The computational modelling of visual attention: saliency models & saccadic models

Olivier Le Meur
olemeur@irisa.fr

IRISA - University of Rennes 1

Januray 14, 2016

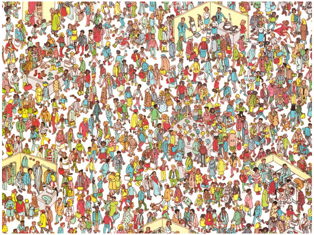
Outline

Visual Attention

1. Visual attention
   - Presentation
   - Overt vs covert
   - Bottom-Up vs Top-Down
   - Bottom-Up overt attention

Introduction to visual attention (1/5)

Natural visual scenes are cluttered and contain many different objects that cannot all be processed simultaneously.



Where is Waldo, the young boy wearing the red-striped shirt...

Amount of information coming down the optic nerve $10^8 - 10^9$ bits per second

Far exceeds what the brain is capable of processing...

Introduction to visual attention (2/5)

WE DO NOT SEE EVERYTHING AROUND US!!!

YouTube link: www.youtube.com/watch?v=ubNF9QNEQLA

Introduction to visual attention (3/5)

Visual attention

Posner proposed the following definition (Posner, 1980). Visual attention is used:

- to select important areas of our visual field (alerting);
- to search for a target in cluttered scenes (searching).

There are several kinds of visual attention:

- Overt visual attention: involving eye movements;
- Covert visual attention: without eye movements (Covert fixations are not observable).

Introduction to visual attention (4/5)

Bottom-Up vs Top-Down

- Bottom-Up: some things draw attention reflexively, in a task-independent way (Involuntary; Very quick; Unconscious);

- Top-Down: some things draw volitional attention, in a task-dependent way (Voluntary; Very slow; Conscious).

Bottom-Up vs Top-Down

- **Bottom-Up**: some things draw attention reflexively, in a task-independent way (Involuntary; Very quick; Unconscious);

- **Top-Down**: some things draw volitional attention, in a task-dependent way (Voluntary; Very slow; Conscious).

Introduction to visual attention (5/5)

Computational models of visual attention aim at predicting where we look within a scene.

In this presentation, we are focusing on Bottom-Up models of overt attention:

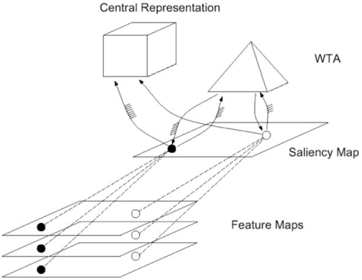
- Low-level visual features (color, luminance, texture, motion,...)
- Mid-level visual features (face, text,...).

Computational models of visual attention

2 Computational models of visual attention
- Main hypothesis
- Taxonomy
- Information theoretic model
- Cognitive model

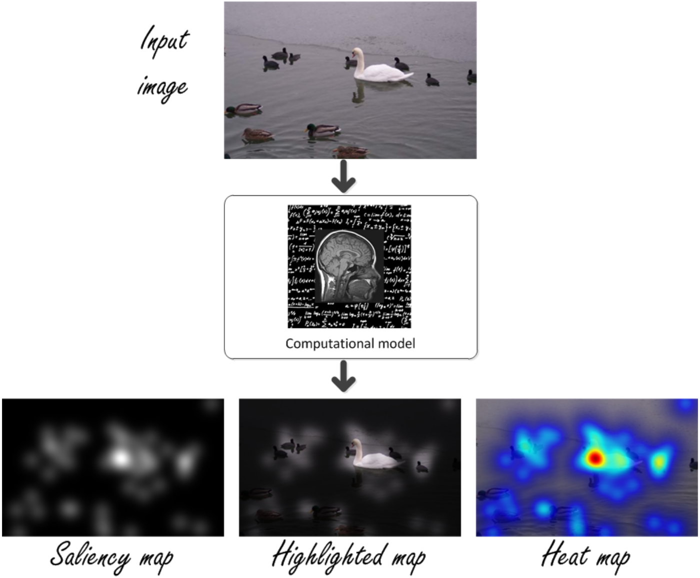
Computational models of Bottom-up visual attention (1/2)

Most of the computational models of visual attention have been motivated by the seminal work of Koch and Ullmann (Koch and Ullman, 1985).



- a plausible computational architecture to predict our gaze;
- a set of feature maps processed in a massively parallel manner;
- a single topographic saliency map.

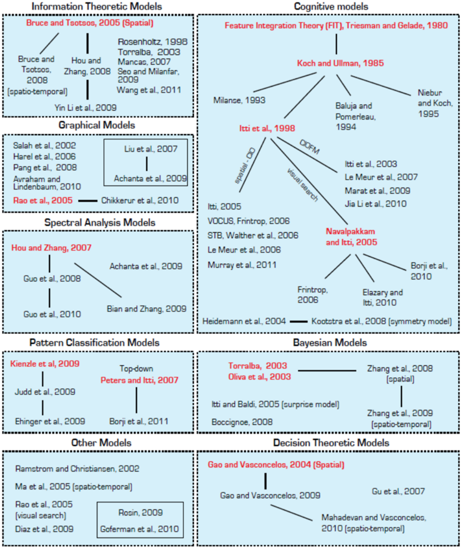
Computational models of Bottom-up visual attention (2/2)



Computational models of Bottom-up visual attention (1/1)

Taxonomy of models:

- Information Theoretic models;
- Cognitive models;
- Graphical models;
- Spectral analysis models;
- Pattern classification models;
- Bayesian models.

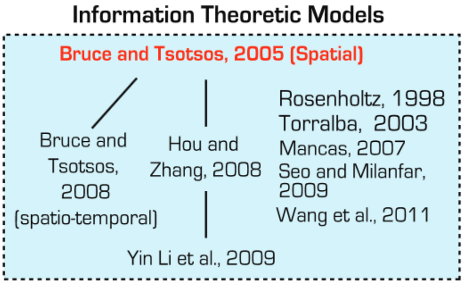


Extracted from (Borji and Itti, 2013).

