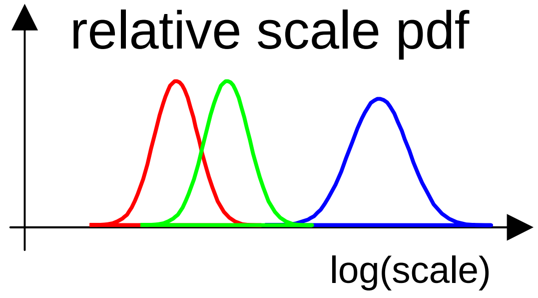
Gaussian shape pdf



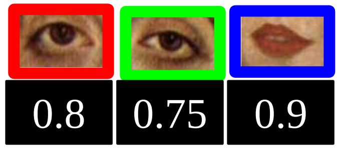
Gaussian part appearance pdf



Gaussian
relative scale pdf

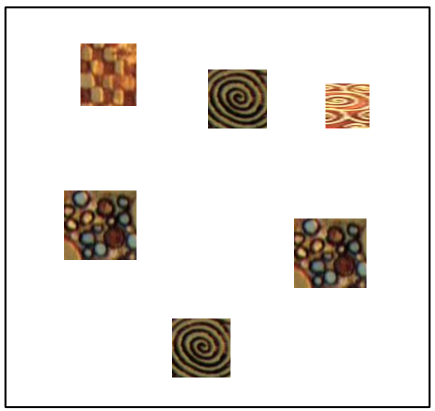


Prob. of detection

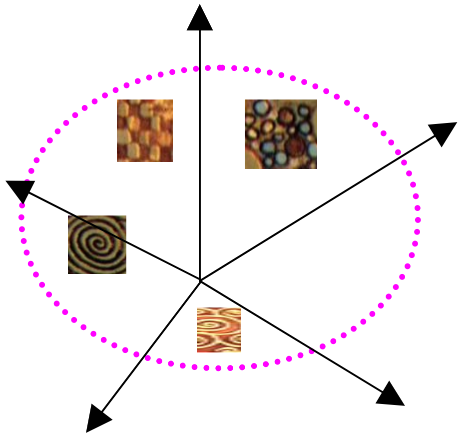


Clutter model

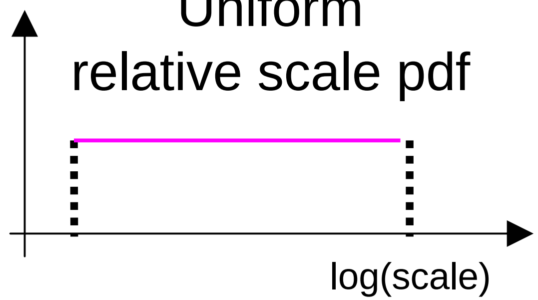
Uniform shape pdf



Gaussian background
appearance pdf



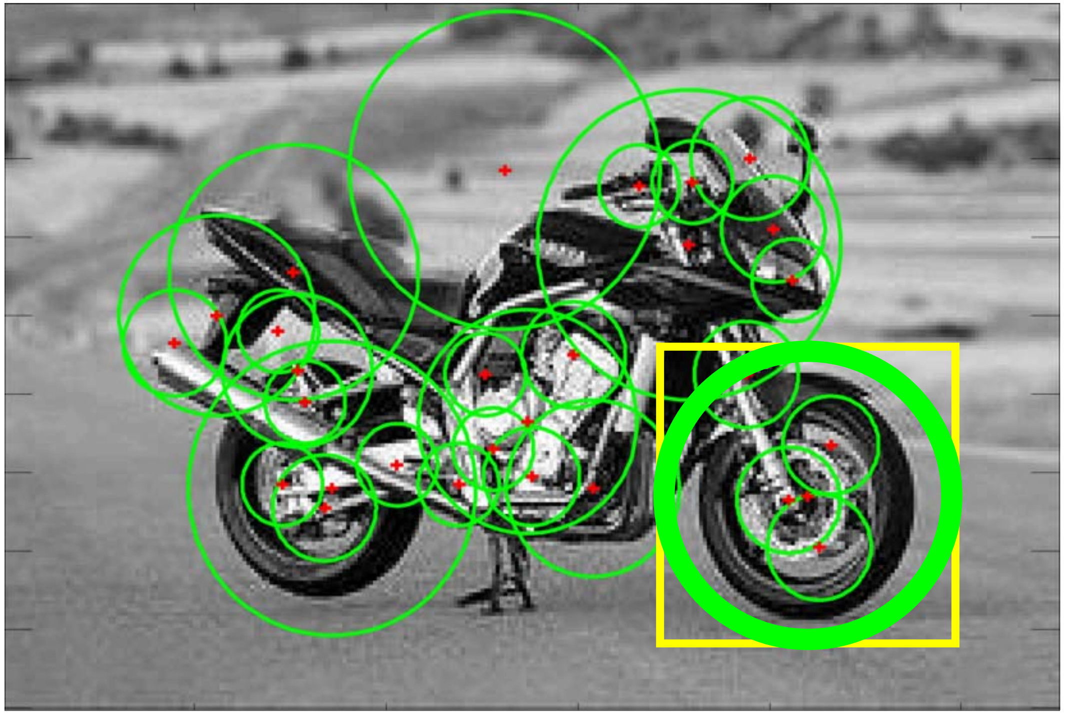
Uniform
relative scale pdf



Poission pdf on #
detections

Recognition

Detection & Representation of regions



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

Location

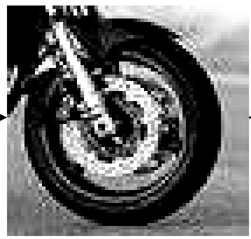
(x,y) coords. of region centre

Scale

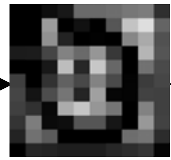
Radius of region (pixels)

