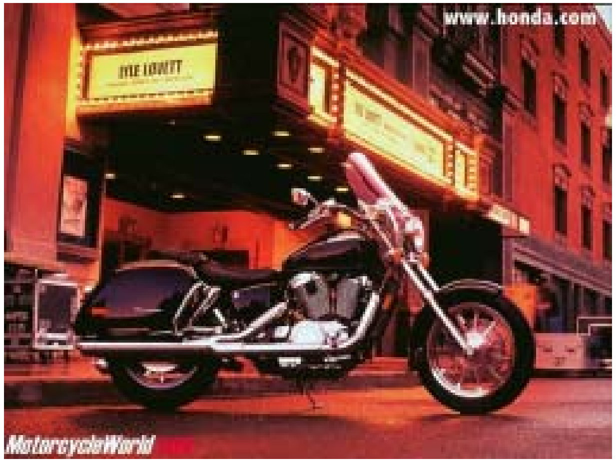
Object Class Recognition by Unsupervised Scale-Invariant Learning

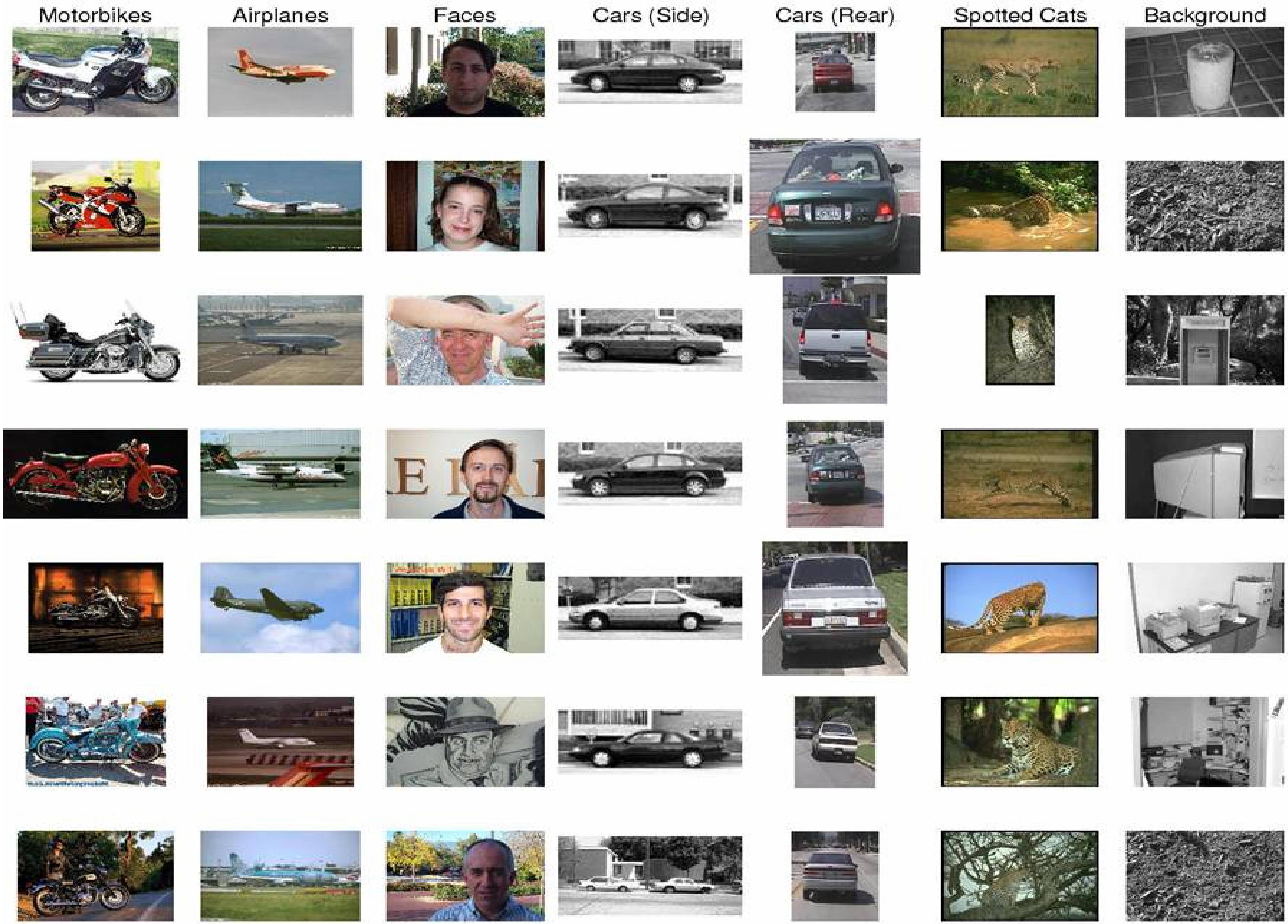
R. Fergus, P. Perona, and A. Zisserman

Presented By Jeff

Goal:

- Enable Computers to Recognize Different Categories of Objects in Images.

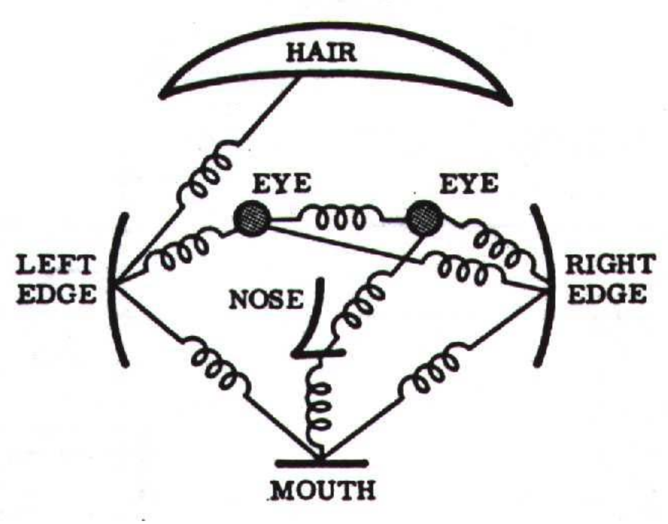



Motorbikes
Airplanes
Faces
Cars (Side)
Cars (Rear)
Spotted Cats
Background

Components

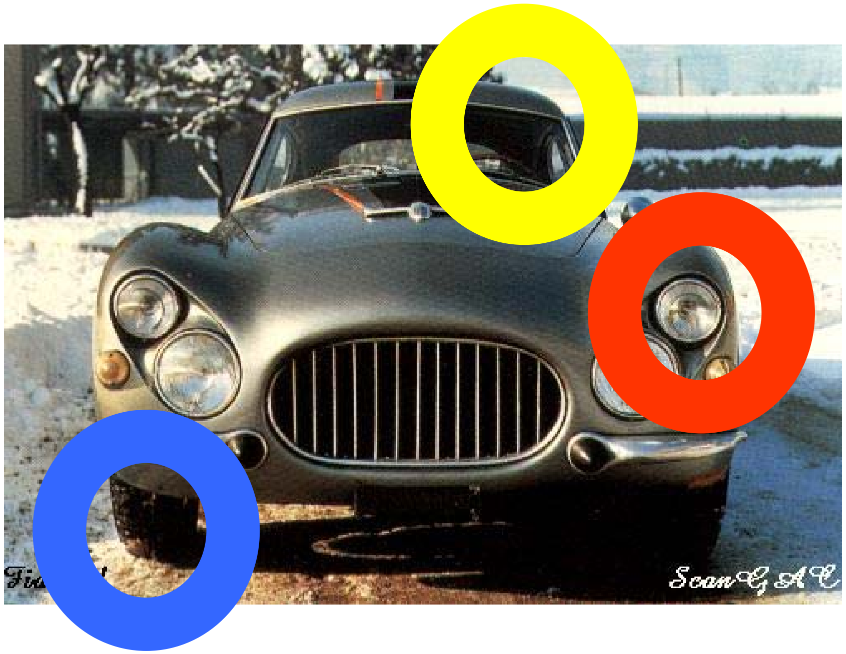
- Model
  - Generative Probabilistic Model
  - Location, Scale, and Appearance
- Learning
  - Estimate Parameters Via EM
- Recognition
  - Evaluate Image Using Model and Threshold

