What Is the Chance of Happening: A New Way to Predict Where People Look

Yezhou Yang, Mingli Song*, Na Li, Jiajun Bu, and Chun Chen

Zhejiang University, Hangzhou, China
brooksong@zju.edu.cn

Abstract. Visual attention is an important issue in image and video analysis and keeps being an open problem in the computer vision field. Motivated by the famous Helmholtz principle, a new approach of visual attention analysis is proposed in this paper based on the low level feature statistics of natural images and the Bayesian framework. Firstly, two priors, i.e., Surrounding Feature Prior (*SFP*) and Single Feature Probability Distribution (*SFPD*) are learned and integrated by a Bayesian framework to compute the chance of happening (*CoH*) of each pixel in an image. Then another prior, i.e., Center Bias Prior (*CBP*), is learned and applied to the *CoH* to compute the saliency map of the image. The experimental results demonstrate that the proposed approach is both effective and efficient by providing more accurate and quick visual attention location. We make three major contributions in this paper: (1) A set of simple but powerful priors, *SFP*, *SFPD* and *CBP*, are presented in an intuitive way; (2) A computational model of *CoH* based on Bayesian framework is given to integrate *SFP* and *SFPD* together; (3) A computationally plausible way to obtain the saliency map of natural images based on *CoH* and *CBP*.

1 Introduction

The surrounding world contains a tremendous amount of visual information which the human visual system (*HVS*) cannot fully process [1]. Therefore, human tends to pay attention to only a few parts while neglect others in front of a scene. This phenomenon is usually called visual attention by psychologists. In order to predict where people look in an image automatically, visual attention analysis has been investigated for dozens of years in computer vision field. But till now it is still an open problem to be tackled. Recently, understanding computer vision problems from the view of psychologist is becoming an important track. As visual attention is also an important issue and has been studied for more than a century in the psychology field, it is reasonable to adopt some useful concepts of psychology to solve the visual attention analysis problem.

By treating the visual attention analysis as a signal processing problem, researchers believe *HVS* functions like a filter. It is natural for them to simulate the psychological mechanism by a filter computationally [2,3]. However, it is

* Corresponding author.

K. Daniilidis, P. Maragos, N. Paragios (Eds.): ECCV 2010, Part V, LNCS 6315, pp. 631–643, 2010.

difficult to strictly simulate the visual attention mechanism in practice because the operating detail of *HVS* is still unknown until now even in the psychology field. Therefore, although some existing approaches [4,5,6] try to build up elegant mapping from psychological theories to computational implementation, they do not match the actual human saccade from the eye-tracking data and are computational consuming in practice.

Another way of approach tries to learn a visual attention model by taking into account of both the low level and the high level features based on the human eye tracking data, In [7], the pixel intensity is extracted as low level feature. And the semantic information is used as the high level features obtained by carrying out face detection, car detection, etc.. However, the study of *HVS* in [8] shows that the visual attention scheme is a kind of pre-processing before semantic analysis. Moreover, it's difficult to always extract the semantic information successfully from the given image.

Different from the aforementioned approaches, Bruce et al. [9,10] try to figure out the correlation between an information theory based definition of visual saliency and the fixation behavior of *HVS*. Then, in order to realize Bruce's theory in a computationally plausible way, some researchers [11,12,13] argue that some natural statistics of visual features, e.g., blue sky seldom exists in the lower part of a scene, should play an important role in the visual attention process. Some methods, such as Independent Component Analysis (ICA), are employed to extract the visual features to achieve the natural statistics. However, such visual features are complex and usually fail to estimate the visual attention efficiently. Moreover, given an image, *HVS* usually tends to pay attention to the central region of the image [14]. This center bias phenomenon of *HVS* is ignored in these methods, which often leads to mismatch between the experimental results and the eye tracking ground truth.

Inspired by the famous Helmholtz principle [15], a new concept called Chance of Happening (*CoH*) is introduced in this paper. As described literally, *CoH* of a point represents how likely this point will exist at a specific location of an image. And *CoH* is believed to be largely determined by a couple of priors, e.g., the relationship between the features of a specific point and its surroundings. These priors are learned from daily life experience by the *HVS*. In our approach, we try to learn these priors in a computationally plausible way and use them to estimate the *CoH* of a point. Both the *CoH* and the center bias are taken into account to compute the saliency map of the image.

