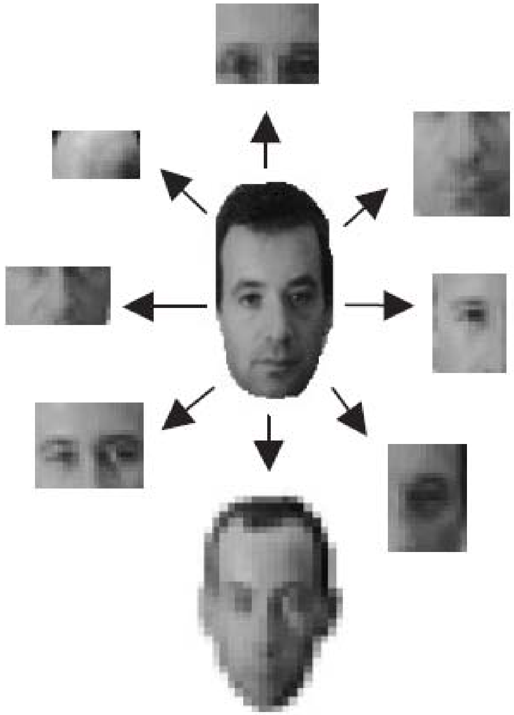
Image Fragments in Object Classification:
Ullman Et Al, 2002

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 *[Hubel, D. H., Wiesel, T. N.: Receptive fields and functional architecture of monkey striate cortex (1968)]*

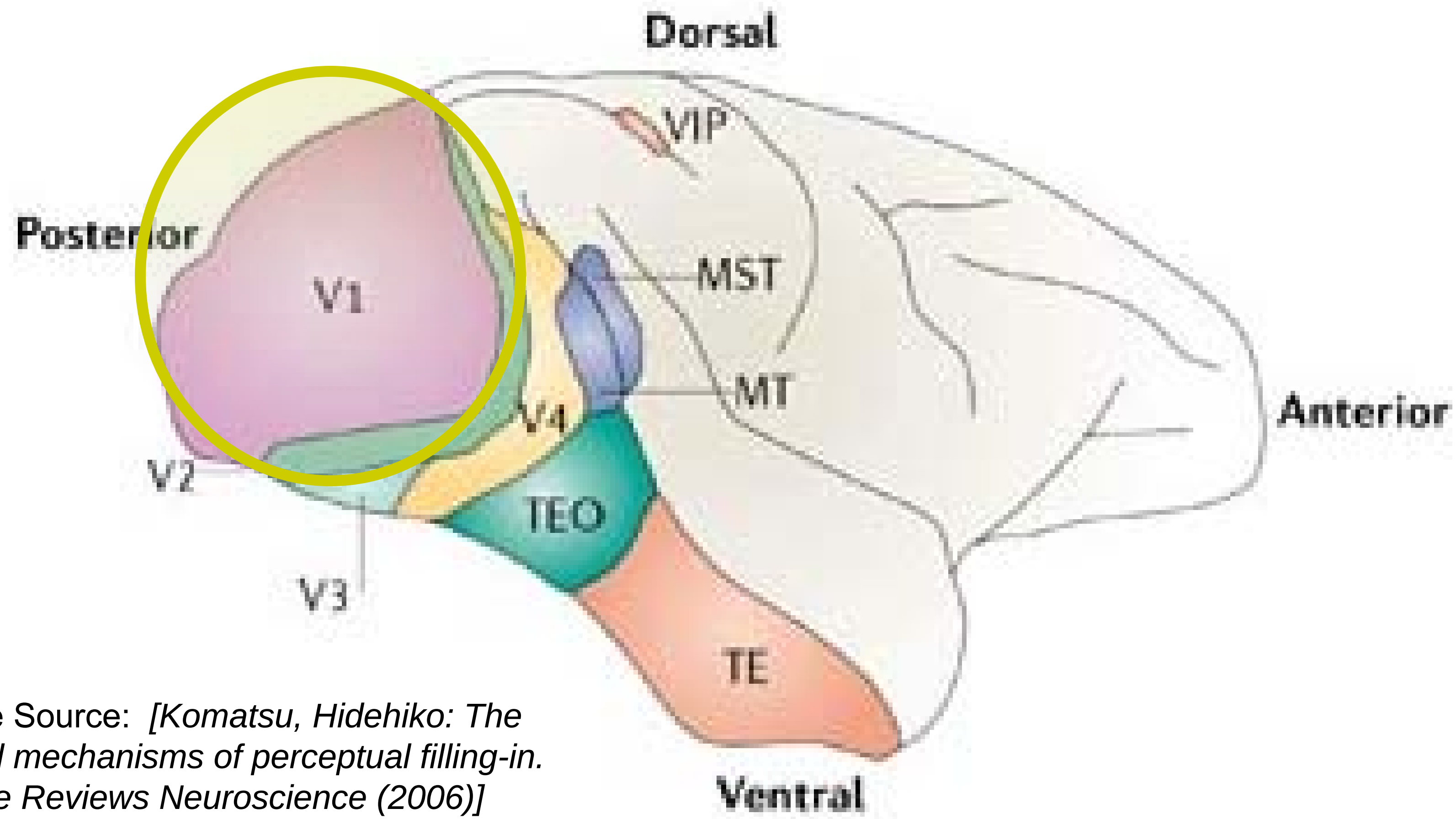


Image Source: *[Komatsu, Hidehiko: The neural mechanisms of perceptual filling-in. Nature Reviews Neuroscience (2006)]*

Neuroscience Background

- V2: Collinear arrangements of features *[Wiskott, L., et al: Face Recognition by Elastic Bunch Graph Matching (1999)]*

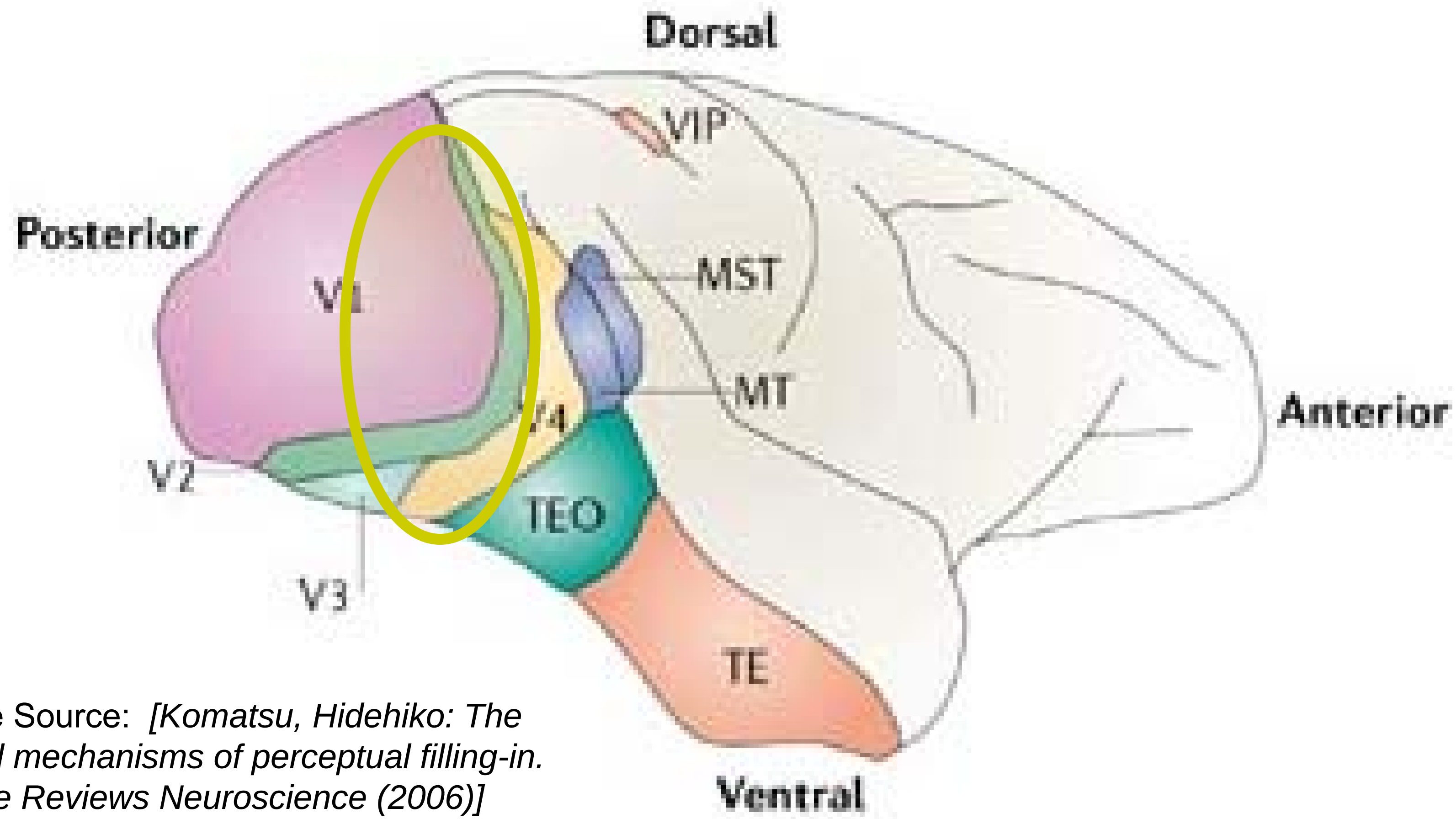


Image Source: *[Komatsu, Hidehiko: The neural mechanisms of perceptual filling-in. Nature Reviews Neuroscience (2006)]*

Neuroscience Background

- V4: Spiral and polar shapes *[Gallant, J.L., et al: Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex (1993)]*

