# What Are We Looking For: Towards Statistical Modeling of Saccadic Eye Movements and Visual Saliency

Xiaoshuai Sun, Hongxun Yao, Rongrong Ji
Dept. of Computer Science, Harbin Institute of Technology, Heilongjiang, China
{xiaoshuaisun, H.yao, rrji}@hit.edu.cn

## Abstract

*In this paper, we present a unified statistical framework for modeling both saccadic eye movements and visual saliency. By analyzing the statistical properties of human eye fixations on natural images, we found that human attention is sparsely distributed and usually deployed to locations with abundant structural information. This new observations inspired us to model saccadic behavior and visual saliency based on Super Gaussian Component (SGC) analysis. The model sequentially obtains SGC using projection pursuit, and generates eye-movements by selecting the location with maximum SGC response. Beside human saccadic behavior simulation, we also demonstrated our superior effectiveness and robustness over state-of-the-arts by carrying out dense experiments on psychological patterns and human eye fixation benchmarks. These results also show promising potentials of statistical approaches for human behavior research.*

## 1. Introduction

Attention guided saccadic eye-moment is one of the most important mechanisms in biological vision systems, based on which the viewer is able to actively explore the environment with high resolution fovea sensors. Benefitting from such unique behavior, human beings, as well as most primates, are able to efficiently process the information from complex environments. For the last four decades, extensive research works have been done by means of theoretical reasoning and computational modeling, trying to uncover the principles that underlie the deployment of gaze. Compared with theoretic hypotheses, computational models of visual attention and saccadic eye-movement not only help us better understand the mechanism of human cognitive behavior but also provide us powerful tools to solve various vision related problems such as video compression [1], scene understanding [2], object detection and recognition [3] *etc.*

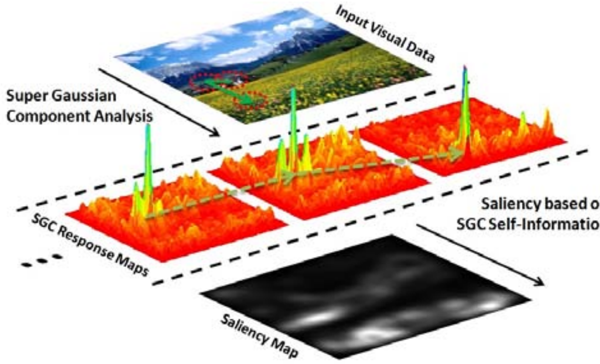

Input Visual Data
Super Gaussian Component Analysis
Saliency based on SGC Self-Information
SGC Response Maps
Saliency Map

Figure 1. What are we looking for when viewing a scene? Our studies suggest that the answer to this question could be revealed via a statistical analysis of human eye fixations. One possible answer named Super Gaussian Component is investigated in this paper.

In this paper, our goal is to establish a statistical framework for both saccadic behavior simulation and visual saliency analysis. Different with previous works that drew inspirations from the existing neurobiological knowledge or mathematical theories, we directly make assumptions based on the statistical analysis of the ground truth human eye-fixations. By means of statistical analysis, we try to find out "what components in visual images draw fixations" which is similar but more reachable compared with the traditional question of "what properties draw attention". The analysis is conducted on eye fixation data captured from human observers using an eye tracking device during task independent free viewing of natural images. In such bottom-up scenario, we have found an interesting phenomenon, which might further be proved as a general principle, that stimuli with a super Gaussian distribution is more likely to gather human gaze. Based on this finding, human saccadic behavior can be modeled as a function of active information pursuit targeting at the statistical components with desired properties such as super Gaussianity.

In our framework (Figure 2), visual data is represented as an ensemble of small image patches. Kurtosis maximization is adopted to search for the Super Gaussian Component

