Signal Processing: *Image Communication* 29 (2014) 434–447

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Bayesian salient object detection based on saliency driven clustering

Lei Zhou [a], Keren Fu [a], Yijun Li [a], Yu Qiao [a], XiangJian He [b], Jie Yang [a,*]

[a] *Shanghai Jiao Tong University, Shanghai, China*
[b] *University of Technology, Sydney, Australia*

ARTICLE INFO

Article history:
Received 5 August 2013
Received in revised form
3 January 2014
Accepted 3 January 2014
Available online 30 January 2014

Keywords:
Saliency object detection
Saliency driven clustering
Regional saliency computation
Bayesian model

ABSTRACT

Salient object detection is essential for applications, such as image classification, object recognition and image retrieval. In this paper, we design a new approach to detect salient objects from an image by describing what does salient objects and backgrounds look like using statistic of the image. First, we introduce a saliency driven clustering method to reveal distinct visual patterns of images by generating image clusters. The Gaussian Mixture Model (GMM) is applied to represent the statistic of each cluster, which is used to compute the color spatial distribution. Second, three kinds of regional saliency measures, i.e, regional color contrast saliency, regional boundary prior saliency and regional color spatial distribution, are computed and combined. Then, a region selection strategy integrating color contrast prior, boundary prior and visual patterns information of images is presented. The pixels of an image are divided into either potential salient region or background region adaptively based on the combined regional saliency measures. Finally, a Bayesian framework is employed to compute the saliency value for each pixel taking the regional saliency values as priority. Our approach has been extensively evaluated on two popular image databases. Experimental results show that our approach can achieve considerable performance improvement in terms of commonly adopted performance measures in salient object detection.

1. Introduction

A human visual system (HVS) often pays more attention to some parts of an image. It is visual attention that allows people to select the information which is most relevant to ongoing behavior. Visual attention has been studied by researchers in physiology, psychology, neural systems, and computer vision for a long time. Extracting objects from an image is a hot research topic and has wide applications, such as content-based image retrieval [1], image/video compression and coding [2], object recognition and scene understanding [3–6] and image segmentation [7,8] in areas of computer vision and computer graphics. Under the mechanism of visual attention, HVS picks out relevant parts of a scene as attention regions corresponding to salient regions in images. In natural scene, salient regions generally stand out relative to its surroundings. This mechanism can be explained by a center-surround difference model [9], which is implemented in the feature spaces of luminance, color and orientation.

In recent years, salient object detection has aroused researches' interest and the related work has been divided into two categories, i.e, approaches of bottom-up category and approaches of top-down category respectively. In bottom-up visual saliency, previous research [10,11] revealed that *contrast* is the most influential factor in low-level visual saliency. By computing the contrast over a pixel domain or region domain, many visual saliency

* Corresponding author.
E-mail address: jieyang@sjtu.edu.cn (J. Yang).

0923-5965/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.image.2014.01.001

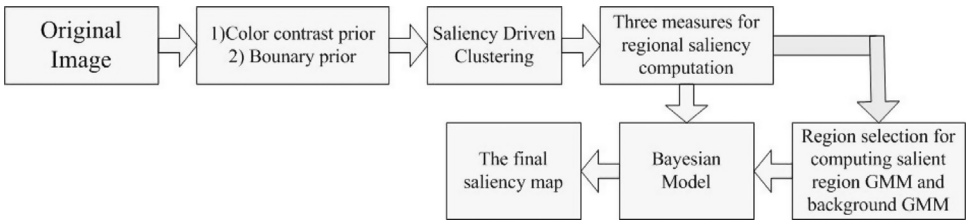

Fig. 1. The flowchart of our proposed model.

models have been proposed over the past year [12–17]. The existing saliency models based on color contrast can simultaneously compute global contrast of an image and spatial coherence between regions and have displayed impressive results. The contrast-based models tell "what the objects look like" by highlighting the pixels with great center-surround difference. However, the performance of the saliency maps that only rely on color contrast will degrade when the images are of confusing pattern or complex scene.

Different from the contrast prior, background prior tackles the salient object detection problem by asking the question "what the background should look like". In [18], two priors, boundary and connectivity prior were used as the priors about common backgrounds in natural images. The boundary prior was discovered from the observations that "the image boundary is mostly background" and "the salient objects seldom touch the image boundary". Even if the boundary priors work for most images, it may fail when objects significantly touch the image boundary or the images are of complex background.

