Objects in Context

Philipp Koralus
Dan O’Shea
Spring 2008

COS/PSY 594

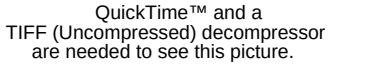
Prof. Fei-Fei Li

Biederman, Mezzanotte, and Rabinowitz (1982)

"Scene Perception: Detecting and Judging Objects Undergoing Relational Violations."

Objects in visual context

- When we look at a scene, we visually comprehend various relations between objects.


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

Claim:

- There is a psychologically real difference between a display of unrelated objects and a well-formed scene.

Linguistic analogy

Colorless green ideas sleep furiously.

He smiled the baby.

Linguistic analogy

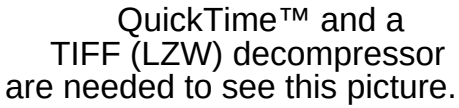
Colorless green ideas sleep furiously.

--> Semantic violation

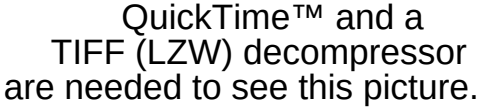
He smiled the baby.

--> Syntactic violation

Do you find this in vision?


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

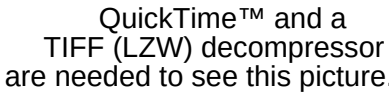
Five relations between objects and scenes



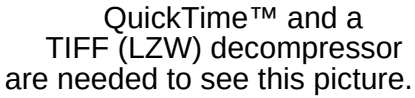
Position Violation

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

Interposition Violation



Probability, Size, and Support Violation



Biederman: Two types of relations

Relations based on object identity

Relations recoverable from low-level features

Biederman: Two types of relations

Relations based on object identity

- -Probability
- -Size
- -Position

Relations recoverable from low-level features

- -Interposition
- -Support

Claim

- If interposition, support, probability, position, and size are all normal, then the scene will not look anomalous as a scene.

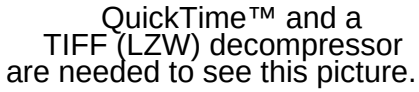
- It will be a “well-formed scene”.

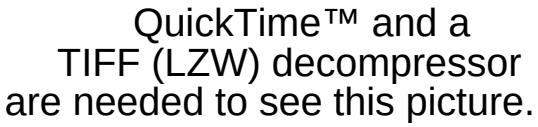
“The bottom-up theory”

Guzman 1968, Waltz 1972, Sugihara 1978

Bottom up line interpretation

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

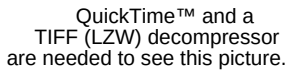

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.


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

Catalogue of natural possibilities

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

Strategy: Start at the outside boundary contours and propagate constraints from catalogue.



- Generally, this approach is sensitive to differences due to support and interposition relations.

Support


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

Interposition

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

Generic Bottom-up model

1. Identify contours.
2. Give the contours a 3D interpretation.
3. Match 3D interpretation with object categories.
4. Determine semantic relations between object categories.

Predictions of generic bottom-up model

- Both interposition and support relations would be determined at an earlier stage of processing than, say, probability.

Predictions of generic bottom-up model

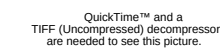
- Both interposition and support relations would be determined faster, at an earlier stage of processing than probability.

- Interposition and support violations should create more recognition problems since they introduce error early in the processing chain.

Biederman et al.(1982):

Is it really true that relations that don’t depend on object identity are processed faster?

Is it really true that relations that don’t depend on object identity have a greater impact on recognition?



Further issues

- Are objects identified prior to determining how they interact?

- If so, then the probability relation which can be gotten just from the identities of a few objects should be computed before the position relation.

- If physical parsing comes before object identification, physical violations should be detected faster.

- If objects are identified before physical parsing, probability violation should be detected faster.

- Also, if violation costs on recognition are due to interference with creating an appropriate frame, the detectability of “innocent bystander” objects should be reduced.

- *Innocent bystander*: target object in a scene containing a violation that does not participate in the violation

- *Frame*: Structure that integrates information about the identity of objects that are most likely to co-occur, together with their relationships.

Look ahead:

- None of the bottom-up expectations are confirmed.

Look ahead:

- None of the bottom-up expectations are confirmed.

- “Physical relations” no more fundamental than “semantic relations”.

Experiment 1

- Determine if the presence of a violation in the relation between an object and its setting affects that object’s perceptibility.

Bottom-up model:

Size, probability, and position only derived following identification, so there shouldn’t be violation effects

Support and interposition should create violation effects.

Stimuli

- 247 scenes
- 42 objects (man, book, car, etc.)
- 17 backgrounds (kitchen, downtown, etc.)
- Each object appears in normal condition at least once.
- Each object appears in one to five slides with a violation.

Stimuli

- 10 violation conditions
  - 5 single violations
  - 4 double violation
  - 1 triple violation

- Efforts were made to equalize camouflage, distance from fixation, and target size.

