Global salient information maximization for saliency detection

Wang Luo [a], Hongliang Li [a,*], Guanghui Liu [a], King Ngi Ngan [b]

[a] *School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu, China*
[b] *Department of Electronic Engineering, Chinese University of Hong Kong, Shatin, N.T. Hong Kong, China*

A R T I C L E   I N F O

A B S T R A C T

In this paper, a new method for saliency detection is proposed. Based on the defined features of the salient object, we solve the problem of saliency detection from three aspects. Firstly, from the view of global information, we partition the image into two clusters, namely, salient component and background component by employing Principal Component Analysis (PCA) and k-means clustering. Secondly, the maximal salient information is applied to find the position of saliency and eliminate the noise. Thirdly, we enhance the saliency for the salient regions while weaken the background regions. Finally, the saliency map is obtained based on these aspects. Experimental results show that the proposed method achieves better results than the state of the art methods. And this method can be applied for graph based salient object segmentation.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Humans and other animals cannot pay attention to more than one or very few items simultaneously. They only could facilitate learning and survival by enabling organisms to focus their limited perceptual and cognitive resources on the most pertinent subset of the available sensory data. This subset can be called as saliency of the sensory data which is the state or quality by which it stands out relative to its neighbors [1], such as an object, a person, a pixel. Usually, the saliency of the sensory data represents the most valuable information from a large amount of the sensory data in the fields of computer vision, design, graphics, and human computer interaction. For computer vision, substantial progresses [2,3] have been made in the psychophysics of visual attention and many computational models [4–24] of visual attention have been proposed in the early years. Visual attention is useful for some applications, such as complex scenes understanding [4,6–11], object detection [20], tracking

[17,19], retargeting [21], and recognition [24]. The salient object detection is one of the most popular applications of visual attention.

Most models of visual attention for saliency detection are biologically inspired and based on a bottom-up computational model [25]. They are usually used to detect the approximate position of the salient region or object, which can be classified as four categories of models of visual attention.

The first is about methods with graph representation. In the early report [26], random walks are employed on image lattice to compute the visual saliency. Harel et al. [18] extend this method by proposing a better dissimilarity measure to model the transition probability between two vertices. Wang et al. [27] utilize fully-connected graph structure to simulate the cortical neuron connection, and yield a new visual saliency measure called Site Entropy Rate.

The second category is about the methods based on information maximization. In these methods, the information is considered as the driving force behind attentive sampling, the visual saliency is measured using the rarity of feature. Bruce et al. [17] employ the self-information of sparse features to measure saliency. Mancas et al. [22] describe a rarity-based visual attention model to provide

---

an approximation of human perception by visualizing its gradual discovery of the visual environment. It is based on the theory of self-information. Hou et al. [19] introduce the incremental coding length (ICL) to allocate different amount of energy to features according to their rarity under the assumption that salient feature can offer entropy gain. However, the obtained salient information is usually local information which has its limits to detect saliency regions accurately. And the obtained saliency map always represents the approximate position of the salient regions or objects, while lot of salient information is ignored.

The third category is about the methods for the center-surround mechanism. These methods model the center-surround mechanism of primary visual cortical cells. A biological-plausible visual saliency model [20] is proposed, which implies that visual perception relies on a linear measure of similarity on intensity, color, and orientation. However, Gao et al. [28,29] find that this is in conflict with the well known properties of high level human judgments of similarity. Therefore, Gao et al. propose a discriminant center-surround hypothesis using mutual information, and it can provide optimal solutions for many other saliency problems for computer vision.

The fourth category is about the methods based on machine learning. Liu et al. [30] combine pixel-based saliency measurements in a CRF and derives a binary segmentation separating the object from the background. Alexe et al. [31] design an objectness measure under a Bayesian framework and explicitly training it to distinguish windows containing an object from background windows.

Recently, a new type of method [21] is proposed for context-aware saliency detection. This method aims at detecting the image regions which represent the scene. It is different from previous definitions whose goal is to either identify fixation points or detect the domain object. Cheng et al. [13] propose a histogram-based contrast method (HC) to measure saliency. HC-maps assign pixel-wise saliency values based simply on color separation from all other image pixels to produce full resolution saliency maps. They also incorporate spatial relations to produce region-based contrast (RC) maps to improve the HC-maps.

From the above analysis, we can see that most of the methods focus on obtaining the saliency map which is made up of parts of the salient region or object. The context-aware method [21] can obtain the saliency map which describes the whole salient region or object. However, it also introduces a lot of false alarms. Therefore, the challenge of saliency detection is to acquire the salient region or object as accurate as possible, not only approximate position of the saliency. Solving this problem is one of the most important goals for scene understanding, which involves interpreting the whole image by recognizing all the objects of interest within an image and their spatial extent or shape.

In this paper, we propose a novel method to saliency detection. Firstly, a definition about the features of the salient objects is introduced. Secondly, we present the method of global salient information maximization which can obtain the salient information from three aspects. (1) Detect global salient component by PCA based method;

(2) extract maximal salient information to eliminate the noise of saliency detection; (3) enhance the salient information. Finally, the optimal saliency map is obtained.

This paper is organized into six sections. The next section introduces features of the salient object used in our paper. Section 3 describes the proposed theory that finds the salient objects. Section 4 provides some experimental results of the proposed approach, and comparison with 11 previous works about saliency detection is also performed in qualitative and quantitative measures. In Section 5, we apply the proposed saliency detection method to graph based image segmentation. Finally, Section 6 draws the conclusion of this paper.

2. The feature of salient object

The goal of our method is to detect salient information. The previous works briefly focused on the field of searching for the approximate position of the saliency. For that reason that neither the specific position nor the size of the salient object is known after the saliency detection, therefore, it is a challenge for these methods to be implemented to realize the accurate saliency detection. There are different applications [4,32] for the saliency detection methods, and there are many different definitions for the salient region or object. In this paper, supported by psychological evidence [33,34], our method can address the above defects to accurately find the salient region or object which possesses three features as follows:

(a) Local low-level factors are considered. The salient region or object should have distinct color, contrast, shape and/or texture in complex scenes.
(b) Global features, such as the size of the salient region or object, are small. The principal component of an image should be the background, therefore, the size of the salient object is smaller than the background comparably.
(c) According to the organization rules, the visual forms may possess one or several centers. The coordinate of the saliency point should be located on the salient object.

Unlike the previous works, in this paper, we aim to extract salient region or object as accurately as possible. In order to achieve this we propose a novel saliency detection method with respect to the three pre-defined features of the salient object, which is based on a method of global salient information maximization.