Appearance

Normalize →



11x11 patch →

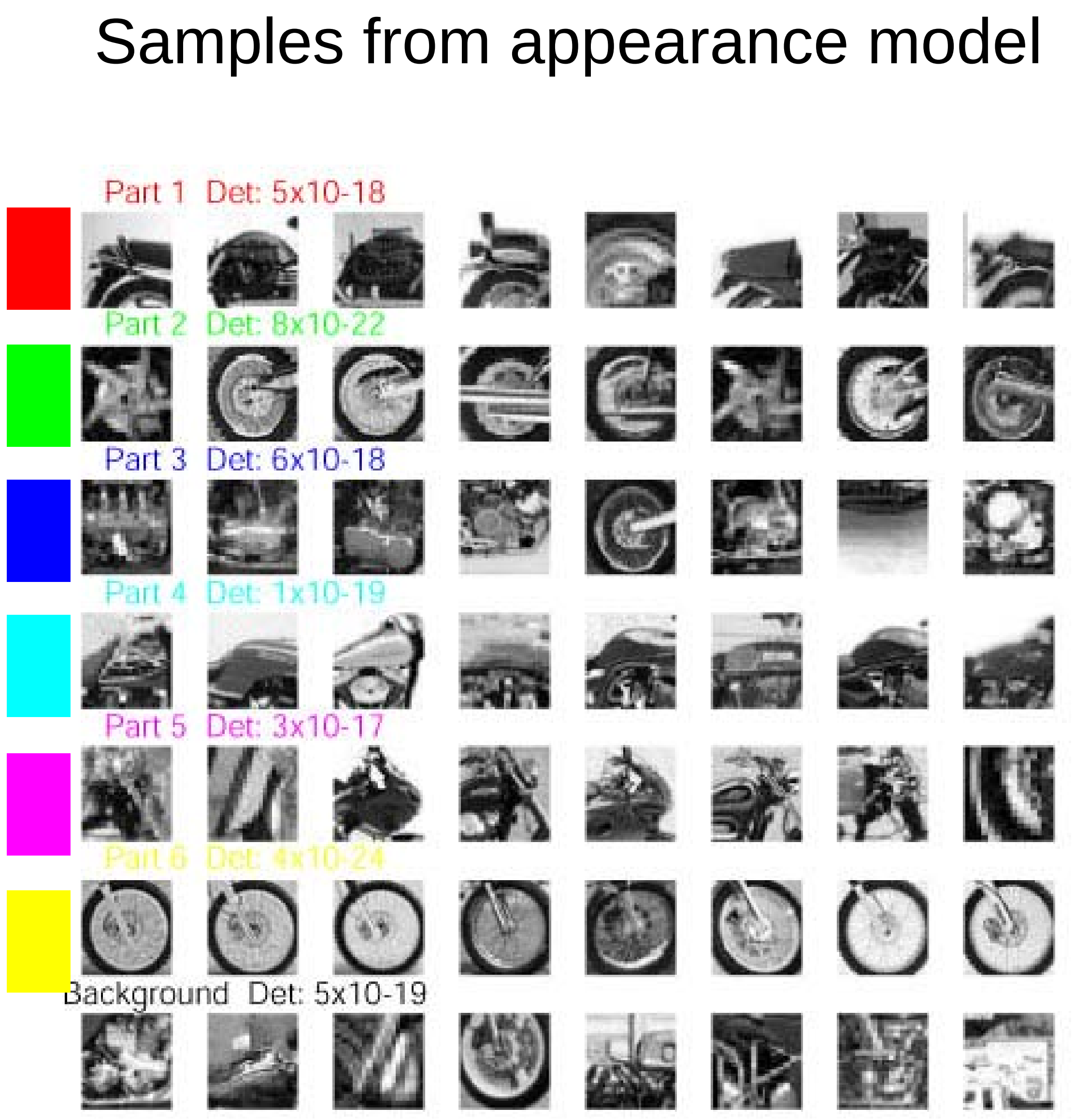


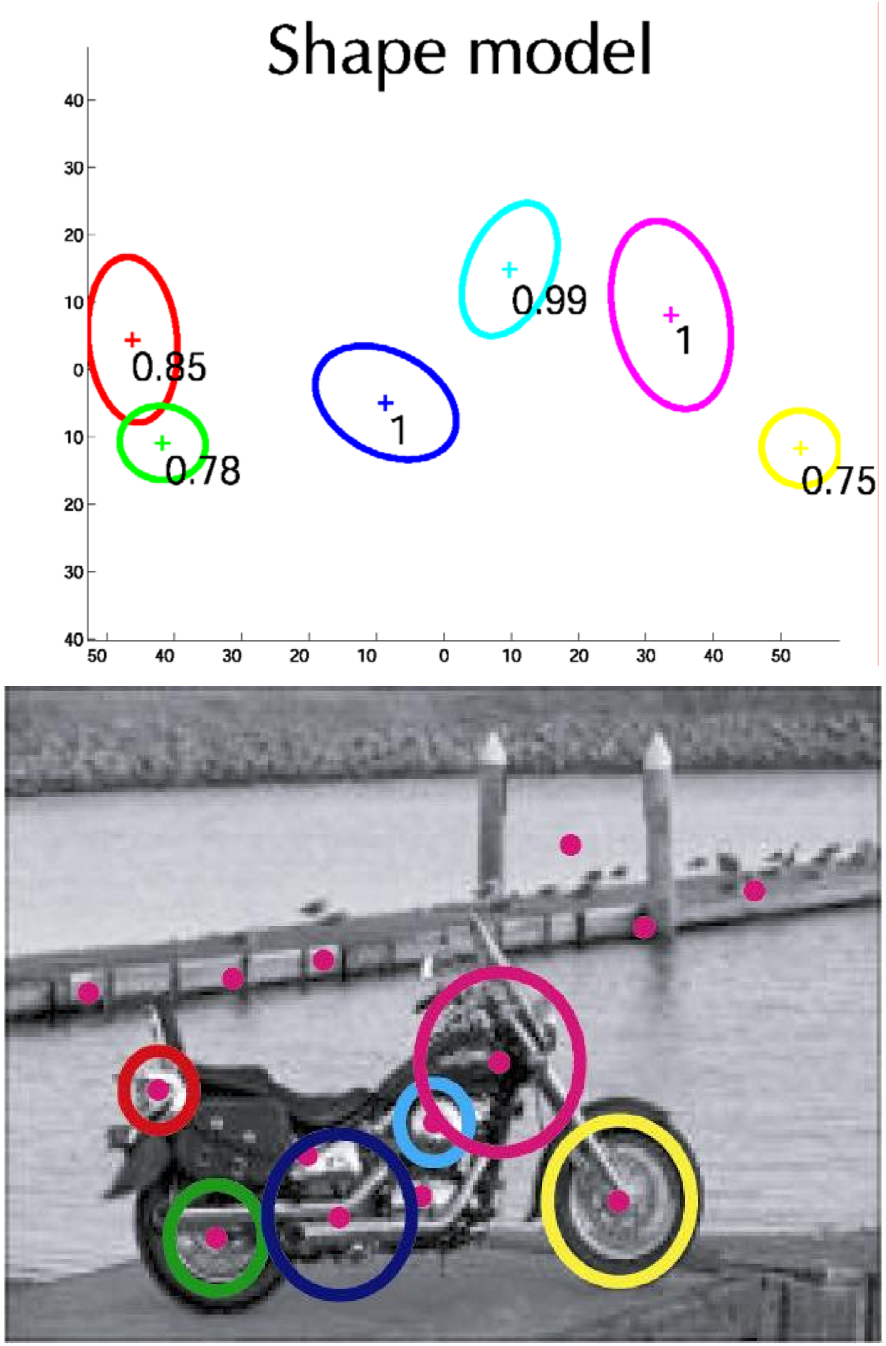
Projection onto PCA basis →

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

Gives representation of appearance in low-dimensional vector space

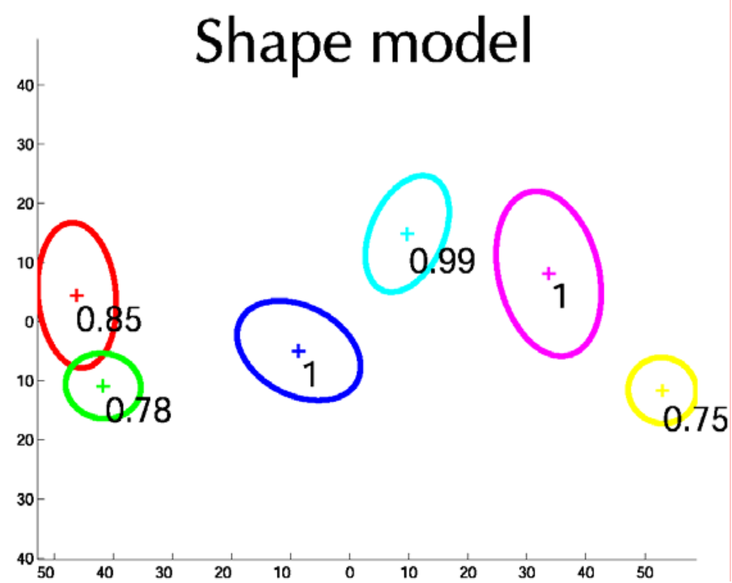
Motorbikes

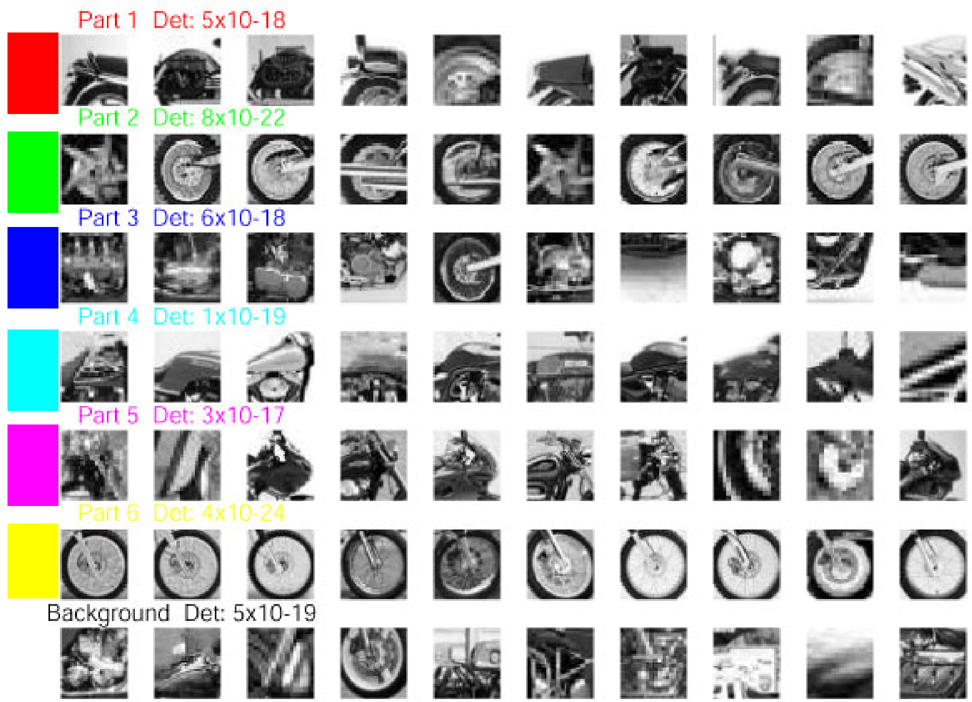

Samples from appearance model
Part 1 Det: 5x10-18
Part 2 Det: 8x10-22
Part 3 Det: 6x10-18
Part 4 Det: 1x10-19
Part 5 Det: 3x10-17
Part 6 Det: 4x10-24
Background Det: 5x10-19


Shape model
0.85
0.78
1
0.99
1
0.75

Detected regions

Recognized Motorbikes





Background images evaluated with motorbike model

Learning

Learning procedure

- Find regions & their location, scale & appearance
  over all training
- Initialize model parameters
- Use EM and iterate to convergence:

  E-step: Compute assignments for which regions are foreground / background

  M-step: Update model parameters

- Trying to maximize likelihood – consistency in shape & appearance

Experiments

Experimental procedure

Two series of experiments:

- Fixed-scale model - Objects the same size (manual normalization)
- Scale-invariant model - Objects between 100 and 550 pixels in width

Training

- 50% images
- No identifcation of object within image

Testing

- 50% images
- Simple object present/absent test

Datasets

Motorbikes

Airplanes

Frontal Faces

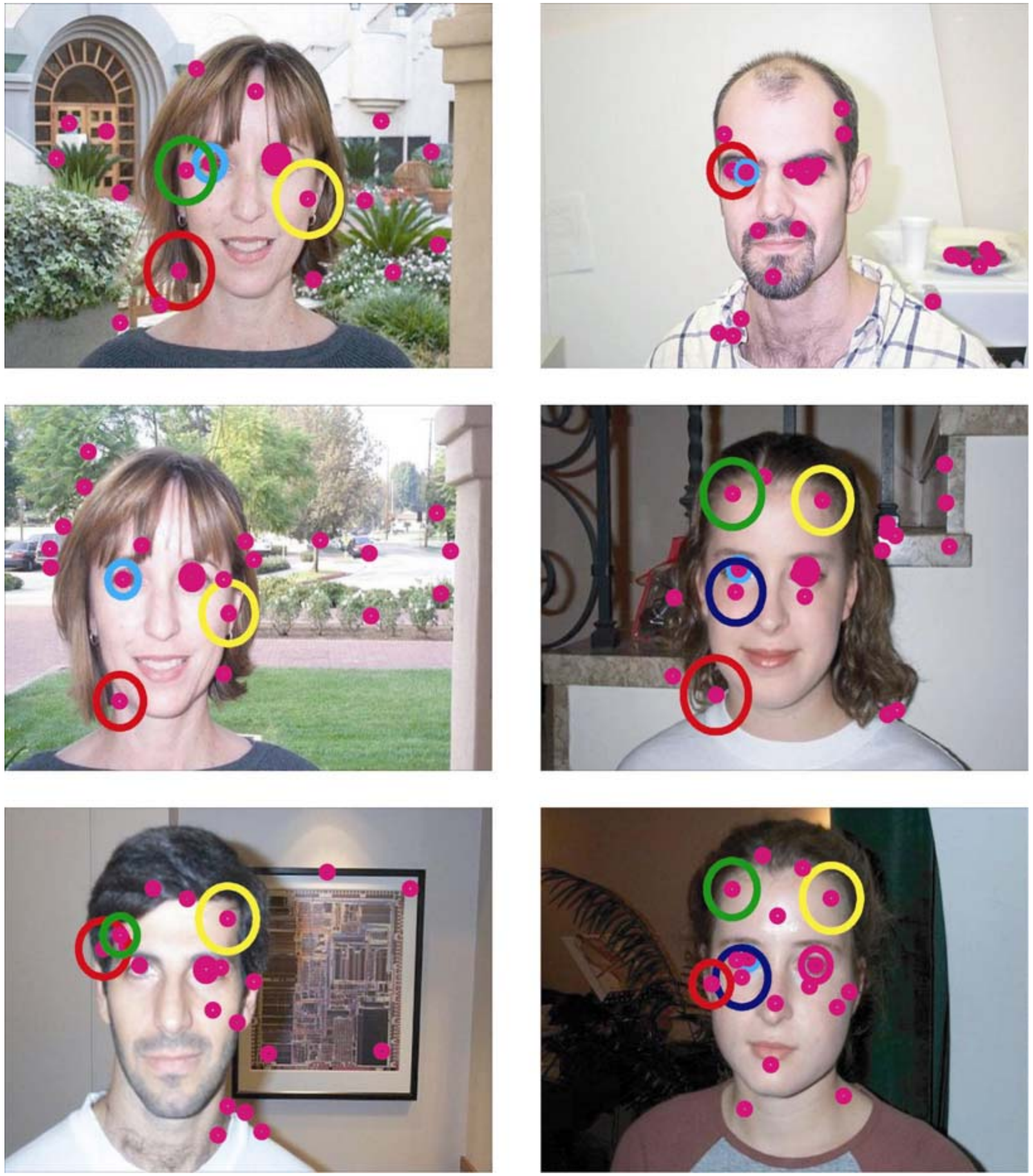
Cars (Side)

