# Multiscale saliency detection using principle component analysis

Jingbo Zhou, Zhong Jin, Jingyu Yang
School of Computer Science and Technology,
Nanjing University of Science and Technology,
Nanjing, China
E-mail: zhoujingbo2006@yahoo.com.cn, zhongjin@mail.njust.edu.cn

*Abstract*— **In this paper, we propose a new multiscale saliency detection algorithm based on principal component analysis. To measure saliency of pixels in a given image, we first segment the image into patches and then use the principal component analysis to reduce the dimensions, in which it throw out dimensions that are noises with respect to the saliency calculation. The saliency of a patch is computed as the dissimilarities of colors and the spatial distance between it and other patches. Finally, we implement our algorithm through multiple scales so it can further decrease the saliency of background. Our method was compared with other saliency detection approaches using two public image datasets. Experimental results show that our method outperforms current state-of-the-art methods on predicting human fixations and salient object segmentation.**

***Keyword: saliency detection, multiscale, principle component analysis***

## I. INTRODUCTION

Humans can identify salient areas in their visual fields with surprising speed and accuracy before performing actual recognition. Computationally detecting such salient in image regions remains a significant goal, as it allows preferential allocation of computational resources in subsequent image analysis and synthesis. There are many applications for visual attention, for example, automatic image cropping [1], adaptive image display on small devices [2], image/video compression, advertising design [3], and image collection browsing. Recent studies [4, 5] demonstrated that visual attention helps object recognition, tracking, and detection as well.

In this paper, we aim to measure the saliency for each patch drawn from an image. Our work is inspired by [6] which defined the saliency in three elements: dissimilarity, spatial distance and central bias. We also measure the patch's saliency value using the information such as dissimilarity and spatial distance. The central bias, proposed by [7], is based on the principle that dominant objects often raise to the center of the image. This underlying hypothesis brings two problems: First, background near the center of image maybe more salient than the foreground which located in the boundary of the image; Second, for a salient object, the part near the center is more salient than that far away from the center. To diminish this effect, we exploit the multiple scales instead of the central bias to decrease the saliency of background patches, improving the contrast between salient and non-salient regions.

In our method, we use the PCA [8] to reduce the dimensionality of each patch which is represented as a vector. Our proposed model is based on the hypothesis similar to [9, 10] that dominant object always smaller than the background. Under this hypothesis, principal components (PCs) according to few largest eigenvalues represent the principal directions of the features that come from the patches of the background. We project the patches to these PCs which not only throw out the dimensions that are noises with respect to the saliency calculation [6], but also throw out the features that can represent the salient object. Considering a patch of background, the sum of differences between it and other patches remains small because the patches of background are more than that of foreground and the difference between patches of background is much small. On the contrary, for a patch of foreground, the sum of difference will be large.

The remainder of the paper is organized as follows: Previous work is discussed in the following section. In Section 3, we state the framework of our saliency detection method in details. In Section 4 we demonstrate our experimental results using three image datasets and compared the results with other state-of-art saliency detection methods. The conclusions are given in Section 5.

## II. RELATED WORK

Itti et al. [11] introduced a saliency model which was biologically inspired. Specifically, they proposed the use of a set of feature maps from three complementary channels as intensity, color, and orientation. The normalized feature maps from each channel were then linearly combined to generate the overall saliency map. Even though this model has been shown to be successful in predicting human fixations, it is somewhat ad-hoc in that there is no objective function to be optimized and many parameters must be tuned by hand. With the proliferation of eye-tracking data, a number of researchers have recently attempted to address the question of what attracts human visual attention by being more mathematically and statistically precise [12-15].

Bruce and Tsotsos [12] modeled bottom-up saliency as the maximum information sampled from an image. More specifically, saliency is computed as Shannon's self-

information $-\log p(f)$, where $f$ is a local visual feature vector (i.e., derived from independent component analysis (ICA) performed on a large sample of small RGB patches in the images.) The probability density function is estimated based on a Gaussian kernel density estimate in a neural circuit.