978-1-4673-1228-8/12/$31.00 ©2012 IEEE

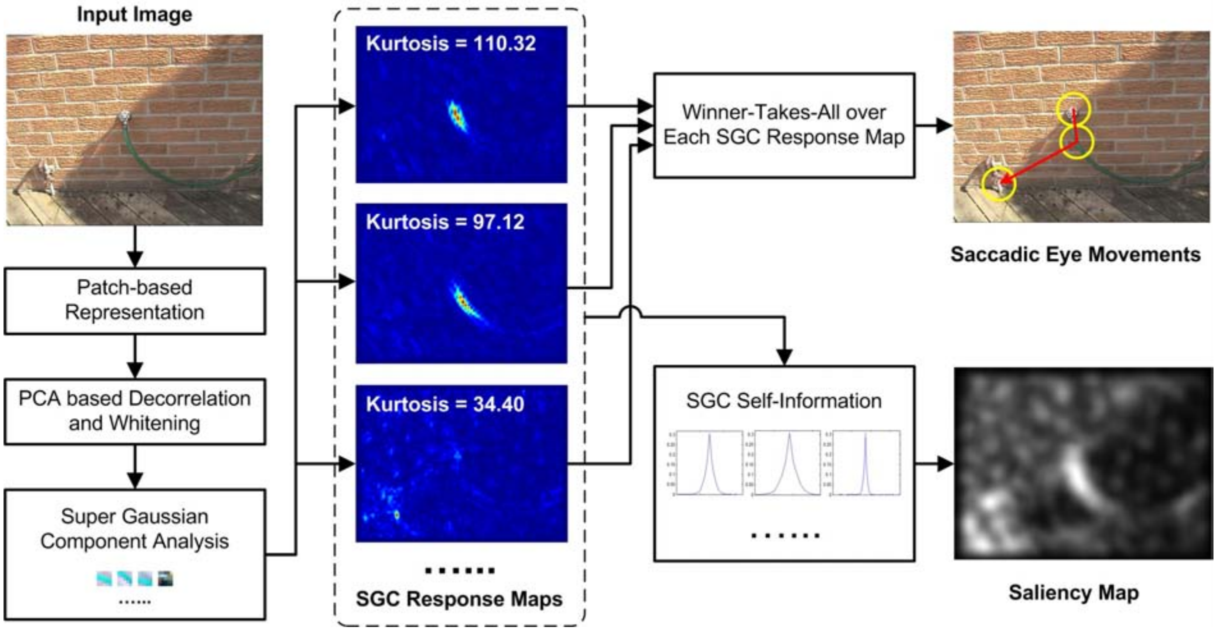


Figure 2. The proposed framework. The input signal is first processed into a patch-based representation. The patch matrix is then whitened in order to simplify the subsequent process. Aiming at super-Gaussianity maximization, projection pursuit is performed iteratively on the whitened data resulting in multiple Super Gaussian Components (**SGC**). Finally, the model generates eye-movements and estimates visual saliency based on the response maps of the SGCs.

(SGC). A response map is then obtained by filtering the original image with the found SGC. Based on the response map, we adopt a well known principle named winner-takes-all (**WTA**) to select and locate the simulated fixation point. Gram-Schmidt orthogonal method is applied at the beginning of each selection to avoid convergence at the same location. Along with the saccadic simulation, a saliency map can also be estimated using either the selected fixations or the response maps. The proposed framework enables fast selection of a small number of fixations, which give processing priority to the most important components of the visual input. Different from low-level feature-based saliency driven approaches, the proposed gaze selection method is guided by high-level feature-independent statistical cues, which is supported by findings observed from real-world fixation analysis.

## 1.1. Related Works

In the literature, it is widely agreed that eye-movement is guided by both bottom-up (stimulus-driven) and top-down (task-driven) factors [2, 4].

The bottom-up stimulus-driven research mainly focuses on saliency-driven approaches, in which a saliency map is pre-computed using low-level image features to guide task independent gaze allocation. These methods have been proven to be very effective in predicting eye fxations captured from human subjects while viewing natural images and video sequences. Itti *et al*. [2] proposed a computational attention model based on Koch and Ullman's attentional selection architecture [5], in which visual saliency is measured by spatial center-surround differences across several feature channels and different scales. In the model of [2, 6], two principles named winner-takes-all (**WTA**) and inhibition-of-return (**IoR**) are adopted to select fixations based on saliency maps. This technique is widely used for scanning visual scene or generating artificial saccades. Bruce and Tsotsos [7] proposed a framework based on image sparse representation and the principle of information maximization, where visual saliency is measured by the self-information of the sparse coefficients. Also based on sparse coding, Hou *et al*. [8] argued that visual saliency should be dynamically measured by the incremental coding length of the sparse features. Wang *et al*. [9] adopt Site Entropy Rate as a saliency measure based on some well acknowledged biological facts with respect to both sparse coding and neuron activities in human vision system. Integrated with more biological factors, Wang *et al*. [10] extended their model to simulate saccadic scanpaths on natural images. Despite the above models, there are also many other works which present insightful saliency measures such as Bayesian Surprise [11], Center-Surround Discriminant Power[12], Spatially Weighted Dissimilarity [13] *etc*.

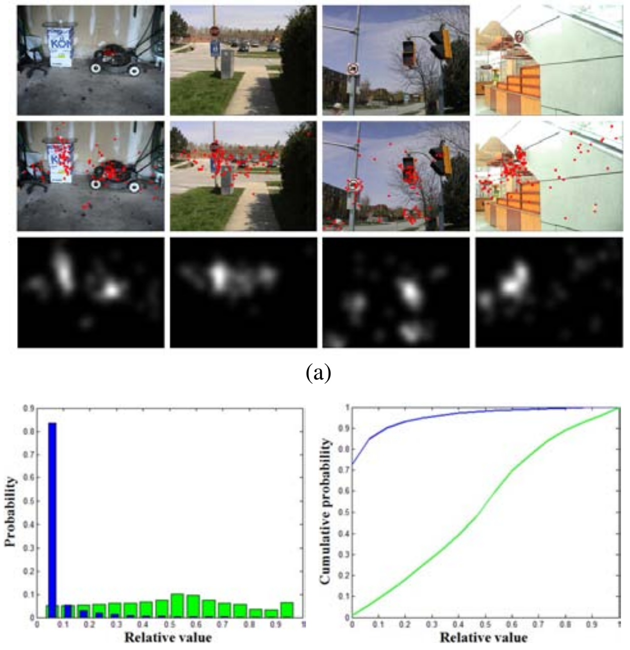
For top-down research, there are also extensive studies of human saccadic behavior during different real-world tasks such as making a sandwich, fixing a cup of tea or learning and matching a shape. Most studies indicate that eye-movements are probably made to collect task-relevant information [14]. Foulsham *et al*. [15] ask the participants to view color photographs of natural scenes in preparation for a memory test. Eye movements were recorded during the viewing and testing process. Analysis on these eye-tracking data indicates that saliency model work much better than random models but still may be missing out on sequential aspects of oculomotor control that could potentially predict fixation much better than saliency alone. Based on these previous findings, we construct our framework not only based on statistical factors but also considering the sequential aspects of human perception.