Model: Constellation Of Parts



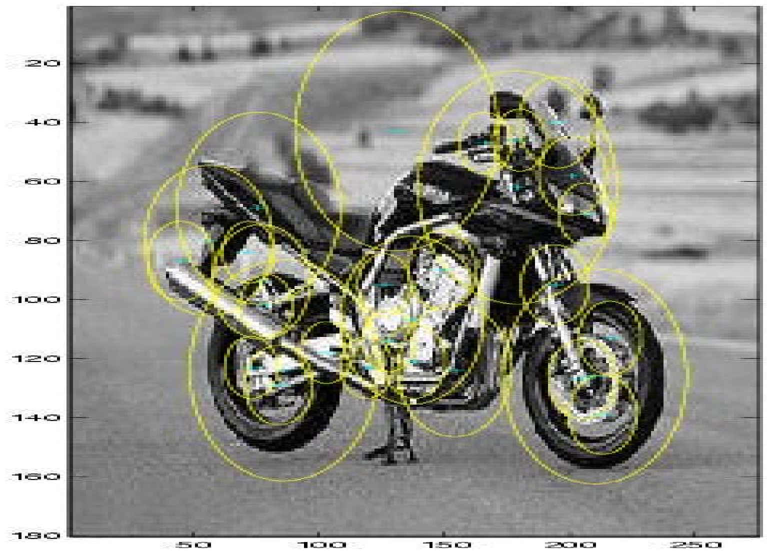
Fischler & Elschlager, 1973

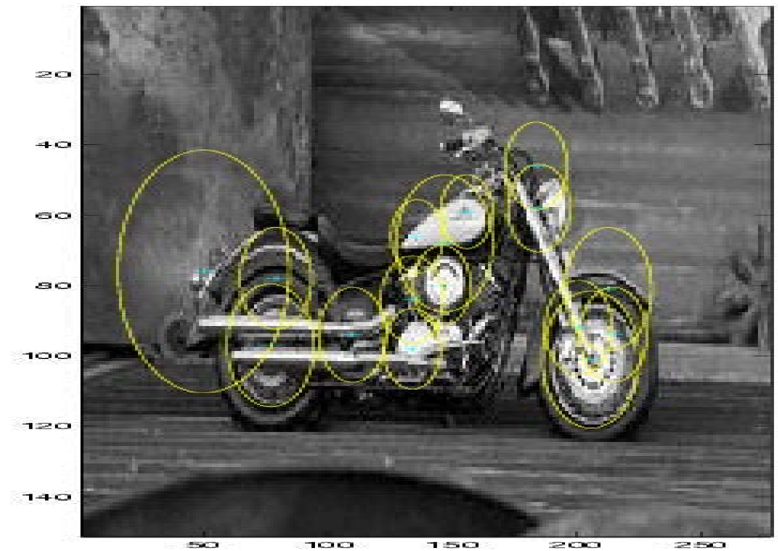
Yuille, 91
Brunelli & Poggio, 93
Lades, v.d. Malsburg et al. 93
Cootes, Lanitis, Taylor et al. 95
Amit & Geman, 95, 99
Perona et al. 95, 96, 98, 00

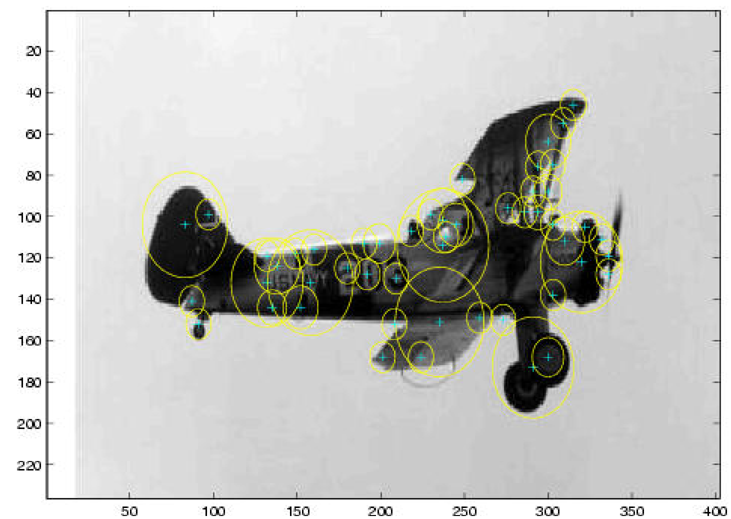


Parts Selected by Interest Operator

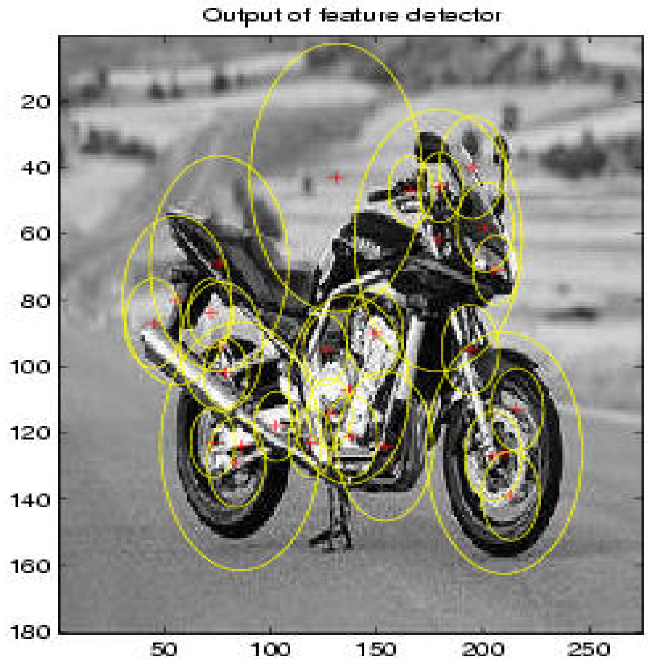
Kadir and Brady's Interest Operator.
Finds Maxima in Entropy Over Scale and Location

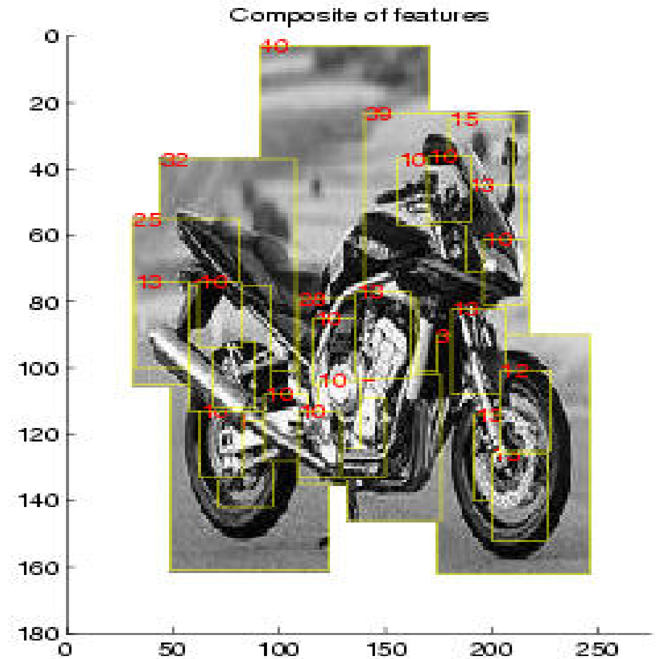




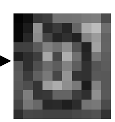


Representation of Appearance

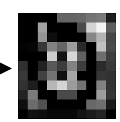




11x11 patch



Normalize



Projection onto PCA basis

$$\begin{pmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \vdots \\ c_{15} \end{pmatrix}$$