Intuitively, if both the information about "what the objects look like" and "what the backgrounds look like" is available, it will be easier to tackle the ill-posed salient object problem. To describe the appearance of salient objects or backgrounds, statistic of images is a powerful feature. From the perspective of statistic theory, many saliency models have been proposed [19–21]. However, the extracted initial salient or background regions are crucial for constructing reliable appearance models. In [20], a coarse salient region was first obtained via a convex hull of interest points and a Laplacian sparse subspace clustering algorithm was presented to obtain the prior map related to the salient object. The algorithm is effective but the cluster method used is of high computation complexity. In [19], each image was divided into initial attention region and initial background region using an adaptive threshold. This method may fail when the salient object possesses a low contrast with the background. In [21], a polygonal potential Region-Of-Interest was extracted through analyzing the edge distribution in an image.

In this paper, we address the salient object detection problem by collecting reliable information about "what the salient objects look like and what the backgrounds look like" by taking the advantages of contrast prior, boundary prior and visual patterns of images. To tell "what the salient objects look like", we propose a method for selecting initial salient and background region based on the techniques like, saliency driven clustering, regional saliency computation and adaptively thresholding. To capture the structural information of an image, the image is first separated into several non-overlapping regions by saliency driven clustering. Then, we utilize three regional saliency measures, i.e, regional color contrast saliency, regional boundary prior saliency and color spatial distribution, to compute the saliency level of each region. Three measures are combined non-linearly to obtain the pixel level saliency values. Then, we classify the pixels of an image into either potential salient region or background region using some adaptive threshold. The color information within the potential salient region and background region is represented using GMMs. Finally, the saliency model is computed by applying a general Bayesian framework. The main contributions of the paper are summarized as follows and the flow chart of our approach is shown in Fig. 1.

- We propose an effective method for implementing the boundary prior saliency. A saliency driven clustering approach is proposed based on the combination of color contrast saliency and boundary prior saliency. The cluster numbers are determined automatically by histogram analysis.
- Color spatial distribution is calculated by analyzing the color statistic of each cluster. The regional saliency values are computed by combining three regional saliency measures, i.e, regional color contrast saliency, color spatial distribution and regional boundary prior saliency.
- An adaptive initial regions selection method is presented. A Bayesian framework is applied to generate the final saliency maps.

The organization of this paper is as follows. The related work is presented in Section 2. We introduce the approach of saliency driven clustering in Section 3. The strategy of region selection based on regional saliency computation is proposed in Section 4, the three measures are introduced in Section 4.2. Our statistic saliency model is presented in Section 5. Experimental results are given in Section 6 and we conclude in Section 7.

2. Related work

Among bottom-up category, as one of the work [15] described, the contrast was often defined over various different classes of image features including color variation of pixels or image patches, spatial frequencies, structure and statistic distribution of image patches, histogram and the combination of all above. In [16], pixel's or region's contrast was computed over all other pixels or regions. Then the compactness in the spatial domain and the cluster contrast

evaluated by the difference between models of different clusters were combined to generate a saliency map. Fu et al. [17] took the advantages of color contrast and color distribution to carry out the saliency detection. Then a saliency measure was obtained by computing two measures of contrast that rated the uniqueness and the spatial distribution of image patches. Furthermore, postprocessing steps were applied to refine the result as well. In [22], a detection algorithm which was based on four principles observed in the psychological literature was presented. The rule of "distinctive colors or patterns" was considered for computing the saliency. In [15], high dimensional Gaussian filters were formulated to generate saliency map in an efficient way. Besides, there are a lot of visual saliency models which measure visual saliency in the frequency domain. In [23], Hou et al. proposed a visual saliency model based on the natural image statistics. In [24], the phase spectrum of quaternion Fourier transform was exploited to evaluate the saliency at block level.

Many works have also been proposed applying information theory [25–28]. Bruce et al. proposed a model [25] in which visual saliency was represented by a local likelihood of each image patch decomposed by the filters learned from natural images. In [27], Dominik et al. proposed a saliency algorithm in which the contrast of the center and surround distribution of features was computed to based on the Kullback–Leibler divergence for salient object detection. In [28], information divergence was used to express the non-uniform distribution of the visual information in an image and it improved the Bayesian surprise model to compute information divergence across an image. A visual saliency map was finally obtained from the information divergence.

