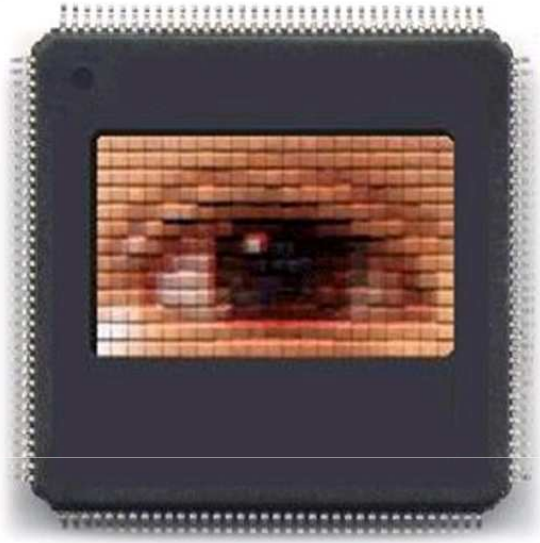
Biological Perspective on Pattern Recognition

Presenter: Jianchao Yang
Course Instructor: Prof. Derek Hoiem

Computer vision follows human vision?



- We want computer to ‘see’ like human, how?
- Do we want the computer to ‘see’ more like a computer or more like a human?

Computer vs. Human

- Modern computers can perform many complex tasks much faster, more efficiently and precisely than human.
- In pattern recognition, a three-year-old can outperform the most sophisticated algorithms available today.

Topics today

- Object recognition with features inspired by visual cortex
- Face recognition by human: 20 results all computer vision researchers should know about.

Object Recognition with Features Inspired by Visual Cortex

Thomas Serre et.al

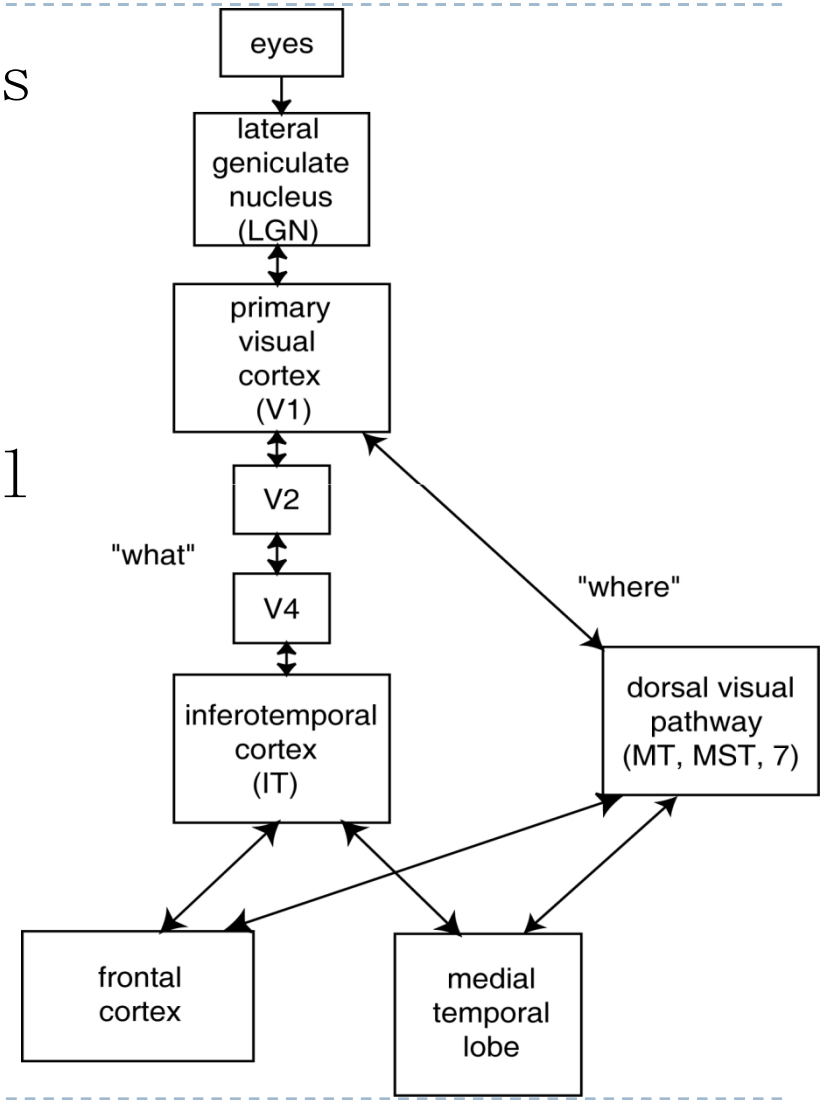
Feature as a trade-off

- Template-based approach
  - Selective, but too rigid to capture the variations of object appearance
- Histogram-based approach
  - Robust to image transforms, but not selective enough
- "Learning the Discriminative Power-Invariance Trade-off"

Feature guided by biological systems

- Computer vision techniques inspired by human vision have never pasted V1

- Follows the standard model in primary visual cortex (V1)



Standard Model of Visual Cortex (V1)

- "Hierarchical models of object recognition in cortex"

- Simple version: four layers of alternating computation units of S units and C units

  - *Simple* S units: Gaussian-like tuning to increase selectivity.
  - *Complex* C units: Maximum pooling to increase invariance

Standard Model of Visual Cortex (cont.)

- Fact 1

  Visual processing is hierarchical, aiming to build invariance to position and scale first and then to viewpoint and other transformations.

- Fact 2

  The receptive fields of the neurons as well as the complexity of their optimal stimuli increases.

Standard Model of Visual Cortex (cont.)