The rest of the paper is organized as follows: section 2 introduces the motivation of the proposed approach. Section 3 describes two low level feature priors: Surrounding Feature Prior (*SFP*) and Single Feature Probability Distribution (*SFPD*). A Bayesian framework is introduced to compute the *CoH* from these two priors. In section 4, Center Bias Prior (*CBP*) is proposed and in section 5, a probabilistic framework is presented to compute the saliency map of an image by integrating the *CoH* and *CBP* together. Experiment is carried out in section 6. And we conclude in section 7.

(a) (b) (c)

Fig. 1. Example of relationship between Chance of Happening (CoH) and visual attention. (a) Scene with hidden regions; (b) Original scene; (c) Saliency map computed directly from eye-tracking dataset.

2 Motivation

"***Whenever some large deviation from randomness occurs, a structure is perceived.***" [15]

–H. von Helmholtz (1821-1894), Psychologist, Germany

H. von Helmholtz, the famous Gestalt psychologist in Germany in 19th century, gave his famous description of human visual perception above, called Helmholtz principle. As a commonsense statement, the Helmholtz principle means that "we immediately perceive whatever could not happen by chance" [15]. This description inspires us with two cues: given an image, first, human's visual attention depends heavily on the chance of happening (CoH) which is estimated based on the previous experience of human visual system (HVS); second, only the region which "could not happen by chance" tends to be "immediately perceived" by HVS, where "could not happen by chance" means the CoH of the region is small. Figure 1 demonstrates a typical instance of Helmholz principle.

In Figure 1, for each blacked out region (left) labeled A, B or C, HVS estimates its CoH based on previous experience. For example, given the surrounding context of blue sky, the existence of dark red spot of region C is a large deviation from our expectation of color intensity distribution according to our previous experience. This large deviation leads region C to a small CoH. In contrast, the CoH of A and B are large because they preserve the consistency to their surroundings as predicted by HVS according to our previous experience. Recalling "we immediately perceive whatever could not happen by chance", C will be "perceived immediately" as its CoH is small. The eye tracking experimental result (Figure 1(c)) also justifies this principle.

Besides CoH, researchers also find that HVS usually tends to pay attention to the center of an image [14,16]. This phenomenon is called center bias. In our approach, both CoH and the center bias are taken into account to carry out the visual attention analysis by computing the saliency map of an image.

Based on the discussion above, three priors, two for *CoH* and one for the center bias, are presented in our approach. And then they are integrated into a computational framework to perform the visual attention analysis.

3 From Low Level Feature Priors to *CoH*

As aforementioned, the visual attention scheme is a kind of pre-processing before the semantic analysis. And the semantic information is high-level which is difficult to extract robustly. In our approach, only the low level feature priors are taken into account to compute *CoH* towards the visual attention analysis.

3.1 Single Feature Probability Distribution

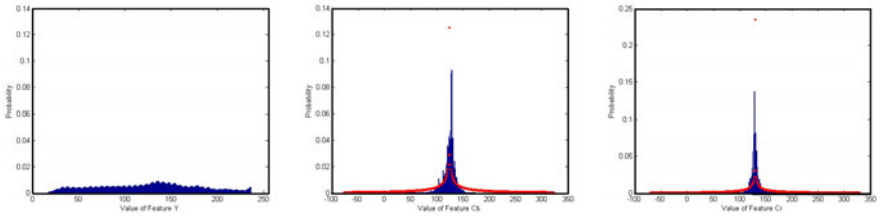
Color intensity is a kind of primary low level features used in computer vision society. In addition, color intensity can be manipulated more efficiently than other features. $YCbCr$ is a preferable representation of the natural images in digital image coding system wherein Y, Cb and Cr are highly uncorrelated components corresponding to luminance, blue difference, and red difference [17]. So it is naturally to adopt $YCrCb$ to learn the low level feature prior for the previous experience of *HVS* on color intensity distribution in natural images. Such low level feature prior is called Single Feature Probability Distribution (*SFPD*) in our approach.