Information theoretic model (1/3)

Information Theory

- Self-information,
- Mutual information,
- Entropy...



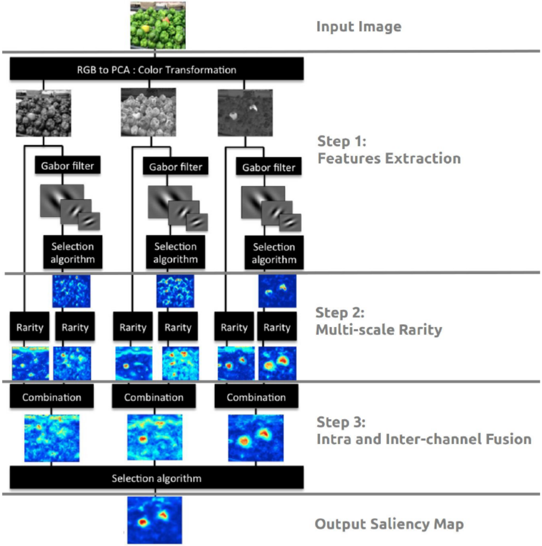
Extracted from (Borji and Itti, 2013).

Self-information is a measure of the amount information provided by an event. For a discrete X r.v defined by $\mathcal{A} = \{x_1, ..., x_N\}$ and by a pdf, the amount of information of the event $X = x_i$ is given by:

$$I(X = x_i) = -log_2 p(X = x_i), \text{ bit/symbol}$$

Information theoretic model (2/3)

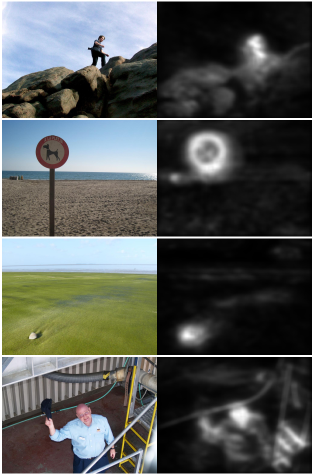
(Riche et al., 2013)'s model (RARE2012) (Extension of (Mancas et al., 2006))



Information theoretic model (3/3)

(Riche et al., 2013)'s model (RARE2012)

- Good prediction:

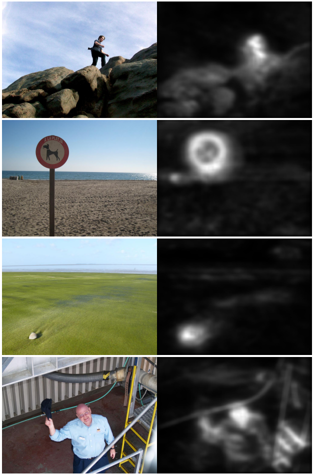


- Difficult cases:

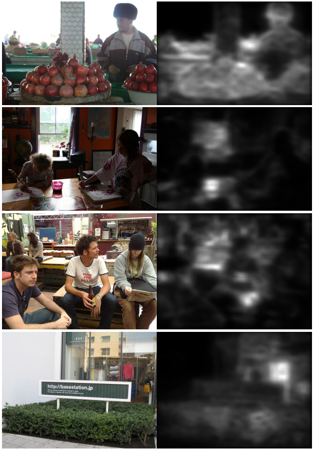
Information theoretic model (3/3)

(Riche et al., 2013)'s model (RARE2012)

- Good prediction:



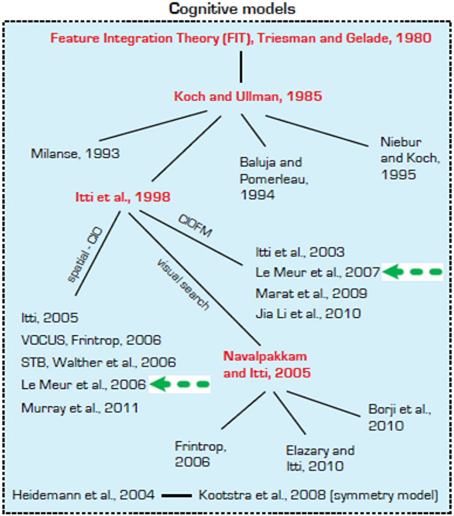
- Difficult cases:



Cognitive model (1/3)

as faithful as possible to the Human Visual System (HVS)

- inspired by cognitive concepts;
- based on the HVS properties.



Extracted from (Borji and Itti, 2013).

Cognitive model (2/3)

(Le Meur et al., 2006)'s cognitive model

In (Le Meur et al., 2006), we designed a computational model of bottom-up visual attention.

INPUT

1. Input color image;
2. Projection into a perceptual color space;
3. Subband decomposition in the Fourier domain;
4. CSF and Visual Masking;
5. Difference of Gaussians;
6. Pooling.

Cognitive model (2/3)

(Le Meur et al., 2006)'s cognitive model

In (Le Meur et al., 2006), we designed a computational model of bottom-up visual attention.

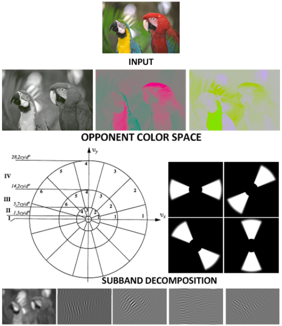
1. Input color image;
2. Projection into a perceptual color space;
3. Subband decomposition in the Fourier domain;
4. CSF and Visual Masking;
5. Difference of Gaussians;
6. Pooling.