Generative Probabilistic Model

Start with Recognition:

$$
\begin{aligned}
R &= \frac{p(\text{Object}|\mathbf{X},\mathbf{S},\mathbf{A})}{p(\text{No object}|\mathbf{X},\mathbf{S},\mathbf{A})} \\
&= \frac{p(\mathbf{X},\mathbf{S},\mathbf{A}|\text{Object})\, p(\text{Object})}{p(\mathbf{X},\mathbf{S},\mathbf{A}|\text{No object})\, p(\text{No object})} \\
&\approx \frac{p(\mathbf{X},\mathbf{S},\mathbf{A}|\,\theta)\, p(\text{Object})}{p(\mathbf{X},\mathbf{S},\mathbf{A}|\theta_{bg})\, p(\text{No object})}
\end{aligned}
$$

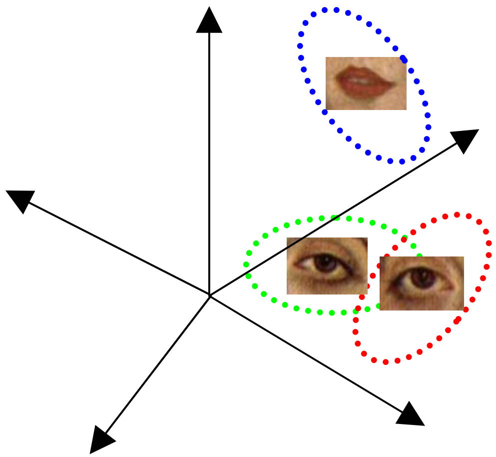
$$
p(\mathbf{X},\mathbf{S},\mathbf{A}|\,\theta) = \sum_{\mathbf{h}\in H} p(\mathbf{X},\mathbf{S},\mathbf{A},\mathbf{h}|\,\theta) =
$$

$$
\sum_{\mathbf{h}\in H} \underbrace{p(\mathbf{A}|\mathbf{X},\mathbf{S},\mathbf{h},\theta)}_{Appearance}\, \underbrace{p(\mathbf{X}|\mathbf{S},\mathbf{h},\theta)}_{Shape}\, \underbrace{p(\mathbf{S}|\mathbf{h},\theta)}_{Rel.\ Scale}\, \underbrace{p(\mathbf{h}|\theta)}_{Other}
$$

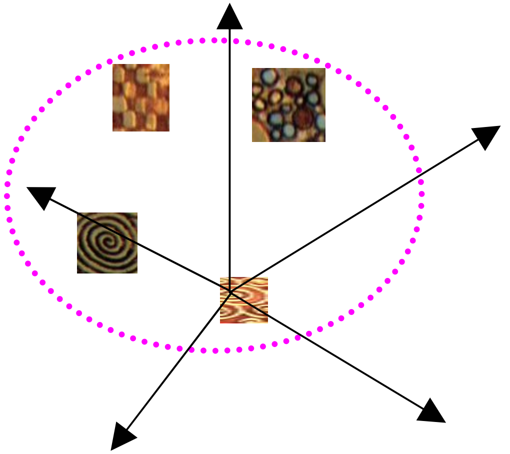
Appearance

$$\frac{p(\mathbf{A}|\mathbf{X},\mathbf{S},\mathbf{h},\theta)}{p(\mathbf{A}|\mathbf{X},\mathbf{S},\mathbf{h},\theta_{bg})} = \prod_{p=1}^{P}\left(\frac{\mathrm{G}(\mathbf{A}(h_p)|\mathbf{c}_p,V_p)}{\mathrm{G}(\mathbf{A}(h_p)|\mathbf{c}_{bg},V_{bg})}\right)^{d_p}$$

Gaussian Part Appearance PDF



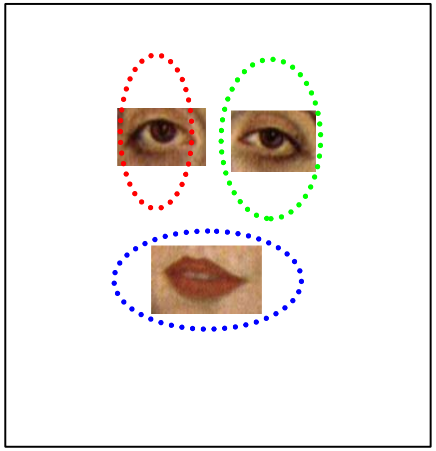
Gaussian Appearance PDF



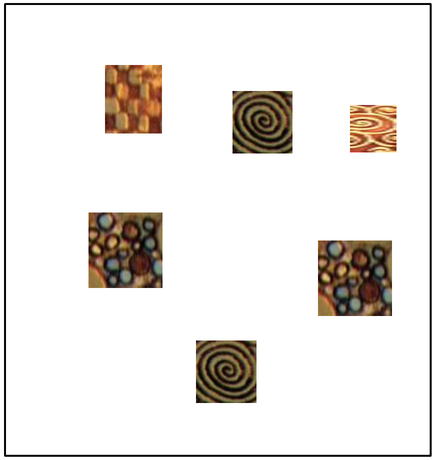
Shape

$$\frac{p(\mathbf{X}|\mathbf{S},\mathbf{h},\theta)}{p(\mathbf{X}|\mathbf{S},\mathbf{h},\theta_{bg})} = \mathrm{G}(\mathbf{X}(\mathbf{h})|\boldsymbol{\mu},\Sigma)\,\alpha^{f}$$

Gaussian Shape PDF



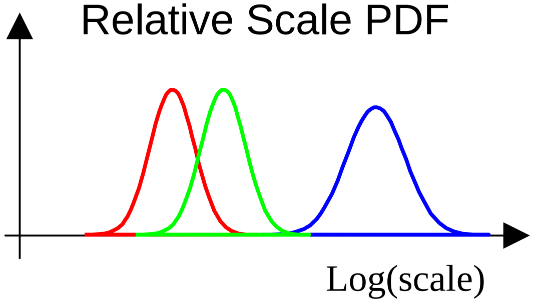
Uniform Shape PDF



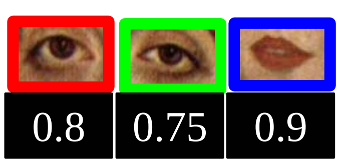
Scale

$$\frac{p(\mathbf{S}|\mathbf{h},\theta)}{p(\mathbf{S}|\mathbf{h},\theta_{bg})} = \prod_{p=1}^{P} \mathrm{G}(\mathbf{S}(h_p)|t_p, U_p)^{d_p}\ r^f$$

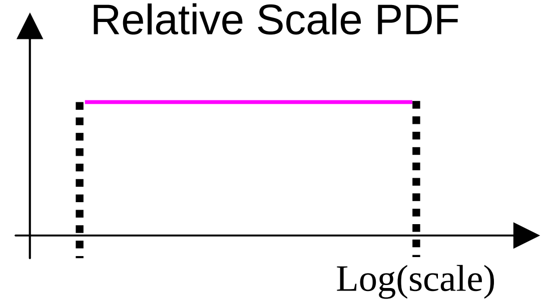
Gaussian
Relative Scale PDF



Prob. of detection



Uniform
Relative Scale PDF



Poission PDF On # Detections

Occlusion and Part Statistics

$$\frac{p(\mathbf{h}|\theta)}{p(\mathbf{h}|\theta_{bg})} = \frac{p_{Poiss}(n|M)}{p_{Poiss}(N|M)} \frac{1}{{}^nC_r(N,f)} p(\mathbf{d}|\theta)$$

