Saliency Detection Based on 2D Log-Gabor Wavelets and Center Bias

Min Wang[1], Jia Li[2,3], Tiejun Huang[1], Yonghong Tian[1], Lingyu Duan[1], Guochen Jia[1]
[1]National Engineering Laboratory for Video Technology (NELVT), School of EE & CS, Peking University, Beijing 100871, China
[2]Key Lab of Intell. Info. Process, Inst. of Comput. Tech., Chinese Academy of Sciences, Beijing 100080, China
[3]Graduate University of Chinese Academy of Sciences, Beijing 100039, China
{mwang, jli, gcjia}@jdl.ac.cn, {yhtian, tjhuang, lyduan}@pku.edu.cn

ABSTRACT
Visual saliency can be a useful tool for image content analysis such as automatic image cropping and image compression. In existing methods on visual saliency detection, most of them are related to the model of receptive field. In this paper, we propose a bottom-up model which introduces 2D Log-Gabor wavelets for saliency detection. Compared with the traditional model of receptive field, the 2D Log-Gabor wavelets can better simulate the biological characteristics of the simple cortical cell in the receptive filed. Moreover, we also incorporate the influence of center bias into our model, which is a common phenomenon that directs visual attention to the center of images in natural scenes. Experimental results show that our approach outperforms three state-of-the-art approaches remarkably.

Categories and Subject Descriptors
I.2.10 [**ARTIFICIAL INTELLIGENCE**]: Vision and Scene Understanding –*Perceptual reasoning*

General Terms
Algorithms, Experimentation,

Keywords
Visual saliency, 2D Log-Gabor wavelets, Center bias

1. INTRODUCTION
Visual attention is a selection mechanism of human visual system (HVS) to quickly focus on the salient visual subsets in a scene. To detect these salient subsets, many approaches tried to simulate the bottom-up and top-down processes in HVS for visual saliency detection. In general, visual saliency detection is challenging but useful in many applications such as automatic image segmentation and image compression.

In existing works on visual saliency detection, most of them focus on the bottom-up processes of HVS. In contrast, the top-down processes often need training and lack the expansibility [1]. In existing bottom-up models, some models are based on information theory [2] or spectrum theory [3, 4], while most of them are driven by the theories from biology, psychology and neuropsychology. Among these models, the most famous one was proposed by Itti *et al.* [5]. They developed the center-surround structure akin to on-type and off-type visual receptive field. In [6], Achanta *et al*. detected the saliency map with a DOG model, which was often used to describe the spatial properties of symmetric receptive field on mammalian retina and lateral geniculate body [7].

Compared with these models, the Gabor model has become prevalent in recent years to describe the receptive field, which is more reasonable to describe the receptive field of simple cells in the Primary Visual Cortex (V1) of primates [8]. Nevertheless, a Gabor filter has some inherent drawbacks. First, the maximum bandwidth of a Gabor filter is limited to approximately one octave. Secondly, a Gabor filter is inadequate for covering the highest frequencies but excessive low-pass over-lapping. Thirdly, a Gabor filter is particularly difficult to cover up the mid frequencies with sufficient uniformity. Fourthly, a Gabor filter still maintains a small DC component in the even-symmetric filter. Therefore, Field *et al*. [9] proposed the Log-Gabor function to avoid the drawbacks of Gabor functions while take advantage of Gabor functions. Compared with the traditional Gabor function, the 2D Log-Gabor filters can be constructed with arbitrary bandwidth. Meanwhile, these filters can better cover up the Fourier domain and recover the highest frequencies without the DC component.

In addition to the bottom-up processes in the receptive field, several studies have observed that subject's visual attention, measured by saccade directions and fixation locations, is often biased toward the center of static image [10]. One of the most interesting causes of such center bias is known as the photographer bias. That is, photographers tend to place the objects-of-interest near the center of their composition in order to enhance their focus and size relative to the background [11].

In this paper, we propose a novel approach for visual saliency detection in natural images. In our approach, the 2D Log-Gabor filters are used to simulate the bottom-up processes in the receptive fields. In order to be as close as possible to the human visual information processing, we should consider the influence of center bias into our model.

