Decomposing the effect of contextual priors in visual search: Where does the Time Go?

Barbara Hidalgo-Sotelo, Aude Oliva

Brain and Cognitive Sciences Department, MIT

Background

Goal Directed Search

Behaviorally relevant visual search is often goal directed (*Where are my keys?*)

Familiarity is established as experience with an environment accumulates

- Features of **environment** may become familiar (*Layout of bedroom*)
- Features of **target** may become familiar (*Keychain has blue fleurs-de-lis*)
- Statistical regularities in environment-target **covariation** may become familiar (*My keys and cell phone are usually on 2nd shelf of bookshelf*)

Where is the vision conference?

Where is Jeremy Wolfe?

Contextual Priors

Top-down information can drive search via 2 types of associative prior knowledge [3]

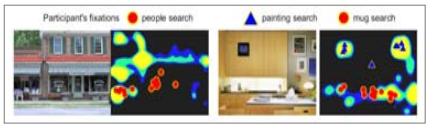
1) Categorical Prior

Knowledge about the relationship between an object's location and a **scene category**

2) Identity Prior

Knowledge about the relationship between a particular object's location and a **particular scene**

Prior studies show that categorical prior information can be integrated prior to the first saccade [4].



Task information for a specific object can change the way that contextual features are used to select relevant image regions. *From [4]*

Questions

- What **stages** of visual search are susceptible to experience-based identity priors (IP)?
- Does the **strength** of environment-target covariation influence contextual guidance by identity priors?

Method

Task

Record eye movements (ISCAN RK-464 eye tracker, 240 Hz)

Stimuli

- 48 photographs of outdoor scenes per block
- 50% Target prevalence per block
- Repeat search over 20 blocks
- Two levels of difficulty: "Easy search" & "Hard search"
- Three levels of identity prior strength

Identity Priors

Unrepeated Scenes:

MAP 0 – No Identity Priors (N = 14)

Repeated Scenes:

MAP 1 – "Weak Identity Priors" (N = 12)

Scene identity predicts P(target) = 50%

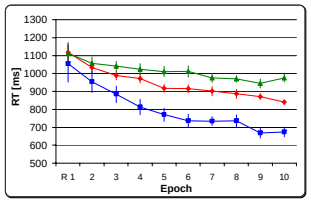
MAP 2 – "Strong Identity Priors" (N = 8)

Scene identity indicates P(target) = 0% or 100%

Building Identity Priors

Results

Reaction Time



Overall Δ

99 ms

221 ms

330 ms

Time Decomposition:

Determine source of RT decrease

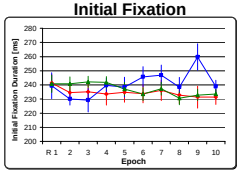
+ **Initial Fixation**

Scan time to enter target region

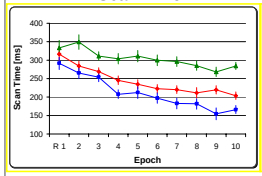
Gaze duration in target region

Q1 – Which **stages** of search are susceptible to IP?

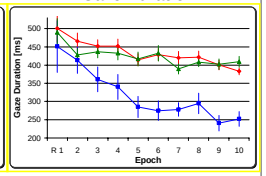
Initial Fixation



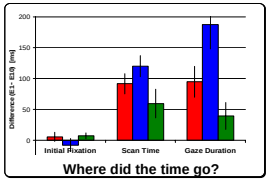
Scan Time



Gaze Duration



Q2 – How does IP **strength** impact context guidance?



Where did the time go?

Weak Identity Priors (ΔRT = 221 ms)

Scan Time → 42% (91 ms)

Gaze Duration → 43% (95 ms)

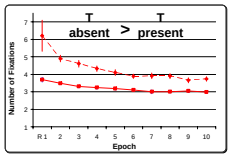
Strong Identity Priors (ΔRT = 330 ms)

Scan Time → 37% (120 ms)

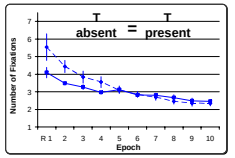
Gaze Duration → 57% (188 ms)

Results

Number of Fixations

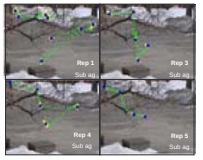


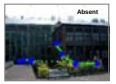
Map 1



Map 2

Scan Paths





Conclusions

Identity priors (specific scene familiarity) can speed search to determine the presence or absence of a familiar target object.

Scan time (search stage) is decreased with identity priors, regardless of the strength of target-scene mapping.

Gaze duration (recognition/response stage) is decreased with identity priors, with strong identity priors yielding a greater magnitude of improvement than weak identity priors.

References

[1] Biederman I, Mezzanotte RJ & Rabinowitz JC (1982) Scene perception: detecting and judging objects undergoing relational violations, Cogn Psychol, 14(2): 142-177.

[2] Chun MM & Jiang Y (1998) Contextual cueing: implicit learning and memory of visual context guides spatial attention, Cogn Psychol, 36(1): 28-71.

[3] Hidalgo-Sotelo B, Oliva A, Torralba A (2005) Human learning of contextual priors for object search: Where does the time go? Proceedings of 3rd Workshop on Attention and Performance in CVPR.

[4] Torralba A, Oliva A, Castelhano M, Henderson J (in press, 2006) Contextual guidance of eye movements and attention in real world scenes: The role of global features on object search, Psych Review

Acknowledgments

Many thanks to Jeremy Wolfe and Antonio Torralba for helpful comments

Funding for this project was provided by a Graduate Research Fellowship from the National Science Foundations (BHS)

For more information contact:

bhs@mit.edu

http://cvcl.mit.edu