FREQUENCY TUNED SALIENT EDGE DETECTION

Yusuf Saber[1] , Matthew Kyan[2]

Ryerson University
Department of Electrical and Computer Engineering
350 Victoria Street, Toronto, Ontario, Canada M5B 2K3

[1]ysaber@ee.ryerson.ca, [2]mkyan@ee.ryerson.ca

ABSTRACT

This paper presents a novel method for salient edge detection based on the frequency tuning principle. Limiting the operation to a single wavelet sub-band, the frequency components of the wavelet are analyzed. The average value in each frequency component (horizontal, vertical, and diagonal) is computed and subtracted from each pixel in its respective component. This creates a variance which allows the salient edges to pop out more significantly. Reconstructing the wavelet using only the frequency components yields a strong response to salient edges. Comparisons are given to classic salient edge detection methods.

Index Terms— Saliency, visual attention, edge detection, spectral residual, wavelets.

1. INTRODUCTION

According to William James, the father of American psychology, a two component framework is implemented in the human visual system (HVS) for attentional deployment[1]. This framework suggests that the HVS directs its attention to an object based on bottom-up cues as well as top-down cues. Bottom-up cues are those that unintentionally grab one's attention, an example of which is face detection. Top-down cues are those that are intentionally looked for, an example of which is face recognition. Hence, the top-down attention model requires prior information, while the bottom-up model does not. Because of this, much research has been explored into modeling the bottom-up concept of attention. To display points of attention, a two dimensional function is developed which maximizes where the attention exists the most. This function is known as a saliency map. In the digital realm, saliency maps are used to determine points of attention in a scene. Finding these points is an important precursor to many applications including segmentation, object recognition, tracking, indexing and image/video informatics. In terms of a human's gaze, the location on the saliency map that yields the highest energy is where the gaze would unintentionally be directed.

Although various methods exist to extract saliency, they all have the same fundamental backbone: to compare local information with global information. The difference between the local and global is directly correlated to the local region's saliency, where pixels that are largely different than the global representation yield a higher saliency. Some of the more basic methods try to extract salient pixels by computing some representation of the overall image and subtracting each pixel in the image from that value [2]. Equation (1) gives this formula, where α is the representation of the entire image (could be the average pixel value of the image for example), s_p is the saliency of pixel p, and i_p is the intensity of pixel p.

$$s_p = |i_p - \alpha| \tag{1}$$

One of the most prominent methods is a biologically inspired one that traverses the image and spatially compares a small local region to a larger local region; doing so with respect to various features of the image such as intensity, colour, orientation, hue, and/or texture[3]. Although this method is quite robust, it requires a lot of processing time. Figure 1b shows the type of saliency map that is extracted by Itti et al.'s method.

Another prominent method extracts sparsely-occurring frequencies from the image[4]. This yields a saliency map which presents high levels of saliency at the edges of the attentive objects as is shown in Figure 1c. The method proposed in this paper further explores the process of extracting certain frequencies as a method of extracting salient edges.

(a) Original (b) Salient regions (c) Salient edges

Fig. 1: Types of saliency maps

In Figure 1b, the saliency map showing salient regions is yielded by Itti et al's method[3]. It is a very low-resolution

CCECE 2011 Niagara Falls, Canada
978-1-4244-9789-8/11/$26 ©2011 IEEE

map containing a blurry blob suggesting the attentive regions. Hou's method [4] gives an edge map of the region of attention (Figure 1c). Each of these maps serves a purpose. Salient region maps are useful for locating focal points in a human gaze and can be used in applications such as robotic vision, where an exact outline of the object of interest is not necessary. Salient edge maps on the other hand can be useful for applications that require a more precise boundary of the region or object of attention, such as image segmentation.

The rest of this paper is organized as follows: Section 2 discusses, in detail, some of the relevant previous work in the literature of saliency maps. Section 3 discusses the proposed method as well as the basis of its construction. Section 4 shows some results and comparisons to previous methods, showing the significance of the proposed method. Section 5 gives some concluding remarks.

2. PAST APPROACHES

2.1. Itti's Method

Itti et al.'s method [3] stands out amongst the rest because it is the first to put the feature integration theory [5] into practise. The feature integration theory states that the HVS scans various features for attentive points, and then combines all the points to provide a perception of attention.

Itti first extracts features from the image (intensity, colour, and orientation, among others) and then performs a center-surround operation on each of the features [3]. The center-surround operation is the basis for all bottom-up attention models. It is used to compare local regions to larger, comparably global, regions. The center-surround operation, when applied to a feature map, yields a conspicuity map of the attentive regions in the image. Itti proposes that many scales be analyzed, therefore all of the conspicuity maps for all of the scales must be normalized and combined to form a final saliency map.

