Representing 3D Objects: An Introduction to Object Centered and Viewer Centered Models

Guangyu Zhu

CMSC828J Advanced Topics in Information Processing: Approaches to Representing and Recognizing Objects

Professor David Jacobs

March 7, 2006

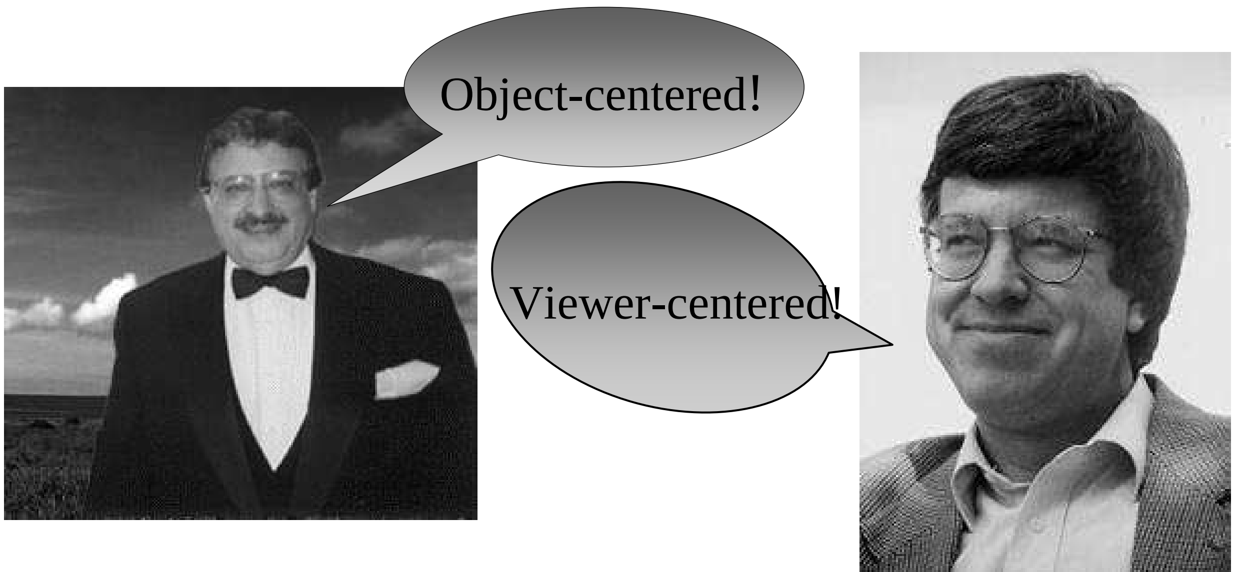
UNIVERSITY OF MARYLAND

Outline

- Review different architectures for the recognition of 3D objects
- Compare the basically different approaches of object centered and viewer centered representations
- Discuss biological findings and computational aspects
- Illustrate the advantages of viewer centered models by some recent recognition systems
- Experimental studies on face perception
- Conclusion

Introduction

- Why we need to talk about the representation schemes of objects
- Object centered (viewpoint invariant) models
- Viewer centered (viewpoint dependent) models



Biederman

Bulthoff

Object centered representations

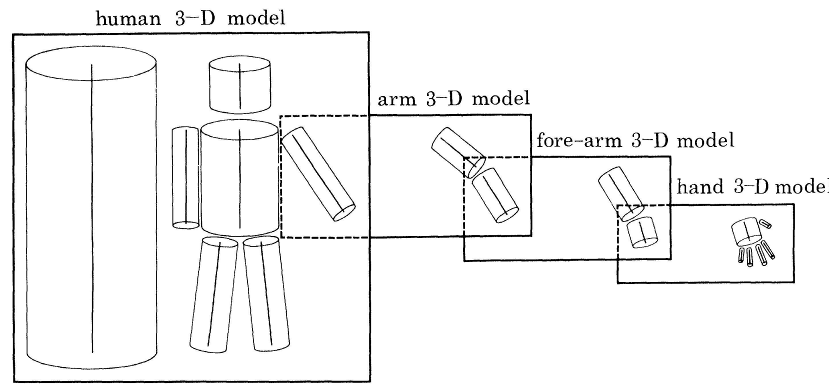
- *Generalized cones* introduced by Marr and Nishihara (1978)
- *Geon structural descriptions (GSD)* proposed by Biederman (1987)
- Thompson and Mundy (1987) and Lowe (1986)
  - based on surface descriptions built upon vertices, edges, and surfaces in conjunction with their connection relation
- Have the one common purpose:
  - The description of objects by high level features which remain stable over all perspectives

Marr's Stages of Visual Processing

- Marr described vision as processing from input of a 2-D visual array (on the retina) to a 3-D description of the world as output. His stages of vision include:
  - **Primal Sketch**: based on feature extraction of fundamental components of the scene, including edges, regions, etc.
  - **2 1/2-D Sketch**: depth and orientation of visible surfaces, shading, texture, motion, binocular disparity; observer-centered
  - **3-D Sketch**: 3-D description of objects independent of view
- Proposed that understanding the brain requires an understanding of the problems it faces and the solutions it finds

Marr & Nishihara (1978)

