# SDSP: A NOVEL SALIENCY DETECTION METHOD BY COMBINING SIMPLE PRIORS

*Lin Zhang*, *Zhongyi Gu*, *and Hongyu Li**[1]
School of Software Engineering, Tongji University, Shanghai, China

## ABSTRACT

Salient regions detection from images is an important and fundamental research problem in neuroscience and psychology and it serves as an indispensible step for numerous machine vision tasks. In this paper, we propose a novel conceptually simple salient region detection method, namely SDSP, by combining three simple priors. At first, the behavior that the human visual system detects salient objects in a visual scene can be well modeled by band-pass filtering. Secondly, people are more likely to pay their attention on the center of an image. Thirdly, warm colors are more attractive to people than cold colors are. Extensive experiments conducted on the benchmark dataset indicate that SDSP could outperform the other state-of-the-art algorithms by yielding higher saliency prediction accuracy. Moreover, SDSP has a quite low computational complexity, rendering it an outstanding candidate for time critical applications. The Matlab source code of SDSP and the evaluation results have been made online available at http://sse.tongji.edu.cn/linzhang/va/SDSP/SDSP.htm.

***Index Terms***— Salient object detection, log-Gabor filter, visual attention

## 1. INTRODUCTION

Human beings can routinely and effortlessly judge the importance of image regions, and focus their attention on important parts. Computationally detecting such salient image regions has vast applications including object-of-interest image segmentation [1], object recognition [2], adaptive image compression [3], content-aware image editing [4], etc. As a result, in the past decade, researchers working in various fields have devoted a great deal of efforts in the area of visual attention modeling.

Most existing visual attention approaches are based on the bottom-up computational framework, where visual attention is supposed to be driven by low-level stimulus in the scene, such as contrast, color, and motion. Since there is an enormous literature in this area, we only review several representative ones of them.

The first influential and best known model in this field was proposed by Itti *et al*. [5]. Itti's model follows the Feature Integration Theory [6] by first decomposing the visual input into separate low-level features maps. Then, normalized center-surround difference maps are computed for individual features and later combined by a weighting scheme to form a saliency map. In [7], following Itti *et al*.'s architecture, Harel *et al*. proposed the graph-based visual saliency (GBVS) model by introducing a novel graph-based normalization/combination strategy. In another work following Itti's framework, Klein and Frintrop [8] modeled the center-surround contrast in an information-theoretic way, in which two distributions of visual feature occurrences are determined for a center and a surround region. In [9], Bruce and Tsotsos modeled the image's saliency as the maximum information that can be sampled from it. In their method, saliency is computed as Shannon's self-information. By analyzing the log-spectrum of the input image, Hou and Zhang [10] proposed a Fourier transform based method to extract the spectral residual of an image in the spectral domain and to construct the corresponding saliency map in the spatial domain; one prominent advantage of this method is its low computational complexity. In Hou's latest work [11], he proposed the *image signature* to approximate the foreground of an image within the theoretical framework of sparse signal mixing. In [12], Achanta *et al*. proposed a conceptually simple approach by combining image's band-pass filtered responses from three $CIEL^*a^*b^*$ channels. This method can provide pleasing results in most cases and it has the advantage of computational efficiency. As an extension to [12], Achanta *et al.* improved their original method by considering the special effects of boundaries in their later work [13]. In [14], Cheng *et al.* proposed a regional contrast based saliency extraction algorithm, which simultaneously evaluates global contrast differences and spatial coherence. In [15], Goferman *et al.* proposed a new type of saliency, namely context-aware saliency, which aims at detecting the image regions that represent the scene. In [16], a conditional random field is learned to combine features, such as multi-scale contrast, center-surround histogram, and color spatial distribution, for salient object detection. In [17], Shen and Wu represent an image as a low-rank matrix plus sparse noises in a learned feature space, where the low-rank matrix explains the non-salient regions while the sparse noises indicate the salient regions. For a more complete survey on modern visual attention models, please refer to [18, 19].