Learning

- Train Model Parameters Using EM:
  - Optimize Parameters
  - Optimize Assignments
  - Repeat Until Convergence

$$\theta = \{\boldsymbol{\mu}, \Sigma, \mathbf{c}, V, M, p(\mathbf{d}|\theta), t, U\}$$

$$\hat{\theta}_{ML} = \underset{\theta}{arg\,max}\; p(\mathbf{X}, \mathbf{S}, \mathbf{A}|\,\theta)$$

Recognition

Make This:

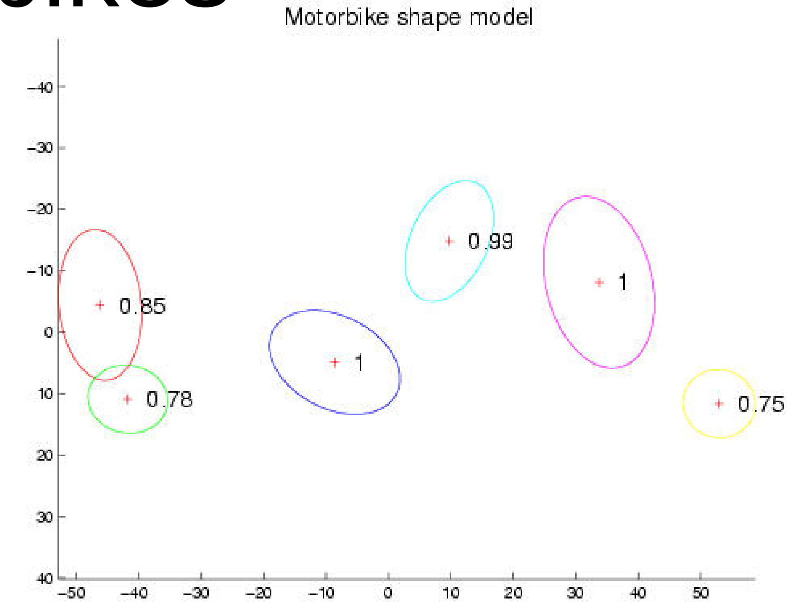
$$
\begin{aligned}
R &= \frac{p(\text{Object}|\mathbf{X},\mathbf{S},\mathbf{A})}{p(\text{No object}|\mathbf{X},\mathbf{S},\mathbf{A})} \\
&= \frac{p(\mathbf{X},\mathbf{S},\mathbf{A}|\text{Object})\,p(\text{Object})}{p(\mathbf{X},\mathbf{S},\mathbf{A}|\text{No object})\,p(\text{No object})} \\
&\approx \frac{p(\mathbf{X},\mathbf{S},\mathbf{A}|\,\theta)\,p(\text{Object})}{p(\mathbf{X},\mathbf{S},\mathbf{A}|\theta_{bg})\,p(\text{No object})}
\end{aligned}
$$

