COLOR SALIENCY MODEL BASED ON MEAN SHIFT SEGMENTATION

Xu Liu[1], *Zengchang Qin*[1*], *Xiaofan Zhang*[1], *and Tao Wan*[2]

[1]Intelligent Computing and Machine Learning Lab
School of ASEE, Beihang University, Beijing, 100191, China
[2]Boston University School of Medicine, MA 02215, USA

ABSTRACT

Saliency detection is one of the extraordinary capabilities of the human visual system (HVS). In this paper, we present a novel saliency detection model to capture visual selective attention of images. The new model does not require prior knowledge of salient regions as well as manual labeling. The mean shift segmentation algorithm and quaternion discrete cosine transform (QDCT) are used to generate a rough saliency map by integrating low-level features and spatial saliency information. In each segmented region, the color saliency is measured based on the probability of its occurrences in foreground and background defined by the rough saliency map. The experimental results on a widely used benchmark database demonstrated that the presented model achieves the best performance in terms of visual and quantitative evaluations compared to existing state-of-the-art saliency detection models.

Index Terms— Saliency detection, quaternion discrete cosine transform, image segmentation, mean shift.

1. INTRODUCTION

Visual system is one of the most important and effective way we use to perceive the world. Human attention is often attracted by salient regions when observe pictures or videos. The human vision system (HVS) has a remarkable ability to identify salient regions from a noisy background [1, 2]. Computational saliency detection is essential in many computer vision tasks, such as object detection [3], object tracking [4], and scene analysis [5]. It remains a challenging task to build a saliency detection model to accurately reflect human attention system due to the complexity of human perception. Fig.1 gives four examples of human attention regions within the sample images. As we can see, the computed salient regions obtained by the presented model capture the shapes of regions of interest (ROI) manually labeled by human subjects.

Two important paradigms of attention process have been extensively studied in literature: the fast, parallel, but simple *pre-attentive* process [6], and the serial, complex, but effective *attention* process [7]. In the *pre-attentive* process, multiple low-level features, such as edges, color, orientation, and intensities, attract more attention of human perception. These regions are the desirable candidates of saliency regions. Rensink [8] introduced a concept of *proto objects* to address these candidates that appear in the first stage of HVS but are not identified as saliency. For the *attention* process, color contrast and spatial information distribution can provide high-level information regarding the salient regions. There are a number of publications in this area, in which the proposed saliency detection models were validated either in the benchmark or real-world problems and reported to be an effective way to identify salient objects [3, 6, 9, 10, 11].

Fig. 1. The top row shows the original sample images. The saliency maps obtained by the presented model and the ground truth binary masks of ROI are shown in middle and bottom rows, respectively.

From an information processing perspective, saliency detection methods can be mainly divided into two categories: top-down and bottom-up. The one of the most popular top-down methods was proposed by Koch and Ullman [12]. Itti *et al.* [5] also developed a bottom-up saliency model according to the human visual process. However, such models tend to rebuild biological functioning of HVS, thus leading to high computational complexity. Recently, Cheng [9] presented a simple and effective saliency region detection method based on a global contrast, which outperformed exiting saliency detection models. In this paper, our main contribution is to es-

* Corresponding author's email: zcqin@buaa.edu.cn. This work is supported by the Innovation & Practical Foundation of BUAA for Graduates, No.YCSJ-02-06, and the NCET Program of MOE of China.

tablish a novel bottom-up model to accurately predict salient regions using a combination of low-level edge features and a color contrast related spatial frequency measure. The detection model presented in this work is derived from the principle of visual contrast, which suggests that human visual system tends to focus on the unique color area in messy segmented regions.

The rest of the paper is organized as follows. Section 2 describes the presented saliency detection model and the detailed algorithm. Experimental results are presented and compared to the previous work in Section 3. Finally, the conclusion is drawn in section 4.

2. SALIENCY DETECTION

2.1. The QDCT-based Saliency Detection Model

Quaternions are the extension of complex numbers. Quaternion discrete cosine transform (QDCT) [13], based on discrete cosine transform (DCT), has been proved to be a very useful tool for digital color image processing [6, 13]. Since a color image can be represented by a quaternion matrix, rather than separating it into three image channels and processing each channel respectively as the traditional methods do, QDCT can handle color image pixels as vectors and process them in a holistic manner. QDCT has been used for saliency detection and showed good results for detecting small salient regions, but performed poorly on big targets [6]. This is due to the fact that QDCT treats the large salient target as part of the gist of image and therefore fails to detect it. The QDCT model has two main disadvantages in saliency detection: (i) it often fails to give the accurate outline of salient object, and (ii) it misclassifies the background pixels with high color contrast as saliency without considering the continuum of salient object. Fig.2 (the 3rd row) shows the intensities of saliency within the original images using QDCT. It clearly indicates that QDCT saliency detection method is able to roughly predict the locations of salient regions, but is insensitive to the edge information, which results in blurry and unshaped predicated region. Therefore, additional information is needed in order to draw clear outlines of salient objects.