## 2. Statistical Analysis of Human Fixation Data

Although there are many research works that address the problem of saliency detection, the statistical analysis of saliency is still non-trivial cause there are no recordable "ground truth" saliency maps. In [7], a fixation density map is produced for each image based on human eye fixation points. The fixation density map comprises the probability of each pixel in the image being sampled by human observers based on their eye fixations. Taking fixation density as the approximation of saliency makes it possible for us to quantitatively analyze the statistical properties of saliency. Specifically, we use eye fixation data from two benchmark datasets (Bruce *et al*. [7] and Judd *et al*. [16]) for intuitive and statistical analysis of human saccades. As discussed in [16], for some images, all viewers fixate on the same locations, while in other images viewers' fixations are dispersed all over the image. To minimize the disturbance caused by the subjects' personal factors, we manually filter out images with large subject-wise inconsistency. Figure 3 shows some example images along with the eye fixation points and the corresponding density maps. We also give a comparison on the probability density distribution between saliency and pixel values. From intuitive and statistical observation, we found two interesting characteristics of visual saliency:

- Saliency is very *sparse*, which means the saliency of most locations is zero and only a small portion of the image has obvious high saliency value;
- High saliency value tends to be located surrounding the regions with abundant structural information.

According to feature integration theory [17], saliency is obtained by integration of multiple feature channels. Thus, features used for saliency detection should share similar statistical characteristics with saliency. From a statistical point of view, the above characteristics of saliency share great similarity with super-Gaussianity, which is synonymous with "sparse" and "structurized" in statistics. Considering the above issues, we proposed the primary assumption that Super Gaussian Components (SGC) of the scene are exactly what we are looking for during the viewing process.


(a)

(b)

(c)

Figure 3. Real-world distribution of bottom-up visual saliency. In (a), Top: natural images; Middle: images covered with human eye fixations (red dots); Bottom: eye-fixation density maps. (b) shows the probability density of saliency (blue) and image pixels (green). (c) is the corresponding cumulative distribution of (b).

## 3. The Model

In this section, we present two major components of our model in details including sequential gaze selection and visual saliency estimation. The sequential aspects missing in previous works and the statistical assumption we made in Section 2 are both considered in our model.

### 3.1. Sequential gaze selection

There is one statistical technique named projection pursuit that shares a similar sequential selection behavior with saccadic eye-movement. As a simple yet powerful tool, projection pursuit is widely used for high-dimensional data visualization and statistical component analysis, *e.g*. ICA [18]. In the case of saccadic modeling, we adopt projection pursuit to search for the **SG** components, which are further used for gaze localization and saliency estimation. From a signal processing point of view, this scheme can also be regarded as unsupervised function that dynamically separates the image data into a salient gaze-favored part and non-salient unattractive part. Technique details are described in 3.1.1 and 3.1.2.

### 3.1.1 Super Gaussian Component Analysis

**Data preparation** Given an image $I$, we first turn it into a patch-based representation $\mathbf{X}$ by scanning $I$ with a sliding window from top-left to bottom-right. $\mathbf{X}$ is stored as a $M \times N$ matrix, in which each column vector corresponds to a reshaped image patch. PCA based decorrelation and whitening are then applied to $\mathbf{X}$, resulting in a new matrix $\mathbf{Z}$ which will simplify the subsequent calculations [18].

**Single SGC pursuit** In statistics, the super-Gaussianity of a random variable is usually measured by the kurtosis function which is defined as:

$$\text{kurt}(y) = \mathbf{E}\{y^4\} - 3(\mathbf{E}\{y^2\})^2, \tag{1}$$

where $y$ is the given random variable, $\mathbf{E}\{.\}$ is the expectation function. If $y$ is a gaussian random variable, kurt$(y)$ will be 0. If kurtosis is positive, the variable is called super-Gaussian which is also an alternative definition of sparsity. For whitened variable $y$, its standard deviation $\mathbf{E}\{y^2\} = 1$. So the kurtosis function can be further simplified as $\mathbf{E}\{y^4\} - 3$. To maximize the kurtosis, *i.e*. maximizing super-Gaussianity, we can start from a random selected projection $\mathbf{w}$, and iteratively change its direction using fixed-point iteration method based on the available samples denoted as $\mathbf{Z}$. Now we give a formal objective function $G_p$ for single SGC pursuit:

$$G_p(\mathbf{w}) = \text{kurt}(\mathbf{w^T Z}). \tag{2}$$

The gradient of $G_p$ has the following form:

$$\frac{\partial G_p}{\partial \mathbf{w}} = 4[\mathbf{E}\{(\mathbf{w^T Z})^\mathbf{3} \mathbf{Z^T}\} - \mathbf{3w^T}\|\mathbf{w^T}\|^\mathbf{2}]. \tag{3}$$