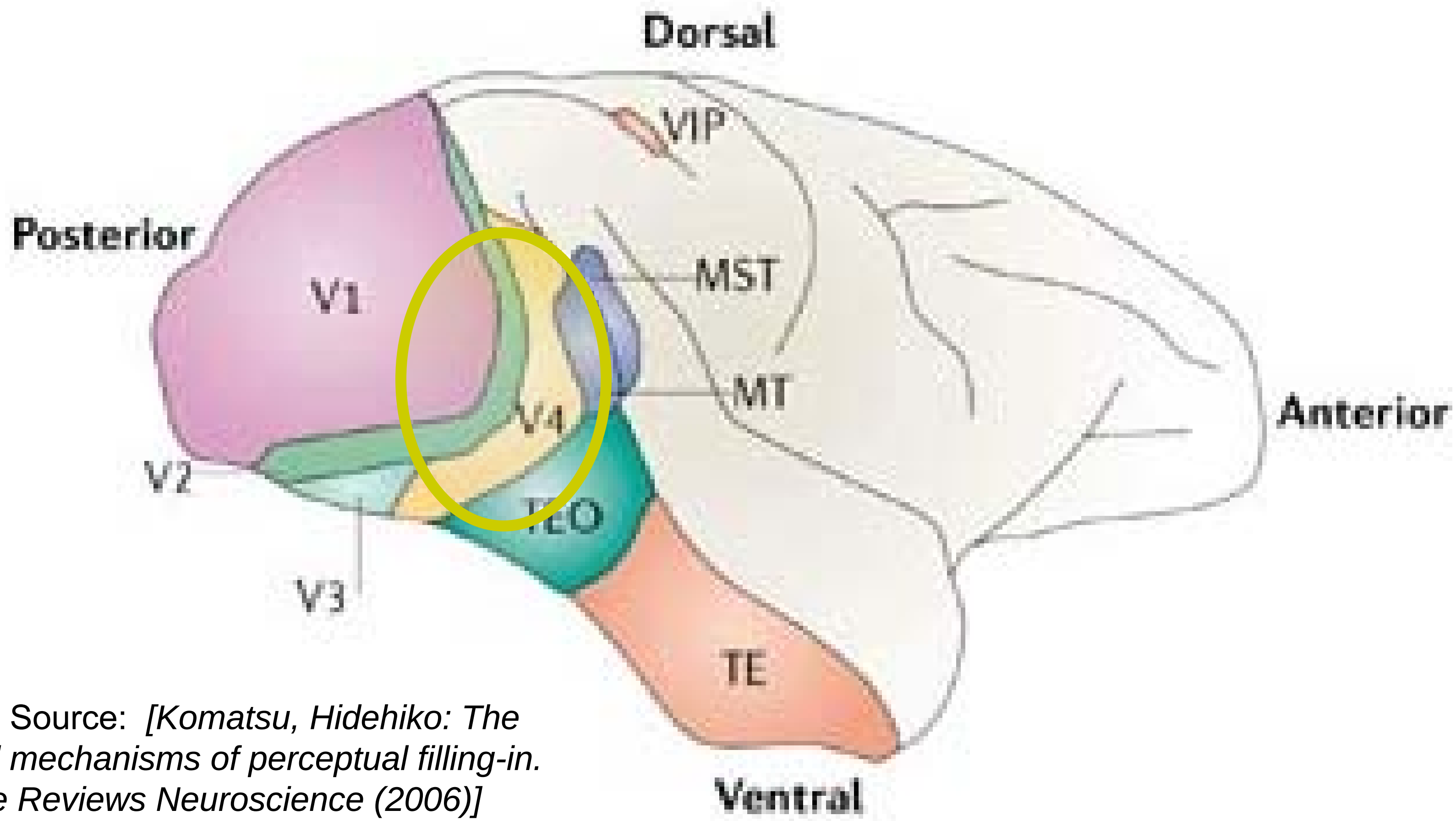


Image Source: *[Komatsu, Hidehiko: The neural mechanisms of perceptual filling-in. Nature Reviews Neuroscience (2006)]*

Neuroscience Background

- TE: Shapes similar to a lip or eyebrow *[Tanaka, K.: Neural Mechanisms of Object Recognition. Science, Vol. 262 (1993)]*

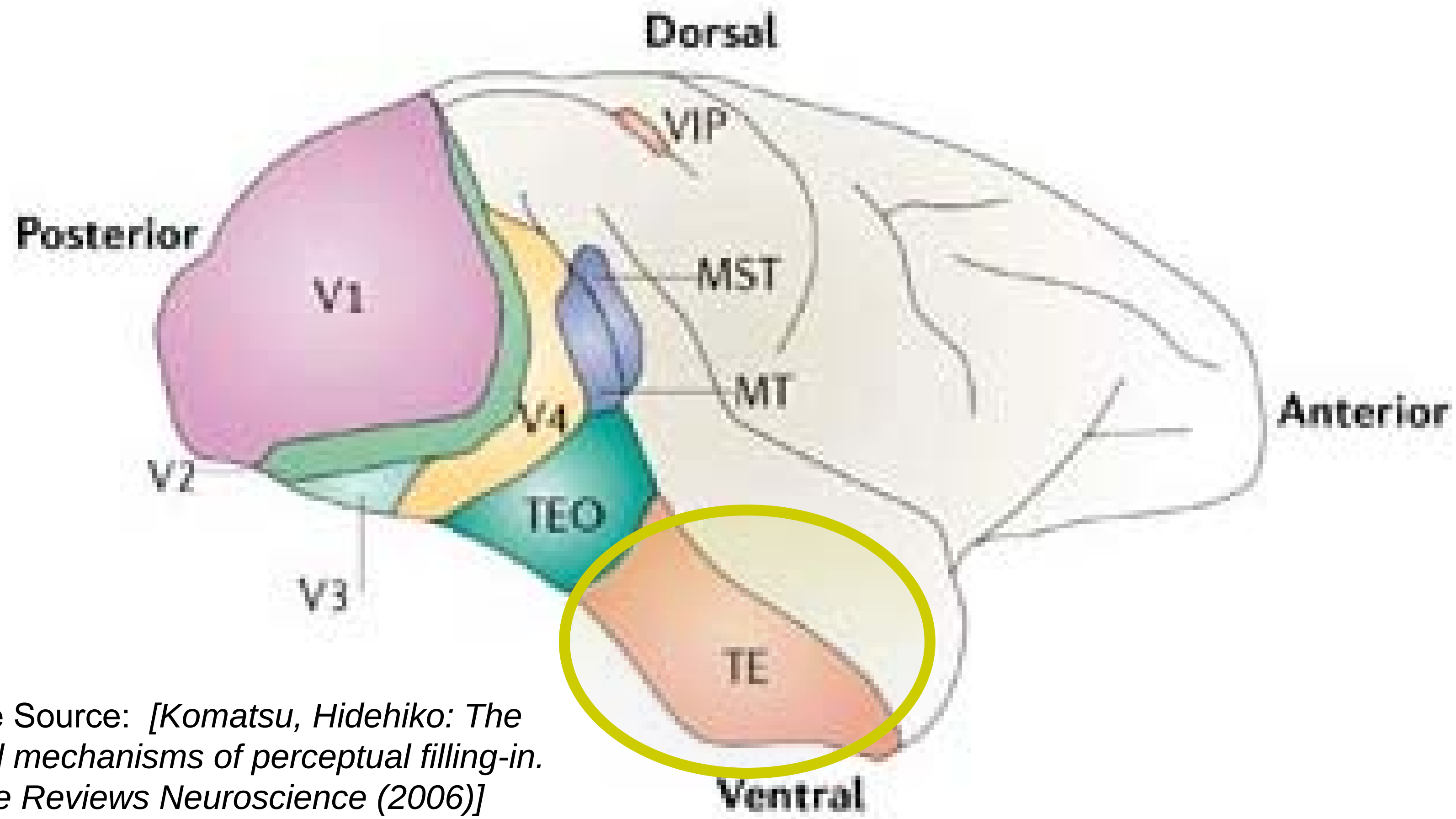


Image Source: *[Komatsu, Hidehiko: The neural mechanisms of perceptual filling-in. Nature Reviews Neuroscience (2006)]*

Neuroscience Background

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

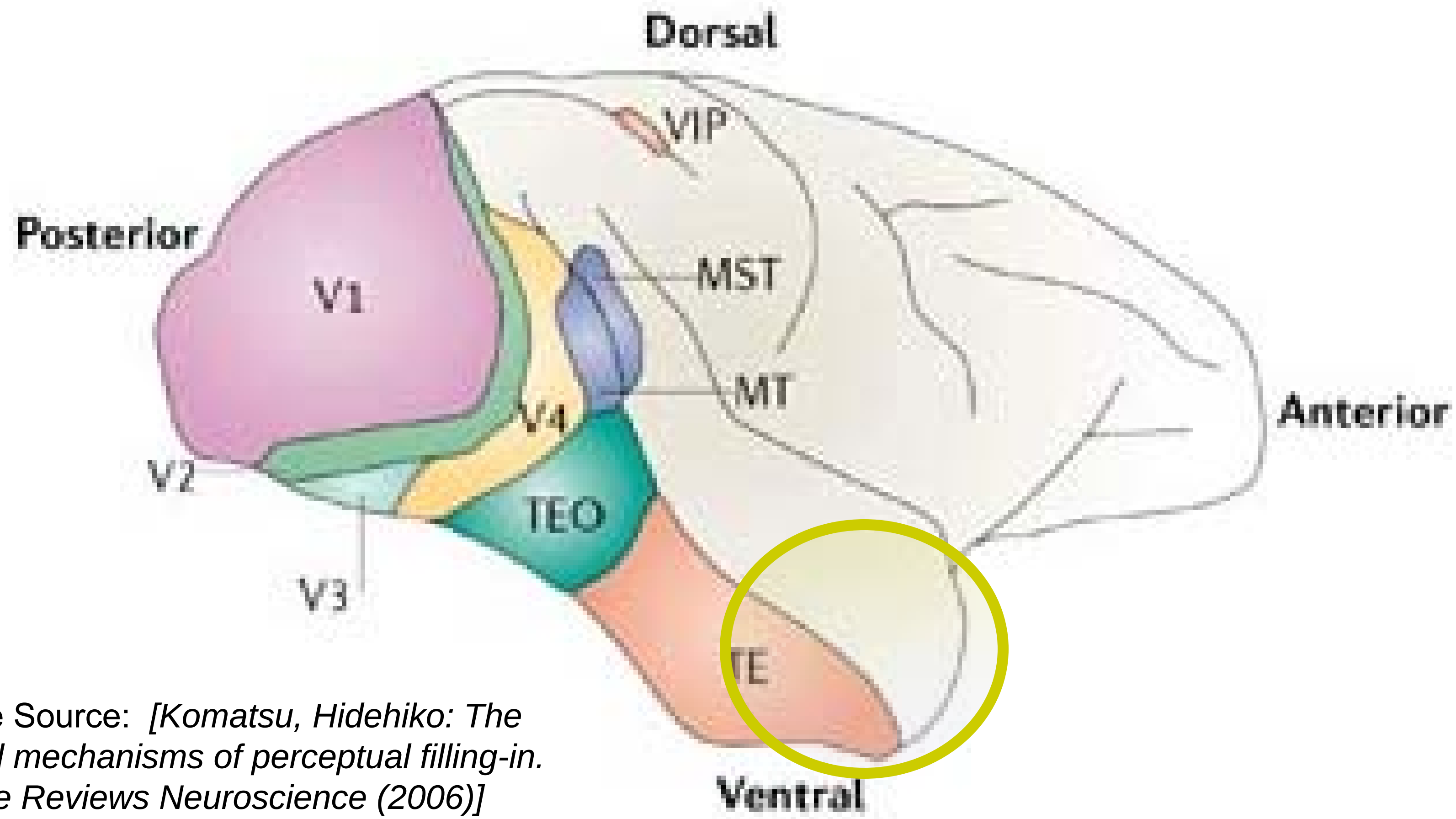


Image Source: *[Komatsu, Hidehiko: The neural mechanisms of perceptual filling-in. Nature Reviews Neuroscience (2006)]*