3. The proposed saliency detection method

In this section we propose an algorithm to realize the extraction of the salient region or object with the features (a)–(c). The PCA based method, which is used to extract the principal component (i.e. background) and minor component (salient objects) of the image, can realize the extraction of the salient object with features (a) and (b). The salient object with features (c) can be refined by the information maximization based method which can be
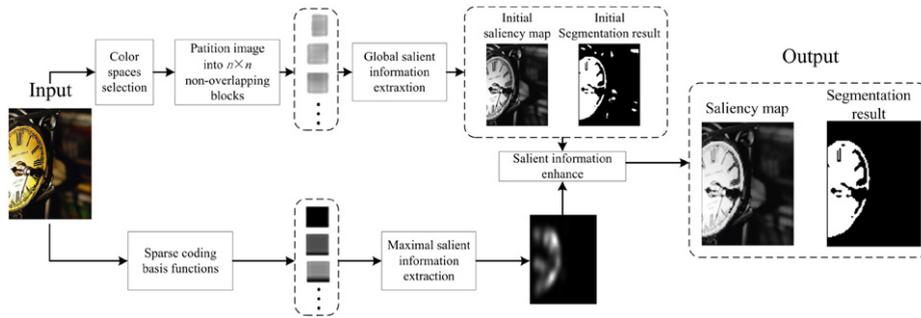
Fig. 1. Flow chart of our proposed approach.

used to eliminate the noise of saliency detection. The salient information enhancement method can be used to enhance the saliency of the salient object with features (a)–(c). The combination of these methods provides the feasibility of detecting the salient object involving the features (a)–(c).

The flow chart of our proposed method is illustrated in Fig. 1. We can obtain the salient information about the clock using the proposed method which mainly contains three steps as follows.

First, we transform the original image to L*a*b and YCbCr color spaces, and partition the image of each color channel into $n \times n$ non-overlapping blocks. The eigenvectors space is created by the orthonormal eigenvectors which are extracted through PCA of this block set. The salient component of each image is obtained by partitioning the feature vector space into two clusters using k-means clustering. The minimum circumscribed rectangular is used to extract the saliency map from these color channels. As shown in Fig. 1, the initial saliency map and segmentation result are obtained in this step.

Second, we represent the image patch as a linear combination of sparse coding basis functions which are referred as features. Then we estimate the position of the salient object by estimating the maximal salient information. The noise introduced from the first step can be eliminated according to the position of the maximal salient information.

Finally, a method of salient information enhancement is proposed to enhance the saliency regions while weaken the background. From Fig. 1, it can be clearly seen that the saliency map and the segmentation result are obtained fully and accurately while the uninteresting background is removed.

Next, we will elaborate the proposed method step by step.

3.1. Global salient information (GSI) extraction

From the view of cognitive science, the image can be decomposed into two parts, i.e., the redundant part and the salient part. Where, the redundant part represents the information with high regularities, and the salient part denotes the novelty part. In this sense, it is known that the part which has low intrinsic dimensionality will alleviate the curse of dimensionality. In other words, the redundant part should be low-rank, and the salient part corresponds to a small number of objects. Intuitively, this

property suggests the distribution of salient part tend to be sparse. We can interpret the problem of saliency detection as a sparse component recovery problem. In this paper, PCA is employed to solve this problem.

PCA is one of the simplest of the true eigenvector-based multivariate analyses methods [36]. It involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principle components. In particular, the first principal component accounts for as much of the variability in the input data as possible, and each of the succeeding component accounts for as much of the remaining variability as possible. In the context of visual attention, the intuition behind the principle component is straight forward: the first principal component corresponds to the redundant (background) data set, whereas, the minor component corresponds to the salient data set. From this analysis, we can employ PCA to acquire the salient regions.

In order to extract the salient region, we aim to use the clustering algorithm to cluster the feature vector space which is obtained by PCA. The early reports [37,38] have revealed the relationship between PCA and k-means clustering. The relaxed solution of k-means clustering, specified by the cluster indicators, is given by the PCA principal components. The subspace of the PCA spanned by the principal directions is identical to the cluster centroid subspace specified by the between-class scatter matrix. Therefore, PCA can automatically project to the subspace where the global solution of k-means clustering lie, then facilitate k-means clustering to find near-optimal solutions.

From the above analysis, in order to obtain the optimal solution of clustering, we can use k-means clustering to cluster the feature vector space which is obtained by PCA to two categories, i.e., salient and background components.

Fig. 2 illustrates the flow chart of global salient information extraction, which mainly consists of three steps, i.e., color spaces selection, feature vector space creation and clustering, and decision, which will be introduced in detail as follows.

3.1.1. Color spaces selection

We first obtain the image of each color channel in the L*a*b and YCbCr color spaces. Experimentally we observed worse results of saliency detection by using PCA and k-means clustering to the RGB image. It is due to the property of RGB color that it does not well approximate
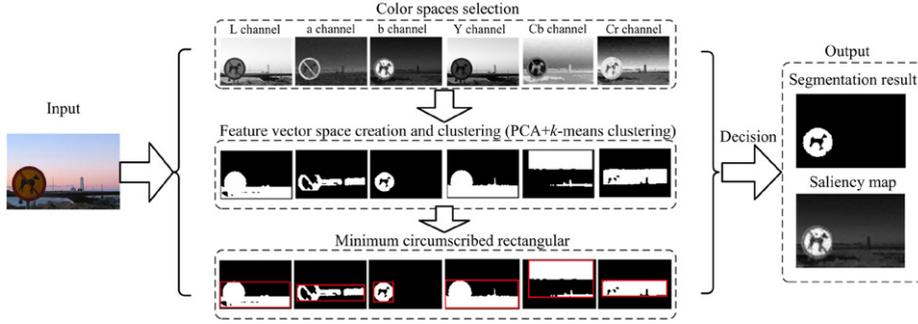
Fig. 2. Flow chart of global salient information extraction.

human vision. We find an interesting phenomenon that salient objects or regions are always very salient in one or several images of the six color channels (i.e., L, a, b, Y, Cb, and Cr channels) by a mount of experiments. It is due to the property of the L*a*b and YCbCr color spaces that they are designed to accurately approximate human vision. Therefore, we separate the L*a*b and YCbCr color spaces to obtain six color channels. Next, we will process each color channel independently, and combine the segmentation results of the six color channels to help to decision.

3.1.2. Feature vector space creation and clustering

To obtain the segmentation results for each channel of the L*a*b and YCbCr color spaces, we first partition each channel X of size $I \times J$ into $n \times n$ non-overlapping blocks and transform them into vector $x(p,q)$. The set of vector $x(p,q)$ is used to create an eigenvector space using PCA. The average vector of the set is calculated as

$$\mu = \frac{1}{M} \sum_{t=1}^{M} X_t \tag{1}$$

where $M = \lfloor (I \times J)/(n \times n) \rfloor$, $\lfloor \rfloor$ denotes a mathematical floor operator which rounds a number down to the nearest integer. p and q represent an index for the vector $x(p,q)$.

