Saliency Detection Based on Frequency and Spatial Domain Analysis

Jian Li[12]
http://www.cim.mcgill.ca/~lijian
Martin D. Levine[2]
levine@cim.mcgill.ca
Xiangjing An[1]
anxiangjing@gmail.com
Hangen He[1]
hehangen@gmail.com

[1] Institute of Automation
National University of Defense Technology
Changsha, PR China

[2] Centre for Intelligent Machines
McGill University
Canada

As a component of low-level vision processing, saliency detection facilitates subsequent processing such as object detection or recognition. In this paper, we argue that a reasonable saliency detector should have the ability to: **(1) Detect both small and large saliency regions.** The size of salient regions vary greatly. When the salient region is large, however, because center-surround algorithms mainly use local information, they will respond heavily in boundary regions, where the texture, intensity or other features are locally different. **(2) Detect saliency in cluttered scenes.** Another drawback of local information-based saliency models is that heavily textured regions are always highlighted. Cluttered scenes are still a challenge for models depending on local information and some based on global information. **(3) Inhibit repeating patterns.** Objects in scenes viewed by the human visual system are thought to compete with each other to selectively focus attention on a subset [5]. These repeating patterns will suppress each other and then be inhibited.

In this paper, inspired by [3, 4], we propose a new saliency detection model by combining global information from frequency domain analysis and local information from spatial domain analysis. In the frequency domain analysis, instead of modeling salient regions, we model the non-salient regions using global information. Thus those so-called repeating patterns that are not distinctive in the scene are suppressed by using spectrum smoothing. In the spatial domain analysis, we enhance those regions that are more informative by using a center-surround mechanism similar to that found in the visual cortex. Finally, the outputs from these two channels are combined to produce the saliency map.

Frequency Domain Analysis Frequency analysis presents an opportunity to deal with the global information in an image. In this paper, we investigate the relationship between the amplitude spectrum and non-salient regions in the image. However, instead of searching for these so-called distinctive patterns, we model the regular patterns (repeating patterns) that would not attract much attention by our visual system. We refer to these as being non-salient. It is argued in [3] that the spectrum residual corresponds to the saliency in an image, while contradictorily in [2], the amplitude information was totally abandoned. However, in this paper, we illustrate that the amplitude spectrum also contains important information corresponding to image saliency. To be more exact, spikes in the amplitude spectrum correspond to repeating patterns, which should be suppressed for saliency detection. A Gaussian kernel h can be employed to suppress spikes in the amplitude spectrum (implemented by a log amplitude spectrum instead of using the amplitude spectrum) as follows:

$$\mathscr{A}_{\mathscr{S}}(u,v) = |\mathscr{F}\{f(x,y)\}| \star h, \tag{1}$$

where h is a Gaussian kernel with a scale σ and $|\mathscr{F}\{f\}|$ is the amplitude spectrum of a signal $f(x,y)$. The resulting smoothed amplitude spectrum $A_S(u,v)$ and the *original* phase spectrum are combined to produce the inverse Fourier Transform, which in turn, yields the saliency map:

$$\mathscr{S} = g \star |\mathscr{F}^{-1}\{\mathscr{A}_{\mathscr{S}}(u,v)e^{i\cdot\mathscr{P}(u,v)}\}|^2. \tag{2}$$

Furthermore, a spectrum scale-space is defined for selecting the best scale of h:

$$\mathscr{L}(u,v;k) = (g(.,.;k) * \mathscr{A})(u,v), \tag{3}$$

given by $k_p = \text{argmin}\,(\text{entropy}(saliencymap(k))$.

Spatial Domain Analysis This approach is used to model salient pixels and regions locally. A center-surround template is usually employed to evaluate the distinctiveness of a local image region by measuring the local contrast. Difference of Gaussian (DOG) and Gabor filters are commonly used to accomplish this. Recently, researchers have also used employed bases as the filters. It has been shown that by training on tens of thousands of natural image patches, the resulting filters turn out to be quite similar to receptive fields found in the visual cortex [1]. In this paper, we take the 192 color features from [4] as the filters and obtain192 response maps. We use entropy to assign a weight to each response map. Given the 192 response maps, we then calculate a weighted sum to obtain a single saliency map. Thus the saliency is defined as:

$$S(x,y) = \sum w_i(f(x,y) * h_i), \tag{4}$$

where the h_i are the local filters obtained by ICA and w_i is given by the following: $w_i = \text{entropy}\,(f(x,y) * h_i)^{-1}$. Unlike the frequency domain analysis which highlights saliency by using global information, spatial domain analysis enhances salient regions that exhibit strong local contrast. Such a "center-surround" model has been adopted in previous work [4].

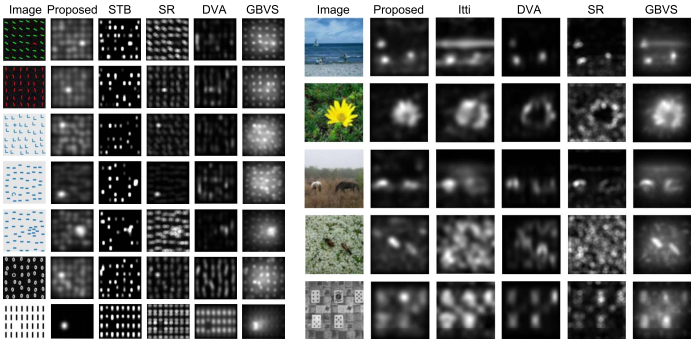

Figure 1: Responses to psychological patterns and natural images

Saliency Map Generation The two processing channels yield two saliency maps, and we denote them as S_g and S_l respectively. Given a color image, it is first represented in an opponent color space: $I = \max\{r,g,b\}$, $RG = r - g$ and $BY = b - \frac{r+g}{2} - \frac{min(r,c)}{2}$. Then a saliency map is computed for each channel.The entropy values for the three best saliency maps are used as weights to produce the final global saliency map S_g. In the case of spatial analysis channel, S_l is computed according to (4). The final saliency map, S_f, is given by: $S_f = S_g + k.\frac{\text{entropy}(S_g)}{\text{entropy}(S_l)}S_l$, where k is a free parameter.

We demonstrate that the proposed model has the ability to highlight both small and large salient regions in cluttered scenes and to inhibit repeating distractors, as shown in Fig.1. Experimental results also show that the proposed model outperforms existing algorithms in predicting objects regions where human pay more attention.

[1] A.J. Bell and T.J. Sejnowski. The "independent components" of natural scenes are edge filters. *Vision research*, 37(23):3327, 1997.

[2] C. Guo, Q. Ma, and L. Zhang. Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In *Proc.* CVPR, 2008.

[3] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In *Proc.* CVPR, 2007.

[4] X Hou and Liqing Zhang. Dynamic visual attention: searching for coding length increments. In *Proc.* NIPS, 2008.

[5] S. Yantis. How visual salience wins the battle for awareness. *Nature neuroscience*, 8(8):975–977, 2005.