Oliva and Torralba [16] proposed a Bayesian framework for the task of visual search (i.e., whether a target is present or not.) They modeled bottom-up saliency as $1/p(f|f_G)$ where $f_G$ represents a global feature that summarizes the appearance of the scene and approximated this conditional probability density function by fitting to multivariate exponential distribution. Zhang et al. [13] also proposed saliency detection using natural statistics (SUN) based on a similar Bayesian framework to estimate the probability of a target at every location. They also claimed that their saliency measure emerges from the use of Shannon's self-information under certain assumptions. They used ICA features as similarly done in [12], but their method differs from [12] in that natural image statistics were applied to determine the density function of ICA features.

Most of the methods [9, 11, 15, 16] based on Gabor or DoG filter responses require many design parameters such as the number of filters, type of filters, choice of the nonlinearities, and a proper normalization scheme. These methods tend to emphasize textured areas as being salient regardless of their context. In order to deal with these problems, [12, 13] adopted non-linear features that model complex cells or neurons in higher levels of the visual system. Kienzle et al. [18] further proposed to learn a visual saliency model directly from human eyetracking data using a support vector machine (SVM).

Different from traditional image statistical models, a spectral residual (SR) approach based on the Fourier transform was recently proposed by Hou and Zhang [19]. Spectral residual does not rely on parameters and detects saliency rapidly. In this approach, the difference between the log spectrum of an image and its smoothed version is the spectral residual of the image. However, Guo and Zhang [20] claimed that what plays an important role for saliency detection is not SR, but the image's phase spectrum.

## III. Proposed Saliency Algorithm

In this section, we will state the framework of our saliency detection method in details. The steps of our algorithm are fourfold: representing the image patches, using PCA to reducing dimensionality, computing each patch's saliency value and implementing our method to multiple scales.

### Image Patches Representation

Given an image $I$ with dimension $H \times W$, non-overlapping patches with the size of $n \times n$ pixels are drawn from it. The total number of patches is $L = \lfloor H/n \rfloor \cdot \lfloor W/n \rfloor$. Denote the patch as $p_i$, $i = 1,2,\cdots,L$. Then each patch is represented as a column vector $x_i$ of pixel values. The length of the vector is $3n^2$ since the color space has three components. Finally, we get a sample matrix $X = [x_1, x_2, \cdots, x_L]$, $L$ is the total number of patches as stated above.

### Dimensionality Reduction

To effectively describe patches in a relatively low dimensional space, we used an equivalent method to PCA to reduce data dimension. Each column in the matrix $X$ subtracts the average along the columns. Then, we calculated the co-similarity matrix $A = (X^T X)/L^2$, so the size of the matrix $A$ is $L \times L$. The eigenvalues and eigenvectors were calculated based on the matrix $A$ selected with their eigenvector $U = [u_1, u_2, \cdots, u_d]^T$ according to the biggest $d$ eigenvalues, where $u_i$ is an eigenvector. The size of the matrix $U$ is $d \times L$.

### Detection of the Patch's Saliency

New algorithm considers two factors for evaluating the saliency: the dissimilarities of color between image patches in a reduced dimensional space, and their spatial distance.

A patch is salient if the color of its pixels is unique. We should not look at an isolated patch, but at its surrounding patches, which lead to a similar meaning of center-surrounding contrast [17] method. Thus, a patch $p_i$ is considered as salient area if the appearance of the patch $p_i$ is distinctive with respect to all other image patches.

Specifically, let $dist_{color}(p_i, p_j)$ be the distance between the patches $p_i$ and $p_j$ in the reduced dimensional space. Patch $p_i$ is considered salient when $dist_{color}(p_i, p_j)$ is high for $\forall j$.

$$dist_{color}(p_i, p_j) = \sum_{n=1}^{d} |u_{ni} - u_{nj}| \quad (1)$$

The positional distance between patches is also an important factor. Generally speaking, background patches are likely to have many similar patches both near and far-away in the image. It is in contrast to salient patches that the latter tend to be grouped together. This implies that a patch $p_i$ is salient when the patches similar to it are nearby, and it is less salient when the resembling patches are far away.

Let $dist(p_i, p_j)$ be the Euclidean distance between the positions of patches $p_i$ and $p_j$, which is represented by the two centers of patches $p_i$ and $p_j$ in the image, normalized by the larger image dimension. Based on the observations above we define a dissimilarity measure between a pair of patches $p_i$ and $p_j$ as:

$$dissimilarity(p_i, p_j) = \frac{dist_{color}(p_i, p_j)}{1 + dist(p_i, p_j)} \quad (2)$$

This dissimilarity measure is proportional to the difference in appearance and inverse proportional to the positional distance.

