Saliency Map


COMPUTATIONAL MODELLING OF VISUAL ATTENTION

Laurent Itti* and Christof Koch‡

NATURE REVIEWS | NEUROSCIENCE VOLUME 2 | MARCH 2001 | 195

5 dimensions importantes issues des modèles de l'attention visuelles

- First, the perceptual saliency of stimuli critically depends on the surrounding context.
- Second, a unique 'saliency map' that topographically encodes for stimulus conspicuity over the visual scene has proved to be an efficient and plausible bottom-up control strategy.
- Third, inhibition of return, the process by which the currently attended location is prevented from being attended again, is a crucial element of attentional deployment.
- Fourth, attention and eye movements tightly interplay, posing computational challenges with respect to the coordinate system used to control attention.
- And last, scene understanding and object recognition strongly constrain the selection of attended locations

Visual Attention - definition

- The most important function of selective visual attention is **to direct our gaze** rapidly towards objects of interest in our visual environment.
- Mais selon Posner (1980)

To illustrate these principles, it is important to divide the attention system into subsystems that *perform different* but interrelated functions. In this chapter, we consider three major functions that have been prominent in cognitive accounts of attention (Posner & Boies, 1971): (1) orienting to sensory events; (2) detecting signals for focal (conscious) processing, and (3) the maintenance of a vigilant or alert

Visual orienting is usually defined in terms of the foveation of a stimulus (overt). Foveating a stimulus improves efficiency of processing targets in terms of acuity, but it is also possible to change the priority given a stimulus by attending to its location covertly without any change in eye or head position (Posner, 1988).

Visual Attention – dual mechanism

- How does it work ?
  - This framework suggests that subjects selectively direct attention to objects in a scene using both bottom-up, image-based saliency cues and top-down, task-dependent cues
    - **bottom-up** = image-based saliency cues
    - **top-down** = task-dependent cues
  - Both mechanisms can operate in parallel

Image-based saliency

- Some **features** are intrinsically salient and automatically and involuntarily attract attention
  - Saliency is independant of the nature of the task
  - Saliency is computed in a pre-attentive manner
  - The speed of this saliency-based form of attention is on the order of 25 to 50 ms per item.

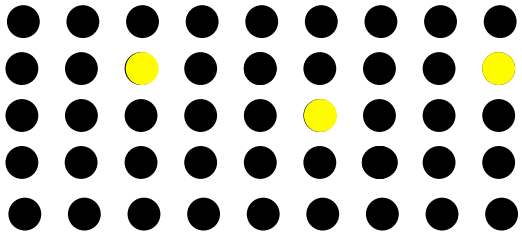
Task-dependent cues

- The second form of attention is a more deliberate and […] volitional
- The expression of this top- down attention is most probably controlled from higher areas, […] which connect back into visual cortex and early visual areas.
  - It takes 200 ms or more [to orient visual attention] — rivals that needed to move the eye

Basis of the model

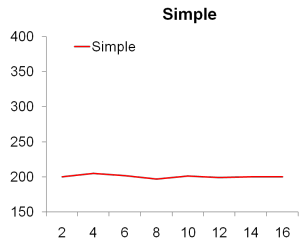
- Treisman and Gelade FIT theory (1980)
  - Distinction between pop-pout and conjunctive search
  - Xp 1

Experience # 1



Pre-attentive mechanism: whatever the number of distractors, the reaction time remains constant.

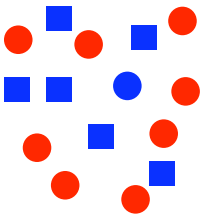
Results



25/01/16 Cognition Adulte – L2 9

Attentive mechanism

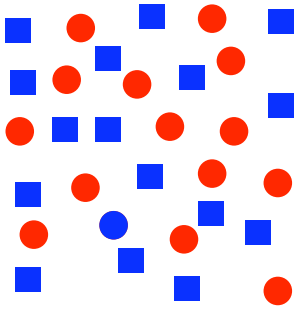
Where is the blue disk?



