Biederman’s Recognition-by-Components (RBC) Theory of Image Recognition

Philipp E. Koralus

Cos 598/Psy 594
Prof. Fei-Fei Li
Spring 2008

Very rough overview of RBC

- Visual input segmented into components

- Components are recognized as falling into different categories of geons

- Recognition memory coded in terms of geons and how they are combined

Very rough overview of RBC

- It’s primarily a theory of “first shot” recognition of novel and familiar objects

- Predicts recognition robustness insofar as components still recognizable

Key questions:

1. What is the relevant visual input to first shot recognition?
2. How is the visual input segmented into parts?
3. What are the parts and how are they recognized?
4. How do the parts combine?

What’s the relevant visual input?

Line layouts!

- We can recognize objects rapidly and normally even if reduced to line drawings.

- Kourtzi and Kanwisher (2000). Adaptation of fMRI BOLD signal in LOC maintained if image changes from gray-level photo to line drawing of the same object.

- Biederman and Ju (1986). Recognition RTs in naming task virtually the same using line drawings and full color photographs.

- Images presented by color photography were 11ms faster than the corresponding drawing but had a 3.9% higher error rate.

Diagnostic colors?

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

- Objects with a diagnostic color did not enjoy any advantage when they were displayed as color slides compared with their line drawing versions.

- *Not even in a “tell me if you see a banana among the following slides” task!*

Key questions

1. What is the relevant visual input to first shot recognition?

Key questions

1. What is the relevant visual input to first shot recognition?

→Line layouts are sufficient

How is the visual input segmented into parts?

- We tend to segment at regions of sharp concavity
- People tend to agree on what the natural components of an object are

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

Qui
TIFF (Uncomp
are neede

Concavities arise where convex volumes are joined (Transversality Principle)

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

- Good continuation is a statistically powerful cue to determining object boundaries in natural scenes. Elder and Goldberg (2002)
- Sharp convexities break good continuation.

Key questions

1. What is the relevant visual input to first shot recognition?

   → **Line layouts are sufficient**

2. How is the visual input segmented in parts?

   → **Segment parts at sharp convexities**

What are the parts and how are they recognized?

Constraints

- We are bad at making absolute judgments in length, tilt or curvature (Beck, Prazdny, and Rosenfeld 1983; Fildes & Triggs, 1985; Garner, 1962; Miller 1956; Virsu, 1971).

- Our memory for shape shows tendency for regularization (Woodworth 1938).

Errors in reproduction:

Slight deviations from symmetrical or regular figures.

Alternatively, irregularities are accentuated as regular subparts

- We tend to see simple shapes even if the metric details rule them out.

QuickTime™ a
TIFF (LZW) decom
are needed to see this

and a
npressor
is picture.

→ Use non-accidental features instead of metric features

Then define parts in terms of those non-accidental features

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

…Step back

- What evidence is there that we are especially sensitive to non-accidental features?

Collinearity

We rarely interpret curved lines as straight or straight lines as curved!

Collinearity distinguished from Curvature

- Search for a letter composed of straight segments such as “Z” is faster if distractor letters are curved “C, Q, G, O” then when distractors are also composed of straight segments” N, W, V, M” (Neisser 1963).

Symmetry and Parallelism

- Ames Room

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

- Symmetrical shapes can be quickly distinguished from asymmetrical ones (Pomeranz 1978).

- We have a strong preference to see Y-junctions as orthogonal, making for blocks with parallel edges(Perkins(1983); Perkins and Deregowski (1982)).

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Cotermination

- If lines co-terminate, we visually interpret them as corresponding to one edge, even if this is geometrically impossible

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

… even if the lines could be the projection of a possible object

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

- We are especially good at finding junctions diagnostic of depth

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Enns and Rensink (1991)

- Upshot: Certain non-accidental features are indeed special. Some plausibility that we can rapidly identify geon features.

Geons

- The values for the generalized cone parameters can be directly detected as contrastive differences in nonaccidental properties:

  straight v. curved, symmetrical v. asymmetrical,
  parallel v. non-parallel.

- Edges and curvature of axis can be determined by collinearity and curvilinearity.

How should we think of Geons?

Bare feature n-tuples?