Neuroscience Background

- Preferred Stimuli: Specific 2D patterns

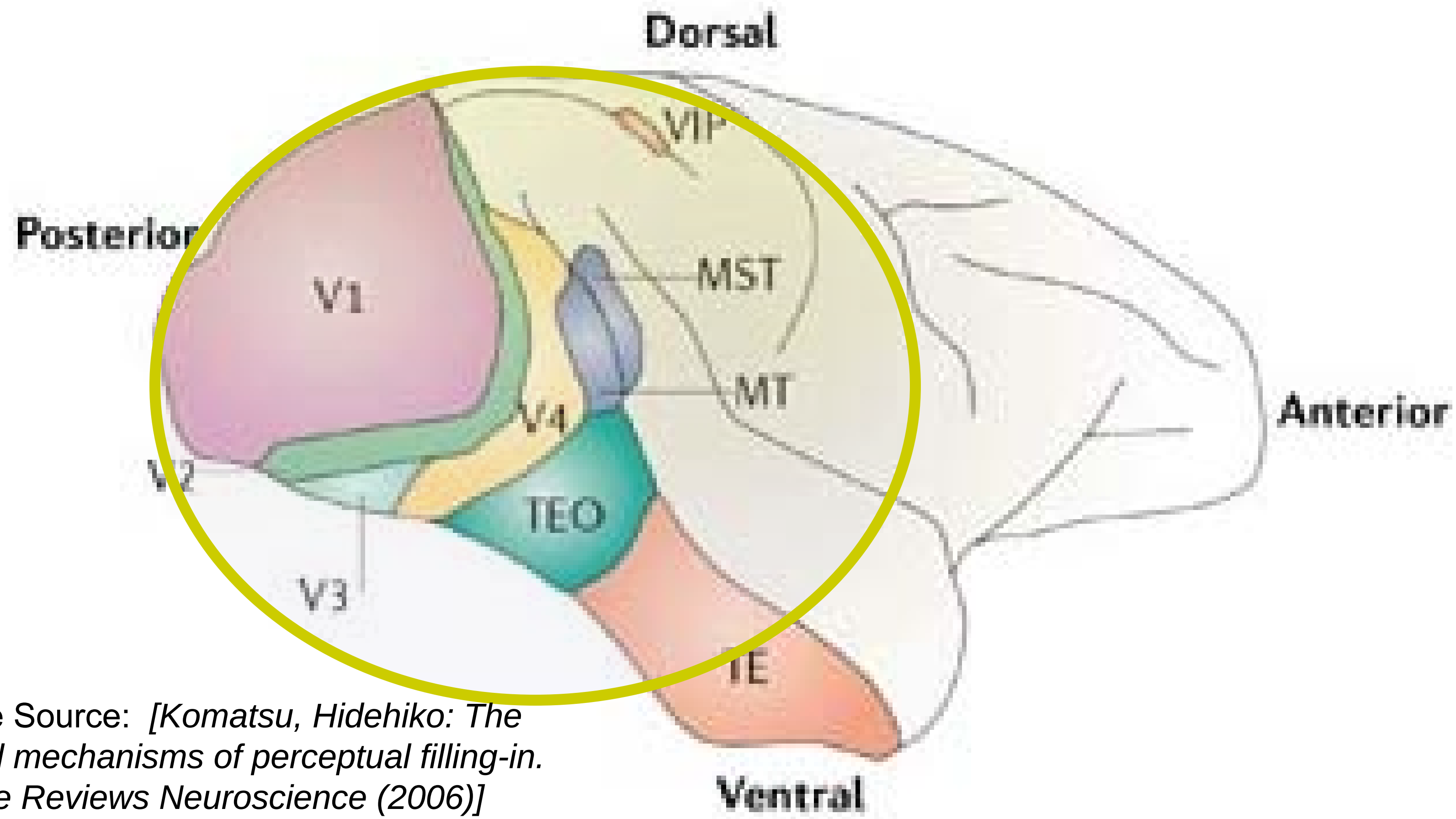


Image Source: *[Komatsu, Hidehiko: The neural mechanisms of perceptual filling-in. Nature Reviews Neuroscience (2006)]*

Neuroscience Background

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

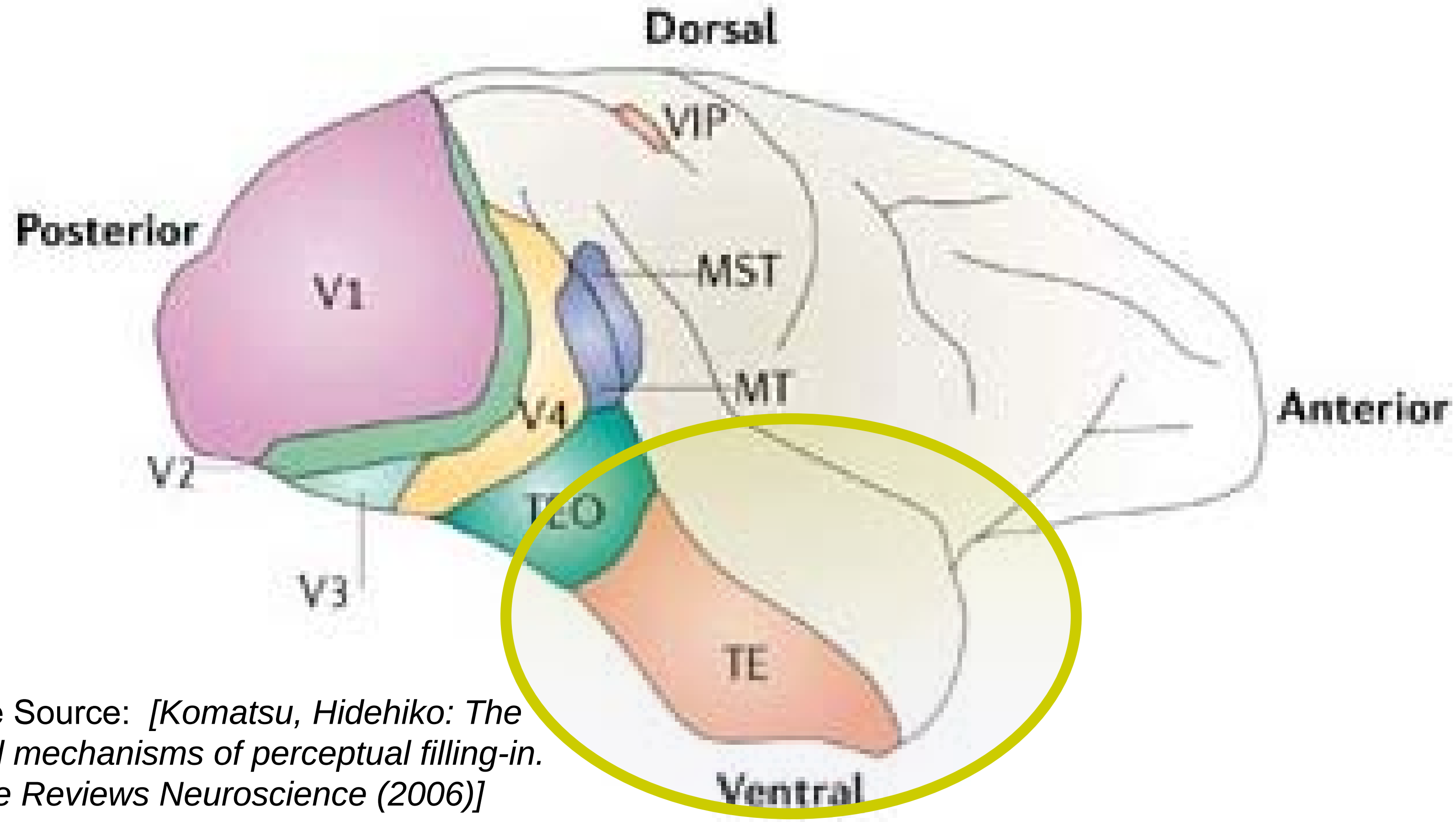


Image Source: *[Komatsu, Hidehiko: The neural mechanisms of perceptual filling-in. Nature Reviews Neuroscience (2006)]*

