Form and Shape Identification: Object Vision

(VisionCourse2009_Object)

Michael P Stryker
Neuroscience 240 Vision Course

Assigned:


Other interesting new papers (not assigned):


One of the most interesting things that we do with vision is to recognize objects. Although each object can give rise to an amazing variety of retinal images, accurate visual object recognition persists under wide variety of conditions.

Physics and optics provides an easy solution to the forward problem of how to compute the image that would be produced by any collection of objects under arbitrary conditions of illumination. This sort of thing is done on an industrial scale all the time by the people at Pixar and Industrial Light and Magic.

Object recognition is the inverse problem, one that is much more difficult. Industrial applications of object recognition work for only the most impoverished images.

We will briefly discuss some of the theories of object recognition, which is the major topic of computer vision.
Our appreciation of the temporal lobe mechanisms involved in object vision originated from neuropsychology—interpretations of seizures and effects of electrical stimulation and lesions on human patients and then on animals.

Charlie Gross, as a new assistant professor, took the neuropsychology seriously and studied inferotemporal (IT) cortex. He found cells that were visually responsive, had fairly large receptive fields that always included the fovea and parts of both sides of the visual field. Most interestingly, some cells in IT cortex appeared to be specialized for faces and a few other complex objects.
Selective Activation of Face Cells in the Inferior Temporal Cortex of a Rhesus Monkey

(A)

Inferior temporal cortex

Record
Selective Activation of Face Cells in the Inferior Temporal Cortex of a Rhesus Monkey
Selective Activation of Face Cells in the Inferior Temporal Cortex of a Rhesus Monkey

(C)
Marr’s overall strategy for a theory of vision

<table>
<thead>
<tr>
<th>Computational Theory</th>
<th>Representation and algorithm</th>
<th>Hardware Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the goal of the computation? Why is it appropriate? What is the logic of the strategy for carrying it out?</td>
<td>How can this computational theory be implemented? What are the representations of input and output, and the algorithm for transformation from input to output?</td>
<td>How can the representation and algorithm be realized physically?</td>
</tr>
</tbody>
</table>
Marr’s Computational Theory of Vision

Grouping Principles

Raw Primal Sketch → Full Primal Sketch → 2.5D-Sketch → 3D Model
Stages in Marr’s Theory of Vision

<table>
<thead>
<tr>
<th>NAME</th>
<th>PURPOSE</th>
<th>PRIMITIVES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image(s)</td>
<td>Represents intensity</td>
<td>Intensity value at each point in the image</td>
</tr>
<tr>
<td>Primal sketch</td>
<td>Makes explicit important information about the two-dimensional image,</td>
<td>Zero crossings, Blobs, Terminations and discontinuities, Groups, Curvilinear organisation, Boundaries</td>
</tr>
<tr>
<td></td>
<td>primarily the intensity changes there and their geometrical distribution and organisation</td>
<td></td>
</tr>
<tr>
<td>2D/4D sketch</td>
<td>Makes explicit the orientation and rough depth of the visible surfaces,</td>
<td>Local surface orientation, Distance from viewer, Discontinuities in depth, Discontinuities in surface orientation</td>
</tr>
<tr>
<td></td>
<td>and contours of discontinuities in these quantities in a viewer-centred co-ordinate frame</td>
<td></td>
</tr>
<tr>
<td>3D model representation</td>
<td>Describes shapes and their organisation in an object centred co-ordinate frame, using a modular hierarchical representation which includes volumetric primitives (i.e. represents volume of space that a shape occupies) as well as surface properties</td>
<td>3-D models arranged hierarchically, each one based on a spatial configuration of a few sticks or axes, to which volumetric or surface shape primitives are attached</td>
</tr>
</tbody>
</table>
Input Picture
(pixel intensities)

Primal Sketch
(lines, points, and terminations)

2-1/2 D Sketch
(surface direction)

3 D Representation
(object centered, modular)
Marr’s Prototype for 3D Representation
Modularity of 3D Representation
Other Ideas About Object Representation

Generalized Cones (Binford, 1971)
Other Ideas About Object Representation

Irving Beiderman has suggested that visual perception encodes objects in a small number of primitive shapes, called *geons*, that can have various spatial relationships. The *geons* form a sort of visual alphabet for the perception of objects. This is much richer than the notion of generalized cones. But singular views are still a problem.