In order to learn *SFPD*, we accumulate the occurrences of different intensities in the collected natural images for Y, Cb and Cr respectively. Thus the probability distributions of Y, Cb and Cr over the observed values is obtained correspondingly. Observing the curves in Fig. 2, we notice that the distribution of Y tends to be uniform, which means it doesn't take function in the *SFPD* computation. (Fig.2(a)). In contrast to Y, the statistics of Cb (Fig. 2(b)) and Cr (Fig. 2(c)) look meaningful and Gaussian-like. Inspired by such observation, both the marginal distributions of Cb and Cr can be formulated as generalized Gaussian densities respectively:

$$P\left(Cb(x,y);\sigma_b,\theta_b\right) = \frac{\theta_b}{2\sigma_b\tau(\frac{1}{\theta_b})}\exp\left(-|\frac{Cb(x,y)}{\sigma_b}|^{\theta_b}\right) \tag{1}$$

$$P\left(Cr(x,y);\sigma_r,\theta_r\right) = \frac{\theta_r}{2\sigma_r\tau(\frac{1}{\theta_r})}\exp\left(-|\frac{Cr(x,y)}{\sigma_r}|^{\theta_r}\right) \tag{2}$$

where τ is the gamma function, $\sigma_{(\cdot)}$ is the scale parameter that describes the standard deviation of the density, $\theta_{(\cdot)}$ is the shape parameter that is inversely proportional to the decreasing rate of the peak. $Cb(x,y)$ and $Cr(x,y)$ are the color intensity values of the pixel (x,y). The model parameter ($\theta_{(\cdot)}$, $\sigma_{(\cdot)}$) can be estimated using the moment matching method [18] or the maximum likelihood rule [19]. In our approach, the estimated parameters are: $\sigma_b = 0.23$, $\sigma_r = 0.041$, $\theta_b = 0.26$, $\theta_r = 0.22$. The estimated distributions are depicted in Figure 2.



(a) (b) (c)

Fig. 2. (a)Single Feature Probability Distribution of Y (b)Single Feature Distribution of Cb (c)Single Feature Distribution of Cr. The blue bars represent the origin distribution and the red dots plot the estimated generalized Gaussian distributions

3.2 Surrounding Feature Prior

The surrounding context of a pixel is another important low level feature that reflects the contrast in an image. In our approach, the previous experience on the surrounding context in natural images is called Surrounding Feature Prior (*SFP*).

Given a surrounding window sized by $w * w$, two distances, i.e., the intensity distance and the location distance, are defined between pixel (x, y) and one of its surrounding pixels $(x_j, y_j), j \in [1, w * w]$. The definition of intensity distance in our approach is given below:

$$D_f^j(x,y) = |Y(x,y) - Y(x_j,y_j)| + |Cb(x,y) - Cb(x_j,y_j)| + |Cr(x,y) - Cr(x_j,y_j)| \tag{3}$$

where $|\cdot|$ is the absolute operation.

And the location distance is defined as follows:

$$D_l^j(x,y) = \max\left(|x - x_j|, |y - y_j|\right) \tag{4}$$

where $\max(\cdot,\cdot)$ returns the largest value of the two inputs.

In order to learn the *SFP*, we count the number of pixels for each D_l varies from 1 to $(w-1)/2$ in the range of D_f based on the collected natural images. By observing Fig. 3, we notice that the probability distributions of intensity distance for different location distances are different from each other, but they are all exponential-like. So we use exponential function to fit these distributions using maximum likelihood rule. Given a pixel (x, y), the existence probability of a pixel (x_j, y_j) in the surrounding window can be formulated as follows: :

$$P\left((x_j,y_j)|(x,y)\right) = \exp\left(-\psi\left(D_l^j(x,y)\right) D_f^j(x,y)\right) \tag{5}$$

where $\psi(\cdot)$ is a enumerate function produces the unique coefficient for each location distance $D_l^j(x,y)$ by exponential regression. In our approach, we set the surrounding window as $81 * 81$, and the estimated parameter $\psi(\cdot)$ varies from 0.955 to 0.472 while the D_l increases from 1 to 40. The original and estimated distributions are depicted in Figure 3.

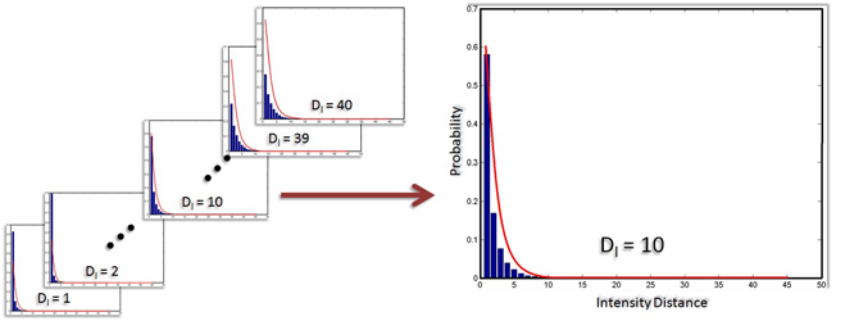


Fig. 3. Surrounding Feature Prior. The blue bars represent the origin distribution and the red lines are estimated exponential probability distribution.