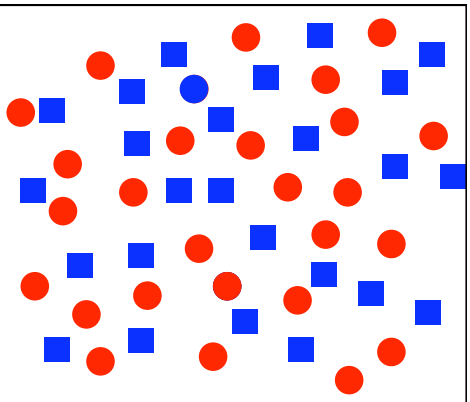
25/01/16 10

Attentive mechanism

Where is the blue disk?

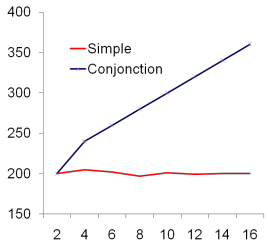


25/01/16 11



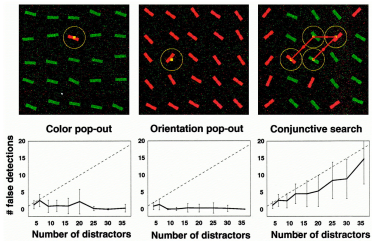
Serial processing: half of the objects are expected to be inspected

results



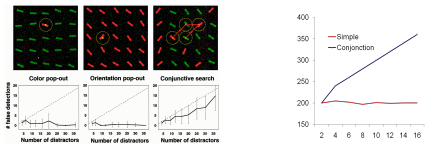
Basis of the model

- Treisman and Gelade FIT theory (1980)
  - Distinction between pop-pout and conjunctive search



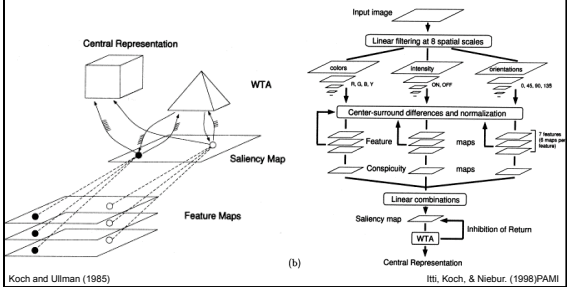
Basis of the model

- Treisman and Gelade FIT theory (1980)
  - Distinction between pop-pout and conjunctive search



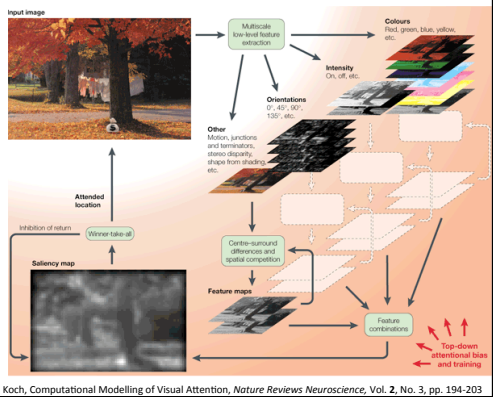
  - This lead to a distinction beetween
    - Pre-attentive features (automatic and parallel search)
    - Attentive features (« volitional » and conjunction search)

Saliency map Archeology



Koch and Ullman (1985) Itti, Koch, & Niebur. (1998)PAMI

Structure du modèle



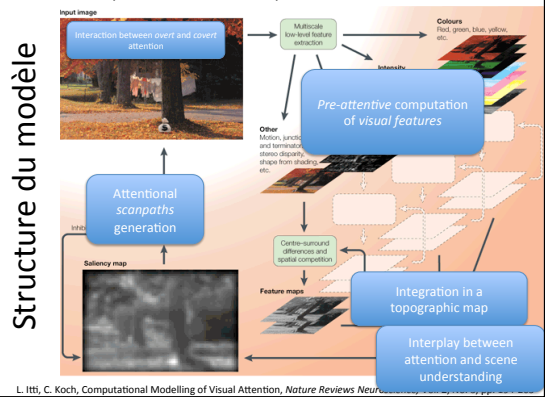
L. Itti, C. Koch, Computational Modelling of Visual Attention, *Nature Reviews Neuroscience,* Vol. **2**, No. 3, pp. 194-203