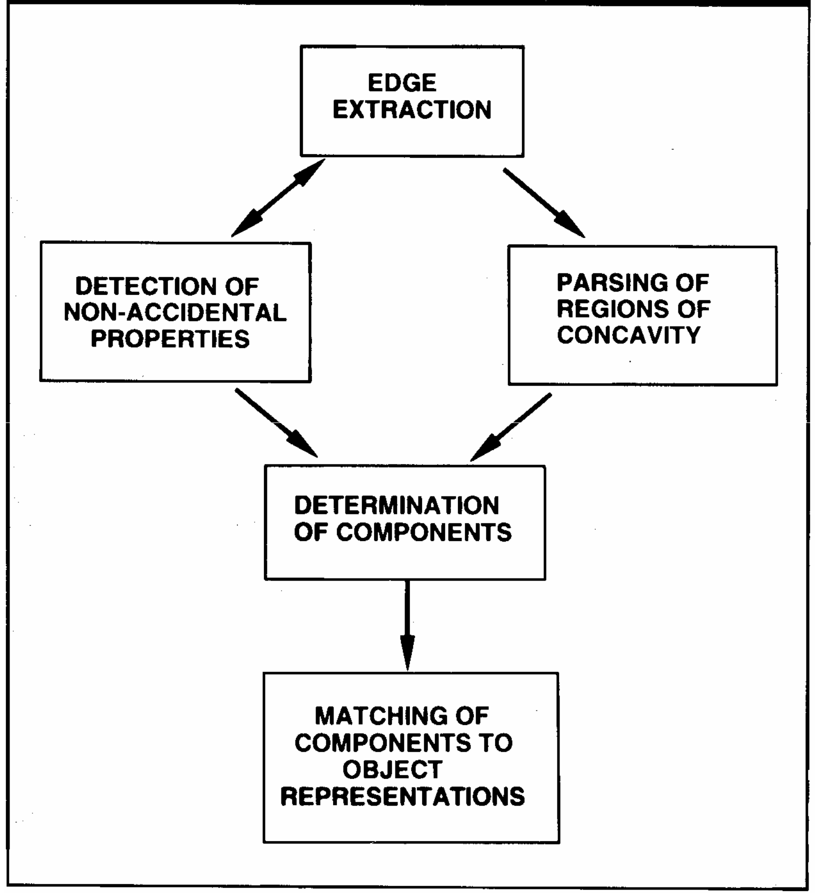
- Development of 3-D sketch based on processing of more elementary shape primitives (basic primitive is a cylinder with a major axis)
- Hierarchical organization of primitives
  - Extended into “recognition by components” (Biederman, 87)
- Concavities important in segmenting parts



Recognition by Parts (Biederman, 87)

- Objects built from primitives called “geons” (n = 36)
- Represent objects by volumetric primitives together with their relations
- Two key components of decision:
  - locating concavity
  - deciding which edge information remains invariant across different viewing angles (invariant properties like curvature, parallelism, etc.)

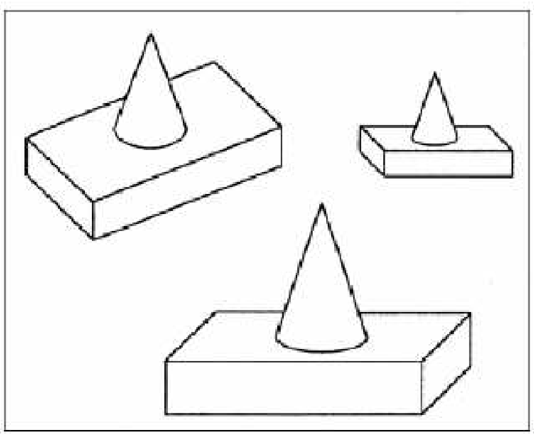
FIGURE 3.7

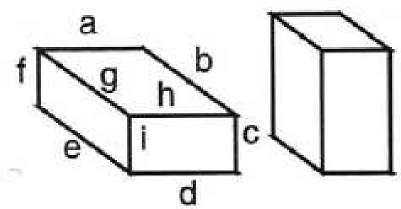


Recognition by Parts (Biederman, 87)

Nonaccidental features for each geon that can be identified in images independent from viewpoint

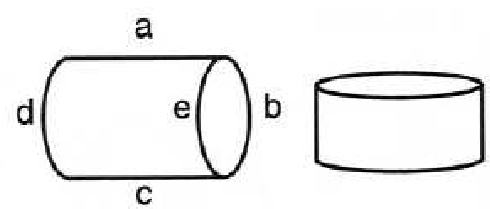
View invariance





"BRICKS"

- 3 sets of 3 parallel edges:
  (a, h, d) (b, e, g) (c, f, i)
- 1 inner Y-vertex:
  (ghi)
- 3 outer arrow vertices:
  (afg) (bch) (dei)



"CYLINDERS"

- 2 parallel straight edges:
  (a, c)
- 2 parallel curved edges:
  (d, e)
- 2 tangent Y-vertices:
  (abe) (bce)

Problems of RBC

- Structural description not enough, also need metric info
- Difficult to extract geons from real images
- Ambiguity in the structural description: most often we have several candidates
- For some objects, deriving a structural representation can be difficult
- Empirically there is view-dependence

Viewer centered models

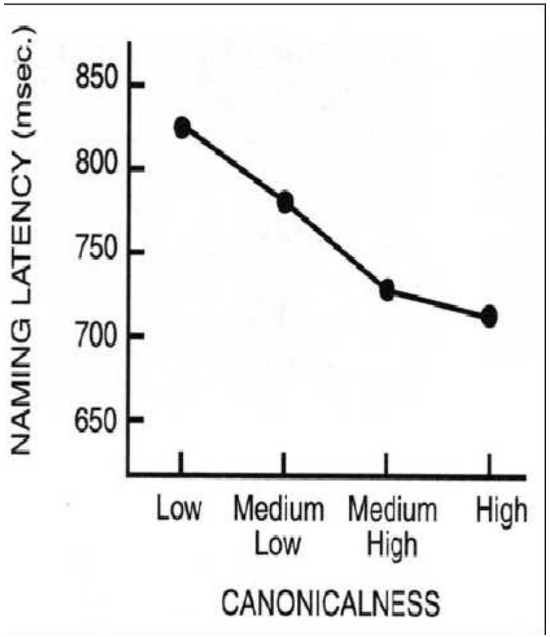
Problem

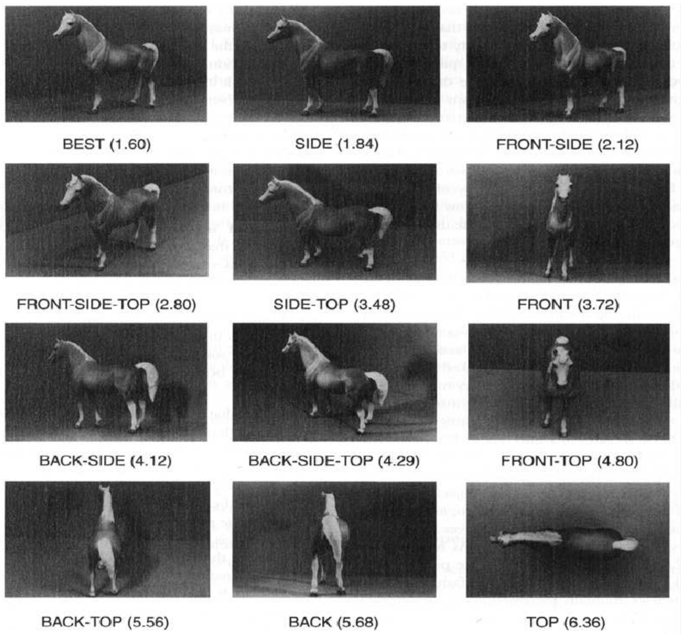
- Very large number of views might have to be stored per object

