Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10)

A Computational Model for Saliency Maps by Using Local Entropy

Yuewei Lin
College of Computer Science
Chongqing University, China
ywlin.cq@gmail.com

Bin Fang
College of Computer Science
Chongqing University, China
fb@cqu.edu.cn

Yuanyan Tang
College of Computer Science
Chongqing University, China
yytang@cqu.edu.cn

Abstract

This paper presents a computational framework for saliency maps. It employs the Earth Mover's Distance based on weighted-Histogram (EMD-wH) to measure the center-surround difference, instead of the Difference-of-Gaussian (DoG) filter used by traditional models. In addition, the model employs not only the traditional features such as colors, intensity and orientation but also the local entropy which expresses the local complexity. The major advantage of combining the local entropy map is that it can detect the salient regions which are not complex regions. Also, it uses a general framework to integrate the feature dimensions instead of summing the features directly. This model considers both local and global salient information, in contrast to the existing models that consider only one or the other. Furthermore, the "large scale bias" and "central bias" hypotheses are used in this model to select the fixation locations in the saliency map of different scales. The performance of this model is assessed by comparing their saliency maps and human fixation density. The results from this model are finally compared to those from other bottom-up models for reference.

Introduction

The visual environment is an enormously rich source of information, and the viewer must select the information that is most relevant at any point in time. The process of selecting which of the information that is entering our eyes receives further processing therefore plays a central role in sensation, serving as the gatekeeper that controls access to our highly evolved visual information processing system (Itti & Koch 2001; Liversedge & Findlay 2000).

Some stimuli are intrinsically conspicuous or salient in a given context. For example, a red dot in a field of green dots, automatically and involuntarily attracts attention. Saliency, which refers to the bottom–up attraction of attention arising from the contrast between the feature properties of an item and its surrounds, is independent of the nature of the particular task, and is primarily driven in a bottom-up manner, although it can be influenced by contextual, figure–ground effects. This suggests that saliency is computed in a pre-attentive manner across the entire visual field, most probably in terms of hierarchical centre-surround mechanisms in the early stages of biological vision (Itti & Koch 2001). Bottom-up saliency models specify early visual filters that quantify the visual conspicuity of each part of the scene in terms of the contrast between a feature and surrounds. These models suggest that low-level feature discontinuities represented in the saliency map can explain a significant part of where people look.

There is a large body of literatures related to biological and theoretical models of attention. The model proposed by Itti et al. (Itti et al. 1998) (to which we will henceforth refer as the "Itti model") is probably the most widely used computational model but it is not uncontroversial. For example, Draper and Lionelle (Draper & Lionelle 2005) show that the Itti model is not scale or rotation invariant, thus questioning the appropriateness of using the Itti model as the basis of computational object recognition systems.

Overview and Contributions of the Proposed Model

In this paper, a computational model is described, which is shown in Figure 1.

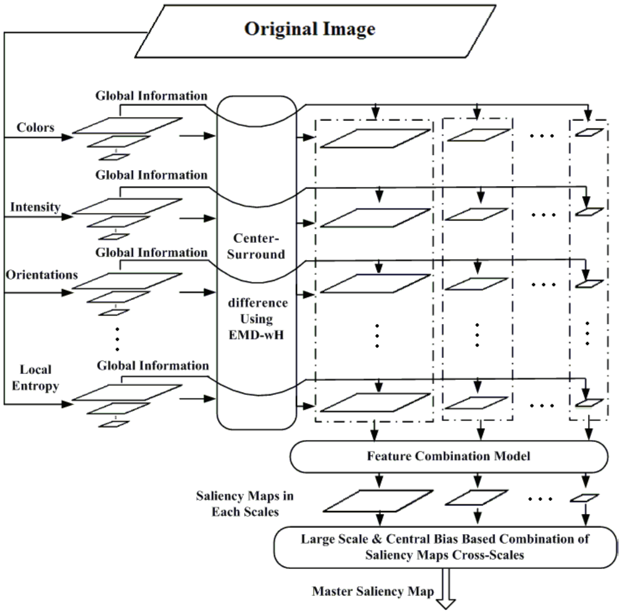


Figure 1: The framework of the proposed model

The major contribution of this model is five fold: first, we introduce the EMD based on weighted-Histogram

(EMD-wH) to measure the center-surround difference; second, we consider both local and global salient information, and combine them together to produce the saliency maps; third, we use the local entropy as a feature in the model; fourth, we propose a general framework to integrate the saliency maps of each feature dimension; fifth, we use the "large scale bias" and "central bias" hypotheses when we get the master saliency map.