3.3 Estimation of *CoH*

CoH of pixel (x, y) depends on not only the previous experience of color intensity distribution but also that of the surrounding context, i.e., *SFPD* and *SFP*. Thus *CoH* can be represented in a probability form and deduced based on Bayesian theorem as follows:

$$\begin{aligned} CoH(x,y) &= P(h(x,y)|\Omega(x,y)) \\ &= P(\Omega(x,y), h(x,y))/P(\Omega(x,y)) \end{aligned} \tag{6}$$

where $h(x, y)$ in the first line means the happening of pixel (x, y). $\Omega(x, y)$ represents the pixel (x, y) and its surrounding window.

Suppose the $w * w$ surrounding pixels are independent of each other. Since $P(\Omega(x, y), h(x, y))$ is the joint probability representing the co-occurrence of the pixel (x, y) and its surrounding context, it can be computed as:

$$\begin{aligned} P(\Omega(x,y), h(x,y)) &= \prod_{j=1}^{w*w} \left(P\left((x_j, y_j), (x, y)\right)\right) \\ &= \prod_{j=1}^{w*w} \left(P\left((x_j, y_j)|(x, y)\right) P(x, y)\right) \\ &= \prod_{j=1}^{w*w} \left(P\left((x_j, y_j)|(x, y)\right) P(Cb(x, y)) P(Cr(x, y))\right) \end{aligned} \tag{7}$$

And $P(\Omega(x, y))$ is the probability representing the occurrence of the surrounding context alone. So it can be computed as:

$$P(\Omega(x,y)) = \prod_{j=1}^{w*w} \left(P(x_j, y_j)\right) = \prod_{j=1}^{w*w} \left(P\left(Cb(x_j, y_j)\right) P\left(Cr(x_j, y_j)\right)\right) \tag{8}$$

Substituting Eq.(7) and (8) into (6), we have

$$\begin{aligned} CoH(x,y) &= P(h(x,y)|\Omega(x,y)) \\ &= \frac{\prod_{j=1}^{w*w}\left(P\left((x_j,y_j)|(x,y)\right)P(Cb(x,y))P(Cr(x,y))\right)}{\prod_{j=1}^{w*w}\left(P(Cb(x_j,y_j))P(Cr(x_j,y_j))\right)} \end{aligned} \tag{9}$$

4 Center Bias Prior

Besides the *CoH*, center bias is another important factor affects visual attention. Liu et al. [16] apply it by setting weight arbitrarily to compute the saliency map. However, it is still unclear that the decay rate of visual attention corresponding to the distance from the center. In order to obtain the decay rate away from the center, we learn a normal Bivariate Gaussian function from eye tracking dataset [10] to model the center bias in this paper. In the dataset, each image is accompanied with eye tracking data which are collected from 20 subjects free-viewing the image. Only 4 seconds are recorded during the free-viewing process by discarding the first several seconds which may introduce the imposed centering operation of the head-mounted eye tracking system. By accumulating all the fixation locations of human eyes in the eye tracking data, we depict the distribution of the fixation locations in Fig. 4. Observing the 3D depiction of the figure, a normal Bivariate Gaussian is defined to fit the eye fixation distribution as follows:

$$CBP(x,y) = \eta \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)}{\sigma_2^2}\right)\right) \tag{10}$$

where $\eta = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}}$. And we use $\rho = 0$ for simplicity. Specifically, the image size used for *CBP* learning is 511×681. And the learned parameters of *CBP* are $\mu_1 = 245.1$, $\mu_2 = 343.6$, $\sigma_1 = 91.2$, $\sigma_2 = 139.5$.

