Twin-Pyramids for Saliency Computation and the Application to Object Proposal Detection

Simone Frintrop, Thomas Werner, and Germán M. García
Institute of Computer Science III
Rheinische Friedrich-Wilhelms-Universität Bonn, Germany
frintrop@iai.uni-bonn.de

This short paper summarizes our recent work on saliency computation [1] and its application for object proposal detection [4, 2].

Saliency Computation with VOCUS2: We introduced in [1] the new saliency system VOCUS2[1], which follows the traditional, biologically-inspired saliency architecture iNVT that was introduced 17 years ago by Itti and colleagues [3]. We have shown that this seminal concept, that follows findings from human perception, is still capable to obtain state-of-the-art performance on current benchmarks for salient object segmentation, if adapted appropriately.

Especially important is the scale space structure: 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. Fig. 1 shows example saliency maps, and Fig. 2 shows the overview over the system, details in [1]. Additionally, Fig. 3 shows how each of our changes with respect to the iNVT improve the performance on the MSRA-1000 dataset. In [1], we show additional results on the MSRA-10k, ECSSD, SED1, SED2, and PASCAL-S dataset and show that our fast and elegant 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 (cf. Fig. 1, bottom). In the following section, we will present an example for such applications.

Object Proposal Generation: The saliency system produces pixel-precise saliency maps, which can serve as starting point for object proposal detection. The idea is that each peak in the saliency map is likely to correspond to an object. Therefore, we extract salient blobs by seeded region growing that starts at the maxima of the saliency map. To obtain precise object boundaries, we compute superpix-

[1] Code: http://www.iai.uni-bonn.de/∼frintrop/vocus2.html

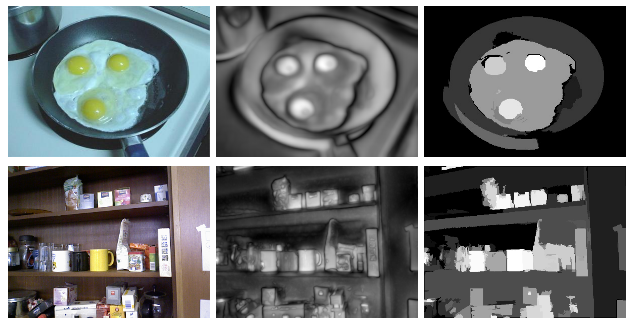

Figure 1. Examples saliency maps. Left: image, middle: VOCUS2 saliency maps, VOCUS2 proposal saliency maps. Fig. from [1].

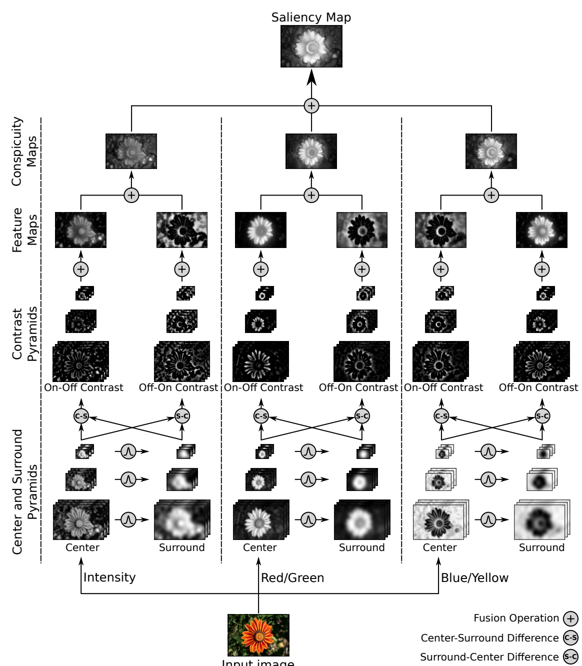

Figure 2. General structure of our saliency system VOCUS2. Fig. from [1].