To evaluate a patch's uniqueness, we can compute the dissimilarity between the patch and all the other patches and take the sum of these dissimilarities as the saliency of related patch. In practice, there is no need to incorporate its dissimilarity to all other image patches. It suffices to consider the K most similar patches that if the most similar patches are highly different from $p_i$, then clearly all image patches are highly different from $p_i$. Hence, for every patch $p_i$, we search for the $K$ most similar patches $\{q_i\}, i = 1,2,\cdots,K$ in the image, according to (2). Under this definition, our algorithm that measures the saliency value from the perspective of global information and local information is different from global saliency detection [21] and Duan's method [6]. A patch $p_i$ is salient when $dissimilarity(p_i, q_k)$ is high for $\forall k \in [1, K]$. The saliency of patch $p_i$ is defined as (we choose $K = 100$ in our experiments):

$$S_i = 1 - \exp\{-\frac{1}{K}\sum_{k=1}^{K} dissimilarity(p_i, q_k)\} \quad (3)$$

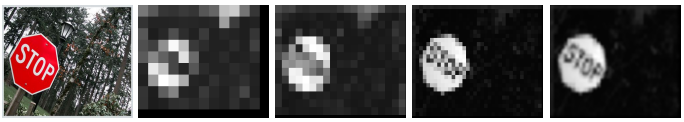
The computational complexity of our algorithm includes twofold: first is the computational complexity to preprocessing, such as dividing original images into patches and PCA; the other time consuming cost is computing dissimilarities between patches. Given an input image with size of $H \times W$ (where $H$ is the height and $W$ is the width) and the patch size of $n \times n$, the computational complexity of our algorithm is $(L^3 + L^2)$, in which $L^3$ and $L^2$ corresponding to the computational cost of preprocessing and dissimilarity calculation respectively, where $L = \lfloor H/n \rfloor \cdot \lfloor W/n \rfloor$. Therefore, with the smaller patch size, the computational complexity will increase.

In addition, large patch size maybe lead to another problem. In the saliency map, the saliency value of all pixels in a patch is decided by the dissimilarity of this patch and all other patches. Therefore, the saliency value in a patch is the same. It is easy to make our algorithm that can not describe the boundary of small salient object when the patch size is larger than the salient object. Therefore, we use (3) to compute saliency value of the original image with different patch sizes can obtain the saliency map with different scales.

### IMPLEMENTATION BY MULTIPLE SCALES

Based on the observation that patches in foreground are likely to have similar patches at multiple scales, which is in contrast to more non-salient patches that could have similar patches at a few scales but not at all of them(It is equal to the principle proposed by [10] that salient object always smaller than the background). Therefore, we wish to incorporate multiple scales to further decrease the saliency of background patches, improving the contrast between salient and non-salient regions.

In addition, the patch with large scale can not describe the boundary of small salient object. So we hope to use different scales that large scale to detect the whole information and the small scale to describe the salient object in details. Last, we compile all saliency value into final saliency. The results of different scales and the final result illustrated in Figure1. The number of PCs set to 4.




(a) (b) (c) (d) (e)

Figure 1. The original image and its different saliency maps with different patch sizes; (a )original image; (b) patch size=30; (c) patch size=20; (d) patch size=10; (e) final result.

For a patch $p_i$ of scale $r$, the saliency value according to (3) is defined as

$$S_i^r = 1 - \exp\{-\frac{1}{K}\sum_{k=1}^{K} dissimilarity(p_i^r, q_k^r)\} \quad (4)$$

We consider the scales $R_c = \{r_1, r_2, \cdots, r_M\}$, using (4) to calculate the saliency of patch $i$ as $\{S_i^{r_1}, S_i^{r_2}, \cdots, S_i^{r_M}\}$. The final saliency is computing as

$$S_i = \frac{1}{M}\sum_{r \in R_c} S_i^r \quad (5)$$

## IV. EXPERIMENT VALIDATION

### PREDICTING HUMAN VISUAL FIXATION DATA

In this section, we show several experimental results on detecting saliency in natural images. We use an image dataset and its fixation data collected by Bruce and Tsotsos [12] as a benchmark for comparison. This dataset contains eye fixation records from 20 subjects for a total of 120 images of size $681 \times 511$. To compare our results with [6], we choose 11 as reduced dimension which is the best value to maximize saliency predictions. For the patch size, we choose $\{30,20,10\}$ because better results are easy to obtain in these values [6]. We obtain an overall saliency map by using YCbCr color space in all experiments. Some visual results of our algorithm are compared with the advanced methods in Figure2.