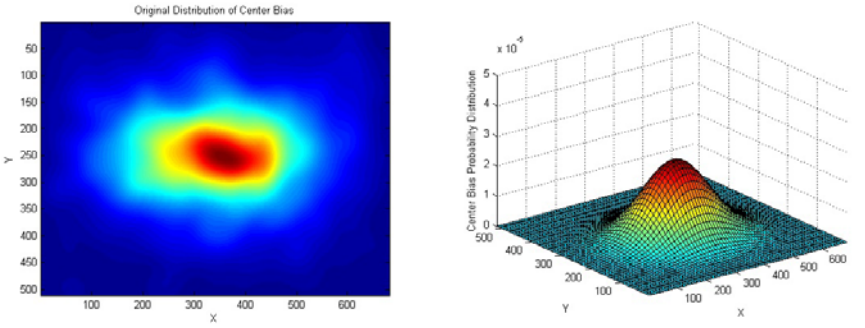
5 Computational Framework for Saliency

As aforementioned, the visual attention of human is influenced by not only the *CoH*, but also the center bias. So in our approach, both the *CoH* and the center bias are taken into account to compute the saliency map of the given image.

As discussed in Section 2, a point will be more attractive to the *HVS* when its *CoH* is relatively smaller. So in our approach, a pixel's saliency is treated inversely proportion to *CoH*. Thus, the saliency $S(x,y)$ of pixel (x,y) can be computed as follows:

$$S(x,y) = \lambda\frac{CBP(x,y)}{CoH(x,y)} \tag{11}$$

where λ is a scalar to perform adjustment.



(a) (b)

Fig. 4. Center Bias Prior: (a)Origin Center Bias Distribution (b)Estimated Center Bias Distribution

By substituting eq. (9), (11) can be rewritten as:

$$S(x,y) = \lambda \frac{\prod_{j=1}^{w*w} \left(P(Cb(x_j,y_j))P(Cr(x_j,y_j))\right) \cdot CBP(x,y)}{\prod_{j=1}^{w*w} \left(P\left((x_j,y_j)|(x,y)\right)P(Cb(x,y))P(Cr(x,y))\right)}. \tag{12}$$

Since logarithm is a monotonically increasing function, we rewrite Eq. (12) to reduce the numerical computational complexity as follows:

$$\begin{aligned} S(x,y) \ &\rightarrow -\log\left(\frac{\prod_{j=1}^{w*w} \left(P\left((x_j,y_j)|(x,y)\right)P(Cb(x,y))P(Cr(x,y))\right)}{\prod_{j=1}^{w*w} \left(P(Cb(x_j,y_j))P(Cr(x_j,y_j))\right)}\right) + \log(CBP(x,y)) \\ &\rightarrow \begin{bmatrix} \sum_{j=1}^{w*w} \left(\psi(D_l^j(x,y))D_f^j(x,y) - |\frac{Cb(x_j,y_j)}{\sigma_b}|^{\theta_b} - |\frac{Cr(x_j,y_j)}{\sigma_r}|^{\theta_r}\right) \\ +w*w*\left(|\frac{Cb(x,y)}{\sigma_b}|^{\theta_b} + |\frac{Cr(x,y)}{\sigma_r}|^{\theta_r}\right) \\ -\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2)}{\sigma_2^2}\right) \end{bmatrix} \end{aligned} \tag{13}$$

where "→" means "depend on".

Fig. 5 depicts the flowchart of saliency map computation. For each input image, the RGB value of each pixel is transformed into YCbCr firstly. Then we calculate the saliency of each pixel by a computational framework in terms of Eq. (13). Thirdly, a 25×25 Gaussian Blur is applied to remove the unwanted noisy saliency value. And histogram equalization is performed for better visualization.

6 Experiments

In our implementation, more than 1000 natural images are collected from internet so as to learn the two low level feature priors for CoH. And another prior CBP is learned based on one part of the eye tracking dataset [10] which consists