Solutions

- Alignment of stored and perceived view
- Generalization/interpolation between learned views

Canonical views





Rated typicality of object views
(Palmer, Rosch and Chase 1981)

Viewer centered models

- Three-Dimensional Models
  - *Recognition by alignment* introduced by Ullman (1989)
- Two-Dimensional Models
  - Ullman and Basri (1991)
    - No long restricted to rigid transformations
    - Does not involve the explicit reconstruction and representation of the 3D structure for the storing the objects
    - Prove that under certain assumptions, all the views of a 3D object can be derived from the linear combination of a few 2D views
    - Presuppose a correspondence between features of the input image and the model views
    - Require the visibility of all object points from every perspective

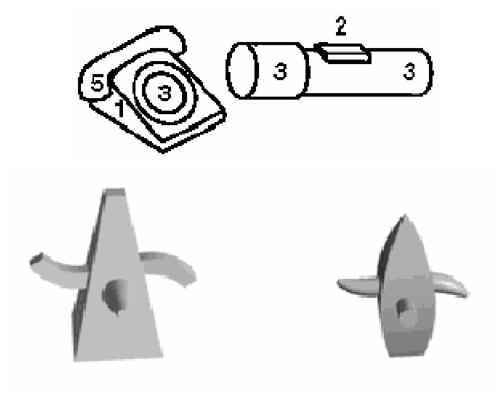
Viewer centered models

- Two-Dimensional Models (cont)
  - Poggio and Edelman (1990)
    - An early implementation of a view based recognition system using an artificial neural network
    - Postulate that for every object an appropriate function can be found which is capable of transforming all possible views into a single standard view
    - The approximations of these functions are expected to be evolved by *RBF* networks (*Radial Basis Functions*)
    - Require a constant number of feature points together with an exact correspondence relation between image and model
  - *CLF* network (*Conjunctions of Localized Features*) suggested by Edelman and Weinshall (1991)
    - Do not need the computation of an explicit correspondence but use topological feature maps

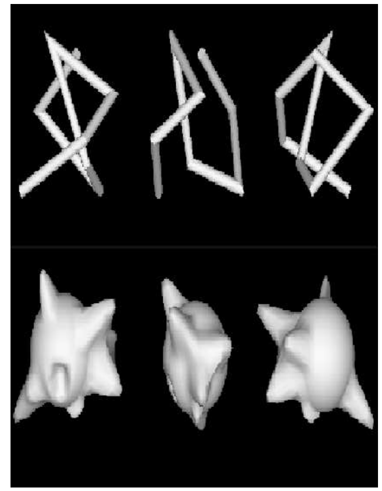
Viewer centered models

- Models Utilizing View Sequences
  - *VIEWNET* architecture (*View Information Encoded with NETworks*) described by Grossberg and Bradski (1995)
    - Demonstrate the advantages resulting from the consideration of view sequences instead of single images
    - Include a biologically motivated preprocessing chain
    - Still neglect the order in which the views appear
  - Evaluation of view sequence by Seibert and Waxman (1992)
  - Darrell and Pentland (1993)
    - Use simple image processing algorithms
    - Training and recognition require an alignment

Psychophysical evidence (Bulthoff et al., 1994)



Biederman



Paper-clip

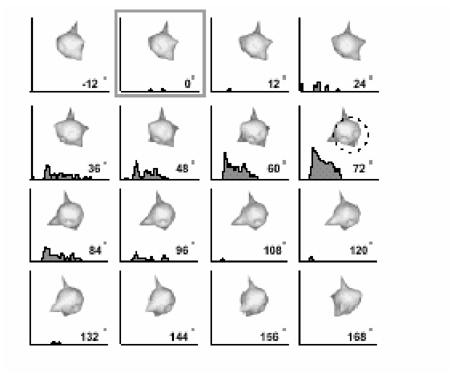
amoeba

Bulthoff, Edelman, Tarr, 1994

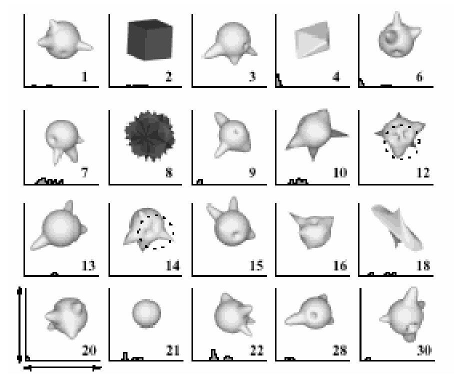
- Subjects presented with realistically rendered images of computer-generated 3D objects
- Tight control of stimulus shapes, surface, illumination, symmetry and viewpoint
- Consistently observed viewpoint dependent

Psychophysical evidence (Bulthoff et al., 1994)

