NONPARAMETRIC SALIENCY DETECTION USING KERNEL DENSITY ESTIMATION

Zhi Liu¹,², Yinzhu Xue¹, Liquan Shen¹,², Zhaoyang Zhang¹,²

¹School of Communication and Information Engineering, Shanghai University, China
²Key Laboratory of Advanced Display and System Application, Ministry of Education, China

ABSTRACT

This paper proposes a nonparametric saliency model based on kernel density estimation (KDE) mainly aiming at content-based applications such as salient object segmentation. A set of KDE models are constructed on the basis of regions segmented using the mean shift algorithm. For each pixel, a set of color likelihood measures to all KDE models are calculated, and then the color saliency and spatial saliency of each KDE model are evaluated based on its color distinctiveness and spatial distribution. The final saliency map is generated by combining saliency measures of KDE models and color likelihood measures of pixels. Experimental results demonstrate the better saliency detection performance of our saliency model.

Index Terms— Saliency detection, kernel density estimation, color saliency, spatial saliency

1. INTRODUCTION

Saliency detection from images plays a key role in a number of content-based applications including salient object segmentation/detection, image retargeting, and region of interesting coding. Generally, saliency is defined as what captures human perceptual attention. Human vision system (HVS) has the ability to effortlessly identify salient objects in a complex scene, since the inherent visual attention mechanism is exploited to rapidly locate the most significant portion of the scene. With the goal to achieve a comparable saliency detection performance of HVS, many computational saliency models have been proposed in the past decades.

Based on Koch and Ullman’s biologically-plausible visual attention architecture [1], Itti et al. [2] proposed a computational saliency model, which first computes feature maps for luminance, color and orientation using a center-surround operator across different scales, and then generates the saliency map by normalization and summation on these feature maps. Itti’s saliency model mainly exploits the center-surround scheme that the center pixel/region is salient if it is discriminated from its surrounding pixels/regions. In the latterly proposed saliency models, the center-surround scheme is realized using different features, such as local contrast [3], multi-scale contrast [4], histogram of filter responses [5], local regression kernels [6], with different mathematical formulations. Saliency detection can also be performed in the frequency domain. Both spectral residual using Fourier transform [7] and the phase spectrum of quaternion Fourier transform [8] are exploited to measure the pixel’s saliency. In [9], a conditional random field is learned to combine the features of center-surround histogram, multi-scale contrast and color spatial distribution for saliency detection. In [10], a frequency-tuned saliency model is proposed to improve the saliency map with well-defined boundaries of salient object. In [11], the distribution of color and orientation over the whole image are exploited to evaluate pixel saliency. In [12], isophotes properties with image curvature and color edges are integrated for saliency estimation.

Among these saliency models, most of them generate spotlight saliency maps [2-8], which usually can only highlight the center portion or the high-contrast boundaries of salient object, while cannot reasonably highlight the complete salient object with accurate boundaries. Spotlight saliency maps are useful for predicting eye fixations and roughly detecting salient objects, but are not sufficient for content-based applications such as salient object segmentation. Although salient objects can be more completely highlighted in the saliency maps generated using other saliency models mentioned above, but the boundaries between salient object and background are usually not accurately preserved, and the contrast between salient object and background is not sufficiently significant in these saliency maps.

In order to provide more appropriate saliency maps for a variety of content-based applications, we propose a nonparametric saliency model based on KDE. Compared with previous saliency models, the distinct characteristic of our saliency model is that color saliency measures and spatial saliency measures of a set of KDE models are efficiently utilized to generate the pixel-wise saliency map.

The rest of this paper is organized as follows. Section 2 describes region based KDE modeling, and Section 3 details our KDE based saliency model. Experimental results and comparisons with previous saliency models are presented in Section 4, and conclusions are given in Section 5.
2. REGION BASED KDE MODELING

The original image is first partitioned into a set of regions using the mean shift algorithm [13] in the Luv color space. The parameters of spatial bandwidth and range bandwidth in the mean shift algorithm are set to the default values. We only adjust the parameter of allowable minimum region area, which controls the balance between over-segmentation and under-segmentation, and set to 3 percent of image size by experiments. Fig. 1(a) shows an example of region segmentation result, in which the boundaries of segmented regions are delineated with white lines.

The pixels in each segmented region $R_i (i = 1, \ldots, n)$ are then used as the samples to construct a KDE based nonparametric model $K_i$. Then for each pixel at $(x, y)$, its color likelihood to each KDE model $K_i$ is defined as

$$C_i(x, y)=\frac{1}{|R_i|}\sum_{(x', y')\in R_i} \kappa_i(e_{x', y'} - e_{x, y})$$

(1)

where $|R_i|$ denotes the number of pixels in $R_i$, i.e., the number of samples in $K_i$. $e_{x, y}$ and $e_{x, y}$ denote the color feature of pixel at $(x, y)$ and $(m, n)$, respectively. Gaussian distribution is selected as the kernel function $\kappa_i$ due to its continuity, differentiability and locality properties [14]. Let $\epsilon_i$ denotes $e_{x, y} - e_{x, y}$, $\kappa_i$ is defined as

$$\kappa_i(\epsilon_i) = \frac{1}{(2\pi)^{\frac{3}{2}}|\mathbf{H}|^{\frac{3}{2}}} \exp\left(-\frac{1}{2} \epsilon_i^T \mathbf{H}^{-1} \epsilon_i \right)$$

(2)

where $\mathbf{H}$ is the bandwidth matrix. Using all the sample pixels in each KDE model, the bandwidth matrix for each kernel function is estimated using the fast binned kernel density estimator [15].