Distance Metric between Center and Surround

The Difference-of-Gaussian (DoG) filter is usually used to compute the center-surround difference. It tends to assign high saliency values to highly textured regions and it is sensitive to small changes since it is implemented as a pixel-by-pixel difference.

Gao and Vasconcelos define bottom-up saliency of a point as the KL divergence between the histogram of the center region and the histogram of the surrounding region (Gao & Vasconcelos 2007). However, as a bin-by-bin dissimilarity measurement, the KL divergence has the major drawback that it accounts only for the correspondence between bins with the same index, and does not use information across bins. To avoid these disadvantages, we employ the EMD to compute the center-surround difference.

EMD Based on Weighted Histogram

The Earth Mover's Distance (EMD) was first introduced and used in some color and texture signature applications by Rubner et al. (Rubner et al. 2000).

However, as a histogram-based method, the EMD may lose the spatial information encoded in the pixels. For example, as illustrated in Figure 2, the two center-surrounding regions have the same EMD if we compute them directly by the traditional histogram. However, we can easily see that the left image is more salient than the right one, this is because the contrast of the regions near the border between the center and the surround in the left is bigger than that in the right one. Therefore, the EMD based on weighted histogram (EMD_wH) is used in our model to reflect this difference.

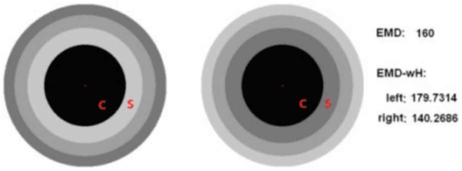


Figure 2: Illumination of the effect of EMD_wH

The normalized weights of the pixels in the center region and surrounding region are defined as follows:

$$\mathrm{weight}_i = \frac{r_i}{\sum_i r_i} \quad i \in \text{centre region} \qquad (1)$$

$$\mathrm{weight}_j = \frac{d_j}{\sum_j d_j} \quad j \in \text{surrounding region} \qquad (2)$$

where r_i denotes the distance from pixel i to the center of the center region, and d_j denotes the distance from pixel j to the outer border of the surrounding region. The histograms of the center and the surrounding region are computed with these weights.

An Efficient Method to Compute EMD_wH

The EMD equals the Wasserstein (Mallows) distance when the two distributions have equal masses (Levina & Bickel 2001). When we consider the two histograms, this requirement is exactly satisfied since the total mass of a histogram is always equal to 1. Therefore, when the ground distance in the EMD is defined as the L1 norm between bins, the EMD between these two histograms exactly equals the Wasserstein distance W_1. Then we produce an efficient way to compute the Wasserstein distance.

Bickel and colleague demonstrated the following equation (Bickel & Freedman 1981):

$$W_1(P,Q) = \int_{-\infty}^{\infty} |F(t) - G(t)|dt \qquad (3)$$

where the $W_1(P,Q)$ denotes the Wasserstein distance between P and Q with exponent 1, F and G are the cumulative distribution functions of P and Q, respectively, and F^{-1} and G^{-1} represent their respective inverse functions. Thus, we can compute the EMD very efficiently. Let the two weighted histograms (256 bins) wH_c and wH_s represent the center and the surrounding regions; then the EMD between them is computed as follows:

$$EMD(wH_c, wH_s) = \sum_{i=0}^{255} \left| \sum_{j=0}^{i} wH_c(j) - \sum_{j=0}^{i} wH_s(j) \right| \quad (4)$$

Local and Global Saliency

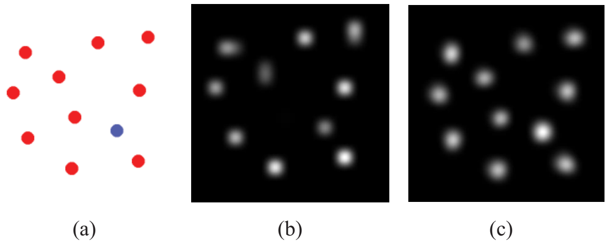


Figure 3: Illustration of the effect of global information

Although the saliency value at each location is essentially the local contrast in most traditional computation models, we cannot ignore the global information which represents the context of the whole scene. For example, in Figure 3 (a), the light blue dish should be more salient than red ones since there is only one of it. The light blue dish, however, appears less salient than the red ones if we only consider the local information, as shown in Figure 3 (b).