During optimization, the iteration reaches convergence when the gradient vector and the projection vector have the same direction. Let Equation 3 be equal to $\mathbf{w}$, we have:

$$\mathbf{w^T} \propto [\mathbf{E}\{(\mathbf{w^T Z})^\mathbf{3} \mathbf{Z^T}\} - \mathbf{3w^T}\|\mathbf{w^T}\|^\mathbf{2}]. \tag{4}$$

Equation 4 leads to a fixed-point iteration algorithm:

$$\begin{aligned} &\mathbf{w} \leftarrow \mathbf{E}\{\mathbf{Z}(\mathbf{Z^T w})^\mathbf{3}\}, \\ &\mathbf{w} = \mathbf{w}/\|\mathbf{w}\|. \end{aligned} \tag{5}$$

Based on Equation 5, we can get a projection vector which maximizes the super-Gaussianity of the projected data. There are two conditions that will make the iteration stop:1. $\|\triangle\mathbf{w}\| < \epsilon$, where $\triangle\mathbf{w}$ is the difference of $\mathbf{w}$ after one iteration and $\epsilon$ is a convergence threshold; 2. Optimization didn't converge within a limited number of iterations.

**Multiple SGC pursuit** Based on the single component pursuit method, multiple SGC pursuit can be implemented by applying the same optimization method under the constrain that the new SGC should be orthogonal to the previous ones. The orthogonal constrain prevents the optimization process from converging on the same directions of the previous pursuit processes. Practically, we use Gram-Schmidt orthogonal method for orthogonalization. Given a set of predefined projections: $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_p$, we ensure the orthogonal constrain by adding the following orthogonalization procedure:

$$\mathbf{w}_{p+1} \leftarrow \mathbf{w}_{p+1} - \sum_{j=1}^{p} (\mathbf{w}_{p+1}^\mathbf{T} \mathbf{w}_j)\mathbf{w}_j. \tag{6}$$

The normalization $\mathbf{w} = \mathbf{w}/\|\mathbf{w}\|$ in Equation 5 can be repositioned at the end of each iteration. Based on Equation 5 and 6, multiple SGC pursuit can be performed in a one-by-one manner.

### 3.1.2 WTA based Gaze Localization

For each SGC, we generate a response map by treating the component as a linear filter.

$$\mathbf{RM}_i = \mathbf{w}_i^\mathbf{T} \mathbf{Z} \tag{7}$$

where $\mathbf{RM}_i(j)$ denotes the response value of $j$th patch for the $i$th SGC. Similar with [2], we select the location with largest response value as the gaze point following the WTA principle. Figure 4 shows the visualized gaze selection process on natural images, along with the attended sub-regions. Similar gaze selection behavior between human observer and the proposed model could be observed in Figure 5.

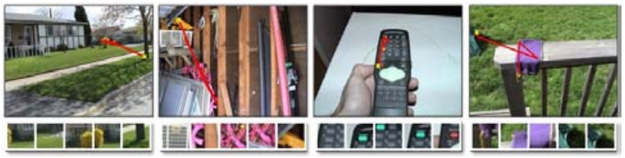


Figure 4. Eye-movements generated by our model. For each image we show the scanpath with five saccades and the corresponding focused regions.

## 3.2. Visual Saliency Estimation

We measure saliency by self-information of the super Gaussian components. As the SC components are acquired sequentially in our model, the saliency map is estimated also in a dynamical manner. The more SG components are involved, the more details will appear in the saliency

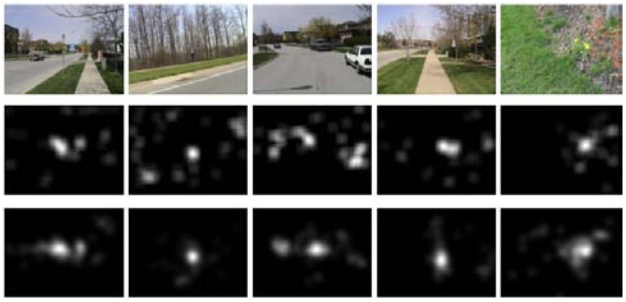


Figure 5. Comparisons of fixation density maps between the proposed model and human observers. Top: input images; middle: fixation density maps generated by our model using 75 fixations per image; bottom: fixation density maps of human eye fixations.

map. Given $k$ response maps: $\mathbf{RM}_1, \mathbf{RM}_2, ..., \mathbf{RM}_k$, the bottom-up saliency of $j$th patch is defined as:

$$\mathbf{S}(j) = -\log \prod_{i=1}^{k} p_i(\mathbf{RM}_i(j)) = -\sum_{i=1}^{k} \log p_i(\mathbf{RM}_i(j)), \tag{8}$$

where $p_i(.)$ is the probability density function of the $i$th S-GC. For simplicity, we estimate $p_i(.)$ by histogram method. An implementation for both gaze selection and saliency estimation is presented in Algorithm 1 with default parameter settings.

