Predicting saliency using two contextual priors: the dominant depth and the horizon line

Olivier Le Meur (olemeur@irisa.fr)

Abstract: A computational model of visual attention using two visual inferences is proposed. The dominant depth and the horizon line position are inferred from low-level visual features. This prior knowledge helps to find salient areas on still color pictures. Regarding the dominant depth, the idea is to favor the lowest spatial frequencies on close-up scenes whereas the highest spatial frequencies are used to predict salient areas on panoramic view. Some studies showed that the horizon line is a natural attractor of our gaze. Horizon detection is then used to improve the saliency prediction. Results show that the proposed model outperforms existing approaches. However, the dominant depth does not bring any gain in the saliency prediction.

Objective

To design a saliency model based on bottom-up features and two contextual priors (depth and horizon line).

Motivation

Eye movements can be guided by various types of information in real-world scenes

Bottom-up information
low-level saliency

Visual inferences based on higher level information
scene context, depth, horizon line…

Task-dependent

Models

A. Bottom-Up Saliency: the proposed computational model is similar to [1].

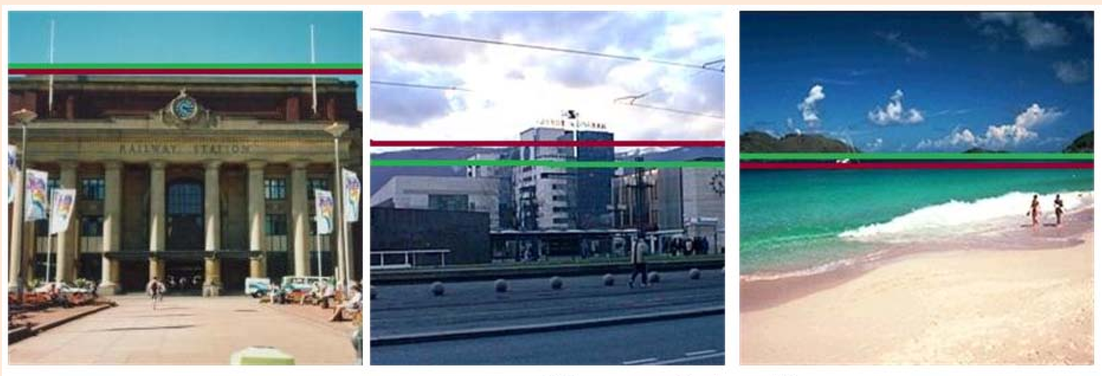
B. Dominant depth estimation:

- Training based on a blockification of subbands
- Training data set composed of 1380 pictures [2]

