Traditional Saliency Reloaded: A Good Old Model in New Shape

Simone Frintrop, Thomas Werner, and Germán Martín García
Department of Computer Science III, University of Bonn, Germany.

Abstract: We show in this paper that the seminal, biologically-inspired saliency model by Itti et al. [3] is still competitive with current state-of-the-art methods for salient object segmentation if some important adaptions are made. We show which changes are necessary to achieve high performance, with special emphasis on the scale-space: we introduce a twin pyramid for computing Difference-of-Gaussians, which enables a flexible center-surround ratio. The resulting system, called VOCUS2, is elegant and coherent in structure, fast, and computes saliency at the pixel level. Some example saliency maps are shown in Fig. 1.

Figure 1: Pixel-precise (middle) and segment-based (right) saliency maps of our VOCUS2 saliency system

System Overview: Fig. 2 shows an overview over the VOCUS2 saliency system. The basic structure is the same as in other systems based on the psychological Feature Integration Theory [5], e.g. Itti's iNVT [3] or our previous VOCUS system [1]: feature channels are computed in parallel, pyramids enable a multi-scale computation, contrasts are computed by Difference-of-Gaussians. The main difference to the above systems is that we introduce a new twin pyramid that enables a flexible center-surround ratio to compute feature contrasts. Instead of computing Difference-of-Gaussians by subtracting layers of the same pyramid, we compute a center and a surround pyramid separately and compute Difference-of-Gaussian contrasts based on these maps. This enables to chose the center-surround ratio in a flexible way instead of being restricted to sigmas available in the pyramid. Since the center-surround ratio is the most crucial parameter of saliency systems, this change has a large effect on the performance. Other changes include a different color space and a different fusion of the channels. Fig. 3 shows how each change with respect to the iNVT improves the performance on the MSRA-1000 dataset.

An additional location prior can be added optionally to the system for specific applications and benchmarks. We added simply a wide Gaussian, centered at the image, which improved the performance on several benchmarks considerably (VOCUS2-LP).

Additionally, we extended the method to obtain segment-based saliency maps by combining the saliency map with a generic object proposal detection method. The resulting object proposals are integrated into a segment-based saliency map (VOCUS2-Prop, cf. Fig. 1, right).

Results: In the full paper, we show results on the MSRA-10k, ECSSD, SED1, SED2, and PASCAL-S datasets and show that our method is competitive with state-of-the-art methods.

Since the system does not rely on center or background priors (although they can be integrated if desired), it is especially well suited to be applied to complex scenes as obtained from mobile devices such as Google Glass or autonomous robots. The second row of Fig. 1 shows an example of such a scene and the corresponding saliency maps. In [2] and [4], we show how such saliency maps can be used for object discovery on mobile systems.

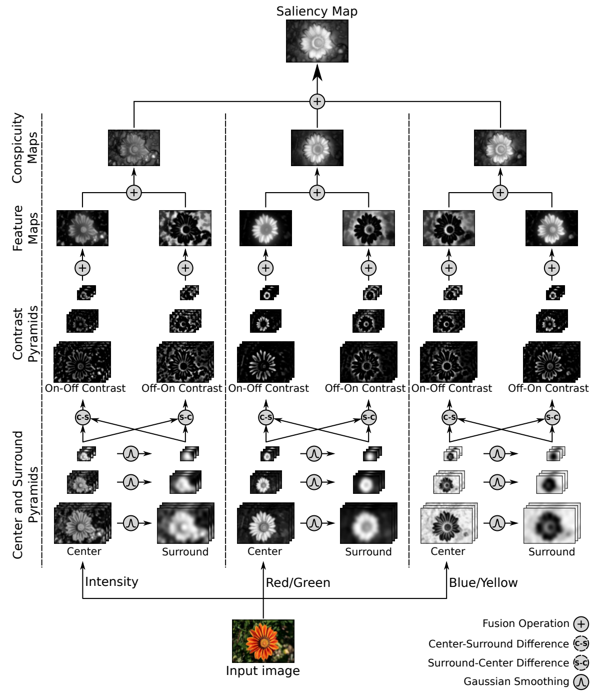

Figure 2: Overview of our saliency system VOCUS2.

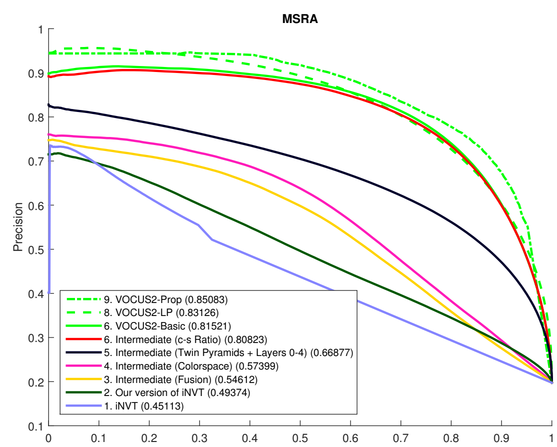

Figure 3: Stepwise improvements of Itti's iNVT saliency system [3] until reaching VOCUS2. AUC values in parentheses.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

[1] Simone Frintrop. *VOCUS: A Visual Attention System for Object Detection and Goal-directed Search*, volume 3899 of *LNAI*. Springer, 2006.

[2] Esther Horbert, Germán Martín García, Simone Frintrop, and Bastian Leibe. Sequence-level object candidates based on saliency for generic object recognition on mobile systems. In *ICRA*, 2015.

[3] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *TPAMI*, 20(11), 1998.

[4] Germán Martín-García, Ekaterina Potapova, Thomas Werner, Michael Zillich, Markus Vincze, and Simone Frintrop. Saliency-based object discovery on RGB-D data with a late-fusion approach. In *ICRA*, 2015.

[5] Anne M. Treisman and Garry Gelade. A feature integration theory of attention. *Cog. Psych.*, 12, 1980.