The differences among different KDE models are then evaluated in both color domain and spatial domain. Specifically, the color distance between a pair of KDE models, $K_i$ and $K_j$, is defined as

$$D_{ij}(x, y) = \frac{\sum_{i, j} C_i(x, y) \left\| \mathbf{d}_i(x, y) \right\| + \sum_{j, k} C_j(x, y) \left\| \mathbf{d}_j(x, y) \right\|}{\sum_{i, j} C_i(x, y) + \sum_{j, k} C_j(x, y)}$$

(3)

where $\mathbf{d}_i(x, y) = e_{x, y} - \mu_i$, and $\mu_i$ is the mean color of the sample pixels in $K_i$. Similarly, the spatial distance between a pair of KDE models, $K_i$ and $K_j$, is defined as

$$D_{ij}(x, y) = \frac{\sum_{i, j} C_i(x, y) \left\| \mathbf{d}_i(x, y) \right\| + \sum_{j, k} C_j(x, y) \left\| \mathbf{d}_j(x, y) \right\|}{\sum_{i, j} C_i(x, y) + \sum_{j, k} C_j(x, y)}$$

(4)

where $\mathbf{d}_i(x, y) = (x - \bar{x}_i, y - \bar{y}_i)^T$. $(\bar{x}_i, \bar{y}_i)$ is the weighted spatial center position for $K_i$, and is defined as

$$\bar{x}_i = \frac{\sum_{(x, y)\in K_i} x \cdot C_j(x, y)}{\sum_{(x, y)\in K_i} C_j(x, y)}, \bar{y}_j = \frac{\sum_{(x, y)\in K_i} y \cdot C_j(x, y)}{\sum_{(x, y)\in K_i} C_j(x, y)}$$

(5)

3. KDE BASED SALIENCY MODEL

The color saliency and spatial saliency of each KDE model are then evaluated based on its color distinctiveness and spatial distribution. In natural images, the colors of salient objects are usually distinctive from background colors, and thus salient object pixels have larger distances to other pixels in the color domain. For KDE models, if the colors covered by $K_i$ are far from the colors covered by other KDE models, the colors covered by $K_i$ are such distinctive colors. The color saliency for $K_i$ is then defined as the sum of weighted color distances between $K_i$ and all the other KDE models

$$KS_i(i) = \sum_{j=1}^{n} \alpha_j \cdot D_{ij}(i, j) - \alpha_i \cdot D_{ii}(i, i)$$

(6)

where the weight $\alpha_j$ is the ratio of the number of samples in $K_j$ to the total number of samples in all KDE models. Eq. (6) indicates that only inter-distances, $D_{ij}(i, j), \forall j \neq i$, are used for evaluating color saliency. Since the intra-distance $D_{ii}(i, i)$ actually represents the color homogeneity of the sample pixels in $K_i$, we should not introduce such a factor that one KDE model covering more colors is more salient than another KDE model covering fewer colors, and thus $D_{ii}(i, i)$ is excluded from our color saliency measure.

For the example in Fig. 1(a), the color saliencies of all KDE models are normalized with $\sum_i KS_i(i) = 1$ and shown in Fig. 1(e), in which each KDE model is represented using a bar with its mean color. We can see that the color saliencies of 6 KDE models (from the 4th to the 9th bar) covering the colors of the butterfly are obviously elevated, while the color saliencies of other KDE models are efficiently suppressed.

In the spatial domain, salient objects are generally surrounded by background regions, and thus the colors of background regions have a wider spatial distribution over...
the whole image than the colors of salient objects. Based on Eq. (4), KDE models that mainly cover the colors of salient objects have shorter spatial distances to other KDE models. Therefore, the spatial saliency for \( K_i \) is defined as the reciprocal of the sum of weighted spatial distances between \( K_i \) and all KDE models.

\[
KS_i(i) = \frac{1}{\sum_{j=1}^{n} \alpha_j \cdot D_i(i,j)}
\] (7)

In contrast with Eq. (6), Eq. (7) includes the intra-distance \( D_i(i,i) \), which actually represents the spatial distribution of colors covered in \( K_i \), and thus it is incorporated in our spatial saliency measure. For the example in Fig. 1(a), the spatial saliencies of all KDE models are normalized with \( \sum_{i} KS_i(i) = 1 \) and shown in Fig. 1(f), in which the spatial saliencies of the same 6 KDE models as in Fig. 1(e) also have higher values, while the spatial saliencies of other KDE models are suppressed.