Cognitive model (2/3)

(Le Meur et al., 2006)'s cognitive model

In (Le Meur et al., 2006), we designed a computational model of bottom-up visual attention.

1. Input color image;
2. Projection into a perceptual color space;
3. Subband decomposition in the Fourier domain;
4. CSF and Visual Masking;
5. Difference of Gaussians;
6. Pooling.

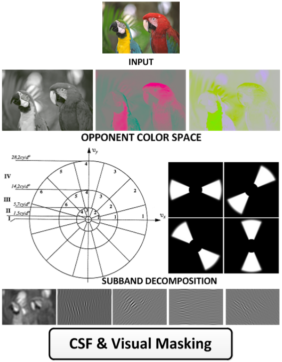


Cognitive model (2/3)

(Le Meur et al., 2006)'s cognitive model

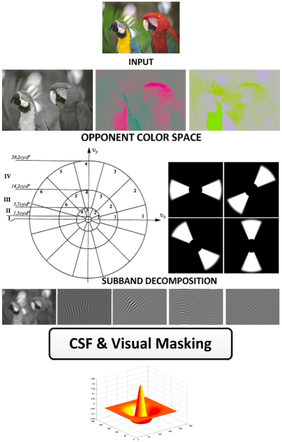
In (Le Meur et al., 2006), we designed a computational model of bottom-up visual attention.

1. Input color image;
2. Projection into a perceptual color space;
3. Subband decomposition in the Fourier domain;
4. CSF and Visual Masking;
5. Difference of Gaussians;
6. Pooling.



In (Le Meur et al., 2006), we designed a computational model of bottom-up visual attention.

1. Input color image;
2. Projection into a perceptual color space;
3. Subband decomposition in the Fourier domain;
4. CSF and Visual Masking;
5. Difference of Gaussians;
6. Pooling.

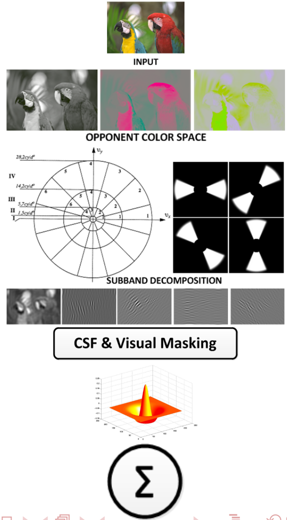


Cognitive model (2/3)

(Le Meur et al., 2006)'s cognitive model

In (Le Meur et al., 2006), we designed a computational model of bottom-up visual attention.

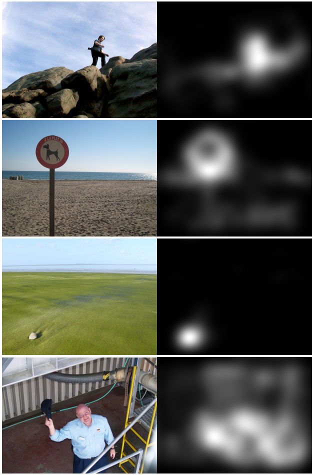
1. Input color image;
2. Projection into a perceptual color space;
3. Subband decomposition in the Fourier domain;
4. CSF and Visual Masking;
5. Difference of Gaussians;
6. Pooling.



Cognitive model (3/3)

(Le Meur et al., 2006)'s cognitive model

- Good prediction:

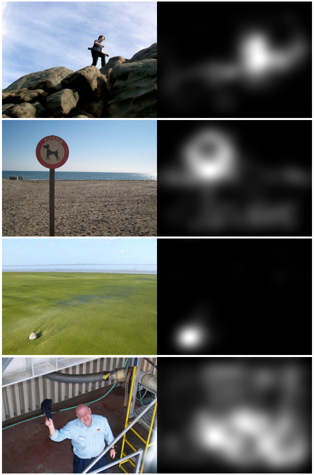


- Difficult cases:

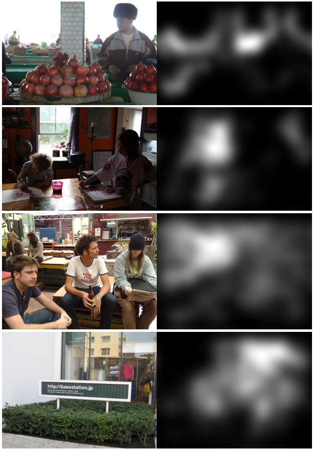
Cognitive model (3/3)

(Le Meur et al., 2006)'s cognitive model

➠ Good prediction:



➠ Difficult cases:



Performances

3. Saliency model's performance
   - Ground truth
   - Similarity metrics
   - Benchmark
   - Limitations and What's next?

Ground truth (1/2)

The requirement of a ground truth

- Eye tracker:

- A panel of observers;
- An appropriate protocol.

Adapted from (Judd et al., 2009).

Ground truth (2/2)

- Discrete fixation map f^i for the i^{th} observer:

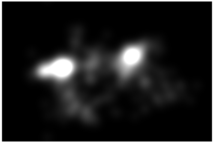
$$f^i(\mathbf{x}) = \sum_{k=1}^{M} \delta(\mathbf{x} - \mathbf{x}_k)$$

where M is the number of fixations and $\mathbf{x}_k$ is the k^{th} fixation.

- Continuous saliency map S:

$$S(\mathbf{x}) = \left(\frac{1}{N} \sum_{i=1}^{N} f^i(\mathbf{x}) \right) * G_\sigma(\mathbf{x})$$

where N is the number of observers.