$<\{S,C\}_{Edge}, \{++,+,-\}_{Symmetry}, \{++,-,--\}_{Size}, \{+,-\}_{Axis}>$

$<\{S,C\}_{\text{Edge}}, \{++,+,-\}_{\text{Symmetry}}, \{++,-,--\}_{\text{Size}}, \{+,-\}_{\text{Axis}}>$

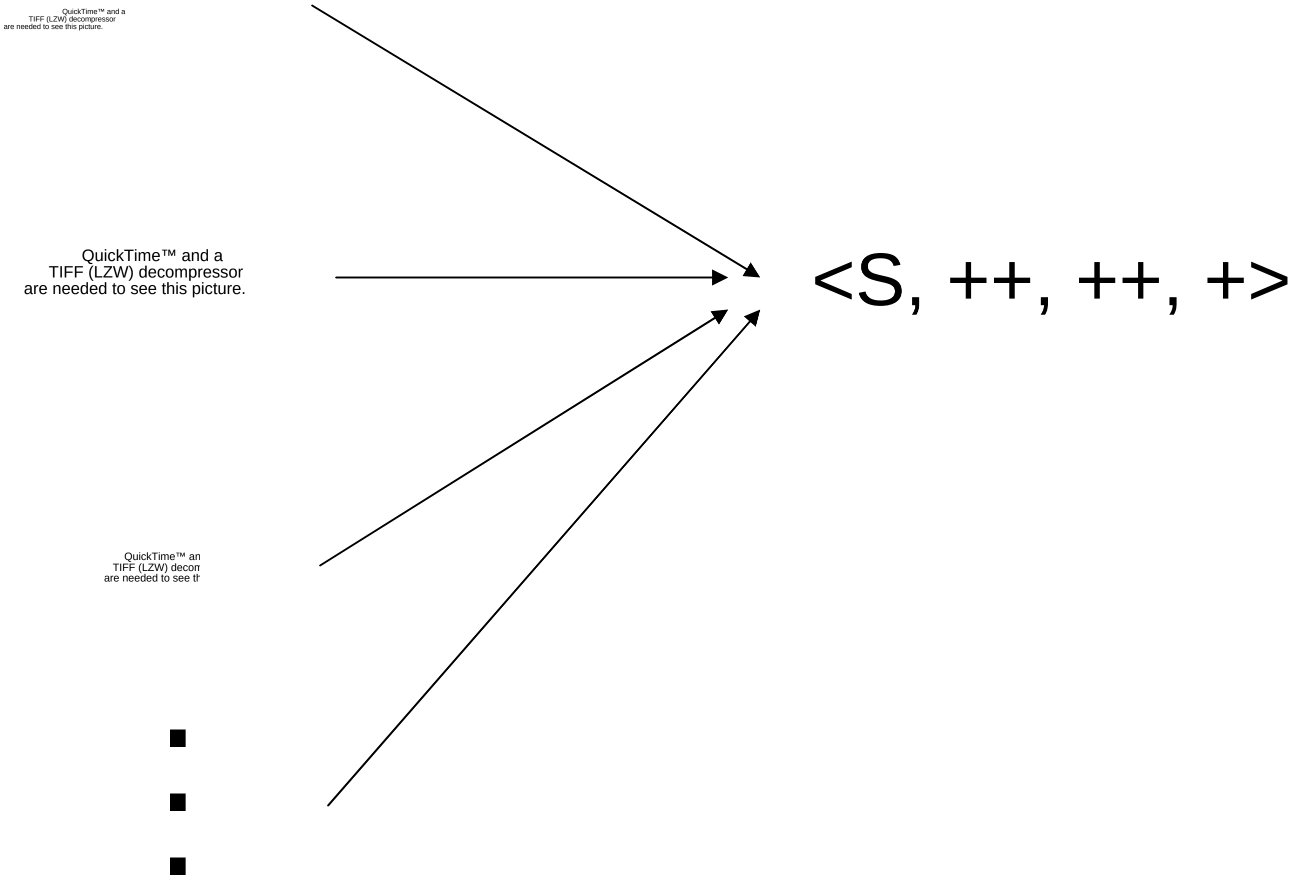
So, the simple block geon *just is*:

<S, ++, ++, +>

- What does <S, ++, ++, +> “look like”?

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

?



QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ an
TIFF (LZW) decom
are needed to see th

⋮

<S, ++, ++, +>

- Any “picture” that fits the parameters is an equally plausible corresponding percept, without further constraints.

- Either say that the percept attaches to the low level features

- Or say that there are further constraints that make geons “feel” more specific than they are.

<S, ++, ++, +, QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture. >

How many Geons are there?