Monster checking for kids under the bed. Soldier and his dog disappearing behind a wall.
Looking at a giraffe through a second story window.
Other Ideas About Object Representation

Steven Zucker et al (1999) has a notion of shapes as defined using a reaction-diffusion process on the borders, that is less *ad hoc* than geons and also richer than generalized cones.
Zucker, 1999. Shock Graphs and Shape Matching
Zucker, 1999. Shock Graphs and Shape Matching Approach does a pretty good job

<table>
<thead>
<tr>
<th>Instance</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.02</td>
<td>2.17</td>
<td>4.48</td>
<td>3.55</td>
<td>2.96</td>
<td>0.21</td>
<td>4.58</td>
<td>14.33</td>
<td>10.01</td>
<td></td>
</tr>
<tr>
<td>2.39</td>
<td>0.10</td>
<td>5.97</td>
<td>15.90</td>
<td>3.98</td>
<td>0.14</td>
<td>26.12</td>
<td>17.28</td>
<td>28.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.89</td>
<td>4.72</td>
<td>2.08</td>
<td>12.24</td>
<td>3.12</td>
<td>2.15</td>
<td>19.73</td>
<td>10.11</td>
<td>12.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.15</td>
<td>6.42</td>
<td>1.19</td>
<td>1.35</td>
<td>5.10</td>
<td>3.38</td>
<td>10.58</td>
<td>11.11</td>
<td>11.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.08</td>
<td>7.72</td>
<td>2.98</td>
<td>1.49</td>
<td>4.26</td>
<td>4.14</td>
<td>26.60</td>
<td>13.54</td>
<td>14.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.77</td>
<td>6.72</td>
<td>5.69</td>
<td>0.36</td>
<td>2.30</td>
<td>5.90</td>
<td>10.58</td>
<td>16.25</td>
<td>19.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.86</td>
<td>8.90</td>
<td>5.94</td>
<td>0.74</td>
<td>1.59</td>
<td>1.10</td>
<td>10.81</td>
<td>10.39</td>
<td>16.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.66</td>
<td>4.23</td>
<td>3.23</td>
<td>6.47</td>
<td>0.62</td>
<td>1.48</td>
<td>11.73</td>
<td>15.38</td>
<td>15.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.18</td>
<td>5.31</td>
<td>1.25</td>
<td>4.64</td>
<td>0.60</td>
<td>1.30</td>
<td>14.18</td>
<td>17.22</td>
<td>9.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.55</td>
<td>0.76</td>
<td>1.32</td>
<td>2.86</td>
<td>1.49</td>
<td>0.11</td>
<td>21.38</td>
<td>15.35</td>
<td>13.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.77</td>
<td>19.46</td>
<td>22.11</td>
<td>13.27</td>
<td>8.21</td>
<td>29.50</td>
<td>0.15</td>
<td>5.12</td>
<td>5.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.73</td>
<td>23.14</td>
<td>31.45</td>
<td>24.41</td>
<td>10.16</td>
<td>31.08</td>
<td>0.18</td>
<td>8.45</td>
<td>7.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.46</td>
<td>19.00</td>
<td>27.40</td>
<td>14.58</td>
<td>24.26</td>
<td>17.10</td>
<td>8.85</td>
<td>7.49</td>
<td>16.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.86</td>
<td>23.07</td>
<td>12.81</td>
<td>11.24</td>
<td>17.48</td>
<td>23.23</td>
<td>6.02</td>
<td>6.92</td>
<td>3.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.73</td>
<td>21.28</td>
<td>14.10</td>
<td>12.46</td>
<td>19.56</td>
<td>19.21</td>
<td>9.53</td>
<td>7.12</td>
<td>5.06</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Tommy Poggio and Shimon Ullman have proposed a radical view, that much of object recognition can be image-based rather than object based, if you can interpolate between the images of various views of an object. That is, they propose that you do not need to represent the object at all. Instead, you can store a few archetypal views and then match an input image directly to some 2-dimensional morphing of the stored views. It works best in limited domains.
Representation is Crucial