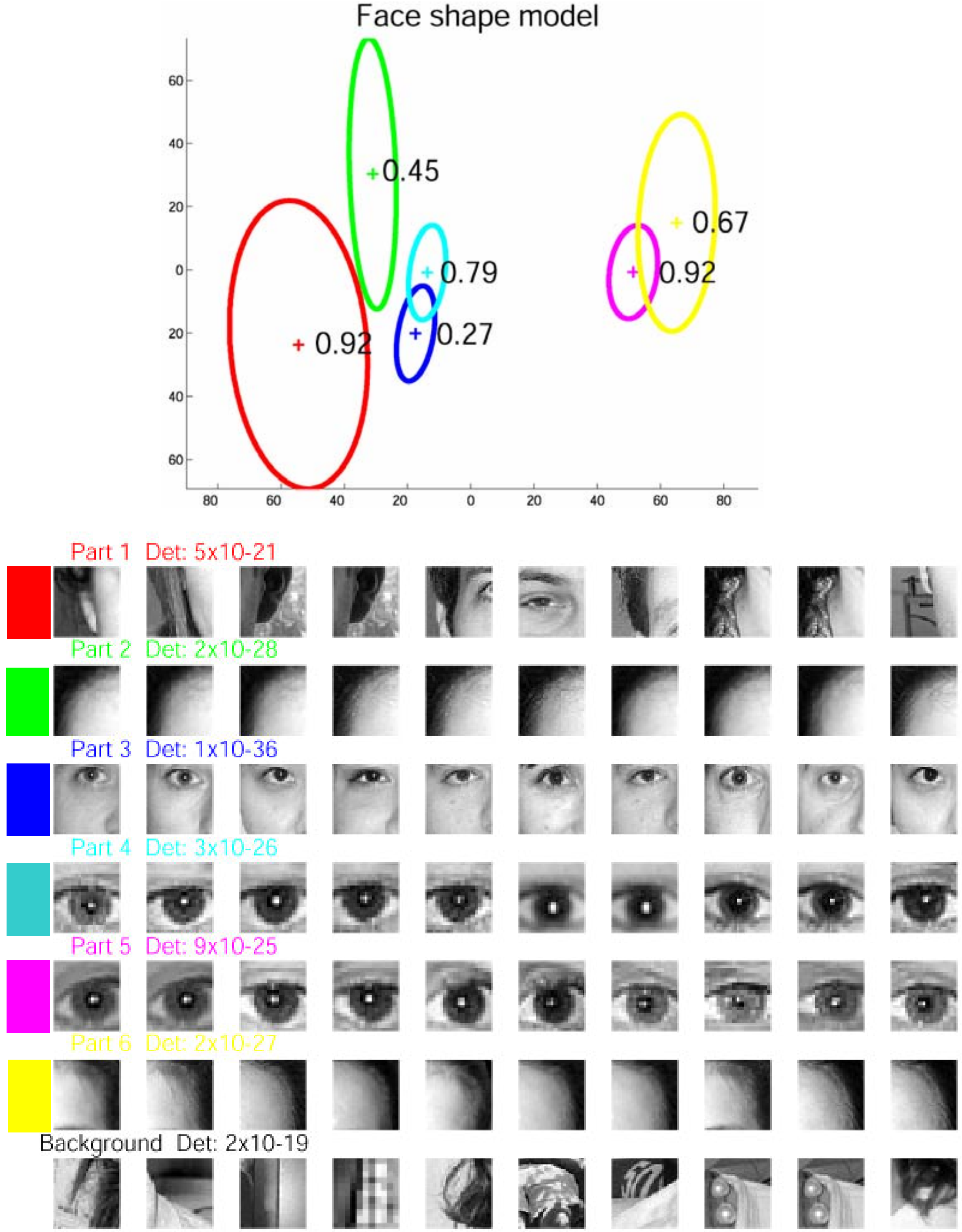
Cars (Rear)

Spotted cats

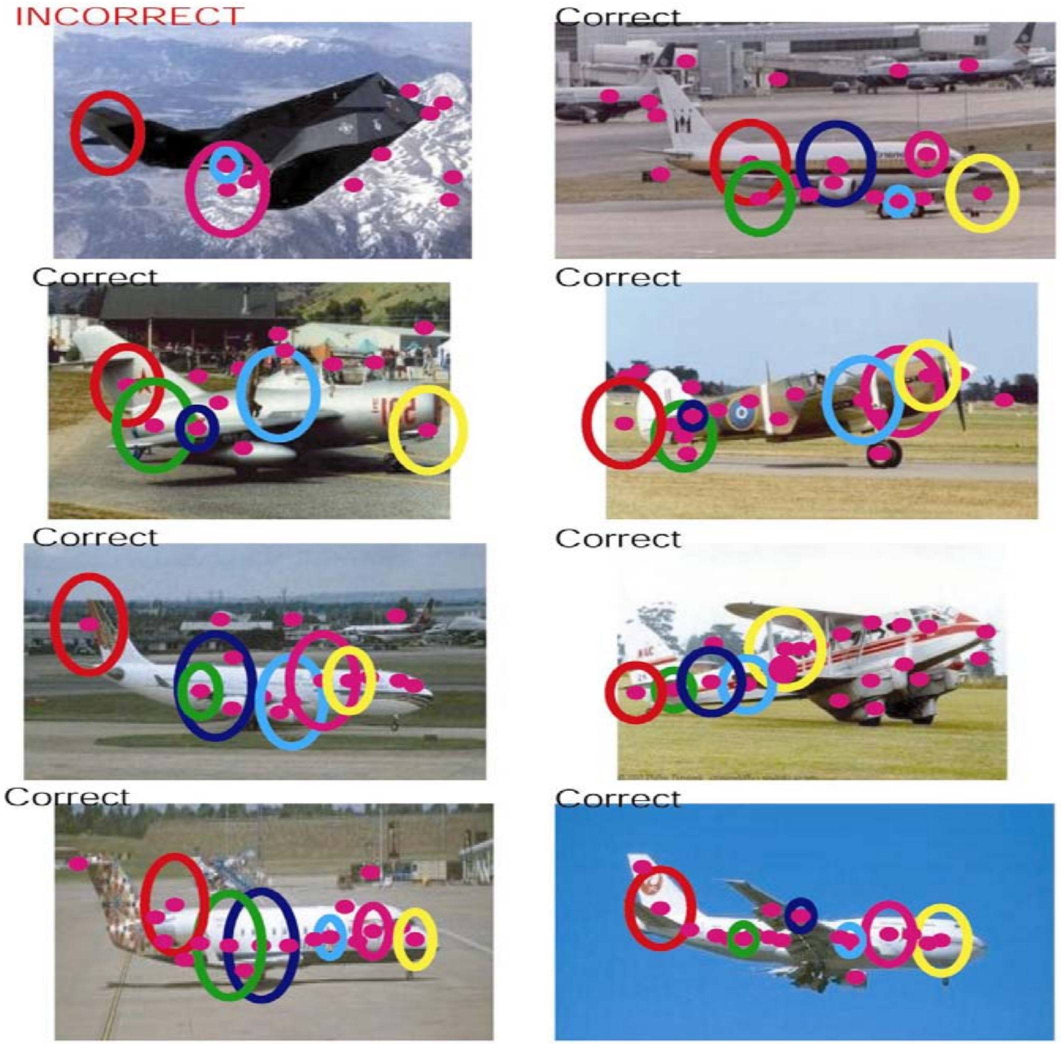
Between 200 and 800 images in each dataset