2.2. Generation of Rough Saliency Maps

Image segmentation is generally used to partition an image into multiple segments, such as mean shift [14], which can provide an accurate shape of object. By combining the spatial saliency information and segmentation, a rough saliency map is generated indicating the locations of salient regions with improved object contours. Given a segmented region $R_k, k = 1, ..., K$ (K is the number of segmented regions), the average saliency intensity of each region R_k is computed based on the corresponding QDCT coefficients within that region. Each

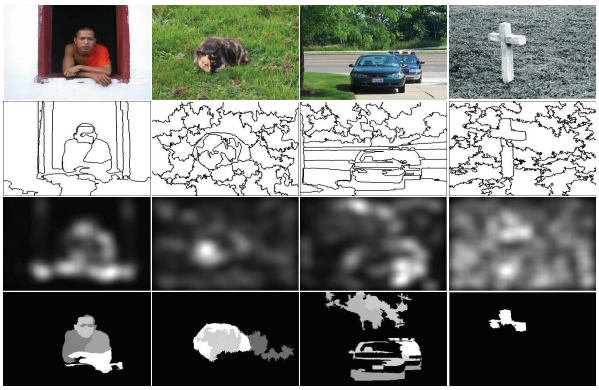

Fig. 2. The top row shows the original images. The segmented images using the mean shift algorithm [14] are displayed in the second row. The third row shows the salient regions generated using QDCT. The rough salient maps shown in the bottom row are generated based on the salient intensities of segmented regions.

pixel $x \in R_k$ is assigned with the average intensity value:

$$x = \sum_{i}^{|R_k|} x_i/|R_k|, \quad \forall x \in R_k, k = 1, ..., K \qquad (1)$$

where $|R_k|$ is the cardinality denoting the total number of pixels in R_k. Fig.2 (the bottom row) shows the obtained rough saliency maps M. It can be seen that by incorporating segmentation results, the outlines of salient objects can be distinguishable in the rough saliency maps.

Given the saliency map M, the foreground and background of image can be separated via a simple thresholding method. In the experiments, the mean QDCT intensity value of the entire image is used as the threshold. However, by examining the rough saliency maps shown in Fig.2, we noted that partial salient objects are neglected because the QDCT-based detection method does not take into account the continuum of the objects, especially in the case of large size of objects or similar color of objects to the background. In addition, the mean shift segmentation algorithm produces an over-segmentation within the objects.

2.3. Improvement of Rough Saliency Maps

To improve the accuracy of the rough saliency maps, Bayesian inference [15] is employed to re-estimate the color saliency of a region by considering the prior distribution of foreground and background colors. For example, the fourth image shown in the bottom row of Fig.2 has a missing bottom of the cross, which was classified as the background. The prior color distribution of foreground $p(F)$ and background $p(B)$ obtained through the rough saliency map M, are used in Bayesian inference. If color of a region is bright, the region more likely

(a) Images (b) GB (c) MT (d) DCT (e) RC (f) SCS (g) SCS-G

Fig. 3. Visual comparison of saliency maps generated from different models. (a) The original images. (b) GB[10]. (c) MT [3]. (d) DCT[11]. (e) RC [9]. (f) SCS. (g) SCS-G.

belongs to foreground, while if the color is dark, the posterior probability of the region belonging to background becomes high.

A color image can be represented as a three-channel image in RGB space. Each channel image contains as many as 256^3 colors. The original RGB color space can be reduced via a quantization method. The colors in each channel are uniformly clustered into 16 colors, making 16^3 colors in total. This 16^3-color space is further normalized into a 336-color space via a nonlinear function [15]:

$$C_k = \lfloor (N_R/16) * 256 \rfloor + \lfloor (N_G/16) * 16 \rfloor + \lfloor N_B/16 \rfloor \quad (2)$$

where $\lfloor \cdot \rfloor$ is a floor function defined as $\lfloor x \rfloor \leq x < \lfloor x \rfloor + 1$, $k = 1, ..., K$ is the index of the segmented regions, and $\{N_R, N_G, N_B\}$ is the normalized 16^3-color space. Given a specified segmented region with color C_k, $k = 1, ..., K$, the probability of the region R_k belonging to the foreground can be computed by the Bayes theorem:

$$p(F|C_k) = \frac{p(C_k|F)p(F)}{p(C_k|B)p(B) + p(C_k|F)p(F)} \quad (3)$$

where $p(F)$ and $p(B)$, representing the prior probabilities of foreground and background, respectively, can be computed by:

$$p(F) = \frac{\phi(x|x \in F)}{\phi(x|x \in F \cup B)} \quad (4)$$

$$p(B) = 1 - p(F) \quad (5)$$

where $\phi(\cdot)$ is a function to count pixel number. $\phi(x|x \in F)$ computes the number of pixels within the foreground F. Similarly, $\phi(x|x \in F \cup B)$ computes the pixel number in both foreground F and background B. $p(C_k|F)$ represents the proportion of color C_k in foreground, which can be defined as:

$$p(C_k|F) = \frac{\phi(x|x \in C_k \cap x \in F)}{\phi(x|x \in F)} \quad (6)$$

Therefore, the rough saliency map M can be improved by refining the probability of this region based on Eq.(3). A new saliency map M_{new} is generated via an iterative refinement process till M_{new} is unchanged. We name this model as *segmentation-based color saliency model* (SCS).

Moreover, according to the center-surrounding property of HVS which suggests that a human being commonly has subconscious mind to focus on the center of an image, a Gaussian filter φ is applied to the QDCT results to enhance the weights of pixels that are located around the center of the salient region, while suppress the weights of pixels far away from the center. φ has the mean μ defined as the center of salient region and the standard deviation σ defined as $1/4$ of the image size. We refer to this model as SCS with Gaussian filter, or shortly SCS-G.

3. EXPERIMENTAL STUDIES

To verify the effectiveness of the new SCS and SCS-G models, we test them on a benchmark dataset *MSRA-1000* [16], which is a subset of Microsoft Research Asia (MSRA) salient object database containing 1000 pictures with human-labeled salient objects or regions. The *MSRA-1000* database has become popular and widely adopted for comparison by many detection methods. The mean shift algorithm was implemented by the *Edge Detection and Image Segmentation* (EDISON) package [17]. The parameter set {*SpatialBandWidth, RangeBandWidth, MinimumRegionArea,*

GradientWindowRadius, MixtureParameter, EdgeStrength-Threshold} defined in [17] was assigned as {*7, 6.5, 2000, 2, 0.3, 0.3*}. We have compared our results with 9 state-of-the-art saliency detection methods that are listed in Table 1.

Fig.3 shows the saliency maps generated using different models. The comparison results demonstrated that both SCS and SCS-G models yielded improved detection performance in terms of clear and accurate outlines of salient objects. For example, the "cross" image presented in the fourth row of Fig.3 has a complete outline of cross due to the refined saliency map generated via the Bayesian inference method. It has been noted that the SCS-G model achieves superior detection results with less interference of noisy background compared to other models.

Quantitative evaluation for saliency detection performance is conducted by using receiver operating characteristic (ROC) analysis. We computed the ratios between the obtained salient region (foreground) to the ground-truth as the true positive (TP) rate and false positive (FP) rate for plotting the ROC curve. The ROC curves of all the detection models are illustrated in Fig.4. It clearly shows that the SCS-G model outperformed all other models. The values of area under curve (AUC) are shown in Table 1. The higher value of AUC indicates better detection performance.

To quantitatively evaluate the influence of segmentation on the detection performance, the SCS-G model was tested under different segmentation degrees by varying the minimum region area (MRA) of the mean shift algorithm [17] to produce various partitions of images. Fig.5 displays average AUC values over the entire database by using $MRA = \{20, 200, 1000, 2000, 3000, 4000, 5000, 6000, 8000, 10000\}$. The trend of the average AUC values shown in Fig.5 demonstrated that the detection performance of the SCS-G model drops down as the images are over-segmented.

4. CONCLUSION AND FUTURE WORKS

In this paper, we presented a novel computational bottom-up saliency model for accurately detecting visual saliency. The new method utilized a mean shift segmentation algorithm and QDCT signatures to generate a rough saliency map in order to separate foreground and background. Bayesian inference was then used to refine the rough saliency map. The experimental results demonstrated that the new model achieves the best performance compared to 9 existing saliency detection models. Moreover, the saliency detection performance of the presented model was evaluated under different degrees of segmentation.

The future work will involve integration of an intelligent prediction method for accurately identifying salient regions to further refine saliency maps, and adoption of a self-adapting Gaussian filter to be applied to the QDCT-based model to improve the detection accuracy.

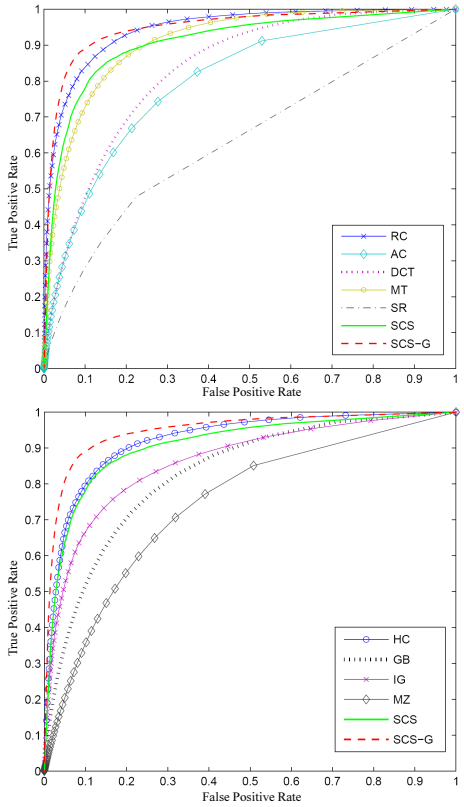

Fig. 4. The ROC curves of the presented models compared to the existing state-of-the-art saliency detection methods.