Ullman's Model

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

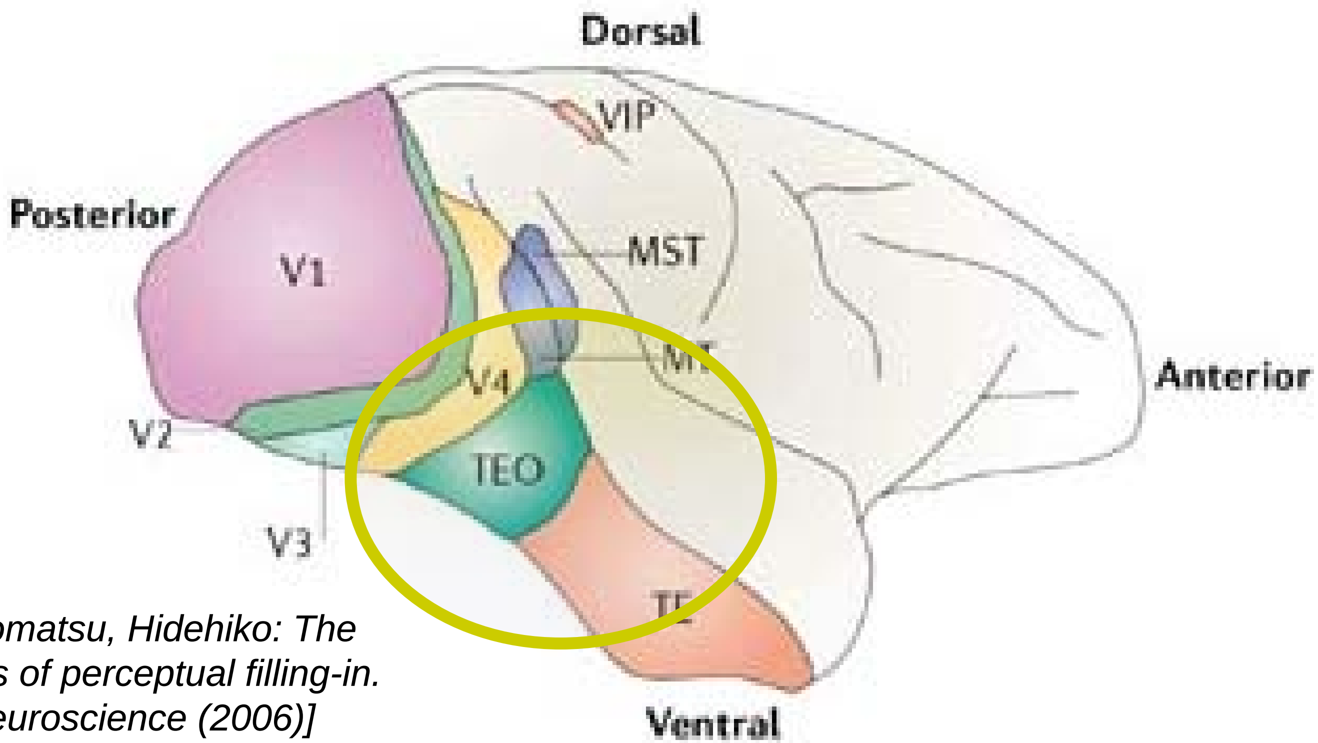


Image Source: *[Komatsu, Hidehiko: The neural mechanisms of perceptual filling-in. Nature Reviews Neuroscience (2006)]*

Computer Science Background

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

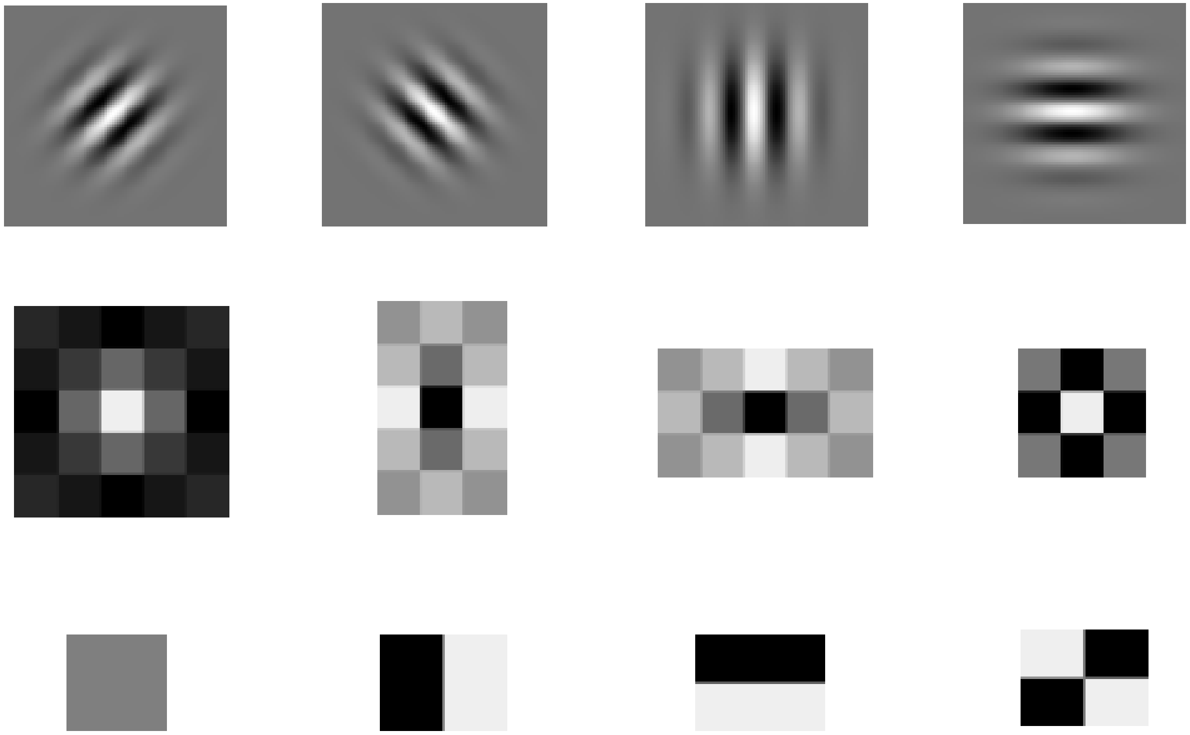


Image Source: *[David Bradley, Object Recognition with Informative Features and Linear Classification (2000)]*

