Role of (Rapid) Top Down Priors on Object Search

Aude Oliva
BCS-MIT
oliva@mit.edu
cvcl.mit.edu
Collaborators

Antonio Torralba
Research Scientist, CSAIL

Barbara Hidalgo-Sotelo,
Graduate Student
BCS

Jeremy Wolfe
Professor
Harvard Medical School

Naomi Kenner
Former research assistant
Graduate Student

Articles and posters related to the present talk are available on the web site http://cvcl.mit.edu
Computational Visual Cognition Laboratory @ MIT
Top Down Priors on Object Search

1- First Study: Effects of scene priors and task constraints on object search

2- Second Study: How learning of scene priors modulates object search

3- Third Study: Automatic activation of scene priors in memory and visual search
Top-down Contextual Priors

- Contextual priors refer to the association between a scene context and an object

Look for a pedestrian

Context prior region

Observers two first fixations

How to define scene context?  
Object-centered view

- Context can be framed as the relationship between objects (Bar, 2004; Biederman 1990; Davenport & Potter, 2004; Henderson et al., 1987).

- Observers may perceive a number of diagnostic objects (e.g., a bed) and infer other objects (e.g., a pillow, Oliva et al., 2006; Greene & Oliva, in preparation; Fei Fei et al., 2006).
How to define scene context?
Scene-centered view

• Scene context is built in a holistic fashion, without recognizing individual objects or regions

• Low level features (e.g. distribution of scales, orientations, texture and colored regions) can predict the basic-level category of the scene.

• Some global properties are more diagnostic of scene categories than others

How to define scene context?
Scene-centered view

Scenes from the same category share similar global properties

Scene Gist Representation Framework

Scene-centered representation
Global Properties

Object-centered representation

From zero to “gist” in 200 msec
Global Feature Representation

$V = \{\text{energy at each orientation and scale}\} = 6 \times 4$ dimensions

80 features

$V_G$

Vector of Global features

Oliva & Torralba (2001)
Global Feature Representation

A global feature represents a *spatial* pattern of orientations and scales across the whole image (or an image region).

Global Features are represented here as the Principal Components of the output magnitudes of a bank of multiscale oriented filters, at various locations.

Oliva & Torralba (in press, 2006). Progress in Brain Research
Global Feature Representation

The “sketch” represents noise images that are coerced to have the same color blobs and the same global features (N=100) as the original image.

Global Feature Representation

- Independent of image clutter and complexity
- No segmentation or grouping mechanisms
- Complementary to object and region analysis
Study 1
Contextual Guidance of Eye Movements & Attention in Real-World scenes

Top-down constraints on Object Search

1- Eye movements are driven by task constraints and scene priors at the beginning of the search

2- Saliency effect is modulated by task constraints and scene priors

“Scene Categorical” Priors in object search

- How far can we go in predicting fixations using only scene priors?
- No representation of object and object template
- A useful strategy when the objects are very small, embedded in clutter or camouflaged.

Look for a cup
Contextual Guidance Model

Scene “gist” Representation

Scene Priors: Top Down, Task dependent module

Saliency map(s)
Koch, Itti & al.

Bottom-up Saliency

• “Saliency” is here defined as how unlikely it is to find a set of local measurements within the image (Koch & Ullman, 85; Itti et al, 98; Rosenholtz, 99; Torralba, 03).

• Saliency computation is independent of the task and target type (e.g. looking for a cup, or looking for a person)
Statistical distribution of object saliency

Torralba, Oliva et al. 2006
The layered structure of scenes

Assuming a human observer standing on the ground

In a display with multiple targets present, the location of one target constraints the ‘y’ coordinate of the remaining targets, but not the ‘x’ coordinate.

Torralba, Oliva et al. 2006
The layered structure of object association

In a display with multiple targets present, the location of one target constraints the ‘y’ coordinate of the remaining targets, but not the ‘x’ coordinate.

Torralba, Oliva et al. 2006
Guidance of attention by context requires a learning stage in which the system learns what are the typical locations of objects in scene.

We trained the model to predict the location of people in the scene.

We used a database of scenes that have been hand-labeled.

The goal is to learn the joint distribution between global image features ($V_c$) and the location of the target.

Hundred of images for which we know the location of People (mug, painting)

Torralba (2003a, 2003b), Torralba, Oliva et al. 2006
Contextual Prototypes

Scene prototypes for the people search task in urban scenes.
Experiment: Three Visual Search Counting tasks

Observers (8 per task) search for a target object and count the number of occurrences of the target object (target present or absent). Eye movements monitored.

