Journal of Information Hiding and Multimedia Signal Processing ©2015 ISSN 2073-4212
Ubiquitous International Volume **6**, Number **4**, July **2015**

Multiple Structure Based Saliency Detection and Its Application in Image Retrieval

Yi Mao, Bao-Long Guo, Yunyi Yan, and Wei Sun

Department of Space Engineering and Technology
Xidian University
No. 2 TaiBai South Road, XiAn,710071,China
olivia.maoy@gmail.com, baolongguo@xidian.edu, yyyan@xidian.edu, wsun@xidian.edu

Received October, 2014; revised February, 2015

Abstract. *Saliency Detection is a hot research topic in both biological and computer vision. Salient structures, edges, regions would greatly contribute to high-level semantics understanding of people's attention and improve retrieval precision, object detection, edge detection and etc. In this paper, based on the biological principle in visual system, we present a saliency detection system which combines global saliency structure, global saliency in color, local saliency, people/cars detection to give a more comprehensive understanding of people attention system. Besides this, ASSOM is imported to provide feature descriptor of each saliency map for further analysis. Experiments show that the proposed method can provide clearer, accurate saliency detection and improve the precision in content based image retrieval.*

Keywords: Saliency detection; Structure saliency; CBIR, ASSOM.

1. **Introduction.** Human beings have the ability to pay attention to the things around with real-time and efficient, but we do not passively accept all information. Instead, we choose a particular portion of the visual scene for further analysis, and ignore other irrelevant parts automatically. This ability is used to help complete almost all kinds of transaction in daily life, so that in a long time, human beings do not seem aware of the existence of this capability, and direct the "notice things" as a real object or objective understanding. In fact, this seemingly simple process is essentially an extremely complex system in human visual system, which adopts a strategy of calculations based on the characteristics of the images, such as through rapid eye movement scan to select a specific area of the image, move the region to high resolution region in brain for better observation and analysis.

Visual attention mechanism refers to the ability of selecting and processing the most relevant visual scene in visual system. Selective visual attention is one of the most important functions of human's vision system which ensures that the visual system can optimize the input information. A large number of psychological studies show that the human visual attention mechanism is broadly divided into two stages: pre-attention stage and attentive stage. In the pre-attention stage, visual processing nerve process the stimulation signals in parallel to obtain a low -level object representation. In attentive stage, specific areas are extracted from the low level representation of objects for further analysis. The extracted regions are so called salient regional or salient objects. Due to a variety of display screens used by users, especially those small display screens in PDA (Personal Digital Assistant)

[1], mobile phone and the size alterable web page browsers, an adaptive image display scheme which can show attention area firstly is urgently needed.

Human visual system is focused on a scene from a series of different factors underlying the decision. Saliency is a significantly important determining factor which refers to the characteristics that can cause visual attention. In visual attention model, saliency map is used to represent the saliency of a visual area. Saliency map is the same size as the original image, in which each pixel value represents the saliency of the original pixel. It not only represents the saliency areas, but also guide the choice for attestation area through its spatial distribution.

In this paper, based on the biological principle in visual system, we address the possibility of utilizing structure saliency and global saliency(SSGS) to detect saliency area. Besides this, Adaptive-Subspace Self-Organizing Map (ASSOM) is used to extract features from different kinds of saliency maps for further analysis. The system covers both local and global saliency detection methods and provide a new way to find the real operating mechanism in our visual system. This paper is organized as follows. Section 2 presents the related work in this area; Section 3 gives the biological background of the proposed saliency detection system. Section 4 describes the saliency detection methods that employed in our system, including structure saliency map for edge and local-global saliency detection strategy. In Section 5, we show how to combine the saliency maps and get the features using ASSOM. Section 6 presents the experimental results on several data sets. In Section 7, we form our conclusions.

2. **Related Work.** Saliency Detection is a hot research topic in both biological and computer vision area. Ma et al., [2] considered the color contrast for saliency detection. It is computational simplicity, and it might not be work for cases where color was not the most useful feature to detect saliency. Itti [3,18] model combines color, orientation, intensity three channels, employs Gaussian pyramid filter in saliency detection. However, both of these two method are more suitable for small target, edge detection than large target, complex background target detection.

Based on Ma's work, Radhakrishna Achanta et al., [4]uses a contrast determination filter that operates at various scales to generate saliency maps. Since it uses average distance instead of Gaussian distance, it works well when detecting large target in high contrast but not suitable in weak local contrast. [5]fuses the color features into quaternion and employs FFT transformation. By retaining the phase information from the inverse transform, they get the spatial saliency map. However, it only works for small target detection.

Based on [5], [6] adds amplitude information which improves the detection performance significantly. But it costs in computation and cannot be used in real-time analysis so far. Sun et al., [7] computed the grouping-based saliency map, but how to group remains an unsolved problem. Hou et al., [8] proposed a spectral residual approach to compute the saliency map. However, it may overlooked the spatial homogeneity of an object without prior knowledge. [9]shows that these methods tend to overemphasize small and local features, making them less suitable for important applications such as image segmentation, object detection, etc.