Several models employ global information to detect the salient locations in a scene. For instance, the Bayesian modeling was proposed to detect the locations of visual attention (Torralba et al. 2006). In this statistical framework, the location of attention is found by computing the posterior probability $\mathrm{P(L=A|f,I)}$, which is the probability of a certain location L being attended (A) in image I given a set of visual features f . By using Bayes' rule, we have:

$$\mathrm{P(L=A|f,I)} = \frac{1}{\mathrm{P(f|I)}} \mathrm{P(f|L=A,I)P(L=A|I)} \quad (5)$$

The first term of the right side of equation (5) 1/P(f|I) is the task-irrelevant (bottom-up) component, and the second and the third term are task-relevant. In terms of the definition of the bottom-up saliency, objects are more salient only when they are sparser in the image. However, this modeling does not consider the local information. To achieve balance between global and local information, in our model we combine them as follows:

$$\mathrm{SaliencyMap_{s,f}(x) = EMD\_wH_{f,s}(x)} \times \log\left(\frac{1}{\mathrm{P(f_s|I)}}\right) \quad (6)$$

where $\mathrm{SaliencyMap_{s,f}(x)}$ denotes the value in the location x in the saliency map with feature f and scale s, $\mathrm{EMD\_wH_{s,f}}$ denotes the EMD_wH map with the feature f and scale s, and $\mathrm{P(f_s|I)}$ denotes the probability of feature f at scale s, which can be computed directly. The saliency maps of each feature dimension at each scale thus consider both local and global information. In this way, the saliency of objects remains low if they are not sparse in the image even if they have strong contrast with their surrounding region, as shown in Figure 3 (c). On the other hand, objects cannot be considered as salient only because they are sparse in the image; their local contrast must also be strong enough.

Feature Set: Local Entropy

Traditional computational models usually employ the features such as colors, orientations, intensity and etc. It is clearly that the saliency regions are highlighted based on the image complexity by computing the centre-surround metric using these features. However, saliency not always be equated with complexity (Gao & Vasconcelos 2007). For example, Figure 4 shows two challenge images containing simple object in a complex background. They contain complex regions, consisting of clustered straw and crayons that are not terribly salient. On the contrary, the much less complex image regions containing the bird or the egg or just the blank appear to be significantly more salient.

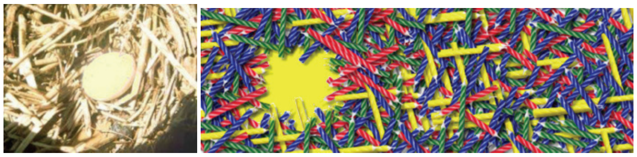


Figure 4: Challenging examples

To handle this problem, we introduce local entropy (Kadir & Brady 2001) as a feature into our model for generating saliency map. Given a location *x*, local Shannon entropy is defined as:

$$\mathrm{E}_x = -\sum_i \mathrm{P}_{\mathrm{N}x}(i)\log_2 \mathrm{P}_{\mathrm{N}x}(i) \qquad (7)$$

where N*x* is the local neighborhood region of *x*, $\mathrm{P}_{\mathrm{N}x}(i)$ is the probability of the pixel value *i* in N*x*.

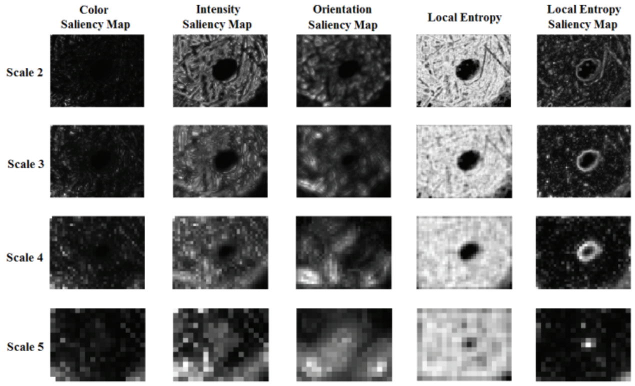


Figure 5: Feature saliency maps in variety scales

Local entropy indicates the complexity or unpredictability of a local region. Regions corresponding to high signal complexity tend to have flatter distributions hence higher entropy and these regions are considered to be the salient region (Kadir & Brady 2001).