The rest of paper is organized as follows. Section 2 presents a bottom-up method based on 2D Log-Gabor wavelets and center bias in detail. Extensive experiments are given in Section 3 and the paper is concluded in Section 4.

MM'10, October 25-29, 2010, Firenze, Italy.

2. OUR METHOD

The framework of our approach is illustrated in Fig. 1. In this framework, we first detect the low-level irregularities in various preattentive feature with the 2D Log-Gabor wavelets. These irregularities are then integrated by considering the center bias to construct the bottom-up saliency map. In the rest of this section, we will first introduce the 2D Log-Gabor wavelets, followed by the details of applying the 2D Log-Gabor wavelets and center bias to detect visual saliency.

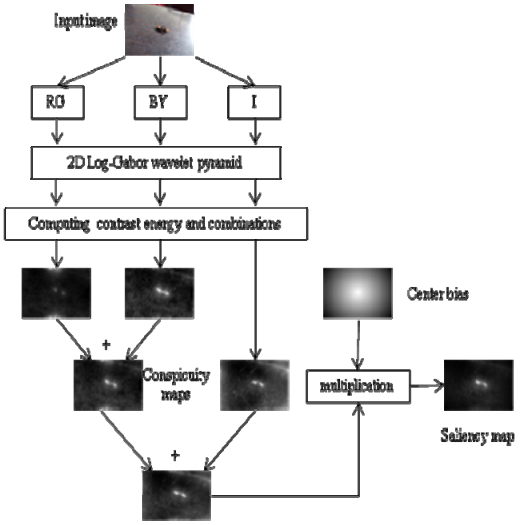

Figure 1: System framework of our approach.

2.1 Introduction to 2D Log-Gabor Wavelets

The Log-Gabor filters are defined in log-polar coordinates of the Fourier domain as Gaussians shifted from the origin [9]:

$$G(\rho,\theta,s,t)=\exp\left(-\frac{(\rho-\rho_s)^2}{2\sigma_\rho^{\,2}}\right)\exp\left(-\frac{(\theta-\theta_{st})^2}{2\sigma_\theta^{\,2}}\right), \qquad (1)$$

where

- (ρ, θ) is the log-polar coordinates;
- s and t indicate the scale and orientation of the filter;
- (ρ_s, θ_{st}) is frequency center of filter and
- $(\sigma_\rho, \sigma_\theta)$ is the bandwidths in ρ and θ.

Similar to the Gabor filter, a Log-Gabor filter also consists of an even-symmetric filter and an odd-symmetric filter, which are represented by the real and imaginary parts, respectively. Fig. 2 shows the shape of a 2D Log-Gabor filter in the spatial domain. To detect all the irregularities in the preattentive features, a set of such filters with different scales and orientations is required. Therefore, we have to first determine the bandwidth of each filter in scale and orientation.

The choice of the bandwidth in orientation is motivated by the orientation resolution of simple cells, which has been evaluated as around 20-40 degrees of full bandwidth at half response [8]. Therefore, it requires around 6 to 13 orientations with such bandwidth to cover the 180 degrees of the plan. Similarly, the bandwidth in scale has been evaluated between 0.6 and 3.2 octaves [13]. One objective of this study is to choose the transform parameters as close as possible to the known physiology of simple cortical cells, so we select 5 scales and 8 orientations to construct a set of 2D Log-Gabor filters

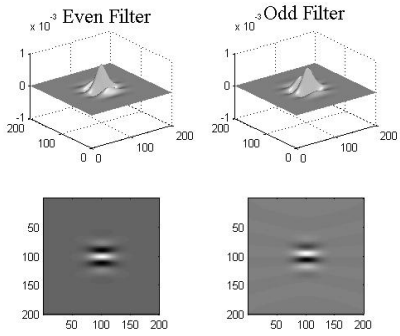

Figure 2: The 2D Log-Gabor filter at 3rd scale and 5th orientation in the spatial domain. (Left) Even-symmetric filter. (Right) Odd-symmetric filter.