Saliency Map Model

- Five component of « *bottom-up* » attention model
  1. *Pre-attentive* computation of *visual features*
  2. Integration in a topographic map
  3. Attentional *scanpaths* generation
  4. Interaction between *overt* and *covert* attention
  5. Interplay between attention and scene understanding

Structure du modèle

Five component of « *bottom-up* » attention model



L. Itti, C. Koch, Computational Modelling of Visual Attention, *Nature Reviews Neur*

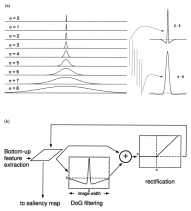


Fig. 2. (a) Gaussian pixel widths for the nine scales used in the model. Scale σ=0 corresponds to the original image, and each subsequent scale is coarser by a factor 2. Two examples of the six center-surround receptive field types are shown, for scale pairs 2...

Laurent Itti, Christof Koch

A saliency-based search mechanism for overt and covert shifts of visual attention

Vision Research, Volume 40, Issues 10–12, 2000, 1489–1506

http://dx.doi.org/10.1016/S0042-6989(99)00163-7

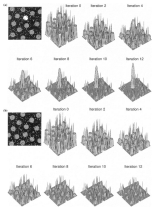


Fig. 3. (a) Iterative spatial competition for salience in a single feature map with one strongly activated location surrounded by several weaker ones. After a few iterations, the initial maximum has gained further strength while at the same time suppressing we...

Laurent Itti, Christof Koch

A saliency-based search mechanism for overt and covert shifts of visual attention

Vision Research, Volume 40, Issues 10–12, 2000, 1489–1506

http://dx.doi.org/10.1016/S0042-6989(99)00163-7

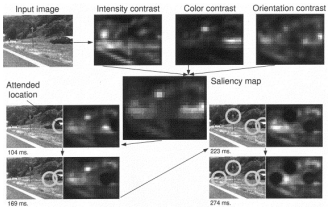


Fig. 4. Example of the working of our model with a 512×384 pixels color image. Feature maps are extracted from the input image at several spatial scales, and are combined into three separate conspicuity maps (intensity, color and orientation; see Fig. 1b) at s...

Laurent Itti, Christof Koch

A saliency-based search mechanism for overt and covert shifts of visual attention

Vision Research, Volume 40, Issues 10–12, 2000, 1489–1506

http://dx.doi.org/10.1016/S0042-6989(99)00163-7

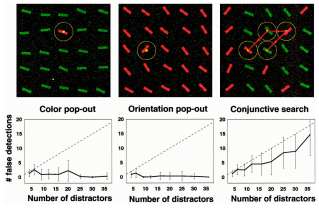


Fig. 5. Model performance on noisy versions of pop-out and conjunctive tasks of the type pioneered by Treisman and Gelade (1980).Stimuli were randomly jittered isoluminant red and green bars with strong speckle noise added. Dashed lines: chance value, based on...

Laurent Itti, Christof Koch

A saliency-based search mechanism for overt and covert shifts of visual attention

Vision Research, Volume 40, Issues 10–12, 2000, 1489–1506

http://dx.doi.org/10.1016/S0042-6989(99)00163-7

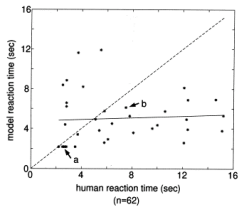


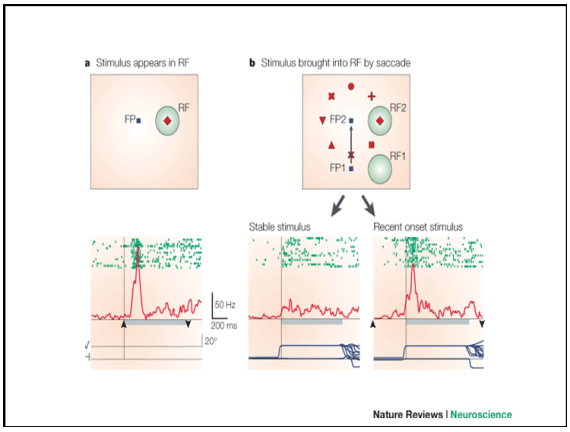
Fig. 8. Mean reaction time to detect the target for 62 human observers and for our deterministic algorithm. Eight of the 44 original images are not included, in which either the model or the humans failed to reliably find the target. For the 36 images studied,...