Object recognition is thought to compare the representation of a present image to that of a stored object, which must have been computed from its image.
Fig. 1. Speculative flow of information processing leading to face recognition. The left hand column gives the computational stage of processing and the right hand side illustrates examples of the type of information available at each stage (for details see text).
Gross, 1984, How selective are IT cells really?

Figure 1. Location of recording sites. Left, Lateral view of macaque brain. The stippled area indicates location of recording sites on inferior temporal gyrus. Recording sites were also located in the lower bank of the superior temporal sulcus adjacent to the stippled area. Right, Section through inferior temporal cortex illustrating reconstructed electrode penetration. The location of the section is indicated by vertical lines on the brain drawing at left. The stippled area delimits the range of all recording sites within the superior temporal sulcus and on the inferior temporal gyrus. The electrode penetration shown is from experiment 2, in which a high proportion of cells responsive to faces was found. All other penetrations in exeriment 2 were located within 3 mm of the penetration shown. ce, central sulcus; Cd, caudate nucleus; H, hippocampus; ip, intraparietal sulcus; l, lunate sulcus; la, lateral sulcus; oi, inferior occipital sulcus; Put, putamen; tma, anterior middle temporal sulcus; tmp, posterior middle temporal sulcus; ts, superior temporal sulcus.
Figure 3. Responses of a unit selective for patterns with irregular edges. Stimuli were presented as dark shadows on a tangent screen, moving at 1.2°/sec. Stimuli with irregular edges elicited the best responses, while stimuli with straight edges were ineffective. See also the legend to Figure 2.
Gross, 1984, fig 5

Figure 5. Responses of a unit that responded more strongly to hands than to any other stimulus tested. Drawings under each histogram were traced from the stimuli. A, Comparison of responses to hands versus other patterns. Stimuli were randomly interleaved and included the front and back of a model of a human hand, white cutouts with the same shape as the human hand model, a cutout with the shape of a monkey hand, a cutout of a monkey's hand with the space between the fingers eliminated, a scrambled photograph of the model of the human hand (10 rearranged pieces), two "grating-like" hands, a model of a human face, and a plain rectangle. Stimuli were moved at 1.2"/sec from the contralateral into the ipsilateral visual field and were visible within a 15" window centered on the fovea. B, Responses to a stimulus with the shape of a hand, in different orientations. The contralateral visual field is represented on the left of each histogram and the ipsilateral field on the right. The arrows indicate the direction of stimulus motion and the direction of time in the histograms. Other conditions were as in A.
Figure 6
Figure 7. Responses of unit that responded better to profiles of faces than to any other stimulus tested. A. Responses to a monkey face in different degrees of rotation. Best responses were to the profiles of the face. B. Responses to profile of face and to profile with components removed or altered. 1. Profile of face; 2. frontal view of face inserted in circle; 3. outline of profile filled with uniform gray; 4. outline of profile filled with “X”s; 5. outline of profile filled with features from the frontal view; 6. anterior portion of face profile; 7. posterior portion of face profile; 8. one half frontal view of face. Removing or altering any of the components of the profile eliminated the response. In both A and B, stimuli were color slides. For conditions of presentation and figure conventions, see Figure 6.
Figure 8. Responses of unit that responded better to profiles of faces than to any other stimulus tested. A, Responses to an upright human face presented with different rotations about the center of the head. 0°, Frontal view; +90°, profile facing into the contralateral field; +180°, back of head; -90°, profile facing into the ipsilateral field. The unit responded to the face in profile view, indicated as P on the bar graph. B, Unit continued to respond to profile view, even with head on its side. C, Responses to a monkey face in profile view, for comparison. A frontal view of a monkey's face without eyes and a brush were also presented, as controls. See also the legend to Figure 6.
Perrett, selectivity of IT cells for complex objects