2.2. Frequency Tuned Method

Achanta et al [6] proposed a method to extract high resolution salient regions. They placed five requirements for their method. They are:

1. Emphasize the largest salient objects
2. Uniformly highlight whole salient regions
3. Establish well-defined boundaries of salient objects
4. Disregard high frequencies arising from texture, noise, and blocking artifacts
5. Efficiently output full resolution saliency maps

These requirements work well for the purpose of salient region detection [6]. However, for the purpose of salient edge detection, we only need requirements 3, 4, and 5 to be met; this will be discussed in Section 3.

To meet all of the aforementioned requirements, Achanta et al define their saliency map function as:

$$S(x, y) = |I_\mu - I'(x, y)| \tag{2}$$

where I_μ is the mean image value and $I'(x, y)$ is the corresponding image pixel value in the Gaussian blurred version of the original image. Blurring the image using a Gaussian kernel allows high frequencies to be disregarded (requirement 4) and hence produces a salient regions map.

Ngau et al. [7] later proposed that instead of blurring the image to remove high-frequency noise and artifacts, one could perform wavelet decomposition to yield an down-sampled approximation of the image. This would allow the image to be reconstructed without really losing any information, and without resorting to blurring the image which, in a sense, is a de-resolution operation and results in loss of salient objects that are very small in size.

Ngau et al. applied the same formula as Achanta et al. (Equation 2) except that they applied it to the approximation of the wavelet decomposition (using only one level). They then reconstructed the image back to its original size. This application allows smaller salient objects to be detected while maintaining the requirements set by Achanta et al.

2.3. Spectral Residual Method

Hou and Zhang [4] discovered that extracting sparsely-occurring frequencies from an image yields a salient region map. This is performed by removing the magnitude spectrum of the image from the so-called expected magnitude spectrum of the image. The expected magnitude spectrum is realized by performing a smoothing operation on the magnitude spectrum. The authors downsample the image to a resolution of 64×64 and hence the yielded saliency map is of very low resolution when resized to that of the original image. Performing the same operation without downsampling produces a salient edges map. This is the process used to compare our method to.

Ma and Zhang [8] later discovered that the same result can be achieved by reconstructing the Fourier transform of the image using only the phase spectrum (i.e. zero-ing the magnitude spectrum and reconstructing). This method is one of the bases for our proposed method.

3. FREQUENCY TUNED SALIENT EDGE DETECTION

Using Achanta et al's requirements as a basis for our method, we set out the requirements we feel are necessary for a similar method as theirs, but that detects salient edges rather than

salient regions. From their requirements, we take the following three:

1. Establish well-defined boundaries of salient objects
2. Disregard high frequencies arising from texture, noise, and blocking artifacts
3. Efficiently output full resolution saliency maps

Our method uses the logic of Achanta et al (frequency-tuning), Ngau et al (using wavelets to attain lossless information processing), as well as Ma et al (frequency-to-time domain reconstruction using only the frequency component).

Fig. 2: A grayscale image (left) and it's wavelet decomposition. The four figures on the right are (clockwise): the approximation (NW corner), the horizontal frequency component (NE corner), the diagonal frequency component (SE corner), and the vertical frequency component (SW corner). Ngau et al. performed frequency-tuning on the approxiamtion, while we perform it on all three frequency components.

From Achanta et al, we realize that frequency-tuning allows salient regions to pop out more (Equations 4, 5, and 6). In order to detect salient edges more clearly, we perform this operation on the frequency components of the wavelet decomposition of the image. From Ngau et al, we realize that performing such an operation using wavelets allows for a reconstruction without any loss of information (Equations 3 and 7). From Ma et al, we realize that reconstructing using only the frequency components yields a map of the salient edges (Equation 7). This is the formulation of our method, which is as follows:

$$[A(f), H(f), V(f), D(f)] = \mathfrak{W}[I(x)] \tag{3}$$

$$H^*(f) = |H_\mu(f) - H'(f)| \tag{4}$$

$$V^*(f) = |V_\mu(f) - V'(f)| \tag{5}$$

$$D^*(f) = |D_\mu(f) - D'(f)| \tag{6}$$

$$S(x) = \mathfrak{W}^{-1}[0, H^*(f), V^*(f), D^*(f)] \tag{7}$$

where $I(x)$ is the input image, $\mathfrak{W}$ and $\mathfrak{W}^{-1}$ are the wavelet decomposition and inverse wavelet decomposition operators, and $A(f), H(f), V(f), D(f)$ are the approximation, horizontal, vertical, and diagonal components of the wavelet decomposition, respectively. $H'(f), V'(f), D'(f)$ are the Gaussian blurred versions of the horizontal, vertical, and diagonal components, respectively. $H_\mu(f), V_\mu(f), D_\mu(f)$ are the mean pixel values of the horizontal, vertical, and diagonal components, respectively. $H^*(f), V^*(f), D^*(f)$ are the horizontal, vertical, and diagonal frequency-tuned components respectively, and $S(x)$ is the saliency map.