2.2 Saliency Detection using the 2D Log-Gabor Wavelet Pyramid

According to studies in [8] and [9], we can use the 2D Log-Gabor filters to describe the receptive field. Therefore, we build the 2D Log-Gabor wavelet pyramid to detect the saliency map. Since visual neurons are often excited by one color and inhibited by opponent color [14], we choose the preattentive features as the red/green (RG) and blue/yellow (BY) color opponencies. As in [5], the intensity (I) is also selected as a preattentive feature. Our model for detecting the saliency map sM can be formulated as follows:

Since the 2D Log-Gabor filters are used in frequency domain, we first calculate the frequency spectrums of these preattentive features using the Fourier Transform. After that, a 2D Log-Gabor pyramid is constructed for each frequency spectrum. Then the irregularities in each preattentive feature can be detected using the Inverse Fourier Transform:

$$\begin{aligned} RG(s,t) &= F^{-1}(LG_{st}(F(RG)) \\ BY(s,t) &= F^{-1}(LG_{st}(F(BY)) \\ I(s,t) &= F^{-1}(LG_{st}(F(I)). \end{aligned} \qquad (2)$$

where

- F and F^{-1} denote the Fourier Transform and Inverse Fourier Transform, respectively,
- LG_{st} is a 2D Log-Gabor filter with scale s and orientation t.

In (2), we observe that the irregularties with different scales and orientations are separated by projecting the preattentive features onto different 2D Log-Gabor basis. After that, we can get 40 feature maps at 5 scales and 8 orientations for each preattentive feature. Since each feature map is a complex image analyzed by a 2D Log-Gabor filter, we can obtain the contrast energy of each feature map. The contrast energy produces salient feature localized with respect to spatial location, orientation and frequency.

$$\begin{aligned} \overline{RG} &= \sum_{s=1}^{5}\sum_{t=1}^{8} \| RG(s,t) \| \\ \overline{BY} &= \sum_{s=1}^{5}\sum_{t=1}^{8} \| BY(s,t) \| \\ \overline{I} &= \sum_{s=1}^{5}\sum_{t=1}^{8} \| I(s,t) \| \\ \overline{C} &= N(\overline{RG}) + N(\overline{BY}), \end{aligned} \qquad (3)$$

where $N()$ denotes the normalization operation to normalize the energy map into [0, 1]. $\bar{I}$ and $\bar{C}$ are the intensity and color conspicuity maps, respectively. At last, these two conspicuity maps can be integrated to calculate the estimated saliency map S:

$$S = \frac{1}{2}(N(\bar{I}) + N(\bar{C})). \quad (4)$$

2.3 Applying the Center Bias

Typically, viewers may reorient to the center of a scene at a greater frequency than to other locations. That is, the closer a pixel is to the center of image, the higher probability it is observed. Therefore, pixels located near to the image center may provide more information than the other pixels, thus becoming more salient. The center bias can influence human eye tracking when observers look in arbitrary natural scenes and it is not tied to a particular stimulus. In order to exactly simulate human visual information processing, we should consider the influences of center bias. In this paper, the distance to the center for each pixel is calculated as the influence of center bias:

$$Dis(x, y) = \frac{\sqrt{(x - x_0)^2 + (y - y_0)^2}}{4\sqrt{W^2 + H^2}}, \quad (5)$$

where W and H are the width and height of the image. $(x_0, y_0) = (W/2, H/2)$ is the center of the image.

2.4 Computing the Saliency Map

We combine the results from 2D Log-Gabor wavelet pyramid and the influences of center bias into the final saliency map which has the same size as the original images:

$$sM(x, y) = (1 - Dis(x, y)) * S. \quad (6)$$

Based on the 2D Log-Gabor wavelet pyramid and center bias, we propose a simple multi-resolution bottom-up model for saliency detection. The proposed method can detect saliency maps that have the same size as the input images, while Itti *et al.* [5] produce saliency maps that are just 1/256 of the original image size in pixel, and Hou *et al*. [3] output maps of size 64 × 64 for any input image size.

