VISUAL ATTENTION SACCADIC MODELS

Taking into account global scene context and temporal aspects of gaze behaviour

Antoine Coutrot[a]* & Olivier Le Meur[b]

[a] CoMPLEX, University College London, United Kingdom
[b] IRISA, University of Rennes 1, France
* corresponding author: a.coutrot@ucl.ac.uk

Introduction

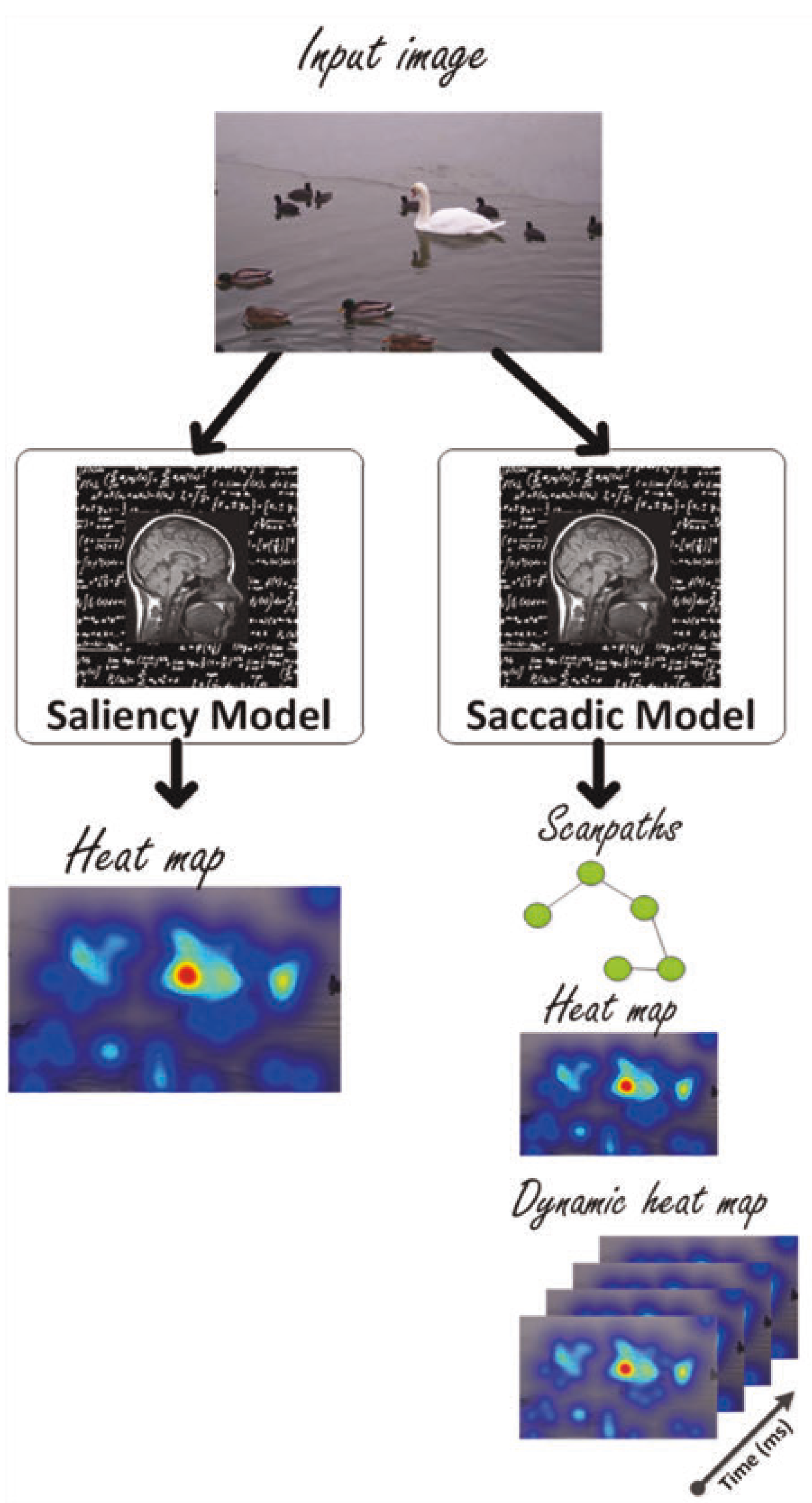

Figure 1 - Conceptual difference between classic saliency models and saccadic models. Classic saliency models output a 2D static saliency map (or heatmap) whereas saccadic models compute visual scanpaths from which static as well as dynamic saliency maps can be computed.