- $<\{S,C\}_{\text{Edge}}, \{++,+,-\}_{\text{Symmetry}}, \{++,-,--\}_{\text{Size}}, \{+,-\}_{\text{Axis}}>$

$$2 \quad * \quad 3 \quad * \quad 3 \quad * \quad 2 \quad = 36$$

- Some distinctions are lost

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Is this a problem?

- Asymmetrical patterns require more time for identification than symmetrical patterns (Checkosky & Whitlock 1973, Pomeranz 1978)

Key questions

1. What is the relevant visual input to first shot recognition?

   → **Line layouts are sufficient**

2. How is the visual input segmented in parts?

   → **Segment parts at sharp convexities**

3. What are the parts?

   → **n-tuples of contrastive non-accidental feature parameters that are rapidly recognized (Geons).**

How do geons combine?

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

- We need to distinguish how different geons are combined (Top/down/side)

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

- We also need to distinguish relative geon sizes!!

- By contrast:

  Aspect ration may matter a little bit but not much.

  Recognition speed is unaffected by variation in aspect ratio across different views of the same object (Bartlett 1976).

Possible further emendations:

- Component terminations
- Planar regions for diagnostic surface features.

How much can we represent with geons in combination?

Homework:

- Can you find two readily distinguishable objects that share all geons and those constraints on combination listed so far?

Key questions

1. What is the relevant visual input to primal first shot recognition?

   → **Line layouts are sufficient**

2. How is the visual input segmented in parts?

   → **Segment parts at sharp convexities**

3. What are the parts?

   → **n-tuples of contrastive non-accidental feature parameters that are rapidly recognized (Geons).**

4. How do Geons combine?

   → **Recognition memory must encode a limited number of relational properties between Geons.**

Empirical evidence?

Predictions

- *Some degree* of perspective invariance in recognition performance.

- Much, much worse performance if we can’t extract geons.

- Recognition should be possible even if only some geons are recoverable

Experiment 1

Biederman, Ju, and Clapper (1985).

- Test whether a few geons are enough for rapid object identification

- Pictures categorized by complexity and by how many components visible

- 100ms presentations followed by mask

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Summary Experiment 1

- Error Rates: Increasing numbers of components resulted in better performance. Complete objects named without error.

- RTs: Slight tendency for more complex objects to have shorter RTs when complete.

Experiment 2

Biederman and Blickle (1985)

Recognition performance with degraded contours

RBC:

Parsing achieved at regions of concavity. Collinearity allows filling in.

→So contour deletion at regions of concavity should be particularly hard, if not impossible to recover from.

→Equivalent deletion in the middle should prove to be less disruptive.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

- Object naming task
- Subjects viewed 35 objects
- Practice with various degraded line dawings

- Three Groups with different preparation:
  1. View 3 sec slide of intact objects. Asked to name them.
  2. Familiarization with only names of objects.
  3. No familiarization with the objects or names.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Summary Experiment 2

- Median error rates for unrecoverable ones was 100%. Most subjects said “don’t know”.

- Most unrecoverable objects show no gain with 5s of exposure.

- Some effect from providing intact versions. No effect from names.

- Error rates at 100ms for recoverable objects averaged to 65%.

Experiment 3

Biederman and Blickle (1985)

Like experiment 2, except with varying degrees of degradation.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Experiment 3 summary

- At most exposure durations and at most amounts of removal, removal of vertices results in more errors and slower RTs than mid-segment removal.

- Improvement with longer exposure.

Experiment 4

Biederman, Beiring, Ju, & Blickle (1985)

- Naming speed and accuracy of six and nine component objects with deleted contours at mid-segment and deletion of contours of particular geon components.

RBC:

- Missing whole components affect matching stage.

- Mid-segment deletion affects determination of components.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Summary experiment 4

- At brief exposures (65ms), both error and RT performance with partial objects better than with objects where same amount of contour was removed midsection.
  - **Brief exposures affect determination of geon stage**

- At longer exposures (200ms), RTs reversed, mid-segment deletion now faster than partial objects.
  - **At longer exposures, determination of geons can recover from deletion and extra geons help speed up recognition**

Further observations

- For recoverable figures with geon cues intact one can restore the view of a three-dimensional object through a masking template.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

- This does not seem to work with unrecoverable objects that have their geon cues deleted

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Further observations

- There seems to be better recognition of common objects from some perspectives rather than others, when the perspectives differ strongly on geons. (Palmer, Rosch, & Chase 1981)

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Further observations

- Different exemplars of certain categories (cars) are less likely to have common components…

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

East German Specimen

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

North American Specimen

….so there should be less priming between different exemplars in a rapid naming task than in a condition where the same objects are presented with identical or similar orientation, but more than for a control with completely unrelated objects.