3. EXPERIMENT RESULTS

In this section, we evaluate the performance of our approach on a public image dataset with recorded human eye fixations. The dataset consists of 1003 images available from [15]. These images are mainly selected from Flicker creative commons and LableMe. The longest dimension of each image is 1024 pixels and the other dimension ranges from 405 to 1024 with the majority at 768 pixels. There are 779 landscape images and 228 portrait images. For these images, the eye tracking data of 15 subjects are recorded with an eye-tracker. They discarded the first fixation from each scan-path to avoid adding trivial information from the initial center fixation.

On this dataset, we choose the ROC curves and ROC areas to evaluate the performance of various saliency models. The ROC curves can be plotted as the False Positive Rate (FPR) vs. Hit Rate (HR). The ROC area can be then calculated as the area under the ROC curve to demonstrate the overall performance of a saliency model. Perfect prediction corresponds to the ROC area of 1, while random prediction generates an ROC area of 0.5.

3.1 Experiment Results of Eye Tracking

In this section, we evaluate our approach and other three state-of-the art bottom-up models [5, 2, 12]. The ROC curves of these models are shown in Fig. 3. The ROC areas of these approaches are shown in Table 1.

From Fig. 3, we can see when FPR is low, the HR of our model reaches the highest and as the increase of FPR, the HR of our model will be the same as the model propose by Wang *et al*. [16]. However, it can be seen from Table 1 that our model has the largest ROC area and achieves the best overall performance.

Table 1. The ROC area comparison on the static image

Model	Itti *et al.*[5]	Bruce *et al.* [2]	Wang *et al.*[12]	Our model
ROC area	0.6736	0.7220	0.7750	0.8176

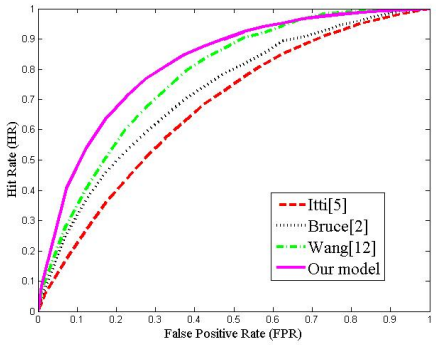

Figure 3: The ROC curves of our model and the other methods on the still image dataset.

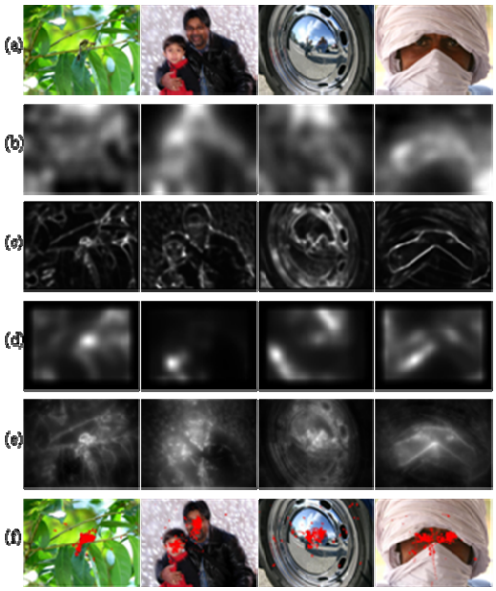

Figure 4: Result of our method in comparison with other methods. From top to bottom: (a) Original images, (b) the saliency maps from [5], (c) the saliency maps from [2], (d) the saliency maps from [12], (e) the saliency maps of ours, (f) the red points indicate human fixations in this image.

Fig. 4 shows the comparison of the saliency maps of our model with the saliency maps from the other models. It can be seen that our model is more robust and can predict where human look more efficiently and more accurately.

3.2 Introducing Center Bias to Other Models

Because we introduce the influence of center bias to our model while the other models don't consider it and the center bias is a common phenomenon when humans look in natural scenes, so it seems that our model considers the priori knowledge of a scene and there is a need to include the influence of center bias to the other models [5, 2, 12].

The criteria of evaluation are also the ROC curves and ROC areas, Table2 and Fig. 5 shows the ROC areas and ROC curves after introducing the center-bias to the other models, respectively. It can be seen that center bias improve the performance greatly. On the other hand, it can prove that a strong bias for human fixations is near the center of the image, therefore when trying to predict where humans look, we should consider the effect of center bias. It can also be seen that our model reaches the highest HR when FPR is low and our model has the largest ROC area. These results indicate our model can also achieve the best performance, even though the other models introduce the influence of center bias.