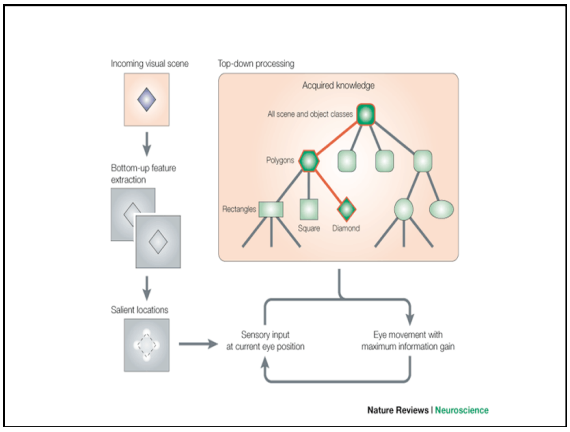
Laurent Itti, Christof Koch

A saliency-based search mechanism for overt and covert shifts of visual attention

Vision Research, Volume 40, Issues 10–12, 2000, 1489–1506

http://dx.doi.org/10.1016/S0042-6989(99)00163-7





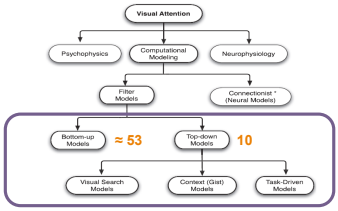
IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 35, NO. 1, JANUARY 2013

State-of-the-Art in Visual Attention Modeling

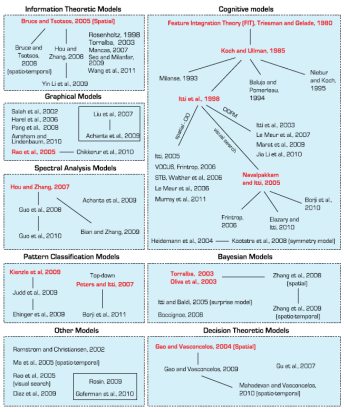
Ali Borji, *Member, IEEE*, and Laurent Itti, *Member, IEEE*

Visual Attention Modelling

- Review from Borji & Itti (2013)



A taxonomy of models:
- Information Theoretic
- Cognitive
- Graphical
- Spectral analysis
- Pattern classification
- Bayesian