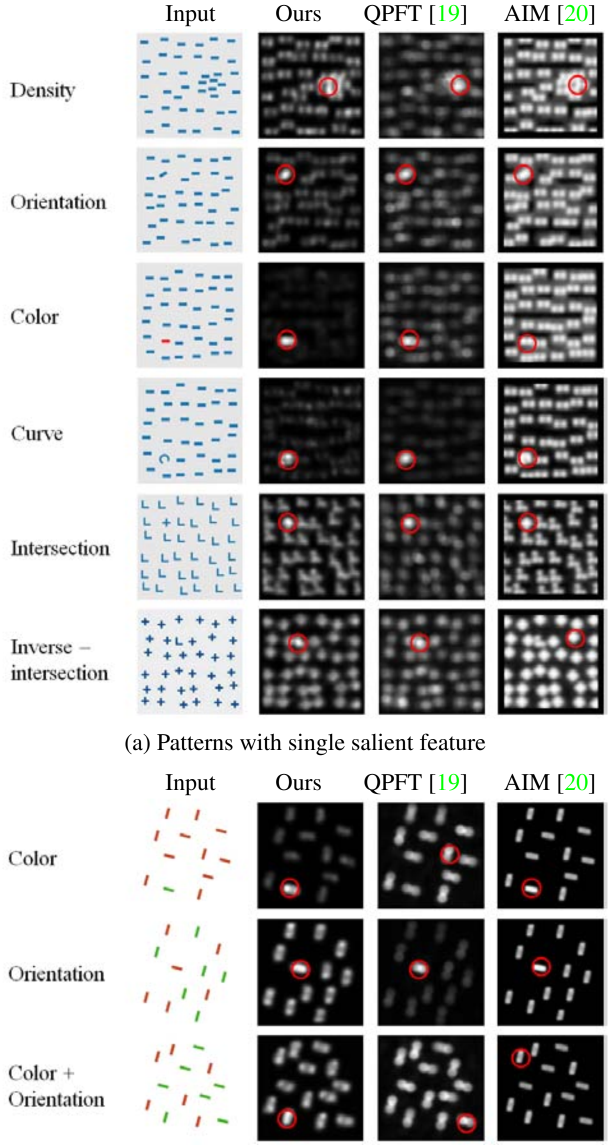
**Algorithm 1:** Gaze selection and saliency estimation

```
Input: $M \times N$ data matrix $\mathbf{Z}$, $M \times M$ zero matrix $\mathbf{B}$,
       maximum iteration $\theta = 500$, convergence threshold
       $\epsilon$=0.0001
Output: Fixation sequence $F$, Saliency map $\mathbf{S}$
Set projection index $k = 1$;
while $k < M$ do
    Generate random vector $\mathbf{w} = [w_1, w_2, ..., w_M]$;
    Orthogonalize $\mathbf{w}$ by $\mathbf{w} = \mathbf{w} - \mathbf{B}\mathbf{B}^{\mathbf{T}}\mathbf{w}$;
    $\mathbf{w} = \mathbf{w}/\|\mathbf{w}\|, j = 1$;
    while $j < \theta$ and $\|\mathbf{w}' - \mathbf{w}\| < \epsilon$ do
        $\mathbf{w}' = \mathbf{w}$;
        $\mathbf{w} = \mathbf{Z}(\mathbf{Z}^{\mathbf{T}}\mathbf{w})^{\mathbf{3}}/N$;
        $\mathbf{w} = \mathbf{w}/\|\mathbf{w}\|$;
        $j = j + 1$;
    end
    Replace the $k$th column of $\mathbf{B}$ by $\mathbf{w}$;
    $\mathbf{RM}_k = \mathbf{w}^{\mathbf{T}}\mathbf{Z}$;
    $F = F \bigcup < k, \text{argmax}\mathbf{RM}_k >$;
    $k = k + 1$;
end
Generate $\mathbf{S}$ based on Equation 8;
Smooth $\mathbf{S}$ with a gaussian filter $(5 \times 5, \sigma = 2)$;
return $F, \mathbf{S}$;
```

# 4. Experiments

## 4.1. Response to psychological patterns

Response to psychological patterns adopted in attention related experiments can indicate the biological plausibility of the tested models. As shown in Figure 6, our method generates reasonable responses to not only normal patterns such as density, orientation, color, curve, insertion and inverse-intersection, but also patterns with conjunctive features.



(a) Patterns with single salient feature

(b) Patterns with conjunctive features

Figure 6. Response to psychological patterns. From left to right, we present the input image, saliency maps generated by our method, QPFT [19] and AIM [20]. Red circles indicate the location of the maximum saliency. Test images are from [19, 20]

## 4.2. Human Eye-fixation Prediction

This experiment is designed for evaluating the consistency between human eye fixations and the saliency map generated by the tested models. Experiments are conducted on two data sets: static image data from Bruce *et al.* [7][1] and dynamic video data from Itti *et al.* [11][2]. Models listed for comparisons are selected by 2 criterions: 1. commonly used benchmark and 2. open source. Following the proposal of [7, 11], we adopt area under the ROC curve (**AUC**) and K-L divergence for quantitative evaluation.

### 4.2.1 Experiment on Static Natural Images

Fixation dataset from Bruce and Tsotsos [7] contains 11,999 eye fixations captured from 20 human subjects free viewing 120 natural images for 4 seconds each. To reduce the influence caused by the subjects' personalized factors, we filter out spatially isolated saccades using the fixation density maps which are included in the dataset package. Each fixation density map is normalized to [0,1]. Fixations with density value greater than 0.5 are preserved, resulting in a sub fixation dataset containing 4339 samples.