A salient region detection model is usually used as a pre-processing component in a machine vision system. Thus,

[1] Corresponding author

a perfect salient region detection algorithm needs to perform well in two aspects. At first, it should have a good saliency prediction performance, which means that salient regions predicted by such an algorithm should be highly correlated with the judgment of human beings. Secondly, to be suitable for real-time applications, the algorithm should have a low computational cost. Nevertheless, through our investigation we find that the computational cost is often ignored when designing such algorithms. Many modern salient region detection methods, such as [7, 9, 15-17], are very complicated and usually are not computationally efficient, which limits their applications in practice.

Based on these considerations, in this paper, we propose a novel salient region detection method having a high prediction performance and a low computational cost simultaneously, namely *Saliency Detection by combining Simple Priors* (SDSP, for short). The proposed SDSP method is constructed by combining three simple priors. At first, the behavior that the human visual system detects salient objects in a visual scene can be well modeled by band-pass filtering. Secondly, people are more likely to pay their attention on the center of an image. Thirdly, warm colors are more attractive to people than cold colors are. We propose to use simple mathematical models to express these three priors efficiently and effectively. By combining cues from three simple priors, the proposed SDSP approach is reached. The performance of SDSP is examined on the benchmark dataset and is compared with other eight state-of-the-art saliency detection methods. Efficacy and efficiency of SDSP are corroborated by the experimental results.

The remainder of this paper is organized as follows. Section 2 describes three simple priors and methods to simulate them. Section 3 presents our salient region detection algorithm SDSP. Section 4 reports the experimental results and Section 5 concludes the paper.

## 2. SIMPLE PRIORS AND THEIR MODELING

In this section, three simple priors, of which our SDSP algorithm is comprised, will be described in detail and methods to model them will be also presented.

### 2.1. Frequency prior

The seminal work from Achanta *et al.* [12] indicates that the salient region detection mechanism can be well approximated by integrating band-pass filtering responses from opponent color channels (such as the $CIEL^*a^*b^*$ color channels). With respect to the band-pass filter, Achanta *et al.* adopted the *Difference of Gaussian* (DoG) filter.

Inspired by Achanta *et al.*'s work [12], in this paper, we also resort to band-pass filtering for saliency detection. However, with respect to the band-pass filter, we adopt the log-Gabor filter [20], instead of DoG. There are some good reasons for our selection. At first, we can construct a log-Gabor filter with an arbitrarily bandwidth and still having no DC component. Secondly, the transfer function of the log-Gabor filter has an extended tail at the high-frequency end, which makes it more capable to encode natural images than other common band-pass filters [20, 21]. The transfer function of a log-Gabor filter $g(\mathbf{x})$ ($\mathbf{x} = (x, y) \in \mathbb{R}^2$) in the frequency domain can be expressed as

$$G(\mathbf{u}) = \exp\left(-\left(\log\frac{\|\mathbf{u}\|_2}{\omega_0}\right)^2 / 2\sigma_F^2\right) \qquad (1)$$

where $\mathbf{u} = (u, v) \in \mathbb{R}^2$ is the coordinate in the frequency domain, $\omega_0$ is the filter's center frequency, and $\sigma_F$ controls the filter's bandwidth. $g(\mathbf{x})$ cannot be analytically expressed due to the singularity in the log function at the origin. Instead, $g(\mathbf{x})$ can only be approximately obtained by performing a numerical inverse Fourier transform to $G(\mathbf{u})$. An example of the 2-D log-Gabor filter in the frequency domain, with $\omega_0 = 1/6$ and $\sigma_F = 0.3$, is shown in Fig. 1.

(a) (b)

**Fig. 1:** An example of the log-Gabor filter in the frequency domain, with $\omega_0 = 1/6$ and $\sigma_F = 0.3$, shown in 3D surface format (a) and in gray-scale image format (b).