Image Fragments

- Overlapping patches of images
- Varying sizes, locations and resolutions

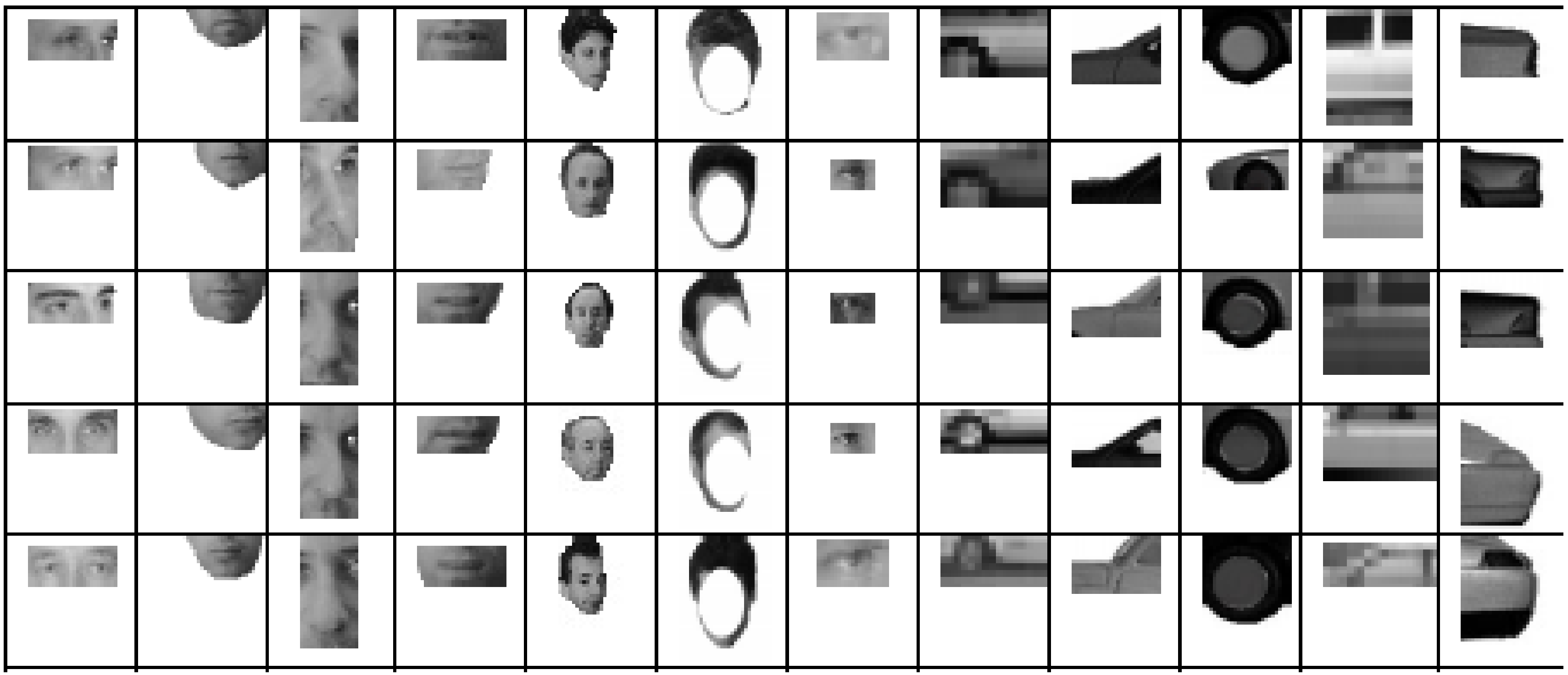


Image Fragment Extraction

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

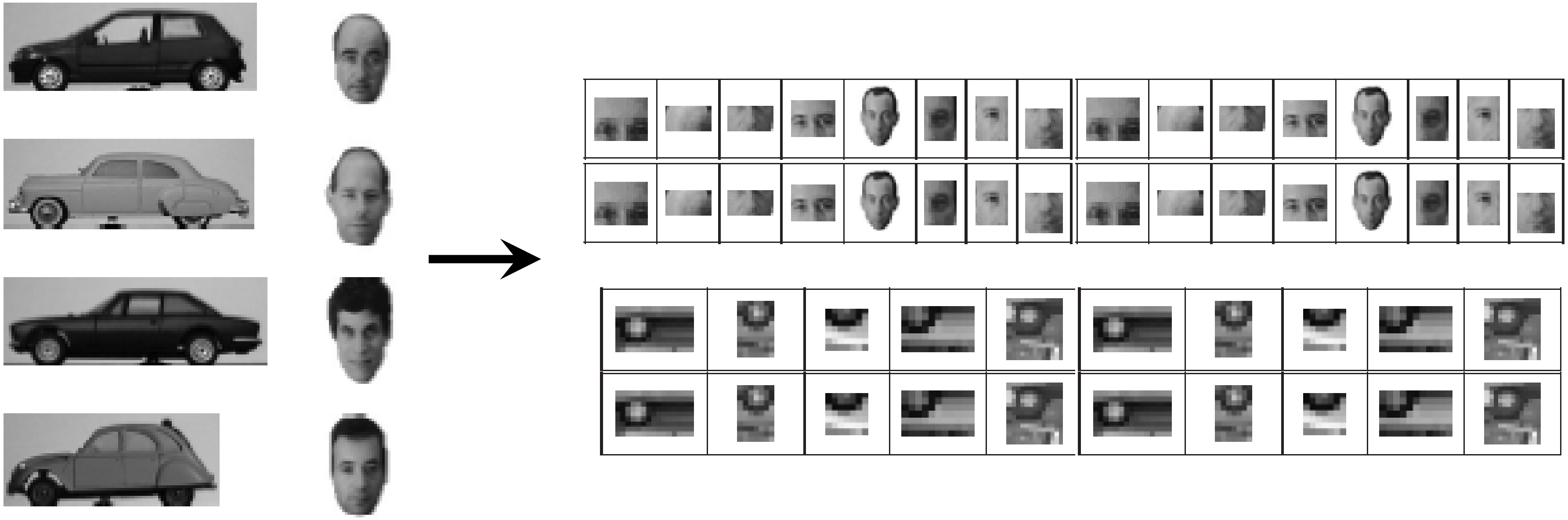
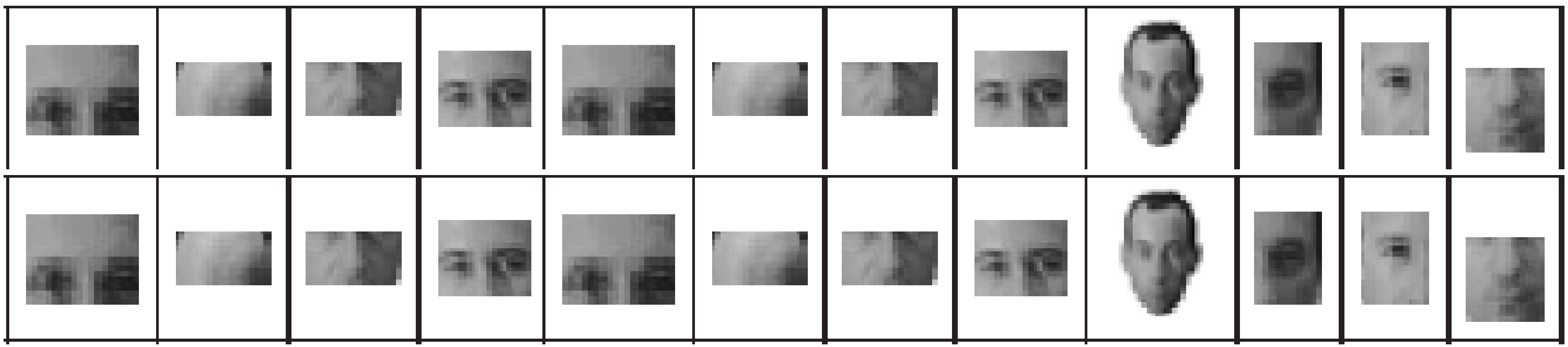
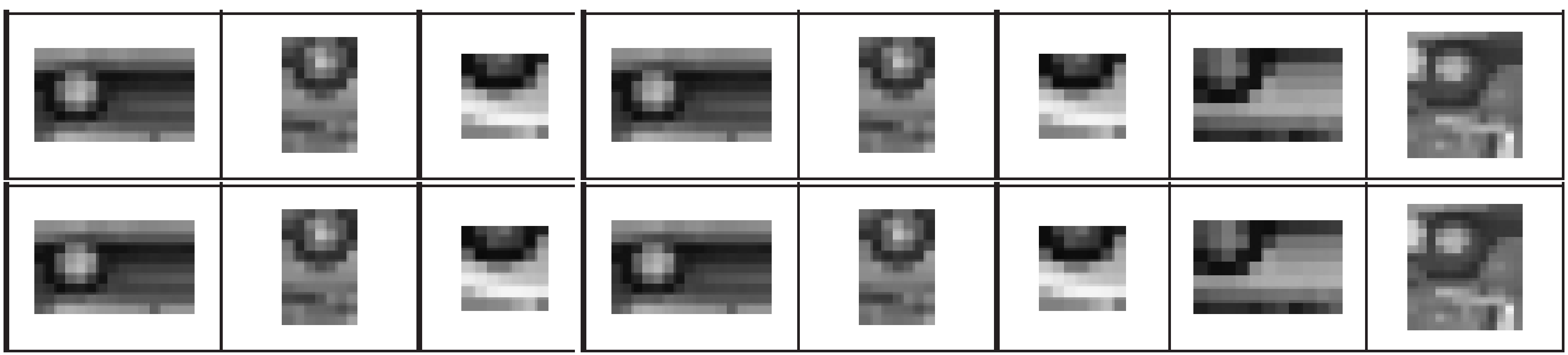


Image Fragment Selection

- Step 1: Remove fragments which only appear once





Aside - Ordinal Measures

R			S		
10	30	70	10	30	70
20	50	80	28	50	80
40	60	100	40	60	120

$$d_m^i = i - \sum_{j=1}^{i} J(s^j \le i)$$

$$s^i = \pi_2^k, \; k = \left(\pi_1^{-1}\right)^i$$

R			S		
1	3	7	1	3	7
2	5	8	2	5	8
4	6	9	4	6	9

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

Image Fragment Comparison

- Difference between fragments F and H:

$$D(F,H) = k_1 \sum_i |d_i| + k_2 |\alpha_F - \alpha_H| + k_3 |G_F - G_H|$$

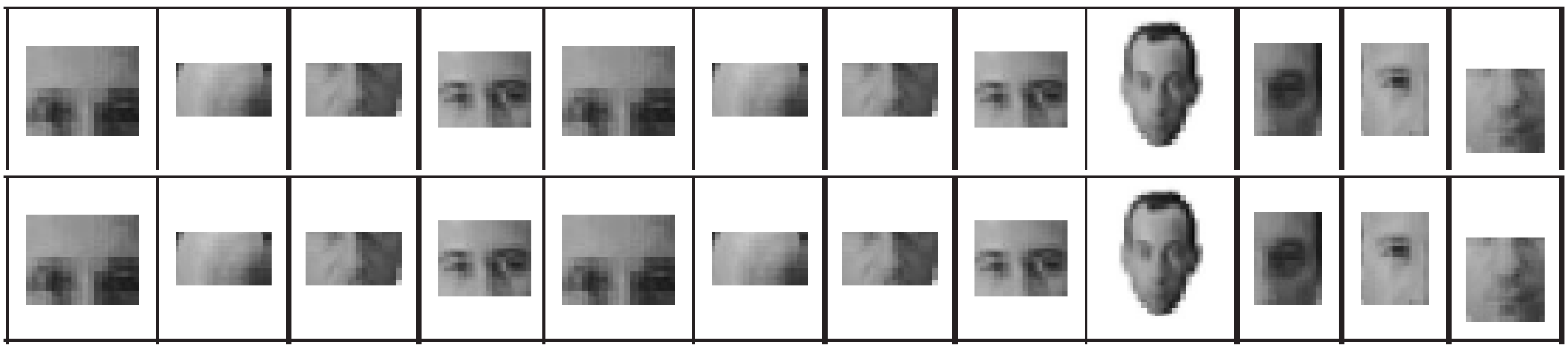
  - α_X: orientation
  - G_X: gradient

- Fragments F and H are the same if:

  D(F,H) < Threshold

Image Fragment Selection

- Step 1: Remove fragments which only appear once



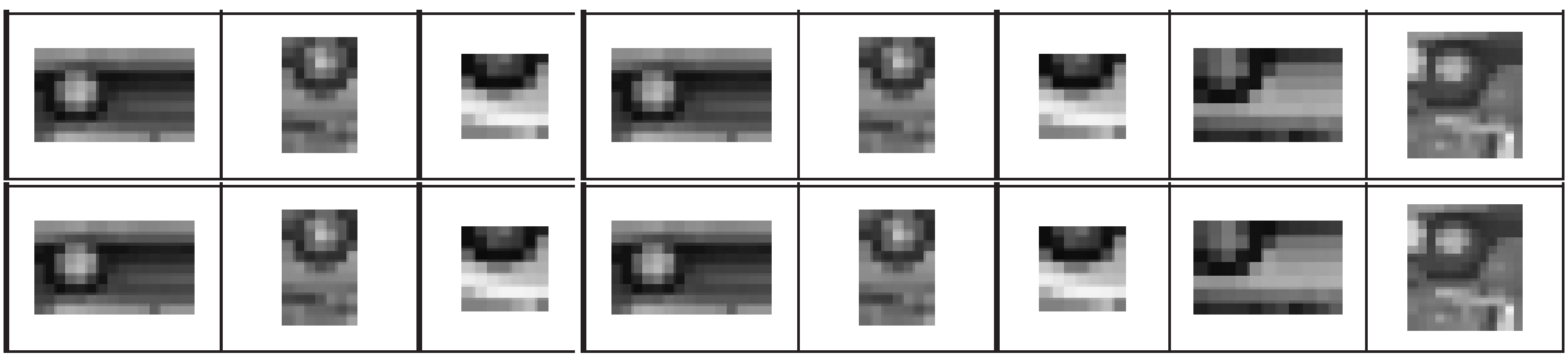
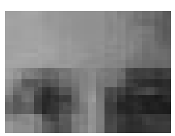
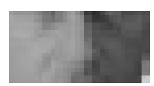
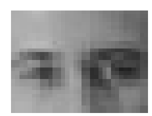
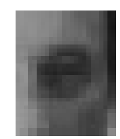
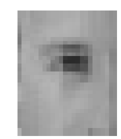
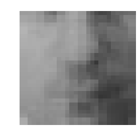