Bruce *et al.* [7] proposed to use Area Under ROC curve (**AUC**) as a quantitative evaluation criterion for this experiment. However, the original **AUC** evaluation is largely affected by the "edge effect" due to center bias caused by the central composition of interesting objects. Zhang *et al.* [21] pointed out that a simple gaussian blob fitted to the eye fixations has a **AUC** score of 0.80 which exceeds most of the reported models on Bruce's data set [7]. To eliminate the interference caused by the "edge effect", we follow the proposal of Zhang *et al.* [21], and use a refined evaluation procedure to compute the **AUC** score. Specifically, we first compute the true positives from the saliency maps based on the human eye fixation points. In order to calculate false positives, we use the human fixation points from other images by permuting the order of images. This permutation of images is repeated for 100 times. Each time, we compute an AUC score by regarding the eye fixations from original image as the positive samples and the fixations from permutated images as the false samples.

We compared our model against the state-of-the-art models including Itti *et al.* [2], Bruce and Tsotsos [20], Zhang *et al.* [21], Hou *et al.* [8], Wang *et al.* [9] and Murray *et al.* [22]. Saliency maps of all tested methods are generated using their default parameter settings. Table 1 shows the mean and standard error of **AUC** scores and KL-divergence. Our method outperforms all state-of-the-art models in both **AUC** and **KL** evaluation. Figure 7 shows more visual comparisons.

| Method | AUC (SE) | KL (SE) |
|---|---|---|
| Itti and Koch [2] | 0.6249 (0.0008) | 0.1300 (0.0026) |
| Bruce and Tsotsos [20] | 0.7547 (0.0013) | 0.4140 (0.0045) |
| Zhang *et al.* [21] | 0.7345 (0.0015) | 0.2972 (0.0050) |
| Hou and Zhang [8] | 0.7708 (0.0013) | 0.5320 (0.0058) |
| Wang *et al.* [9] | 0.7594 (0.0012) | 0.4812 (0.0052) |
| Murray *et al.* [22] | 0.7707 (0.0013) | 0.4528 (0.0056) |
| Our Method | 0.**7903 (0.0012)** | **0.5374 (0.0054)** |

Table 1. Experimental results on static natural images

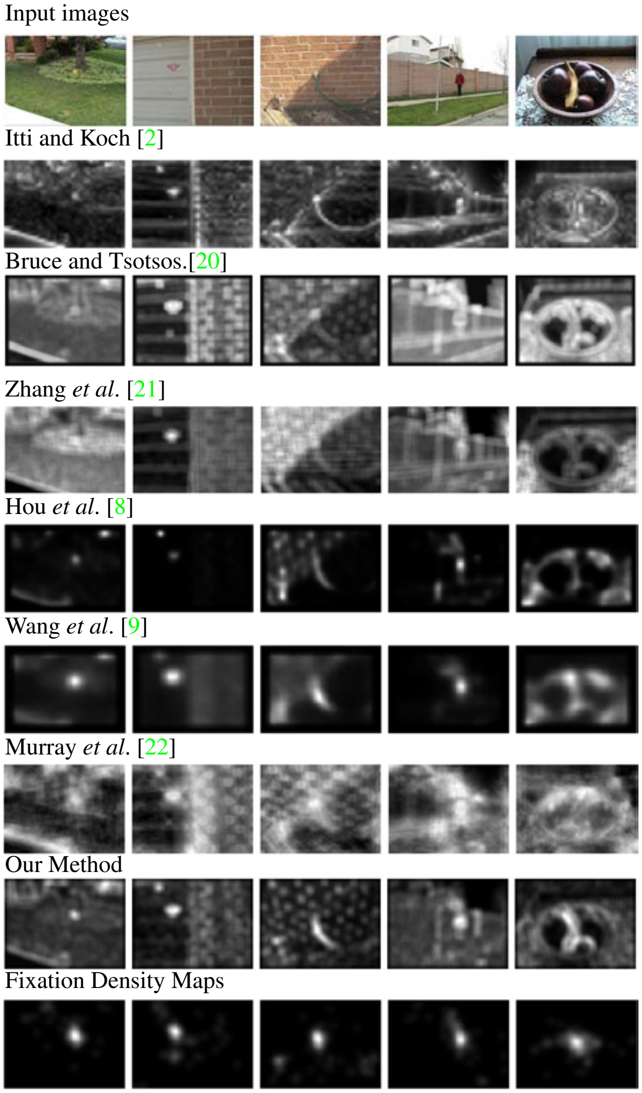


Figure 7. Visual comparisons of different saliency detection methods. From top to bottom: input images, saliency maps generated by previous models [2, 8, 9, 20, 21, 22], saliency maps generated by our model and fixation density maps generated by eye fixations.