Given an image $\{\mathbf{f}(\mathbf{x}):|\ \mathbf{x} \in \Omega$, where $\Omega$ denotes the image spatial domain$\}$ in RGB color space ($\mathbf{f}(\mathbf{x})$ is actually a vector, containing three values representing R, G, and B intensities at the position $\mathbf{x}$), its saliency map $\{S_F(\mathbf{x})\}$ modeled by band-pass filtering can be obtained as the follows, which is similar to [12]. At first, $\mathbf{f}(\mathbf{x})$ needs to be converted to $CIEL^*a^*b^*$ space, which is actually an opponent color space. The three resulting channels are denoted by $f_L(\mathbf{x})$, $f_a(\mathbf{x})$, and $f_b(\mathbf{x})$. Then, the saliency $S_F(\mathbf{x})$ is defined as

$$S_F(\mathbf{x}) = \left(\left(f_L * g\right)^2 + \left(f_a * g\right)^2 + \left(f_b * g\right)^2\right)^{\frac{1}{2}}(\mathbf{x}) \qquad (2)$$

where * denotes the convolution operation.

### 2.2. Color prior

Some studies [17] find from daily experiences that warm colors, such as red and yellow, are more pronounced to the human visual system than cold colors, such as green and blue. In this paper, we propose a simple yet effective method to model this prior.

For a given image $\{\mathbf{f}(\mathbf{x})\}$ in the *RGB* color space, at first, it will be converted to the $CIEL^*a^*b^*$ color space. $\{f_L(\mathbf{x})\}$, $\{f_a(\mathbf{x})\}$, and $\{f_b(\mathbf{x})\}$ represent $L^*$-channel, $a^*$-channel, and $b^*$-channel, respectively. $CIEL^*a^*b^*$ is an opponent

color system, in which $a^*$-channel represents green-red information while $b^*$-channel represents blue-yellow information. If a pixel has a smaller (greater) $a^*$ value, it would seem greenish (reddish). With the same manner, if a pixel has a smaller (greater) $b^*$ value, it would seem bluish (yellowish). Hence, if a pixel has a higher $a^*$ or $b^*$ value, it would seem "warmer"; otherwise, it would seem "colder".

Based on the aforementioned analysis, we devise a metric to evaluate the "color saliency" for a given pixel. At first, we perform linear mappings $f_a(\mathbf{x}):\mapsto f_{an}(\mathbf{x})\in[0,1]$ and $f_b(\mathbf{x}):\mapsto f_{bn}(\mathbf{x})\in[0,1]$ by

$$f_{an}(\mathbf{x})=\frac{f_a(\mathbf{x})-mina}{maxa-mina}, f_{bn}(\mathbf{x})=\frac{f_b(\mathbf{x})-minb}{maxb-minb} \quad (3)$$

where *mina* (*maxa*) is the minimum (maximum) value of $\{f_a(\mathbf{x}):|\mathbf{x}\in\Omega\}$ and *minb* (*maxb*) is the minimum (maximum) value of $\{f_b(\mathbf{x}):|\mathbf{x}\in\Omega\}$. Thus, each pixel $\mathbf{x}$ can be mapped to one point in the color plane ($f_{an}$, $f_{bn}$)$\in$[0, 1]×[0, 1]. Intuitively, in this color plane, the point ($f_{an}$=0, $f_{bn}$=0) is the "coldest" point and thus it is the "least salient" one. Therefore, we define the color saliency of a point $\mathbf{x}$ in a straightforward manner as

$$S_C(\mathbf{x})=1-\exp\left(-\frac{f_{an}^2(\mathbf{x})+f_{bn}^2(\mathbf{x})}{\sigma_C^2}\right) \quad (4)$$

where $\sigma_C$ is a parameter.