From the perspective of statistical theory, several saliency models have been proposed. The global information of an image was applied to generate a saliency map of high quality as shown in [29,30], the distinctness of different statistic models representing different clusters showed great importance in measuring the saliency of a region. Zhang et al. [31] posited a saliency detection problem by representing visual information of a specific class of object using Bayesian framework and the information of a known target class was modeled by a likelihood function. There are also many works that compute contrast based on image regions' natural statistics. In [32], Bayesian theory was applied to describe the interaction between top-down and bottom-up information. It evaluated and selected visual features before saliency estimation. In [19], the attention Gaussian mixture model (GMM) for salient object and background GMM were constructed based on the image clustering result, and pixels were classified under the Bayesian framework to obtain the salient object. In [29], the framework of mixture of Gaussian in H–S space was used to compute the distance between clusters and color spatial distribution. Then, color and orientation distributions in images are fully utilized to selectively generate a saliency map. In [30], a kernel density estimation (KDE) based nonparametric model was constructed for each segmented region, and color and spatial saliency measures of KDE models were evaluated and exploited to measure saliency of pixels. In [16], the histograms of regions were exploited to generate the saliency map at pixel-level and region-level respectively. In [20], a Bayesian framework was proposed to combine the low level cues (coarse saliency region obtained via a convex hull of interest points) and mid level cues (saliency information provided by superpixels) to generate a saliency map.

The significant difference between the proposed model and previous statistical theory based models such as [19–21] is that the initial salient and background regions are extracted by a region selection strategy which integrates color contrast prior, boundary prior and visual patterns information of images. Then, the regional saliency values are taken as the prior probability of Bayesian model and likelihood probability is computed by analyzing statistic of the adaptively selected initial regions.

3. Saliency driven clustering

Intuitively, image clusters can reveal distinct visual patterns of an image. In this section, we present a saliency driven clustering method. First, the basic notations of the color contrast saliency and boundary prior saliency are introduced. Then, the clustering technique which is based on the analysis of combined saliency map's histogram is introduced.

3.1. Color contrast saliency

Color contrast is inspired by the observation that color components of salient objects may have a strong contrast to their surroundings. Assume that an image with size N is divided into M regions (or superpixels) and the regions are represented as $R_i, i \in \{1,2,3,\ldots,M\}$. Then, region R_i's color contrast saliency S_i^{Rcon} is computed according to the definition in [16,17]:

$$S_i^{Rcon} = \sum_{j \neq i} D_c(R_i, R_j) D_s(R_i, R_j), \tag{1}$$

where $D_c(R_i, R_j) = \| c_i - c_j \|$ is the color distance between the two regions. c_i represents the average color information in region i, i.e., $c_i = \sum_{I_k \in R_i} I_k^C / |R_i|$, for $i \in \{1,2,\ldots,M\}$, where I_k^C (or I_k) is the color feature vector at pixel k. $|R_i|$ is the size of region R_i. $D_s(R_i, R_j)$ in Eq. (1) stands for the spatial distance between regions R_i and R_j, which is defined as

$$D_s(R_i, R_j) = e^{-\alpha \| p_i - p_j \|^2}, \tag{2}$$

where α is a parameter to control the contrast's sensitivity to spatial distance and p_i describes the average position of region R_i, i.e., $p_i = \sum_{I_k \in R_i} I_k^P / |R_i|$, where I_k^P is the coordinate vector at pixel k. In the experiment, 300 superpixels are generated using SLIC [33] and we set $\alpha = 1 \times 10^{-3}$. Finally, the pixel level color contrast saliency is given as $S_i^{con} = S_j^{Rcon}, i \in R_j$. We normalize S^{con} to the range [0,1] through $S^{con} = (S^{con} - \min(S^{con}))/(\max(S^{con}) - \min(S^{con}))$.

3.2. Boundary prior saliency

As stated in [34], the boundary prior is an important and helpful measure for salient object detection. For an

image, the pixel-level undirected weighted graph is represented as $G=\{V,\varepsilon\}$. In the graph, the boundary nodes (Ω_B) are selected using the strategy similar with [34]. Then, the geodesic distance of a pixel m to the boundary nodes is computed as the shortest distance to all the background nodes [35]

$$g(m)=\min_{s\in\Omega_B} d_g(s,m), \tag{3}$$

where $d_g(s,m)$ is the geodesic distance between two nodes s and m, which is computed based on the length of a discrete path [35]:

$$d_g(s,m)=\min_{\Gamma\in P_{s,m}} L(\Gamma), \tag{4}$$

where $P_{s,m}$ stands for the set of paths between node s and m, and the length L of a discrete path Γ is defined as [35]

$$L(\Gamma)=\sum_{i=1}^{n-1}\sqrt{(1-\gamma_g)d(I^i,I^{i+1})^2+\gamma_g\|\nabla(\Delta^i)\|^2}, \tag{5}$$

where Γ is an arbitrary discrete path composed with n pixels $\{I^1,\ldots,I^n\}$. The term $d(I^i,I^{i+1})$ is Euclidean distance between I^i and I^{i+1}) and $\|\nabla(\Delta^i)\|$ is a finite difference approximation of the image gradient between I^i and I^{i+1}. We use the parameter γ_g to weight two kinds of distances: the Euclidean distance and the geodesic distance. The role of γ_g has been studied in [35] and we set $\gamma_g=0.2$ in the experiments. In our implementation, the paths in Eq. (5) are computed using fast marching algorithm [36]. Then, the boundary prior saliency is defined as $S_i^{Boundary}=g(i)$, $i\in[1,\ldots,N]$ and it is normalized to the range [0,1] as well.