Greater Than Threshold

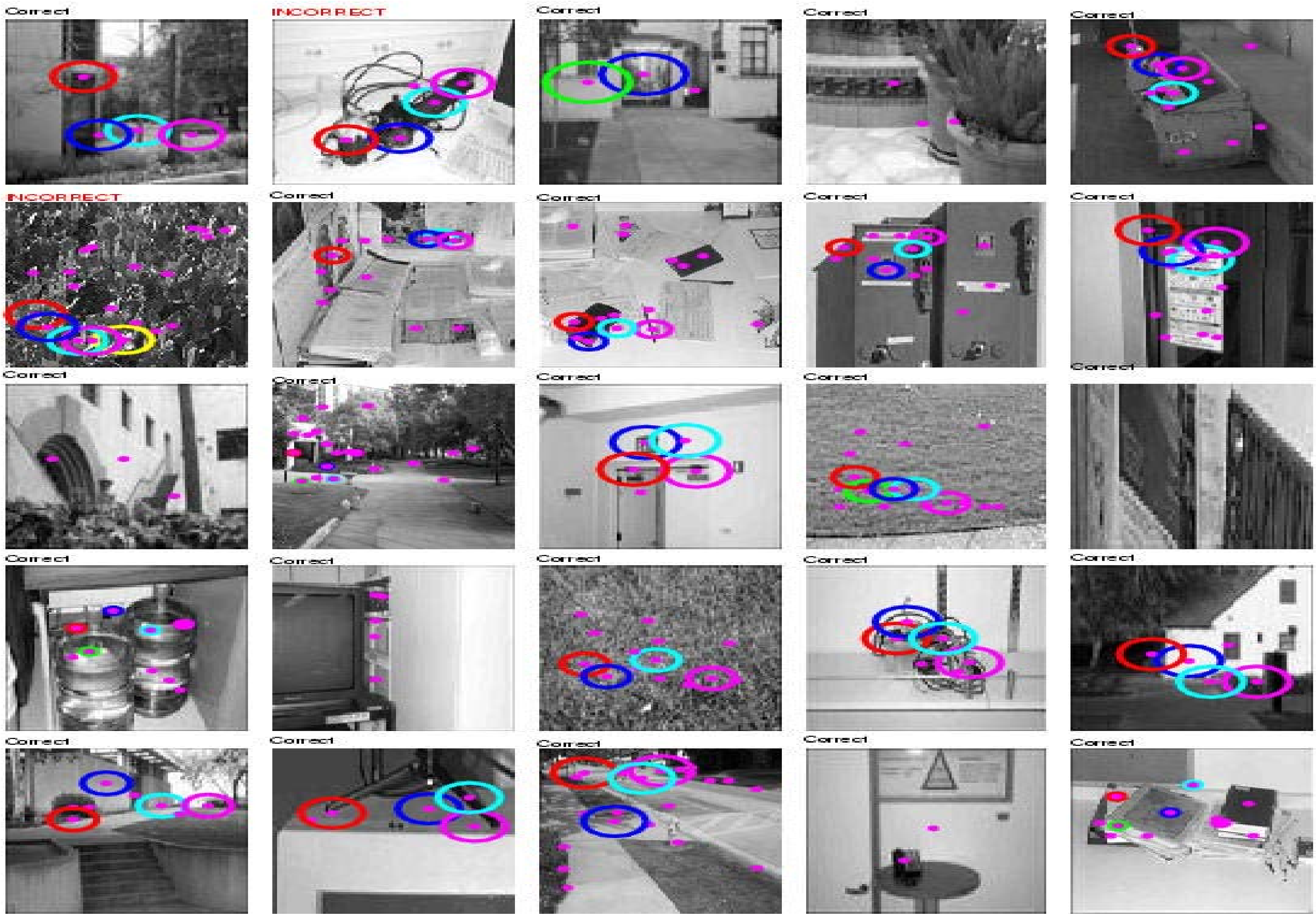
RESULTS

Motorbikes

Equal error rate: 7.5%

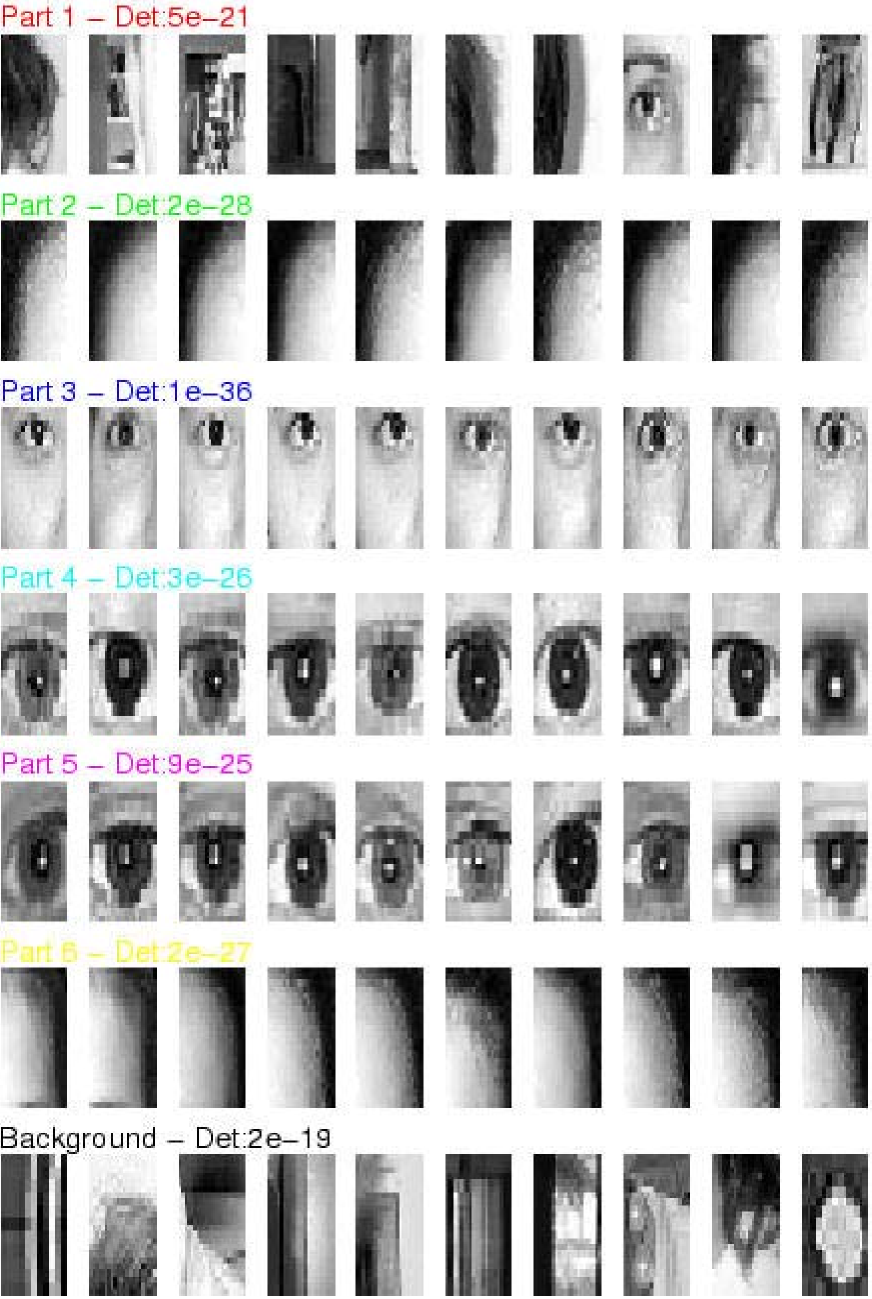


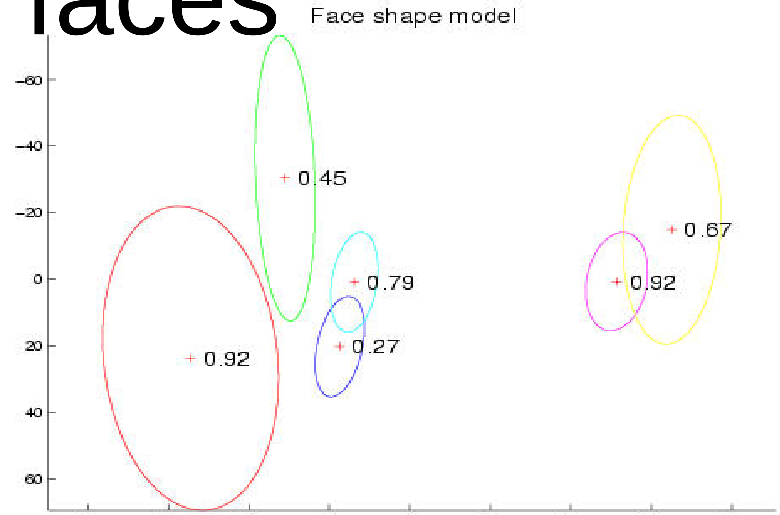
Background Images



Frontal faces

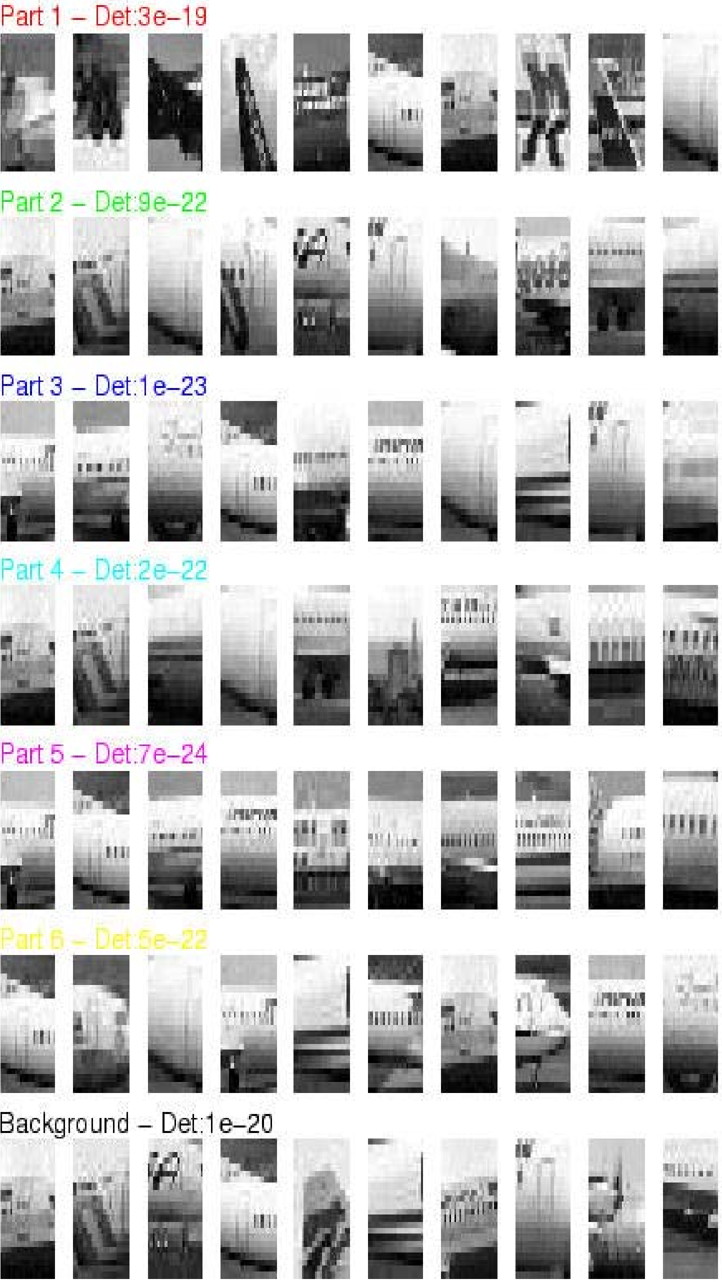
Equal error rate: 4.6%

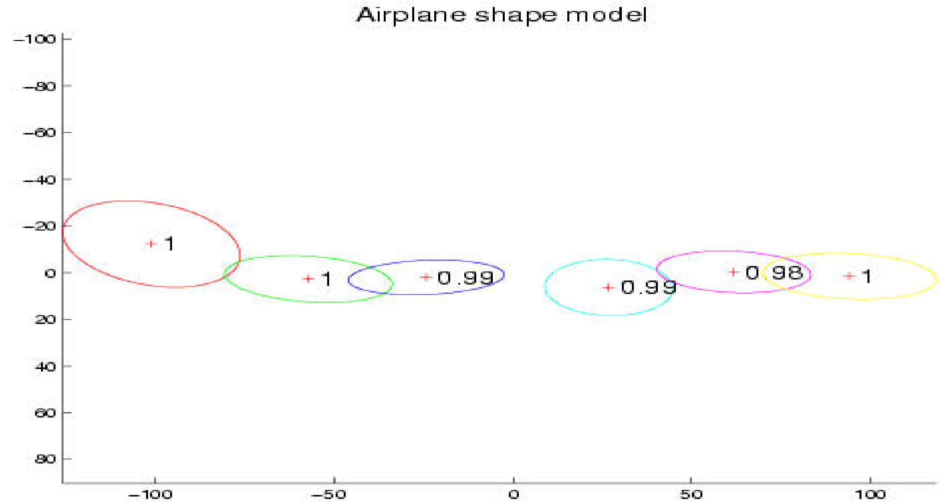


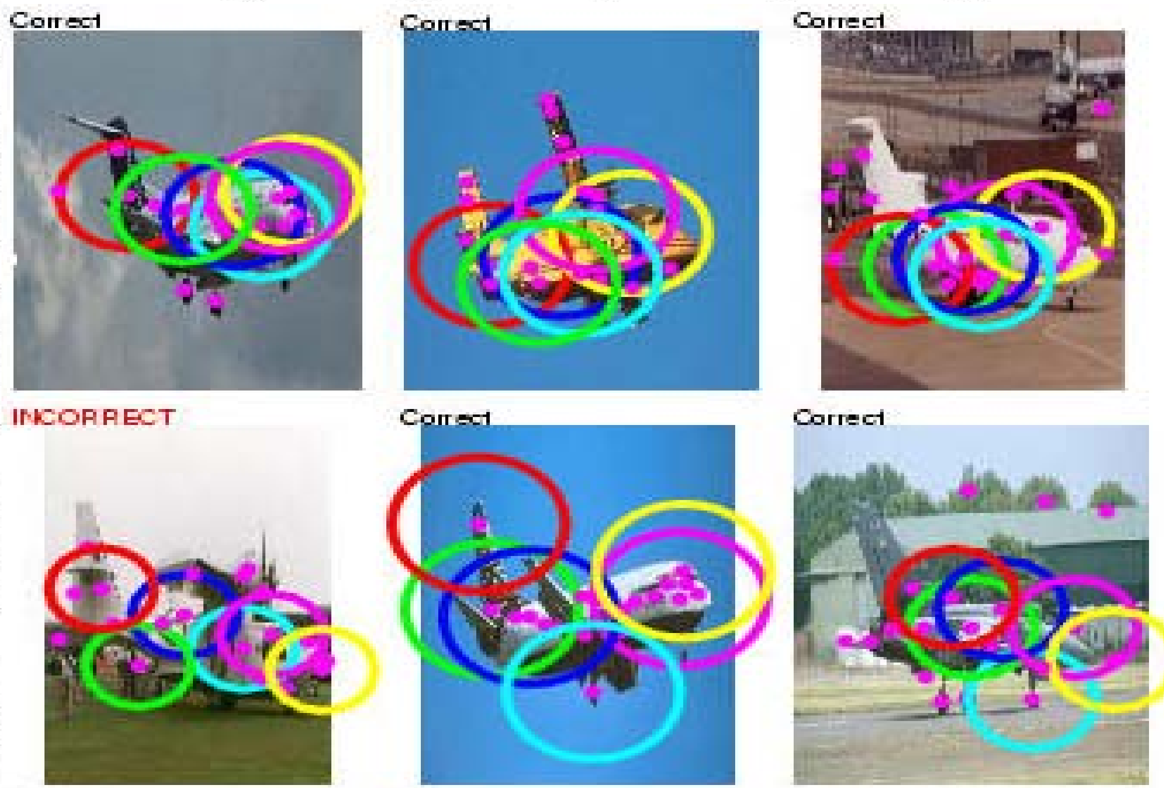