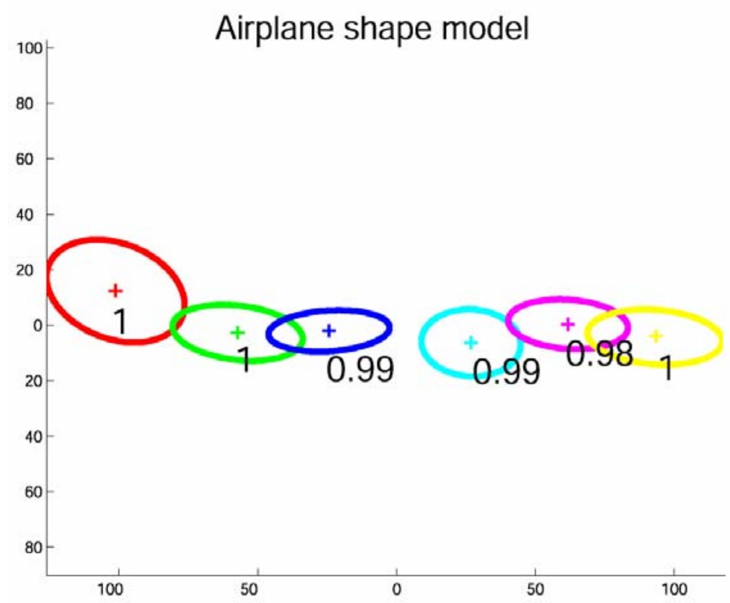
Frontal faces

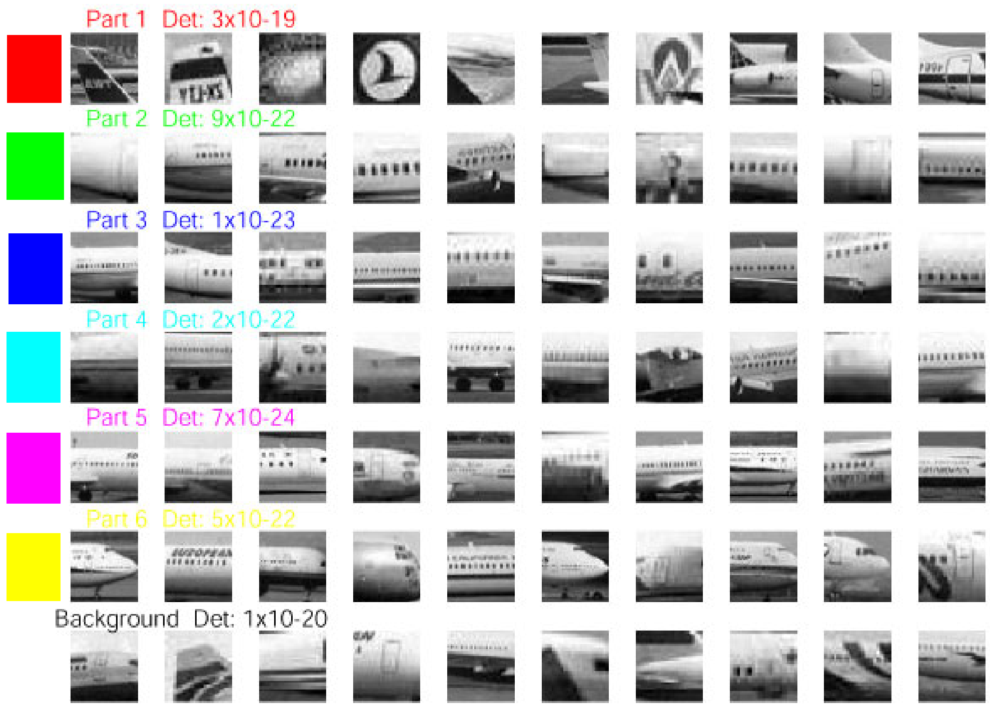




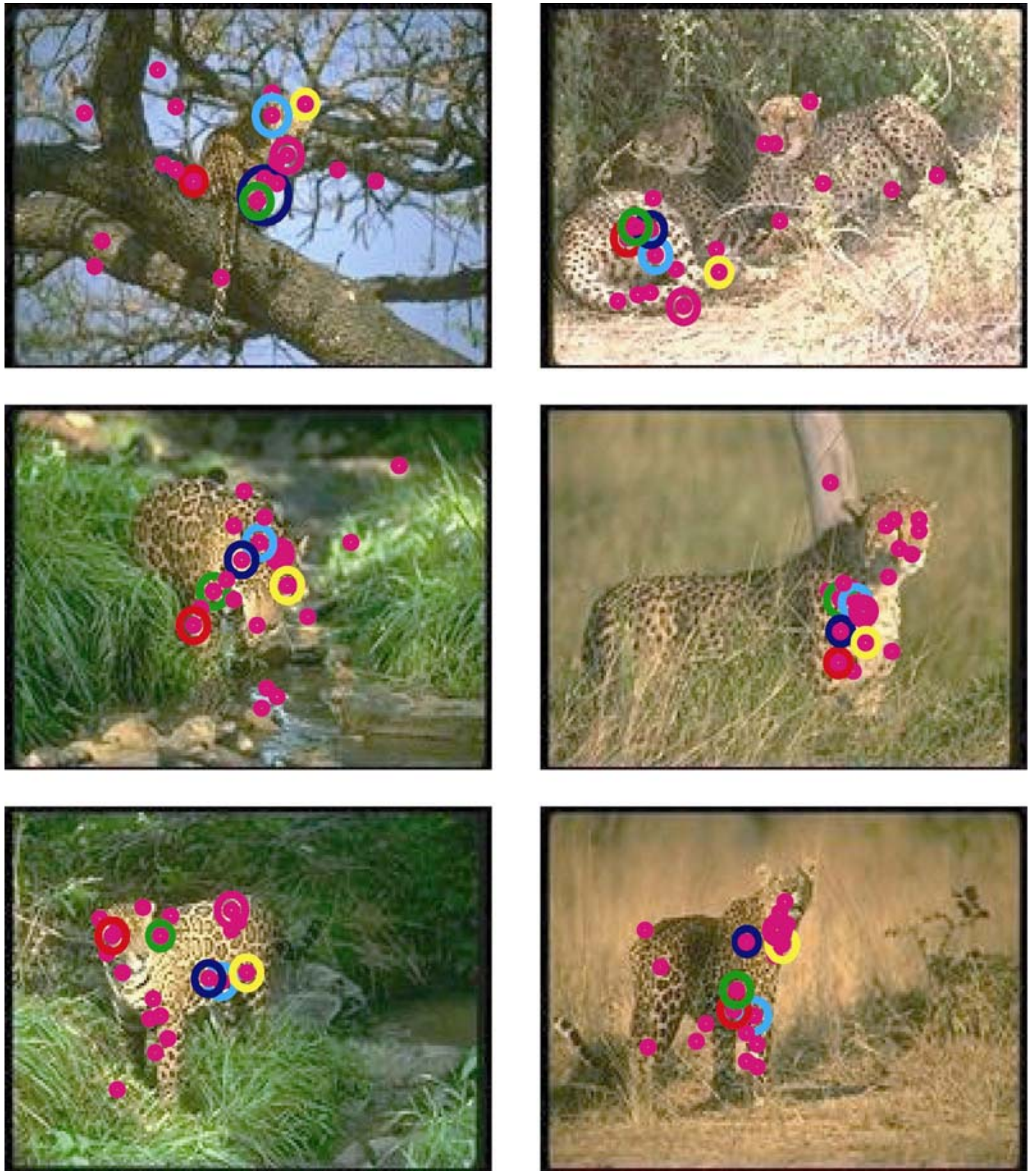
Airplanes

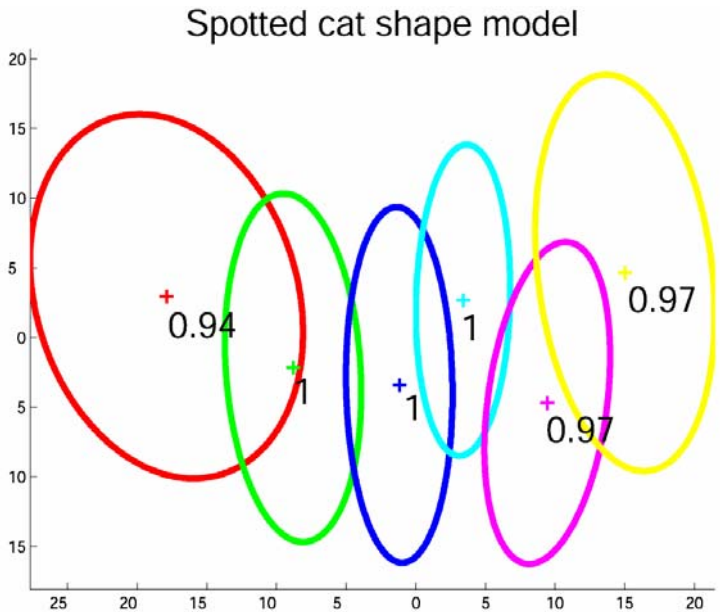

INCORRECT
Correct
Correct
Correct
Correct
Correct
Correct
Correct

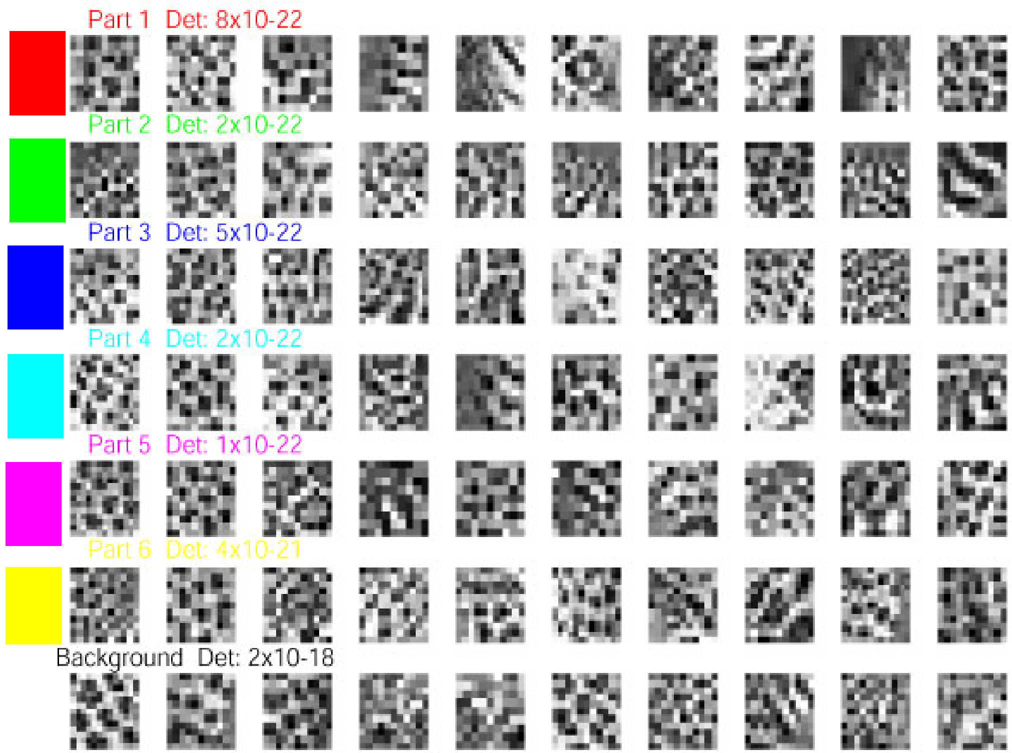

Airplane shape model
1
1
0.99
0.99