Object

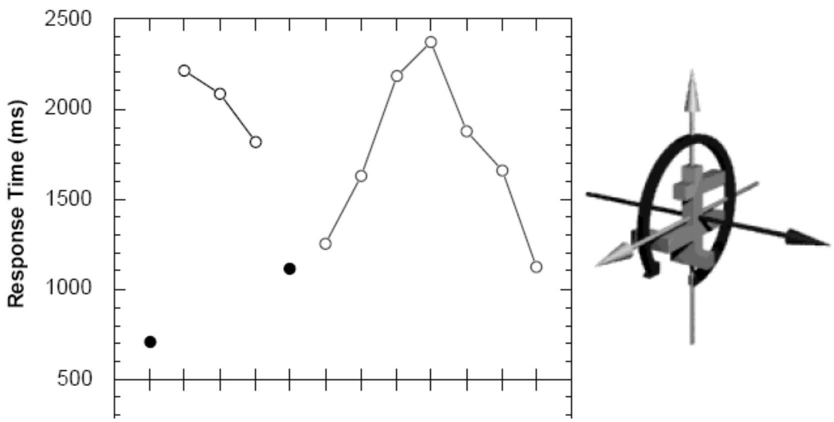


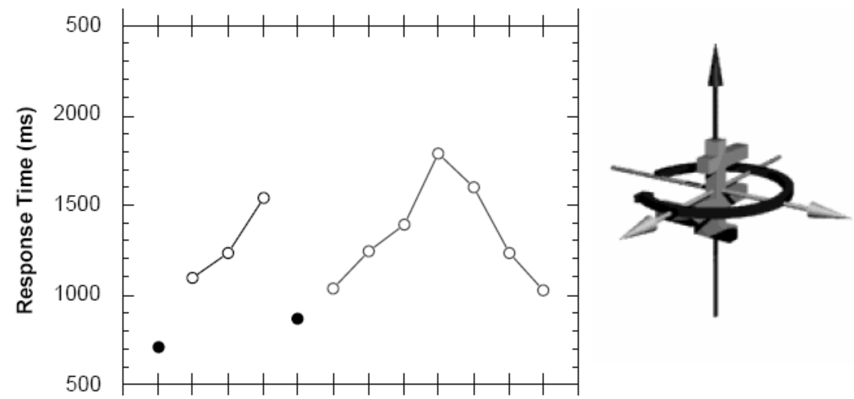
Distractors



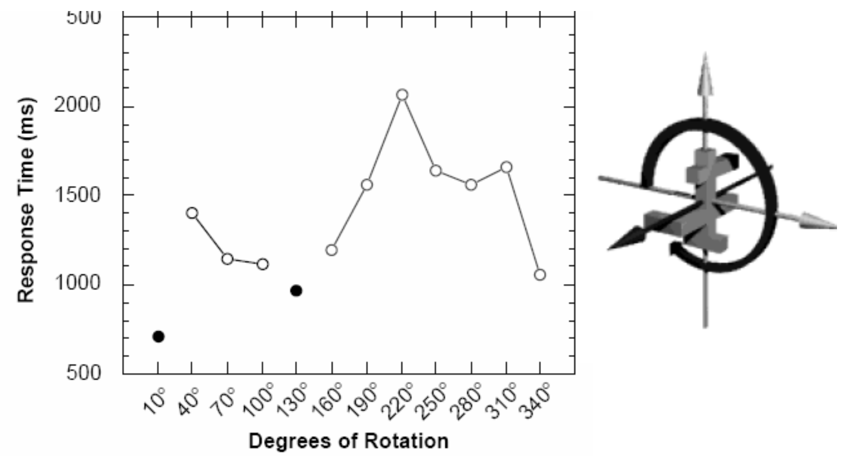
Biological evidence (Bulthoff et al., 1994)

Mean response time in familiar and unfamiliar viewpoints





Filled data points represent familiar viewpoints learned during training, open points represent unfamiliar viewpoints introduced in the “surprise" phase of the experiment



Biological evidence (Bulthoff et al., 1994)

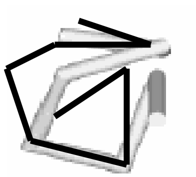
- Object represented as a collection of structurally stored specific views
- Input stimulus activates the representation of that view

V = (o1, o2, o3,…on)

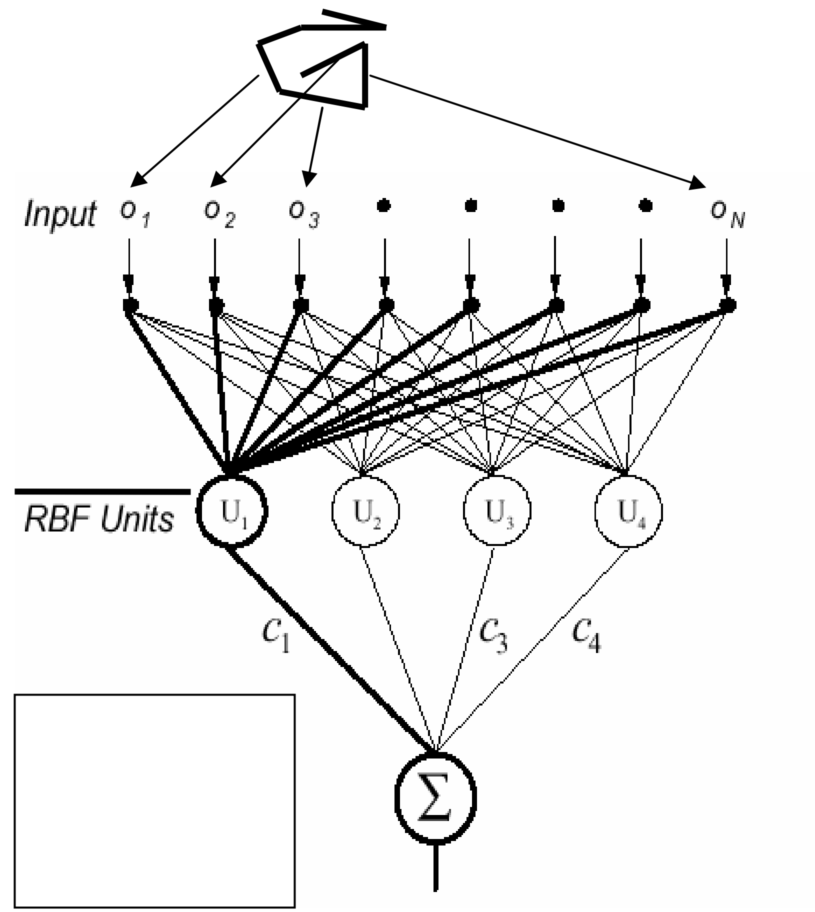
T = (Ot1, Ot2 …Otn)

RBF networks

$$U_i = \exp\left(\frac{-\|\mathbf{V} - \mathbf{T}_i\|}{\mathbf{V}\sigma_i^2}\right)$$