Instead of using local entropy directly to measure the complexity of a region in an image and defining saliency as complexity (Kadir & Brady 2001), we use the local entropy to generate the saliency map by using the centre-surround metric. In other words, we don't care what the value of the local entropy of a region is; we only consider the difference of the local entropies between the region and its surround. For example, the regions of straw in the left image of Figure 4 have large local entropy, and the region of egg has a small one, see the fourth column of Figure 5. By using the EMD_wH, we can get the salient region in the local entropy saliency map related to the egg; see the fifth column of Figure 5, while the other feature saliency maps fail to relate to the egg.

Feature Integration

There are mainly two types of feature integration models, those proposed by Itti & Koch and Li (called the "V1 Model"), respectively. We here propose a more general model to integrate features.

Itti's Feature Integration Model and V1 Model

Itti's feature integration model was developed by Itti and Koch (Itti & Koch 2001; Itti et al. 1998; Walther & Koch 2006). In this model, among all of the chosen features, saliency maps are generated by extracting the feature strength at several scales and combining them in a center-surround approach that highlights the regions that stand out from their neighbors. Then, the individual feature saliency maps are summed to generate a master saliency map. We call this a summation model since all the feature maps are directly summed together.

The primary visual cortex (V1) is the simplest, earliest cortical visual area. The V1 model is a biologically based model of the preattentive computational mechanisms in the primary visual cortex which was proposed and developed by Li (Li 2002; Koene & Zhaoping 2007), showing how V1 neural responses can create a saliency map that awards higher responses to more salient image locations. Each location evokes responses from multiple V1 cells that have overlapping receptive fields (RFs) covering this location; these V1 cells include many types, each tuned to one or more particular features. The saliency value of a location x is determined by the V1 cell with the greatest firing rate that responds to x. We refer to the cells tuned to more than one feature as feature conjunctive cells (Koene & Zhaoping 2007); e.g., there are cells that can respond to both color and orientation (we call them CO cells).

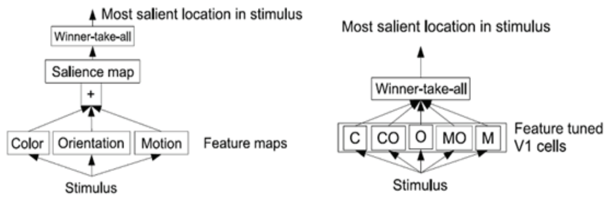


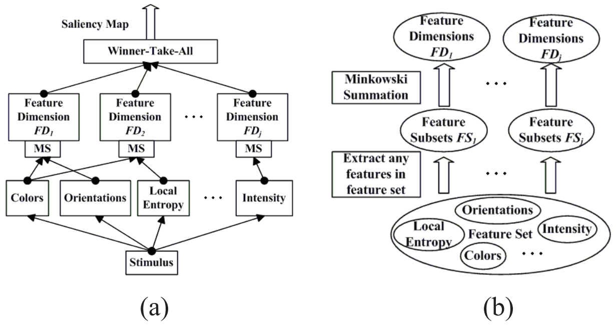
Figure 6: The difference between Itti model and V1 model. Figure adapt from (Koene & Zhaoping 2007).

The key differences between the V1 model and the summation model are: first, that the V1 model does not sum the separate feature-based information, and second, that the summation model only includes the single features, whereas the V1 model includes combinations of features to which V1 cells are turned, as shown in Figure 6 (Koene & Zhaoping 2007)

General Model of Feature Integration

The summation model was shown to be problematic in some literatures (such as Riesenhuber & Poggio 1999). Theoretical investigations into possible pooling mechanisms for V1 complex cells support a maximum-like pooling mechanism.

Here, we use a more general framework for Feature Integration. A feature set is generated first, which includes all the features we want to analyze the image for, such as colors, orientations, intensity, etc. We can extract any features in the feature set to obtain particular subsets. Each feature dimension is generated by combining the features in one feature subset using Minkowski Summation (MS) (To et al. 2008) as (8); see Figure7 (a). All the feature dimensions are then used generate the saliency map by using a Winner-Take-All (WTA) approach, see Figure7 (b).