0.98
1


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

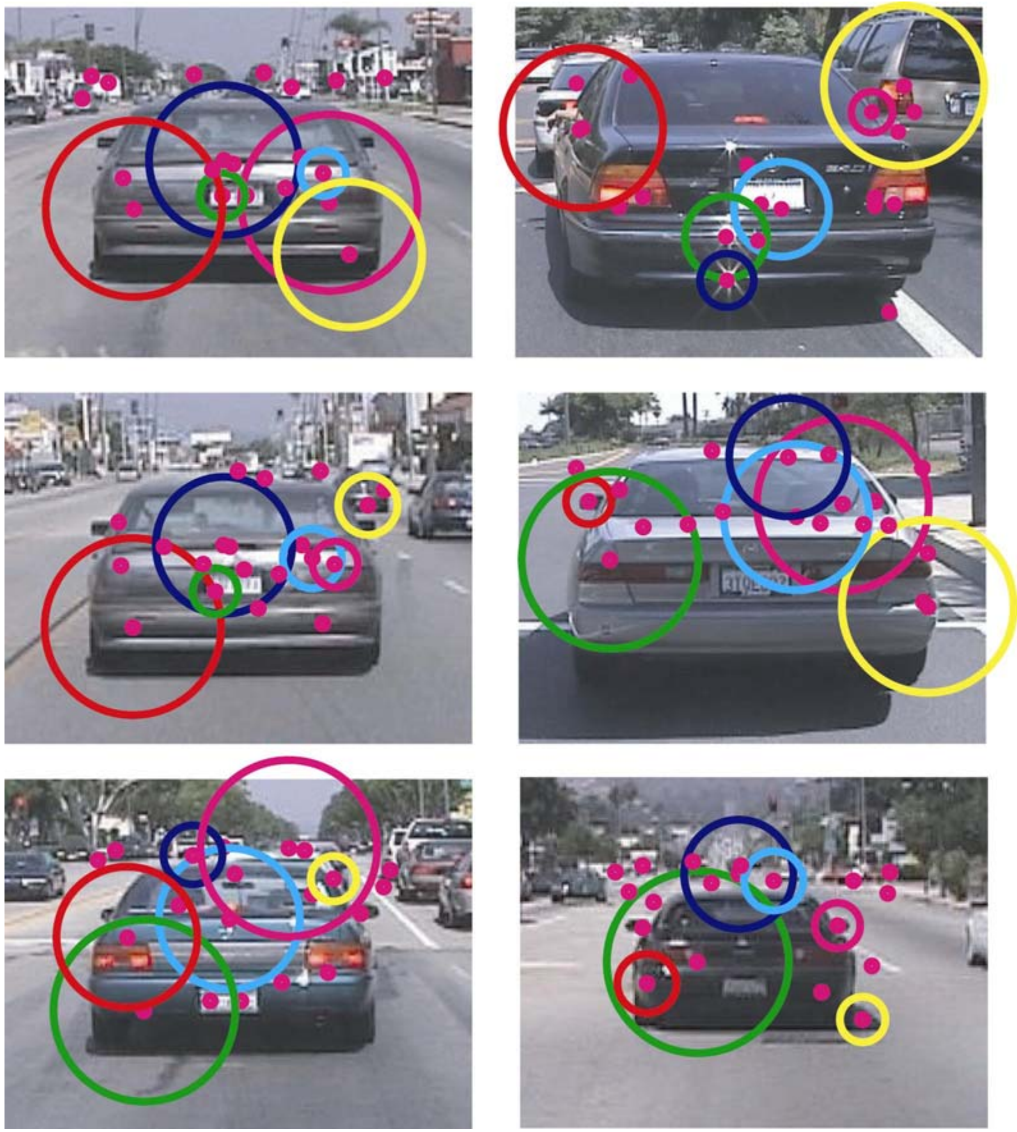


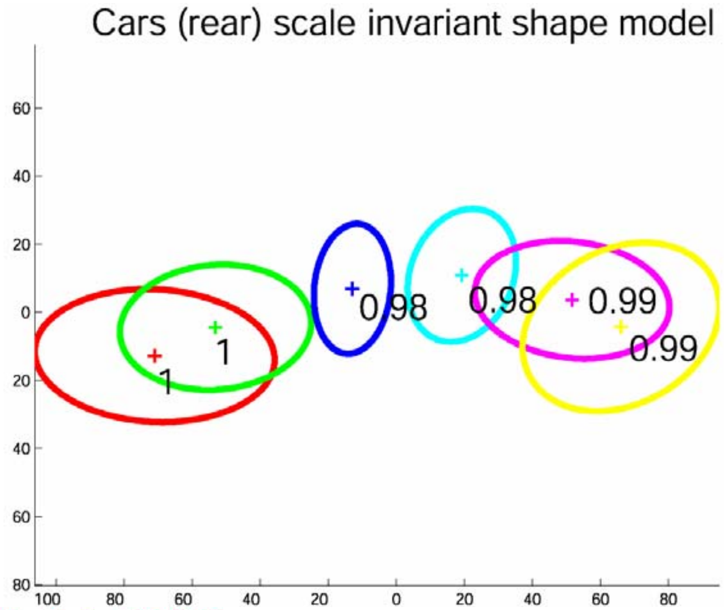


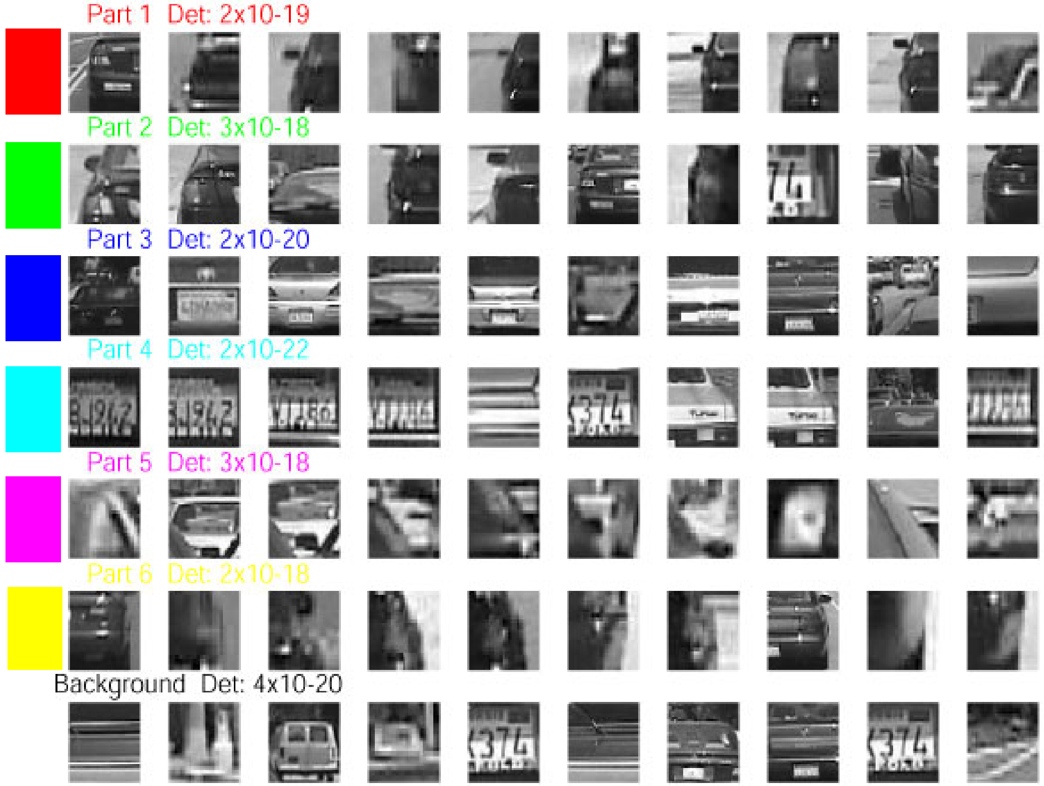


Cars from rear

- Scale invariant







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

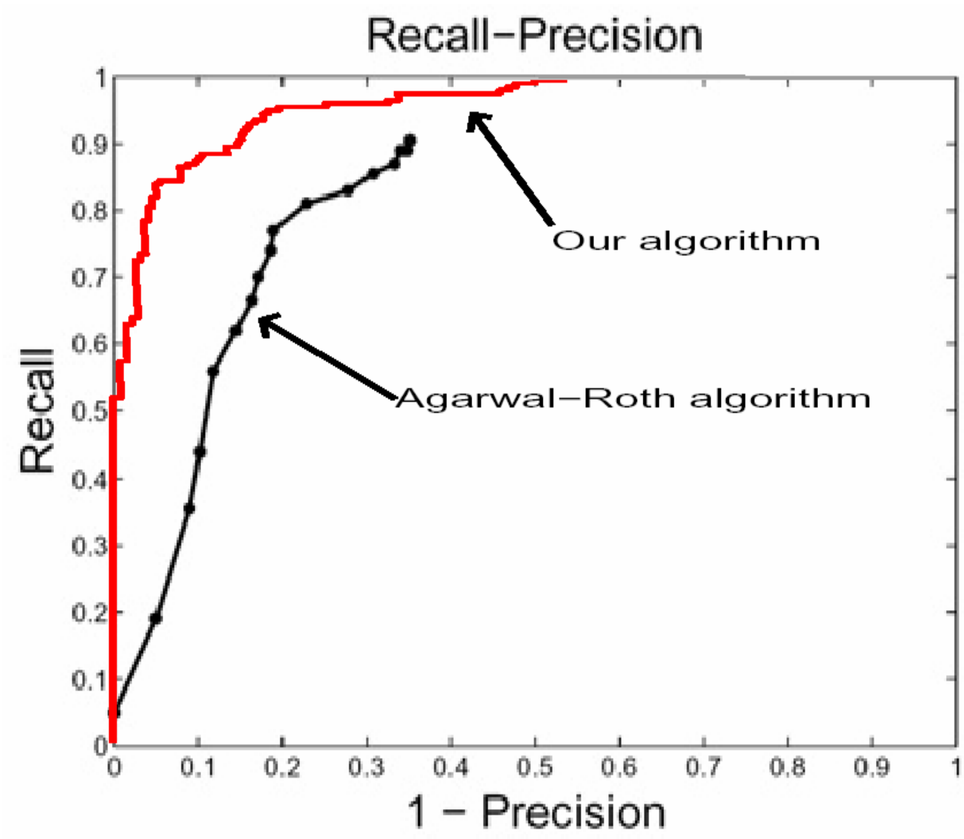