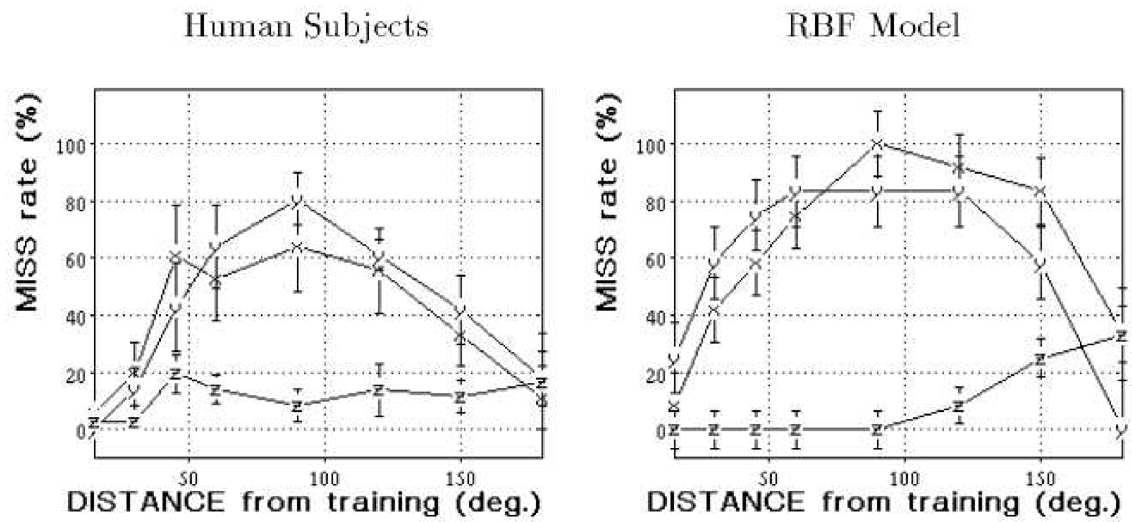


Template matching



Biological evidence (Bulthoff et al., 1994)

Performance of Human and RBF network



Biological evidence (Bulthoff et al., 1994)

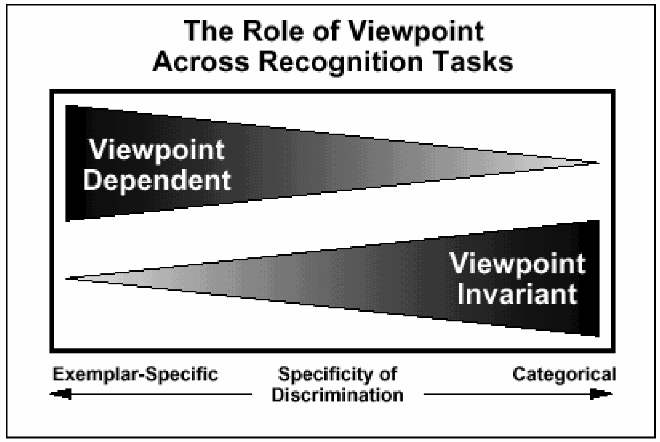
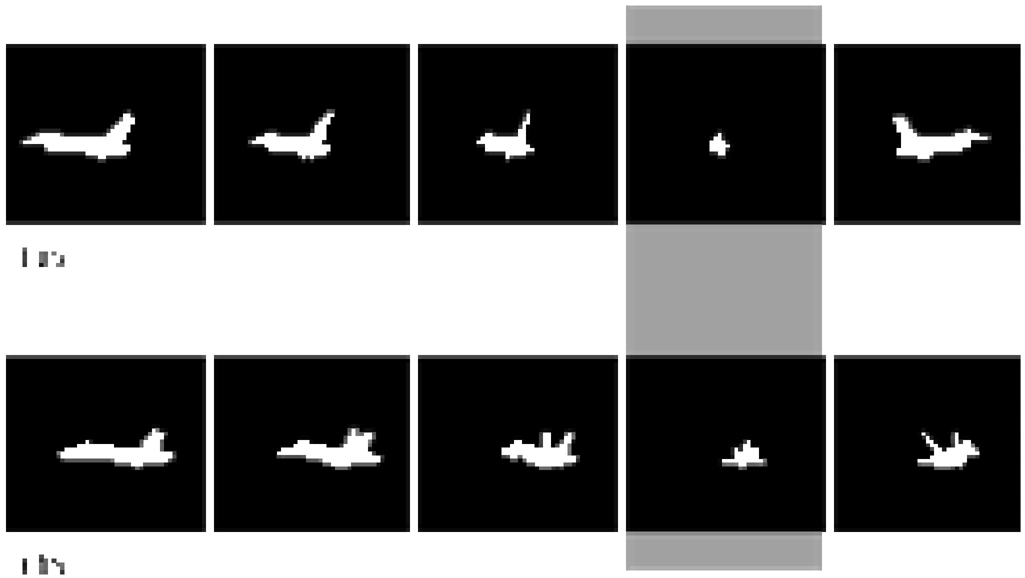


Figure 4:

- Viewpoint dependence most strongly demonstrated in subordinate-level recognition
- Extreme viewpoint dependence and extreme viewpoint invariance lie at two ends of a contiuum, with appropriate mechanisms and features
- Question viewpoint invariant model even in entry-level recognition

Utilizing temporal associations



Seibert and Waxman (1992)

Summary on object centered models

- Already involve a high degree of complexity in representation
- Recognition process is tedious
- Only models of Lowe (1986) and Thompson and Mundy (1987) were realized as functioning object recognition systems
- Implementation limited to relatively simple and completely specified primitive objects
- Applicability to real world remain questionable

Summary on viewer centered models

- Can be built upon 2D instead of 3D views
- Closer relationship to biological findings
- Lead to successful technical implementations in practice

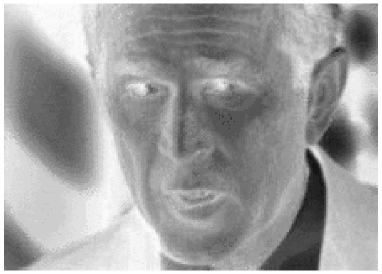
Face perception

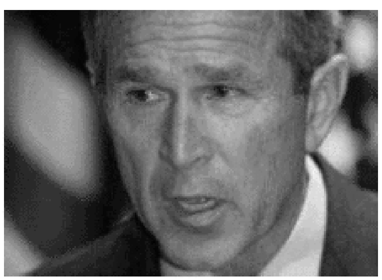
- How does the brain understand and interpret faces
- An important site for the identification of others
- Convey significant social information
- Early development
  - Innate tendency to pay attention to faces from birth
- Adult face perception

Questions

- Do we genuinely develop specific skills for understanding faces or is it just part of a general skill for making within-category discriminations?