The comparison results show that the most salient locations on our saliency maps are more consistent with the human fixation density maps. Note that our method is much less sensi-

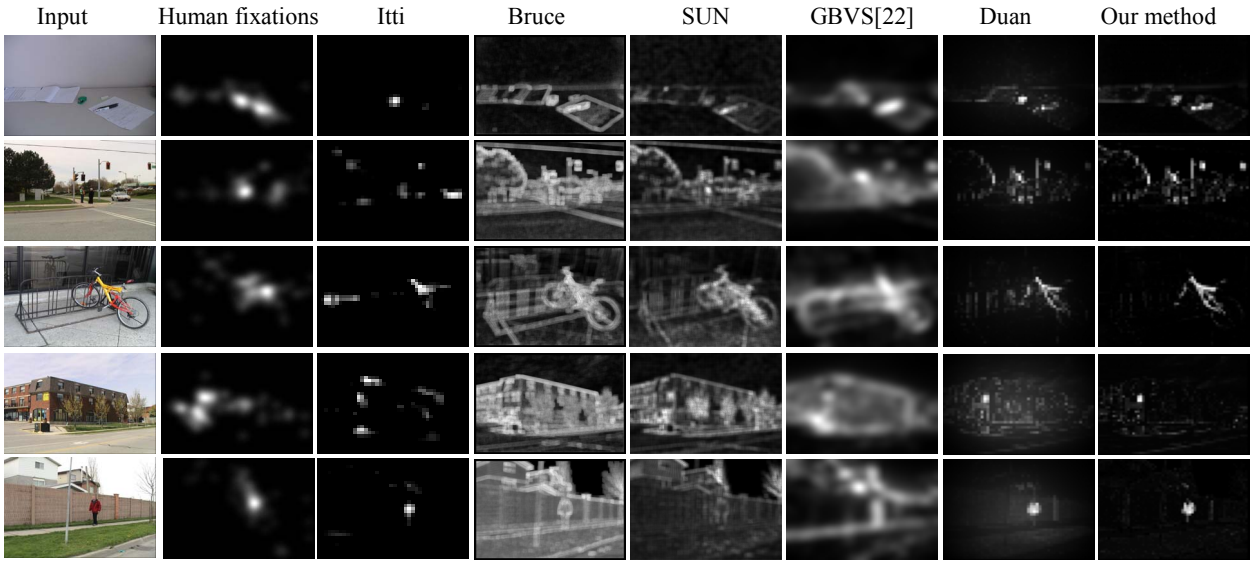



Figure 2. Results on predicting human visual fixation data

TABLE I. PERFORMANCE ON PREDICTING HUMAN VISUAL FIXATION DATA

| Attention Model | AUC |
|---|---|
| Itti [11] | 0.7049 |
| Bruce [12] | 0.7613 |
| SUN [13] | 0.7974 |
| GBVS [23] | 0.8021 |
| Duan [6] | 0.8333 |
| Our method | 0.8346 |

tive to background texture, which is different from Bruce method and SUN. Duan's method, which used the center bias mechanism indicating a strong bias to the center of the image, is easy to detect the salient object near the center of image. Our algorithm detects the dominant object not only in the center but also far away from the center. To compare the saliency maps with the human fixations, we use the popular validation approach as Bruce et al. introduced in [12]. The area receiver operating characteristic (ROC) curve, i.e., the area under the curve (AUC) to quantitatively evaluate the algorithm performance (reported in Table 1). In Table 1, the results between Duan's method and our method are similar (corresponding AUC are 0.8333 and 0.8346 respectively), both are better than other four models. In [13], Zhang et al. point out that the dataset collected by Bruce [12] is center-biased. The characteristic of Bruce database maybe benefits Duan's method and other saliency models that based on the center bias mechanism.

