Image Fragments in Object Classification: Ullman Et Al, 2002

Mike Onorato
COS 598B
Overview

- Intermediate-complexity features
  - Image “fragments”
    - Used in object classification
Part 1: Image Fragments
Neuroscience Background

- V1: Simple lines, edges or small regions  

Neuroscience Background


Neuroscience Background


Neuroscience Background

- **TE:** Shapes similar to a lip or eyebrow  

Neuroscience Background

- Anterior IT: Complete or partial object views [Logothetis, et al: View-dependent object recognition in monkeys (1994)]

Neuroscience Background

- Preferred Stimuli: Specific 2D patterns

Neuroscience Background

- Preferred Stimuli: Dependent on training stimuli & Independent of position and orientation

Ullman’s Model

- Preferred Stimuli:
  - Specific 2D patterns
  - Dependent on training stimuli
  - Position and orientation independent

Computer Science Background

- Class-independent small features:
  - Wavelets & Gabor functions

Image Source: [David Bradley, Object Recognition with Informative Features and Linear Classification (2000)]
## Image Fragments

- Overlapping patches of images
- Varying sizes, locations and resolutions
Image Fragment Extraction

- Extract many hundreds of features from each image
- Never explain how or exactly how many
Image Fragment Selection

- Step 1: Remove fragments which only appear once
Aside - Ordinal Measures

\[ R \]
\[
\begin{array}{ccc}
10 & 30 & 70 \\
20 & 50 & 80 \\
40 & 60 & 100 \\
\end{array}
\]
\[ S \]
\[
\begin{array}{ccc}
10 & 30 & 70 \\
28 & 50 & 80 \\
40 & 60 & 120 \\
\end{array}
\]

\[ d_m^i = i - \sum_{j=1}^{i} J(s^j \leq i) \]

\[ s^i = \pi_k^2, \quad k = \left(\pi_1^{-1}\right)^i \]

Source: [Bhat, D., Nayar, K.: Ordinal Measures for Image Correspondence (1998)]
Image Fragment Comparison

- Difference between fragments F and H:
  \[ D(F, H) = k_1 \sum_{i} d_i + k_2 |\alpha_F - \alpha_H| + k_3 |G_F - G_H| \]
  - \(\alpha_X\): orientation
  - \(G_X\): gradient

- Fragments F and H are the same if:
  \[ D(F, H) < \text{Threshold} \]
Image Fragment Selection

- Step 1: Remove fragments which only appear once
Image Fragment Selection

- Step 2: Select the 8 most informative fragments

<table>
<thead>
<tr>
<th></th>
<th>1st</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merit</td>
<td>0.20 0.18 0.18 0.17 0.16 0.11 0.10 0.09</td>
</tr>
<tr>
<td>Weight</td>
<td>6.5 5.5 6.45 5.45 3.52 2.9 2.9 2.86</td>
</tr>
</tbody>
</table>
Aside – Information Theory

- **Entropy:**
  - Amount of information transmitted.

  \[
  H(X) = - \sum_{i=1}^{K} P(x_i) \log(P(x_i))
  \]
  Where X is r.v.
Aside – Information Theory

- Mutual Information:
  - The amount of information about $X$ given by $Y$.

- $I(X,Y) = H(X) - H(X|Y)$
  - Where $X$ and $Y$ are r.v.’s
Step 2: Select the 8 fragments with highest $I(C,F)$

$I(C,F) = H(C) - H(C|F)$

- $C = \text{object is in the class}$
- $F = \text{fragment is in the image}$

<table>
<thead>
<tr>
<th>1st</th>
<th>Merit</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.20</td>
<td>6.5</td>
</tr>
<tr>
<td></td>
<td>0.18</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>0.18</td>
<td>6.45</td>
</tr>
<tr>
<td></td>
<td>0.17</td>
<td>5.45</td>
</tr>
<tr>
<td></td>
<td>0.16</td>
<td>3.52</td>
</tr>
<tr>
<td></td>
<td>0.11</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td>2.86</td>
</tr>
</tbody>
</table>
Image Fragment Selection

- Step 3: Select more fragments of the same 8 types

| 1st | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|
| Merit | 0.20 | 0.18 | 0.18 | 0.17 | 0.16 | 0.11 | 0.10 | 0.09 |
| Weight | 6.5 | 5.5 | 6.45 | 5.45 | 3.52 | 2.9 | 2.9 | 2.86 |

| 2nd | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|

| 3rd | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|

| 4th | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|
Fragment Selection - Results

- Dataset of 138 faces and 40 cars
- Resultant fragments had intermediate size:
  - Median: 11% object size
  - SD: 16% object size
- All had intermediate size or resolution
Fragment Selection - Results

- Information peaks at intermediate size:
Fragment Selection - Results

- Mutual information peaks at intermediate resolution.
Fragment Selection - Analysis

- These fragments provide best compromise between:
  - Specificity
  - Relative frequency
Part 2: Classification Algorithm
Classification Algorithm

- Extracted fragments from a training set
- Classify objects in new images
Classification Algorithm

- Step 1: Detect fragments
  - Extract candidate fragments H from the image:
    - Size: 0.5-2 times area of F
    - Location: Steps of 3 pixels
    - Resolution: 1x to 1/10x in steps of 1/20x

Example H's
Classification Algorithm

- Step 2: Local search around detected fragments
  - Slight adjustments in size, location and resolution
Classification Algorithm

- Likelihood ratio of the image belonging to class C:
  \[ R(F) = \frac{P(F \mid C)}{P(F \mid \bar{C})} \]

- F = fragment detected in image
Classification Algorithm

- Step 3: Sum likelihood ratios

\[ \sum_k w_{ik} \max(F_{ik}) > \theta \]

- \( F_{ik} = i\text{-th fragment of } k\text{-th type} \)
- \( w = \log_2(R(F)) \)
- \( \theta = \text{threshold} \)
Classification Algorithm

- To detect faces of varying sizes, test images are rescaled at multiple levels
Classification Performance

- 200 face images & 200 non-face images
- Results:
  - 97% detection
  - 2.1% false detection.
- Comparable to best preexisting systems
Classification Performance

- “Optimal size” fragments:
  - 95.6% face detection
  - 0% false alarms

- Smaller fragments (4% of average face area):
  - 97% face detection
  - 30.4% false alarms

- Larger fragments (33% of average face area):
  - 39% face detection
  - 0% false alarms
Arrangement Specificity
Part 3: Other Things
Other Things

- Matching:

<table>
<thead>
<tr>
<th>Fragments</th>
<th>Novel</th>
<th>Full face</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
</tbody>
</table>

- *No numerical data about the 8 observers’ judgments is provided
Other Things

- Fragments used in back prop neural net.
- Improved classification performance of net.
- *No numerical data is given
Conclusions

- IC fragments are most informative fragments.
- Fragments are good at classification.
- Similar to human visual pathway.
Why Are Fragments Good?

- Similar to cortex
  - Features learned from experience
  - Intermediate complexity
  - Independent of position and some rotation
- Perform global search on large set of potential features.
  - Back-propagation models start from randomly selected features and perform local search.
Why Are Fragments Bad?

- Cannot generalize to large changes in rotation:
  - No 3D information
- Rectangular
Additional Critiques

- As a computer science paper:
  - Qualitative comparison with other methods
  - Test on more difficult object classes

- As a neuroscience paper:
  - Neurons that respond to the extracted fragments?
  - No additional work