Image Fragment Selection

- Step 2: Select the 8 most informative fragments

1^{st}								
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

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

Image Fragment Selection

- Step 2: Select the 8 fragments with highest I(C,F)
  - I(C,F) = H(C) - H(C|F)
    - C = object is in the class
    - F = fragment is in the image

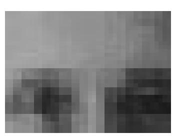
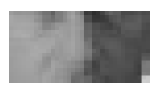
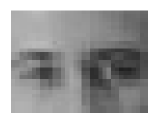
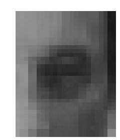
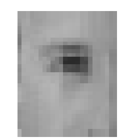
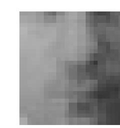
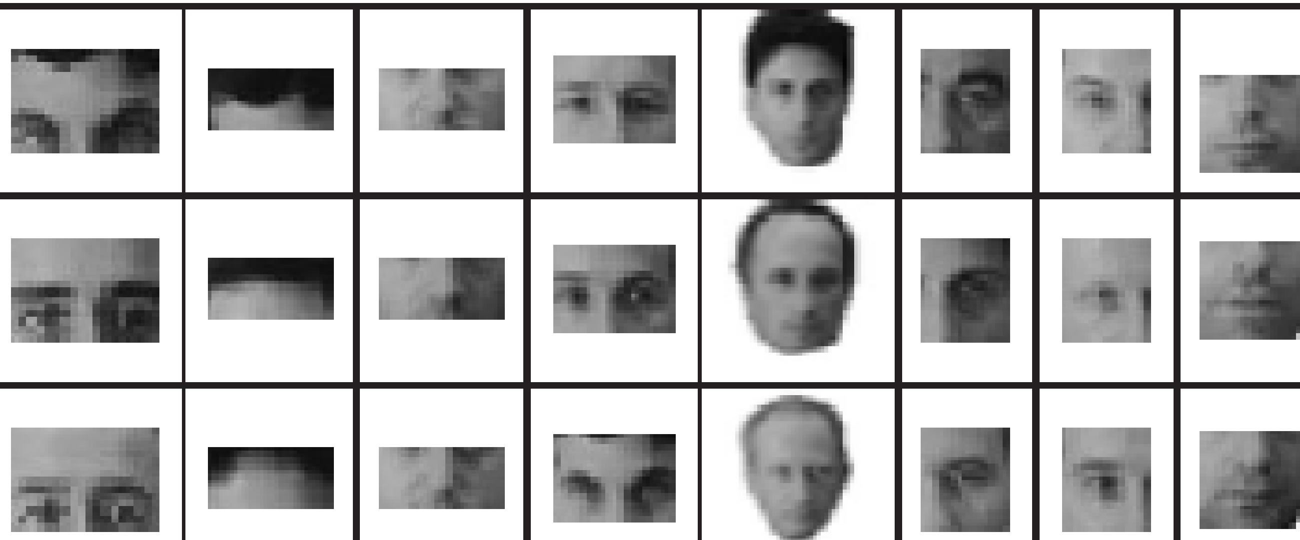
1st								
Merit	0.20	0.18	0.18	0.17	0.16	0.11	0.10	0.09
Weight	6.5	5.5	6.45	5.45	3.52	2.9	2.9	2.86

Image Fragment Selection

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

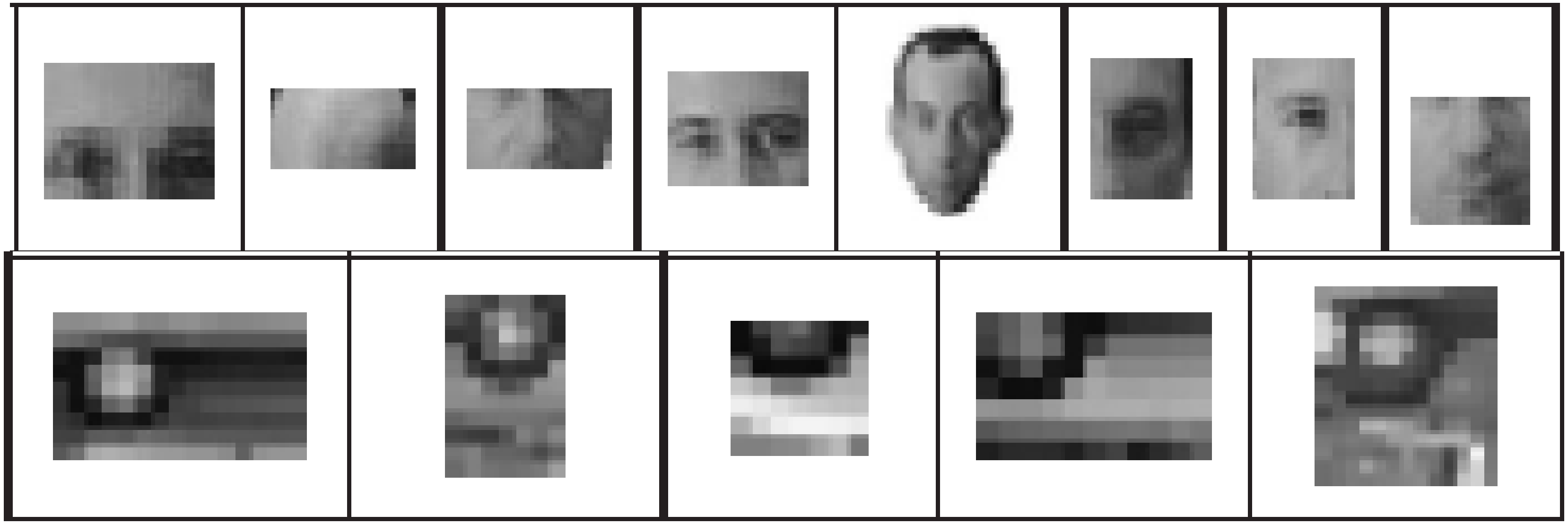
1st								
Merit	0.20	0.18	0.18	0.17	0.16	0.11	0.10	0.09
Weight	6.5	5.5	6.45	5.45	3.52	2.9	2.9	2.86



2nd								
3rd								
4th								

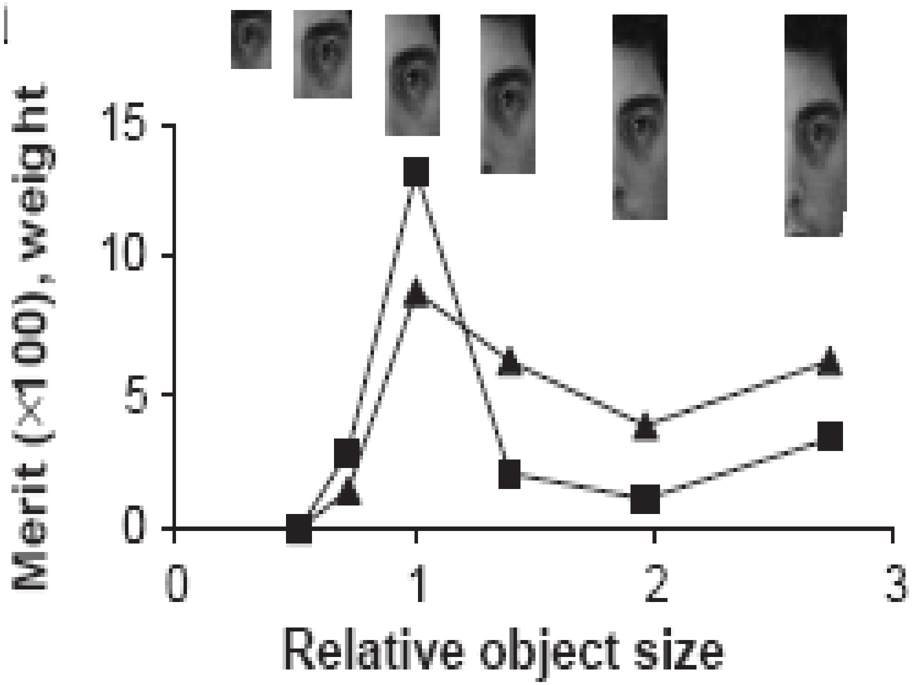
Fragment Selection - Results

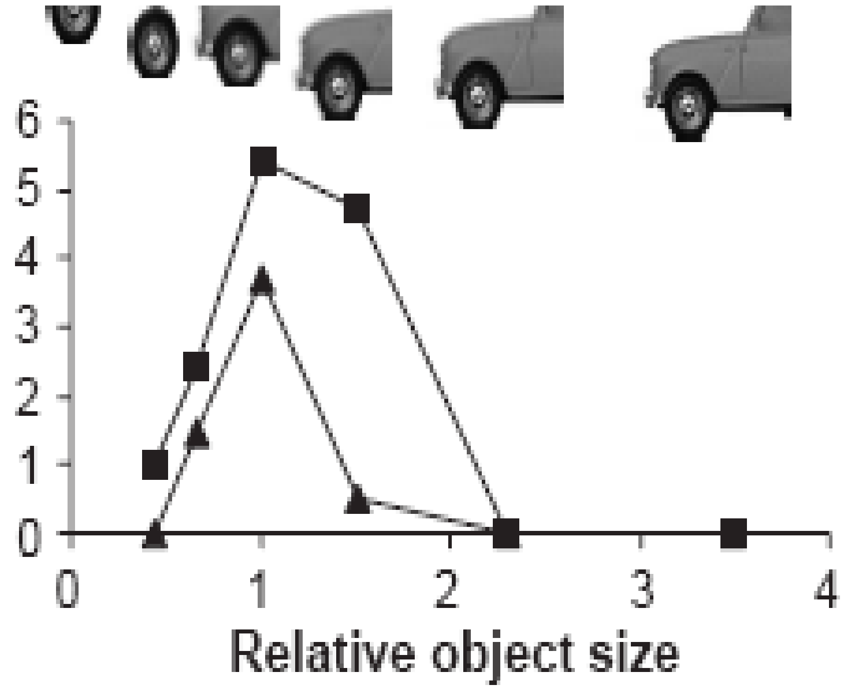
- Dataset of 138 faces and 40 cars
- Resultant fragments had intermediate size:
  - Median: 11% object size
  - SD: 16% object size
- All had intermediate size or resolution