Liu et al., [10] use the Gaussian pyramid images in Itti and histogram in calculating local contrast, use color distribution in calculating the global contrast, and use CRFs(Conditional Random Field) integrate them. Due to the use of CRF integration, the importance of each saliency map is not clear. The merged saliency map did not work well in small target detection. [11, 12]are based on finding regions in the image which imply unique frequencies in the Fourier domain. Therefore, they are able to quickly locate visual pop-outs that can serve as candidates for salient objects. These methods are very fast, but since they are based on global considerations, they do not detect object boundaries accurately. [13] divides the image to patches and calculates the similarity between them. Since it keeps the information of the scenario, the saliency object usually is big area. Cheng [9] use color histogram to calculate the difference between regions. Similarly to [4], it works better in strong contrast cases because it employs global contrast. However, due to the use of image segments, saliency cues like spatial distribution cannot be easily formulated.

[14] employs the number of saliency regions, the variance of the saliency regions, and the respective sizes of the most conspicuous saliency regions as the features to pre-classify the query images into attentive class or non-attentive class. [3] proposes using multi-scale image features to create topographical saliency map and use it in rapid scene analysis. Salient edge histogram and salient adjacency graphs are used in Content Based Image Retrieval (CBIR) [15, 19, 20] which suggests that saliency model based on Itti is significantly better than other low-level visual descriptors. But the limitation form Itti model itself made target and background contrast is not high enough for further analysis. Besides this, the spatial relationship between pixels is not included in the model.

[16,17] indicates that without edge information, the saliency map was only a blur map which could only provide the location of the attention, but not exact the scope of the region. In this paper, besides low-level visual descriptors, we employed global structure saliency in edge and global saliency(SSGS) in color space to get the edge information and the relationship between pixels for saliency analysis.

3. **Biological Background.**

Attention is the nexus between cognition and perception. The control of attention may be goal-driven and stimulus driven which corresponds to top-down and bottom-up processes in human perception, respectively. The study of attention involved in a few fields, including biology, psychology, neuron-psychology, cognitive science and computer vision. Although the attention mechanism is not completely understood yet, some proven conclusions can be used to guide its applications. Earlier attention research began with William James, who was the first person to outline a theory of human attention [21]. Successively, Broadbent proposed his filter theory of attention in an attempt to explain many of the existing experimental results [22]. The response selection theory of attention was proposed by Deutsch [23], who indicated that a part of attention involves high level processing which is called late selection in later studies. In 1960s, Treisman proposed a series of models that combined early and late selection into a model known as Feature Integration Theory (FIT) [24]. Treismans recent study believes that early selection is most active when the perceptual load is high, whereas late selection (object-based and location-based) is used when perceptual load is low [25]. Besides, advances in a neurophysiological model of attention were also made by Koch [26, 27].

Human visual attention principles we used in this paper are as follows: 1) Local saliency/low-level considerations, including conspicuous local property such as contrast and color, orientation, etc. [3, 8, 2, 29, 30, 31, 32] can easily attract attention for its distinguishing feature. For example, a red item among green ones immediately attracts attention by virtue of its unique color [28, 33]. 2) Global considerations, which suppress frequently-occurring features, while maintaining features that deviate from the norm [8, 34, 5]. As shown in [36], what are informative are deviations from the norm, and only unexpected input features are signaled to the next stage for process in predictive coding. Contours are more perceptually important than other points related features such as corners. Salient points detected by corner detector may be gathered in small image region in the case of textured or noisy images, resulting in a very local image description [37]. 3) High-level factors: prior knowledge/understanding on the salient object location and object detection, such as faces, cars, texts, etc.

4. **Structure Saliency.**

Certain salient structures can easily capture our immediate attention with the first glance of the image. Saliency network proposed by Ullman [38, 39, 40] is a well-known approach to extract salient curves while filling the gap between segment. It is attractive for the following several reasons. First, the saliency function encourages the long, closed curves while inhibit the short, wiggly ones [41]. Secondly, it offers a locally connected network by simple iterative scheme using the locally connected processing elements. Third, globally salient structures are extracted through local processing with a small number of iterations. Finally, curves are smoothed and gaps are filled in through a by-product of the computations.

Local saliency occurs when an element/pixel becomes prominent by having simple different local property such as color, contrast, orientation, etc. Structure saliency occurs when the structure are salient in a global manner. Even elements/pixels are not salient in local; the arrangement of them may also make the structure salient. Pyramid techniques [42] may not suitable in discovering structure saliency since it assumes that a salient curve is composed of salient sub-parts. The compactness of a structure, the degree of symmetry it contains, properties of the curves that are involved are all good way to evaluate the salient of the structure.

4.1. **Structure Saliency Map for Edge.**

Edge information is a basic feature of images since human eyes are sensitive to edge features for image perception. It contains contour information of the image and can be used to represent the image content, recognize the objects and further for object-based image retrieval [43]. Although some edge detectors such as Canny [44] can filter out part of the background edges of an image, not all the extracted edges derived from edge detector are beneficial to describe the image content. In the case of natural images, humans perceive a textured region as a whole and thus attend more to the texture discontinuities (changes in the textures) instead of the textured region [44]. Some graph theory based methods have been developed to extract salient boundaries from images. Most of the existing methods are mainly based on the well-known Gestalt laws of closure, proximity and continuity [45, 38]. Elder et al., [46] used the shortest-path algorithm to connect fragments to form salient closed boundaries. Wang et al., [47] proposed the ratio cut algorithm which formulated the salient boundary detection problem into a problem for finding an optimal cycle in an undirected graph. Nevertheless, these algorithms are too time-consuming to be used in content based image retrieval.