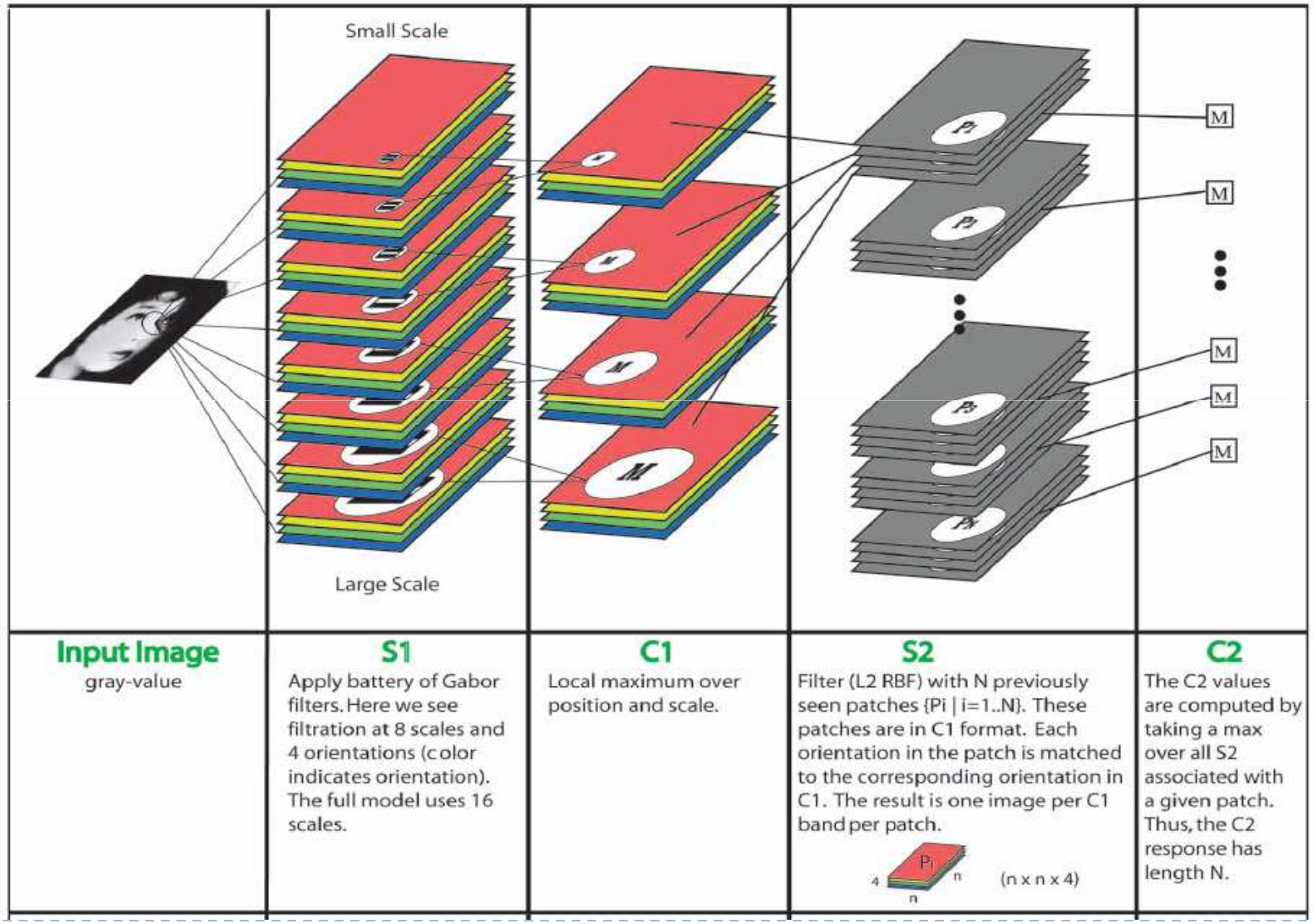
- Fact 3

  The initial processing of information is feed-forward.

- Fact 4

  Plasticity and learning probably occurs at all stages and certainly at the level of IT and PFC, the top layers of the hierarchy.

System overview



S_1 Units

- Multi-dimensional array of filters, arranged by scales and orientations
- Take the form of Gabor functions

$$F(x,y) = \exp\left(-\frac{(x_o^2 + \gamma^2 y_o^2)}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda} x_o\right), \quad \text{s.t.}$$

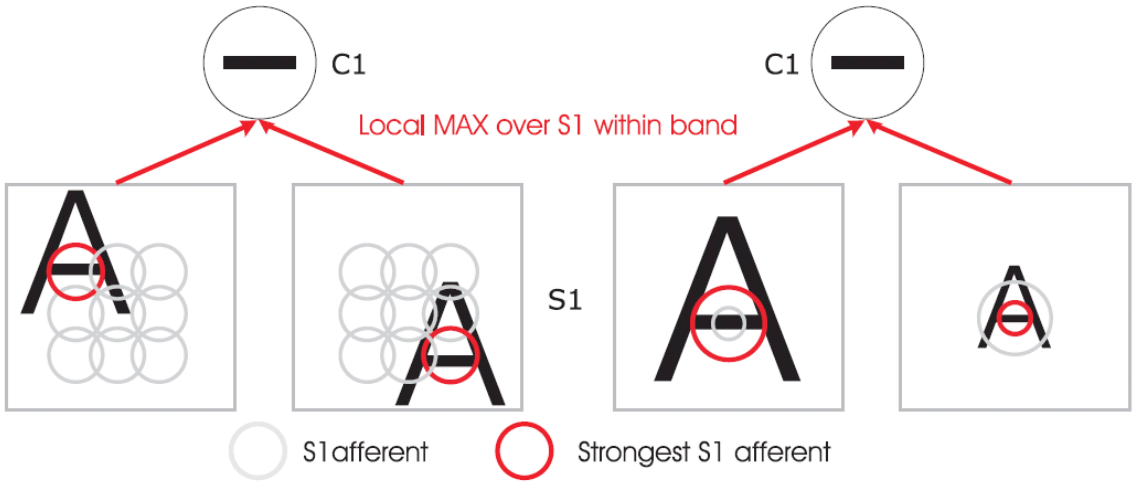
$$x_o = x\cos\theta + y\sin\theta \text{ and } y_o = -x\sin\theta + y\cos\theta.$$

- 16 Scales (7-by-7 to 37-by-37 pixels) and four orientations (0, 45, 90 and 135 degrees) used

C_1 Units

- Larger receptive fields (twice as large as S_1)
  - Add more tolerance to shift and size

- C_1 units spatially pool over S_1 inputs by max operation for each band and each orientation.