Is face recognition special?





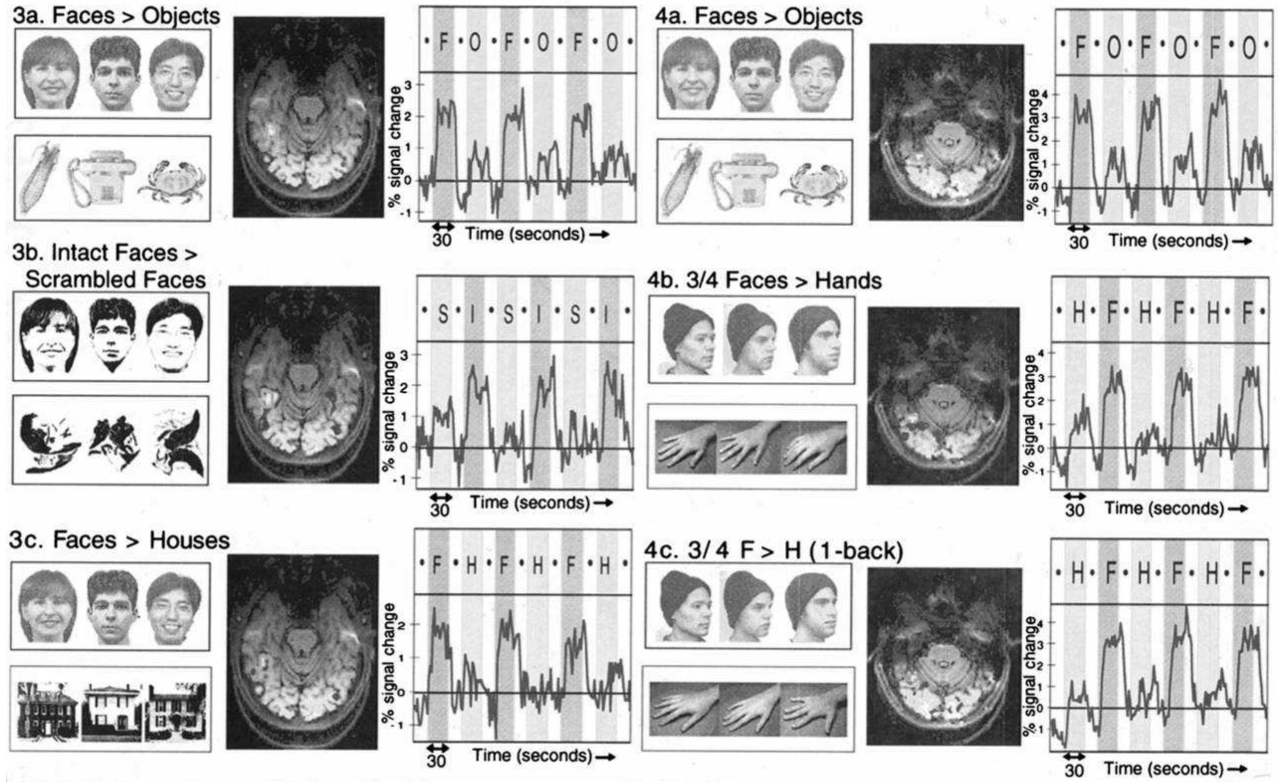
Why should it be special ?

- Inverted and wrong-color images very difficult to recognize
- Stronger orientation dependence than for other objects
- Seems to be more holistic
- Dissociations (object / face perception deficits) in patients with localized lesions
- We have much more experience with faces than with other objects.

Thatcher illusion

Thompson (1980)

Response properties of human fusiform face area



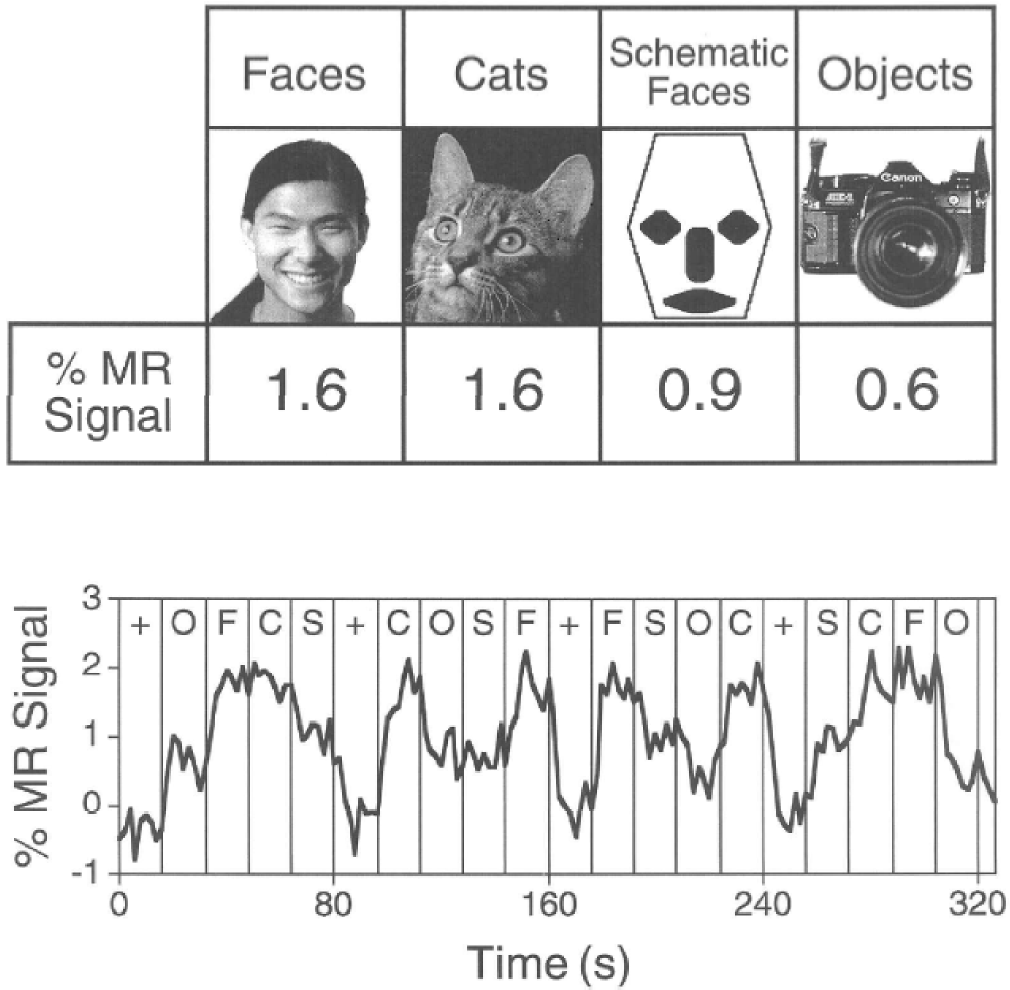
- FFA is a module in human extrastriate cortex specialized for face perception (Kanwisher et al. 1997)