Stimuli

- Two judges rated degree of masking of target's critical features by the adjacent contours (camouflage)
  - Ratings on 10 point scale.
    - Mean camouflage rating 4.27
    - Inter-rater correlation .793
    - Correlation between camouflage and size = -0.146

Camouflage rating for man = 8.0

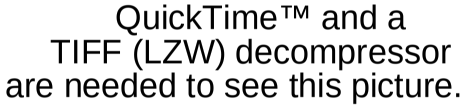
QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Camouflage rating for fire hydrant = 5.5

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

Camouflage rating for couch = 2.0
Camouflage rating for fire hydrant = 2.0

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

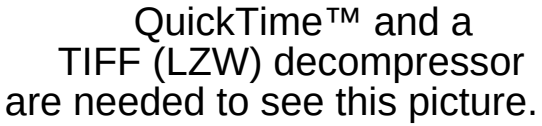

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Stimuli

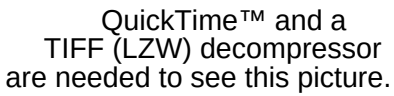
- Two judges rated degree to which targets violated various relations.
  - Scale from 1 to 10.
    - Inter-rater correlation .873 for size, .928 for support, .950 for interposition and probability and .970 for position

Stimuli

- Two judges rated degree to which targets violated various relations.
- Scenes of mean degree of violation 8.9 were selected.
- For multiple violations, largest violation and violation sums were taken.


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

Experiment 1 task design



- On half of the trials, cue dot corresponds to target object center.
- In “cue on target” trials, subject says “yes” into voicekey.
- In “cue off target” trials, subject says “no” into voicekey.

- 96 subjects.
- 247 slides, in 12 blocks of 18-22 scenes.
- Objects, violation conditions, and backgrounds were homogenously spread across blocks.
- All scenes had the same mean serial position (123.5).

- For each violation slide, there were two possible cues:
  - One: cue indicating the violating object.
  - Two: cue indicating a different object.

- For each cue, two target-object labels:
  - One: naming the cued object (for “Yes” responses)
  - Two: naming a different object (for “No” responses)

- The condition in which the target object is normal for the background but there is a violation involving a different object tests for innocent-bystander effects.

- For each base-slide, there were two conditions:
  - Cue matching label (“Yes” condition)
  - Cue not matching label (“No” condition)

Experiment 1: Results

- Block sequence assignment negligible.
- Overall decrease in RT and error.
- Overall error 31.2%.
- Miss rate higher than false alarm rate.

- Miss rate higher than false alarm rate.
  - “No” when target cued in 43.2% of trials.
  - “Yes” when target not cued in 19.2% of trials.

- Miss rate higher than false alarm rate.
  - “No” when target cued in 43.2% of trials.
  - “Yes” when target not cued in 19.2% of trials.

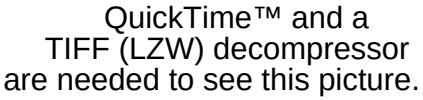
- Mean correct RT = 999msec.

Violation costs:

- A target which violated a relation was more likely to be missed than the same target in base position.

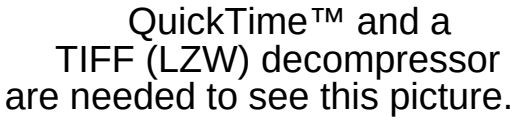
Violation costs:

- Average miss rate (violation condition) = 45%
- Average miss rate (base condition) = 24.9%
- Increased violations produce higher miss rates.
- False alarms higher for violation condition.

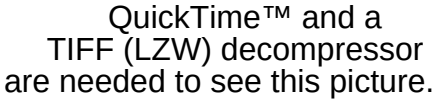

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

- No significant cost of violations on detection of other objects not undergoing violation.

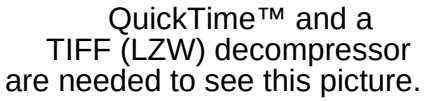
- This suggests the problem isn’t a general effect on elicitation of a frame.


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

- False alarms go down if target improbable.


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

- The further an object is from fixation, the smaller its size, or the greater its camouflage, the more likely it is to be missed.


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

- Violation costs are apparent for all conditions except the interposition condition.

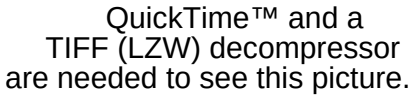
- Violation costs are incurred even when the target is at cue location.

- Next: Regress out various parameters:
  - Distance from fixation
  - Size
  - Camouflage

    - Regression analysis with those variables as predictors and residuals as corrected scores.

- For each object, calculate difference between violation and base residual miss rates to remove effects due to object drawings.

Results after regression-cleanup



- The results remain after regressing out physical parameters.