In order to describe the distribution of the data better, the PCA is applied on the difference vector set $\Delta_t = x_t - \mu$ to seek two data, i.e., a set of orthonormal vector e_l (i.e. eigenvectors) and their associated scalars λ_l (i.e., eigenvalues) of the covariance matrix $C = (1/M) \sum_{t=1}^{M} \Delta_t \Delta_t^T$. The feature vector space is created by projecting $x(i,j)$ to eigenvector space for each pixel located at (i,j), i.e.

$$v(i,j) = [v_1, v_2, \ldots, v_L]^T \tag{2}$$

where $1 \leq L \leq n^2$ and $v_l = e_l^T(x(i,j) - \mu)$, $1 \leq l \leq L$.

The method of k-means clustering is then used to cluster the feature vector space to two categories, i.e., the salient component ω_s and the background component ω_b. Let $v_{\omega s}$ and $v_{\omega b}$ be cluster mean feature vectors for classes ω_s and ω_b. According to the features of salient object, the area of salient region is smaller than background region. Therefore, the region whose area is larger in the clustering result is assigned as the ω_b class, while the other region is assigned as ω_s class. Then, a binary clustering result can be created with respect to $v_{\omega s}$ and $v_{\omega b}$, in which "1"

indicates that the corresponding pixel location involves a saliency region, whereas "0" involving background region. For the six color channels, we can obtain six clustering results which are shown in the middle of Fig. 2.

3.1.3. Decision

With the six clustering results, we start to decide which image should be selected to represent the result of global salient information extraction. According to the feature of the salient object, the area of salient regions A_{region} should be lower than the area of background. In this paper, we set a threshold T_A ($=0.4$), and assume that the ratio of the area of salient regions A_{region} to the area of original image is larger than the threshold T_A. Moreover, according to the features of the salient object, the salient object is small and concentrate, which can be approximated to a rectangular. Therefore, we use the ratio R_d of two values (i.e., the area of the salient regions A_{region}, and the area of the minimum circumscribed rectangular $A_{\text{rectangular}}$ for the salient regions) to help to decision

$$R_d = \frac{A_{\text{region}}}{(A_{\text{rectangular}})^{(1+\gamma)}} \tag{3}$$

where γ is set to 0.5 in this paper. The segmentation result which has the highest values of R_d will be selected as the final saliency segmentation result. Note that noise usually exists in the image, which would affect the generation of minimum circumscribed rectangular. Therefore, we choose the minimum rectangular which covers 80% pixels of the salient regions to represent the minimum circumscribed rectangular. The color channel corresponding to the selected image is used to represent the initial saliency map.

From Fig. 2, we can see that there are six different clustering results obtained from PCA and k-means clustering. After decision, we obtain the segmentation result of GSI, and the saliency map is represented by the corresponding b channel.

3.2. Global salient information maximization (GSIM)

To improve the result of global salient information extraction, we eliminate the noise by incorporating the maximal salient information. We also propose the method of salient information enhancement to highlight the saliency for the result of global salient information extraction.

3.2.1. Maximal salient information extraction

The GSI described above is used to obtain the salient component of the image. But the GSI has limit to eliminate the noise with large size. Note that the larger the value of n (i.e., the size of the block), the fewer the number of the noise. However, large value of n will introduce more miss detections. Therefore, in this section, we aim to incorporate a method to eliminate the noise with large size. Many early works are used to eliminate the noise, such as DOG, Gabor filter, and entropy maximization [19]. In this paper, we incorporate entropy maximization method [19] to eliminate the noise. This method can optimize the immediate energy distribution in the system to achieve an energy-economic representation of its environment, which is introduced in the following.

The method [19] imagines that each image patch, represented by the vector $\boldsymbol{x}$, has been formed by the linear combination of N basis functions. The algorithm of independent component analysis (ICA) [35] is used to learn the matrix of filters $\boldsymbol{F} = [f_1, f_2, \ldots, f_k, \ldots]^T$. Given an image $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, \ldots]$. The coefficient $s = \boldsymbol{FX}$ can be computed, which could be used to represent the early visual feature response, and to decide where the salient regions are.

Given the coefficient s which could encode visual information of all features, the coding strategy that makes equal use of all possible feature response levels should be the most efficient. In order to achieve this optimality, the entropy $H(\boldsymbol{P})$ is introduced. The model can be solved by maximizing of the entropy $H(\boldsymbol{P})$.

In order to detect the change of entropy $H(\boldsymbol{P})$ with respect to the feature activity probability increment $p_i = \sum_k |f_i \mathbf{x}_k| / \sum_i \sum_k |f_i \mathbf{x}_k|$, the incremental coding length

(ICL) [19] is introduced

$$ICL(p_i) = \frac{\partial H(\boldsymbol{P})}{\partial p_i} = -\frac{\partial p_i \log p_i}{\partial p_i} - \frac{\partial \sum_{j \neq i} p_i \log p_i}{\partial p_i} \tag{4}$$

where $ICL(p_i) > 0$ denotes that the feature of the region is salient. Let $S_{\text{ICL}} = \{i | ICL(p_i) > 0\}$ represent the salient feature set, the partition $\{S_{\text{ICL}}, \overline{S}_{\text{ICL}}\}$ implies whether successive observations of feature i would increase the entropy $H(\boldsymbol{P})$.

Therefore, the saliency map $\boldsymbol{M} = [\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots, \boldsymbol{m}_n]$ can be quantified as

$$\boldsymbol{m}_k = \sum_{i \in S_{\text{ICL}}} \frac{ICL(p_i)}{\sum_{j \in S_{\text{ICL}}} ICL(p_j)} f_i \boldsymbol{x}_k \tag{5}$$

Note that non-salient features are automatically neglected by setting $\boldsymbol{m}_k = 0$. From (5), we can see that the saliency of a patch may vary over space and time.

The ICL based method [19] provides the possibility to obtain the salient features by measuring the perspective entropy gain of each feature. Some examples of ICL based method are shown in Fig. 3. The first row of Fig. 3 shows that a white salient object can be detected in the grassland. However, ICL based method cannot detect salient feature in some complex scenes in which background is large and bright, and foreground is black and smooth. As shown in the second row of Fig. 3, the monk in the complex scenes cannot be detected by ICL based method. Therefore, ICL has its limits. In this paper, we only incorporate ICL based method to obtain the maximal salient information to eliminate the noise obtained from the GSI, and the process is as follows.