S_1 and C_1 Parameters

C_1 layer			S_1 layer		
Scale band $\mathcal{S}$	Spatial pooling grid ($N_S \times N_S$)	Overlap Δ_S	filter size s	Gabor σ	Gabor λ
Band 1	8×8	4	7×7 9×9	2.8 3.6	3.5 4.6
Band 2	10×10	5	11×11 13×13	4.5 5.4	5.6 6.8
Band 3	12×12	6	15×15 17×17	6.3 7.3	7.9 9.1
Band 4	14×14	7	19×19 21×21	8.2 9.2	10.3 11.5
Band 5	16×16	8	23×23 25×25	10.2 11.3	12.7 14.1
Band 6	18×18	9	27×27 29×29	12.3 13.4	15.4 16.8
Band 7	20×20	10	31×31 33×33	14.6 15.8	18.2 19.7
Band 8	22×22	11	35×35 37×37	17.0 18.2	21.2 22.8

S_2 Units

- Behaves as radial basis function units
- The response r of the corresponding S_2 unit for an input patch X from C_1 is given by:

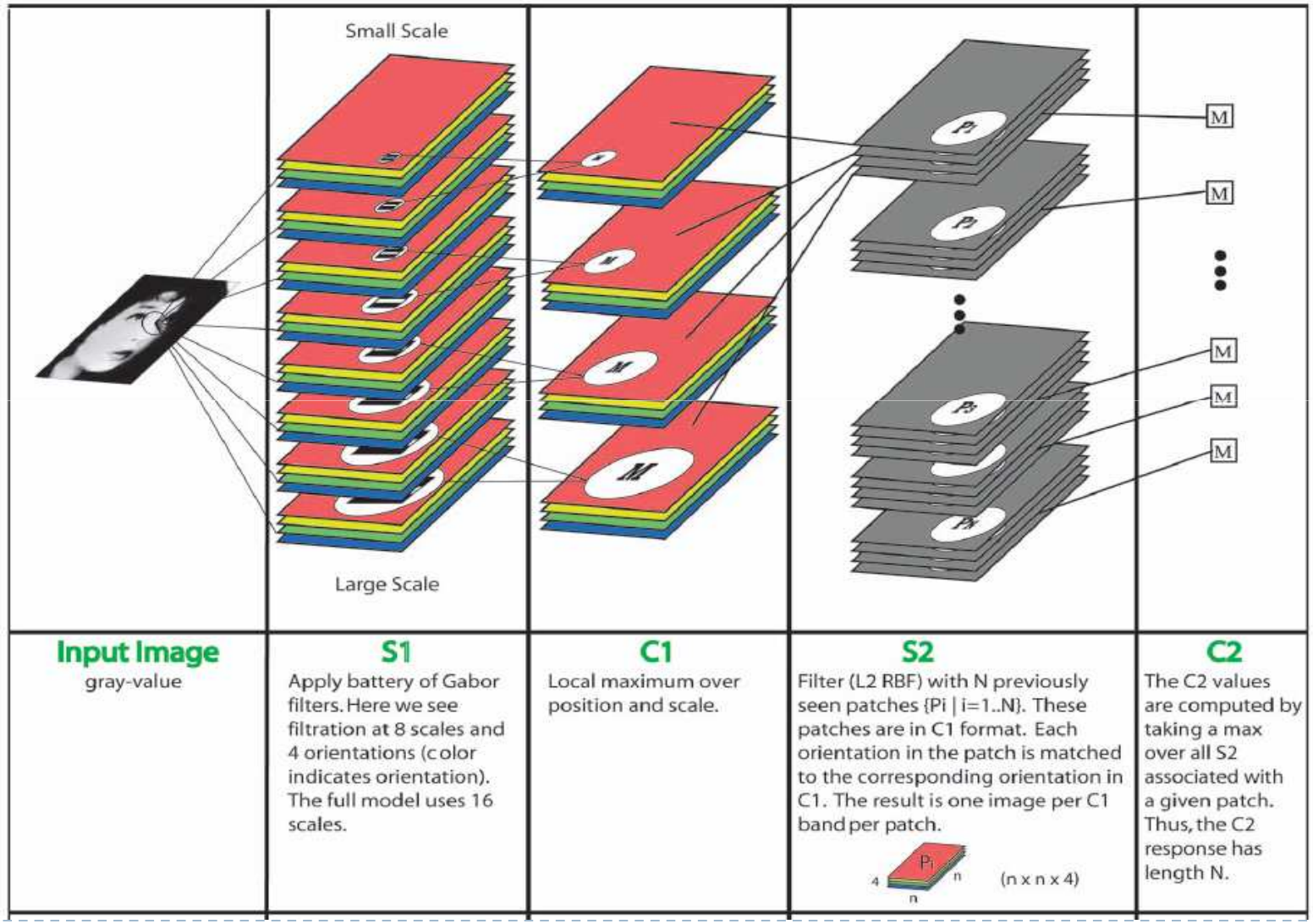
$$r = \exp\left(-\beta \|\mathbf{X} - \mathbf{P_i}\|^2\right)$$

- P_i is the learned prototype. S_2 maps are computed across all positions for each of the eight scale bands.