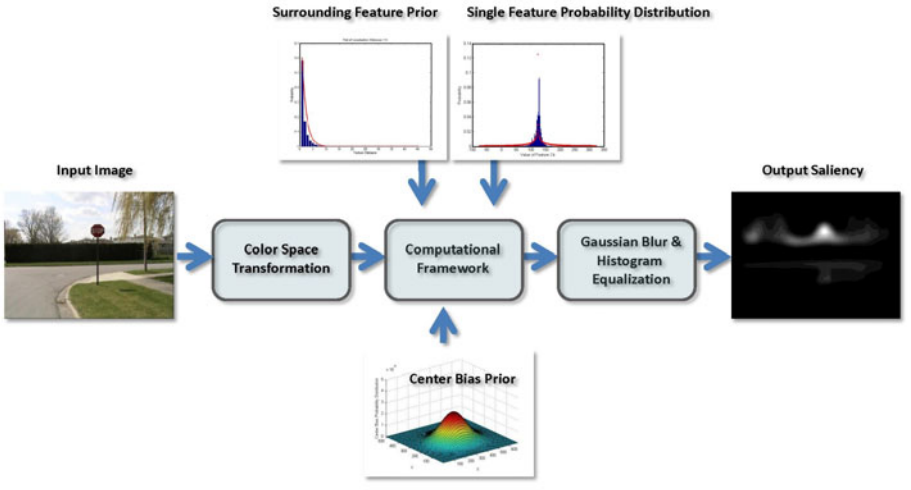


Fig. 5. The computational framework of our approach

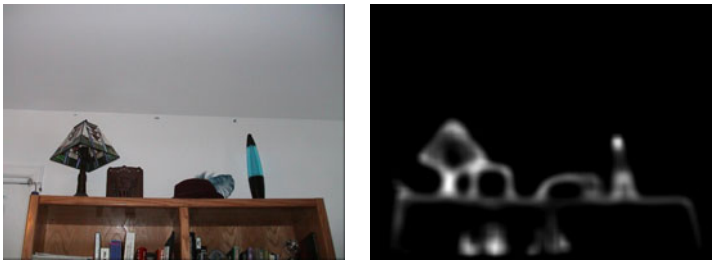
80 records. The other part of the eye tracking dataset is used as testing data for evaluation.

The computational complexity of our method is $O(w^2 \times N)$ which is lower than most of the conventional approaches, e.g., Itti's method. N is the number of pixels in the image. In addition, both look-up table and Integral Image [20] techniques are adopted by our approach to further speed up the computation of saliency map significantly. The experiment is carried out on a system with 2.8GHz processor and 8 Gigabyte memory. It takes our approach less than 1 minute on average for 511×681 images with 81×81 surrounding window. Under the same condition, it usually takes Itti's method 10 minutes to deduce the saliency map.

It is noticeable that different sizes of the surrounding window lead to different results (Fig. 6). When the surrounding window is too small, the computed saliency map usually stays away from the ground truth (Fig. 6(b)). On the contrary, the computed saliency map will be closer to the ground truth when the size of surrounding window becomes larger (Fig. 6(d)). But it will take more time to produce the result as a tradeoff. As a compromise, we choose 81×81 for 511×681 images in our approach (Fig. 6(c)).

6.1 Qualitative Evaluation

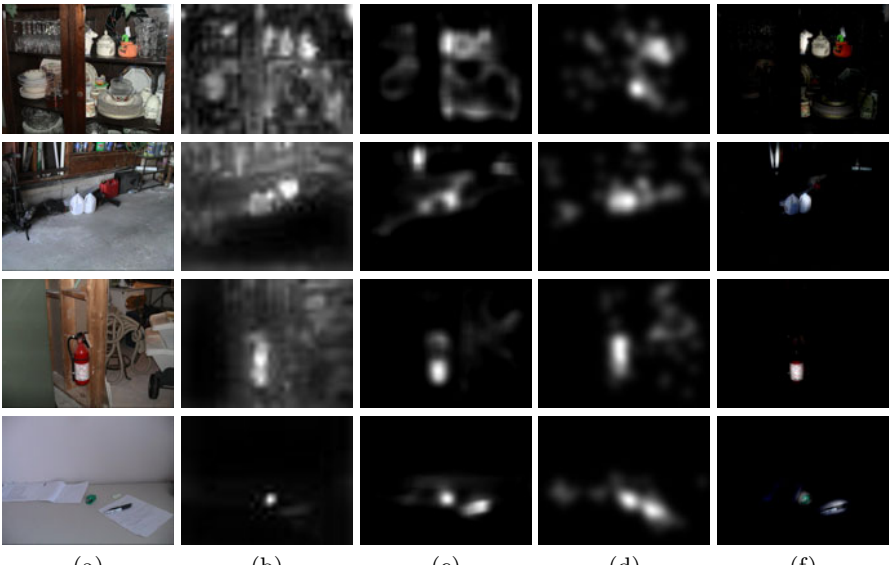
The saliency maps are depicted in Fig. 7 and 8 for the outdoor and the indoor images separately. The ground truths for evaluation are obtained from the eye tracking dataset. And the testing image are modulated by the output saliency of our method for better qualitative evaluation. It is noticeable that our method



