Cognitive Neuroscience Society Ninth Annual Meeting, April14 - 16, San Francisco

Modeling of Attentional Modulation Effects in Object Recognition

Dirk Walther[1], Maximilian Riesenhuber[2], Tomaso Poggio[2], Laurent Itti[3], Christof Koch[1]

[1]California Institute of Technology, [2]Massachusetts Institute of Technology, [3]University of Southern California

Introduction

The work presented here connects two previously separate models - the model for **Saliency-based Bottom-up Attention** *by Koch and Itti [1] and the hierarchical model for object recognition* **HMAX** *by Riesenhuber and Poggio [2].*

[1] Itti, L. & Koch, C. 2001, *Nat. Rev. Neurosci.*, **2** (3), 194-203
[2] Riesenhuber, M. & Poggio, T. 1999, *Nature Neurosci.* **2** (11) 1019-1025

Saliency-based Attention Model

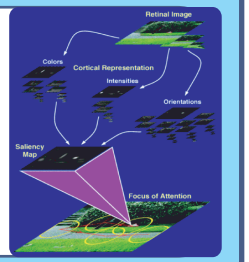
1. extraction of intensity, orientation and color (R-G, B-Y) feature maps.

2. local competition within each feature map.

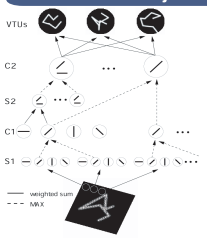
3. combination of feature maps into a unique saliency map.

4. WTA competition between candidate locations - winner is attended to.

5. inhibition of return, go to 4.

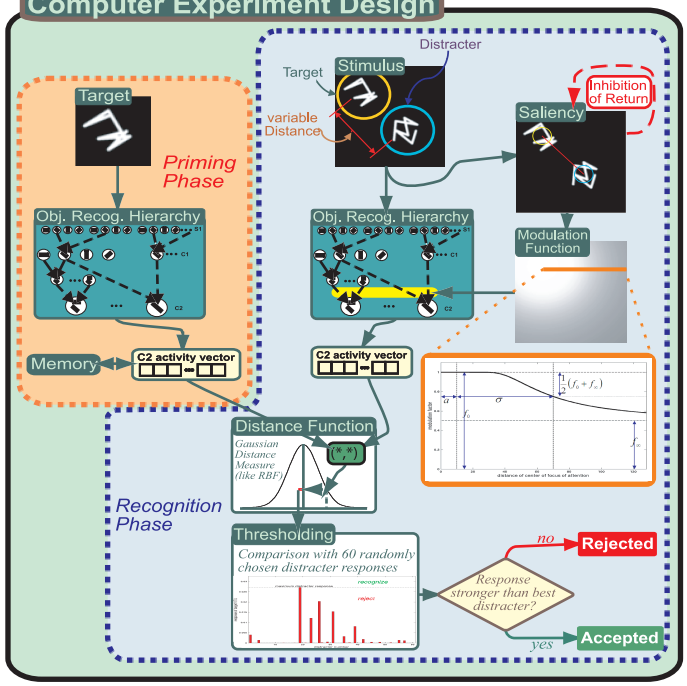


HMAX Model of Object Recognition



- features extraction in "simple cells" (S1)
- max-like pooling of S1 responses in. "complex cells" (C1).
- combination of C1 responses in "composite feature cells" (S2).
- max-like pooling of S2 responses in "complex composite cells" (C2).
- comparison of C2 response vector with view-tuned units (VTUs).

Computer Experiment Design

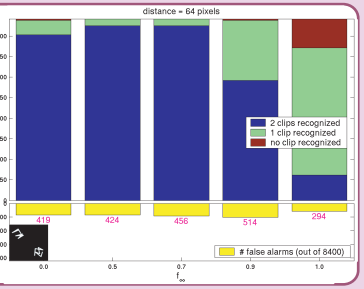


Results 2

How many paperclips does the model recognize in a 2-paperclip display?