Fig. 2. Effects of face identify. (A) Examples of the different views of the faces of the two experiments PS (upper row) and DP (lower row). (B) The mean response to these views (± 1 se, n = 3–10) is illustrated for cell MO55. The cell responds to the face of PS (White bars) under a variety of viewing conditions: when his face was presented upright, horizontal, inverted, at an increased distance, with changed expression or through a green filter. Comparable views of the face of DP (grey bars) produced less response.
Another candidate representation

Gross 1983, Fourier Descriptors

Fig. 1. Examples of FD stimuli varying in frequency (2–64 cycles per perimeter) and amplitude (0.8 and 1.6).

Fig. 2. Responses of IT neurons to FD stimuli. Responses are plotted as percent change in firing rate over the mean spontaneous rate. Each point represents the mean of 10 presentations. (A) Effect of stimulus size and FD amplitude (AMP) on frequency (FRFQ) tuning. Note that the shape of the tuning curve remains similar over changes in stimulus size and amplitude. (B) Mean tonic increase in response with increase in FD amplitude. Stimulation was all 32 cycles per perimeter. (C) Effect of stimulus size on frequency tuning. The shape of the tuning curve remains similar at different stimuli. (D) Effect of location of stimulus on frequency tuning. Stimuli were centered on the fovea (F) or 5 deg into the upper (U), lower (L), contralateral (C), or ipsilateral (I) visual field. Note that the frequencies eliciting the maximal and minimal responses at each location remain the same but the magnitude of the response is greatest at the fovea and least in the ipsilateral field. (E) Effect of contrast on frequency tuning. Light pattern on dark background; D, dark pattern on light background. Note that the best frequency is independent of contrast. For A–E, unless otherwise indicated, the amplitude of the FD stimuli was 1.6, the area was 20 deg², and they were light on a dark background.
Miyashita, 1994, Can IT Receptive Fields Learn Fourier Descriptor Tuning?
Geometric Transform    Parametric Transforms

QuickTime™ and a decompressor are needed to see this picture.

Sakai & Miyashita (1994) *Neuroreport* 5: 829-832
Summary Data: Geometric Transforms Generally Better Than parametric Transforms