Equal error rate: 9.8%

Airplanes

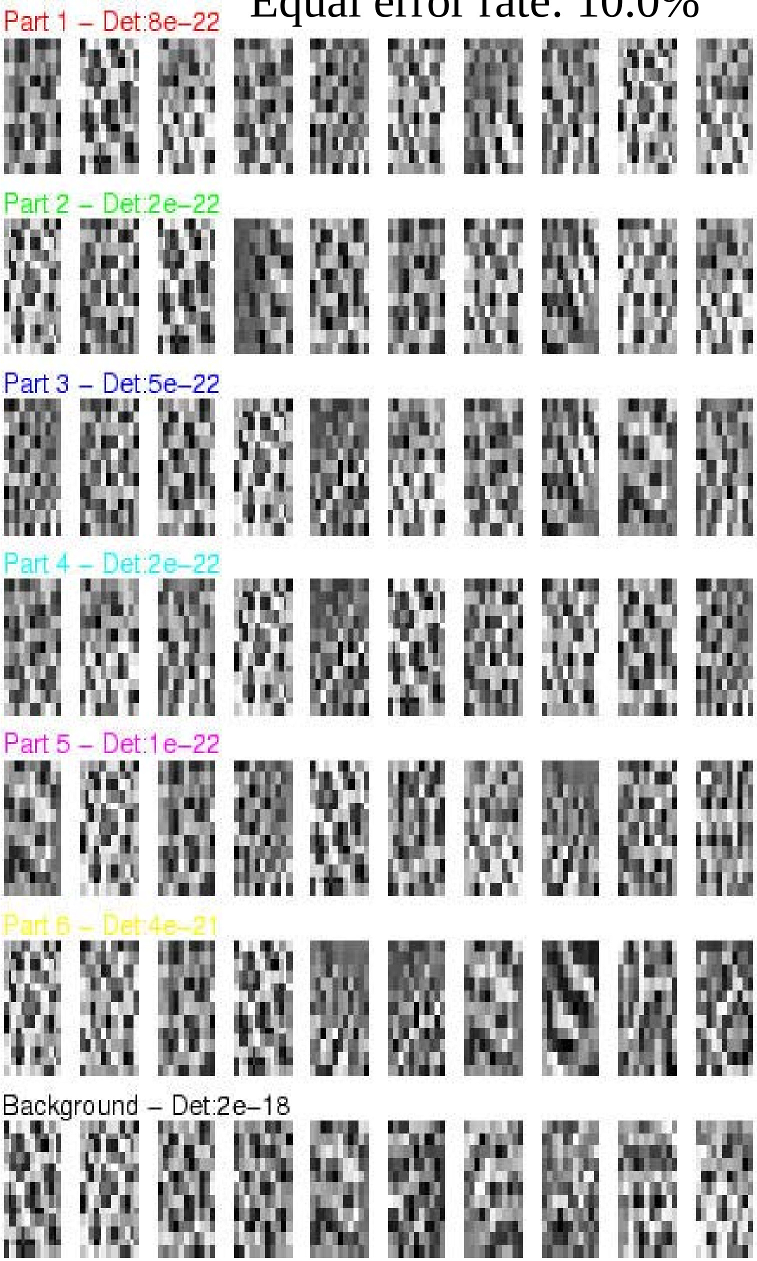


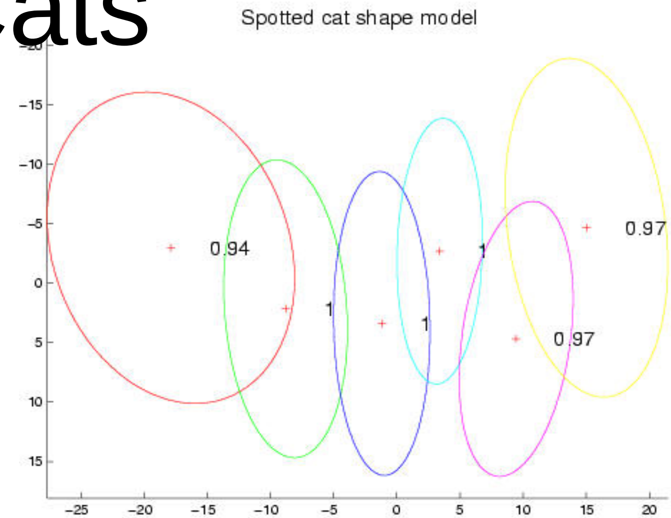


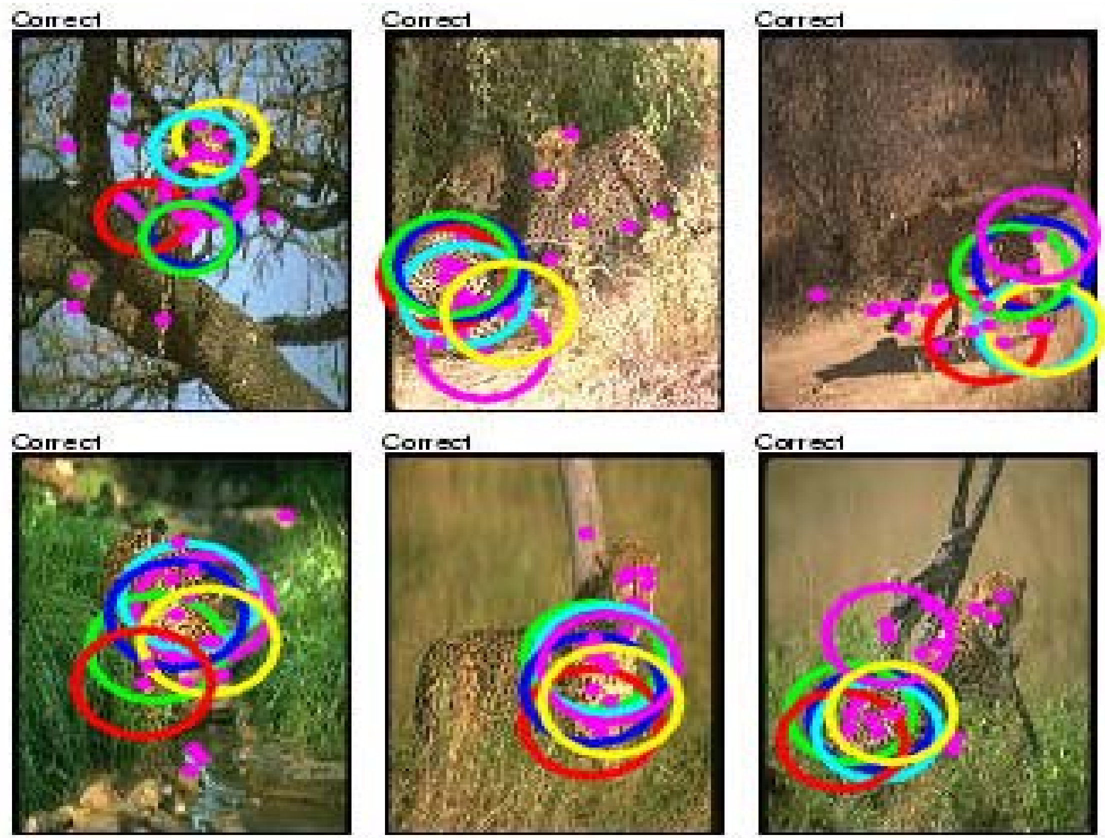


Scale-Invariant Cats

Equal error rate: 10.0%

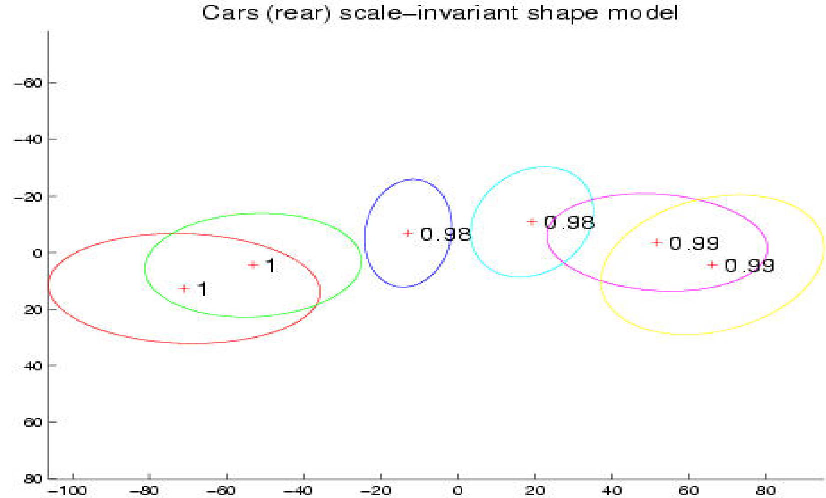


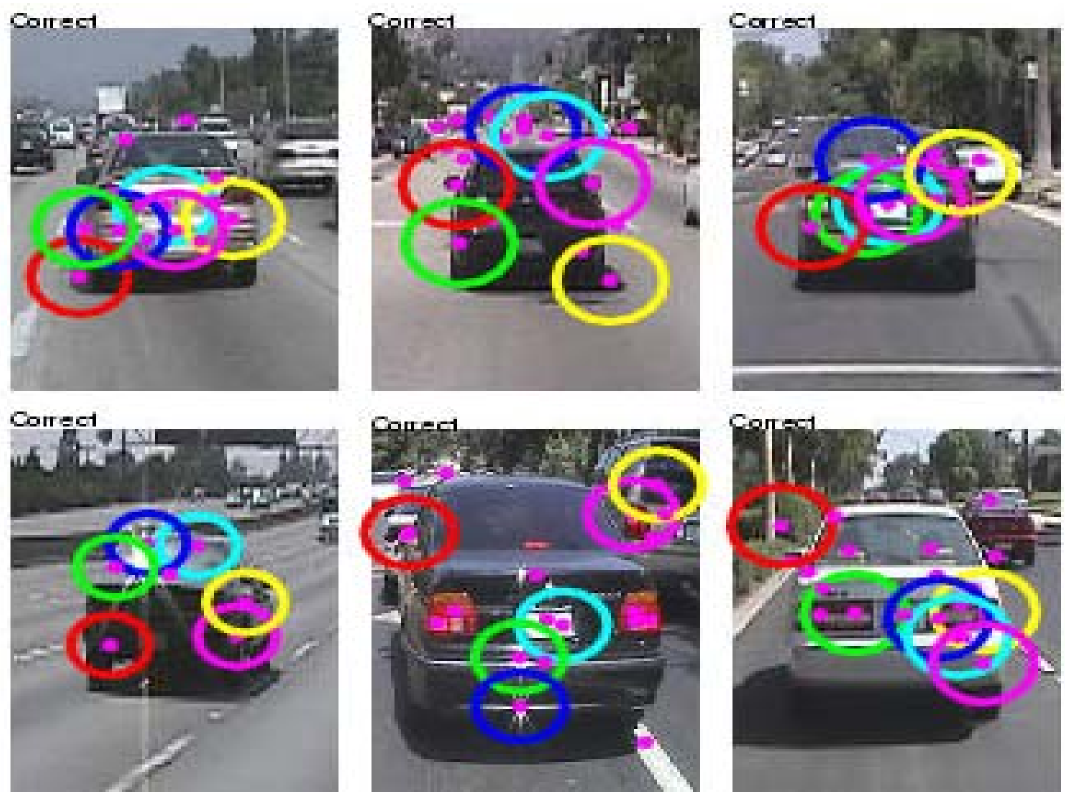




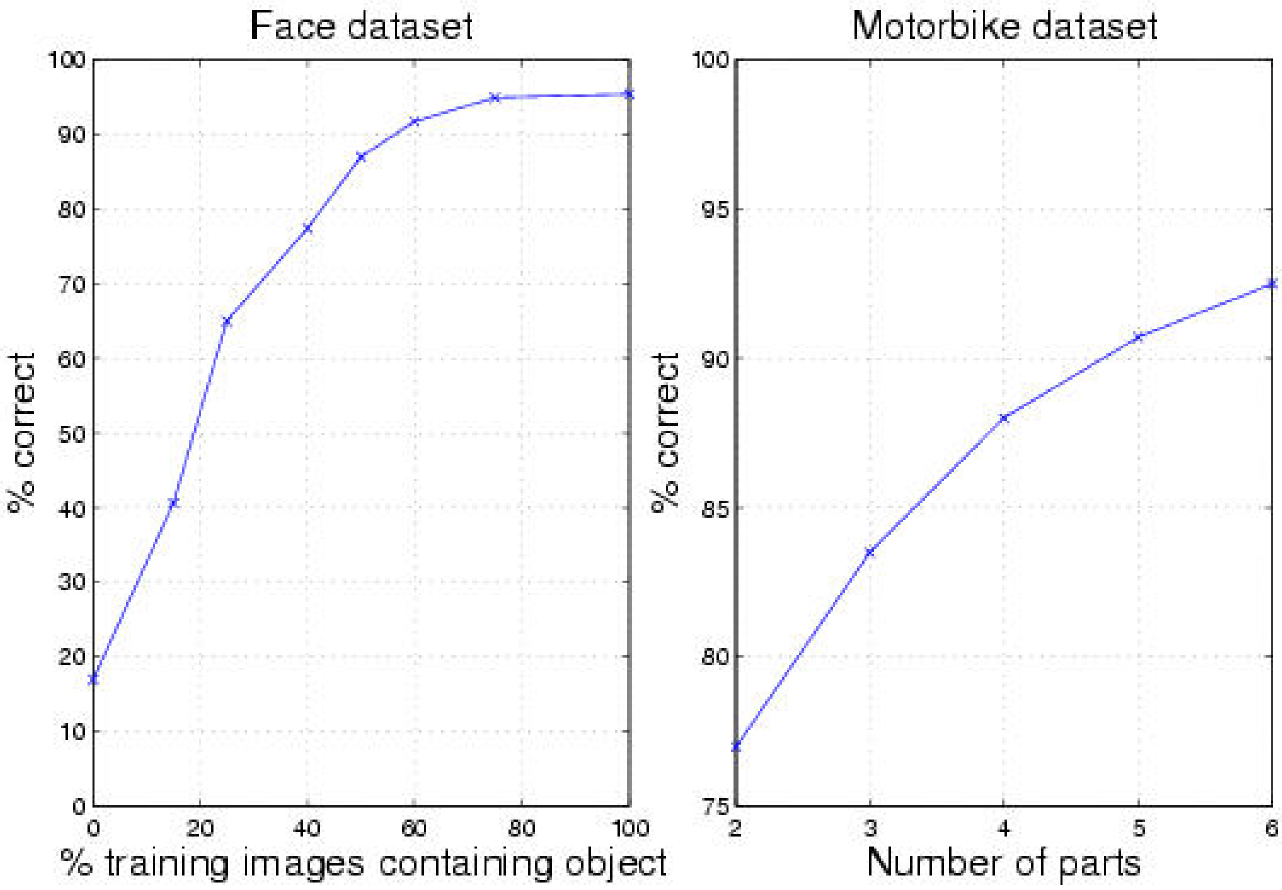
Scale-Invariant cars

Equal error rate: 9.7%