Let $R_{\text{ICL}} = \{R_{\text{ICL}}^1, R_{\text{ICL}}^2, \ldots, R_{\text{ICL}}^N\}$ represents the saliency region obtained by thresholding the saliency map of ICL
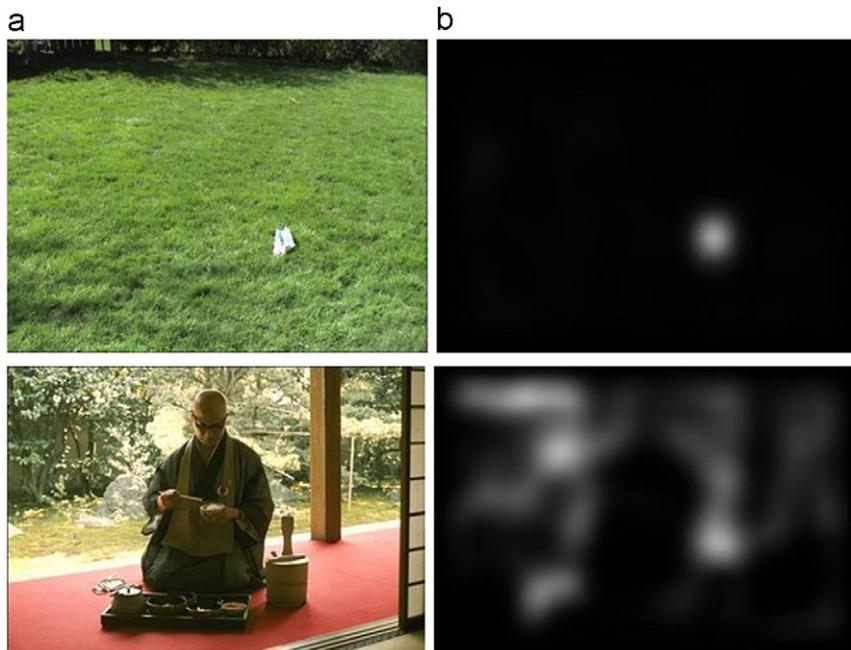
Fig. 3. The performance of ICL for salient object detection. (a) Raw images. (b) Results of ICL.

based method with threshold T_{ICL} (here, it is 0.6), and $R_{GSI} = \{R_{GSI}^1, R_{GSI}^2, \ldots, R_{GSI}^M\}$ represents the saliency region obtained from the segmentation result of GSI. R_{GSI}^B denotes the biggest region in R_{GSI}. For the region R_{GSI}^j whose area is smaller than the area of R_{GSI}^B, if $R_{ICL}^i \cap R_{GSI}^j = \emptyset$, R_{GSI}^j can be regarded as noise which can be eliminated, otherwise, if $R_{ICL}^i \cap R_{GSI}^j \neq \emptyset$, R_{GSI}^j can be regarded as salient region which should be retained. The obtained result is used to represent the segmentation result of GSIM. In other words, the ICL based method is used to obtain the position of the maximal salient information. The small regions for the segmentation result of GSI which are not located at this position will be regarded as noise. The remaining regions construct the saliency regions of GSIM.

3.2.2. Global salient information enhancement

In order to make the result of saliency detection better, many methods can be used to enhance the saliency, such as the theory of information maximization [17], context-aware [21]. In this paper, we propose a method to enhance the salient information.

With the segmentation result of GSIM, we define the white regions of the segmentation result as the foci of attention. The areas whose colors are close to the foci of attention should be explored significantly more than far away regions. Let x_i be the pixel of the input image. $g(x_i)$ is the gray value in the obtained initial saliency map for GSI, which is in the range [0,1]. The enhanced saliency value of pixel x_i is defined as

$$S(x_i) = 1 - \exp\left(-\frac{g(x_i)}{(1+c)}\right) \qquad (6)$$

where c is experimentally set to 0.3 if x_i exists in the salient region of the segmentation result of GSIM, and c is set to 0 if x_i exists in the background region of the segmentation result of GSIM. Note that the larger $S(x_i)$ is, the more salient pixel x_i is. Thus, the salient information is maximized while the background information is minimized.

An example of GSIM which involves maximal salient information extraction and salient information enhancement is shown in Fig. 4. First, we obtain the initial saliency map and segmentation result of GSI, which is embedded with noise in the bottom right corner. Second, we incorporate the thresholding result of ICL (see Fig. 4(f)) to locate the maximal salient region of Fig. 4(d) and eliminate the noise, thus, the segmentation result of GSIM is obtained, which is shown in Fig. 4(g). Finally, salient information enhancement method is used to highlight the salient regions and weaken the background regions. The

final saliency map of GSIM is obtained, which is shown in Fig. 4(h). Compared with ground truth in Fig. 4(b), it is clear that we can obtain good segmentation result and saliency map which are shown in Fig. 4(g) and (h).

4. Experimental results

In this section, to evaluate the performance of the proposed method, we compare the proposed method with 11 state of the art methods in three cases in the qualitative measure. The quantitative evaluation is also obtained by comparing precision and recall curves on the database.

4.1. Data description and parameters setting

We verify the performance of our proposed saliency detection method on 1500 images selected from the database of MSRA [30]. To the best of our knowledge, the database of MSRA [30] is the largest database for saliency detection. To obtain the ground truth database, we manually segmented all images into salient regions and backgrounds, which are labeled as one and zero in the ground-truth mask. The images and ground-truth masks can be found from our future website.

To compute the segmentation result of GSI, the value of n (i.e., size of the block) is set to 3 in this paper. To calculate the saliency map of ICL, a set of basis functions are learned to yield a sparse representation of natural image patches. And the basis functions are considered as features to analyze the visual attention. Here, as shown in Fig. 5, a set of $8 \times 8 \times 3 = 192$ basis functions is learned from 120,000 $8 \times 8 \times 3$ RGB image patches extracted randomly from natural scenes[16,39].

4.2. Comparison with other methods

To demonstrate the performance of our proposed saliency detection method, we perform saliency detection on all of 1500 images by using our proposed method and several different saliency detection methods [12–21]. Table 1 compares the average time taken by each method. Most images in the database have resolution 400×300, and these methods were tested using an Dual Core 2.8 GHz machine with 2 GB RAM. The methods namely FT [12], HC [13], LC [14], RC [13], and AC [15] are implemented in C++, the other methods namely SR [16], AIM [17], GB [18], ICL [19], IT [20], CA [21] and our method are implemented in Matlab. Our GSI method needs O(N) computation time and is sufficiently efficient for real-time application. In contrast, our GSIM method is
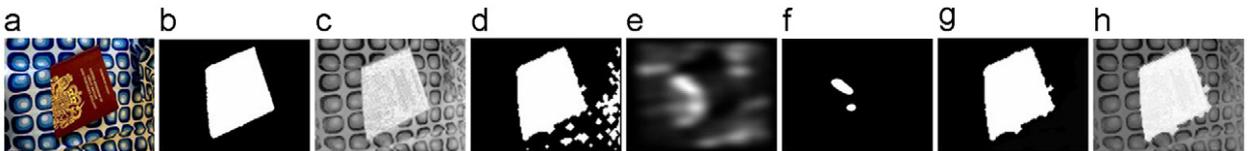
Fig. 4. Maximal salient information extraction and salient information enhancement. (a) Raw image. (b) Ground truth. (c) Initial saliency map of GSI. (d) Segmentation result of GSI. (e) Result of ICL [19]. (f) Thresholding result of (e). (g) Segmentation result of GSIM which involves maximal salient information extraction and salient information enhancement. (h) Final saliency map of GSIM.

slower as it requires the result of ICL [19] and the operation of salient information enhancement. But the latter method produces superior quality saliency maps. Next, we qualitatively compare the results of the proposed method with methods [12–21] between three cases.