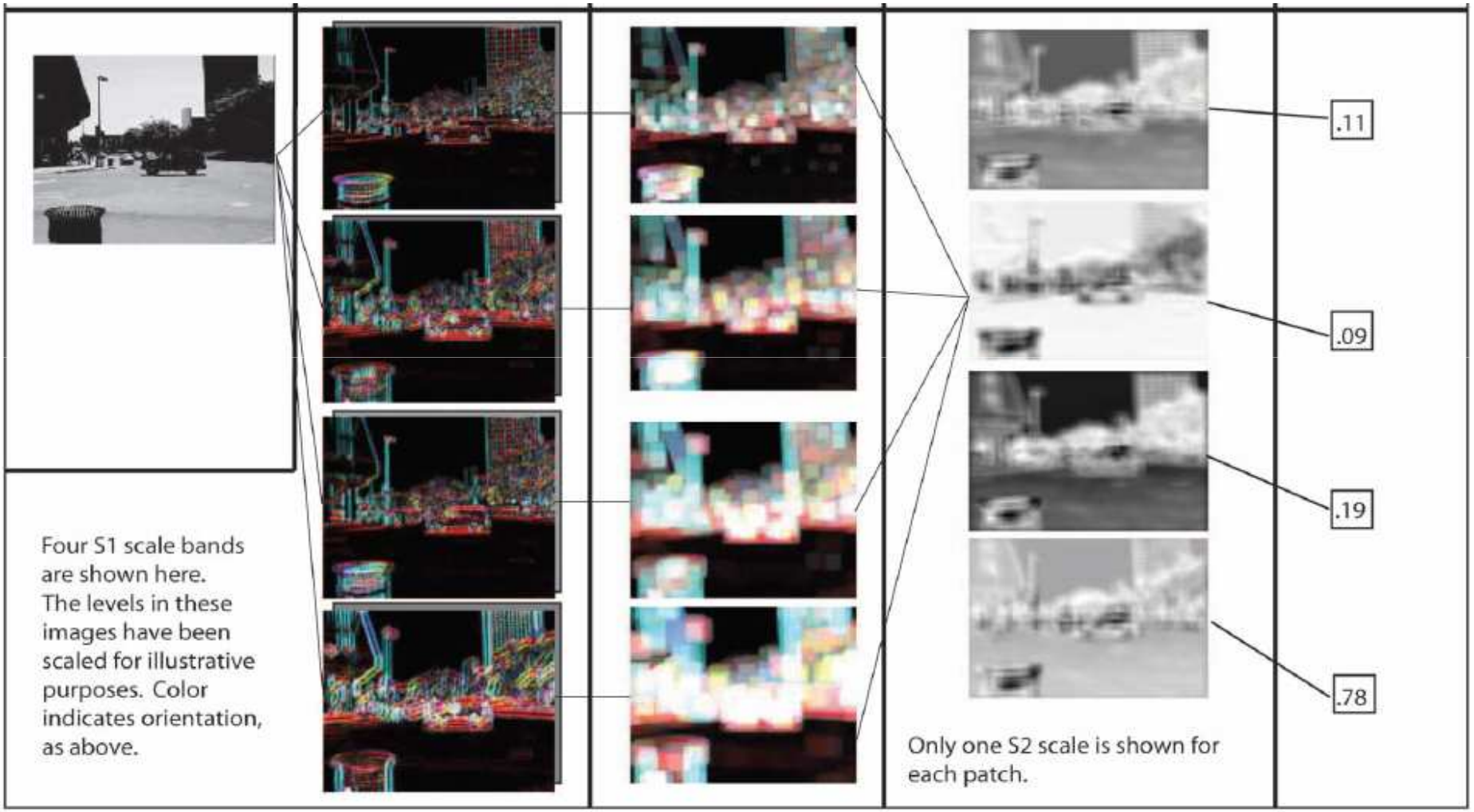
C_2 Units

- Max pooling over all scales and positions for each S_2 type over the entire S_2 lattice.
- C_2 is totally invariant to shift and scale changes.
- Final feature is the combination of C_1 and C_2 (local invariance and global invariance)