Saccadic models output plausible visual scanpaths, i.e. having the same peculiarities as human scanpaths.

For a *probabilistic tour of visual attention models*, cf. Boccignone's review, arXiv:1607.01232, 2016.

How do saccadic models work?

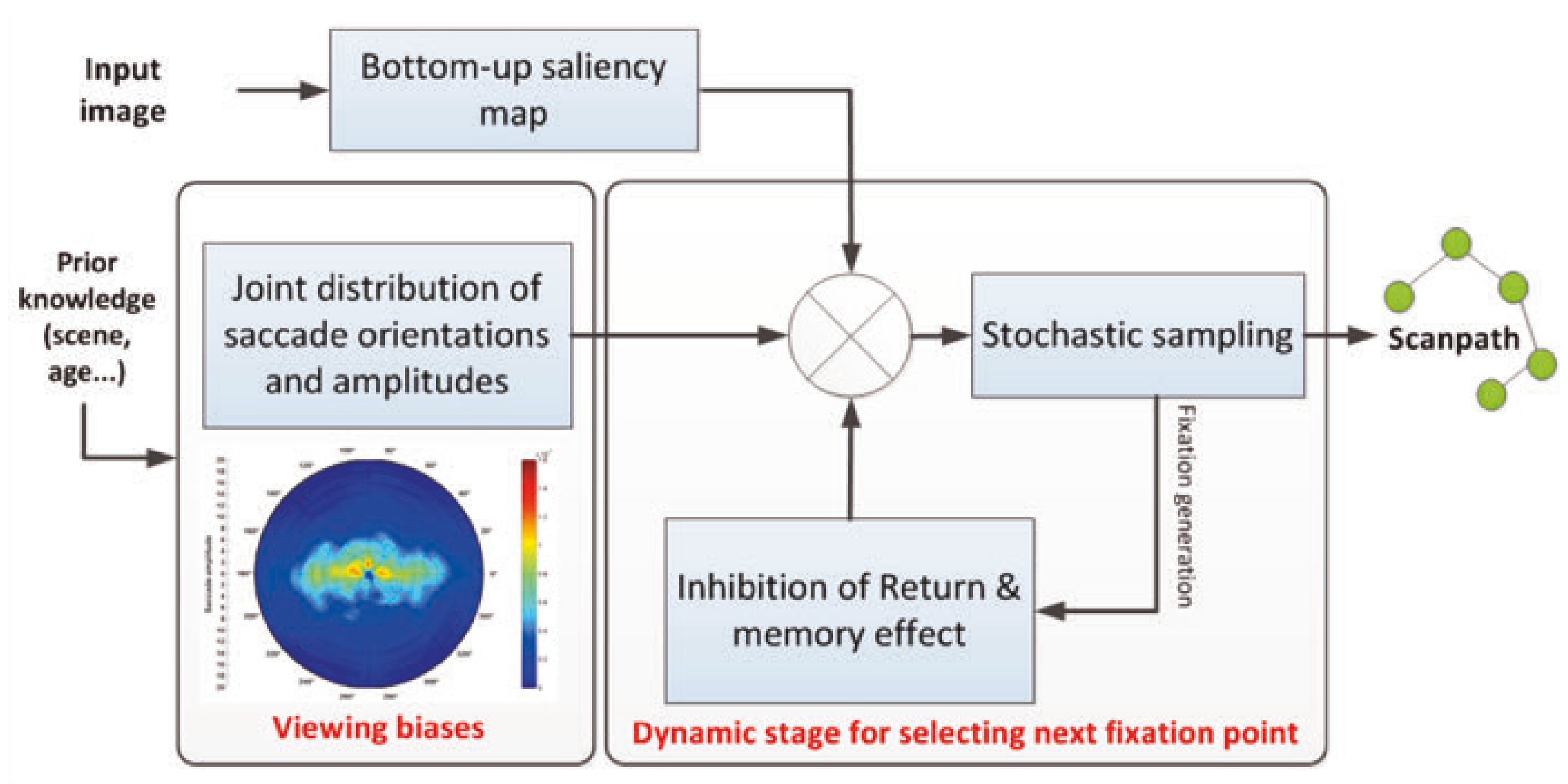

Figure 2 - Saccadic model flow chart. Predicted scanpaths result from the combination of three components: 1- bottom-up saliency map, 2- viewing biases and 3- memory mechanism.