Fragment Selection- Results

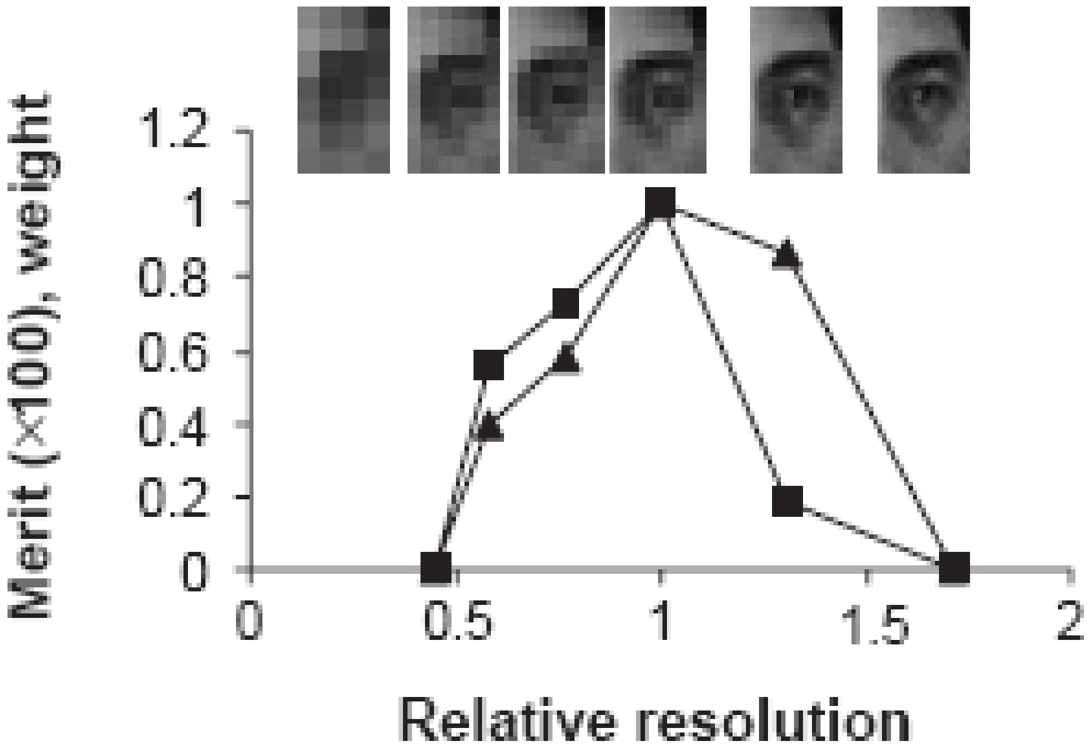
- Information peaks at intermediate size:





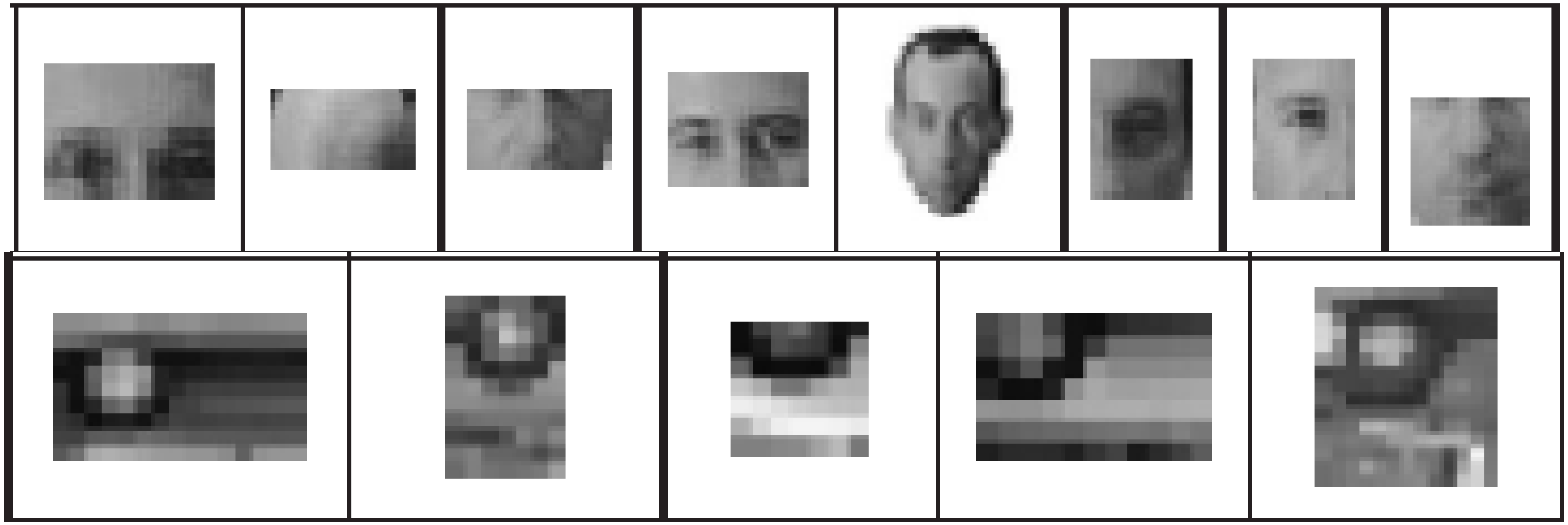
Fragment Selection- Results

- Mutual information peaks at intermediate resolution.



Fragment Selection - Analysis

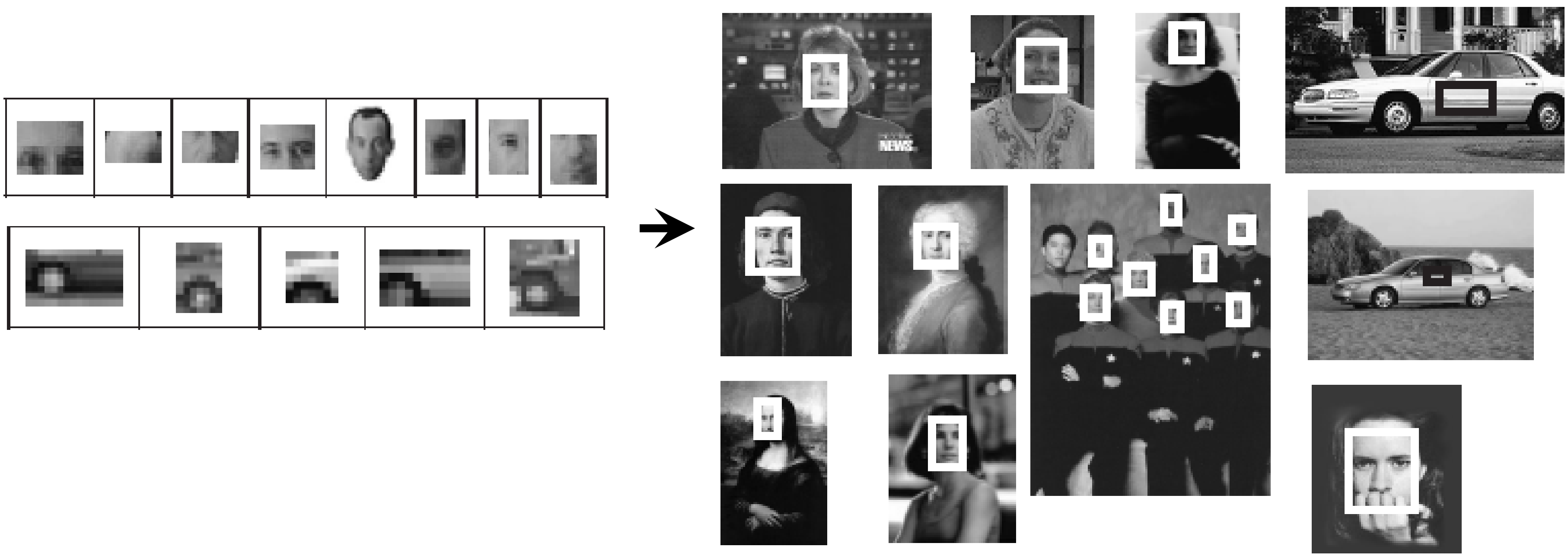
- These fragments provide best compromise between:
  - Specificity
  - Relative frequency



Part 2: Classification Algorithm

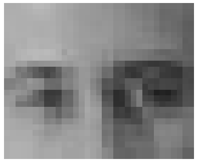
Classification Algorithm

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

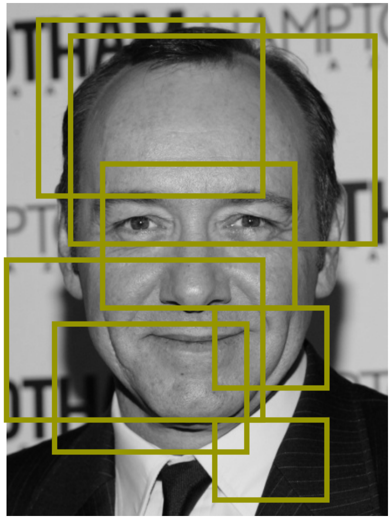


Classification Algorithm

- 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



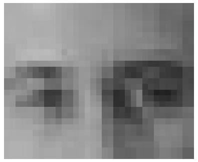
F



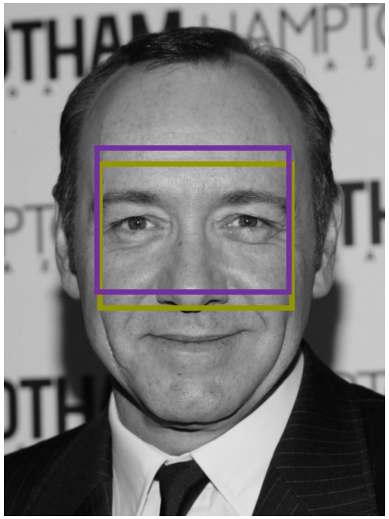
Example H’s

Classification Algorithm

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



F



H

Classification Algorithm

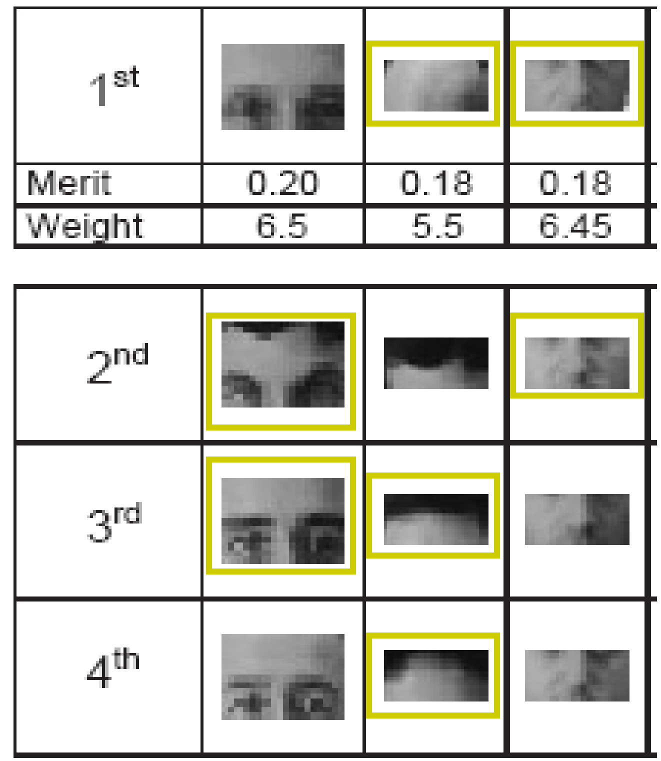
- Likelihood ratio of the image belonging to class C:
  - $R(F) = \dfrac{P(F \mid C)}{P(F \mid \overline{C})}$
  - F = fragment detected in image

