Modeling Attention to Salient Proto-objects

Dirk Bernhardt-Walther

Collaborators:
Laurent Itti
Christof Koch
Pietro Perona
Tommy Poggio
Maximilian Riesenhuber
Ueli Rutishauser

Beckman Institute,
University of Illinois
at Urbana-Champaign
Attention and Object Recognition

Parsing a scene into objects:

- Selective attention used to serialize perception
- How can objects be attended *before* they are perceived?

- Our solution: select salient proto-object regions *likely* to contain an object, based on bottom-up visual information.
Outline

1. A model for attending to salient proto-object regions.

2. Application to biologically plausible object recognition.

3. Application to computer vision.
Proto-objects in Coherence Theory

“Quick and dirty“ interpretation

Only structures above primary line are “visible” to focused attention

“Quick and clean” measurement

Secondary processing stage
- proto-objects (rapid vision)
- local interpretation

Primary processing stage
- edges (linear filtering)
- local inhibition/excitation

Transduction stage
- pixels (photoreception)
- minimal interactions

Attending to proto-objects

Rensink, Vision Res. 2000

Nexus
- collects information from proto-objects
- basis of high-level decisions

Links
- feedback/forward paths
- provide coherence for selected proto-objects

Proto-objects
- retinopic array of constantly-regenerating volatile structures
Related Ideas

- Coherence Theory (Rensink 2000 a,b)

- Object files (Kahneman and Treisman 1984, 1992)

- Attention spreading over objects (Egly et al. 1994)

- Spreading of inhibition of return over objects (Jordan and Tipper 1999)
Bottom-up attention to locations

Koch and Ullman 1985

Itti, Koch, and Niebur 1998

Itti and Koch 2001
Saliency-based region selection

Matlab code:
http://www.saliencytoolbox.net
Outline

1. A model for attending to salient proto-object regions.

2. Application to biologically plausible object recognition.

3. Application to computer vision.
Proto-objects in Coherence Theory


Only structures above primary line are “visible” to focused attention

“Quick and dirty” interpretation

Primary line

“Quick and clean” measurement

Secondary processing stage
- proto-objects (rapid vision)
- local interpretation

Primary processing stage
- edges (linear filtering)
- local inhibition/excitation

Transduction stage
- pixels (photoreception)
- minimal interactions
Attending to proto-objects

Rensink, Vision Res. 2000

**Nexus**
- collects information from proto-objects
- basis of high-level decisions

**Links**
- feedback/forward paths
- provide coherence for selected proto-objects

**Proto-objects**
- retinopic array of constantly-regenerating volatile structures
Attentional modulation in HMAX

(with Maximilian Riesenhuber and Tomaso Poggio, MIT)

Walther et al. *BMCV* 2002
Walther & Koch, *Neural Netw.* 2006

(Riesenhuber & Poggio 1999)
Attentional modulation of V4 neurons

Delayed match-to-sample task

Firing rate of two neurons

McAdams & Maunsell 2000

21 %  38 %
Deployment of attention in area V4

Modulation of V4 neurons in macaques due to spatial attention

Spitzer et al. 1988
Connor et al. 1997
Luck et al. 1997
Reynolds et al. 2000
McAdams & Maunsell 2000
McAdams & Maunsell 2000 (*)
Chelazzi et al. 2001

(*) spatial and feature-based attention
Modeling Results – paperclips

mean ROC area

Modulation Strength $\mu$

no attention 0 0.2 0.4 0.6 0.8 1

error bars: s.e.m. over 21 paperclips

paperclips (Logothetis et al. 1994)

64 pixels separation
48 pixels separation
32 pixels separation
16 pixels separation
0 pixels separation

Walther & Koch, NN 2006
Modeling Results – MPI Tübingen faces

error bars: s.e.m. over 21 faces

faces (Vetter & Blanz 1994)
Deployment of attention in area V4

Modulation of V4 neurons in macaques due to spatial attention

- Spitzer et al. 1988: 18%
- Connor et al. 1997: 39%
- Luck et al. 1997: 30-42%
- Reynolds et al. 2000: 51%
- McAdams & Maunsell 2000: 31%
- McAdams & Maunsell 2000 (*): 54%
- Chelazzi et al. 2001: 39-63%

(*) spatial and feature-based attention

Walther and Koch, Neural Netw. 2006
Summary – biologically motivated vision

- Can use spatial attention to enable detection of multiple objects in a biologically realistic model of object recognition in cortex.

- 20-40 % modulation of neural activity is sufficient in our experiments → compatible with neurophysiology
Attention Deployment in Machine Vision

(with Ueli Rutishauser and Pietro Perona, Caltech)

- Spatial grouping and preferential processing of a likely object regions for:
  - Learning and recognition of several independent objects from a single image,
  - Pruning away clutter.

- All-or-none modulation preferable to save resources, i.e., $\mu = 1$.

- Apply to recognition system by Lowe (2004):
  1. Detect scale invariant features (SIFT),
  2. Store and match constellations of SIFT features.
Learning and recognizing multiple objects

Walther, Rutishauser, Koch, and Perona, CVIU 2005
Grouping of keypoints
Learning and recognizing multiple objects

- 1 training image, 101 test images
- 2 objects learned from training image
- Book present in 24, box in 23 images, in four of these both objects are present
- Recognition results:

<table>
<thead>
<tr>
<th>object</th>
<th>hits</th>
<th>misses</th>
<th>false positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>box</td>
<td>21 (91%)</td>
<td>2 (9%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>book</td>
<td>14 (58%)</td>
<td>10 (42%)</td>
<td>2 (2.6%)</td>
</tr>
</tbody>
</table>

Objects in cluttered scenes

- Using the relative object (ROS) size as an inverse measure of the amount of clutter.

\[
ROS = \frac{\# \text{ pixels } \_ \text{ object}}{\# \text{ pixels } \_ \text{ image}}
\]
Recognition performance in clutter

(true positive rate < 0.07 % in all cases)
Conclusions

☐ Suggest a simple bottom-up way to attend to proto-object regions,

☐ Demonstrate applications in biological modeling and in computer vision.

Not covered:

☐ Non-spatial attention (e.g. feature or object-based attention to overlapping stimuli),

☐ Top-down biases (see, e.g., Navalpakkam and Itti 2005).
Acknowledgements

- Collaborators:
  - Caltech: Christof Koch, Ueli Rutishauser, Pietro Perona
  - USC: Laurent Itti
  - MIT: Maximilian Riesenhuber (now Georgetown), Tomaso Poggio

- Funding: Beckman Foundation, NSF, NIMH, Sloan-Swartz Foundation, Keck Foundation

- Matlab code available at: http://www.saliencytoolbox.net