[1] http://www-sop.inria.fr/members/Neil.Bruce

[2] http://crcns.org/data-sets/eye

### 4.2.2 Experiment on Dynamic Videos

Video processing is slightly different from static image processing: 1. the size of each image patch is [3 × 3 × 3 × 3] (Height×Width×Color×Frame) which is also reshaped into 1-D vector during the feature decomposition stage; 2. to compare saliency of the same visual content over time, we use Equation 9 to normalize the saliency maps.

$$\mathbf{S}(x) = \delta(\eta \mathbf{S}(x)/\bar{\mathbf{S}}),$$
$$\delta(x) = \begin{cases} 1 & \text{If } x > 1 \\ x & \text{else} \end{cases} . \tag{9}$$

where $\bar{\mathbf{S}}$ is the mean value of $\mathbf{S}$, $\eta = 0.3$ is a scale parameter and fixed in the following experiment. Eye tracking data from Itti *et al*. [11] are recorded from 8 human subjects aged at 23-32 with normal vision. 50 video clips consisting various categories of dynamic scenes, including outdoor scenes, television broadcast and video games, are used for constructing the data set. 7 video clips of Berkeley outdoor scene consisting of 568 saccade points are used for evaluation. The KL score produced by our model is $0.692 \pm 0.053$ which is much better compared with $0.530 \pm 0.045$ for Itti's saliency [2] and $0.589 \pm 0.045$ for surprise [11]. The AUC score of our model is $0.803 \pm 0.009$ which also outperforms the other two models ($0.775 \pm 0.011$ for Itti's saliency, $0.776 \pm 0.010$ for surprise).

## 4.3. Proto-Object Detection

A candidate that have been detected but not yet identified as an object is defined as a *proto-object*. In this experiment, we test the model's ability of detecting proto-objects in unconstrained natural scenes. The image data set, human label masks and evaluation codes used for this experiment are provided by Hou *et al*. [23]. Hit Rate (HR) and False Alarm Rate (FAR) are used for evaluating the saliency maps. Higher HR and lower FAR imply a better detection performance. Quantitative comparisons between different saliency models are shown in Table 2. We give two groups of results, one with the fixed HR and the other the fixed FAR. Our model provides an overall better performance compared to Hou *et al*. [23] and Seo *et al*. [24]. Figure 8 shows some visual examples of the detection results.

## 4.4. Robustness Test

We test our model with manually modified images which contain commonly encountered visual distortions. As demonstrated in Figure 9, our model is basically not influenced by various distortions including salt noise, gaussian noise and brightness change. For the case of contrast change, the mean value of the resulting saliency map became larger, so the target region was not very salient against the surrounding regions compared with the other cases. Practically, it is acceptable because the distortion of low contrast has similar interference to human vision system. In addition, the most salient region indicated by our saliency map remains the same despite the contrast changes.

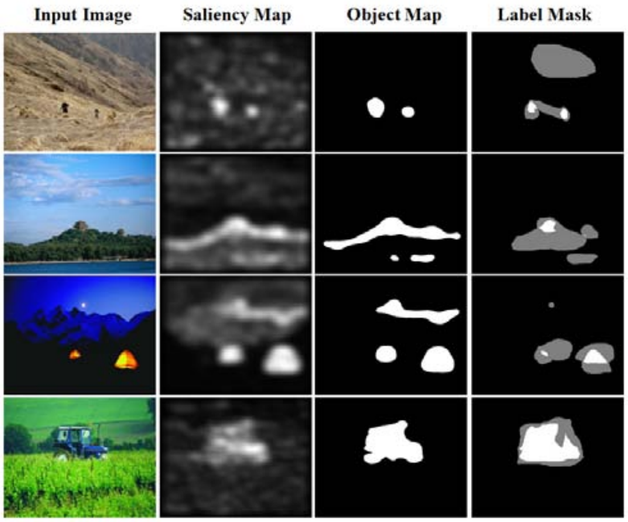


Figure 8. Examples of proto-object detection. From left to right: input image, saliency map, proto-object mask, human label mask.

Table 2. Proto-Object Detection

| | Our model | Seo *et al*. [24] | Hou *et al*. [23] |
|---|---|---|---|
| HR | 0.7227 | 0.5933 | 0.4309 |
| Fixed FAR | 0.1433 | 0.1433 | 0.1433 |
| Fixed HR | 0.5076 | 0.5076 | 0.5076 |
| FAR | 0.0816 | 0.1048 | 0.1688 |

## 5. Conclusions and Future Works

In this paper, we present an unified statistical model for saccadic eye movements and visual saliency. Different from previous works that mostly aim to reproduce the exact mechanisms of visual perception, we draw inspirations from the statistical characteristics of real-world human behavior. Experimental results demonstrate our superior performance over the state-of-the-art approaches and implies the promising potential of statistical models for human behavior analysis. In further studies, we will continue our effort to analyze human saccadic behavior considering other factors such as scale change and individual differences. Applying the framework to other computer vision problems such as anomaly detection and pattern discovery *etc*. will be another direction of our future works.

## 6. Acknowledgments

The work was supported in part by the National Science Foundation of China (No. 61071180 and No. 61133003).

