Object Recognition with Interest Operators

• Object recognition started with line segments.
  
  - Roberts recognized objects from line segments and junctions.
  
  - This led to systems that extracted linear features.
  
  - CAD-model-based vision works well for industrial.

• An “appearance-based approach” was first developed for face recognition and later generalized up to a point.

• The interest operators have led to a new kind of recognition by “parts” that can handle a variety of objects that were previously difficult or impossible.
Object Class Recognition by Unsupervised Scale-Invariant Learning

R. Fergus, P. Perona, and A. Zisserman
Oxford University and Caltech

CVPR 2003
won the best student paper award
CVPR 2013
won the best 10-year award
Goal:

• Enable Computers to Recognize Different Categories of Objects in Images.
Approach

• An object is a constellation of parts (from Burl, Weber and Perona, 1998).

• The parts are detected by an interest operator (Kadir’s).

• The parts can be recognized by appearance.

• Objects may vary greatly in scale.

• The constellation of parts for a given object is learned from training images
Components

• Model
  – Generative Probabilistic Model including Location, Scale, and Appearance of Parts

• Learning
  – Estimate Parameters Via EM Algorithm

• 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

121 dimensions was too big, so they used PCA to reduce to 10-15.
Learning a Model

• An object class is represented by a generative model with $P$ parts and a set of parameters $\theta$.

• Once the model has been learned, a decision procedure must determine if a new image contains an instance of the object class or not.

• Suppose the new image has $N$ interesting features with locations $X$, scales $S$ and appearances $A$. 
Probabilistic Model

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

- X is a description of the shape of the object (in terms of locations of parts)
- S is a description of the scale of the object
- A is a description of the appearance of the object
- \( \theta \) is the (maximum likelihood value of) the parameters of the object
- h is a hypothesis: a set of parts in the image that might be the parts of the object
- H is the set of all possible hypotheses for that object in that image.
- For N features in the image and P parts in the object, its size is \( O(N^P) \)
The appearance (A) of each part \( p \) has a Gaussian density with mean \( c_p \) and covariance \( V_p \).

Background model has mean \( c_{bg} \) and covariance \( V_{bg} \).
Object shape is represented by a joint Gaussian density of the locations (X) of features within a hypothesis transformed into a scale-invariant space.
The relative scale of each part is modeled by a Gaussian density with mean $t_p$ and covariance $U_p$. 

![Gaussian Relative Scale PDF](image)

Prob. of detection

| 0.8 | 0.75 | 0.9 |
Occlusion and Part Statistics

This was very complicated and turned out to not work well and not be necessary, in both Fergus’s work and other subsequent works.
Learning

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

\[ \theta = \{ \mu, \Sigma, c, V, M, p(d|\theta), t, U \} \]

\[ \hat{\theta}_{ML} = \arg \max_{\theta} p(X, S, A | \theta) \]
Recognition

Make this likelihood ratio:

\[
R = \frac{p(\text{Object}|X, S, A)}{p(\text{No object}|X, S, A)} = \frac{p(X, S, A|\text{Object}) p(\text{Object})}{p(X, S, A|\text{No object}) p(\text{No object})} \approx \frac{p(X, S, A|\theta) p(\text{Object})}{p(X, S, A|\theta_{bg}) p(\text{No object})}
\]

greater than a threshold.
RESULTS

• Initially tested on the Caltech-4 data set
  – motorbikes
  – faces
  – airplanes
  – cars

• Now there is a much bigger data set: the Caltech-101
  http://www.vision.caltech.edu/archive.html
Motorbikes

Equal error rate: 7.5%

Part 1 – Det: 5e-18

Part 2 – Det: 8e-22

Part 3 – Det: 8e-18

Part 4 – Det: 1e-19

Part 5 – Det: 3e-17

Part 6 – Det: 4e-24

Background – Det: 5e-19

Motorbike shape model
Background Images

It learns that these are NOT motorbikes.
Equal error rate: 4.6%

Frontal faces
Equal error rate: 9.8%
Scale-Invariant Cats

Equal error rate: 10.0%

Part 1 – Det:8e–22

Part 2 – Det:2e–22

Part 3 – Det:5e–22

Part 4 – Det:2e–22

Part 5 – Det:1e–22

Part 6 – Det:4e–21

Background – Det:2e–18
Scale-Invariant Cars

Equal error rate: 9.7%

Part 1 – Det 2e–19

Part 2 – Det 3e–18

Part 3 – Det 2e–20

Part 4 – Det 2e–22

Part 5 – Det 3e–18

Part 6 – Det 2e–18

Background – Det 4e–20
# Accuracy

## Initial Pre-Scaled Experiments

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Ours</th>
<th>Others</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motorbikes</td>
<td>92.5</td>
<td>84</td>
<td>[17]</td>
</tr>
<tr>
<td>Faces</td>
<td>96.4</td>
<td>94</td>
<td>[19]</td>
</tr>
<tr>
<td>Airplanes</td>
<td>90.2</td>
<td>68</td>
<td>[17]</td>
</tr>
<tr>
<td>Cars(Side)</td>
<td>88.5</td>
<td>79</td>
<td>[1]</td>
</tr>
</tbody>
</table>