Let x_{t-1} be a fixation point at time t-1. The next fixation point x_t is determined by sampling the 2D discrete conditional probability $\mathbf{p}(x \mid x_{t-1})$

$$\mathbf{p}(x \mid x_{t-1}) = \mathbf{p_{BU}}(x) \cdot \mathbf{p_B}(d(x,x_{t-1}), \phi(x,x_{t-1})) \cdot \mathbf{p_M}(x,t \mid T)$$

To implement the stochastic nature of visual exploration, Nc points are randomly drawn from $\mathbf{p}(x \mid x_{t-1})$. The next fixation point corresponds to the highest value.

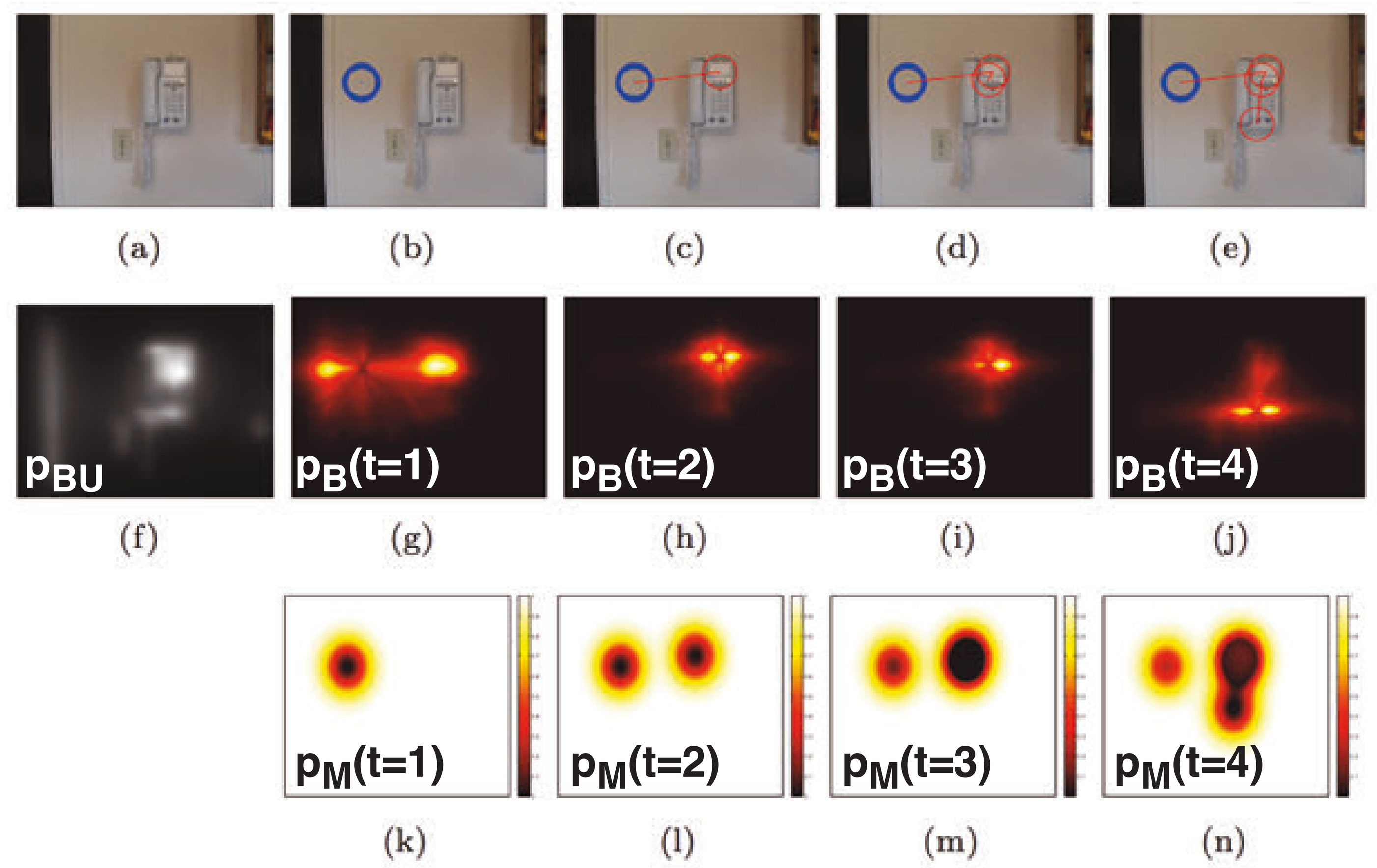

Figure 3 - Scanpath generation. (a) Original image and (f) its saliency map computed by the GBVS model (Harel *et al.*, 2006). (b)–(e) Sequence of fixations (g)–(j) Temperature plot of the joint probability p_B weighted by the saliency p_{BU}. (k)–(n) Temperature plot of the memory effect and inhibition of return p_M . Adapted from [1].

Saccadic models can be easily tuned to emulate a specific visual behavior.

For instance the joint distribution $\mathbf{p_B}$ can be adapted to the semantic visual category of the stimulus.