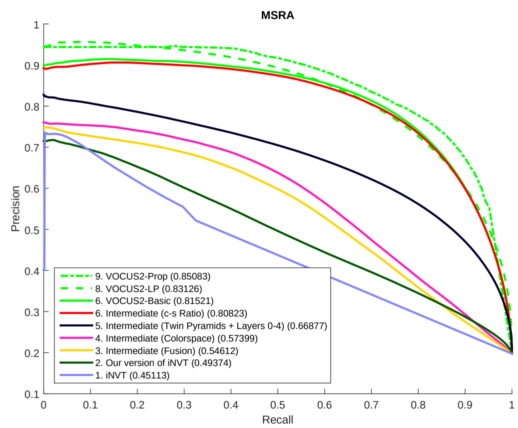

Figure 3. Stepwise improvements of Itti's iNVT saliency system [3], until finally obtaining our VOCUS2 system. Evaluation on MSRA-1000 dataset, AUC values in parentheses. Details in [1].

els with a standard segmentation method (we experimented with MeanShift in [1] and with the segmentation of Felzenszwalb & Huttenlocher in [4, 2]) and combine them with the saliency map to obtain object proposals. For each salient blob, all segments are assembled that overlap with the blob to a certain amount.

In [4] we show how this method can be extended to RGB-D data to exploit the fact that color and depth are complementary data channels. Surface clustering on normals enables a segmentation in the depth data. Several shape measures, such as convexity, are used to determine the quality of the proposals and to obtain a better ranking. This is especially important for real-time systems that aim to prioritize processing and focus on a subset of proposals. We evaluated our method with several other approaches on the well-known Washington RGB-D object dataset[2], an example result is shown in Fig. 4.

Figure 4. Object candidates on the Washington dataset, generated from color and depth data. Details in [4].

In [2], we have extended the proposal generation method to video sequences, to obtain sequence-level object candidates by tracking proposals over time. This improves the recall, since tracking helps to sort out bad quality candidates. But more important is an aspect that becomes relevant when using the proposal detection as pre-processing for object recognition. In such a case, ideally only one object view per sequence should be sent as request to a classifier, instead of one request per frame for each object candidate. This is enabled by the sequence-level proposals and the generation of representative views for each tracklet. This strongly reduces the amount of required classifier requests while improving on recall as well as precision. Fig. 5 shows an example result on the Kitchen Object Discovery (KOD) dataset[3].

Figure 5. Correct sequence-level object candidates on the Kitchen Object Discovery dataset. Details in [2].

Combining Saliency and Proposal Generation to Segment-based Saliency Maps: Interestingly, the two described approaches of saliency computation and saliency-based proposal generation can be combined to obtain segment-based saliency maps in a simple and fast manner. In [1], we rank the object proposals by average saliency, perform non-maxima suppression, and overlay the ranked proposals by taking for each pixel the maximum of all proposals covering this pixel. This gives us segment-based saliency maps that perform very well on salient object segmentation benchmarks (see [1]). Two examples of such proposal saliency maps are shown in Fig. 1, right.

References

[1] S. Frintrop, T. Werner, and G. Martín García. Traditional saliency reloaded: A good old model in new shape. In *accepted for the Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, USA, June 2015. 1, 2

[2] E. Horbert, G. Martín García, S. Frintrop, and B. Leibe. Sequence-level object candidates based on saliency for generic object recognition on mobile systems. In *accepted for Proc. of the IEEE International Conference Robotics and Automation (ICRA)*, 2015. 1, 2

[3] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on PAMI*, 20(11), 1998. 1, 2

[4] G. Martín-García, E. Potapova, T. Werner, M. Zillich, M. Vincze, and S. Frintrop. Saliency-based object discovery on RGB-D data with a late-fusion approach. In *accepted for the IEEE Int'l Conf. on Robotics and Automation (ICRA)*, 2015. 1, 2

[2] http://rgbd-dataset.cs.washington.edu/

[3] http://www.mmp.rwth-aachen.de/projects/kod/