(a) (b)

Figure 7: Overview illustration of proposed framework and feature dimension generation.

Note that there are no intensity-tuned or entropy-tuned cells, to the best of our knowledge. However, these features are effective in computer vision, so we use them all the same, but apply a weight of 0.6 and 0.4 to them, respectively, see (10) and (11).

We employ MS of each feature in a given feature subset as a feature dimension in our model. The jth feature dimension $\mathrm{FD_j}$ is produced by combining features in the feature subset $\mathrm{S_j}$ using MS, as follows:

$$\mathrm{FD_j} = \left(\sum_{i=1}^{n} \mathrm{F_i^m}\right)^{1/m} \qquad \mathrm{F_i} \in \mathrm{S_j} \qquad (8)$$

where $\mathrm{F_i}$ is a single feature belongs to $\mathrm{S_j}$, n is the number of features in $\mathrm{S_j}$, and m is the Minkowski exponent. Clearly $\mathrm{FD_j} \geq \max_i(\mathrm{F_i})$, i.e., MS will consider all the responses from each single feature and biasd towards the larger one. In our model, we set m = 2.8. The "MAX" operation is then used as the "WTA" mechanism to generate the saliency map from the feature dimensions.

It is obvious that the Itti model is a particular case of this general model, in which only one feature dimension is considered, which includes the orientation, color and intensity, and the Minkowski power is set to m=1.

We select three feature dimensions as follows:

$$\text{Feature dimension } 1 = \text{Color} \oplus \text{Orientation} \quad (9)$$

$$\text{Feature dimension } 2 = 0.6 \times \text{Intensity} \quad (10)$$

$$\text{Feature dimension } 2 = 0.4 \times \text{Local\_Entropy} \quad (11)$$

where the $\oplus$ denotes Minkowski summation.

Biased Competition

According to biased competition theory, attentional selection operates in parallel by biasing an underlying competition between multiple objects in the visual field toward one object or another at a particular location or with a particular feature (Desimone and J. Duncan 1995; Deco & Rolls 2005). In our model, "large scale bias" and "central bias" hypotheses are used as biased competition schemes when we get the master saliency map.

Large Scale Bias

A visual system works under a certain scale at one time. For example, in a large scale, one may perceive a football field as an object, but in a smaller scale it is very likely that a player on the field, or even a football will be popped out as an object. Draper and Lionelle proposed a model called SAFE (Draper & Lionelle 2005). Unlike those in the Itti model, saliency maps in SAFE are not combined across scales within a feature dimension. Instead, saliency maps are combined across dimensions within each scale, producing a pyramid of saliency maps. In other words, SAFE treats every scale independently and equally.

Intuitively, however, we are more likely to capture the large object than a small one if they have similar saliency values in their own scale. Therefore, we assume "large scale bias" in attention tasks. We employ following equation to determine the most suitable scale for attention:

$$j = \arg\max_s \left(w_s \times \text{maxvalue}(i)\right) \quad (12)$$

where the function $\text{maxvalue}(s)$ denotes the largest value in the saliency map of scale s, and w_s is the weight of scale s. w_s increases as the scale gets larger as discussed above, and it is set to be an exponential function of the scale s: $w_s = \alpha^s$, where the bias parameter α is set to be 1.1 in our model. That means that the weight of a larger scale is 1.1 times that of the next smaller one.

Central Bias

When people view images, they have a tendency to look more frequently around the center of the image (Tatler 2007).

Avraham and Lindenbaum experimented with many different image sets, and all the results show the preference for the center (Avraham & Lindenbaum 2010). In order to simulate this central bias, our model employs a two-dimension anisotropic Gaussian function with standard deviations (σ_x, σ_y):

$$f_s(x,y) = \exp\left\{-\left(\frac{(x-x_0)^2}{2\sigma_x^2} + \frac{(y-y_0)^2}{2\sigma_y^2}\right)\right\} \quad (13)$$

where (x_0, y_0) denotes the center of image. We set $\sigma_y = \text{Height}_s/5$, and $\sigma_x = 1.3 \times \text{Width}_s/5$, where Height_s and Width_s denote the height and the width of the image at scale s, respectively. Then the central bias function is used as weighted sum with the saliency maps in each scale to produce the final saliency maps in each scale $\text{SaliencyMap}'_s = \text{SaliencyMap}_s + \ w \times f_s$, and we set w=0.5.