System Revisit

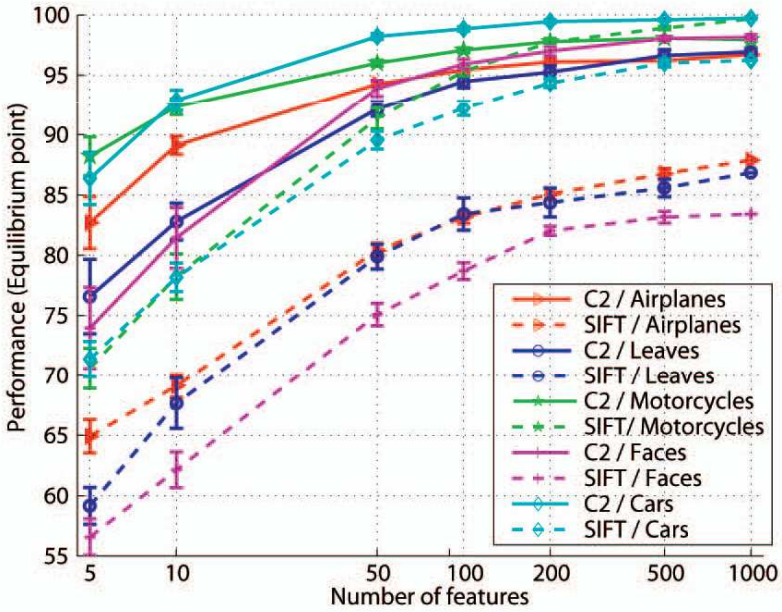


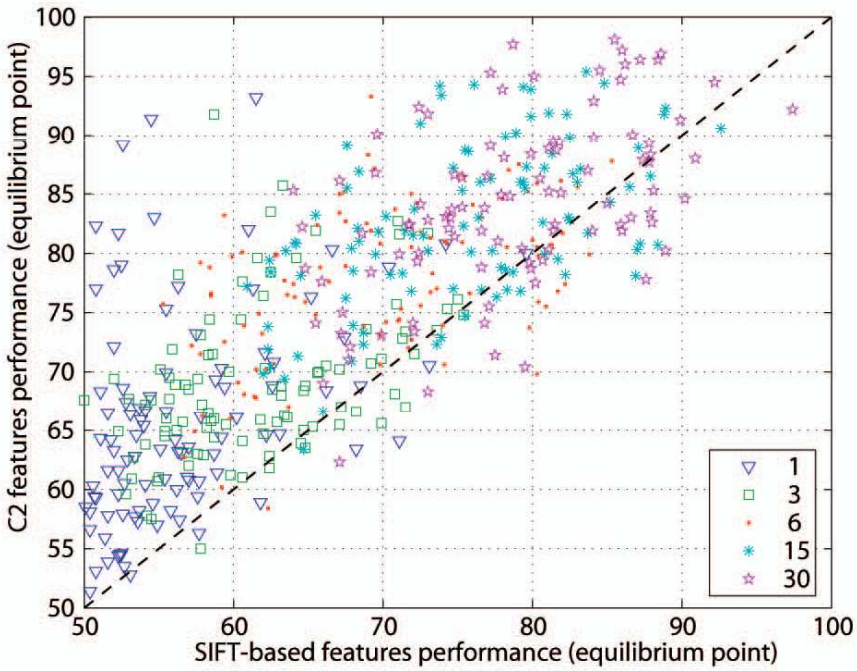
System Revisit: example



Evaluation

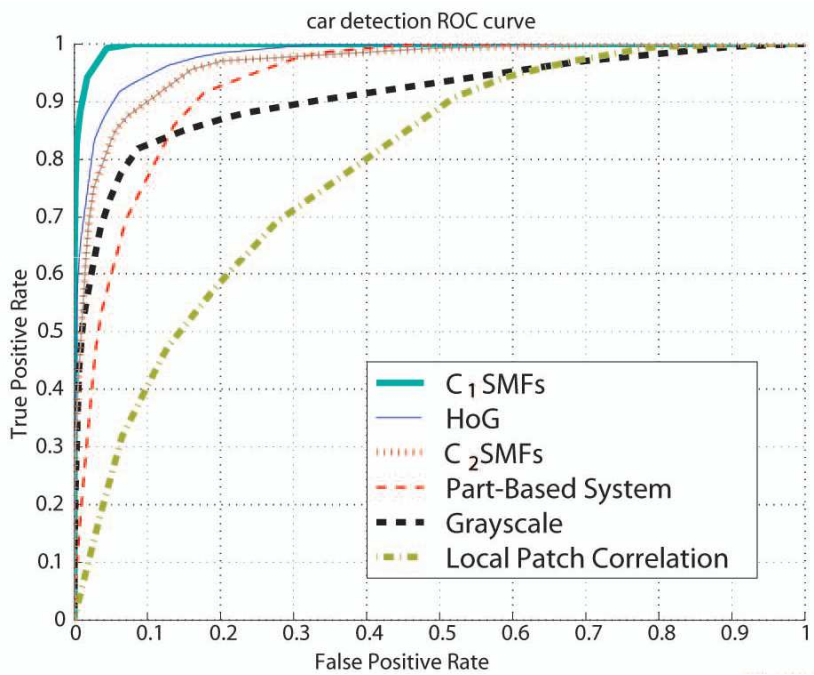
- Detection on five classes of Caltech 101
- Compared with key-point SIFT features

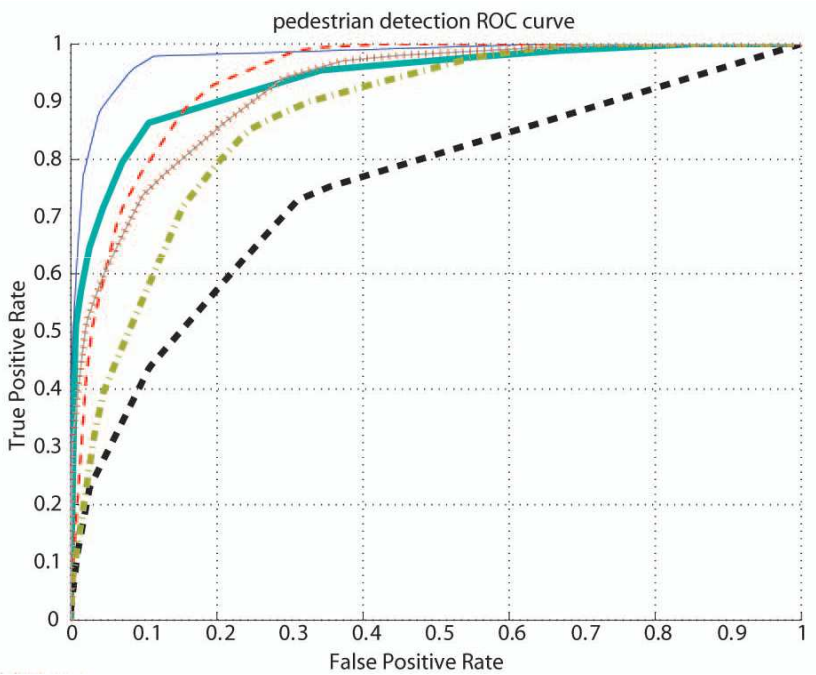




Evaluation (cont.)

- Street scene database





My comments about the paper

- Use neuroscience as a guide to build a complex system for object recognition.
- The feature is somewhat robust to shift and scale changes. C1 is still too rigid on spatial variations.
- C_2 is somewhat like the histogram feature.

Questions to ask

- Should we use neuroscience as a guide for computer vision?