In the first case, we use an image with a single salient object over a non-interesting background. Fig. 6(a) shows a flower in a simple background. Fig. 6(m) and (n) shows the results of GSI and GSIM using our proposed method. It can be seen that our proposed method is able to extract the flower with the original color images. In particular, the result of GSIM is better than the result of GSI. It is due to the effect of maximal salient information extraction and salient information enhancement. The results using other different methods [12–21] are shown in Fig. 6(b)–(l), which are not so good as the result of GSIM. For example, as
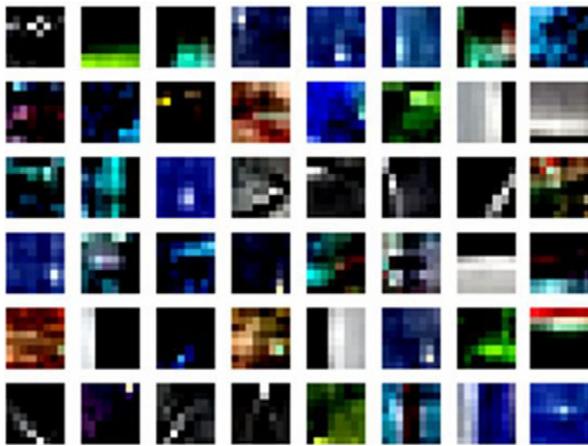
shown in Fig. 6(e), RC [13] is able to detect the flower, but the rachis is also false detected. Note that the petal of the flower is not so salient as the saliency map of GSIM by using the method of RC [13].

In the second case, we select an image with multiple salient objects over simple scenes. In this case, all of the objects are silent; therefore, we expect to extract all of them. Fig. 7(a) shows several leaves in a simple background, which consists of two components, i.e., white plane and the shadow of the leaves. Fig. 7(b)–(l) shows the results of the different saliency detection methods [12–21]. From these figures, we can see that the shadow is false detected by FT [12], HC [13], LC [14], RC [13], AC [15], AIM [17], IT [20], and CA [21]. The methods of SR [16], GB [18], and ICL [19] only detect parts of the salient object. Our proposed method is able to achieve the best saliency detection results which are shown in Fig. 7(m) and (n). It can be clearly seen that the leaves are detected accurately while the shadow is ignored by our proposed method.

The third case includes images of complex scenes shown in Figs. 8 and 9. Fig. 8 shows a guideboard among a complex background, which consists of five components, i.e., sky, forest, building, ground, and river. The color, shape, and texture of each component are different from those of other components. These figures clearly show that our proposed method is able to achieve the best result toward the other methods. For instance, as shown in Fig. 8(l), CA [21] obtains the guideboard but introduce the noise of building. Note that ICL [19] detects the position of the guideboard, but only parts of the salient object are obtained. Fig. 9 shows a toy on an alphabet board. The results are interesting. Our proposed method which considers the features of the human vision is able to accurately detect the toy, while most of the other methods falsely recognize the alphabet board as the salient objects.

Fig. 5. The 48 components of the basis function.

Table 1
Average time taken to compute a saliency map for images in the testing database.