Applying the frequency-tuning operation while in the decomposed state allows the salient edges to be further accentuated, while suppressing the unattentive points.

4. RESULTS AND ANALYSIS

To test the validity of the proposed method, a few test images were taken from the MSRA Salient Object Database[9]. Both the spectral residual (SR) and our method were run on the sample images. The results show that our method responds better to edges in general than SR.

In the case of a salient object having mostly sharp edges, SR tends to pick up only certain edges, not all of them (see Figure 5). In the case of a salient object having mostly blurry edges, SR does not have any response whereas our method responds appropriately, picking up the salient edges (see the red and green peppers in Figure 3).

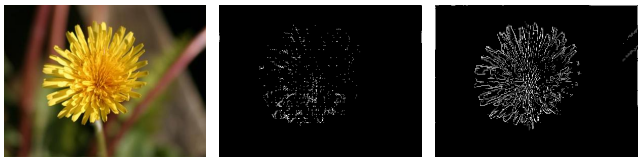
In the case where there exists a mix of both sharp and blurry edges, SR tends to detect most of the salient object, but not entirely. Our method yields an edge map of the salient object entirely. Since SR and our method yielded different dynamic ranges of saliency maps, all the results were thresholded using Otsu's method to provide a fair comparison.

(a) Original (b) SR (c) Our method

Fig. 3: Here, the SR only extracts the middle pepper while our method extracts all three.

(a) Original (b) SR (c) Our method

Fig. 4: Here, the SR does a decent job and extracting portions of the bird, while our method extracts the bird edges entirely.



(a) Original (b) SR (c) Our method

Fig. 5: Here, the SR extracts traces of the flower, while our method extracts almost every pedal.

(a) Original (b) SR (c) Our method

Fig. 6: Here, the SR extracts traces of the boy, while our method extracts him entirely.

(a) Original (b) SR (c) Our method

Fig. 7: Here, the SR map does not contain a recognizable object, while ours provides a complete outline of the salient object.

5. CONCLUSIONS

While biologically inspired saliency detection methods clearly have their advantages, it appears that frequency-tuning is the clear winner in the pool of computationally inspired methods as it provides computationally efficient, high resolution saliency maps. Achanta et al. showed that frequency-tuning works well for salient region detection while we show it can also be applied for salient edge detection.

Combining Ngau et al's method with ours could potentially improve the results of a salient region map. Requirement 3 (Establish well-defined boundaries of salient objects) on Achanta's list could be improved by our method.

In our method, we only used one level of wavelet decomposition. This allows for computational efficiency. However, if processing time is not a factor in some particular application, more levels of decomposition can be analyzed to allow for a saliency map with more edges. It should be noted though that the first level of decomposition extracts the most salient points, with further levels extracting points that are less salient.

There are two avenues that can be explored for future work. From SR, we know that decomposing and recomposing using only phase/frequency yields salient edges; and we now know that wavelets perform better than FT. A better method of time-frequency conversion would potentially yield even better results than those shown in this paper. Also, once decomposed, performing an edge-enhancing operation such as frequency-tuning clearly accentuates the edges which allows for a better saliency map. Other operations for edge-accentuation may also potentially yield better saliency maps.

Wavelets have potential in other areas in visual attention. One of our current research areas is using wavelets with Itti's method to construct a lossless, biologically inspired salient region map.

6. REFERENCES

[1] W. James, "The principles of psychology," *Harvard University Press, Cambridge, Massachussetts*, 1980/81.

[2] Q. Zhang and H. Xiao, "Extracting regions of interest in biomedical images," *2008 International Seminar on Future BioMedical Information Engineering*, 2008.

[3] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol. 20, No. 11.*, 1998.

[4] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR '07.*, 2007.

[5] A. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology, vol. 12, no. 1*, 1980.

[6] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[7] C. Ngau, L. Ang, and K. Seng, "Bottom-up visual saliency map using wavelet transform domain," *3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT)*, 2010.

[8] C. Guo, Q. Ma, and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform," *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[9] T. Liu, J. Sun, N. Zheng, X. Tang, and H. Shum, "Learning to detect a salient object," *Proc. IEEE Cont. on Computer Vision and pattern Recognition (CVPR), Minneapolis, Minnesota*, 2007.