For each pixel in the image, we define a set of orientation elements connecting the pixel to its neighbors. Orientation element is called ”active” if it lies on the edge of the image, else it is called ”gap”. Given a curve Γ emanating at p_i composed of $textN+$ 1 orientation elements, the saliency of Γ is defined as:

$$\Phi(\Gamma) = \sum_{j=i}^{i+N} \sigma_j \rho_{ij} C_{ij} \tag{1}$$

To each elementp_i, we associate its local saliency σ_i. If p_i is active, σ_i is set to be a positive value, and for gap is set to 0. $\sum \sigma_j$ is a sum of local saliency values depending on the active elements. In order to penalize the gaps, ρ_{ij} ,which is an attenuation function is defined as:

$$\rho_{ij} = \prod_{k=i} j\rho_k = \rho^{g_{ij}} \tag{2}$$

Where$g_{ij}$is the number of gap elements between $p_i$andp_j, which attenuates the contribution of the elements when it is too fragmented.ρdepends on the decay speed along the curves. C_{ij}gives a weight to each local saliency value along the curve which relies on the total curvature of the curve.

$$C_{ij} = e^{-K_ij}, K_ij = \int_{p_i}^{p_j} (\frac{d\theta}{ds})^2 ds \tag{3}$$

$\frac{d\theta}{ds}$ is the curvature at position s. To get a discrete approximation for (3), we denote α_k the orientation between the k' th element and its active element, and Δs the length of an orientation element. Then the total curvature square is approximated by:

$$\frac{2\alpha_k tan\frac{\alpha_k}{2}}{\Delta s} \tag{4}$$

then

$$C_{ij} = \prod_{k=i}^{j-1} f_{k,k+1} \tag{5}$$

where $f_{k,k+1} = e^{\frac{2\alpha_k tan\frac{\alpha_k}{2}}{\Delta s}}$

Let$\varsigma_i$be the set of curves terminating fromp_i. The saliency of$p_i$is defined as the maximum saliency over all curves emanating from it:

$$\Phi(i) = max_{\Gamma\in\varsigma_i}\Phi(\Gamma) \tag{6}$$

$\Phi(i)$is updated by the following computation:

$$\Phi(i)^0 = \sigma_i,\ \Phi(i)^{(n+1)} = \sigma_i + \rho_i max_{p_i\in\delta(i)} C_{ij}\Phi(j)^{(n)} \tag{7}$$

Where$\delta(i)$is the set of all neighbor elements ofp_i. Denote$\Phi_{N(i)}$as the saliency of the most salient curve of length $N+1$emanating fromp_i:

$$\Phi(i)^{(N)} = \sigma_i + \rho_i max_{p_i\in\delta(i)} C_{ij}\Phi(j)^{(N-1)} \tag{8}$$

FIGURE. 1 shows a simple example of for edges, in which (a) and (c) are synthetic image with a fragmented circle/triangle among a background of randomly place and oriented short curves. Structure saliency map (b) and (d) can easily catch the hidden circle/triangle immediately from the structure which is in agreement with our visual system. Furthermore, we conduct experiment on a natural image as shown in (e). Although the corresponding edge map (with Canny operator) is fragmented with noisy background, the hidden structure–cars can also "pop-out" as shown in (g).

4.2. **Global saliency in color space.** [48] incorporated visual attention with seeded region growing to extract the attention objects. However, since finding the best seed areas is crucial issue in region growing, the blurred saliency map could not always provide such reliable information. Fu et al., [49] proposed an iterative object popping-out algorithm to obtain the combined regions with maximum attention value at each iteration step. Nevertheless, the computational complexity is 2^N. It is computationally expensive especially when the number of segments N is large.