Method	FT [12]	HC [13]	LC [14]	RC [13]	AC [15]	SR [16]	AIM [17]	GB [18]	ICL [19]	IT [20]	CA [21]	GSI	GSIM
Time (s)	0.015	0.018	0.017	0.251	0.107	0.062	9.994	1.620	0.264	0.607	51.200	0.419	0.823
Code	C++	C++	C++	C++	C++	Matlab	Matlab	Matlab	Matlab	Matlab	Matlab	Matlab	Matlab
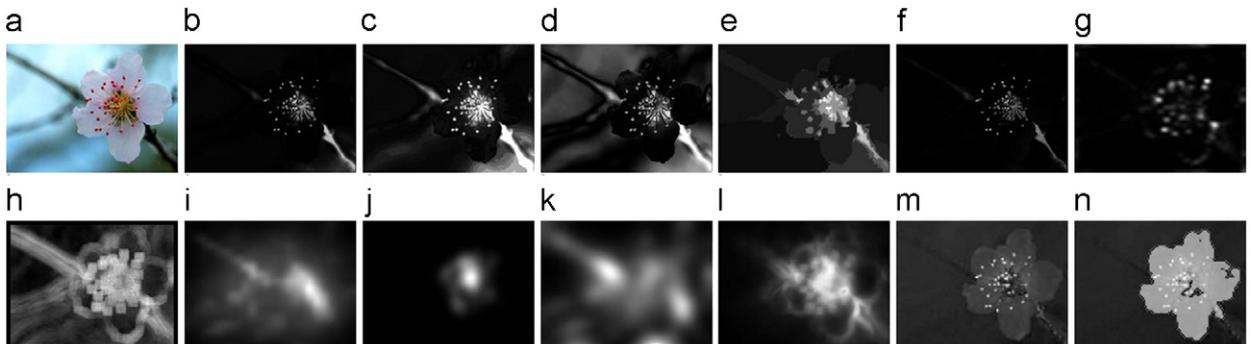
Fig. 6. Comparison of different methods on images with one object and simple background. (a) Original image, saliency maps produced using (b) FT [12], (c) HC [13], (d) LC [14], (e) RC [13], (f) AC [15], (g) SR [16], (h) AIM [17], (i) GB [18], (j) ICL [19], (k) IT [20], (l) CA [21], (m) GSI, (n) GSIM.
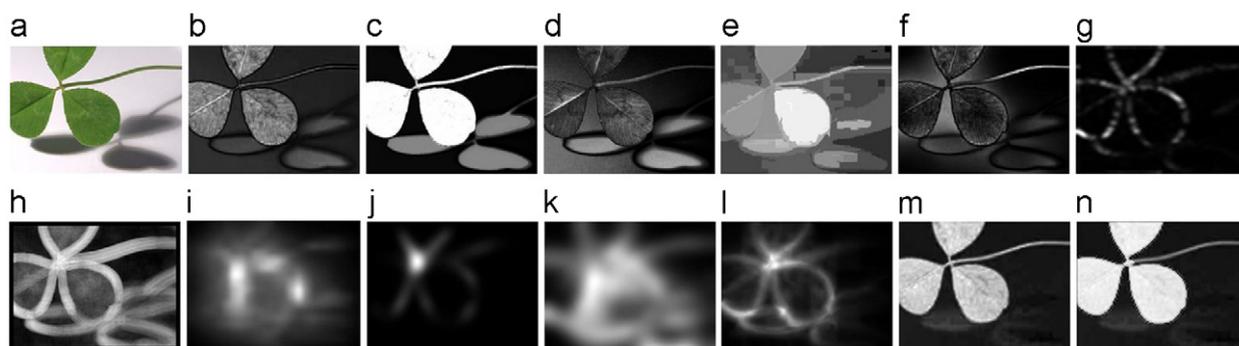
Fig. 7. Comparison of different methods on images with multiple objects and simple background. (a) Original image, saliency maps produced using (b) FT [12], (c) HC [13], (d) LC [14], (e) RC [13], (f) AC [15], (g) SR [16], (h) AIM [17], (i) GB [18], (j) ICL [19], (k) IT [20], (l) CA [21], (m) GSI, (n) GSIM.
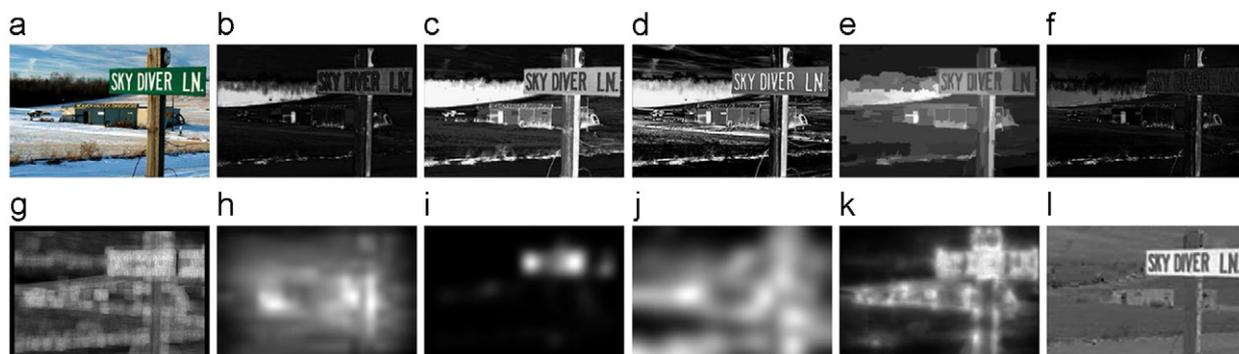
Fig. 8. Comparison of different methods on images in complex scene. (a) Original image, saliency maps produced using (b) FT [12], (c) HC [13], (d) LC [14], (e) RC [13], (f) AC [15], (g) SR [16], (h) AIM [17], (i) GB [18], (j) ICL [19], (k) IT [20], (l) CA [21], (m) GSI, (n) GSIM.
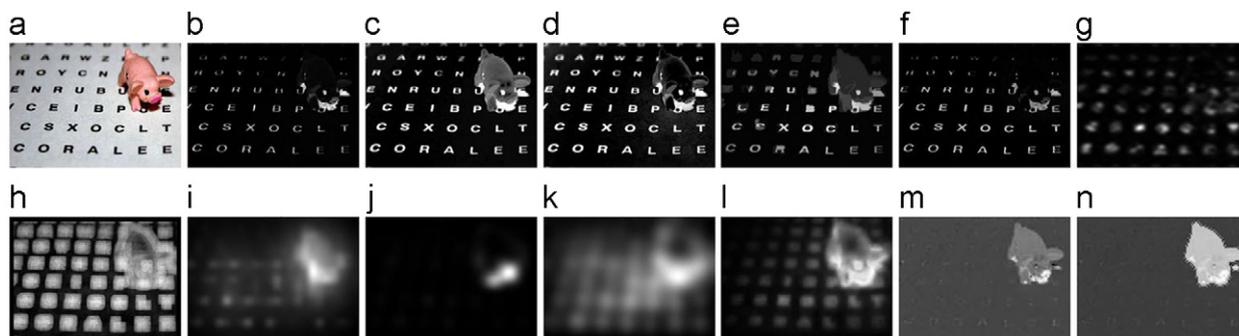
Fig. 9. Comparison of different methods on images in complex scene. (a) Original image, saliency maps produced using (b) FT [12], (c) HC [13], (d) LC [14], (e) RC [13], (f) AC [15], (g) SR [16], (h) AIM [17], (i) GB [18], (j) ICL [19], (k) IT [20], (l) CA [21], (m) GSI, (n) GSIM.

Note that the method of ICL [19] can obtain the position of the toy, but only parts of the salient object are detected.

In order to evaluate the proposed method in quantitative measure, we compare precision and recall curves on the database in which there are 1500 images that are selected from the database of MSRA [30]. To obtain precision and recall curves, we binarize the saliency map using each possible fixed threshold T which is in the range [0,255]. To compare how well various saliency detection methods highlight salient regions in images, we vary the threshold T from 0 to 255. The precision and recall curves in Fig. 10 clearly show that our proposed method GSIM is significantly superior to the other methods. The extremities of the precision and recall curves are interesting. For instance, on the left hand of each figure in Fig. 10, the minimum recall values of our method are higher than those of the other method. It is due to reason that the saliency maps computed by our methods contain more pixels with the saliency value 255. On the right hand of each figure in Fig. 10, the value of recall is
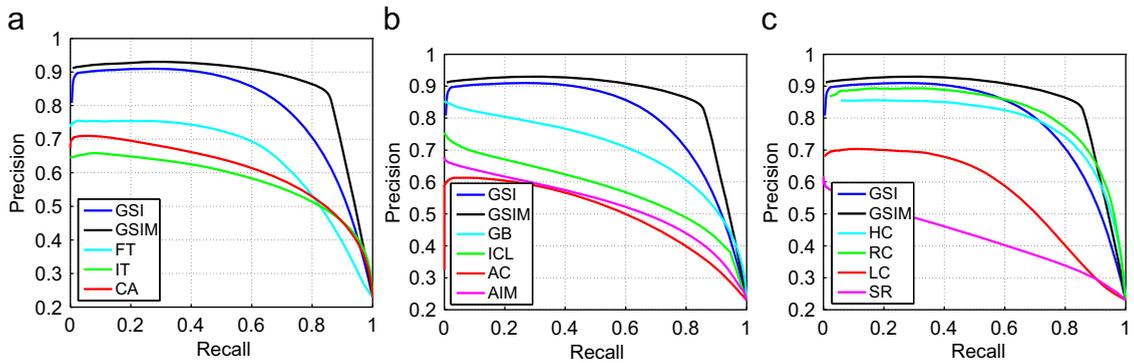
Fig. 10. Precision-recall curves for the naive thresholding of saliency maps using 1500 benchmark images which are selected from the databases of MSRA.

maximal when $T=0$, all pixels are considered to be salient regions. Therefore, all the methods have the same precision and recall values.