(a) Input Image (b) 41*41 surrounding

(c) 81*81 surrounding (d) 161*161 surrounding

Fig. 6. Comparison of saliency maps computed by various surrounding sizes



(a) (b) (c) (d) (f)

Fig. 7. Saliency comparison on indoor images (a) Original image (b)Itti et al. saliency (c) Saliency of our method (d) Experimental saliency map (f) Modulated image

performs better than Itti's method [4] on both the indoor and the outdoor images. Especially for the outdoor images, Itti's method does not match the ground truth very well, e.g., the white car in the first row.

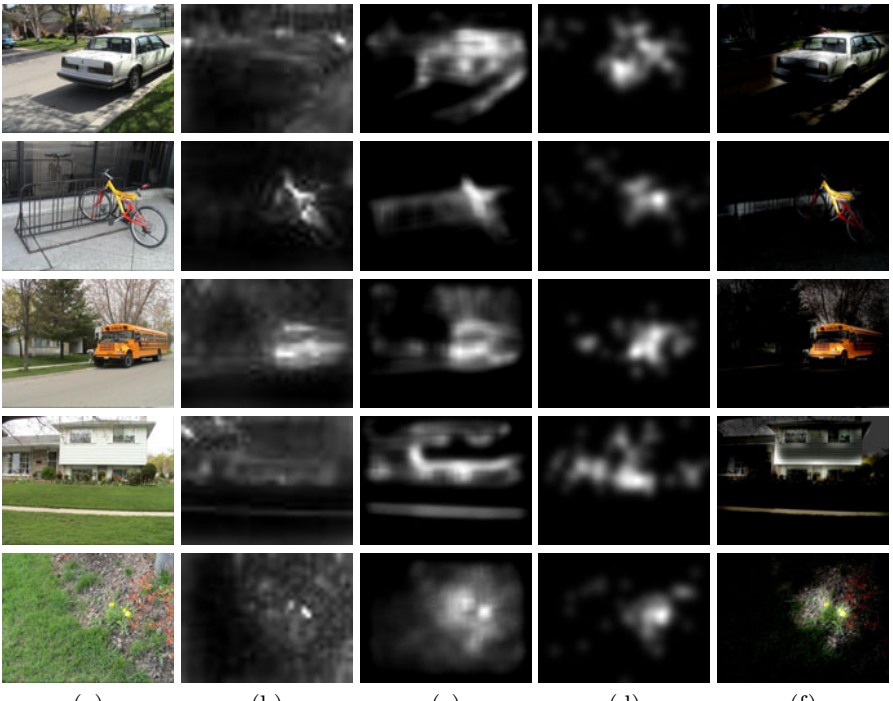


Fig. 8. Saliency comparison on outdoor images (a) Original image (b)Itti et al. saliency (c) Saliency of our method (d) Experimental saliency map (f) Modulated image

6.2 Quantitative Evaluation

Recently, the receiver operating characteristic curve (*ROC*)[7,10] becomes a widely adopted metric to evaluate eye fixation prediction quantitatively. The *ROC* metric treats the saliency map as a binary classifier in the image, wherein pixels with saliency value larger than a threshold are identified as "fixated" while the others "non-fixated". By varying the threshold and testing the true positive rate for each threshold, an *ROC* curve can be drawn and the area under the curve indicates how well the saliency map predicts eye's fixations.

Figure 9 shows the comparison between the proposed method and the conventional ones. In our evaluation, the threshold of visual attention region varies from top 5% to top 30% part of the saliency map. The *ROC* curves in the figure depict quantitatively that the proposed method takes advantage over the conventional ones.

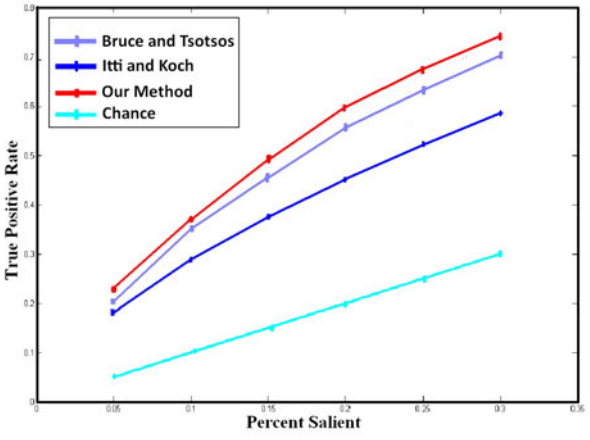


Fig. 9. The ROC curves of performances for Itti [4], Bruce/Tsotsos [13] and our method. We also plot chance for comparison.