% equal error rate

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

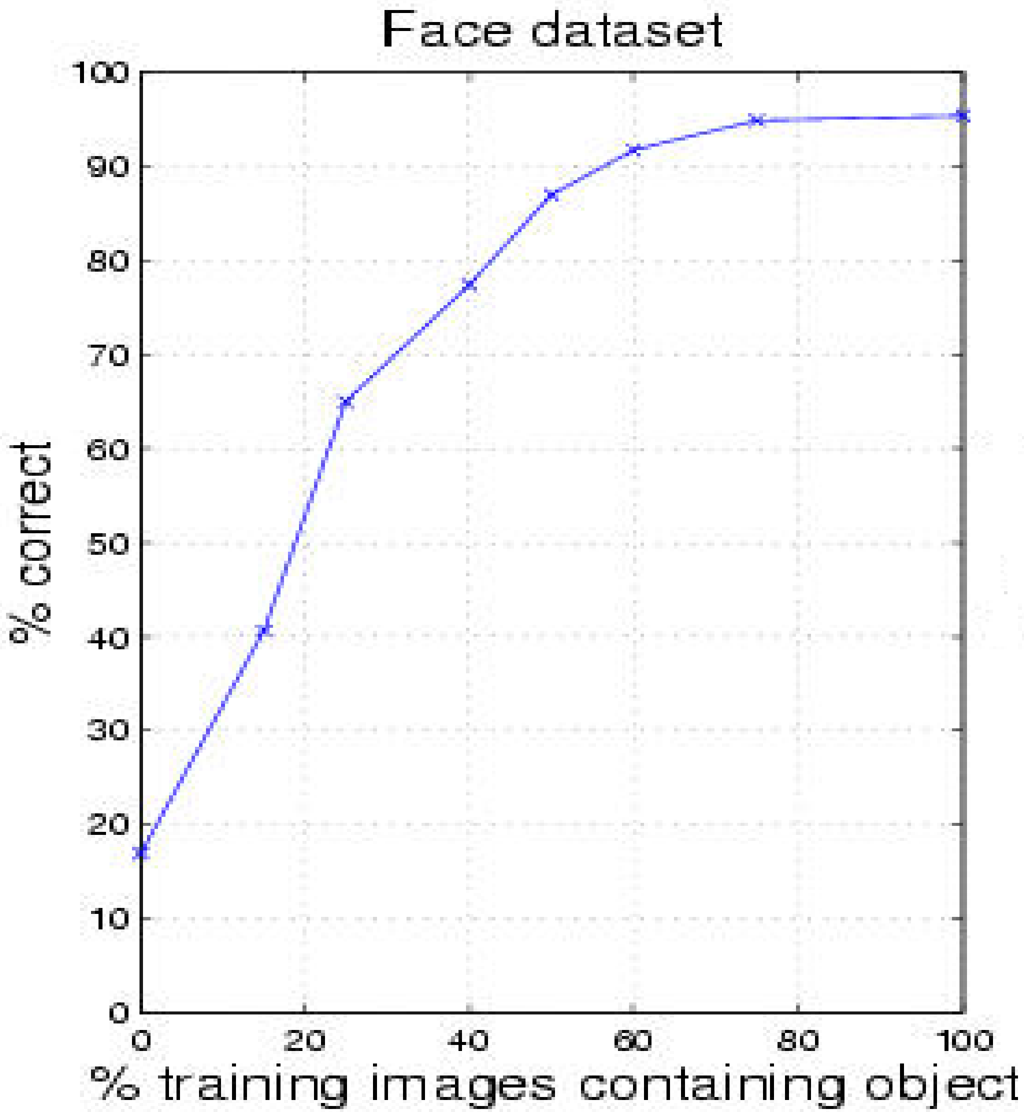
Comparison to other methods

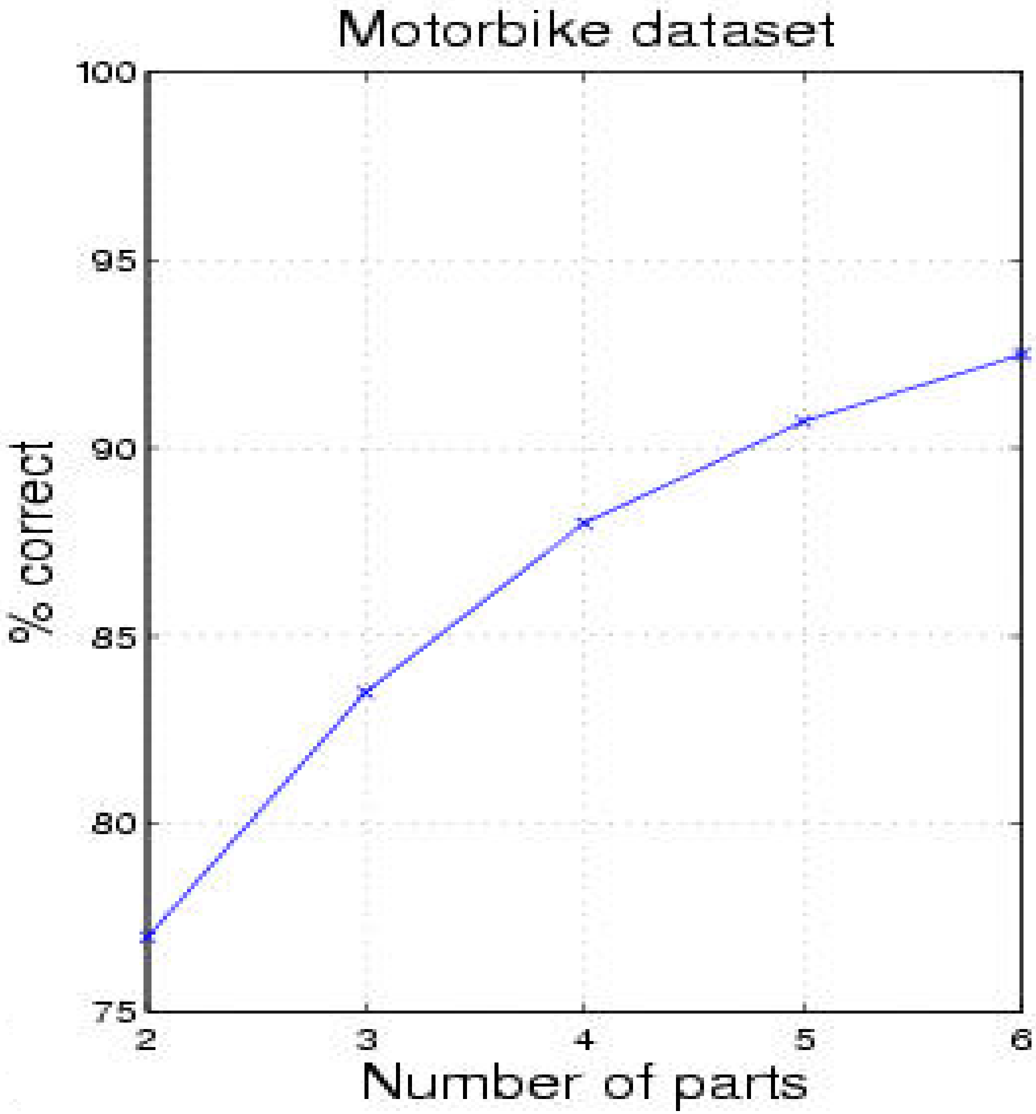
Dataset	Ours	Others	
Motorbikes	7.5	16.0	Weber et al. [ECCV '00]
Faces	4.6	6.0	Weber
Airplanes	9.8	32.0	Weber
Cars (Side)	11.5	21.0	Agarwal Roth [ECCV '02]