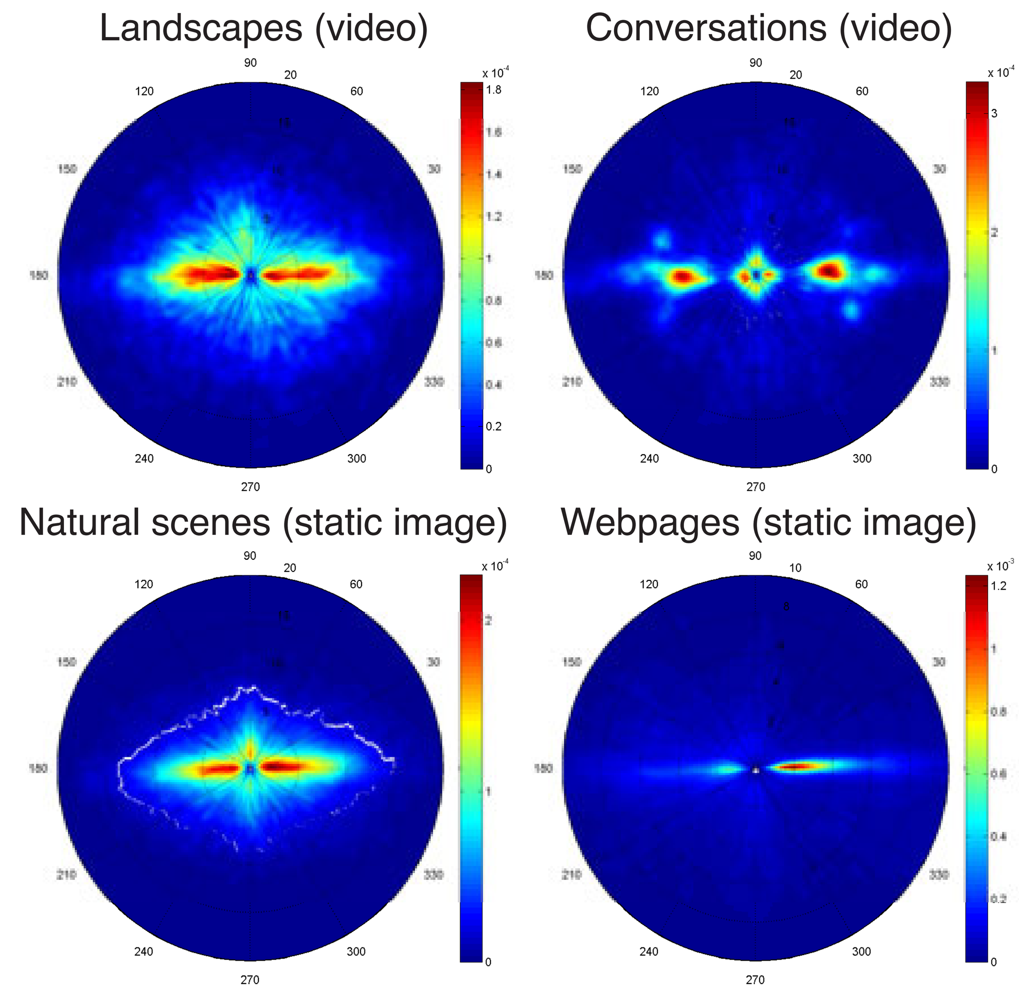

Figure 4- Joint distribution of saccade orientations and amplitudes according to 4 visual categories of stimulus. Landscapes and conversations videos are from [3]; natural static scenes are from [4,5,6]; webpages are from [7].

$\mathbf{p_B}$ can also be adapted to be spatially-variant. This allows to naturally take into account the center bias.

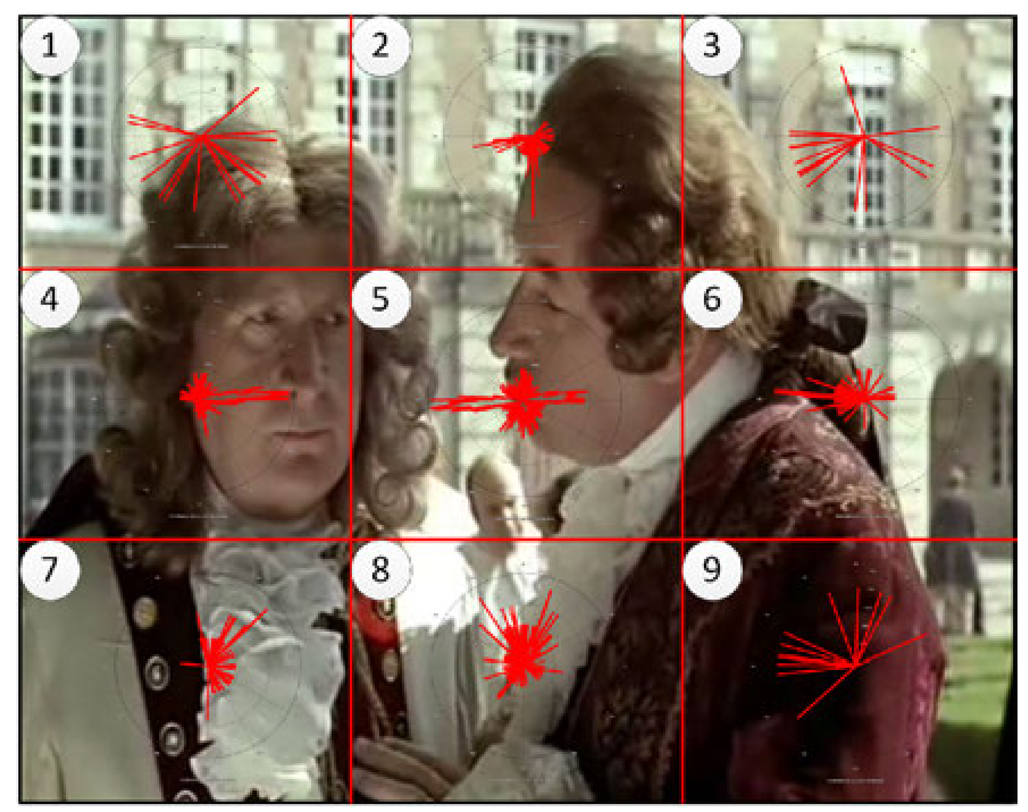

Figure 5 - p_B is spatially-variant. Distributions of saccade orientations and amplitudes are computed over each base frame (1-9). Extracted from [2].

Model Evaluation

Evaluation is performed over Bruce's dataset (120 natural images) [4]. For each image, 100 scanpaths of 10 fixations are computed from the saccadic model, and added up into a heatmap. We repeat the operation with the $\mathbf{p_B}$ distribution from 4 visual categories. These heatmaps are compared with the output of a classic bottom-up only saliency model (a combination of [8] and [9]).