- 12.6% greater miss rate for violation condition (significant $t(205)=7.13$, $p < .001$)

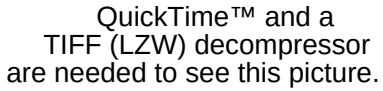
- The picture stays more or less the same if you include false alarms.

- Calculate d’ from residual difference scores from misses and false alarms (adding base condition data)

- Mean d’ for support and interposition violations were 1.48
- Mean d’ for the three semantic violations was .98
- Violation of probability less disruptive (1.42) than violation of position (.98)
- Low detectibility of objects with size violation (.61)

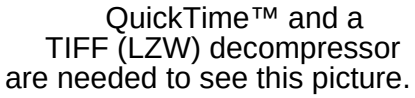
A d’ of 1.0 would correspond to about 69%

correct for both “Yes” and “No” trials.

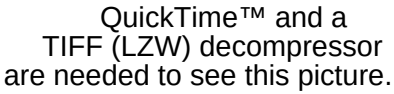

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

- Reaction times

It takes 31msec longer to detect an object in violation than an object in base condition


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

- As before, regress out physical parameters.


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

- Reaction times get you a similar ordering as the error data.

- Precludes general explanation in terms of speed-for-accuracy tradeoff for miss-rates.

- Violations of the physical relations did not yield larger violation costs on RTs than violations of semantic relations.

- Increasing the number of violations increases violation cost. Why?
  - More violations might make it more likely that a misleading relation is detected before the object is identified.
  - Misleading relations could decrease plausibility of the target.

Experiment 2

- Determine relative detection speeds of the violations themselves.

- Acceptability judgment task: is a given target undergoing a violation?

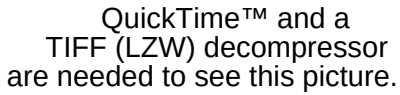
- Bottom up model:
  - Interposition and support should be detected more rapidly than violations in position or size.
  - Adding semantic violation to physical violation should not make combined violation detectable faster.

- Assumption: If position violation can be accurately judged from a single fixation, then scene comprehension is present.

Method

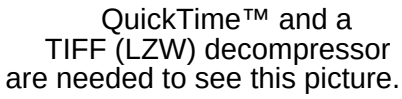
- Very similar to object detection task.
- Positional cue *precedes* the object and always corresponds to target named.
- Finger key for “normal” and “violation” responses.

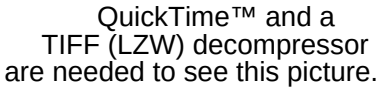
- 48 subjects, 246 scenes used in analysis.
- 150msec presentations.


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

Experiment 2: Results

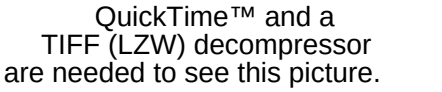
- Subjects were able to detect violations with a single glance, except for the interposition violation.
- Overall hit-rate 88%
- False alarm rate 10.3%
- Correct RT = 851msec.

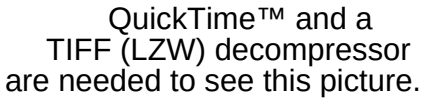

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.


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

- Camouflage only physical variable that significantly correlated with RTs and errors.

- r’s = .318 and .319 respectively. $P < .001$, $df = 244$


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


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

- Both original and corrected data show no evidence for a consistent advantage in the accessibility of the support and interposition relations relative to size position and probability relations.

- Interposition had a much higher miss rate than other violations.

- Probability relations were not more readily detected than position violations.

- On the bottom-up model, support and interposition are processed before probability, position and size.
- Hence, no gain in speed should be there when a violation of a semantic relation is added to the violation of support.
- But *there is* a gain of detection speed.

General discussion

- Objects undergoing violation are generally harder to identify than objects in a base condition

- Violations of semantic relations are at least as disruptive as violations of support and interposition.

- Moreover, the addition of a violation of a semantic relation to a violation of support tended to result in a greater violation cost in object detection and better violation detection than just the support violation by itself.

- The results are not compatible with a model that says that physical parsing precedes the interpretation of semantic relations.

- Semantic relations are accessed at least as quickly as relations defined by physical parameters.

- Also note lack of violation effect for interposition.

- Interestingly, violations of probability were not more disruptive than violations of position and size.

- This suggests that an object's semantic relations are processed simultaneously with its own identification.

- “Instead of a 3D parse being the initial step, the pattern recognition of the contours and the access to semantic relations appear to be the primary stages. In this respect, the detection of violations of support may simply be a special case of the detection of violations of position in real world scenes.”

- *Alternative explanation 1*: Guess “No” when subjects can’t detect target but detect violation. This would reduce false alarm rate.

- But false alarm rates were slightly higher for objects undergoing violation.

- *Alternative explanation 2*: Violations disrupt creation of frame for the scene.

- But this would have also affected innocent bystanders.

- *Alternative explanation 3*: Maybe it's a global plausibility measure that's at work instead of those individual violation conditions.

- But you still need an account of how global plausibility is determined, and what features influence it.

The End.