Category	Application	References
Computer Vision and Graphics	**Image segmentation**	Mishra and Aloimonos, 2009, Mali et. al., 2000
	Image quality assessment	Ma and Zhang, 2008, Ninassi et. al., 2007
	Image matching	Walther et al., 2006, Siagian and Itti, 2009, Frintrop and Jensfelt, 2008
	Image rendering	DeCarlo and Santella, 2002
	Image and video compression	Duerhani et al., 2003, Itti, 2004, Guo and Zhang, 2010.
	Image thumbnailing	Marchesotti et al., 2009, Le Meur et al., 2006, Suh et al., 2003
	Image super-resolution	Jacobson et al, 2010
	Image re-targeting (thumbnailing)	Setlur et al., 2005, Chamaret et al., 2008, Goferman et al., 2010, Achanta et al., 2009, Marchesotti et al., 2009, Le Meur et al., 2006, Suh et al., 2003
	Image superresolution	Sadaka and Karam, 2009
	Video summarization	Manat et al., 2007, Ma et al., 2005
	Scene classification	Siagian and Itti, 2009
	Object detection	Frintrop, 2006, Navalpakkam and Itti, 2005, Fritz et al., 2005, Butko and Movellan, 2009, Viola and Jones, 2004, Ehinger et al., 2009.
	Salient object detection	Liu et al, 2007, Goferman et al, 2010, Achanta et al., 2009, Rosin, 2009.
	Object recognition	Sailah et al., 2002, Walther et al., 2006 and 2007, Frintrop, 2006, Moi et al., 2005, Gao and Vasconcelos, 2004 and 2009, Han and Vasconcelos 2010, Paletta et al., 2005.
	Visual tracking	Mahadevan and Vasconcelos, 2009, Frintrop, 2010
	Dynamic lighting	Seif ElNasr, 2009
	Video shot detection	Bocognone et al., 2005
	Interest point detection	Kadir and Brady, 2001, Kienzle et al., 2007.
	Automatic collage creation	Goferman et al., 2010, Wang et al., 2006.
	Face segmentation and tracking	Li and Ngan, 2006
Robotics	**Active vision**	Mertsching et al., 1999, Vijaykumar et al., 2001, Dankers, 2007, Borji et al., 2010
	Robot Localization	Siagian and Itti, 2009, Duerhani et al., 2005
	Robot Navigation	Baluja and Pomerleau, 1997, Schwier and Egner, 1997
	Human-robot interaction	Breazeal, 1999, Heidemann et al., 2004, Belandinelli, 2008, Nagai, 2009, Muhl, 2007
	Synthetic vision for simulated actors	Courty and Marchand, 2003
Others	**Advertising**	Rosenholtz et al, 2011, Liu et al., 2008
	Finding tumors in mammograms	Hong and Brady, 2003
	Retinal prostheses	Parikh et al., 2010

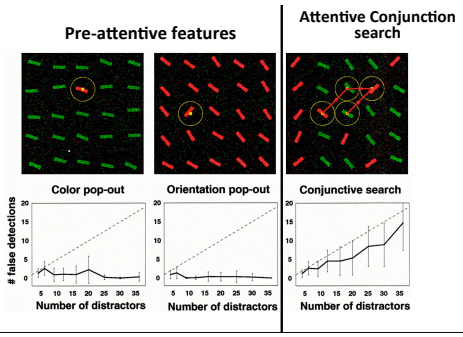
Antonio Torralba

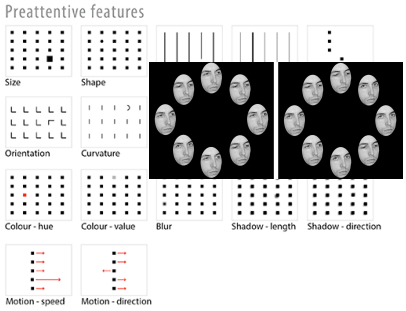
http://saliency.mit.edu/

The goal of this website is to be the most up-to-date, online source of saliency model performances and datasets. We believe that a continuously updated all-in-one comparison page will serve as an essential resource to document and promote progress in the field of saliency modeling.

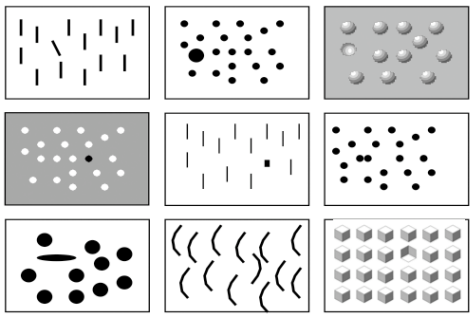
Limits of FIT theory

Limits of FIT theory





Primitives



Primitives

- Ligne
- Longueur
- Largeur
- Taille
- Courbure
- Nombre
- Terminaisons
- Intersections
- Clôture
- Couleur (teinte)
- Intensité
- Scintillement
- Mouvement
- Profondeur
- Stéréoscopie
- illumination

Conjonctions

- Il existe des exceptions
  - Couleur et stéréopsie
  - Couleur et mouvement
  - Couleur et position
  - Forme et position

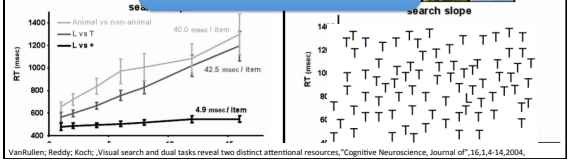
Limites of FIT theory

Task 1 : detection of a target among distractors, stimuli stay on screen until subjects response

