

Object class recognition using unsupervised scale-invariant learning

Rob Fergus
Pietro Perona
Andrew Zisserman

Oxford University
California Institute of Technology

Overview

Task: Recognition of object categories



3 main issues:

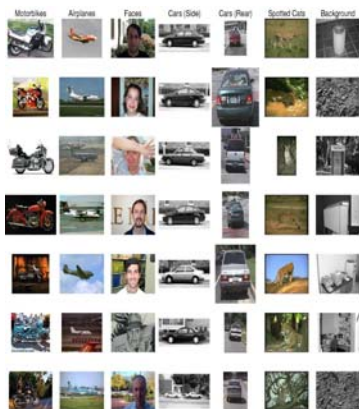
- 1 Representation
- 2 Learning
- 3 Recognition

Some object categories

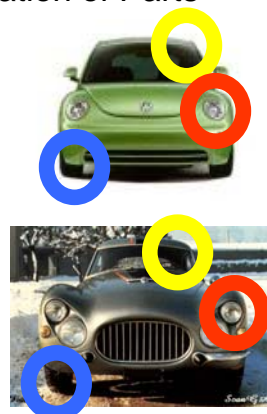
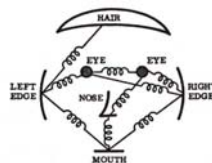
Learn from just examples

Difficulties:

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



Model: constellation of Parts



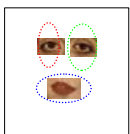
Fischler & Elschlager, 1973

- Yuille, 1991
- Brunelli & Poggio, 1993
- Lades, v.d. Malsburg et al. 1993
- Cootes, Lanitis, Taylor et al. 1995
- Amit & Geman, 1995, 1999
- Perona et al. 1995, 1996, 1998, 2000

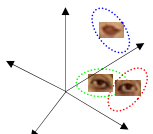
Generative probabilistic model

Foreground model

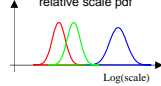
Gaussian shape pdf



Gaussian part appearance pdf



Gaussian relative scale pdf



Prob. of detection

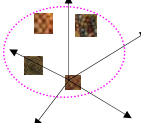


Clutter model

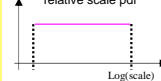
Uniform shape pdf



Gaussian appearance pdf



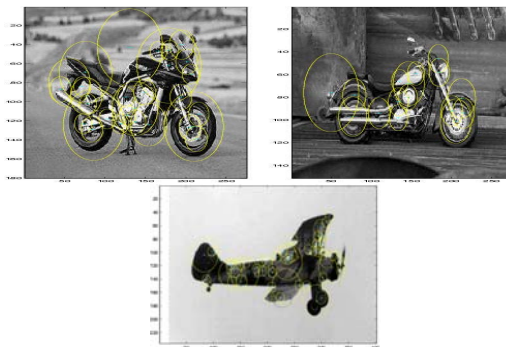
Uniform relative scale pdf



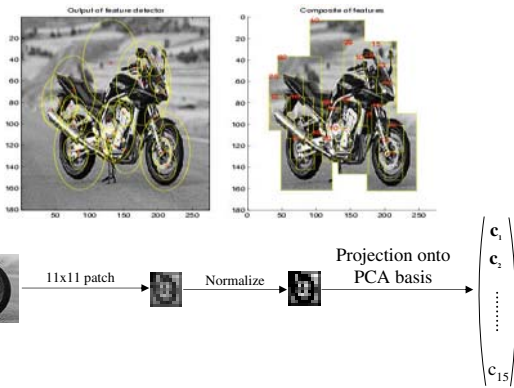
Poisson pdf on # detections

Interest Operator

Kadir and Brady's interest operator.
Finds maxima in entropy over scale and location



Representation of appearance



Experimental procedure

Two series of experiments:

- 1 Scale variant (using pre-scaled images)
- 2 Scale invariant

Datasets:

- ▣ Motorbikes, Faces, Spotted cats, Airplanes, Cars from behind and side
- ▣ Between 200 and 800 images in size



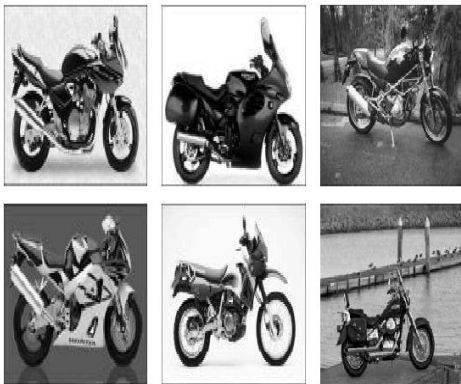
Training

- ▣ 50% images
- ▣ No identification of object within image

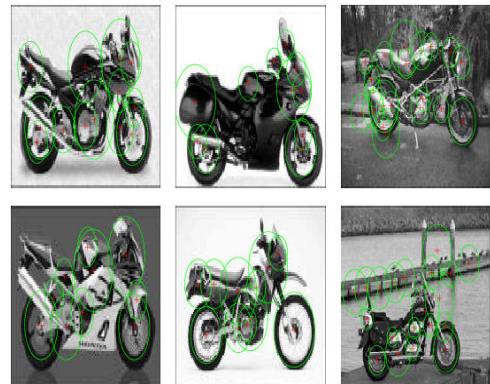
Testing

- ▣ 50% images
- ▣ Simple object present/absent test
- ▣ ROC equal error rate computed, using background set of images