Experimental Results

Our experiments consist of two parts: first, we give qualitative comparison to prove the usefulness of the feature of local entropy; second, we compared the saliency maps generated by using our method to those generated by using the Itti model (saliencytoolbox, Walther & Koch 2006), AIM (Bruce & Tsotsos 2009), and Esaliency (Avraham & Lindenbaum 2010) using images from a image set (Bruce & Tsotsos 2009).

First, we need combine the saliency maps at all the scales to form a master saliency map, i.e. $\text{MasterMap} = \sum_s (w_s \times \text{saliencyMap}_s)$, where w_s is the weight of scale s which defined above. Then we can compare the master maps produced using the different methods.

The master maps generated by AIM, Esaliency and our model should be normalized before comparison. First we employ the soft threshold shrink to process the master maps with setting the threshold as the average intensity of its master map for AIM and twice the average intensity for Esaliency and our model, respectively.

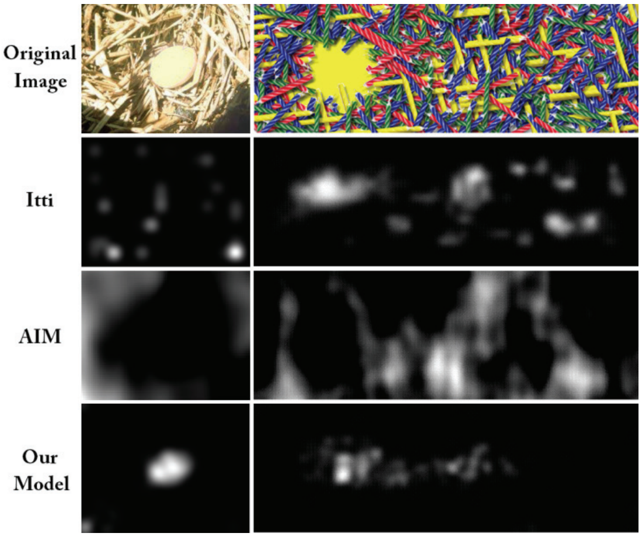


Figure8: Comparisons to prove the usefulness of the local entropy

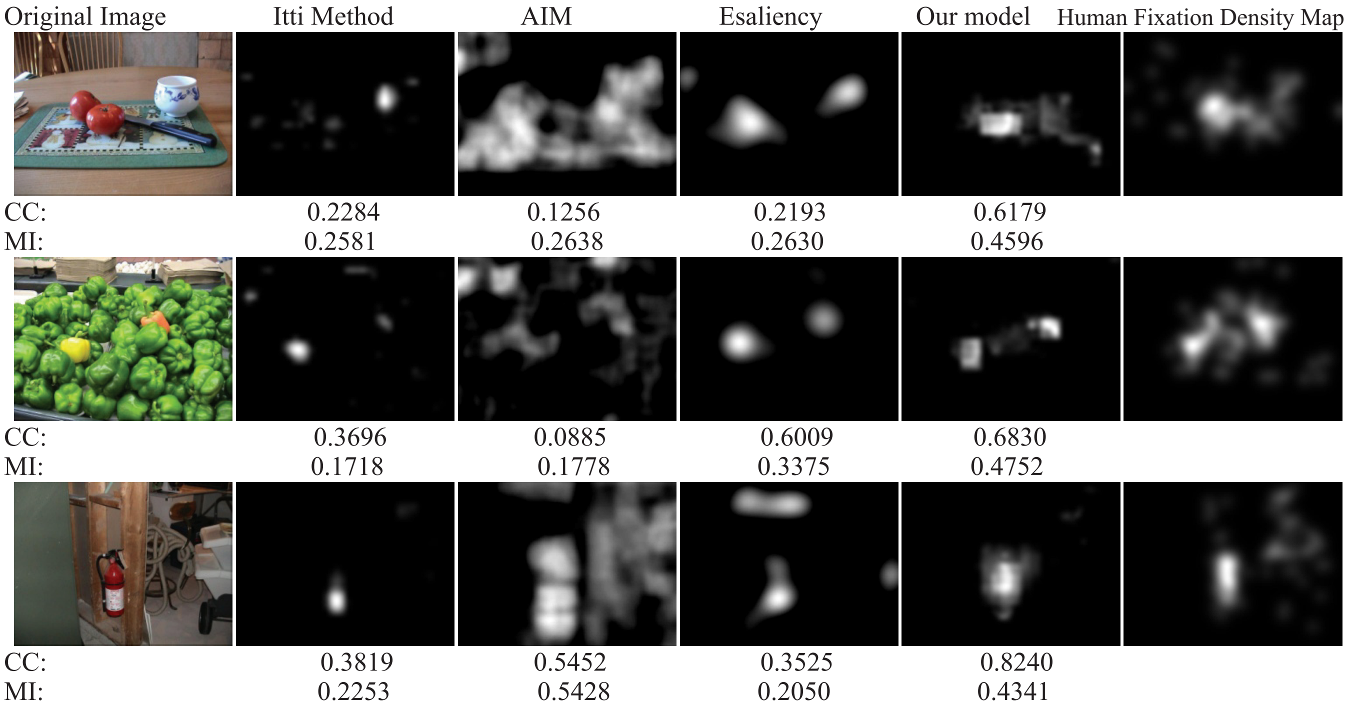

Original Image
Itti Method
AIM
Esaliency
Our model
Human Fixation Density Map
CC: 0.2284 0.1256 0.2193 0.6179
MI: 0.2581 0.2638 0.2630 0.4596
CC: 0.3696 0.0885 0.6009 0.6830
MI: 0.1718 0.1778 0.3375 0.4752
CC: 0.3819 0.5452 0.3525 0.8240
MI: 0.2253 0.5428 0.2050 0.4341

Figure 9: Qualitative and quantify comparison of the saliency map by using different models

The experimental data includes 120 different color images and 120 corresponding human eye tracking density maps. The human eye tracking density maps depict the average extent to which each pixel location was sampled by human observers.