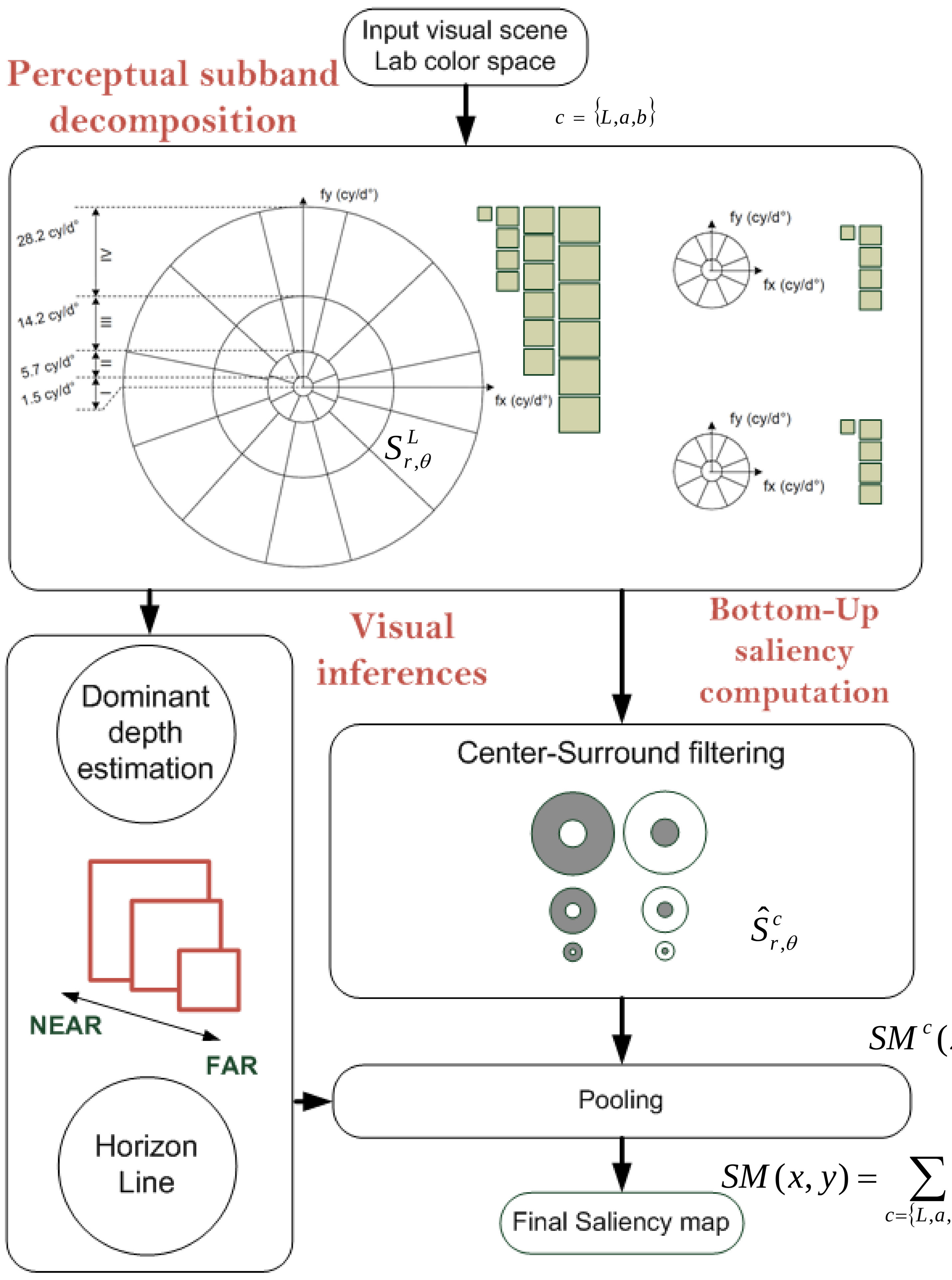
D=3.5	D=3.94	D=1.93	D=3.38
D_{est}=4.34	D_{est}=4.15	D_{est}=2.06	D_{est}=4.95

C. Horizon line position estimation:

- Training based on a blockification of subbands
- 213 natural outdoor visual scenes



Ground truth
Prediction



More examples on author's website

$$SM^{c}(x,y) = \sum_i \alpha_i \max_\theta \left(\hat{S}_{i,\theta}(x,y)\right)$$

$$SM(x,y) = \sum_{c=\{L,a,b\}} \sum_{p,m} d\left(SM^{c}(x,y), SM^{c}(p,m)\right)$$

Pooling

- **Depth-based pooling:**

Hypothesis: for a close-up scene, we favor the lowest spatial frequency whereas, for a panoramic scene, the highest spatial frequency are favored.

→ in function of the estimated depth, the computation of the saliency map is performed on a limited range of spatial frequencies.

- **Facilitation based on the spatial position h of the horizon line:**

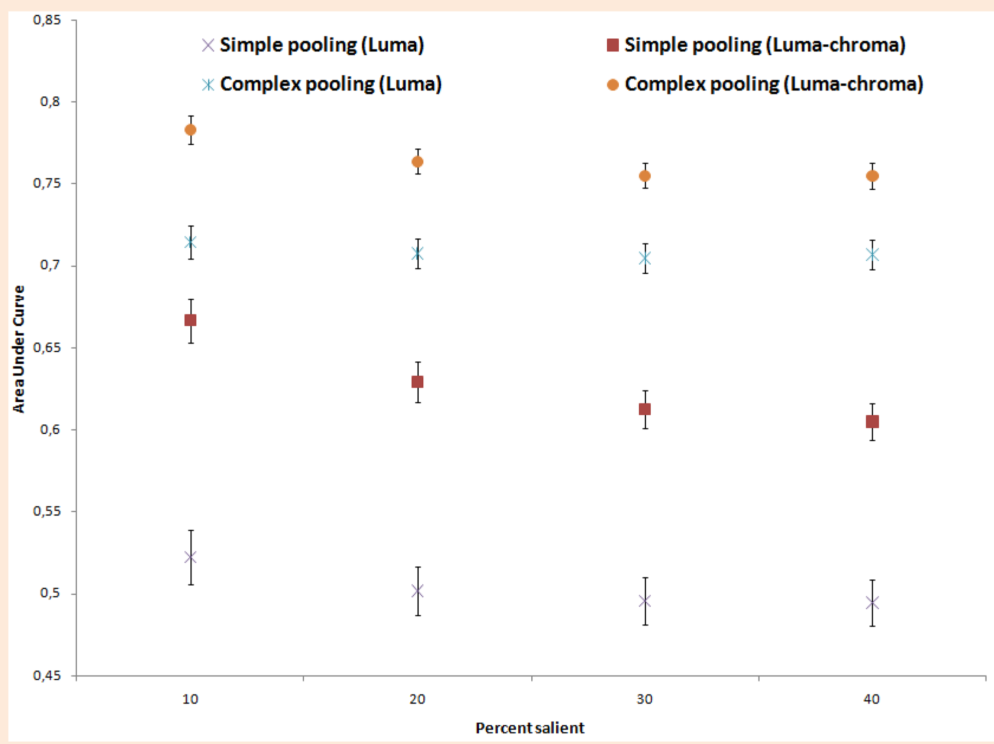
Hypothesis: we favor locations located near the horizon line.