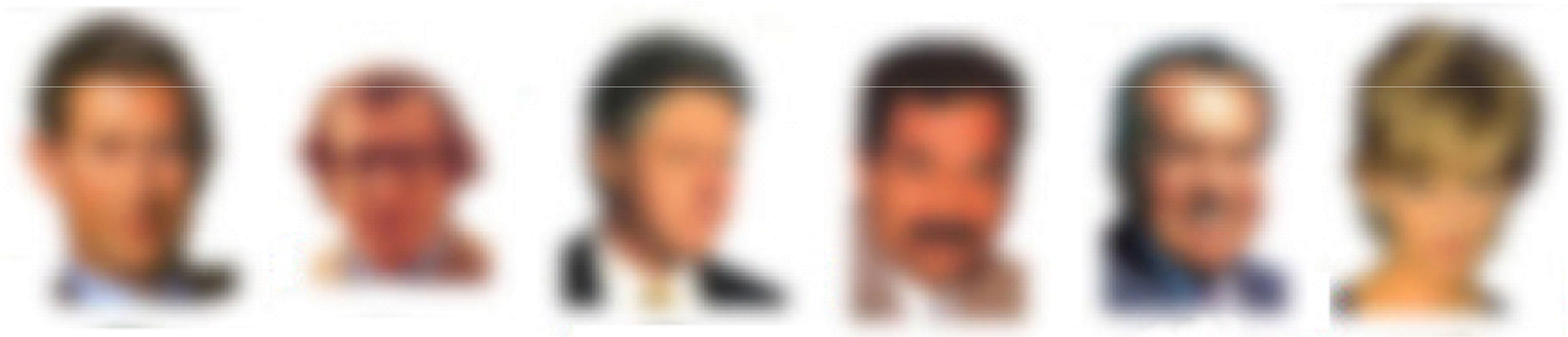
- If not, what should be the principle for constructing a complicated system that work for real applications?

Face Recognition by Humans

20 Results all Computer Vision Researcher Should Know About

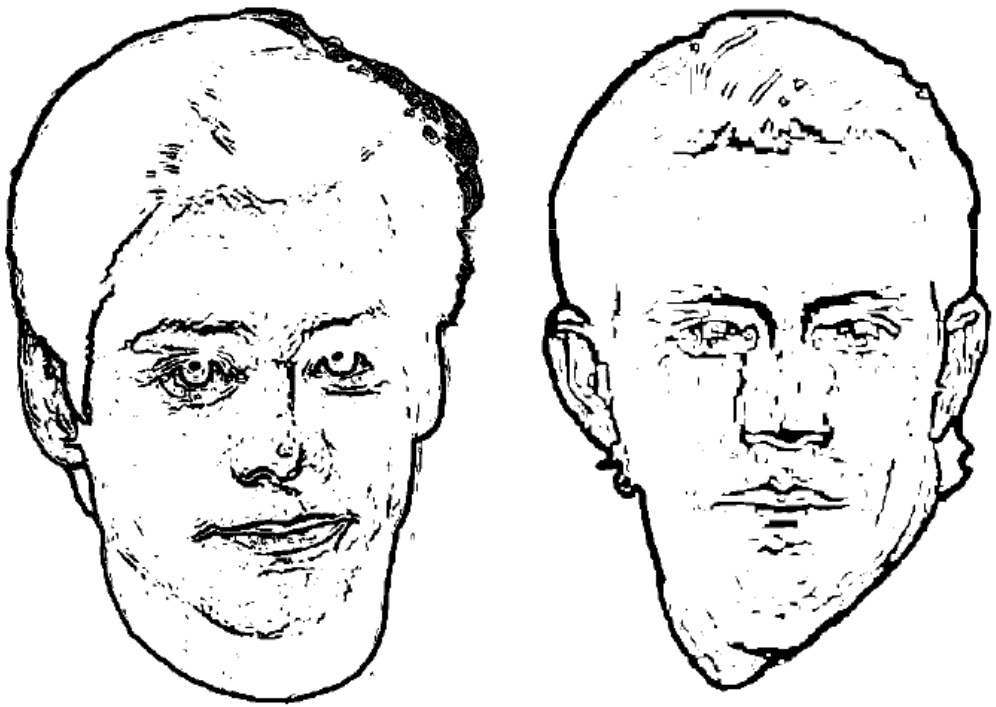
Result 1

- Humans can recognize faces in extremely low resolution images.