% equal error rate

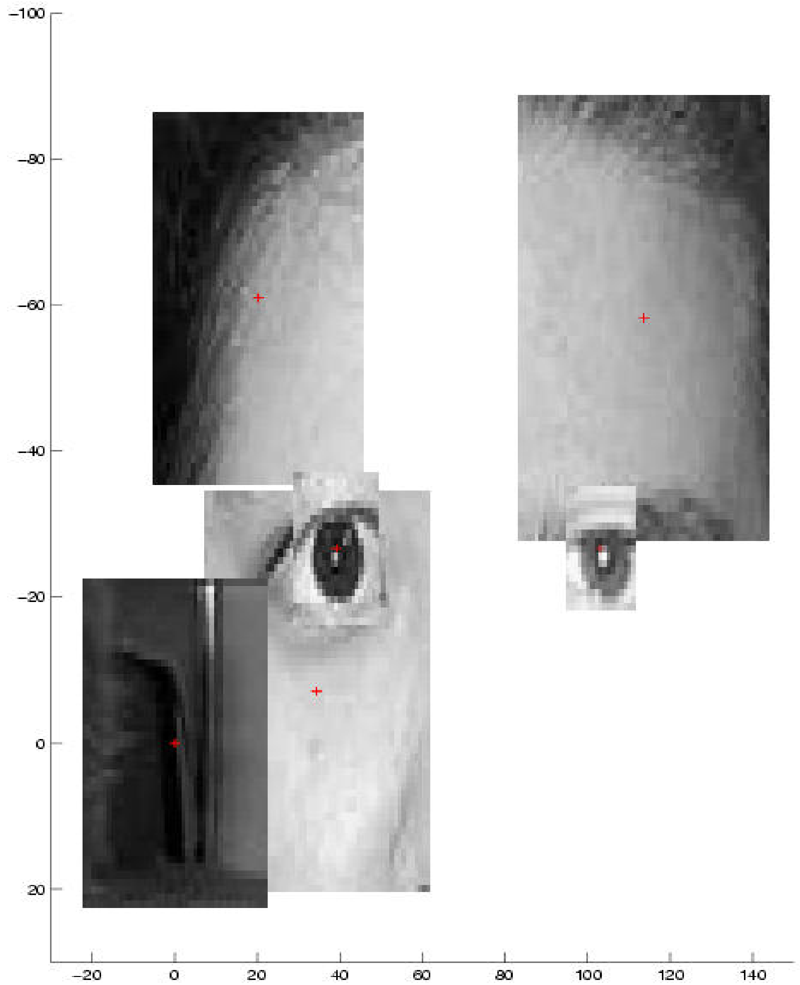


Robustness of Algorithm

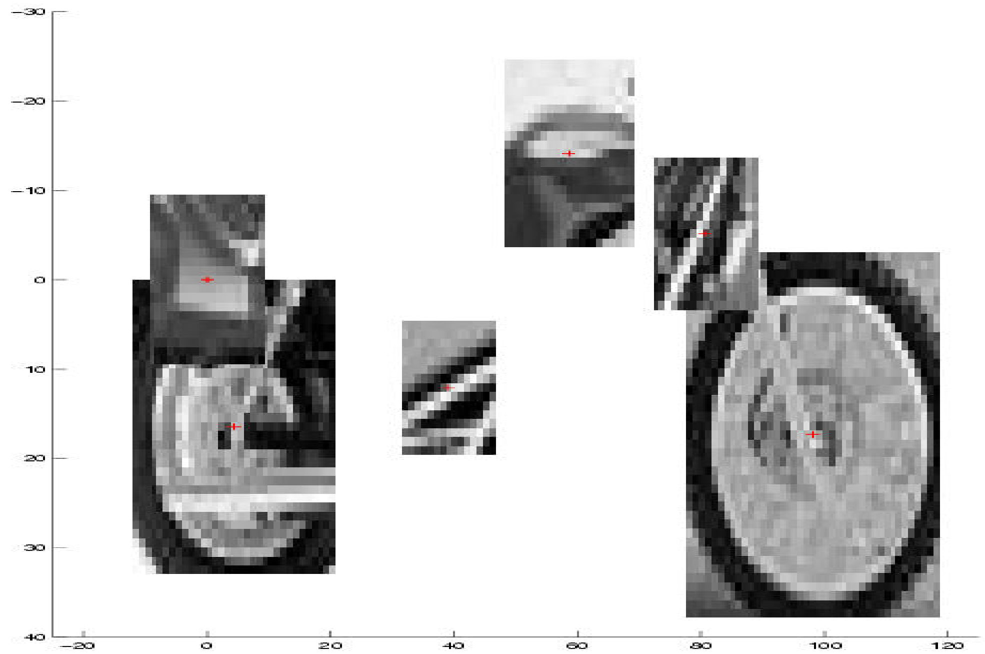




Sampling from models



Faces

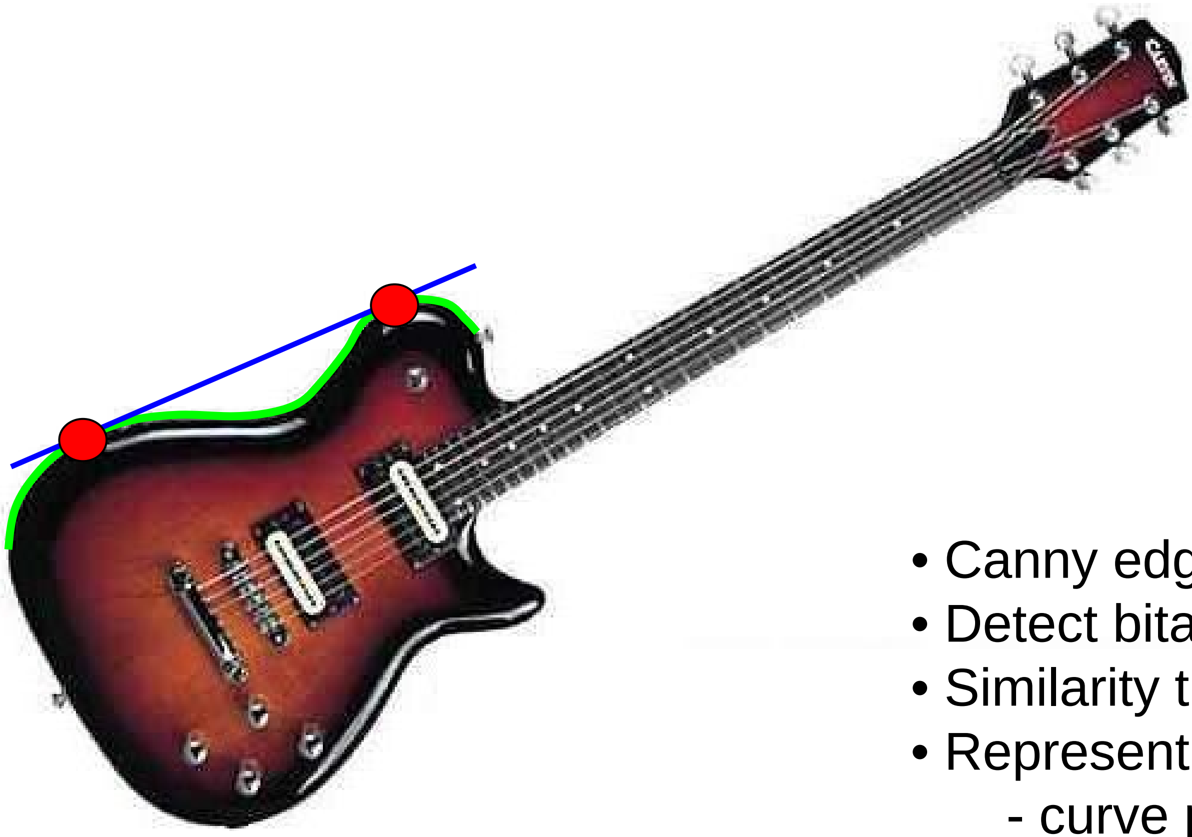


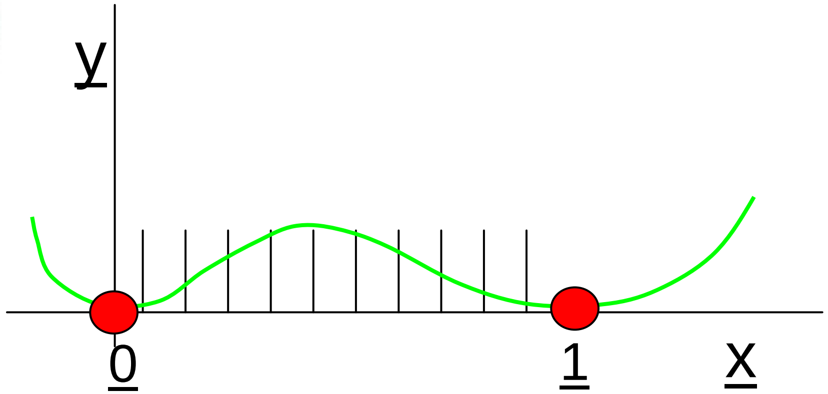
Motorbikes

Extending the Model

Two types of parts:

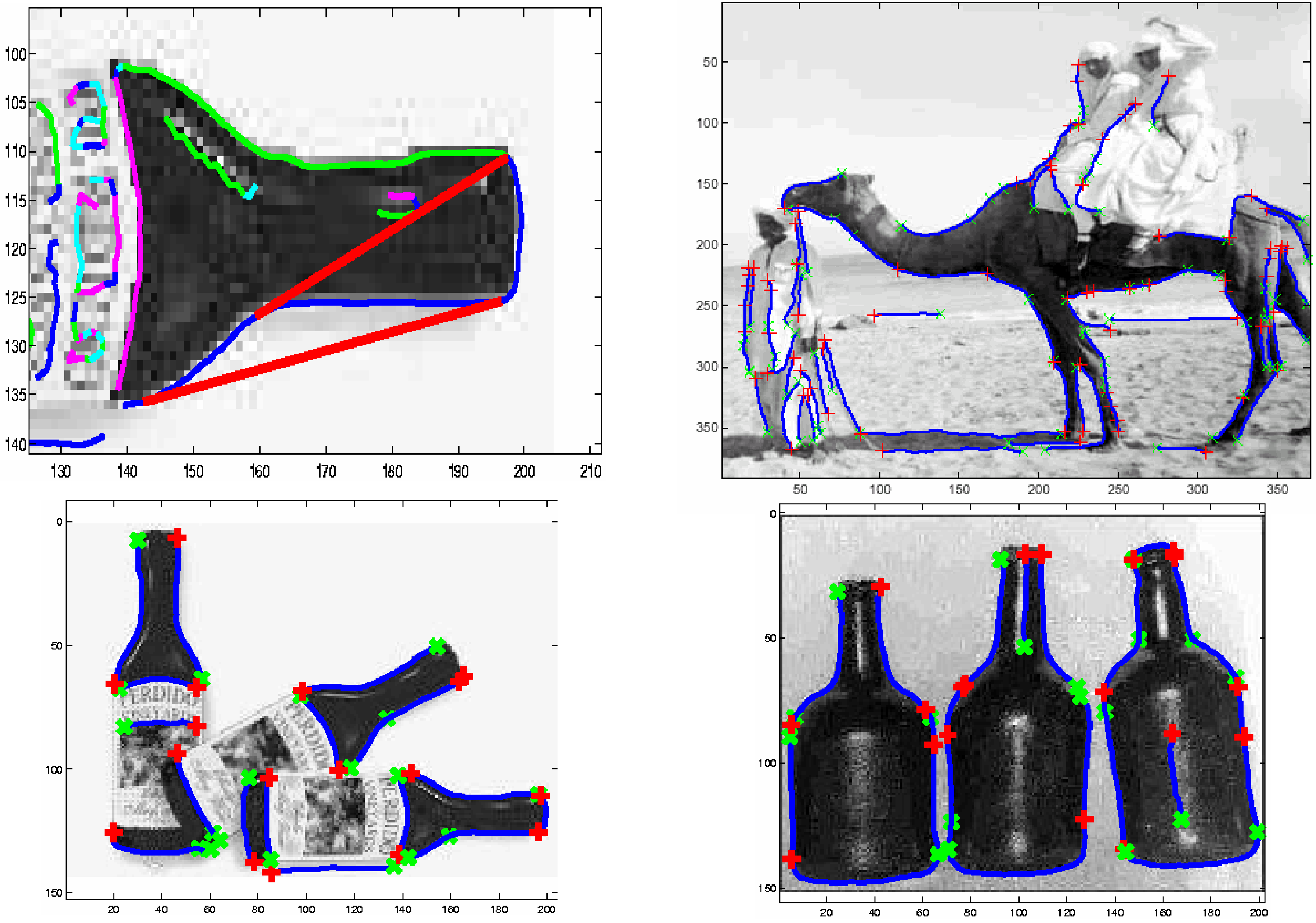
- Appearance patch - scale invariant region operator
- Curve segment - similarity invariant detection and representation





- Canny edge detection – gives edgel chains
- Detect bitangent points
- Similarity transform curve segment
- Represent:
  - curve position (x,y coords. of centroid)
  - curve scale (distance btw. bitangent points)
  - curve shape by 10-vector of y values

Example curves



Fitting the extended model

- Learn models with different combinations of patches and curves
- Choose between models using a validation set
- For the experiments the image datasets are divided into the ratio:
    - 5/12 training
    - 1/6 validation
    - 5/12 testing