### 2.3. Location prior

Several previous studies have demonstrated that objects near the image center are more attractive to people [22]. That implies locations near the center of the image will be more likely to be "salient" than the ones far away from the center. This prior can be simply and effectively modeled as a Gaussian map. Suppose $\mathbf{c}$ is the center of the image $\{\mathbf{f}(\mathbf{x})\}$. Then, the "location saliency" at $\mathbf{x}$ under the "location prior" can be expressed as a Gaussian map

$$S_D(\mathbf{x})=\exp\left(-\frac{\|\mathbf{x}-\mathbf{c}\|_2^2}{\sigma_D^2}\right) \quad (5)$$

where $\sigma_D$ is a parameter.

## 3. SDSP: SALIENCY DETECTION BY COMBINING SIMPLE PRIORS

Based on three simple priors discussed in Section 2, we can derive our proposed saliency detection method, namely *Saliency Detection by combining Simple Priors* (SDSP). Suppose that from the given image $\mathbf{f}(\mathbf{x})$, we have computed three saliency maps, $S_F(\mathbf{x})$, $S_C(\mathbf{x})$, and $S_D(\mathbf{x})$ by using the three simple priors, respectively. The image's final saliency map can be naturally defined as

$$SDSP(\mathbf{x})=S_F(\mathbf{x})\cdot S_D(\mathbf{x})\cdot S_C(\mathbf{x}) \quad (6)$$

The procedures to compute SDSP is illustrated using a sample image taken from [12] in Fig. 2.