Task 2 : detection of a target among distractors. Stimuli are displayed for 200 ms then masked masked

Pre-attentive features

Attentive Conjunction search

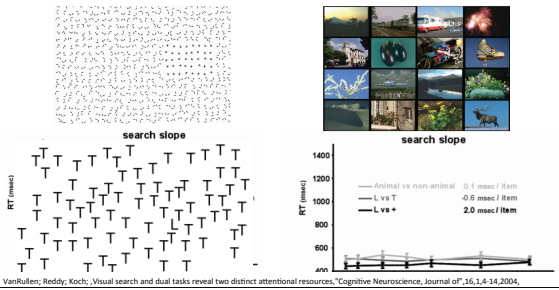


VanRullen; Reddy; Koch; ,Visual search and dual tasks reveal two distinct attentional resources,"Cognitive Neuroscience, Journal of",16,1,4-14,2004,

Limites of FIT theory

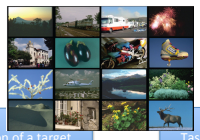
Task 2 : detection of a target among distractors. Stimuli are displayed for 200 ms then masked masked

Pre-attentive features



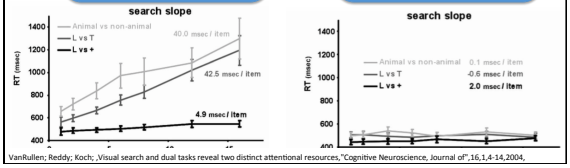
VanRullen; Reddy; Koch; ,Visual search and dual tasks reveal two distinct attentional resources,"Cognitive Neuroscience, Journal of",16,1,4-14,2004,

Limites of FIT theory



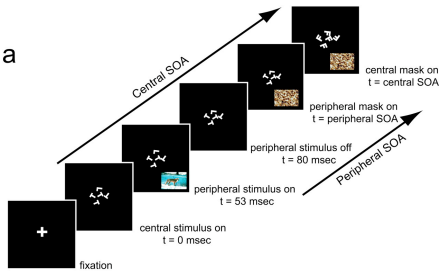
Task 1 : detection of a target among distractors, stimuli stay on screen until subjects response

Task 2 : detection of a target among distractors. Stimuli are displayed for 200 ms then masked masked



VanRullen; Reddy; Koch; ,Visual search and dual tasks reveal two distinct attentional resources,"Cognitive Neuroscience, Journal of",16,1,4-14,2004,

New protocol

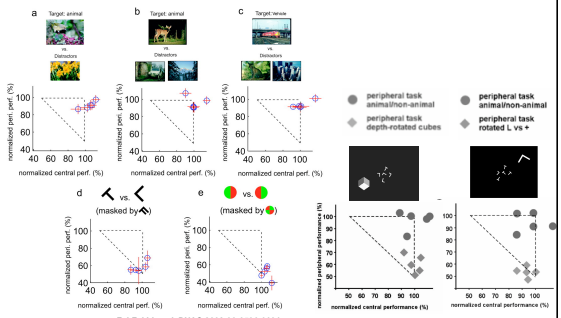


Li; VanRullen; Koch; Perona; ,Rapid natural scene categorization in the near absence of attention,Proceedings of the National Academy of Sciences, 99,14,9596-9601,2002,

©2002 by National Academy of Sciences

PNAS

Results

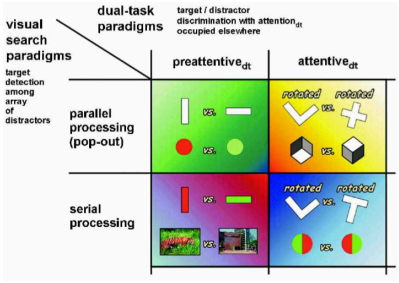


Fei Fei Li et al. PNAS 2002;99:9596-9601

©2002 by National Academy of Sciences

PNAS

Summary



Extension of Model