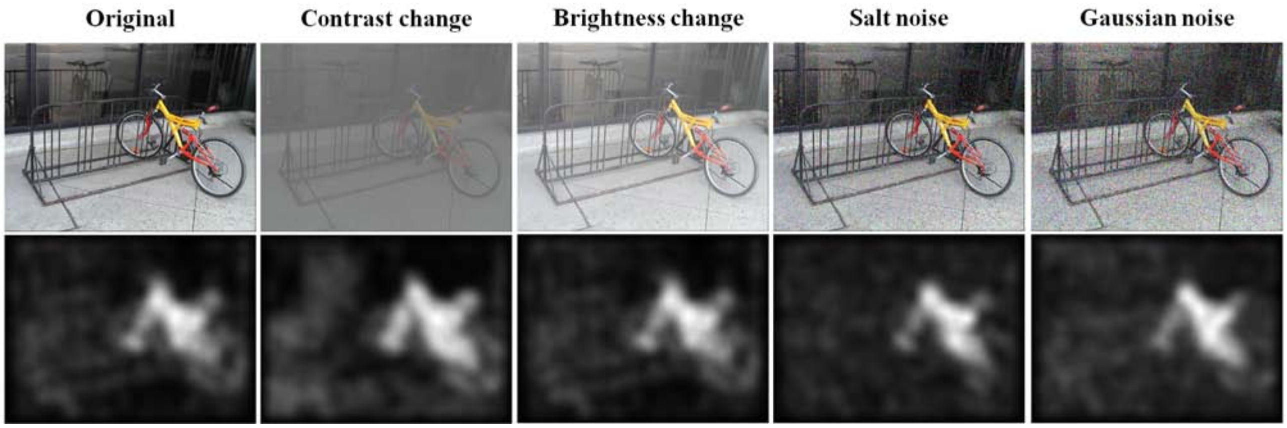


Figure 9. Our model is robust to various distortions including contrast and brightness change, salt and gaussian noise.

## References

[1] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention", *TIP*, vol. 13, no. 10, pp. 1304–1318, 2004.

[2] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis", *TPAMI*, vol. 20, no. 11, pp. 1254–1259, 1998.

[3] D. Gao, S. Han, and N. Vasconcelos, "Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition", *TPAMI*, 31(6):989–1005, 2009.

[4] Tsotsos, J.K. and Culhane, S.M. and Kei Wai, W.Y. and Lai, Y. and Davis, N. and Nuflo, F., "Modeling visual attention via selective tuning", *Artificial Intelligence*, vol. 78, no. 1, pp. 507–545, 1995.

[5] C. Koch, S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry", *Human Neurobiology*, 4 (4), pp. 219–227, 1985.

[6] L. Itti and C. Koch, "Computational modelling of visual attention", *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

[7] N. Bruce and J. Tsotsos, "Saliency based on information maximization," *NIPS*, 2006, pp. 155–162.

[8] X. Hou and L. Zhang, "Dynamic visual attention: searching for coding length increments", *NIPS*, 2008, pp. 681–688.

[9] W. Wang, Y. Wang, Q. Huang, and W. Gao, "Measuring visual saliency by site entropy rate", *CVPR*, 2010, pp. 2368–2375.

[10] W. Wang, C. Chen, Y. Wang, T. Jiang, F. Fang, and Y. Yao, "Simulating human saccadic scanpaths on natural images", *CVPR*, 2011, pp. 441–448.

[11] L. Itti and P. F. Baldi, "Bayesian surprise attracts human attention", *NIPS*, 2006, pp. 547–554.

[12] D. Gao, V. Mahadevan, and N. Vasconcelos, "The discriminant center-surround hypothesis for bottom-up saliency," *NIPS*, 2007, pp. 1–8.

[13] L. Duan, C. Wu, J. Miao, L. Qing and Y. Fu, "Visual saliency detection by spatially weighted dissimilarity", *CVPR*, 2011, pp. 441–448.

[14] L. Renninger, P. Verghese, and J. Coughlan, "Where to look next? eye movements reduce local uncertainty", *Journal of Vision*, vol. 3, no. 6, pp. 1–17, 2007.

[15] T. Foulsham and G. Underwood. "What can saliency models predict about eye movements? spatial and sequential aspects of fixations during encoding and recognition," *Journal of Vision*, 2008.

[16] Judd, T. and Ehinger, K. and Durand, F. and Torralba, A., "Learning to predict where humans look", *ICCV*, 2009, pp. 2106–2113.

[17] Treisman Anne M. and Gelade Garry, "A feature-integration theory of attention", *Cognitive Psychology*, vol. 12, no. 1, pp. 97–136, 1980.

[18] Hyvärinen, A. and Hurri, J. and Hoyer, P.O., "Natural Image Statistics: A probabilistic approach to early computational vision", in *Springer-Verlag New York Inc*. 2009

[19] C. Guo, Q. Ma, and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform", *CVPR*, 2008, pp. 1–8.

[20] N. Bruce and J. Tsotsos, "Saliency, attention, and visual search: An information thretic approach", *Journal of Vision*, vol. 9, no. 3, pp. 1–24, 2009.

[21] L. Zhang, M. Tong, T. Marks, H. Shan, and G. Cottrell, "Sun: A bayesian framework for saliency using natural statistics", *Journal of Vision*, vol. 8, no. 7, pp. 1–20, 2008.

[22] N. Murray, M. Vanrell, X. Otazu and C. Parraga, "Saiency estimation using a non-parametric low-level vision model", *CVPR*, 2011, pp. 433–440.

[23] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach", *CVPR*, 2007, pp. 1–8.

[24] H. Seo, P. Milanfar, Nonparametric bottom-up saliency detection by selfresemblance, *IEEE Computer Vision and Pattern Recognition, 1st International Workshop on Visual Scene Understanding*, 2009, 45 - 52.