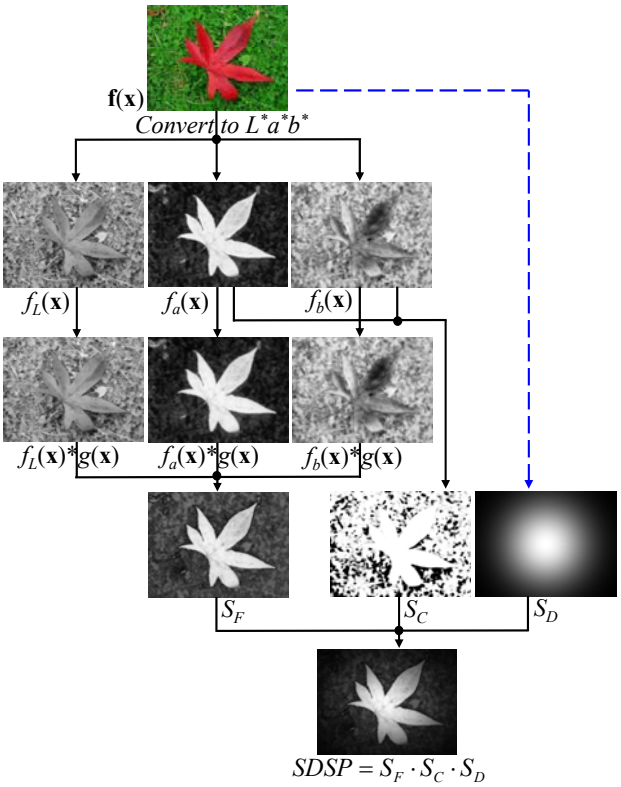


**Fig. 2:** Illustration for the computation process of SDSP.

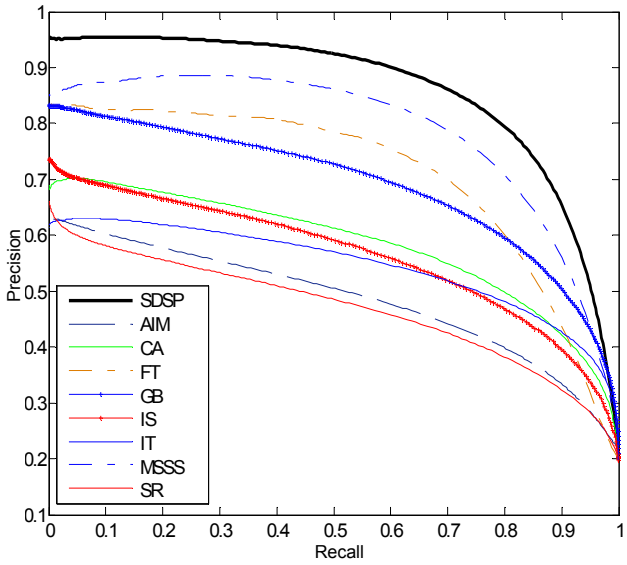


**Fig. 3:** Precision-recall curves obtained by using various saliency detection methods on the benchmark dataset.

## 4. EXPERIMENTAL RESULTS

### 4.1. Dataset and parameter settings

We exhaustively compared our approach SDSP with other eight state-of-the-art saliency detection methods on a publicly available dataset [12] comprising 1000 images with binary ground truth. The methods used for comparison

included AIM [9], CA [15], FT [12], GB [7], IS [11], IT [5], MSSS [13], and SR [10].

Parameters in our SDSP approach were empirically tuned as: $\omega_0 = 0.002$, $\sigma_F = 6.2$, $\sigma_D = 114$, and $\sigma_C = 0.25$.

### 4.2. Segmentation by fixed thresholding

Similar to [12], we evaluated the performance of a saliency detection algorithm in the context of salient object segmentation. For a given saliency map with values in the range [0, 255], the simplest way to get a binary segmentation of the salient object is to threshold the saliency map at a threshold $T_f \in [0, 255]$. When $T_f$ varies from 0 to 255, different precision-recall pairs are obtained, and a precision-recall curve can be drawn. The average precision-recall curve is generated by averaging the results from all the 1000 test images. The resulting curves are shown in Fig. 3.

### 4.3. Segmentation by adaptive thresholding

Table 1: *F*-measure for each algorithm

| Method | *F*-measure |
|---|---|
| AIM [9] | 0.4317 |
| CA [15] | 0.5528 |
| FT [12] | 0.6700 |
| GB [7] | 0.6186 |
| IS [11] | 0.5020 |
| IT [5] | 0.4959 |
| MSSS [13] | 0.7417 |
| SR [10] | 0.4568 |
| SDSP | **0.7758** |

In this experiment, we used an image dependent adaptive threshold to segment objects in the image. Such a strategy was proposed in [12]. Specifically, such an adaptive threshold $T_a$ is determined as twice the mean saliency of the image by

$$T_a = \frac{2}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} S(x, y) \tag{7}$$

where $W$ and $H$ are the width and height of the saliency map, respectively, and $S(x, y)$ is the saliency value of the pixel at the position $(x, y)$.

Using the adaptive threshold, we could obtain binarized maps of salient objects extracted by each of the saliency detection algorithm. Then, for each algorithm, for each image, we can compute the *F*-measure, which is defined as

$$F_\beta = \frac{(1 + \beta^2) \cdot Precison \cdot Recall}{\beta^2 \cdot Precison + Recall} \tag{8}$$

Similar to [12, 17], we set $\beta^2 = 0.3$ in our experiments. *F*-measure can reflect the overall prediction accuracy of an algorithm. Averaged *F*-measure over 1000 images achieved by each saliency detection algorithm is listed in Table 1.

### 4.4. Computational cost

In addition to the saliency prediction accuracy, the computational costs of various methods were also evaluated. Experiments were performed on a standard HP Z620 workstation with a 3.2GHZ Intel Xeon E5-1650 CPU and an 8G RAM. The software platform was Matlab R2012a. The time cost consumed by each evaluated saliency detection method for processing one 400×300 color image is listed in Table 2.

Table 2: Time cost of each method

| Method | Time (seconds) |
|---|---|
| AIM [9] | 5.118 |
| CA [15] | 33.662 |
| FT [12] | 0.045 |
| GB [7] | 0.464 |
| IS [11] | 0.022 |
| IT [5] | 0.134 |
| MSSS [13] | 0.784 |
| SR [10] | 0.013 |
| SDSP | 0.039 |

### 4.5. Discussions

From precision-recall curves shown in Fig. 3 and *F*-measures listed in Table 1, it can be seen that with respect to the saliency region detection accuracy, the proposed method SDSP performs consistently and significantly better.