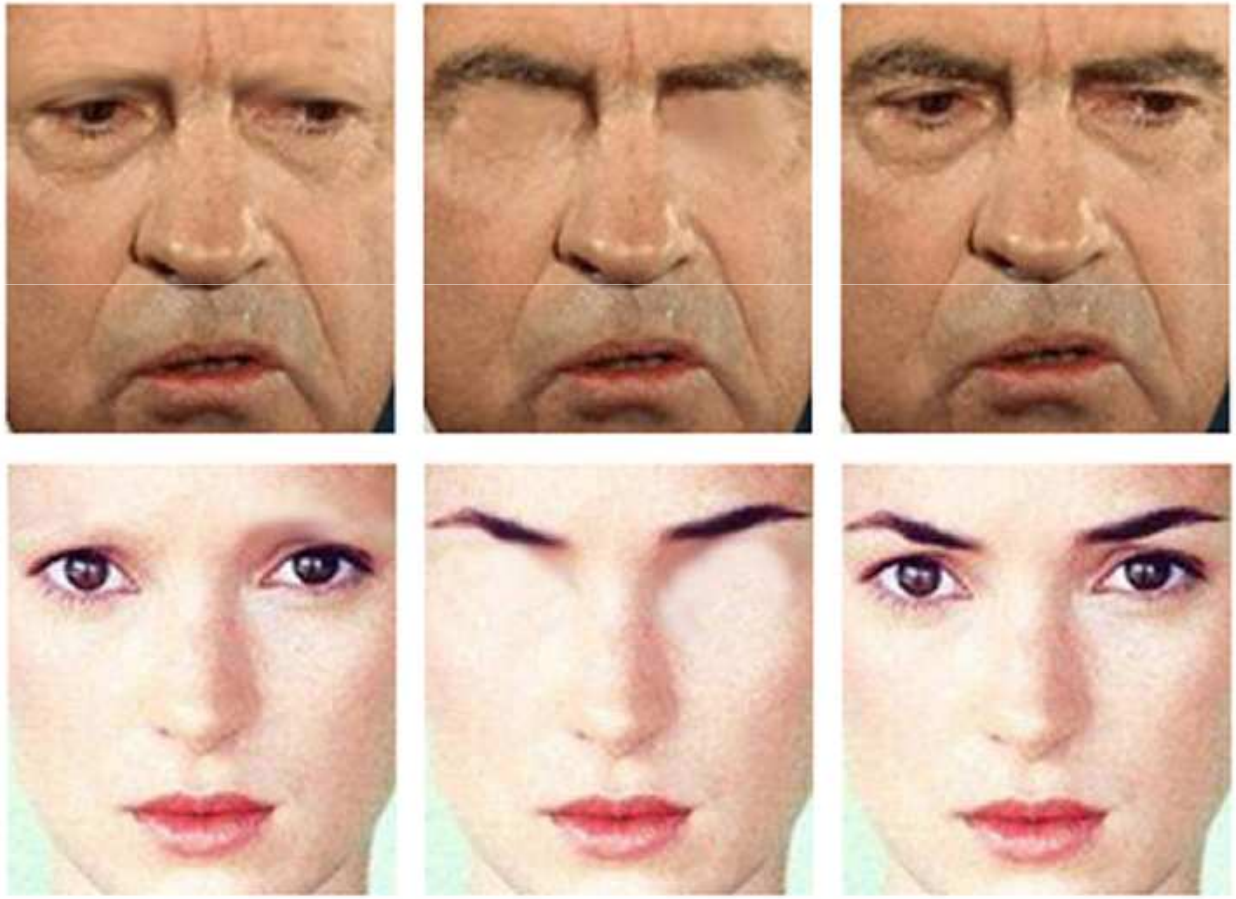
Result 3

- High-frequency information by itself does not lead to good face recognition performance



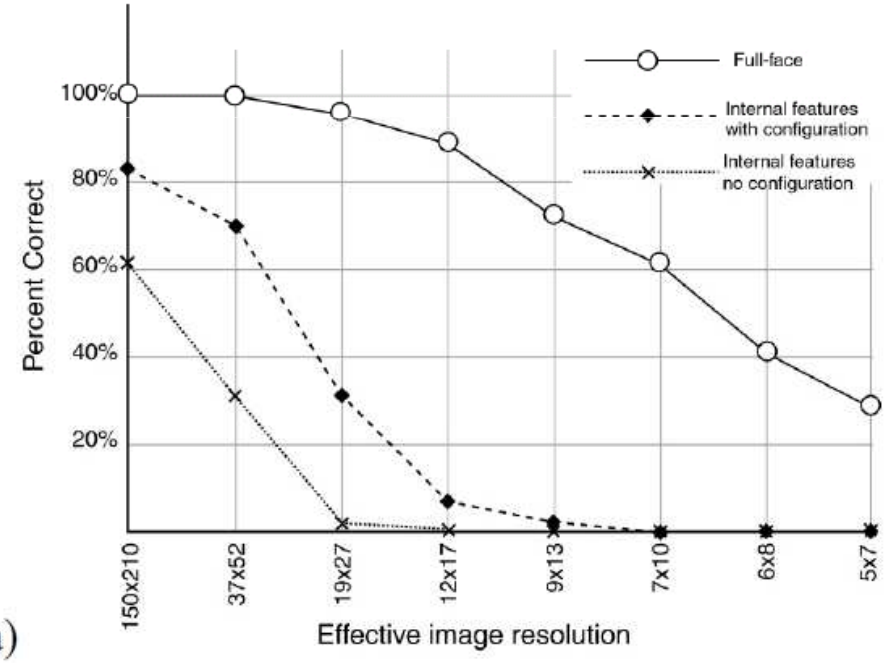
Result 5

- Eyebrows are among the most important for recognition



Result 6

- Both internal and external facial cues are important and they exhibit non-linear interactions



(a)

(b)

Result 7

- The important configural relations appear to be independent across the width and height dimensions

Result 8

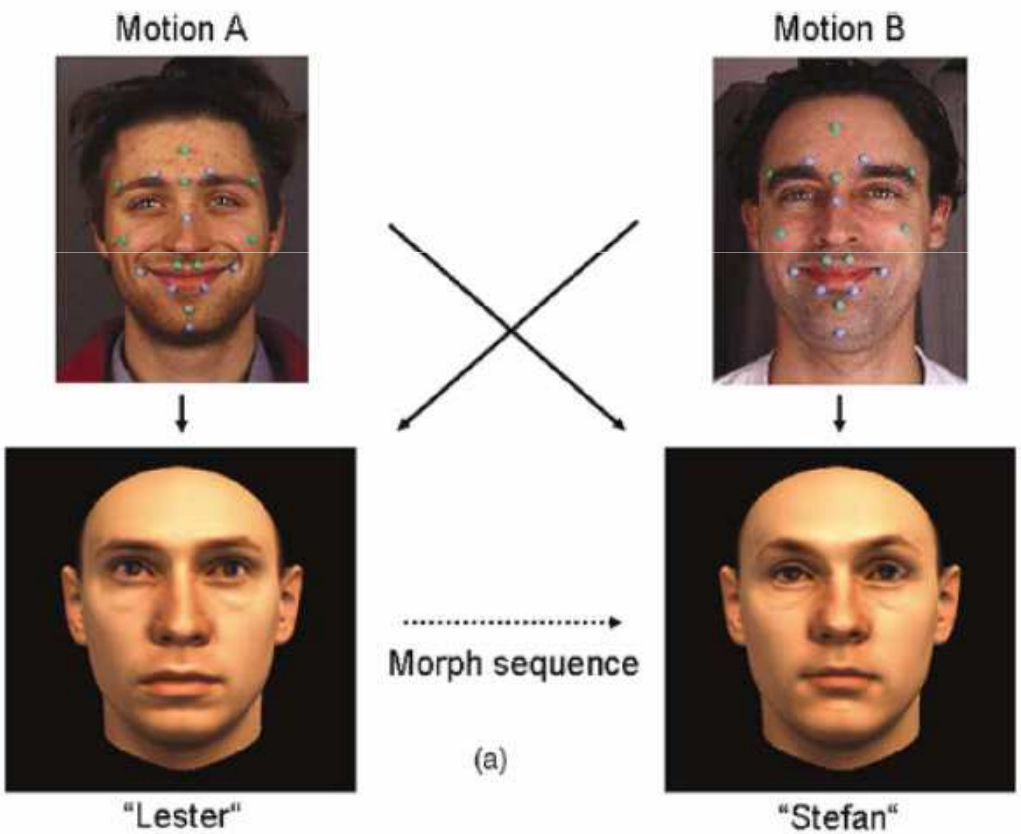
- Vertical inversion dramatically reduces recognition performance

