Saliency Density Maximization for Efficient Visual Objects Discovery

Zhang, Liqiang

May 15,2018

1 Maximum Saliency Density

THE STUDIES of psychology and cognitive science have shown that the human perception is attentional and selective [2]. For example, when watching images or videos, not every pixel is of equal importance to us. In most cases, our visual attention mainly focuses on a salient sub-region of an image, e.g. the salient object, and we follow this salient object in an image sequence. The major limitation is that it is troublesome to train the model or difficult to predefine the amount of saliency the salient object should contain, as it depends on the size and the shape of the salient object, as well as how cluttered the background is. To address the above mentioned problems, the author propose a novel unsupervised method to discover salient objects from various saliency maps of images or videos. They propose to locate the object by finding a bounding box of the maximum saliency density (MSD).

Figure 1: The object detection result using the saliency map proposed in [3].

2 Major Contribution

1) The author first formulate the task of the unsupervised salient object detection as the maximum saliency density discovery problem, which well balance the size of the saliency object and the saliency it contains [1]. It is also compatible to different types of saliency maps or a fused saliency map.

2) To obtain the global optimal solution of the saliency density maximization problem, we propose an efficient branch-and-bound search algorithm, which is based on the saliency density rather than the traditional classifi- cation scores. Derivation of the upper-bound estimation and the average convergent time (referred to table II) show its efficiency on salient object detection in both images and videos.

3 Related Work

1) $Minimum Rectangle with Fixed Saliency (MRFS)$: the binary saliency map is adopted and the formulation of salient object detection by M-RFS can be written as in Eq. 1:

$$W* = argminA(h(W)), \qquad (1)$$

where $A(h(W))$ measures the area size of $h(W)$ and S_b is the binary image of S. W is the sub-window of the whole image region I and λ λ is the fixed percentage.

2) $Maximum Saliency Region (MSR)$: the idea of salient object detection in [40] can be formulated as in Eq. 2:

$$h(W) = \sum_{S_b(x,y) \in W} S_b(x,y), \qquad (2)$$

where $S_b(x,y)$ is obtained in a similar way in Eq. 1 with a slight difference that $S_b(x,y) = -1$ when $S(x,y) < \tau$ From Eq. 2, the salient object is located with the region $W*$ that contains the maximum of saliency.

ro=9:16	Frame	Global
precision	73.69%	82.30%
recall	93.36%	86.67%
F?measure	79.26%	83.71%

Table 1: PRECISION, RECALL AND F-MEASURE FOR FRAME SALIENT OBJECT AND GLOBAL SALIENT SUB-IMAGE DETECTION IN TENNIS VIDEO CLIP.

The performance on precision, recall and F-measure for the method in the tennis video is given in table I, from which we can see the detection method can correctly locate the salient object either in a single frame or in a video shot.

References

[1] L Itti. Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Trans Image Process*, 13(10):1304–1318, 2004.

[2] Tie Liu, Jian Sun, Nan Ning Zheng, Xiaoou Tang, and Heung Yeung Shum. Learning to detect a salient object. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8, 2007.

[3] Yu Fei Ma and Hong Jiang Zhang. Contrast-based image attention analysis by using fuzzy growing. In *Eleventh ACM International Conference on Multimedia*, pages 374–381, 2003.