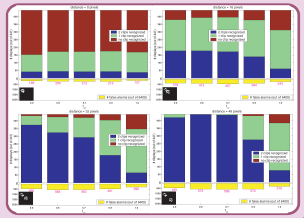
- with attention: in most cases both clips recognized.
- without attention: in most cases only one clip, in some none.

d = 64 pixels	Without attention	With attention	Single stimuli
2 clips recognized	62	425	--
1 clip recognized	309	16	21
0 clips recognized	70	0	0
False alarms (of 8400 combi-nations)	294	424	11 (of 420 com-binations)

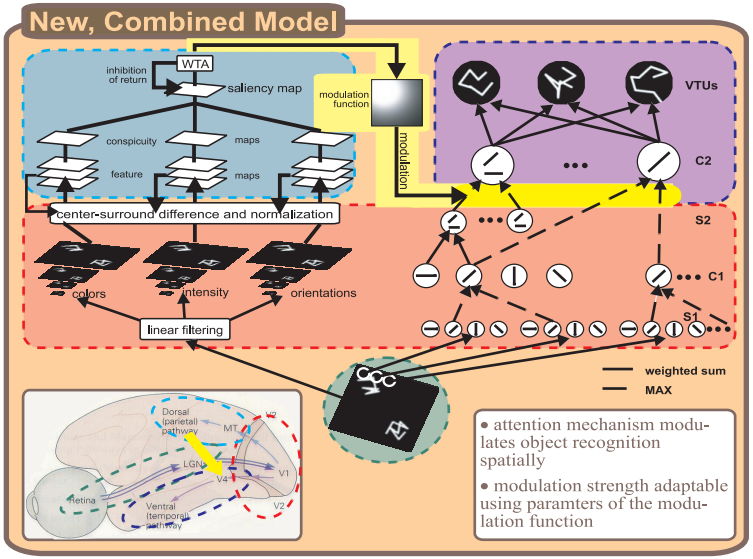


- performance drops when target and distracter are close
- but similar robustness for varying f_{inf}

- modulation strength varied by adjusting f_{inf}.
- $f_{inf} = 1$ means no modulation; $f_{inf} = 0$ means strong suppression of region outside of FOA.
- performance robust from $f_{inf} = 0.0$ through $f_{inf} = 0.7$
- performance drops for weak modulation ($f_{inf} = 0.9$)



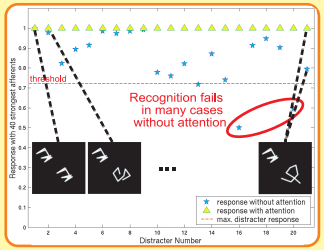
New, Combined Model

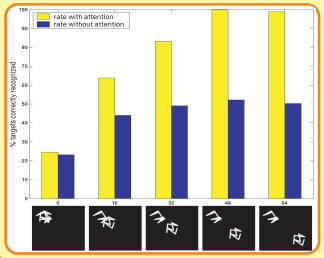


- attention mechanism modulates object recognition spatially
- modulation strength adaptable using paramters of the modulation function

Results 1

- VTU trained with ★
- tested with 21 different 2-clip stimuli
- ★ responses **without** attentional modulation
- ▲ responses **with** attentional modulation
- recognition more reliable with attention





- test with all 441 2-clip stimuli
- vary the target-distracter distance
- ■ with attn. modulation - up to 100%
- ■ without attn. modulation - only 50% recognition performance

Discussion and Outlook

Discussion

We are presenting first results from simulation experiments connecting a model for saliency-based bottom-up attention with a hierarchical model for object recognition (*HMAX*). In these experiments, the spatial information obtained from the saliency model was used to modulate the activity of a processing stage in the *HMAX* hierarchy that is approximately equivalent to V4 in monkey cortex.

Several physiological studies ([3-5]) indicate attentional modulation of V4. It has yet to be tested, whether modulation of processing stages equivalent to earlier visual areas like V1 yields similar results.

The paperclip stimuli used in these experiments are very artificial. It is imperative to test the model with more natural stimuli, which might necessitate an extension of the feature extraction units (at the S1 level) in the *HMAX* model.

More complex interactions between the two models for the *"where"* and the *"what"* pathways of visual processing are planned.

[3] Connor et al. 1997, *J. Cogn. Neurosc. 17*(9): 3201-3214
[4] Luck et al. 1997, *J. Neurophysiol. 77*(1): 24-42
[5] He et al. 1996, *Nature 383*: 334-337

Outlook

Next steps for the interactions between the saliency-based model and HMAX:

- Extract size information from saliency map and modulate scale bands in HMAX.
- Test model with cluttered natural scenes.

Long-term plans:

- Bias saliency map towards certain features for visual search
- Add volitional top-down control
- Adapt shape and size of IOR mask to the results of HMAX
- Close local integration of the neural circuitry of the two models
- Model saccade planning and execution

Acknowledgement

Financial support for this work is provided by the National Science Foundation.

corresponding author: Dirk Walther <walther@caltech.edu>