Encoding based Saliency Detection for Videos and Images

Thomas Mauthner, Horst Possegger, Georg Waltner, Horst Bischof
Institute for Computer Graphics and Vision, Graz University of Technology.

We present a novel encoding based saliency (EBS) detection method to support human activity recognition and weakly supervised training of activity detection algorithms in videos and salient object detection in images. Recent research has emphasized the need for analyzing salient information in videos to minimize dataset bias or to supervise weakly labeled training of activity detectors. In contrast to previous methods we do not rely on training information given by either eye-gaze or annotation data, but propose a fully unsupervised algorithm to find salient regions within videos.

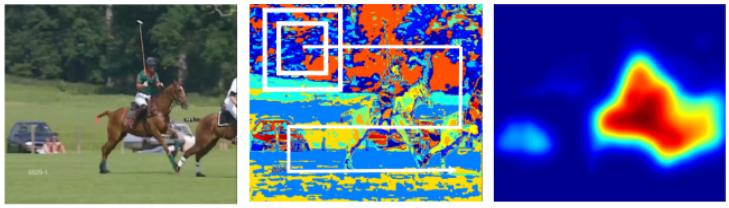
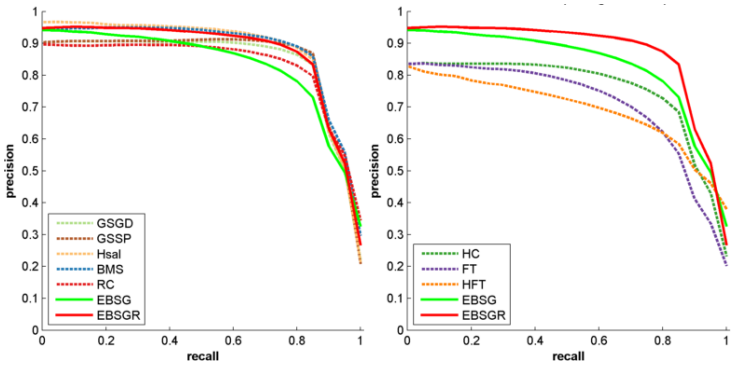


Figure 1: Left: input image. Middle: pixels are represented by a small set of encoding vectors (color coded representation of membership) and saliency is estimated locally (in a sliding-window fashion). Right: final saliency result as a weighted combination of appearance and motion saliency maps.

Our encoding approach allows us to efficiently compute saliency by approximating joint feature distributions per pixel. This approximation of the joint feature distribution is based on analyzing the image or video content, respectively. This efficient representation allows us to scan images on several scales and to estimate foreground distributions locally instead of relying on global statistics only (see Figure 1). In addition, we enforce the Gestalt principle of figure-ground segregation for both appearance and motion cues locally on all scales, without any assumptions on image centered objects. Finally, we propose a measurement of salience quality that allows us to dynamically weight and combine the results of different saliency maps, e.g. appearance and motion.



(a) (b)

Figure 2: Comparisons on the ASD dataset.

We evaluate our approach on several datasets, including challenging scenarios with cluttered background and camera motion, as well as salient object detection in images. On the UCF sports dataset [3], we demonstrate favorable performance compared to state-of-the-art methods in estimating both ground-truth eye-gaze and activity annotations. This dataset depicts challenging scenarios including camera motion, cluttered backgrounds, and non-rigid object deformations. Furthermore, it provides ground-truth bounding box annotations for all activities. We compare against a variety of saliency approaches including Rathu *et al.* [2] and Zhou *et al.* [6]. Exemplar results for the UCF sports dataset are depicted in Figure 3.

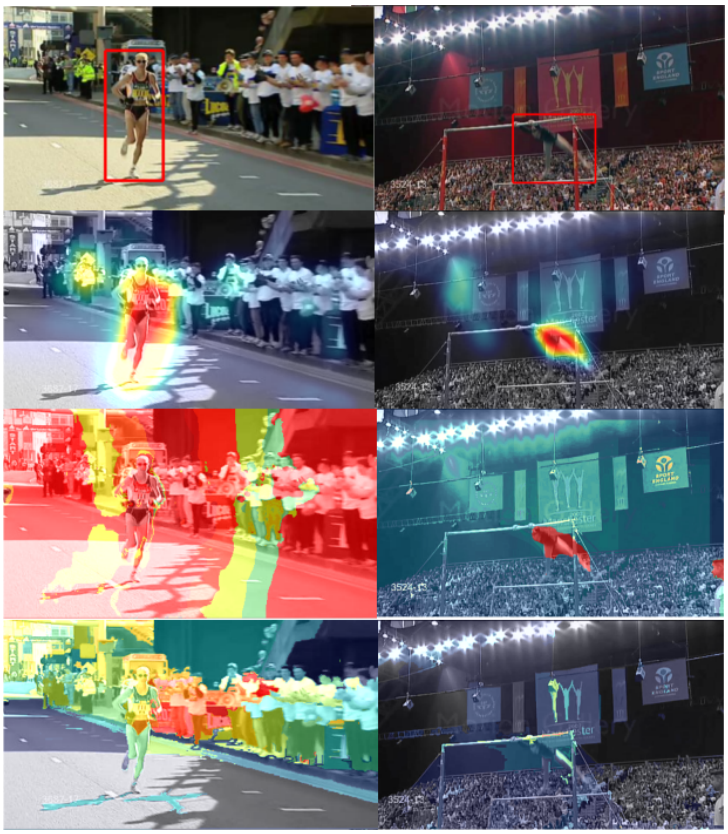


Figure 3: Exemplary results on the UCF sports dataset. From top to bottom: ground truth annotation, our results, Rathu *et al.* [2], and Zhou *et al.* [6].

Although the focus of our work is saliency estimation in activity videos, EBS can easily be applied to standard image saliency tasks by switching off the motion components. We benchmark against recent state-of-the-art approaches such as FT [1], BMS [5], and Hsal [4] on the ASD dataset [1]. Our proposed EBSG and EBSGR methods perform better or equal compared to approaches without explicit segmentation steps and on par in comparison with segmentation-dependent methods or centered object assumptions (see Figure 2). In addition, we compare the robustness of methods concerning off-centered objects within our paper.

[1] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Süsstrunk. Frequency-tuned Salient Region Detection. In *CVPR*, 2009.

[2] Esa Rahtu, Juho Kannala, Mikko Salo, and Janne Heikkilä. Segmenting Salient Objects from Images and Videos. In *ECCV*, 2010.

[3] Mikel D. Rodriguez, Javed Ahmed, and Mubarak Shah. Action MACH - A Spatio-temporal Maximum Average Correlation Height Filter for Action Recognition. In *CVPR*, 2008.

[4] Qiong Yan, Li Xu, Jiangping Shi, and Jiaya Jia. Hierarchical Saliency Detection. In *CVPR*, 2013.

[5] Jiangming Zhang and Stan Sclaroff. Saliency Detection: A Boolean Map Approach. In *ICCV*, 2013.

[6] Feng Zhou, Sing Bing Kang, and Michael F. Cohen. Time-Mapping using Space-Time Saliency. In *CVPR*, 2014.

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