Result 12

- Contrast polarity inversion dramatically impairs recognition performance, possibly due to compromised ability to use pigmentation cues

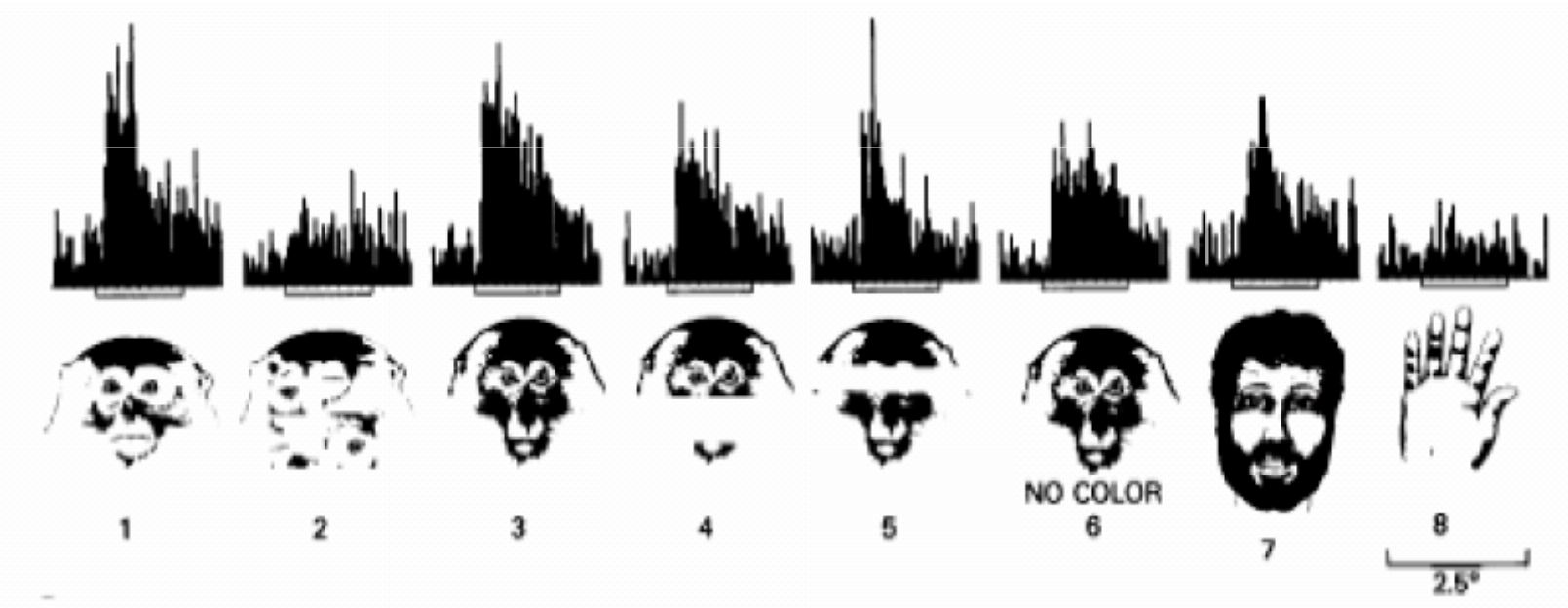
Result 15

- Motion of faces appears to facilitate subsequent recognition



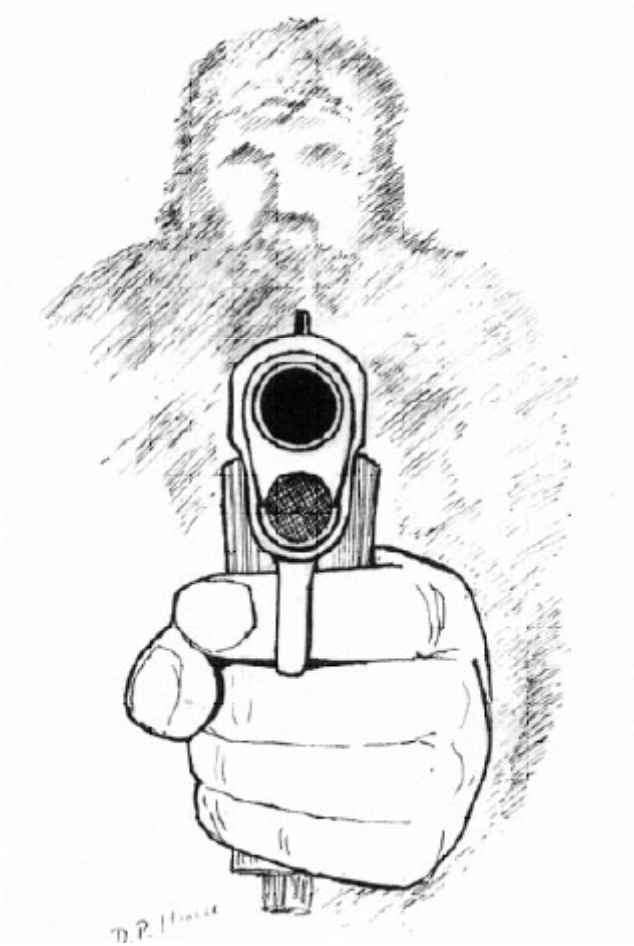
Result 19

- Latency of responses to faces in IT cortex is about 120 ms, suggesting a largely feed-forward computation



Result 20

- Human memory for briefly seen faces is rather poor



Discussion