People search task  Mug and painting search tasks
Results: People search

Saliency predictions

Contextual Guidance Model
(saliency * scene priors)

Dots correspond to human fixations 1-4
Results: People search

1- The best predictor of eye movements during search are the observers themselves (consistency across participants)
2- The contextual guidance model predicts fixations better than saliency alone
Mug and painting search

Task: painting search

Task: mug search

Saliency
Target (a man or a woman?)

Salient distractor

in or out context region

Experimental Procedure

Instructions: Find the person in the scene and report as quickly and accurately as possible whether they are male or female by pressing the assigned keys.
Where will the eye be drawn?

Distractor INSIDE Context region

Distractor OUTSIDE Context region

Saliency * Context

% trials on which eyes drawn to the distractor

First fixation drawn to the Salient object WITHIN the Contextual region

Study 2

Effects of Learning Identity Priors on object search: Where does the time go?

Hidalgo-Sotelo at al (2005)
Hidalgo-Sotelo & Oliva (2006-Poster, in preparation)

CogLunch Talk by Barbara Hidalgo-Sotelo October 24, Noon, BCS
Contextual Priors in Object Search

- **Categorical priors**
  - Association between a *scene category* and an object

- **Identity priors**
  - Association between a *specific scene exemplar* and an object

Novel scene

After learning (20 repetitions)
Role of Identity Priors in Object Search

- **Where** in the search process do identity priors influence visual search?
- How do the **strength** of identity priors affect visual search efficiency?
Where did the time go?

- Do contextual identity priors benefit object search at an 
  early or late stage?

**Three stages:**

- Central Fixation
- Scan time to enter target region
- Gaze duration in target region
Strength of Identity Priors for target presence

**Weak Identity Priors**-
- \( P(\text{target} \mid \text{scene } X) = 0.5 \)

**Strong Identity Priors**-
- \( P(\text{target} \mid \text{scene } X) = 1.0 \)
- \( P(\text{target} \mid \text{scene } Y) = 0 \)

+ Control condition where scenes are never repeated (categorical priors)
Experimental Procedure

48 scene stimuli-
50% target present, 50% target absent
20 Repetitions of 48 scenes (10 epochs)
The classical result - People learn

1) Search efficiency *improves* over many exposures

2) *Strength* of identity prior influences Reaction Time gain.

Categorical Priors (control)
(145 ms Epoch 1 – 10)

Weak Identity Priors
(293 ms Epoch 1 – 10)

Strong Identity Priors
(337 ms Epoch 1 – 10)
Initial Fixation Duration

At the earliest search stage: similar behavior

Categorical Priors  Weak IP  Strong IP
Scan Time

Scan Time – Target Present

Categorical Priors

Weak Identity Priors ~

Strong Identity Priors
Gaze Duration on Target Object

Gaze Duration – Target Present

Categorical Priors - control
(target presence probability 0.5)

Weak Identity Priors (target presence probability 0.5)

Strong Identity Priors (target presence is 1)
Where do Top-down Identity Priors affect visual search?

- And much more results ....

- Predictability of scene-target association differentially modulates search efficiency

- The nature of efficiency improvements can be manifested at different stages of visual search
Study 3

Automatic activation of scene priors in visual and memory search


Oliva & Wolfe (in preparation)
Panorama Procedure

- The window slides back and forth and at each instant, it shows a specific view of the panoramic scene.
- E.g. a visible set of 5 objects and an Hidden set of 3 objects.
Visible set size

Slope: a cost of 10 msec per object in the view

Efficient search: no slope

Memory window
Effect of identity priors on visual and memory search

Phase 1
Visual search clutter dependent (slope)
Memory search clutter independent (no slope)

Phase 2

Phase 3
?
Memory search No slope
Automatic effect of top-down identity priors

Phase 1
Visual search (slope of 10 msec)
Memory search (no slope)

Phase 2

Phase 3
Ghost clutter (Phase 3)
Slope = 9 msec

Ghost set size
Automatic effect of top-down identity priors

Visible Target (10 msec)
Hidden Target – Memory Search - (flat)

Number of objects

New scene
Memory search

Number of ghosts

Oliva & Wolfe (in preparation)
Conclusion – Study 3

- Mandatory “scene context” effect on search:
  The scene identity may activate a “search strategy” that appears to be independent of volitional control, even if another strategy (memory search) would have been more efficient.

* A scene context priors influences (learning the statistical regularities between a pattern and the location of a target) might occur before the end of the glance (~ 100 msec. cf. Chaumon et al. 2006 – MEG study)
Acknowledgments

Barbara Hidalgo-Sotelo, Jeremy Wolfe, Antonio Torralba, Michelle Greene, George Alvarez, Talia Konkle, Todd Horowitz, Naomi Kenner

Sponsors
National Science Foundation CAREER award
NEC award in computers and communication

Contact
Aude Oliva (oliva@mit.edu)
Computational Visual Cognition Lab
Web site: cvcl.mit.edu