Conclusions: peak in parameter space must have been learned by experience

FIG. 4. Neuronal tuning as a local maximum of responses. Schematic single-cell responses $f(x_1^o, \ldots, x_n^o)$ (see text for definition) to various forms are shown as the function of a FD parameter $x_k$, where $(x_1^o, \ldots, x_n^o)$ is the point of an original pattern that elicited the strongest response among the 24 original FDs in the screening set. These 24 patterns provide only discrete points (three are shown), from which the graph of $f(x_1, x_2, \ldots, x_n)$ is estimated; two exemplars of interpolated graphs are shown as thin and thick curved lines. Here we assume that the middle and the points most to the right give strong responses, whereas the point most to the left meets the spontaneous discharge level. The parametric transformation corresponds to a local perturbation around $x_1^o$, as denoted by ↔ symbol. The local change of $f(x_1, x_2, \ldots, x_n)$ in the region of ↔ symbol may be a monotonic decrease as shown by a thin curved line, where half of the time the parametric transformation would cause an increase in neuronal responses. The experimental results indicate that any tested points of original patterns give local maxima of $f(x_1, x_2, \ldots, x_n)$ as shown by a thick curved line, in which $x_1^o$ gives a local maximum. Accordingly, neuronal responses should be optimally tuned to the original patterns.
Fig. 2. *Top:* Primary cells dominated in the prelunate gyrus and posterior IT but were few in anterior IT. An open circle indicates an oriented cell without color selectivity, a star a color-selective cell without orientation selectivity, and a star within a circle an oriented color-selective cell. Cells without either orientation or color selectivity are differentially indicated by their selectivity for the size: a downward open triangle indicates a cell preferring a small size, an upward filled triangle a cell preferring a large size, and a small filled circle a cell without size preference. Broken line indicates border between anterior and posterior IT, which was determined on the basis of the ratio of Primary cells to Elaborate cells recorded in single penetrations, and the dotted line indicates the border between posterior IT and the prelunate gyrus. Inh. cell was inhibited by the stimulus. Cells marked in a cluster were recorded within a single penetration. *Bottom:* Elaborate cells dominated anterior IT, but were few in the posterior regions. Filled square, an identified Elaborate cell; open square, an unidentified Elaborate cell. A face cell. Border between the prelunate gyrus and posterior IT is indicated only by an arrowhead in these maps and the following figures.
The reduction method used by Tanaka lab to find the qualities of complex objects that inferotemporal cortex cells respond to.
## TABLE 1. Classification of cells

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Prelunate Gyrus</th>
<th>Posterior IT</th>
<th>Anterior IT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>109 (70%)</td>
<td>101 (72%)</td>
<td>53 (12%)</td>
</tr>
<tr>
<td>Oriented</td>
<td>59</td>
<td>49</td>
<td>16</td>
</tr>
<tr>
<td>Color</td>
<td>7</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>Oriented color</td>
<td>7</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Nonoriented noncolor</td>
<td>36</td>
<td>30</td>
<td>25</td>
</tr>
<tr>
<td>Texture</td>
<td>8 (5%)</td>
<td>3 (2%)</td>
<td>27 (6%)</td>
</tr>
<tr>
<td>Elaborate</td>
<td>3 (2%)</td>
<td>12 (9%)</td>
<td>193 (45%)</td>
</tr>
<tr>
<td>Identified</td>
<td>1</td>
<td>7</td>
<td>82</td>
</tr>
<tr>
<td>Face</td>
<td>0</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>Unidentified</td>
<td>2</td>
<td>4</td>
<td>81</td>
</tr>
<tr>
<td>Others</td>
<td>3 (2%)</td>
<td>1 (1%)</td>
<td>8 (2%)</td>
</tr>
<tr>
<td>Weak</td>
<td>19 (12%)</td>
<td>14 (10%)</td>
<td>68 (16%)</td>
</tr>
<tr>
<td>Unresponsive</td>
<td>13 (8%)</td>
<td>10 (7%)</td>
<td>80 (19%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>155 (100%)</td>
<td>141 (100%)</td>
<td>429 (100%)</td>
</tr>
</tbody>
</table>

Values are numbers of cells; numbers in parentheses are percentages of column totals. IT, inferotemporal cortex.
Tanaka 1991, Some Primary Cells

**FIG. 4.** Cell 1: Primary cell with orientation selectivity, recorded in posterior IT. Thin broken lines indicate the receptive field. Response was moderately inhibited by lengthening the stimulus further beyond the optimal (C). Cell was activated by a white bar on a gray background but not by a black bar on the gray background (D and E). In this and following figures, white patterns lighter than the background are indicated by drawing the outline (except for in D). Figures above each PSTH indicate the magnitude of the response (see METHODS for the way of calculating the magnitude) normalized to the maximal response to the optimal stimulus. * and + indicate that the response is significantly smaller than the maximum response, with a possible error <1% for * and 5% for +. Cell 2: Primary cell with color selectivity, recorded in posterior IT. Stimuli were paper cuts presented on a gray background. The color (F–I) but not the shape (K and L) of stimulus was crucial. Response was inhibited by increasing the size of stimulus (M).
Tanaka 1991, Some Elaborate Cells