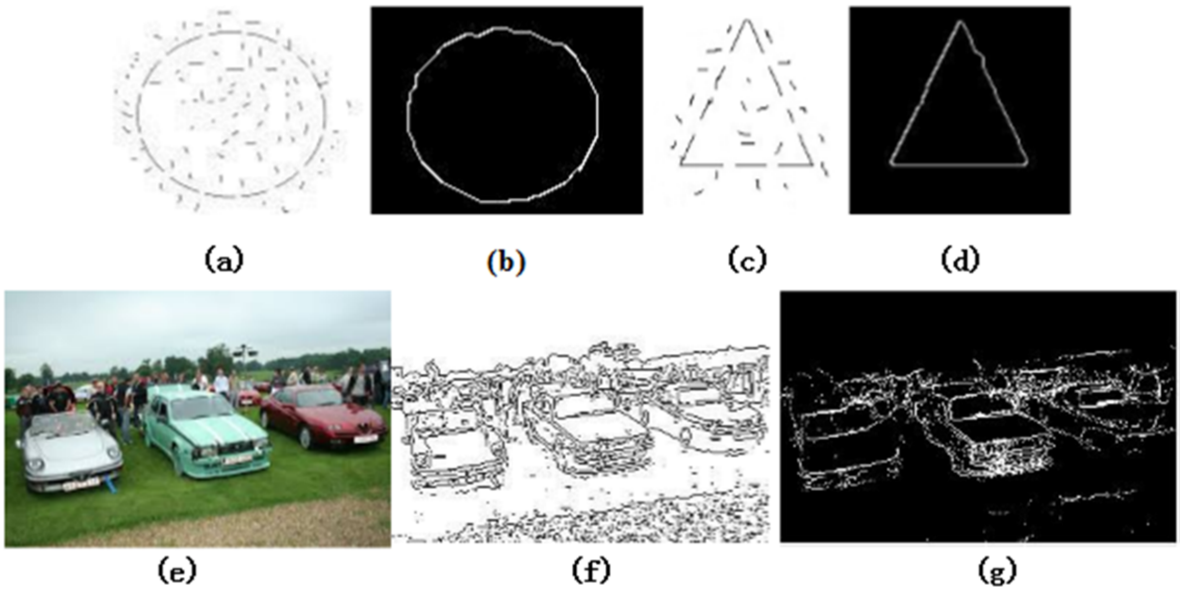

FIGURE 1. Structure saliency map for edge ((a), (c), (e) are original images; (f)is edge of the original image (e); (b), (d), (g) are edge maps of the original images

We consider a single patch of scale r at each pixel. Thus, a pixel i is considered salient if the appearance of the patch o_i centered at pixel i is distinctive with respect to all other image patches. Specifically, let $d_{color}(o_i, o_j)$ be the Euclidean distance patches o_i and o_j in $CIEL*a*b$ color space. Pixel i is considered salient when $d_{color}(o_i, o_j)$ is high when considering all possible j . In practice, to evaluate a patchs uniqueness, there is no need to incorporate its dissimilarity to all other image patches. Usually, consider the K most similar patches is success since if the most similar patches are highly different from o_i, then clearly all image patches are highly different from o_i. Hence, for every patch o_i, we search for the K most similar patches $q_k(k = 1, \cdots, K)$. A pixel i is salient when $d_{color}(o_i, q_k)$ is high for all q_k(in all our experiments). Also, distance between the patches is another important factor in saliency. Patch o_i is salient when the patches similar to it are nearby, and it is less salient when the resembling patches are far away. Let $d_{position}(o_i, q_k)$ be the Euclidean distance between the positions of patches o_i and q_k, then the dissimilarity measure between o_i and q_k is defined as:

$$d(o_i, q_k) = \frac{d_{color}(o_i,\ q_k)}{1 + c * d_{position}(o_i,\ q_k)} \tag{9}$$

This dissimilarity measure is proportional to the difference in appearance and inverse proportional to the positional distance. The saliency value of pixel i is defined as:

$$\phi(i) = 1 - exp\{-\frac{1}{K}\sum_{k=1}^{k=K} d(o_i,\ q_k)\} \tag{10}$$

Similarly as in [3], different salient maps are combined as follows:

$$\phi(i) = \sum_k w_k \phi_k(i) \tag{11}$$

$\phi_k(i)$represents the saliency value as shown in (8),(10). $w_k$is the corresponding coefficient that can be tuned. Since we set it to be[0.2,.0.4,0.4].

We simulate this visual contextual effect in two steps. First, the most attended localized areas at each scale are extracted from the saliency maps produced by Equation (11). A pixel is considered attended if its saliency value exceeds a certain threshold. Then, each pixel outside the attended areas is weighted according to its Euclidean distance to the closest attended pixel. Let $d_{foci}(i)$ be the Euclidean positional distance between pixel i and the closest focus of attention pixel. The saliency of pixel is redefined as:

$$S(i) = \phi(i)(1 - d_{foci}(i)) \tag{12}$$

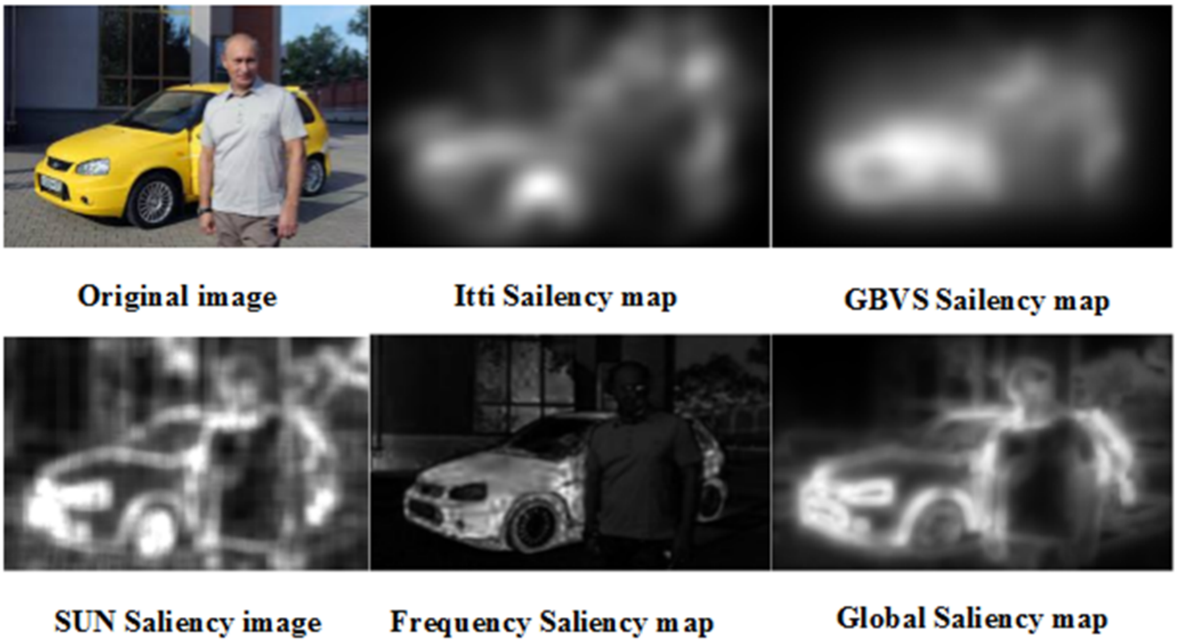

Figure 2. Global saliency in color space

FIGURE. 2 compares the results of global saliency with local contrast approach: Itti [3], GBVS [31]; global approach such as: frequency-tuned saliency detection [4]; SUN-saliency using natural statistics [11]. We could find out that compared to other methods, the global saliency detection method mentioned in out paper can provide a clearer, sharply focused saliency map. A saliency map based on the distance of each pixel to the center of the image provides a better prediction of the salient object than many previous saliency models. [50, 11] indicate a strong bias for human fixations to be near the center of the image. Inspired by this we further incorporate a center prior to our saliency estimation. Let $G(\sigma_x, \sigma_y)$ be a two-dimensional Gaussian positioned at the center of the image, which centered in box detected as shown in FIGURE. 3.

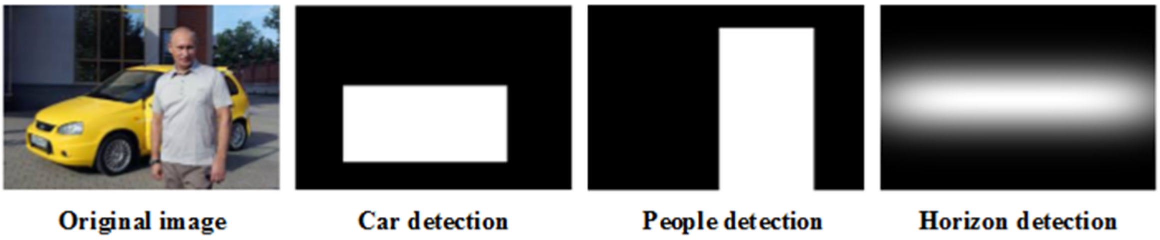

Figure 3. High Level Features

saliency of pixel i is updated as:

$$\hat{S}(i) = S(i)G_1(i)G_2(i) \tag{13}$$

4.3. **ASSOM: Adaptive-Subspace Self-Organizing Map.** Unlike existing methods exclusively using global features or local features for image retrieval, we fuse edge, local, global, high level features (as shown in FIGURE. 3) into an integrated comprehensive feature for more accurate image retrieval by ASSOM. OM neural network firstly is proposed by T.Kohonen in 1981 [51], which is composed of two connected layer: input layer and competitive layer. It can map any arbitrary input patter into a two-dimensional discrete graphics while keep the structure. The basic idea is: for each input, only some weights need to be adjusted, the adjustment goal is to make the weight vector closer or far further from the former ones. When the input vector input to the self-organizing neural networks, network randomly selects weights for computing, and then finds the winning neuron. Through the effects of Feedback, the small area near the winning neuron will be excited, and the connections weights within the region will be adjusted towards to the more competitive orientation. The repeated learning to the input makes probability density distribution of the connection between weight vector and the input consistent, which means, the connection weights reflect the spatial distribution of the statistical characteristics of the input. ASSOM [52] is a useful tool for invariant feature generation and visualization, which is an alternative to the standard Principal Component Analysis (PCA) method of feature extraction.

Basically, ASSOM is a combination of SOM and the subspace method. The single weight vectors at map units in the SOM are replaced by sets of basis vectors that span some linear subspaces in the ASSOM. By setting filters to correspond to pattern subspaces, some transformation groups, such as translation, rotation and scaling, can imported. By using the spatial interactions between processing units of the network, ASSOM can learn topologically ordered filters corresponding to feature subspaces. It has been widely applied to speech processing [53], texture segmentation [54], hand-written digit recognition [55] and retrieving faces [56], and etc. The ASSOM will be trained to generate the appropriate feature filters based on the patches around salient points, which are supposed to carry essential information for image description and consequently for classification. Let p_k is the salient patches detected from the image I. p_k then is fed into ASSOM which is trained in all categories of images in the training set. For each patch p_k , the module i generates energy $\hat{p}_k L_i$, with $\hat{p}_k L_i$ being the orthogonal projection of p_k on the subspace L_i of the module . Energies generated by all the modules construct an activity map as a vector:

$$A_k = [\hat{p}_k L_1, \cdots, \hat{p}_k L_i, \cdots, \hat{p}_k L_{|I|}] \tag{14}$$