Classification Algorithm

- Step 3: Sum likelihood ratios

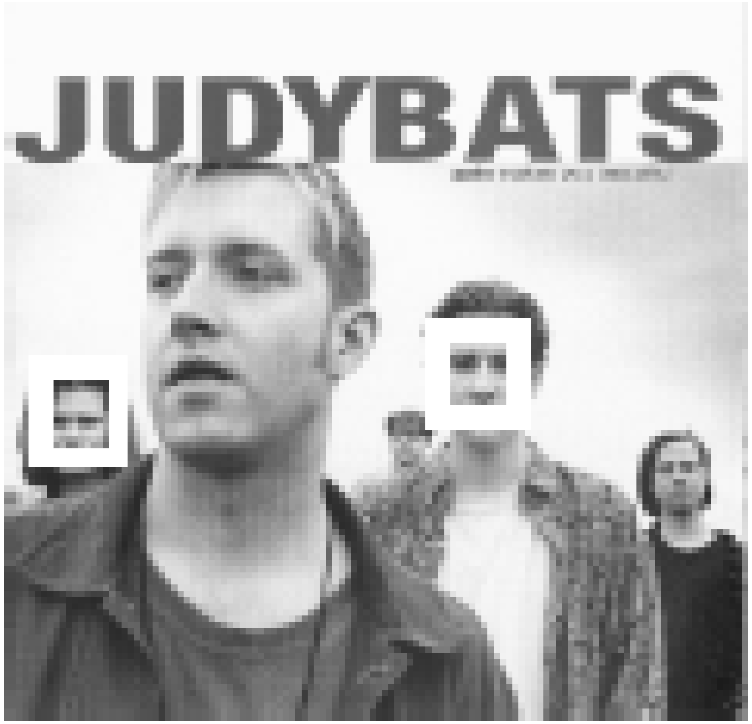
$$\Sigma_k w_{ik} \max(F_{ik}) > \theta$$

  - F_{ik} = *i*-th fragment of *k*-th type
  - $w = \log_2(R(F))$
  - θ = threshold



Classification Algorithm

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



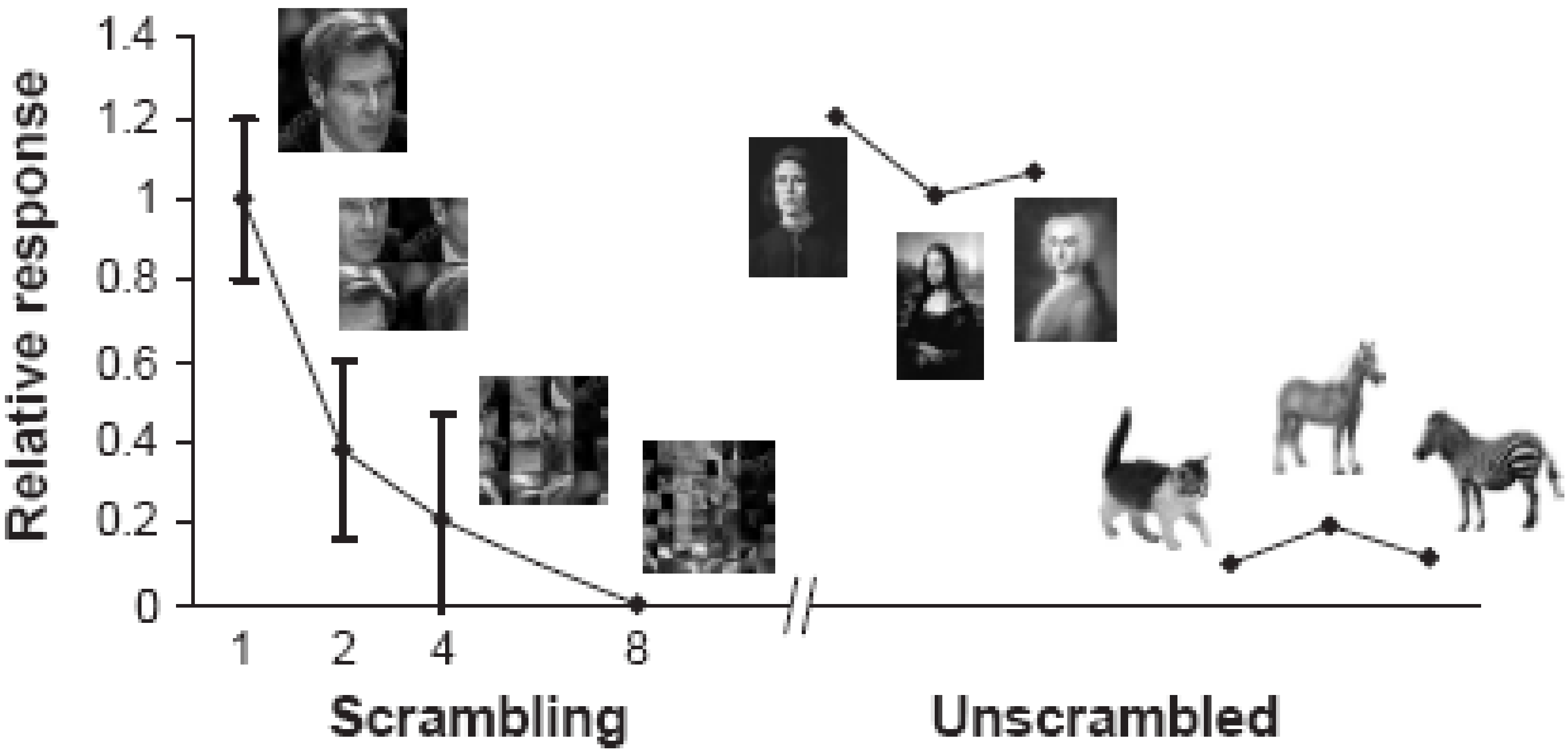
Classification Performance

- 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

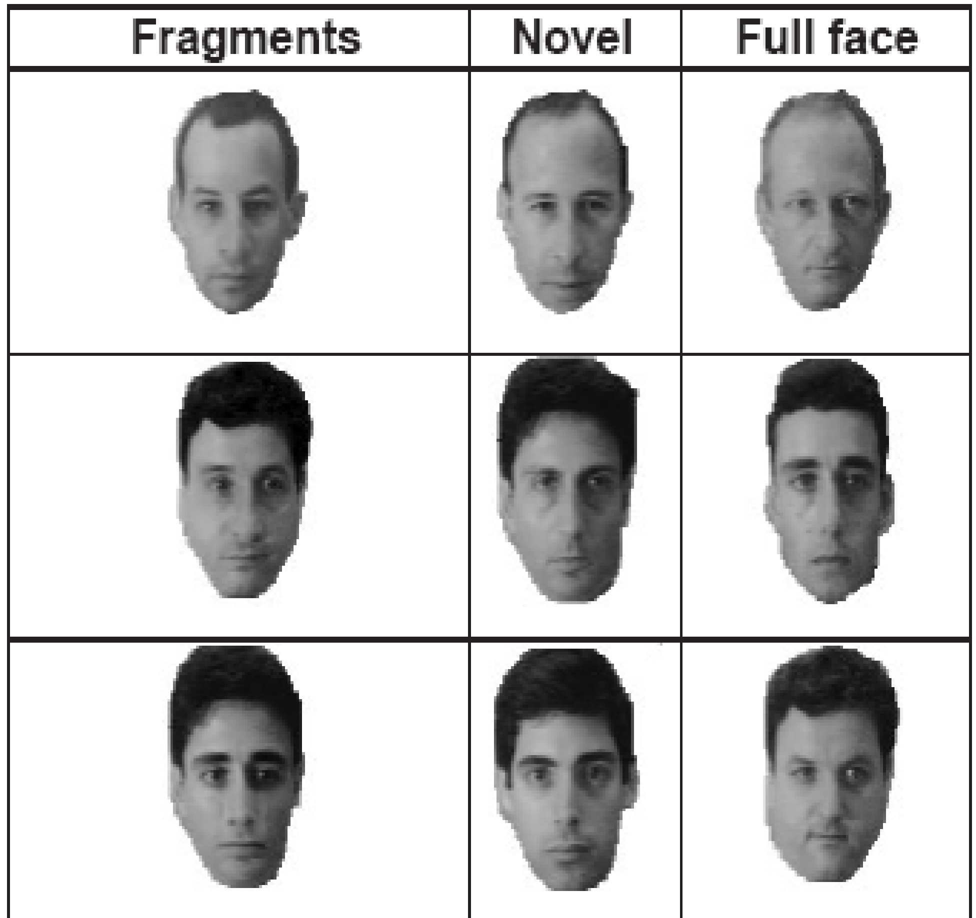
Arrangement Specificity



Part 3: Other Things

Other Things

- Matching:



- *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