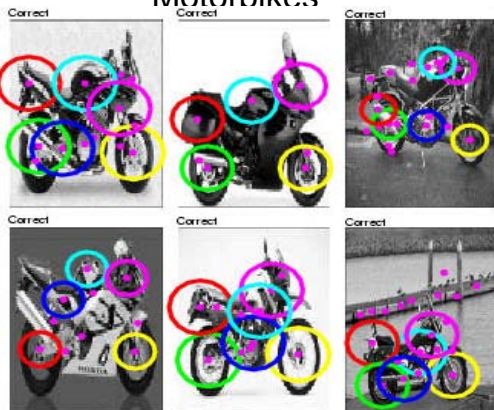
Motorbikes



Motorbikes



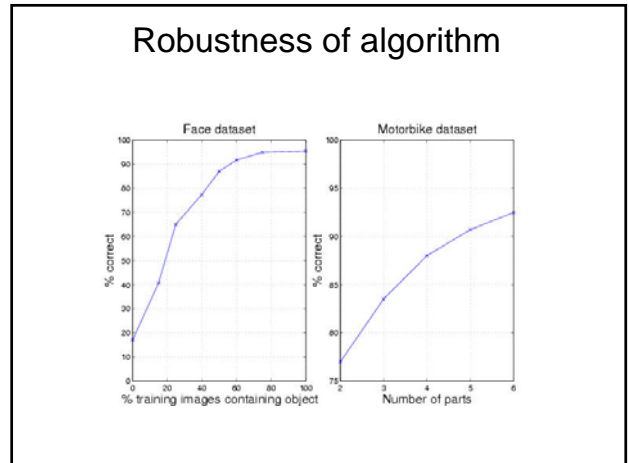
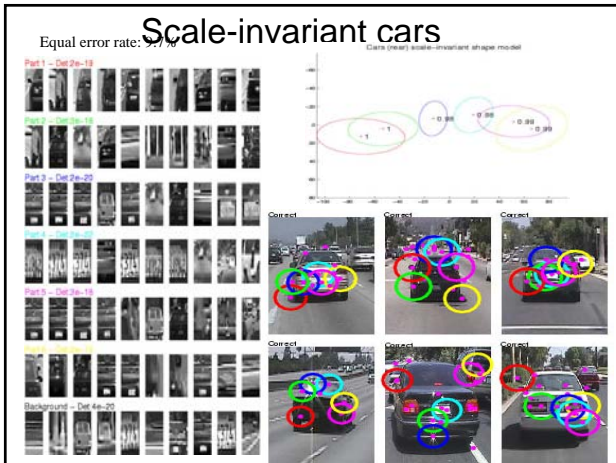
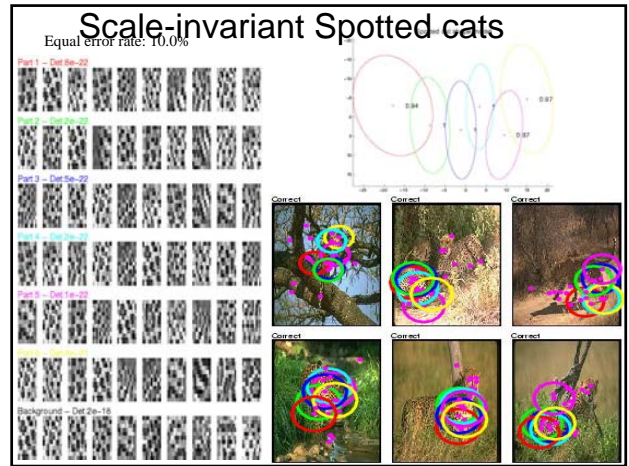
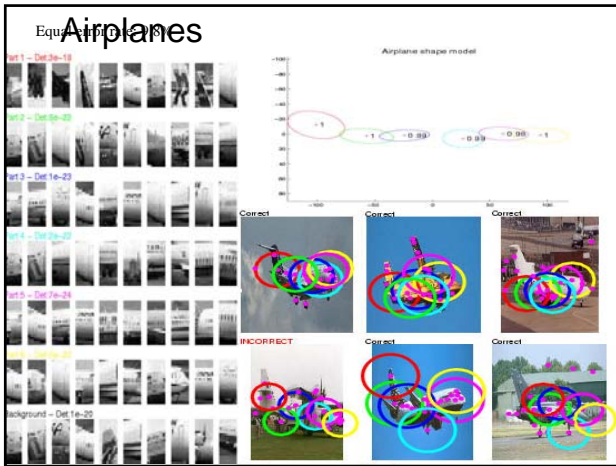
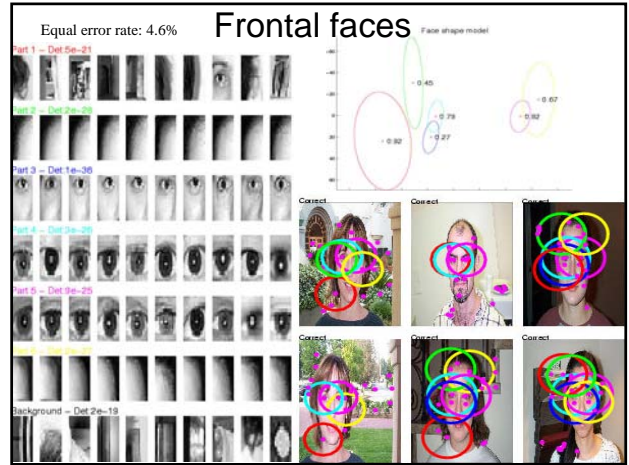
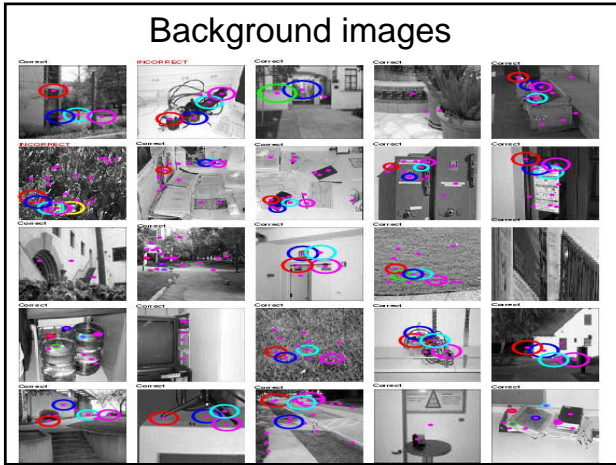
Motorbikes