- Confirmed. Bartram (1974)

Further observations

RBC predicts that children should be predisposed to employ different labels for objects with different geon descriptions.

→Children seem to distinguish tigers and male lions from cats before they distinguish them from female lions

Further observations

- RBC predicts that we can make sense of novel objects and compare them component wise with objects we know

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

Biederman and Bar (1999)

Two theories of recognizing objects at arbitrary rotations in depth:

- Generalization from templates specified by metric properties (MPs)
- Recognition by matching to descriptions in terms of non-accidental properties (NAPs)
  - →The RBC theory

- Task: judge sequential pair of images of novel objects for sameness or difference under rotation

- Two manipulations: Change objects by NAPs or MPs

- NAP and MP differences scaled to be equally detectable when objects were at the same orientation.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

- Further condition: same task without rotation.

Summary

- Detection of MP differences seems to come at a higher cost than detection of NAP differences.

Vogels et al. 2001

- There are cells in macaque IT that respond more strongly to a change in NAP than to a change in MP.

background

- Tanaka 1993: There are cells in macaque IT that are shape selective for shapes of moderate complexity.
- Esteky and Tanaka (1998): Metric variation has minimal effect on IT cell activity.

- Single cell recordings from macaque IT. Objects calibrated in terms of metric shape properties and non-accidental properties.

- Object changes equalized for difference in terms of pixel differences and wavelet measure differences.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Kayaert et al. (2003)

- Effect of greater modulation from NAP compared to MP changes. MP image changes (by pixel energy measure) had to be approx 50% larger than NAP changes to produce the same degree of modulation.
- Amount of modulation produced by depth rotation equivalent to modulation produced by non-rotated MP changes when the two conditions were equated according to the magnitude of image change.

Kayaert, Biederman, Op De Beek, Vogels (2005)

- Macaques viewed 2d regular and irregular shapes while TE neurons were recorded

Difference in regular shape (circle v. square) produced markedly more absolute modulation than change in highly irregular shape, where two types of change were matched with respect to pixel similarity.

Kayaert, Biederman, Vogels (2005)

- Population code of IT neurons represents independent dimensions of generalized cylinders.

- Cells found that respond to highly curved axis of shape independently of its taper, aspect ration, or curvature of its sides.

- Cells tend to be tuned to one end of a dimension or the other, with very few cells preferring intermediate values.

- Some cells respond predominantly to a highly curved axis while others respond to straight axis with the firing declining as the axis curvature is changed away from the maximally preferred value.

Hayworth and Biederman (2005)

- Subjects presented with two-frame “flip movies” in fMRI. In the movies, one part of a two-geon object cycled between two different shapes so that a cylinder on top of a brick could change into a pyramid and back again for several cycles.

- 24-sec block of trials consisted of three such geon change movies with shape change varying between movies.

- In another block, geon would retain shape but vary in relations, (top or side).

- Magnitude of image changes equated with respect to pixel energies.

- According to RBC, there is a difference between recognizing components and combining them.

- MT was equally affected b the different kinds of changes. For every subject, for every voxel in LOC (Lateral Occipital complex), greater activity associated with change in part shape compared to change in relations.

- A region in intraparietal sulcus showed markedly greater activity to the relations condition than the shape part condition.

Conclusion

- Geon based recognition by components view is theoretically well motivated
- Seems to sit well with results from psychophysics
- There is evidence for relevant neural correlates
- Clearly, as Biederman and Logothetis both note, RBC is not all there is to visual recognition.

Replies on behalf of Biederman

Paper clips

Logothetis et al. 1995.
Evidence for IT neurons coding “blurred templates” not geometric features

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

- How does Logothetis’ result bear on RBC?

"We can look at the zig-zag horizontal brace as a texture region or zoom in and interpret it as a series of connected blocks" (Biederman)

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Regarding objects like cork-screws: “those regions are represented through the statistical processing that characterizes their texture.” not in terms of volumetric components unless we “zoom in”. (Biederman)

- --Maybe those object just aren’t recognized using geon differences

- Irrelevant as a test of RBC

Amoebas

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Again, those differences are not geon based

Could be due to episodic memory.

RBC is primarily about first shot recognition.

Who knows what the brain does after 700,000 trials with one object type in isolation.