In addition, from Table 2, it can be seen that the computational costs of different saliency detection methods vary greatly. SDSP runs only a little slower than the methods SR [10] and IS [11] while it works much faster than all the other methods evaluated. However, it should be noted that in terms of the saliency detection accuracy, SDSP performs much better than SR and IS.

Thus, we can conclude that the proposed SDSP method could achieve the best saliency detection accuracy while it has a quite low computational complexity, which renders it a better candidate for time critical applications.

## 5. CONCLUSION

In this paper, we proposed a novel salient region detection method, namely SDSP, by combining three simple priors in a straightforward manner. SDSP is conceptually simple and can be easily implemented. Experimental results indicate that SDSP could yield statistically better saliency detection accuracy than all the other competing methods evaluated. Moreover, SDSP has a very low computational complexity, making it the best candidate for real-time applications.


## ACKNOWLEDGEMENT

This work is supported by the Fundamental Research Funds for the Central Universities under grant no. 2100219033, the Natural Science Foundation of China under grant no. 61201394, and the Innovation Program of Shanghai Municipal Education Commission under grant no. 12ZZ029.

## 6. REFERENCES


[1] B.C. Ho and J. Nam, "Object-of-interest image segmentation based on human attention and semantic region clustering," *J. Opt. Soc. Am. A*, vol. 23, pp. 2462-2470, 2006.

[2] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition," *CVPR'04,* pp. 37-44, 2004.

[3] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 still image coding system: an overview," *IEEE Trans. Consumer Elec.*, vol. 46, pp. 1103-1127, 2002.

[4] Y. Wang, C. Tai, O. Sorkine, and T. Lee, "Optimized scale-and-stretch for image resizing," *ACM Trans. Graph.*, vol. 27, pp. 118.1-8, 2008.

[5] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. PAMI*, vol. 20, pp. 1254-1259, 1998.

[6] A.M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, pp. 97-136, 1980.

[7] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Adv. Neural Information Process. Syst.*, vol. 19, pp. 545-552, 2007.

[8] D.A. Klein and S. Frintrop, "Center-surround divergence of feature statistics for salient object detection," *ICCV'11*, pp. 2214-2219, 2011.

[9] N. Bruce and J. Tsotsos, "Saliency based on information maximization," *Adv. Neural Information Process. Syst.*, vol. 18, pp. 155-162, 2006.

[10] X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," *CVPR'07*, pp. 1-8, 2007.

[11] X. Hou, J. Harel, and C. Koch, "Image signature: highlighting sparse salient regions," *IEEE Trans. PAMI*, vol. 34, pp. 194-201, 2012.

[12] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," *CVPR'09*, pp. 1597-1604, 2009.

[13] R. Achanta and S. Susstrunk, "Saliency detection using maximum symmetric surround," *ICIP'10*, pp. 2653-2656, 2010.

[14] M. Cheng, G. Zhang, N.J. Mitra, X. Huang, and S. Hu, "Global contrast based salient region detection," *CVPR'11*, pp. 409-416, 2011.

[15] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *CVPR'10*, pp. 2376-2383, 2010.

[16] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Shum, "Learning to detect a salient object," *IEEE Trans. PAMI*, vol. 33, pp. 353-367, 2011.

[17] X. Chen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," *CVPR'12*, pp. 853-860, 2012.

[18] A. Toet, "Computational versus psychophysical bottom-up image saliency: a comparative evaluation study," *IEEE Trans. PAMI*, vol. 33, pp. 2131-2146, 2011.

[19] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Trans. PAMI*, vol. 35, pp. 185-207, 2013.

[20] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am. A*, vol. 4, pp. 2379-2394, 1987.

[21] P. Kovesi, "Image features from phase congruency," *Videre: J. Comp.Vis. Res.*, vol. 1, pp. 1-26, 1999.

[22] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," *ICCV'09*, pp. 2106-2113, 2009.