$|I|$ is the number of modules in ASSOM. Activity maps of all patches extracted from I are then accumulated to form a cumulative activity map C_I:

$$C_I = \sum_{k=1}^{K} A_k \tag{15}$$

5. **Experiments.** Our system will ultimately combine different kinds of saliency maps into a single feature space. In order to optimize the parameters of the system and compare it to other systems, we need a methodology for judging the quality of the proposed combination method. First, we select HFT public test database: Saliency database for saliency comparison since previous database as Bruce's dataset, Hou's dataset, Harel's dataset is simply a collection of pictures. Saliency database contains 235 images of natural

Table 1. AUC of compared algorithms

Method	Large target	Medium target	Small target	Cluttered back-ground	Repeating dis-tracters	Different size	All
ITTI	0.8808	0.9004	0.9328	0.8156	0.8667	0.9344	0.8994
SUN	0.8212	0.8455	0.8843	0.6992	0.8064	0.8777	0.8402
AC	0.9125	0.8904	0.8292	0.8481	0.9268	0.9144	0.8808
IG	0.9347	0.923	0.9204	0.9393	0.9641	0.9362	0.9322
GBVS	0.8716	0.9016	0.9474	0.8229	0.8677	0.9373	0.9012
MZ	0.8316	0.8804	0.9508	0.7661	0.8096	0.8517	0.8747
RC	0.9264	0.7973	0.6902	0.9203	0.89036	0.8429	0.8132
SSGS	0.9535	0.9382	0.9499	0.9565	0.9594	0.9581	0.9446

scenes, with size of 80 pixels * 640 pixels, which are all from Google search and recent articles. It also provides each picture corresponding manual calibration saliency map. The dataset can use in all kinds of natural scenes significant target for testing. Comparative experiments use AUC values (The Area under the Receiver Operating Characteristic, ROC curve area, ROC curve is a graphical plot which illustrates the performance of a binary classifier system as its discrimination threshold is varied. It is created by plotting the fraction of true positives out of the total actual positives (TPR = true positive rate) vs. the fraction of false positives out of the total actual negatives (FPR = false positive rate), at various threshold settings), the closer to 1 the value, the better the performance of the algorithm. Table 1 is a comparison of AUC in different methods. ITTI [3], SUN [11], AC [4], IG [57], GBVS [31], MZ [2], RC [9] are used as the comparison. We could find out that IG is good at finding repeating distracters while MZ performs better in small target detection. It shows that our proposed combination of structure saliency and global saliency outperforms almost all algorithms in this dataset.

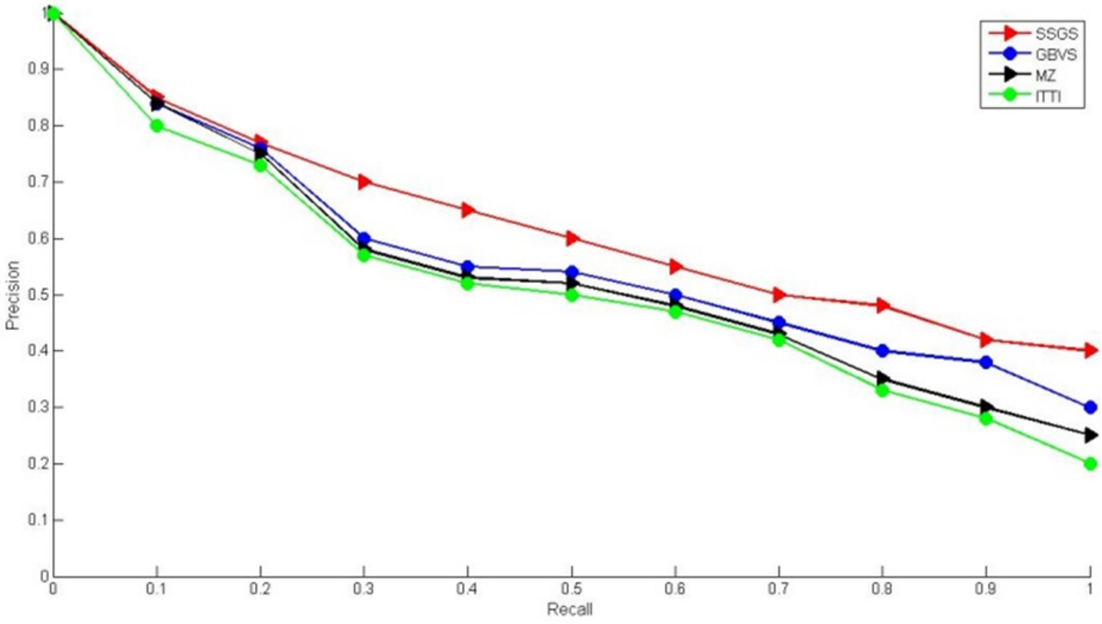

Figure 4. Comparison result of PR curves

Then we formulate saliency detection as a classification problem and apply precision-recall framework using Caltech 101 classified dataset as the ground truth. Caltech data set consists of a total of 9146 images, split between 101 different categories, cars, owners, as well as an additional background/clutter category. FIGURE. 4 shows precision-recall comparison between four different methods: SSGS that is proposed in our paper; GBVS [31], MZ [2], ITTI [3]. Form the results, we could find compared to other salient based method, SSGS can offer a much better classification results.