Table 2. The ROC area after introducing center bias

Model	Itti *et al.*[5]	Bruce *et al.* [2]	Wang *et al.*[12]	Our model
ROC area	0.7763	0.7897	0.8090	0.8176

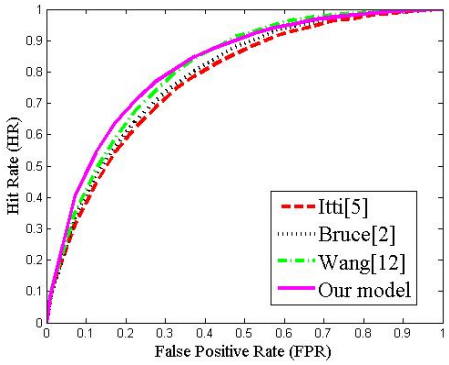

Figure 5: The ROC curves of the proposed model and the other methods after introducing center-bias.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we propose an effective multi-resolution bottom-up approach for saliency detection by using 2D Log-Gabor which is the known mechanism of receptive field of the primate visual cortex and the common phenomenon–center bias. Our model is able to provide full resolution saliency maps. It can be seen that our model has the best performance and the highest ROC area in comparison with other three state-of-the art bottom-up techniques [5, 2, 12] on the dataset of static images. Even after adding the center-bias to the other models, our model still has the best performance. In this work, we just consider the still feature and test our model in static images. For future work, we will consider the motion feature in temporal domain and then test our improved model in video.

5. ACKNOWLEDGMENTS

This work is supported by grants from the Chinese National Natural Science Foundation under contract No. 60973055 and No. 90820003, National Basic Research Program of China under contract No. 2009CB320906, and Fok Ying Dong Education Foundation under contract No. 122008.

6. REFERENCES

[1] R. Fergus, P. Perona and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In CVPR 2003.

[2] N. D. Bruce and J. K. Tsotsos. Saliency based on information maximization. In NIPS, 2006.

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

[4] C. Guo, Q. Ma and L. Zhang, Spatio-temporal saliency detection using phase spectrum of Quaternion Fourier transform. In CVPR, 2008.

[5] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. In TPAMI, 20(11):1254-1259, 1998.

[6] R. Achanta, S. Hemami, F. Estrada and S. Süsstrunk. Frequency-tuned salient region detection. In CVPR, 2009.

[7] L. G. Roberts. Machine Perception of Three-Dimension Solids. In Optical and Electro-Optimal Information Processing[C]. Cambridge, MA: MIT Press, 1965.

[8] Daugman, J. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. In J. Opt. Soc. Am. A, 2(7):1160-1169, 1985.

[9] Field. D. J. Relation between the statistics of natural images and the response properties of cortical cells. In J. Opt. Soc. Am. A, 4(21):2379-2394, 1987.

[10] Foulsham, T., and Underwood, G. What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. In Journal of Vision, 8(2):6,1-17, 2008.

[11] P.H. Tseng, R. Carmi, L.G.M. Cameron, D. P. Munoz, and L. Itti. Quantifying center bias of observers in free viewing of dynamic natural scenes. In Journal of Vision, 9(7):4,1-16, 2009.

[12] W. Wang, Y. Wang, Q. Huang, and W. Gao. Measuring visual saliency by site entropy rate. In CVPR, 2010.

[13] DeValois, R.L., Albrecht, D.G., and Thorell, L.G. Spatial frequency selectivity of cells in macaque visual cortex. In Vision Research, 22:545-559, 1982.

[14] S. Engel, X. Zheng, and B. Wandell. Colour Tuning in Human Visual Cortex Measured with functional magnetic resonance imaging. In Nature, Vol.388, no.6,637, pp.68-71, July 1997,3.

[15] T. Judd, K. Ehinger, F. Durand and A.Torralba. Learning to predict where humans look, In CVPR, 2005.