We use the correlation coefficients (CC) and mutual information (MI) between the human fixation density map and the master saliency map produced using each method to compare the effects of each method. The correlation coefficient is a well-known metric to measure the strength of a linear relationship between two images. It is defined as:

$$\mathrm{cc(hmap, smap)} = \frac{\mathrm{cov(hmap, smap)}}{\sigma_{\mathrm{hmap}}\sigma_{\mathrm{smap}}} \qquad (14)$$

where hmap and smap represent the human fixation density map and the master saliency map, respectively; $\mathrm{cov(hmap, smap)}$ is the covariance value between hmap and smap; σ_{h} and σ_{p} are the standard deviation for the hmap and smap, respectively.

Mutual Information (MI) has been widely used because of the robustness of MI to occlusion, noise, and its tolerance of nonlinear intensity relationships. The mutual information of hmap and smap (256 levels) is defined as:

$$\mathrm{MI(hmap, smap)} = \sum_{\mathrm{i=0,j=0}}^{\mathrm{i=255,j=255}} \mathrm{p(i,j)log}\frac{\mathrm{p(i,j)}}{\mathrm{p(i)p(j)}} \qquad (15)$$

where $\mathrm{p(i), p(j)}$denote the distribution of the grey levels of two images, respectively. And $\mathrm{p(i,j)}$ denotes the joint distribution of the two images' gray levels $\mathrm{p(i,j)}$. It is a measure of dependence between hmap and smap.

	CC	MI
Itti	0.2128±0.2061; 0.1667	0.1625±0.0616; 0.1641
AIM	0.2795±0.1977; 0.2998	0.3220±0.1338; 0.3008
Esaliency	0.3068±0.2621; 0.2749	0.2171±0.1104; 0.2009
Our model	0.4945±0.2290; 0.5426	0.3673±0.1368; 0.3784

Table 1: The Mean, Standard Deviations and Median Number of CC and MI Which Simulate All 120 Images

Figure 8 represents the comparisons to prove the usefulness of the local entropy. Figure 9 represents the master saliency maps using four different models. Also, it gives the CC and MI between human fixation density map and master saliency maps. Table 1 represents the mean, standard deviations and median number of 120 CC and MI obtained from all the images in the dataset, respectively.