6. **Conclusions.** In this paper, we proposed combining structure saliency and global saliency(SSGS) to detect saliency area based on the biological principle in visual system. Adaptive-Subspace Self-Organizing Map (ASSOM) is used to extract features from different kinds of saliency maps for image retrieval. This new system covers both local and global saliency detection methods and provide a new way to find the real operating mechanism in our visual system. The proposed framework is good in precision. However, it still needs to be improved. In the future, we would like to make further study in biological principle in visual system and find better, faster model in saliency detection.

Acknowledgment. This work is partially supported by national Natural Science Foundation of China(61105066, 61201290, 61305041, 61305040).The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] X. Fan W.Y. Ma H.J. Zhang L.Q. Chen, X. Xie and H.Q. Zhou, A visual attention model for adapting images on small displays, *ACM Multimedia Systems Journal*, 2003.

[2] H.Zhang Y.Ma, Contrast-based image attention analysis by using fuzzy growing, *ACM Multi-media*, pp. 374-381, 2003.

[3] Ernst Niebur Laurent Itti, Christof Koch, A model of saliency-based visual attention for rapid scene analysis, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp.1254-1259, 1998.

[4] P. Wils, R. Achanta, Francisco Estrada and Sabine Susstrunk, *Salient region detection and segmentation, Proceedings of International Conference on Computer Vision System*,pp. 66-75, 2008.

[5] C. Guo and L. Zhang, A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression, *Image Processing, IEEE Trans.*, vol. 19, no.1, pp.185-198, 2010.

[6] J. Li, M. D. Levine, X. An, X. Xu, H. He, Visual saliency based on scale-space analysis in the frequency domain, *Pattern Analysis and Machine Intelligence, IEEE Trans*, vol. 35, no.4, pp.996-1010, 2013.

[7] R. Fisher, Y. Sun, Object-based visual attention for computer vision, *Articial Intelligence*, no. 146, pp. 77-123, 2003.

[8] X. Hou and L. Zhang, Saliency detection: A spectral residual approach, *CVPR*,pp. 18, 2007.

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

[10] T. Liu, J. Sun, N. N. Zheng, X. Tang, and H. Y. Shum, Learning to detect a salient object, *In Computer Vision and Pattern Recognition, IEEE Conference on CVPR*, pp. 1-8, 2007.

[11] TK. Marks, H. Shan, GW. Cottrell, L. Zhang, MH. Tong, Sun: A bayesian framework for saliency using natural statistics, *Journal of Vision*, vol. 8, no. 7, pp. 1-20, 2008.

[12] N. Bruce and J. Tsotsos, Saliency based on information maximization, *Advances in neural information processing systems*, no. 18, pp. 155-167, 2006.

[13] A. Tal, S. Goferman, Z. M. Lihi, Context-aware saliency detection. CVPR, 2010.

[14] Z. Chi, D. Feng, Z. Liang, H. Fu, Image pre-classication based on saliency map for image retrieval, *ICICS*, 2009.

[15] X. Yang, S. Feng, D. Xu, Attention-driven salient edges and regions extraction with applification to cbir, *Signal Processing*, vol. 90, no. 1, pp. 1-15, 2010.

[16] W. Ma, D. Rajan, L. Chia, Y. Hu, X. Xie, Salient object extraction combining visual attention and edge information, *Technical Report*, 2004.

[17] T. Judd, K. Ehinger, F. Durand, A. Torralba, Learning to predict where humans look, Computer Vision, IEEE 12th international conference, pp. 2106-2113, 2009.

[18] A. Borji and L. Itti, State-of-the-art in visual attention modeling, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 185-207, 2013.

[19] D. N. Sihite, A. Borji and L. Itti, Salient object detection: A benchmark, ECCV, 2012.

[20] D. Sihite, A. Borji and L. Itti. Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study, *IEEE TIP*, 2012.

[21] W. James, The principles of psychology, Harvard University Press, 1890.

[22] D.E. Broadbent, Perception and communication, *Pergamon Press, Oxford*, 1958.

[23] D. Deutsch, J. Deutsch, Attetnion: Some theoretical considerations, *Psychological Review*, no.70, pp. 80-90, 1963.

[24] S. Gormican A. Treisman. Feature analysis in early vision: evidence from search asymmetries. Psychology Review, 95:15-48, 1988.

[25] A. Treisman, Perception of features and objects, *Visual Attention*, 1998.

[26] S. Ullman, C. Koch, Shifts in selective visual attention: towards the underlying neural circuitry, Human Neurobiology, no. 4, pp. 219-227, 1985.

[27] C. Koch, F. Crik, Some reflections on visual awareness, *Proceedings of the Cold Spring Harbor Symposia on Quantitative Biology*, 1990.

[28] A. Treisman and G. Gelade, A feature-integration theory of attention, Cognitive Psychology, vol. 12, no. 1, pp. 97136, 1980.