→ Weighting function $W(x,y) = \exp\left(-\frac{(y-h)^2}{2\sigma_h^{\ 2}}\right)$

Performance on Bruce's dataset

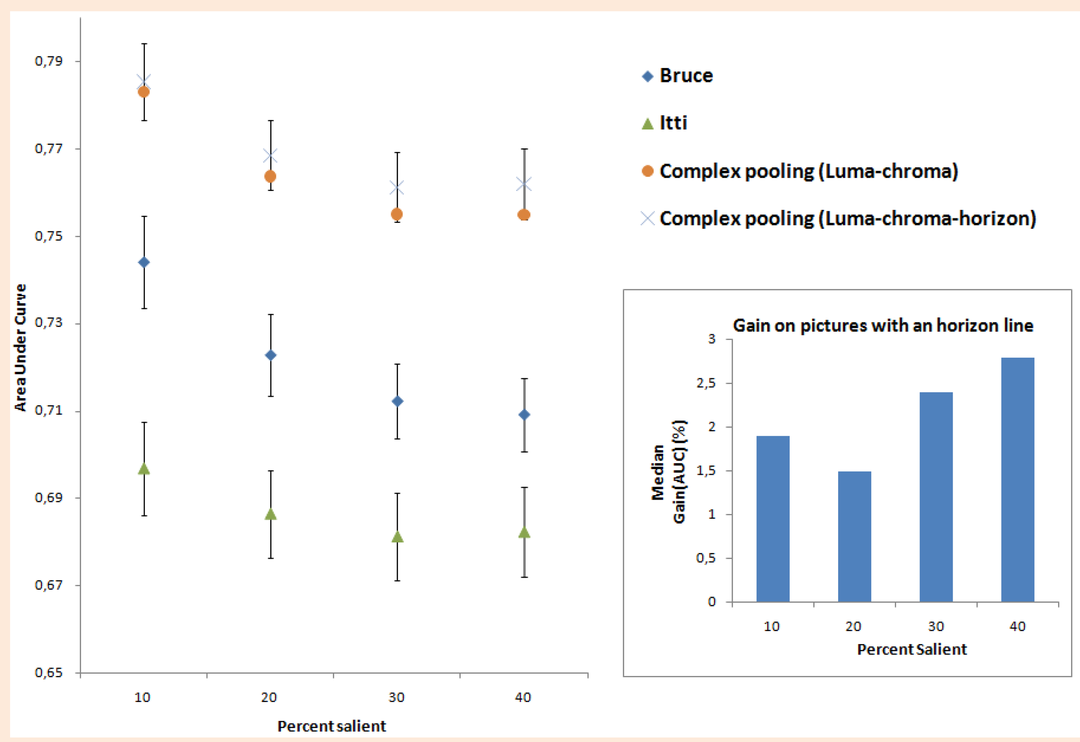
- Importance of the pooling strategy
→ Simple linear combination
→ Complex combination based on the sum of dissimilarities between a point and all other points.

- Importance of the chromatic component



- Combined source model (bottom-up saliency + horizon line)

- Comparison with state-of-the-art approaches



Conclusion: In the proposed model, saliency is based on low-level visual features combined with the extraction of global features. They provide layout information and contextual priors to bias the saliency map to certain locations. We use machine learning to train models to infer the dominant depth and the position of the horizon line. These inferences are based on the low-level visual features. The proposed model is compared to both purely bottom-up model and existing models. We found that the dominant depth doesn't bring any improvement when compared to a naïve model. Regarding the horizon line, the median gain is of 2% on pictures for which there is an horizon line.

[1] 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, vol. 28, pp. 802–817, 2006.
[2] M.G. Ross and A. Oliva, "Estimating perception of scene layout properties from global image features," Journal Of Vision, vol. 10, no. 1, pp. 1–25, 2010.

ICME 2011