Similarity metrics (1/2)
For comparing two maps

- The linear correlation coefficient, $cc \in [-1, 1]$;
- The similarity metric sim uses the normalized probability distributions of the two maps (Judd et al., 2012). The similarity is the sum of the minimum values at each point in the distributions:

$$sim = \sum_{\mathbf{x}} \min \left(pdf_{map1}(\mathbf{x}), pdf_{map2}(\mathbf{x}) \right) \tag{1}$$

  - $sim = 1$ means the pdfs are identical, $sim = 0$ means the pdfs are completely opposite.
- Earth Mover's Distance metric EMD is a measure of the distance between two probability distributions. It computes the minimal cost to transform one probability distribution into another one.
  - $EMD = 0$ means the distributions are identical, i.e. the cost is null.

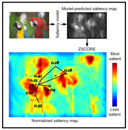
Matlab software is available on the following webpage: http://saliency.mit.edu/.

Similarity metrics (2/2)

For comparing a map and a set of visual fixations

- Receiver Operating Analysis;
- Normalized Scanpath Saliency (Parkhurst et al., 2002, Peters et al., 2005);
- Percentile (Peters and Itti, 2008);
- The Kullback-Leibler divergence (Itti and Baldi, 2005).

See the review:



Le Meur, O. & Baccino, T., Methods for comparing scanpaths and saliency maps: strengths and weaknesses, Behavior Research Method, 2013.

Benchmark (1/1)

More recently, two new *online* benchmarks (http://saliency.mit.edu/): MIT300 and CAT2000.

Dataset	Citation	Images	Observers	Tasks	Durations	Extra Notes
MIT300	Tilke Judd, Fredo Durand, Antonio Torralba. A Benchmark of Computational Models of Saliency to Predict Human Fixations [MIT tech report 2012]	**300** natural indoor and outdoor scenes *size:* max dim: 1024px, other dim: 457-1024px 1 dva* ~ 35px	**39** *ages:* 18-50	free viewing	3 sec	This was the first data set with held-out human eye movements, and is used as a benchmark test set. *eyetracker:* ETL 400 ISCAN (240Hz) Download 300 test images.
CAT2000	Ali Borji, Laurent Itti. CAT2000: A Large Scale Fixation Dataset for Boosting Saliency Research [CVPR 2015 workshop on "Future of Datasets"]	**4000** images from 20 different categories *size:* 1920x1080px 1 dva* ~ 38px	**24** per image (120 in total) *ages:* 18-27	free viewing	5 sec	This dataset contains two sets of images: train and test. Train images (100 from each category) and fixations of 18 observers are shared but 6 observers are held-out. Test images are available but fixations of all 24 observers are held out. *eyetracker:* EyeLink1000 (1000Hz) Download 2000 test images. Download 2000 train images (with fixations of 18 observers).

To perform a fair comparison, download the images, run your model and submit your results...

Limitations (1/1)

The picture is much clearer than 10 years ago!
BUT...

Important aspects of our visual system are clearly overlooked

- Current models implicitly assume that eyes are equally likely to move in any direction;
- Viewing biases are not taken into account;
- The temporal dimension is not considered (static saliency map).

Limitations (1/1)

The picture is much clearer than 10 years ago!
BUT...

Important aspects of our visual system are clearly overlooked

- Current models implicitly assume that eyes are equally likely to move in any direction;
- Viewing biases are not taken into account;
- The temporal dimension is not considered (static saliency map).

Limitations (1/1)

The picture is much clearer than 10 years ago!
BUT...

Important aspects of our visual system are clearly overlooked

- Current models implicitly assume that eyes are equally likely to move in any direction;
- Viewing biases are not taken into account;
- The temporal dimension is not considered (static saliency map).

Limitations (1/1)

The picture is much clearer than 10 years ago!
BUT...

Important aspects of our visual system are clearly overlooked

- ✗ Current models implicitly assume that eyes are equally likely to move in any direction;
- ✗ Viewing biases are not taken into account;
- ✗ The temporal dimension is not considered (static saliency map).

Saccadic model

4. Saccadic model
   - Presentation
   - Proposed model

Presentation (1/2)

- Eye movements are composed of fixations and saccades. A sequence of fixations is called a visual scanpath.
- When looking at visual scenes, we perform in average 4 visual fixations per second.

Saccadic models are used:

1. to compute plausible visual scanpaths (stochastic, saccade amplitudes / orientations...);
2. to infer the scanpath-based saliency map ⇔ to predict salient areas!!

Presentation (2/2)

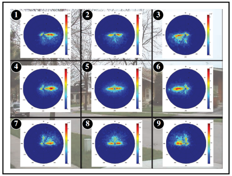
Saccadic model to infer the saliency map

The fundamental assumption is that scanpaths can be described by a Markov process, i.e. each eye fixation only depends on the previous ones.

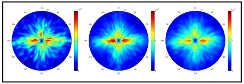
- The seminal work of (Ellis and Smith, 1985, Stark and Ellis, 1981) described a probabilistic approach where the eye movements are modelled as a first-order Markov process.

Proposed model (1/7)

O. Le Meur & Z. Liu, Saccadic model of eye movements for free-viewing condition, Vision Research, 2015



O. Le Meur & A. Coutrot, Introducing context-dependent and spatially-variant viewing biases in saccadic models, Vision Research, 2016.



O. Le Meur & A. Coutrot, How saccadic models help predict where we look during a visual task? Application to visual quality assessment, SPIE Electronic Imaging, Image Quality and System Performance XIII, 2016.

Proposed model (2/7)

So, what are the key ingredients to design a saccadic model?

- The model has to be stochastic: the subsequent fixation cannot be completely specified (given a set of data).
- The model has to generate plausible scanpaths that are similar to those generated by humans in similar conditions: distribution of saccade amplitudes and orientations, center bias...
- Inhibition of return has to be considered: time-course, spatial decay...
- Fixations should be mainly located on salient areas.

Proposed model (3/7)

Let $\mathcal{I} : \Omega \subset \mathcal{R}^2 \mapsto \mathcal{R}^3$ an image and $\mathbf{x}_t$ a fixation point at time t.