5. Graph based salient object segmentation

As mentioned in Section 3, we propose the method of GSIM, which can be used to produce good segmentation results for most of the images in the testing database. However, due to the linear property of PCA, there are some false and miss detections. Therefore, we propose a graph based salient object segmentation method (GGSIM) to improve the segmentation results of GSIM.

There are many traditional methods of saliency detection which are used in unsupervised object segmentation. For example, Ma and Zhang [41] propose a method to find rectangular salient regions by fuzzy region growing on their saliency maps. The method of Itti et al. [20] produces saliency maps to perform unsupervised object segmentation. With Itti's saliency map, Han et al. [10] use a Markov random field to combine low-level features of color, texture, and edges to extract the salient object regions. Ko and Nam [40] employ a Support Vector Machine which is trained on image segment features to extract the salient regions by using Itti's maps. Then, the extracted salient regions are clustered to obtain the salient objects. Achanta et al. [12] average saliency values within image segments which is produced by mean-shift segmentation, and then find salient objects by identifying image segments that have average saliency value above a threshold which is set to be twice the mean saliency value of the entire image. Cheng et al. [13] iteratively apply GrabCut to refine the segmentation result initially obtained by thresholding the saliency map.

In this paper, we propose an automatic image segmentation method by iterated graph cut. Similar to the grab cut based segmentation algorithm [8], we also employ the method of iterated graph cuts. The grab cut method, which is based on graph cut and Gaussian mixture models, needs to put a light load on the user to drag a rectangle around the desired object. In this paper, instead of manually inputting a rectangular region to initialize the process, we use the segmentation result of GSIM to represent the initialize image for Graph cut.

In this method, once initialized, we iteratively apply graph cut to improve the saliency segmentation result. Note that we employ closing operation, which is a dilation followed by an erosion operation, on the current segmentation result to get a new trimap for the next graph cut iteration. For each trimap, foreground, background, and unknown regions are set as follows: the region inside the eroded region is set to foreground, while the region outside the dilated region is regarded as background, and the remaining regions are set to unknown regions. Graph cut helps to refine salient object regions at each step. Regions near to an initial salient object region are more likely to be part of the salient object than far away regions. Due to the good segmentation result of GSIM, the graph based salient object segmentation algorithm can convergence in less than three iterations. Some examples of graph based salient detection are shown in Fig. 11. For instance, we obtain the segmentation result of GSIM shown in the second row of Fig. 11(d). However, this segmentation result is not accurate while the nose of the dog is ignored. After graph based salient object segmentation, the nose of the dog is retrieved, and the accurate segmentation result is obtained, which is shown in Fig. 11(f).

To evaluate the proposed graph based method for salient object segmentation, we compare our results with results obtained by other two methods, i.e., FT [12] and RCC [13]. The segmentation results have been compared against the ground truth of 1500 images of the database by qualitative and quantitative measures. Fig. 11 shows some examples for the quantitative comparison. We can see that our proposed method can achieve better segmentation results than the other methods.

A quantitative comparison of the results is shown in Fig. 12. The value of average precision, recall and F-measure are compared over the entire ground truth database. The F-measure is defined as

$$F_\beta = \frac{(1+\beta^2)precision \times recall}{\beta^2 \times precision + recall} \tag{7}$$

where β^2 is set to 0.3 to weight precision more than recall [12]. From Fig. 12, we can clearly see that the proposed method which is based on the result of GSIM outperforms other methods. Compared with the results by Cheng et al.
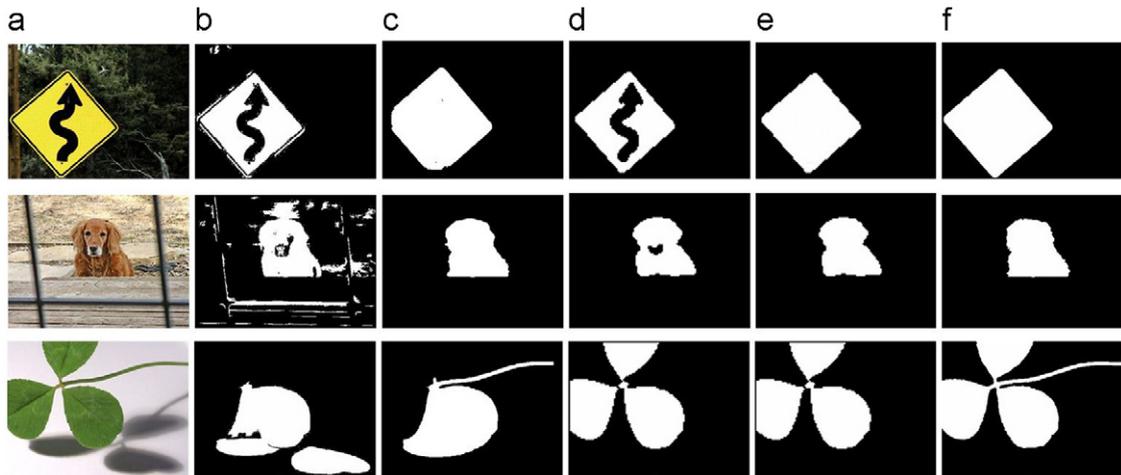
Fig. 11. Salient region segmentation using different methods. (a) Original image. The segmental results produced using (b) FT [12], (c) RCC [13], (d) segmentation result of GSIM, (e) segmentation result of graph based GSIM (GGSIM). (f) Ground truth.
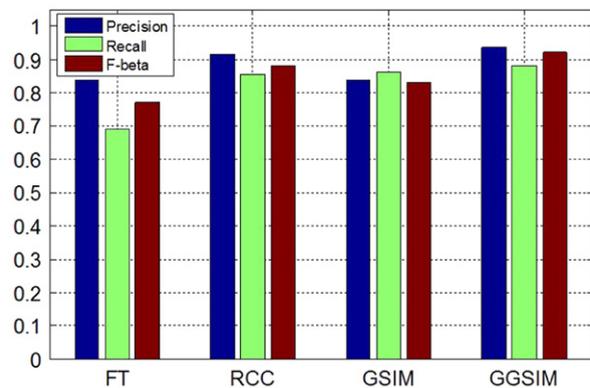
Fig. 12. Precision-recall bars for the results of different saliency segmentation methods over 1500 images.

[13] (precision=91.52%, recall=85.62%), we achieved better accuracy (precision=93.36%, recall=88.28%).

6. Conclusion

In this paper, we propose a novel method for saliency detection which is based on the features of the salient object. This method can extract the global salient information by PCA based method. And we use two methods, i.e., maximal salient information extraction and salient information enhancement methods, to effectively eliminate the noise and enhance the salient information respectively. Graph based salient object segmentation is also proposed to extract the salient object. The experimental results demonstrate that the proposed method achieves better results than the state of the art methods.