[29] L. Itti and C. Koch, Computational modelling of visual attention, *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194-204, 2001.

[30] D. Walther and C. Koch, Modeling attention to salient proto-objects, Neural Networks, vol. 19, no. 9, pp. 1395-1407, 2006.

[31] C. Koch, J. Harel and P. Perona, Graph-based visual saliency, *Advances in neural information processing systems*, no. 19, pp. 545-560, 2007.

[32] D. Barba, O. L. Meur, P. L. Callet and D. Thoreau, A coherent computational approach to model bottom-up visual attention, *PAMI*, vol. 28, no.5, pp. 802-817, 2006.

[33] B Julesz, Textons, the elements of texture perception and their interactions, *Nature*, no. 290, pp. 91-97, 1981.

[34] J. Wolfe, Guided search 2. 0. a revised model of visual search, *Psychonomic bulletin & review*, vol. 1, no. 2, pp. 202-238, 1994.

[35] Q. Ma, C. Guo and L. Zhang, Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform, *CVPR*, pp. 18, 2008.

[36] C. Koch, and T. Poggio, Predicting the visual world: silence is golden, *Nature Neu-roscience*, no. 2, pp.9-10, 1999.

[37] C. Laurent, N. Laurent, M. Maurizot, and T. Dorval, In depth analysis and evaluation of saliency-based color image indexing methods using wavelet salient features, *Multimedia Tools Appl*, vol. 31, no. 1, pp. 73-94, 2006.

[38] A. Sha'ashua and S. Ullman, Structure saliency: The detection of globally salient structures using a locally connected network, *ICCV, Weizmann Institute of Science Report*, pp. 321-327, 1988.

[39] S. Ullman, Filling-in the gaps: the shape of subjective contours and a model for their generation, Biological Cybernetics, no. 25, pp. 1-6, 1976.

[40] T.D. Alter and R. Basri, Extracting salient curves from images: An analysis of the saliency network, International Journal of Computer Vision, vol. 27, no. 1, pp. 51-69, 1998.

[41] I. Kovacs and B. Julesz. A closed curve is much more than an incomplete one: effect of closure in gure-groun segmentation, *Proc. Nat'l Academy of Sciences*, U.S.A., vol. 90, no. 16, pp. 7495-7497, 1993.

[42] A. Rosenfeld, Some pyramid techniques for image segmentation, *CS-TR-1664*, University of Maryland, 1986.

[43] T. Huang, X. Zhou, Edge-based structural features for content-based image retrieval, *Pattern Recognition Letters*, vol. 22, no. 5, pp. 457-468, 2001.

[44] E. Erdem and A. Erdem, Visual saliency estimation by nonlinearly integrating features using region covariances, *Journal of Vision*, vol. 13, no. 4, pp. 1-20, 2013.

[45] G Humphrey, The psychology of the gestalt, Journal of Educational Psychology, vol. 15, no. 7, pp. 401-412, 1924.

[46] S. Zucker, J. Elder, Computing contour closure, *European Conference on Computer Vision*, pp. 399-412, 1996.

[47] J. Siskind, J. Wang, S. Wang, T. Kubota, Salient closed boundary extraction with ratio contour, *IEEE Trans. on Pattern Analysis and Machine Intelligence*,vol. 27, no. 4, pp. 546-561, 2005.

[48] M. J. Li, H. J. Zhang ,J. Han, K. N. Ngan, Unsupervised extraction of visual attention objects in color images, *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 16, no. 1, pp.141-145, 2006.

[49] D. Feng, H. Fu, Z. Chi, Attention-driven image interpretation with application to image retrieval, *Pattern Recognition*, vol. 39, no.9, pp. 1604-1621, 2006.

[50] B. W. Tatler, The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions, *Journal of Vision*, vol. 7, no. 14, pp. 117, 2007.

[51] T. Kohonen, Self-organized formation of topologically correct feature maps, Biological Cybernetics, vol. 43, no. 1, pp.59-69, 1982.

[52] T. Kohonen, The adaptive-subspace som (assom) and its use for the implementation of invariant feature detection, *Artificial Neural Networks*, no. 1, pp. 3-10, 1995.

[53] H. Matsuyama, H. Tokutaka, H. Kishida, S. Hase, Speech signal processing using adaptive subspace som (assom), *Technical Report NC95-140, The Inst. of Electronics, Information and Communication Engineers*, Tottori University, Koyama, Japan, 1996.

[54] J. Ruiz-del, Solarm, Texsom: Texture segmentation using self-organizing maps, *Neurocomputing*, vol. 21, no. 2, pp. 7-18, 1998.

[55] B. Zhang, M. Fu, H. Yan, and M. A. Jabri, Handwritten digit recognition by adaptive-subspace self-organizing map (assom), *Neural Networks, IEEE Trans.*, vol. 10, no. 4, pp.939-945, 1999.

[56] Z. Q. Liu, Retrieving faces using adaptive subspace self-organising map, *International Symposium on Intelligent Multimedia, Video and Speech Processing*, pp. 377-380, 2001.

[57] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, Frequency-tuned salient region detection, *In Computer Vision and Pattern Recognition, IEEE Conference*, pp. 1597-1604, 2009.