We consider the 2D discrete conditional probability:

$$p\left(\mathbf{x}|\mathbf{x}_{t-1}, \ldots, \mathbf{x}_{t-T}\right) \propto p_{BU}(\mathbf{x}) p_B(d, \phi) p_M(\mathbf{x}|\mathbf{x}_{t-1}, \cdots, \mathbf{x}_{t-T})$$

- $p_{BU} : \Omega \mapsto [0, 1]$ is the grayscale saliency map;
- $p_B(d, \phi)$ represents the joint probability distribution of saccade amplitudes and orientations. d is the saccade amplitude between two fixation points $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$ (expressed in degree of visual angle), and ϕ is the angle (expressed in degree between these two points);
- $p_M(\mathbf{x}|t-1, \ldots, t-T)$ represents the memory state of the location $\mathbf{x}$ at time t. This time-dependent term simulates the inhibition of return.

Proposed model (4/7)

Bottom-up saliency map

$$p\left(\mathbf{x}|\mathbf{x}_{t-1},\ldots,\mathbf{x}_{t-T}\right) \propto p_{BU}(\mathbf{x})p_B(d,\phi)p_M(\mathbf{x}|\mathbf{x}_{t-1},\cdots,\mathbf{x}_{t-T})$$

- p_{BU} is the bottom-up saliency map.
  - Computed by GBVS model (Harel et al., 2006). According to (Borji et al., 2012)'s benchmark, this model is among the best ones and presents a good trade-off between quality and complexity.
  - $p_{BU}(\mathbf{x})$ is constant over time. (Tatler et al., 2005) indeed demonstrated that bottom-up influences do not vanish over time.

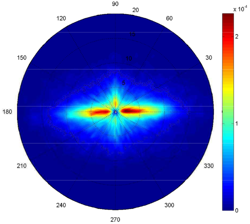
Proposed model (5/7)

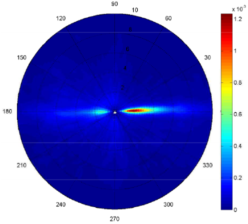
Viewing biases

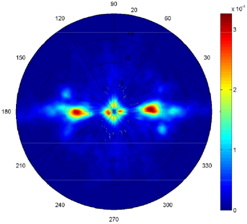
$$p\left(\mathbf{x}|\mathbf{x}_{t-1},\ldots,\mathbf{x}_{t-T}\right) \propto p_{BU}(\mathbf{x})p_B(d,\phi)p_M\left(\mathbf{x}|\mathbf{x}_{t-1},\cdots,\mathbf{x}_{t-T}\right)$$

- $p_B(d, \phi)$ represents the joint probability distribution of saccade amplitudes and orientations.
  d and ϕ represent the distance and the angle between each pair of successive fixations, respectively.



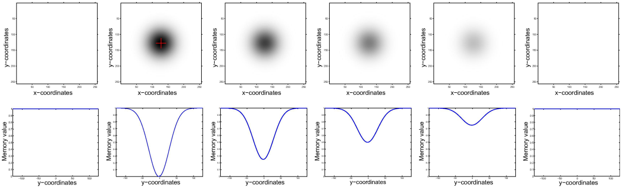







$$p\left(\mathbf{x}|\mathbf{x}_{t-1},\ldots,\mathbf{x}_{t-T}\right) \propto p_{BU}(\mathbf{x})p_B(d,\phi)p_M(\mathbf{x}|\mathbf{x}_{t-1},\cdots,\mathbf{x}_{t-T})$$

- $p_M(\mathbf{x}|\mathbf{x}_{t-1},\cdots,\mathbf{x}_{t-T})$ represents the memory effect and IoR of the location $\mathbf{x}$ at time t. It is composed of two terms: Inhibition and Recovery.



(a) Initialization (b) Inhibition (c) Recovering $\frac{1}{4}$ (d) Recovering $\frac{2}{4}$ (e) Recovering $\frac{3}{4}$ (f) Recovering $\frac{4}{4}$

- The spatial IoR effect declines as a Gaussian function $\Phi_{\sigma_i}(d)$ with the Euclidean distance d from the attended location (Bennett and Pratt, 2001);
- The temporal decline of the IoR effect is simulated by a simple linear model.

Proposed model (7/7)

Selecting the next fixation point

$$p\left(\mathbf{x}|\mathbf{x}_{t-1},\ldots,\mathbf{x}_{t-T}\right) \propto p_{BU}(\mathbf{x})p_B(d,\phi)p_M(\mathbf{x}|\mathbf{x}_{t-1},\cdots,\mathbf{x}_{t-T})$$

- Optimal next fixation point (*Bayesian ideal searcher* proposed by (Najemnik and Geisler, 2009)):

$$\mathbf{x}_t^* = \arg\max_{\mathbf{x}\in\Omega} p\left(\mathbf{x}|\mathbf{x}_{t-1},\cdots,\mathbf{x}_{t-T}\right) \quad (2)$$

Problem: this approach does not reflect the stochastic behavior of our visual system and may fail to provide plausible scanpaths (Najemnik and Geisler, 2008).

- Rather than selecting the best candidate, we generate $N_c = 5$ random locations according to the 2D discrete conditional probability $p\left(\mathbf{x}|\mathbf{x}_{t-1},\cdots,\mathbf{x}_{t-T}\right)$.
The location with the highest saliency gain is chosen as the next fixation point $\mathbf{x}_t^*$.

Saccadic model performance

5. Saccadic model's performance
    - Plausible scanpaths?
    - Similarity between human and predicted scanpaths
    - Saliency map and randomness
    - Limitations (1/1)
    - Extensions

Results (1/8)

The relevance of the proposed approach is assessed with regard to **the plausibility**, **the spatial precision** of the simulated scanpath and ability **to predict saliency areas**.