## SALIENT OBJECT SEGMENTATION DATABASE

We have evaluated the results of our approach on the publicly available database provided by Achanta et al. [9]. According to the best of our knowledge, the database is the largest of its kind, and has ground truth in the form of accurate human-marked labels for salient regions. We compared the proposed method with state-of-the-art saliency detection methods.

We used our methods and the others to compute saliency maps for all the 1000 images in the database. To reliably compare how many PCs choose to get best results, we vary the number of PCs from 2 to 20. Figure 3 shows the resulting recall vs. number of PCs. From Figure 3, we choose 4 as PCs number. With the number of PCs increase, precision is decrease. The patch size set to be the same to last section, which is still the best parameter in this database.

Visual comparison of saliency maps obtained by the various methods can be seen in Figure 4. Note that Duan's method often detect the foreground in the image, however, the saliency just focus on the center of the image. This characteristic might be lead to that parts of the foreground near the center is more salient than that far from the center or the parts of background near the center is more salient than the foreground in the image's boundary. Our method overcomes this problem that can detect the saliency in the whole image. In addition, comparing to Duan's method, our algorithm can detect the salient object with precision boundary.

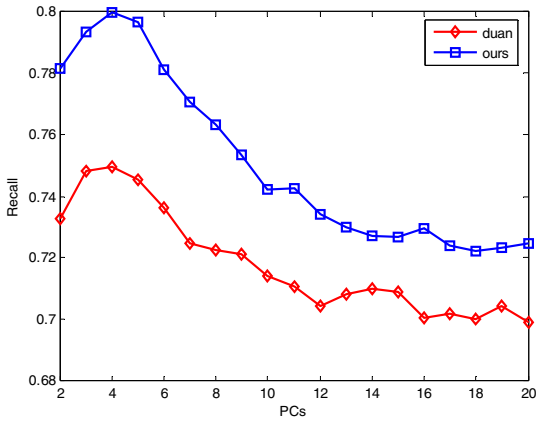



Figure 3. The relationship between the Recall and the dimension when patch size set to 14.