**Cell 1**

- A
- B
- C
- D
- E
- F
- G
- H

![Graphs showing responses to different stimuli for Cell 1](image)

**Cell 2**

- I
- J
- K
- L
- M
- N
- O
- P

![Graphs showing responses to different stimuli for Cell 2](image)

**FIG. 8.** *Cell 1:* Elaborate cell responding to a combination of a disk and a bar projecting from the disk. Selectivity of this cell for the orientation of the stimulus is shown in Fig. 12G. *Cell 2:* Elaborate cell responding to a T shape.

**FIG. 9.** *Cell 1:* Elaborate cell responding to a combination of a dark disk and a light disk above the dark disk. For the orientation selectivity of the response, see Fig. 12C. *Cell 2:* Elaborate cell responding to a combination of a dark ovoid and a light disk within the ovoid.
Tanaka 1991, More Elaborate Cells

**Cell 1**

- **A**
- **B**
- **C**
- **D**
- **E**
- **F**
- **G**
- **H**

**Cell 2**

- **I**
- **J**
- **K**
- **L**
- **M**
- **N**
- **O**
- **P**

*Fig. 10. Cell 1: Elaborate cell that required integration of texture and shape. The critical feature was a vertical grating within a triangle directed to the right. Interval of the grating was 1°, and width of black lines was 0.4°. Cell 2: Elaborate cell that required integration of color and shape. The critical feature was a green star.*

**Cell 1**

- **A**
- **B**
- **C**
- **D**

**Cell 2**

- **G**
- **H**
- **I**
- **J**

*Fig. 15. Cell 1: example of an unidentified Elaborate cell. Bottom of a pineapple (A), a potato (B), and a brown sphere (C) activated the cell, whereas a brown disk was not effective (D). We attached various textures on the brown disk, but none of them activated the cell (E and F). Cell 2: another example of an unidentified Elaborate cell. A dot pattern (G) and a radial pattern like a windmill (H) activated the cell, whereas a pattern composed of concentric rings (I), vertical stripes (J), or horizontal stripes (K) failed to activate the cell. Interspot interval, interval between neighboring rings, and interstripe interval were 4°. Diameter of spots, width of light rings, and width of light bars were 2°. Patterns were projected within a stimulus window 40° in diameter.*
FIG. 12. Tuning curves of 8 Elaborate cells for the orientation of the stimulus. Abscissas are deviations from optimal orientations. Positive value is in the clockwise direction. Ordinates are magnitudes of responses normalized by the maximum response to the optimal orientation. Thin horizontal lines are drawn at 0.5 to visualize the steepness of tuning. A, C, E, and G are for the same cells as cell 2 of Fig. 7, cell 1 of Fig. 9, cell 1 of Fig. 11, and cell 1 of Fig. 8, respectively. H is the most flat tuning out of 28 cells tested quantitatively. * indicates that the response is significantly smaller than the maximum response with a possible error <1%. 
FIG. 16. Distribution of subgroups of Elaborate cells. Two types of Elaborate cells are subdivided by their critical features into shape only, shape + texture, shape + color, texture + color, shape + texture + color, and face. There was no clear separation between cells requiring shape only and cells requiring shape + something, although the face cells clustered at the most posterior part of anterior II (around anterior S). Bottom: identified Elaborate cells are subdivided into single-segment cells, the critical features of which can be assumed to be comprised of a single segment, multisegment cells, the critical features of which can be divided into >2 segments, and face cells. Filled and open triangles indicate cells that were classified, with uncertainty, into single-segment and multisegment cells. There was no clear separation between single-segment and multisegment cells.
Fig. 17. Five penetrations in which cells showed related cortical features. Positions of recording of cells are indicated by short horizontal lines on the electrode track (vertical line). Top is the surface of the cortex. Elab., identified Elaborate cells; unidentified, unidentified Elaborate cells. Because we did not determine the position of the transition from the gray matter to the white matter, we cannot determine the layer localization of recording positions. Position of the entrance of these penetrations are indicated in Fig. 1 by 17.1-17.5.
FIG. 18. Averaged square root of the area of the receptive field. Areal extent of the receptive fields was measured by assuming they were elliptic in shape, and the square root of the areal extent was averaged over cells recorded in a single penetration. Circles indicate the averaged value by their diameter and the position of the penetration by the position of their center. Figures attached to circles indicate the number of cells over which the averaging was made.
Tanaka 1992, Inferotemporal Cortex Columns?
Columnar organization of the IT cortex. Columns of cells selective for a range of similar stimuli (for example, the coloured bars) are interspersed among columns unresponsive to those stimuli but selective for other complex visual qualities.
Tanaka Comparison of Selectivity with Optical Imaging and Single Units
Tanaka Similar Poses are Represented Near One Another
Logothetis, Amoebas
Logothetis, Paper Clips