- Do the generated scanpaths present the same oculomotor biases as human scanpaths?
- What is the similarity degree between predicted and human scanpaths?
- Could the predicted scanpaths be used to form relevant saliency maps?

Results (2/8)

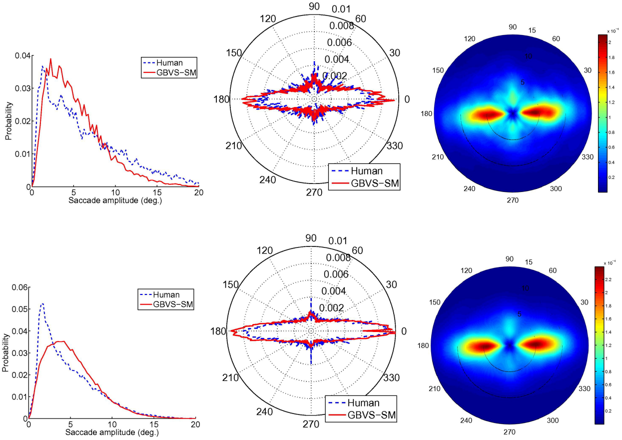
Are the simulated scanpaths plausible?

➠ Protocol:

- We assume that the simulated scanpaths are obtained in a context of purely free viewing ⇒ top-down effects are not taken into account.
- For each image in Bruce's and Judd's datasets, we generate 20 scanpaths, each composed of 10 fixations ⇒ 224600 generated visual fixations.
- We assume that the visual fixation duration is constant. So, considering an average fixation duration of 300ms, 10 fixations represent a viewing duration of 3s.
- Bottom-up saliency maps are computed by GBVS model (Harel et al., 2006).

Results (3/8)

Are the simulated scanpaths plausible?



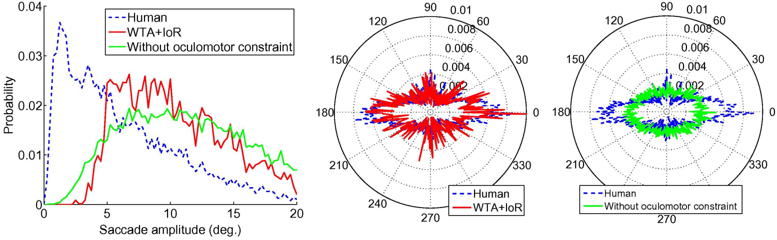
(a) (b) (c)

Top row: Bruce's dataset. Bottom row: Judd's dataset.

Results (4/8)

Are the simulated scanpaths plausible?

Impact of the oculomotor constraints (spatial and orientation), WTA+IoR



(a) (b) (c)

- Model WTA+IoR: $p_M(\mathbf{x}, t)$ is just composed of the inhibition term, i.e. re-fixation is not possible. In addition, we pick the location having the highest probability (deterministic model);
- Model without oculomotor constraint: we replace the joint probability distribution $p_B(d, \phi)$ by a 2D uniform distribution.

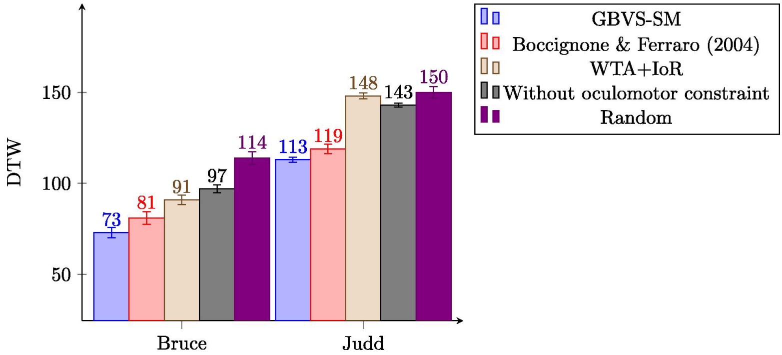
There are few methods for comparing scanpaths: string-edit (Privitera and Stark, 2000), Dynamic Time Warp algorithm (DTW) (Gupta et al., 1996, Jarodzka et al., 2010). More details in (Le Meur and Baccino, 2013).

- We use DTW's method.
- For a given image, 20 scanpaths each composed of 10 fixations are generated. The final distance between the predicted scanpath and human scanpaths is equal to the average of the 20 DTW scores.

  The closer to 0 the value DTW, the more similar the scanpaths.

Results (6/8)

What is the similarity degree between predicted and human scanpaths?

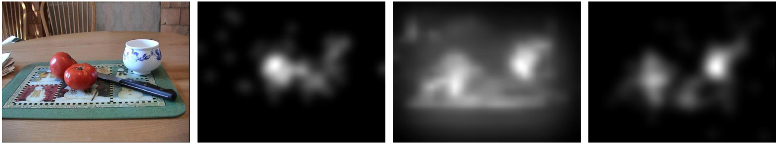


- Five models are evaluated.
- The error bars correspond to the SEM (Standard Error of the Mean).
- $DTW = 0$ when there is a perfect similarity between scanpaths.
- There is a significant difference between the performances of the proposed model and (Boccignone and Ferraro, 2004)'s model (paired t-test, $p << 0.01$).
- As expected, the lowest performances are obtained by the random model.

Results (7/8)
Scanpath-based saliency map

- We compute, for each image, 20 scanpaths, each composed of 10 fixations.

- For each image, we created a saliency map by convolving a Gaussian function over the fixation locations.



(a) (b) (c) (d)

(a) original image; (b) human saliency map; (c) GBVS saliency map; (d) GBVS-SM saliency maps computed from the simulated scanpaths.