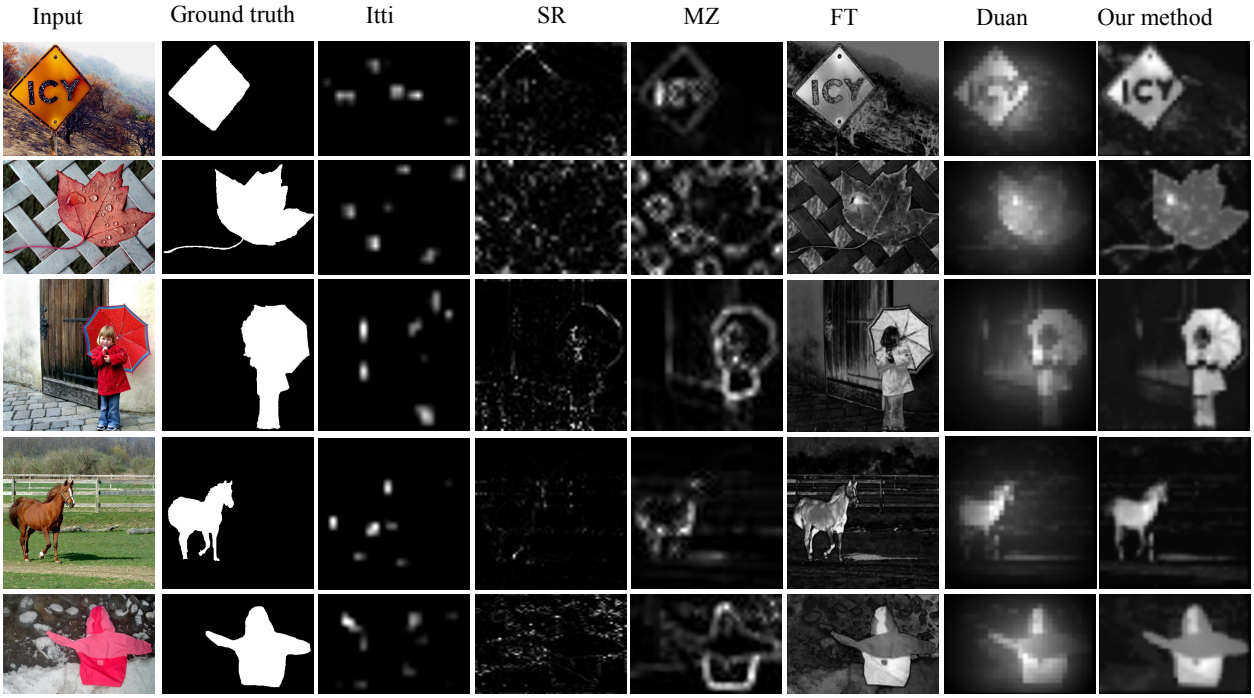



Figure 4. Saliency maps from different saliency detection models on segmentation image database.

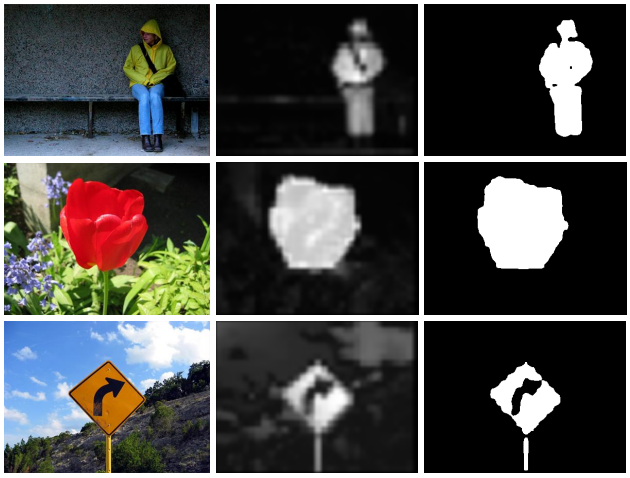


Figure 5. Binary masks for saliency maps by the proposed algorithm.

In order to comprehensively evaluate the accuracy of our method for salient object segmentation, we use the mean-shift segmentation algorithm [22] to obtain a binary mask for given saliency map (see Figure5). The parameters of meanshift and the adaptive threshold value are set to the same as [9].

Using this approach we obtain binarized maps of salient object from each of the saliency algorithms. Average values of precision, recall and F-measure are obtained over the ground-truth database.

$$F_\beta = \frac{(1+\beta^2)\,precision \times recall}{\beta^2 \times precision + recall} \qquad (6)$$

We use $\beta^2 = 0.3$ in our work to weigh precision more than recall. The comparison is shown in Figure 6. Itti's method and SR show a high precision but very poor recall, indicating that they are better suited for gaze-tracking experiments, but perhaps not well suited for salient object segmentation. FT, which is better than other methods except ours, is sensitive to the background textures. Duan's method, which precision is the same as our method, has the poor recall values. Our algorithm shows the highest F-measure and recalls which indicating that the salient object boundary can be better preserved.

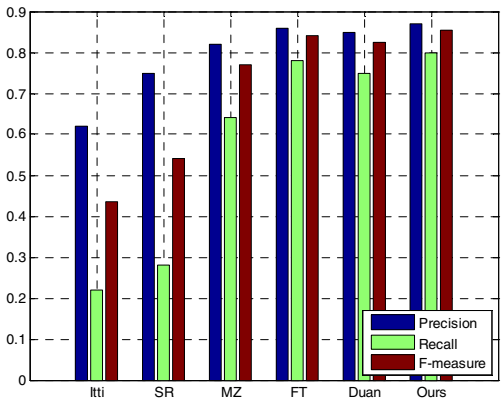



Figure 6. The experiment results for the comparison between our proposed model and other state-of-art methods.

## V. CONCLUSION

We presented a multiscale saliency detection algorithm based on PCA to detect the saliency object in the color image. Our saliency algorithm is based on three elements: the color dissimilarity of patches, the spatial distance and multiple scales. We evaluated our method on two publicly available data sets and compared our scheme with other models. The resulting saliency maps are a little improvement on the database with center bias mechanism, however, are better suited to salient object segmentation which preserving more fine details, demonstrating both higher precision and recall than other state-of –art models.