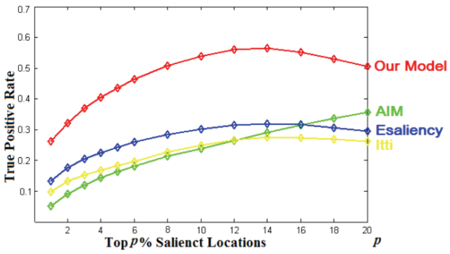


Figure 10: Performance comparison of different saliency models.

Performance comparison is also made as follow: the saliency maps produced by using all the four models and the human fixation density maps are binarized to show the top p percent salient locations, respectively. In our experiment, we set p =1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 18 and 20. In Figure 10, the curves denote the averaged true positive rate of each model with different p.

Discussion and Future Works

In this paper we introduced a computational model for saliency maps. The proposed model was compared both qualitatively and quantitatively to reference models. The proposed model outperforms the reference models in majority of cases based on the experimental results.

Note that this model only considers four features, intensity, color, and orientation and local entropy. It would be possible to improve its performance by considering other visual features. However, how to choose the features and how to combine each feature dimension is an open question; to solve it we would need more evidence from neuroscience to demonstrate that there are particular feature tuned or conjunctive cells for these features.

Acknowledgment

This work is supported in part by the National Natural Science Foundations of China (60873092&90820306), and the Program for New Century Excellent Talents of Educational Ministry of China (NCET-06-0762). The authors would like to thank Tamar Avraham for her kindly help and Neil D. B. Bruce for his very helpful advice.

References

Avraham T., and Lindenbaum, M. 2010. Esaliency (Extended Saliency): Meaningful Attention Using Stochastic Image Modeling. *IEEE Transaction on Pattern Analysis Machine and Intelligence.*

Bickel, P., and Freedman, D. A. 1981. Some asymptotic theory for the bootstrap. *Annals of Statistics.*

Bruce N. D. B., and Tsotsos, J. K. 2009. Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision.*

Deco, G., and Rolls, E. T. 2005. Neurodynamics of Biased Competition and Cooperation for Attention: A Model with Spiking Neurons. *Journal of Neurophysiology.*

Desimone, R., and Duncan, J. 1995. Neural mechanisms of selective visual attention. *Annual Review Neuroscience.*

Draper, B. A., and Lionelle, A. 2005. Evaluation of selective attention under similarity transformations. *Computer Vision and Image Understanding.*

Gao, D., and Vasconcelos, N. 2007. Bottom-up saliency is a discriminant process. *In Proceeding of IEEE International Conference Computer Vision.*

Itti, L., and Koch, C. 2001. Computational modeling of visual attention. *Nature Review Neuroscience.*

Itti, L., Koch, C., and Niebur, E. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transaction on Pattern Analysis Machine and Intelligence.*

Kadir, T., and Brady, M. 2001. Scale, saliency and image description. *International Journal of Computer Vision*.

Koene, A. R., and Zhaoping, L. 2007. Feature-specific interactions in salience from combined feature contrasts: Evidence for a bottom–up saliency map in V1. *Journal of Vision.*

Levina, E., and Bickel, P. 2001. The Earth Mover's Distance is the Mallows Distance: Some Insights from Statistics. *In Proceeding IEEE International Conference of Computer Vision.*

Li, Z. 2002. A saliency map in primary visual cortex. *Trends in Cognition Science.*

Liversedge, S. P., and Findlay, J. M. 2000. Saccadic eye movements and cognition. *Trends in Cognition Science.*

Riesenhuber, M., and Poggio, T. 1999. Hierarchical models of object recognition in cortex. *Nature Neuroscience.*

Rubner, Y., Tomasi, C., and Guibas, L. J. 2000. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision.*

Tatler. B. W. 2007. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision.*

To, M., Lovell, P. G., Troscianko, T., and Tolhurst, D. J. 2008. Summation of perceptual cues in natural visual scenes. *Proceedings of the Royal Society B.*

Torralba, A., Oliva, A., Castelhano, M. S., and M.Henderson, J. 2006. Contextual guidance of eye movements and attention in real-world scenes: The role of global features on object search. *Psychology Review.*

Walther, D., and Koch, C. 2006. Modeling attention to salient proto-objects. *Neural Networks.*