(a)

-72°  -60°  -48°  -36°

-24°  -12°  0°  12°

24°  36°  48°  60°

72°  84°  96°  108°

142 spikes sec⁻¹

600 msec

Wire 526, cell 202
Miyashita, 1991
Memory for Visual Paired Associates

Fig. 3. Visual paired associates. Twelve pairs of pictures for stimuli in the PA task. When one member of each pair is shown, trained monkeys can recall and select the other member of the paired associates.

Fig. 4. Responses of a pair-coding neuron. Typical responses of pair-coding neuron that exhibited form-selective activity during the cue period. (A) Trials for cue 6' that elicited the strongest cue response. (B) Trials for cue 6 that elicited the second strongest cue response.
Observation: Only a few out of 100 slides of arbitrary stimuli give good responses, and a different few for each cell.

No geometrical feature is in common between the stimuli that give the best responses for each cell.
Serial position in the order the stimuli were shown during training explains the stimuli that give the best responses for each cell: a temporal association.
Shape recognition seems to be common to all humans rather than cultural.


**ABSTRACT**—Many of the phenomena underlying shape recognition can be derived from the greater sensitivity to nonaccidental properties of an image (e.g., whether a contour is straight or curved), which are invariant to orientation in depth, than to the metric properties of an image (e.g., a contour’s degree of curvature), which can vary with orientation. What enables this sensitivity? One explanation is that it derives from people’s immersion in a manufactured world in which simple, regular shapes distinguished by nonaccidental properties abound (e.g., a can, a brick), and toddlers are encouraged to play with toy shape sorters. This report provides evidence against this explanation. The Himba, a seminomadic people living in a remote region of northwestern Namibia where there is little exposure to regular, simple artifacts, were virtually identical to Western observers in their greater sensitivity to nonaccidental vs. metric properties.
Fusiform face area represents images of faces rather than individual people.

Viewing a sequence of faces of two different people results in a greater Blood Oxygenation Level Dependent (BOLD) response in FFA compared to a sequence of identical faces. Changes in identity, however, necessarily involve changes in the image. Is the release from adaptation a result of a change in face identity, per se, or could it be an effect that would arise from any change in the image of a face? Subjects viewed a sequence of two faces that could be of the same or different person, and in the same or different orientation in depth. Critically, the physical similarity of view changes of the same person was scaled, by Gabor-jet differences, to be equivalent to that produced by an identity change. Both person and orientation changes produced equivalent releases from adaptation in FFA (relative to identical faces) suggesting that FFA is sensitive to the physical similarity of faces rather than to the individuals depicted in the images.
Reviews


Faces and forms: Gross and Perrett


Faces and forms revisited: Rolls and Logothetis


A Hubel and Wiesel Job on the Inferotemporal Cortex: Tanaka


Memory is the Central Problem of Recognition: Miyashita