Results (8/8)
Scanpath-based saliency map

Table 2
Performance of different models over two datasets using the linear correlation coefficient (CC), KL-divergence (KL), hit rate (HR) and the normalized scanpath saliency (NSS). The last column indicates the average rank $\bar{R}$ for the two tested datasets. The best rank is 1 whereas the lowest is 12. Note that the models proposed by Garcia-Diaz et al. (2012), Riche et al. (2013) and Harel et al. (2006) are commonly called AWS, RARE2012 and GBVS, respectively.

Method	Bruce's dataset				Judd's dataset				$\bar{R}$
	CC	KL	NSS	HR	CC	KL	NSS	HR	
Itti et al. (1998)	0.27	15.74	0.08	0.74	0.22	21.24	0.11	0.75	11.8
Hou and Zhang (2007)	0.30	13.72	0.10	0.76	0.24	19.5	0.13	0.76	9.9
Bruce and Tsotsos (2009)	0.31	16.18	0.09	0.79	0.23	22.15	0.12	0.78	10.6
Le Meur et al. (2006)	0.37	11.42	0.11	0.79	0.31	17.85	0.16	0.79	7.6
Garcia-Diaz et al. (2012)	0.43	13.14	0.14	0.80	0.32	18.89	0.17	0.80	6.4
Riche et al. (2013)	0.55	10.39	0.17	0.84	0.37	17.17	0.20	0.83	3.9
Judd et al. (2009)	0.42	14.56	0.13	0.82	**0.41**	20.24	**0.21**	**0.88**	5.4
Harel et al. (2006)	0.56	10.97	0.17	0.86	0.40	17.56	**0.21**	0.86	3.1
	Trained with Judd's dataset				Trained with Bruce's dataset				
GBVS-SM	**0.59**	**5.08**	**0.19**	0.86	0.40	**12.95**	**0.21**	0.85	**1.9**
	Trained with the four eye tracking datasets								
GBVS-SM	0.56	5.05	0.19	0.87	0.40	12.37	0.21	0.85	–
WTA and IoR	0.47	5.51	0.15	**0.88**	0.26	17.09	0.14	0.79	4.8
Without oculomotor constraints	0.42	12.51	0.13	0.81	0.32	19.2	0.17	0.81	6.8
Boccignone and Ferraro (2004)	0.36	7.45	0.13	**0.88**	0.22	14.49	0.13	0.86	6.0

Bold values represent the highest performance (one bold value per column).

Saliency map and randomness (1/2)

➟ Influence of the saliency map:

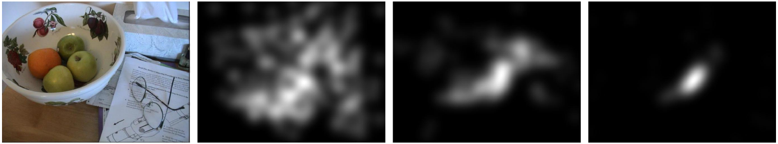
Table 3
Performance of different models over Bruce's datasets using the DTW metric, the linear correlation coefficient, KL-divergence, hit rate and the normalized scanpath saliency (NSS). Itti-SM, AWS-SM and Top2-SM represent the saliency maps generated by the proposed approach when Itti, AWS and Top2 saliency maps are the input of the proposed saccadic model, respectively.

Method	Bruce's dataset				
	DTW	CC	KL	NSS	HitRate
GBVS-SM	73.80	0.59	5.08	0.19	0.86
Itti-SM	95.34	0.28	12.10	0.09	0.75
AWS-SM	91.18	0.45	8.56	0.14	0.82
Top2-SM	72.57	0.64	4.37	0.20	0.87

Top2-SM: we aggregated the saliency maps of GBVS and RARE2012 models through a simple average. (Le Meur and Liu, 2014) demonstrated that a simple average of the top 2 saliency maps, computed by GBVS and RARE2012 models, significantly outperforms the best saliency models.

Saliency map and randomness (2/2)

- Randomness:



(a) (b) $N_c = 1$ (c) $N_c = 5$ (d) $N_c = 50$

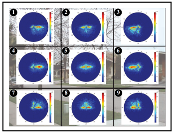
The maximal randomness is obtained when $N_c = 1$.

Limitations of the proposed model

Still far from the reality...

- We do not predict the fixation durations. Some models could be used for this purpose (Nuthmann et al., 2010, Trukenbrod and Engbert, 2014).
- Second-order effect. We assume that the memory effect occurs only in the fixation location. However, are saccades independent events? No, see (Tatler and Vincent, 2008).
- High-level aspects such as the scene context are not included in our model.
- Should we recompute the saliency map after every fixations? Probably yes...
- Randomness (N_c) should be adapted to the input image. By default, $N_c = 5$.
- Is the time course of IoR relevant? Is the recovery linear?
- Foveal vs peripheral vision? Cortical magnification...

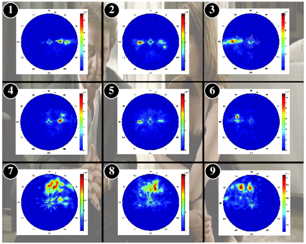
Extensions (1/2)



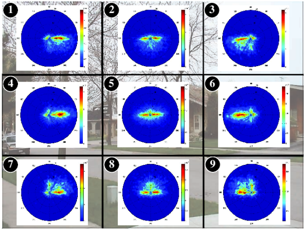
O. Le Meur & A. Coutrot, Introducing context-dependent and spatially-variant viewing biases in saccadic models, Minor Revision in Vision Research.

Spatially-variant and context dependent joint distribution $p_B(d, \phi, \mathbf{x})$

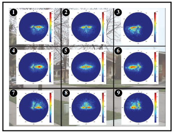
Conversational videos



Natural scenes



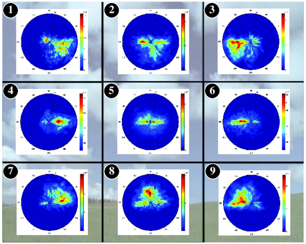
Extensions (2/2)



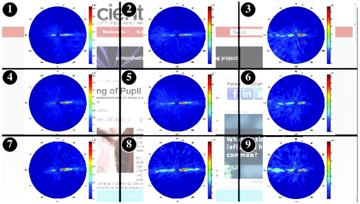
O. Le Meur & A. Coutrot, Introducing context-dependent and spatially-variant viewing biases in saccadic models, Minor Revision in Vision Research.