In the future, we plan to investigate the practicability of the proposed saliency maps can be used for efficient object detection, reliable image classification, robust image scene analysis, leading to improved image retrieval.

## ACKNOWLEDGEMENT

This work is supported by Natural Science Foundation of China (No.90820306 and No.KT06015).

## REFERENCES

[1] A.Santella, M.Agrawala, D.Decarlo, D.Salesin, and M.Cohen. Gaze-based interaction for semi-automatic photo cropping. ACM Human Factors in Computing Systems (CHI), pages 771–780, 2006.

[2] L.Chen, X.Xie, X.Fan, W.Ma, H.Shang, and H.Zhou. A visual attention mode for adapting images on small displays. Technical report, Microsoft Research, Redmond, WA, 2002.

[3] L.Itti. Models of Bottom-Up and Top-Down Visual Attention. PhD thesis, California Institute of Technology Pasadena, 2000.

[4] V.Navalpakkam and L.Itti. An integrated model of top-down and bottom-up attention for optimizing detection speed. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2049–2056, 2006.

[5] U.Rutishauser, D.Walther, C.Koch, and P.Perona. Is bottom-up attention useful for object recognition? IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 37–44, 2004.

[6] L. Duan, C.Wu, J. Miao, L. Qing and Y. Fu, Visual Saliency Detection by Spatially Weighted Dissimilarity, IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pages 21-23, 2011.

[7] B. W. Tatler. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. Journal of Vision, 7(14):4, pages 1-17, 2007.

[8] J. Yang, D. Zhang, A. F. Frangi and J. Y. Yang. Two-dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. IEEE Trans. Pattern Analysis and Machine Intelligence, vol.26, no.1, pages 131-137, 2004.

[9] R.Achanta, S.Hemami, F.Estrada, and S.S¨ usstrunk. Frequency-tuned salient region detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1597–1604, 2009.

[10] T.Avraham and M.Lindenbaum, "Esaliency (Extended Saliency): Meaningful Attention Using Stochastic Image Modeling," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 32, no. 4, pages 693-708, 2010.

[11] L.Itti, C.Koch and E.Niebur. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Analysis and Machine Intelligence, vol.20, no.11, pages 1254-1259, 1998.

[12] N. Bruce and J. Tsotsos. Saliency based on information maximization. In Advances in Neural Information Processing Systems, pages 155–162, 2006.

[13] L.Zhang, M.Tong, T.Marks, H.Shan, and G. Cottrell. SUN: A Bayesian framework for saliency using natural statistics. Journal of Vision, 8(7):32, pages 1–20, 2008.

[14] D.Gao and N.Vasconcelos. Discriminant saliency for visual recognition from cluttered scenes. In Advances in Neural Information Processing Systems, pages 481–488, 2004.

[15] Pierre Baldi, Laurent Itti. Of bits and wows: A Bayesian theory of surpise with applications to attention. Neural Networks, 23(5), pages 649-666, 2010

[16] A.Oliva, A. Torralba,M. Castelhano, and J. Henderson. Top-down control of visual attention in object detection. In Proceedings of International Conference on Image Processing, pages 253–256, 2003.

[17] D. Gao, V. Mahadevan, and N. Vasconcelos. On the plausibility of the discriminant center-surround hypothesis for visual saliency. Journal of Vision, 8(7):13, pages 1–18, 2008.

[18] W.Kienzle, F.Wichmann, B.Scholkopf, and M. Franz. A nonparametric approach to bottom-up visual saliency. In Advances in Neural Information Processing Systems, pages 689–696, 2007.

[19] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1–8, 2008.

[20] 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 (CVPR), pages 1–8, 2008.

[21] M. Mancas et al., A Rarity-Based Visual Attention Map Application to Texture Description, In Proceedings of International Conference on Image Processing, pages 445-448, 2006.

[22] C.Christoudias, B.Georgescu, and P. Meer. Synergism in low level vision. In Proceedings of International Conference on Pattern Recognition, pages 150-155, 2002.

[23] J.Harel, C.Koch, and P.Perona, Graph-Based Visual Saliency, Advances in Neural Information Processing Systems, pages 545-55， 2003