Robustness of Algorithm



ROC equal error rates

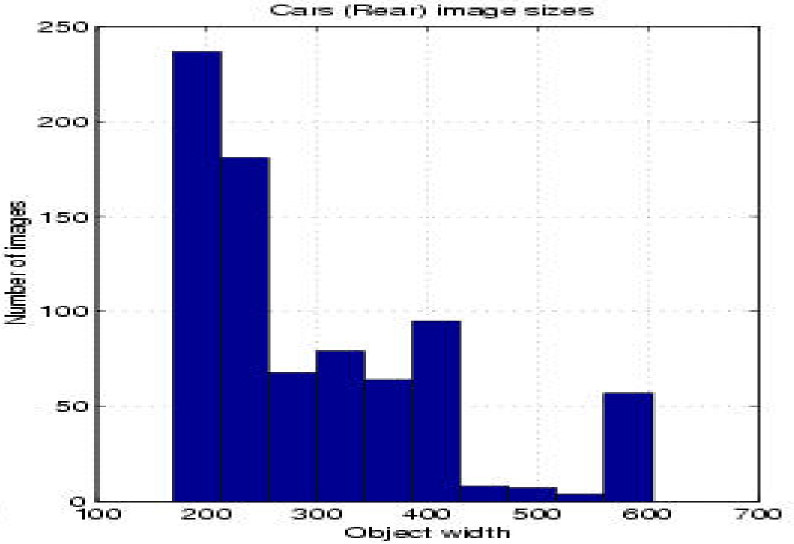
Pre-Scaled Data (Identical Settings):

				Model		
Dataset	Total size of dataset	$\sim$ Object width (pixels)	Motorbikes	Faces	Airplanes	Spotted Cats
Motorbikes	800	200	92.5	50	51	56
Faces	435	300	33	96.4	32	32
Airplanes	800	300	64	63	90.2	53
Spotted Cats	200	80	48	44	51	90 .0

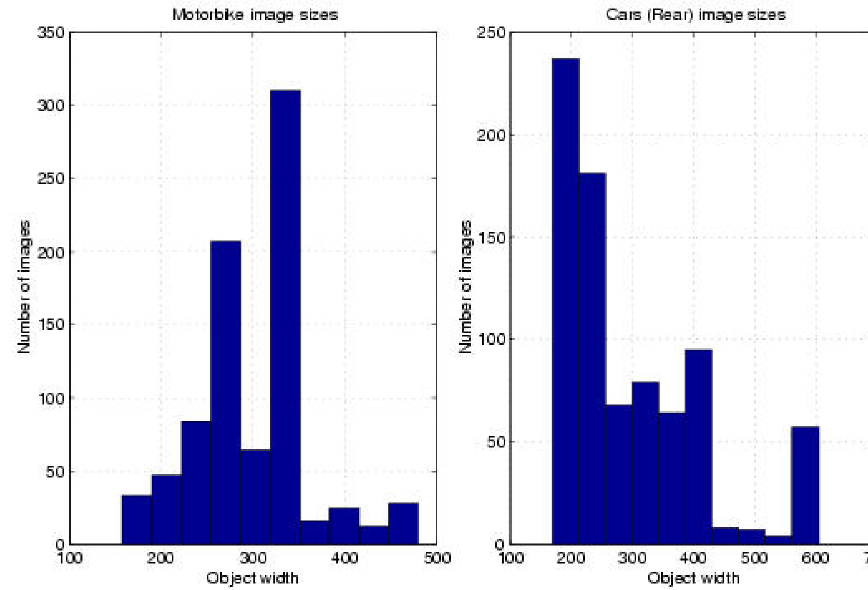
Scale-Invariant Learning and Recognition:

Dataset	Total size of dataset	Object size range (pixels)	Pre-scaled performance	Unscaled performance
Motorbikes	800	200-480	95.0	93.3
Airplanes	800	200-500	94.0	93.0
Cars (Rear)	800	100-550	84.8	90.3

Scale-Invariant Cars



The End


Motorbike image sizes
Number of images
Object width
Cars (Rear) image sizes
Number of images
Object width