Response properties of human fusiform face area

- Kanwisher et al. 1997
  - low-level feature extraction
  - allocation of attention to faces due to a general attentional bias towards faces
  - subordinate level recognition of category exemplars
  - recognition of any animate (or perhaps only human) objects

Response properties of human fusiform face area

- Tong, Nakayama, Moscovitch, Weinrib, Kanwisher, 2000



Holistic vs. Piecemeal

The whole is greater than the sum of its parts

Latency of responses to faces suggests a largely feed-forward computation

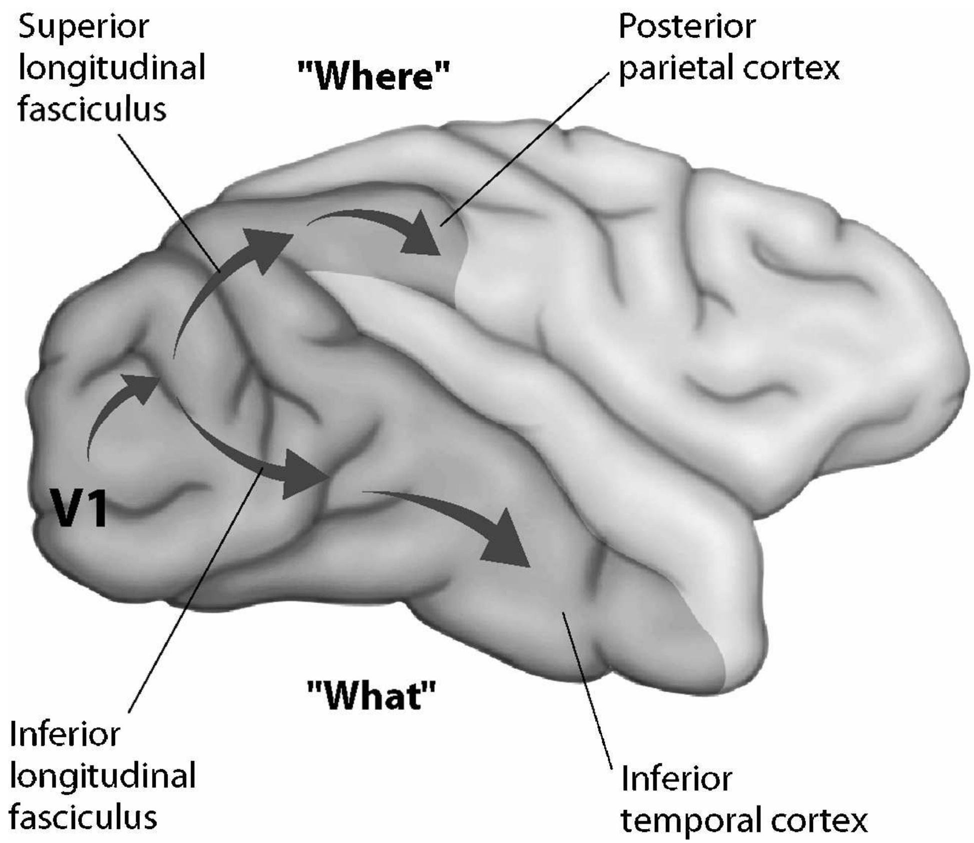


Image-based face recognition system

Approach	Representative Works
Holistic methods	
Principal Component Analysis (PCA)	
Eigenface	Direct application of PCA
Fisherface/Subspace LDA	FLD on eigenspace
SVM	Two-class problem based on SVM
ICA	ICA-based feature analysis
Other Representations	
LDA/FLA	FLD/LDA on raw images
PDBNN	Probabilistic decision based NN
Feature based methods	
Pure geometry methods	Earlier methods, recent methods
Dynamic Link Architecture	Graph matching methods
Convolution Neural Network	SOM learning based CNN methods
Hybrid methods	
Modular eigenface	Eigenface & eigenmodules
Hybrid LFA	Local & global feature method
Component-based	Face region and components

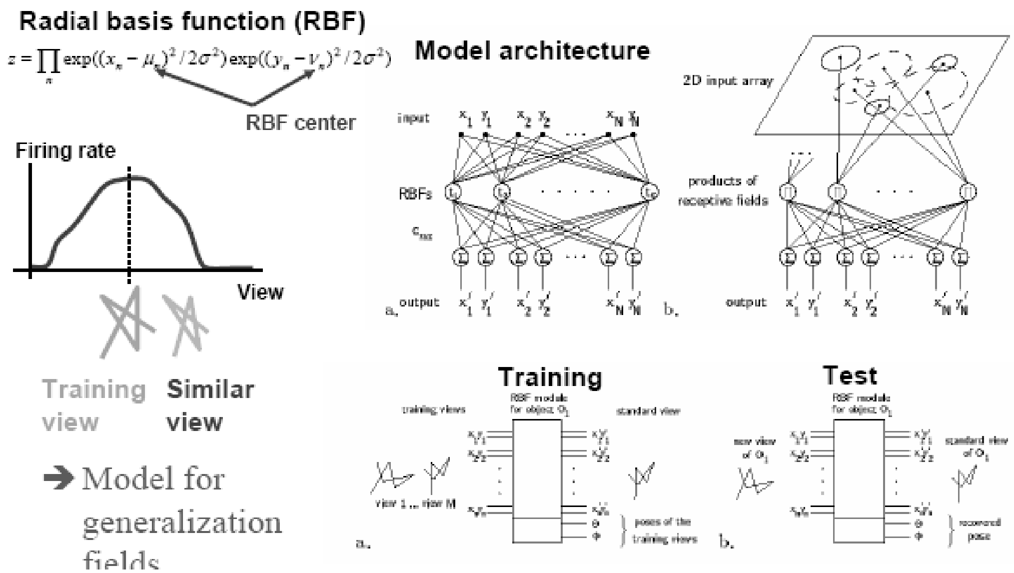
Zhao, Chellappa, Rosenfeld and Phillips, 2003

Question?