Based on color saliencies and spatial saliencies of all KDE models, the pixel-wise color saliency map \( S_c \) and spatial saliency map \( S_s \) are generated as follows

\[
S_c(x,y) = \sum_{i=1}^{n} C_i(x,y) \cdot KS_i(i)
\] (8)

where the subscript \( a \) may denote \( c \) or \( s \). Eq. (8) indicates that the color/spatial saliency for each pixel is the sum of color/spatial saliencies of all KDE models weighted by its color likelihood measures. Using Eq. (8), the global color information of the image is incorporated into the saliency calculation for each local pixel. By integrating color saliency map with spatial saliency map, the final saliency map \( S \) is generated as follows

\[
S(x,y) = S_c(x,y) \cdot S_s(x,y)
\] (9)

Based on color/spatial saliencies of KDE models shown in Fig. 1(e) and (f), the color saliency map, the spatial saliency map and the final saliency map are shown in Fig. 1(b), (c) and (d), respectively. The three saliency maps are normalized into the range of [0, 255] for display. It can be seen that the complete salient object region is highlighted and most background regions are suppressed in Fig. 1(b) and (c). Moreover, the color saliency map and the spatial saliency map can complement each other to generate a more reasonable final saliency map as shown in Fig. 1(d).

4. EXPERIMENTAL RESULTS

We evaluate the performance of our saliency model on 1000 test images with the manually segmented ground truths for salient objects [10]. These test images are selected from a publicly available saliency dataset containing 5000 high-quality images [9]. We compare the performance of our saliency model implemented using Matlab against three previous saliency models, i.e., Itti’s model [2], Hou’s model [7] and Achanta’s model [10]. We use the Matlab code [16] from http://www.saliencytoolbox.net/ for Itti’s model, the Matlab code from http://www.its.caltech.edu/~xhou/ for Hou’s model, and the C++ code from http://ivrg.epfl.ch/supplementary_material/RK_CVPR09/index.html for Achanta’s model, respectively. Since Itti’s model and Hou’s model generate block-level saliency maps with variable ranges, these saliency maps are upsampled to full resolution and normalized into the same range of [0, 255]. In order to reduce the computation cost without noticeable quality degradation of saliency maps, the input images are first resized to half width and half height for our saliency model, and the generated saliency maps are finally upsampled to the full resolution of the original image.

Experimental results on some test images are shown in Fig. 2, in which the original images, the ground truths, and the four classes of saliency maps generated using Itti’s model, Hou’s model, Achanta’s model and our model are shown from the 3rd column to the 6th column in turn. It can be seen from Fig. 2 that the saliency maps generated using our model can generally highlight the complete salient objects with accurate boundaries and effectively suppress background regions. In contrast, salient objects cannot be completely highlighted in the saliency maps generated using Itti’s model and Hou’s model, and the background regions cannot be effectively suppressed in some saliency maps generated using Achanta’s model and Hou’s model. Compared with Itti’s and Hou’s saliency maps, salient objects are more completely highlighted in Achanta’s
Fig. 3. ROC curves for four saliency models.

saliency maps, but the contrasts between salient object regions and background regions in Achanta’s saliency maps are not as distinct as that in our saliency maps. In particular, it can be observed from Fig. 2 that our saliency model can efficiently handle such images, in which backgrounds are clutter (see the 3th and the 4th row) or with similar colors as salient objects (see the 7th and the 8th row). It should be noted that any region with distinctive colors is generally highlighted in our saliency maps as well as other saliency maps, but such a region may not belong to any salient object as defined in the ground truth (see the last row).

Using the four classes of saliency maps and the ground truths for all test images, we can make an objective comparison of saliency detection performance among different saliency models. The binary segmentation of salient objects can be obtained by performing the thresholding operation on the saliency map. By incrementing the threshold from 0 to 255, we can obtain 256 binary masks of salient object for each saliency map. Assume that one binary mask of salient object generated by thresholding a saliency map is denoted by $B_{xy}$ and the corresponding ground truth is denoted by $G_{xy}$. The true positive rate (TPR) and the false positive rate (FPR) are defined as

$$\text{TPR} = \frac{\sum_{(x,y)} B_{xy} \cdot G_{xy}}{\sum_{(x,y)} G_{xy}}$$

$$\text{FPR} = \frac{\sum_{(x,y)} B_{xy} \cdot [1 - G_{xy}]}{\sum_{(x,y)} [1 - G_{xy}]}$$

The saliency detection performance of different models is evaluated using the receiver operating characteristic (ROC) curve. As shown in Fig. 3, the ROC curve for each saliency model plots the average TPR value versus the average FPR value of 1,000 saliency maps at each threshold. These ROC curves objectively demonstrate that our saliency model outperforms the other three saliency models.

5. CONCLUSIONS

We have presented an efficient KDE based nonparametric saliency model for content-based applications. Due to the integration of color saliency and spatial saliency of a set of KDE models with the color likelihood measures of pixels, our saliency model can generally highlight the complete salient objects with accurate boundaries, and thus can provide an overall better saliency detection performance than previous saliency models.

6. ACKNOWLEDGMENTS

This work is sponsored by Shanghai Rising-Star Program (No. 08QA14031), and supported by National Natural Science Foundation of China under Grant No. 60602012 and Innovation Program of Shanghai Municipal Education Commission (No. 09YZ02).

7. REFERENCES