Equal error rate: 7.5%

Motorbikes





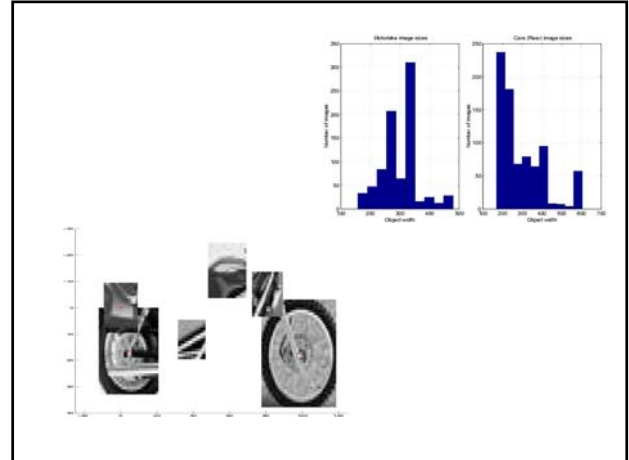
ROC equal error rates

Pre-scaled data (identical settings):

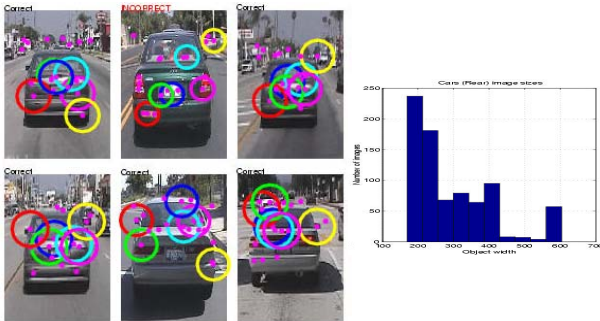
Dataset	Total size of dataset	~ Object width (pixels)	Model			
			Motorbikes	Faces	Airplanes	Spotted Cats
Motorbikes	800	200	92.5	50	51	56
Faces	435	300	33	96.8	32	32
Airplanes	800	300	64	63	90.2	53
Spotted Cats	200	80	48	44	51	90.0

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



Discussion Questions

1. What prior work does this paper build on (besides Kadir's) ?
2. What was significant about the training for object vs. background?
3. What are the main contributions of the Fergus work?
4. What are the drawbacks to this approach?