References

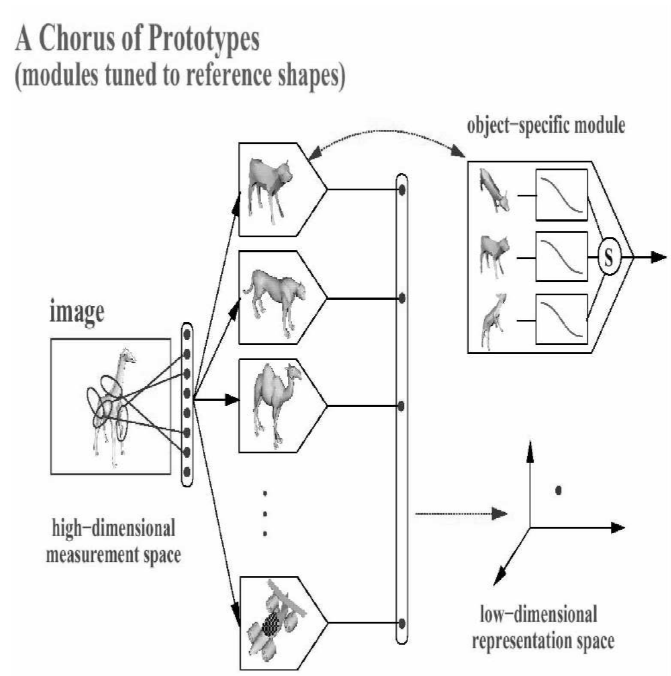
1. **Biederman, I. (1987):** Recognition by components: A theory of human image understanding. *Psychol. Review* 94: 115-147.
2. **Bulthoff, H., Edelman S., and Tarr M. (1994):** How are three-dimensional objects represented in brain?, MIT AI Memo #1479.
3. **D. Lowe (1987):** Three-dimensional object recognition from single two-dimensional images. Artificial Intelligence, 1987.
4. **Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (2000):** Response properties of the human fusiform face area. Cogn. Neuropsychol., 17, 257-279.
5. **Marr, D. and Nishihara, H. K. (1978):** Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. Lond.* 200: 269-294.
6. **Thompson, P. (1980):** Margaret Thatcher: a new illusion. *Perception*. **9**(4):483-4.
7. **Ullman, S. and Basri, R. (1991):** Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(10): 992-1006.
8. **Sinha, P., Balas, B.J., Ostrovsky, Y., & Russell, R (2005):** Face recognition by humans: twenty results all computer vision researchers should know about.

Poggio & Edelman (1990)



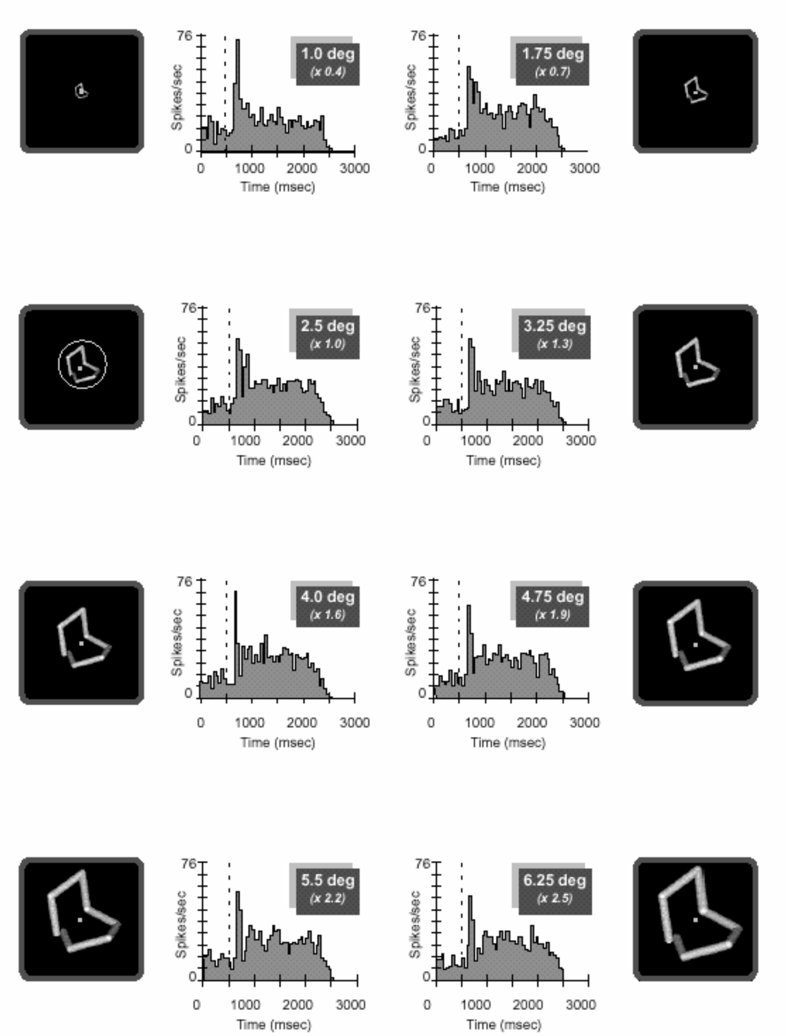
View-based categorization, Edelman (1999)

- Each category represented by ensemble of views
- New categories encoded by distribution of activation over prototypical neurons that represent different categories



Biological evidence (Bulthoff et al., 1994)

Size invariance



Position invariance