Acknowledgement

References

[1] A. Berengolts, M. Lindenbaum, On the distribute of saliency, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (12) (2006) 1973–1990.
[2] A. Treisman, G. Gelade, A feature-integration theory of attention, Cognitive Psychology (1980).
[3] A. Yarbus, Eye Movements and Vision, Plenum, NY, 1967.
[4] U. Rutishauser, D. Walther, C. Koch, P. Perona, Is bottom-up attention useful for object recognition?, in: Proceedings of CVPR, 2004, pp 37–44.
[5] N. Murray, M. Nanrell, X. Otaze, C.A. Parraga, Saliency estimation using a non-parametric low-level vision model, in: Proceedings of CVPR, Colorado Springs, Colorado, USA, June 2011, pp. 21–23.
[6] J. Freixenet, X. Munoz, D. Raba, J. Marti, X. Cufi, Yet another survey on image segmentation: region and boundary information integration, in: Proceedings of ECCV, 2002, pp. 408–422.
[7] S. Avidan, A. Shamir, Seam carving for content-aware image resizing, ACM Transactions on Graphics 26 (3) (2007).
[8] C. Rother, V. Kolmogorov, A. Blake, GrabCut: interactive foreground extraction using iterated graph cuts, ACM Transactions on Graphics 23 (3) (2004) 309–314.
[9] L. Lubor, S. Paul, A. Karteek, R. Chris, H.S. Torr Philip, What, where & how many? Combining object detectors and CRFs, in: Proceedings of ECCV, 2010.
[10] J. Han, K.N. Ngan, M. Li, H.J. Zhang, Unsupervised extraction of visual attention objects in color images, IEEE Transactions on Circuits and Systems for Video Technology, USA 16 (1) (2006) 141–145.
[11] H. Li, K.N. Ngan, Saliency model based face segmentation in head-and-shoulder video sequences, Journal of Visual Communication and Image Representation, Elsevier Science 19 (5) (2008) 320–333.
[12] R. Achanta, S. Hemami, F. Esgtrada, S. Süsstrunk, Frequency-tuned salient region detection, in: Proceedings of CVPR, 2009, pp. 1597–1604.
[13] M.-M. Cheng, G.-X. Zhang, N.J. Mitra, X. Huang, S.-M. Hu, Global contrast based salient region detection, in: Proceedings of CVPR, Colorado Springs, Colorado, USA, June 2011, pp. 21–23.
[14] Y. Zhai, M. Shah, Visual attention detection in video sequences using spatiotemporal cues, in: ACM Multimedia, 2006, pp. 815–824.
[15] R. Achanta, F. Estrada, P. Wils, S. Süsstrunk, Salient region detection and segmentation, in: Proceedings of ICVS, 2008, pp. 66–75.
[16] X. Hou, L. Zhang, Saliency detection: a spectral residual approach, in: Proceedings of CVPR, 2007, pp. 1–8.
[17] N. Bruce, J. Tsotsos, Saliency based on information maximization, Proceedings of NIPS, vol. 18, 2006, pp. 155–162.

[18] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, in: Proceedings of NIPS, 2006.

[19] X. Hou, L. Zhang, Dynamic visual attention: searching for coding length increments, in: Proceedings of NIPS, 2008.

[20] L. Itti, C. Koch, E. Niebur, A model of saliency based visual attention for rapid scene analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence (1998).

[21] S. Goferman, L. Zelnik-Manor, A. Tal, Context-aware saliency detection, in: Proceedings of CVPR, 2010.

[22] M. Mancas, B. Gosselin, B. Macq, Perceptual Image Representation, EURASIP, 2007.

[23] O. Le Meur, P. Le Callet, D. Barba, D. Thoreau, A coherent computational approach to model bottom-up visual attention, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (5) (2006) 802–817.

[24] Y. Yu, G.K.I. Mann, R.G. Gosine, An object-based visual attention model for robotic applications, IEEE Transactions on Systems, Man, and Cybernetics-part B: Cybernetics 40 (5) (2010).

[25] L. Itti, C. Koch, A saliency-based search mechanism for overt and covert shifts of visual attention, Vision Research 40 (10–12) (2000) 1489–1506.

[26] L. Costa, Visual saliency and attention as random walks on complex networks, Physics 0603025 (2007).

[27] W. Wang, Y. Wang, Q. Huang, W. Gao, Measuring visual saliency by site entropy rate, in: Proceedings of CVPR, 2010.

[28] D. Gao, V. Mahadevan, N. Vasconcelos, The discriminant center-surround hypothesis for bottom-up saliency, in: Proceedings of NIPS, 2007.

[29] D. Gao, N. Vasconcelos, Bottom-up saliency is a discriminant process, in: Proceedings of ICCV, 2007.

[30] T. Liu, J. Sun, N. Zheng, X. Tang, H. Shum, Learning to detect a salient object, in: Proceedings of CVPR, 2007.

[31] B. Alexe, T. Deselaers, V. Ferrari, What is an object?, in: Proceedings of CVPR, 2010.

[32] B. Suh, H. Ling, B.B. Bederson, D.W. Jacobs, Automatic thumbnail cropping and its effectiveness, in: Proceedings of UIST, 2003, pp. 95–104.

[33] J. Wolfe, Guided search 2.0. A revised model of visual search, Psychonomic Bulletin & Review 1 (2) (1994) 202–238.

[34] C. Koch, T. Poggio, Predicting the visual world: silence is golden, Nature Neuroscience 2 (1999) 9–10.

[35] A. Bell, T. Sejnowski, The independent components of natural scenes are edge filters, Vision Research 37 (23) (1997) 3327–3338.

[36] T. Celik, Unsupervised change detection in satellite image using principal component analysis and k-means clustering, IEEE Geoscience and Remote Sensing Letter 6 (1) (2009) 33–37.

[37] H. Zha, C. Ding, M. Gu, X. He, H.D. Simon, Spectral relaxation for k-means clustering, in: Proceedings of NIPS, 2001, pp. 1057–1064.

[38] C. Ding, X. He, K-means clustering via principal component analysis, in: Proceedings of ICML, 2004, pp. 225–232.

[39] J. Hateren, A. van der Schaaf, Independent component filters of natural images compared with simple cells in primary visual cortex, in: Proceedings of the Royal Society of London Series B: Biological Sciences, 1998.

[40] B.C. Ko, J.Y. Nam, Object-of-interest image segmentation based on human attention and semantic region clustering, The Journal of the Optical Society of America 23 (10) (2006) 2462–2471.

[41] Y.-F. Ma, H.-J. Zahng, Contrast-based image attention analysis by using fuzzy growing, in: ACM Multimedia, 2003, pp. 374–381.