7 Conclusions

In this paper, a new computational approach is presented for visual attention analysis. Two low level feature priors, *SFPD* and *SFP*, are learned based on natural images and integrated by a Bayesian framework to compute the *CoH* at each pixel. Then another prior, *CBP*, is learned from the eye tracking dataset. Finally, the saliency of each pixel is obtained by taking consideration of both *CoH* and *CBP*. By using the proposed approach, visual attention analysis becomes more effective and more efficient than the existing low level feature based approaches and produces better matching to human eye-tracking data.

We also notice some limitations of our method. For example, the size of the surrounding window is set arbitrarily. In the future, we will develop a multi-scale approach to figure out the optimal size of the surrounding window. In addition, more low level feature priors for *CoH*, especially those biologically plausible priors, will be investigated. It may also be possible to employ a GPU-based implementation to parallelize the computation. However, these points do not impact on the conclusions of this paper or the theory presented.

Acknowledgment

This paper is supported by the National Natural Science Foundation of China under Grant 60873124, by the Natural Science Foundation of Zhejiang Province under Grant Y1090516, and by the Fundamental Research Funds for the Central Universities under Grant 2009QNA5015.

References

1. Tsotsos, J.: Analyzing vision at the complexity level. Behavioral and Brain Sciences 13, 423–445 (1990)
2. Wolfe, J.M., Cave, K.: Deploying visual attention: The guided search model. John Wiley and Sons Ltd., Chichester (1990)
3. Tsotsos, J.K., Culhane, S.M., Wai, W.: Modeling visual attention via selective tuning. Artificial Intelligence 78, 507–545 (1995)
4. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence (1998)
5. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research 40, 1489–1506 (2000)
6. Yang, Y., Song, M., Li, N., Bu, J., Chen, C.: Visual attention analysis by pseudo gravitational field. ACM Multimedia, 553–556 (2009)
7. Judd, T., Ehinger, K., Durand, F., Torralba, A.: Learning to predict where humans look. In: International Conference on Computer Vision, Acceptance (2009)
8. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiology 4, 219–227 (1985)
9. Bruce, N.D.B.: Features that draw visual attention: An information theoretic perspective. Neurocomputing 65-66, 125–133 (2005)
10. Bruce, N.D.B., Tsotsos, J.K.: Saliency based on information maximization. Advances in Neural Information Processing Systems 18, 155–162 (2006)
11. Itti, L., Baldi, P.: Bayesian surprise attracts human attention. Advances in neural information processing systems 19, 547–554 (2006)
12. Zhang, L., Tong, M.H., Marks, T.K., Shan, H., Cottrell, G.W.: Sun: A bayesian framework for saliency using natural statistics. Journal of Vision 8, 1–20 (2008)
13. Bruce, N.D.B., Tsotsos, J.K.: Saliency, attention, and visual search: An information theoretic approach. Journal of Vision 9(3), 1–24
14. Meur, O.L., Callet, P.L., Barba, D., Thoreau, D.: A coherent computational approach to model bottom-up visual attention. IEEE Trans on Pattern Analysis and Machine Intelligence 28, 802–816 (2006)
15. Desolneux, A., Moisan, L., Morel, J.: From Gestalt Theory to Image Analysis, A Probabilistic Approach. Springer Science and Business Media, Heidelberg (2008)
16. Liu, H., Jiang, S., Huang, Q., Xu, C., Gao, W.: Region-based visual attention analysis with its application in image browsing on small displays. ACM Multimedia (2007)
17. Zhai, G., Chen, Q., Yang, X., Zhang, W.: Scalable visual sensitivity profile estimation. In: ICASSP, pp. 876–879 (2008)
18. de Wouwer, G.V., Scheunders, P., Dyck, D.V.: Statistical texture characterization from discrete wavelet representations. IEE Trans. Image Processing 8, 592–598 (1999)
19. Poor, H.V.: An Introduction to Signal Estimation and Detection, 2nd edn. Springer, Heidelberg (1994)
20. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 511–518 (2001)