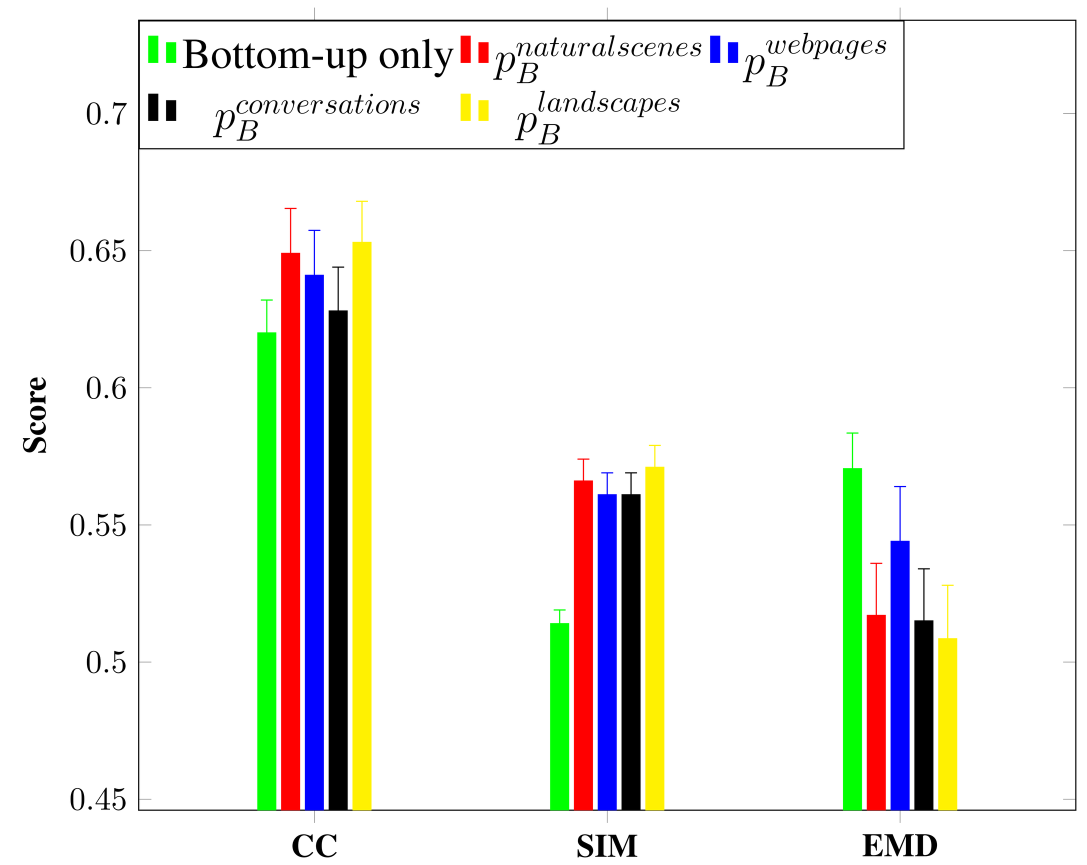

Figure 6- Evaluation is performed with 2 similarity metrics (CC and SIM) and one dissimilarity metric (EMD). EMD has been scaled down by a factor 4. Error bars represent standard errors.

Conclusions

Visual attention saccadic models take into account the temporal dimension of visual exploration. They provide an efficient framework to integrate in a data-driven fashion variables as different as bottom-up saliency, spatial bias, context and scene composition, as well as oculomotor constraints. They will allow to tailor saliency model for specific populations (e.g. for different age groups, tasks, states of health...).

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

References

[1] Le Meur O, Liu Z. Vision Research 2015.
[2] Le Meur O, Coutrot A. Vision Research 2016.
[3] Coutrot A, Guyader N. Journal of Vision 2014.
[4] Bruce, N. D., & Tsotsos, J. K. Journal of vision 2009.
[5] Kootstra, G., de Boer, B., & Schomaker, L. R. Cognitive computation 2011.
[6] Judd T, Ehinger K, Durand F, Torralba A. IEEE International Conference on Computer Vision 2009.
[7] Shen, C., & Zhao, Q. European Conference on Computer Vision 2014.
[8] Harel J, Koch C, Perona P. Advances in Neural Information Processing Systems 2006.
[9] Riche N, Mancas *et al.*, Signal Processing : Image Communication 2013.