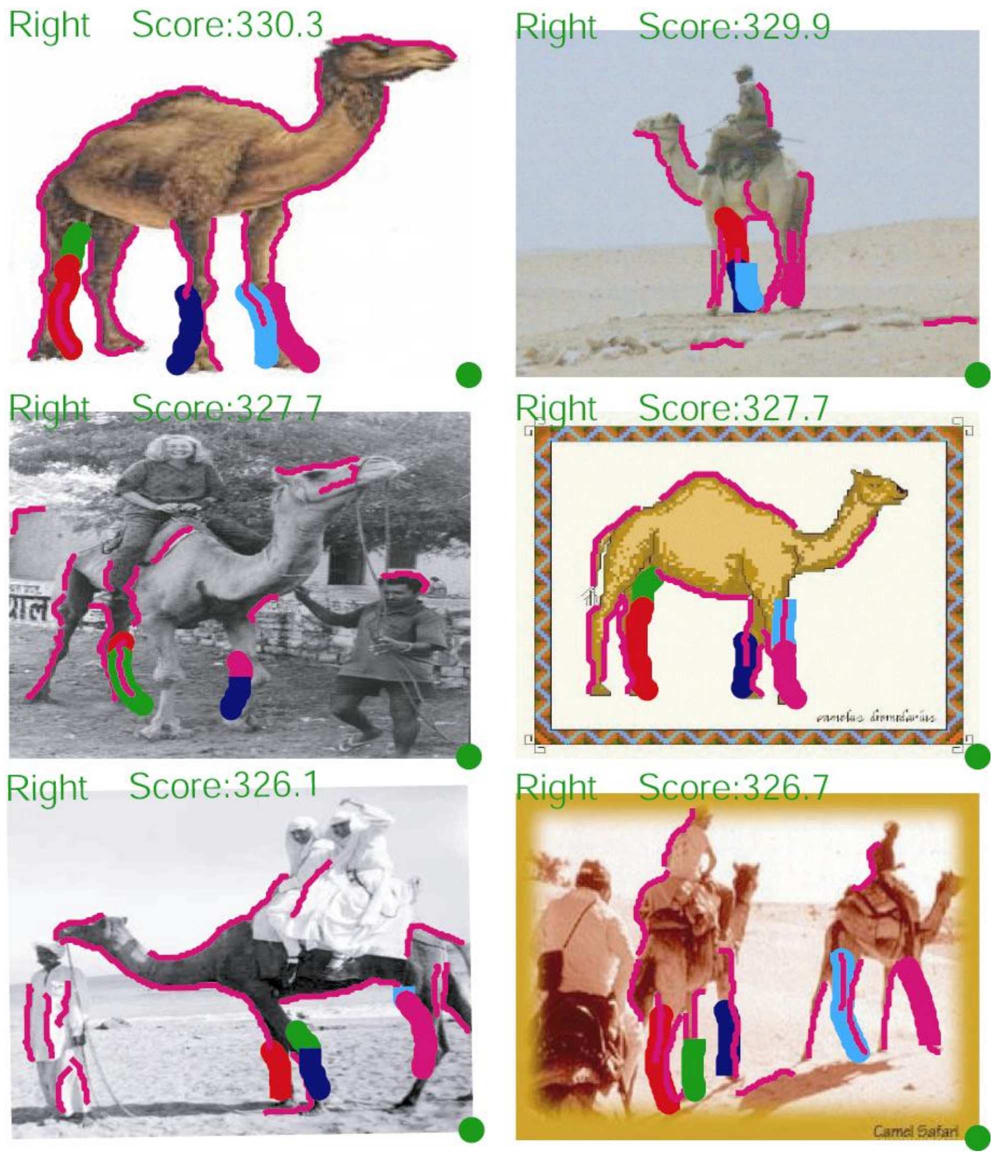
Example datasets



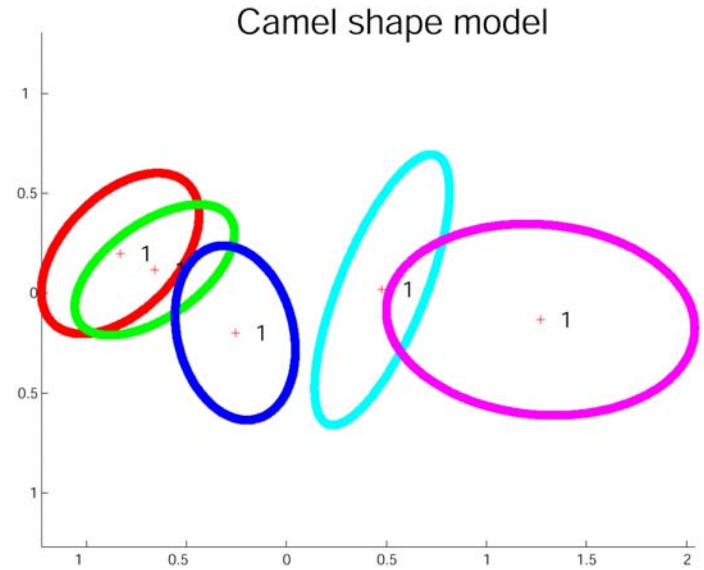


Camels Bottles Zebras

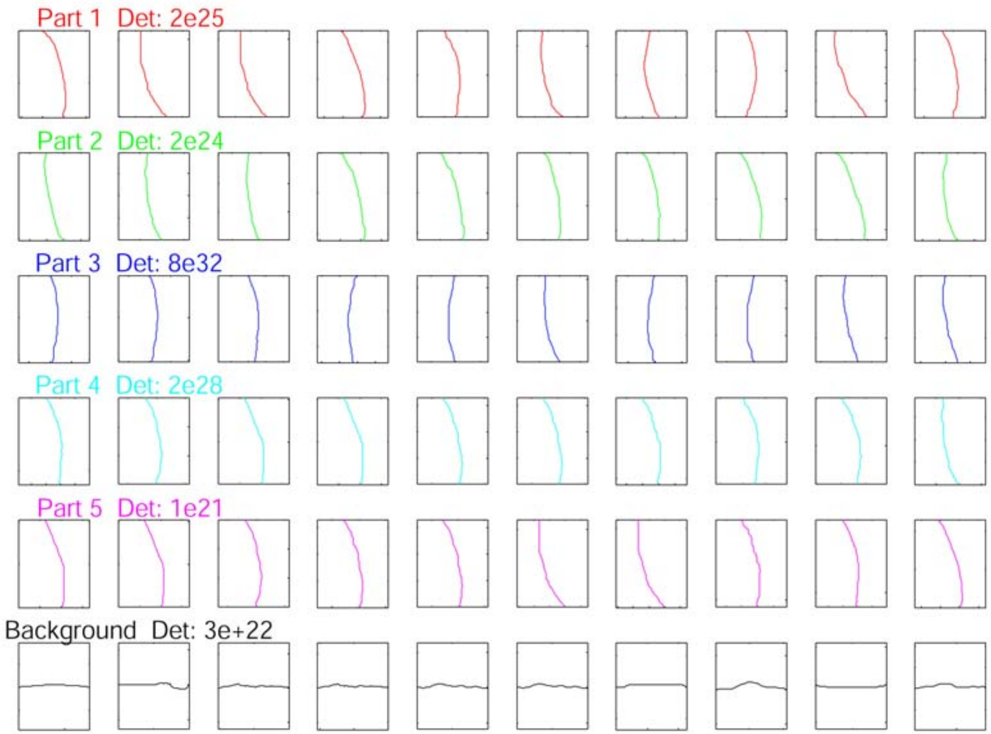
Camels



Right Score:330.3
Right Score:329.9
Right Score:327.7
Right Score:327.7
Right Score:326.1
Right Score:326.7
Camel Safari

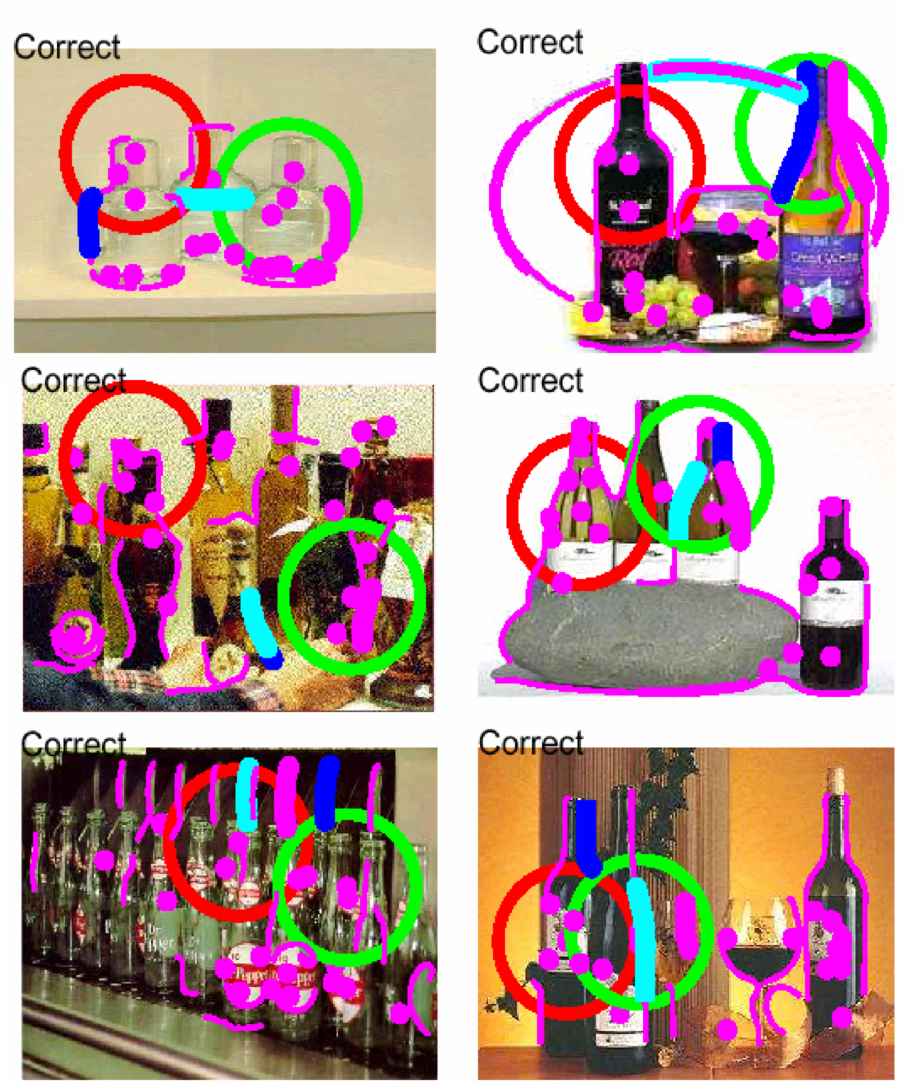


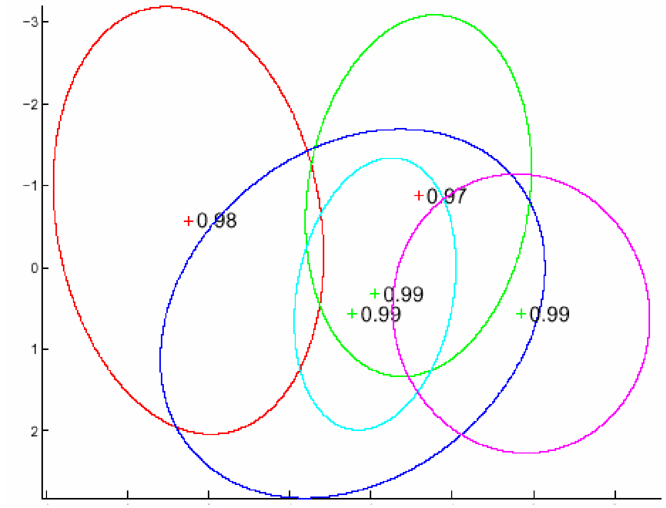
Camel shape model
1
0.5
0
0.5
1
1
0.5
0
0.5
1
1.5
2

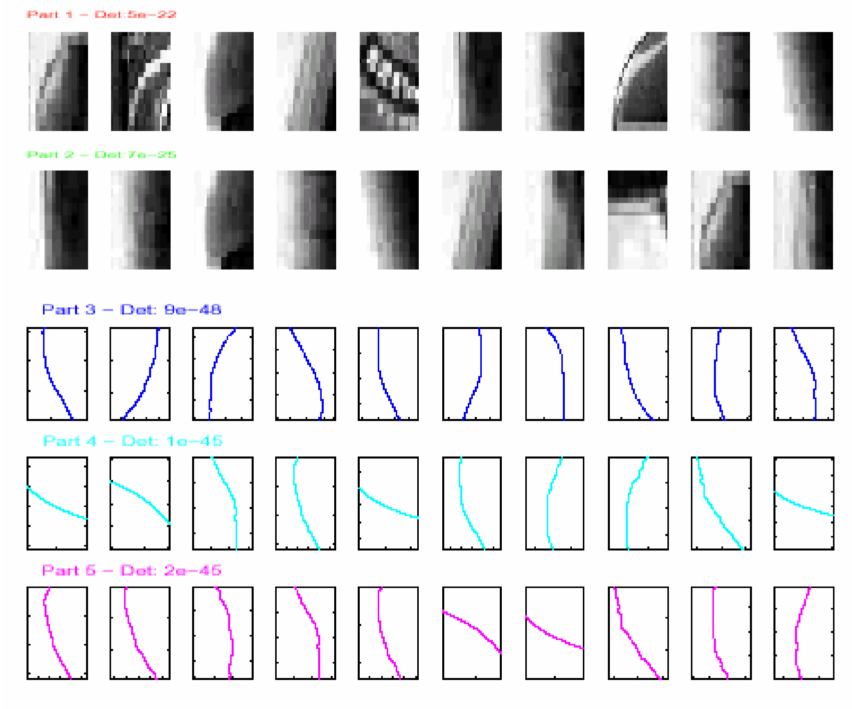


Part 1 Det: 2e25
Part 2 Det: 2e24
Part 3 Det: 8e32
Part 4 Det: 2e28
Part 5 Det: 1e21
Background Det: 3e+22

Bottles using patches and curves







Summary

- Comprehensive probabilistic model for object classes
- Learn appearance, shape, relative scale, occlusion etc. simultaneously in scale and translation invariant manner
- Same algorithm gives <= 10% error across 5 diverse datasets with identical settings

Future work

- Invariance to (affine) viewpoint changes
- Extend to 100’s of object categories
- Reduce training requirements - fewer images

  Use Bayesian methods – ICCV ’03 paper