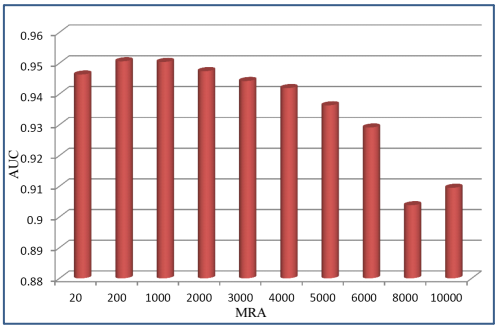

Fig. 5. The AUC values with different degrees of segmentation using the SCS-G model.

Table 1. The comparison of area under curve (AUC) values of different models.

Model	AUC value	Model	AUC value
RC [9]	0.9449	AC [18]	0.8055
HC [9]	0.9226	GB[10]	0.8312
MT [3]	0.9118	MZ [19]	0.7509
IG [16]	0.8644	SR [6]	0.6401
DCT [11]	0.8331	SCS	0.9088
SCS-G	0.9504		

5. REFERENCES

[1] B. Alexe, T. Deselaers, and V. Ferrari, What is an object? *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 73-80, 2010.

[2] T. Judd, K. Ehinger, F. Durand, and A. Tortabla, Learning to predict where humans look, *International Conference on Computer Vision (ICCV)*, pp. 2106-2113, 2009.

[3] C. T. Vu and D.M. Chandler, Main subject detection via adaptive feature selection, *International Conference on Image Processing (ICIP)*, pp. 3101-3104, 2009.

[4] V. Mahadevan and N. Vasconcelos, Saliency-based discriminant tracking, *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1007-1013, 2009.

[5] L. Itti, C. Koch, and E. Niebur, A model of saliency-based visual attention for rapid scene analysis, *Transactions on Pattern Analysis and Machine Intelligence*, pp. 1254-1259, 1998.

[6] X. Hou and L. Zhang, Saliency detection: A spectral residual approach, *Computer Vision and Pattern Recognition (CVPR)*, pp. 1-8, 2007.

[7] A. Tresman and G. Gelade, A feature integration theory of attention, *Cognitive Psychology*, pp. 97-136, 1980.

[8] R. Rensink, Seeing sensing and scrutinizing, *Vision Research*, pp. 1469-87, 2000.

[9] M. M. Cheng, G. X Zhang, N. J. Mitra, X. Huang, and S. M. Hu, Global contrast based salient region detection, *Computer Vision and Pattern Recognition (CVPR)*, pp. 409-416, 2011.

[10] J. Harel, C. Koch, and P. Perona, Graph-based visual saliency, *Neural Information Processing Systems (NIPS)*, pp. 545-552, 2006.

[11] X. Hou, J. Harel, and C. Koch, Image signature: Highlighting sparse salient regions, *Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, pp. 194-201, 2011.

[12] C. Koch and S. Ullman, Shifts in selective visual attention: Towards the underlying neural circuitry, *Human Neurobiology*, pp. 219-227, 1985.

[13] B. Schauerte and R. Stiefelhagen, Predicting human gaze using quaternion DCT image signature saliency and face detection, *Workshop on the Applications of Computer Vision (WACV)*, pp.137-144, 2012.

[14] D. Comaniciu and P. Meer, Mean shift: A robust approach toward feature space analysis, *Pattern Analysis and Machine Intelligence (PAMI)*, pp. 603-619, 2002.

[15] Y. Xie and H. Lu, Visual saliency detection based on Bayesian model, *International Conference on Image Processing (ICIP)*, pp. 645-648, 2011.

[16] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, Frequency-tuned salient region detection, *Computer Vision and Pattern Recognition (CVPR)*, pp. 1597-1604, 2009.

[17] EDISON package, http://coewww.rutgers.edu/riul/research/code.html.

[18] R. Achanta, F. J. Estrada, P. Wils, and S. Susstrunk, Salient region detection and segmentation, *International Conference on Computer Vision Systems (ICVS)*, pp. 66-75, 2008.

[19] Y.-F. Ma and H.-J. Zhang, Contrast-based image attention analysis by using fuzzy growing, *ACM International Conference on Multimedia* , pp. 374-381, 2003.