Spatially-variant and context dependent joint distribution $p_B(d, \phi, \mathbf{x})$

Landscapes



Webpages



6 Conclusion

Conclusion (1/2)

Two contributions:

- A new saccadic model performing well to:
  - produce plausible visual scanpaths;
  - detect the most salient regions of visual scenes.
- Signature of viewing tendencies. This signature is spatially-variant and context-dependent;

Conclusion (2/2)

Future works:

- Dealing with the limitations of the current implementation;
- Spatio-temporal signature of viewing tendencies:
  - for healthy people (according to gender, sex...);
  - for visually impaired people (use eye-movement to detect degenerative diseases).
- Longitudinal studies from childhood to adulthood.

Better signature of viewing tendencies can be used to screen mental health... (see (Itti, 2015))

References

P. J. Bennett and J. Pratt. The spatial distribution of inhibition of return:. *Psychological Science*, 12:76–80, 2001.

G. Boccignone and M. Ferraro. Modelling gaze shift as a constrained random walk. *Physica A: Statistical Mechanics and its Applications*, 331(1Ű2):207 – 218, 2004. ISSN 0378-4371. doi: http://dx.doi.org/10.1016/j.physa.2003.09.011.

A. Borji and L. Itti. State-of-the-art in visual attention modeling. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 35: 185–207, 2013.

A. Borji, D. N. Sihite, and L. Itti. Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Transactions on Image Processing*, 22(1):55–69, 2012.

S. R. Ellis and J. D. Smith. *Patterns of statistical dependency in visual scanning*, chapter Eye Movements and Human Information Processing, pages 221–238. Elsevier Science Publishers BV, (eds) Amsterdam, North Holland Press, 1985.

L. Gupta, D. L. Molfese, R. Tammana, and P. G. Simos. Nonlinear alignment and averaging for estimating the evoked potential. *IEEE Transactions on Biomedical Engineering*, 43(4):348–356, 1996.

J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *Proceedings of Neural Information Processing Systems (NIPS)*, 2006.

L. Itti. New eye-tracking techniques may revolutionize mental health screening. *Neuron*, 88(3):442–444, Nov 2015.

Laurent Itti and Pierre F Baldi. Bayesian surprise attracts human attention. In *Advances in neural information processing systems*, pages 547–554, 2005.

H. Jarodzka, K. Holmqvist, and K. Nystrom. A vector-based, multidimensional scanpath similarity measure. In *ETRA*, pages 211–218, 2010.

T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where people look. In *ICCV*, 2009.

T. Judd, F. Durand, and A. Torralba. A benchmark of computational models of saliency to predict human fixation. Technical report, MIT, 2012.

C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4: 219–227, 1985.

O. Le Meur and T. Baccino. Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior Research Method*, 45(1):251–266, 2013.

O. Le Meur and Z. Liu. Saliency aggregation: Does unity make strength? In *ACCV*, 2014.

O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau. A coherent computational approach to model the bottom-up visual attention. *IEEE Trans. On PAMI*, 28(5):802–817, May 2006.

M. Mancas, C. Mancas-Thillou, B. Gosselin, and B. Macq. A rarity-based saliency map - application to texture description. In *ICIP*, 2006.

J. Najemnik and W.S. Geisler. Eye movement statistics in humans are consistent with an optimal strategy. *Journal of Vision*, 8(3): 1–14, 2008.

J. Najemnik and W.S. Geisler. Simple summation rule for optimal fixation selection in visual search. *Vision Research*, 42: 1286–1294, 2009.

A. Nuthmann, T. J. Smith, R. Engbert, and J. M. Henderson. CRISP: A Computational Model of Fixation Durations in Scene Viewing. *Psychological Review*, 117(2):382–405, April 2010. URL http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ884784.

D. Parkhurst, K. Law, and E. Niebur. Modelling the role of salience in the allocation of overt visual attention. *Vision Research*, 42: 107–123, 2002.

R. J. Peters and L. Itti. Applying computational tools to predict gaze direction in interactive visual environments. *ACM Transactions on Applied Perception*, pages 1–21, 2008.

R. J. Peters, A. Iyer, L. Itti, and C. Koch. Components of bottom-up gaze allocation in natural images. *Vision Research*, 45(18): 2397–2416, 2005.

M. I. Posner. Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32:3–25, 1980.

C. M. Privitera and L. W. Stark. Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(9):970–982, 2000.

N. Riche, M. Mancas, M. Duvinage, M. Mibulumukini, B. Gosselin, and T. Dutoit. Rare2012: A multi-scale rarity-based saliency detection with its comparative statistical analysis. *Signal Processing: Image Communication*, 28(6):642 – 658, 2013. ISSN 0923-5965. doi: http://dx.doi.org/10.1016/j.image.2013.03.009.

L. W. Stark and S. R. Ellis. *Scanpaths revisited: cognitive models direct active looking*, pages 193–226. Lawrence Erlbaum Associates, Hillsdale, NJ, 1981.

B.W. Tatler and B.T. Vincent. Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, 2:1–18, 2008.

B.W. Tatler, R. J. Baddeley, and I.D. Gilchrist. Visual correlates of fixation selection: effects of scale and time. *Vision Research*, 45:643–659, 2005.

Hans A Trukenbrod and Ralf Engbert. Icat: A computational model for the adaptive control of fixation durations. *Psychonomic bulletin & review*, 21(4):907–934, 2014.