3.3. Saliency driven clustering

To integrate the complementary strength of two kinds of saliency maps, the color contrast saliency and boundary prior saliency are combined nonlinearly,

$$S_i^{cb}=S_i^{con}*S_i^{Boundary},\quad i\in[1,\ldots,N]. \tag{6}$$

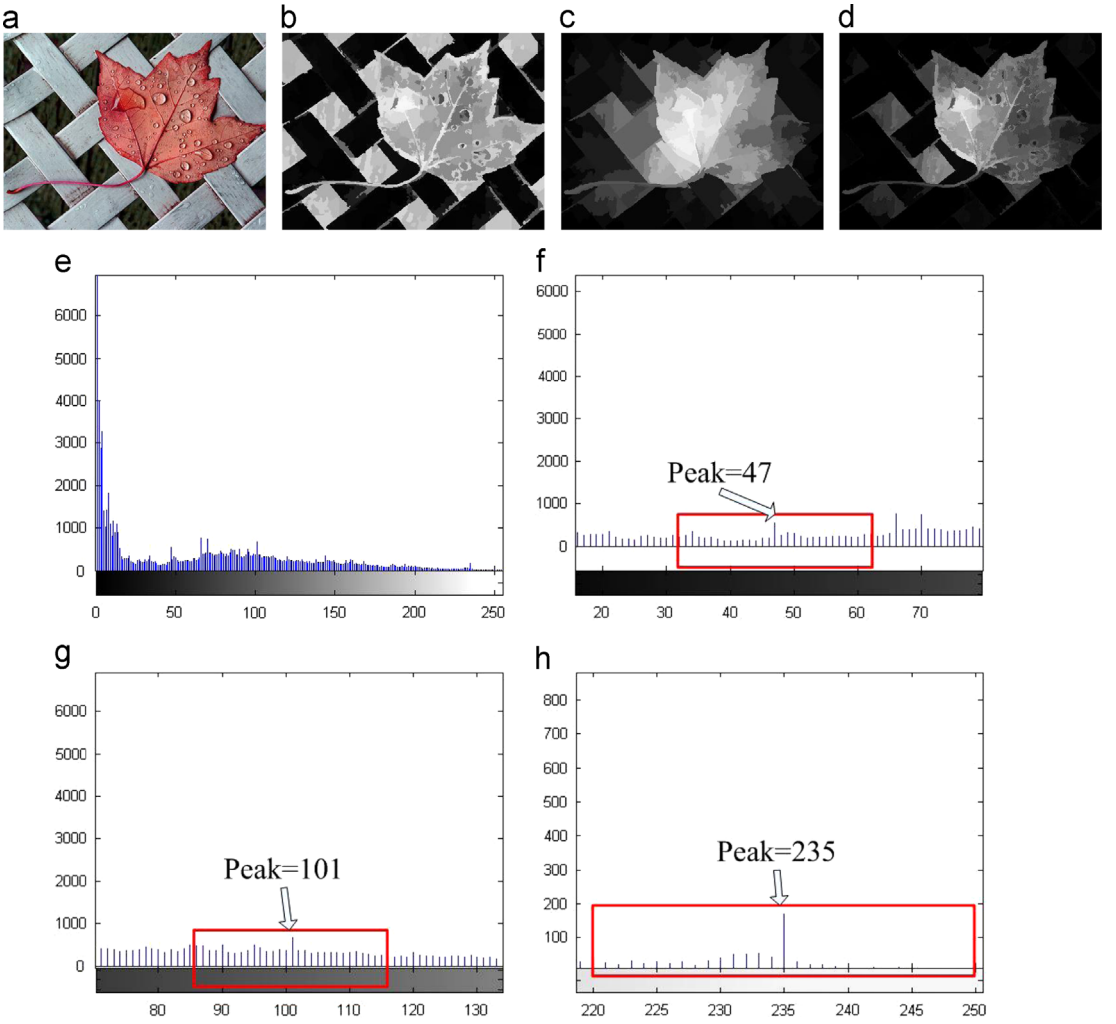

Fig. 2. An example illustrates the procedure of peak selection by histogram analysis. Nine peaks [1, 47, 66, 101, 144, 160, 194, 235, 255] are located by the Hill climbing algorithm. (a) The source image; (b) S^{con}; (c) $S^{boundary}$; (d) S^{cb}; (e) histogram ranges from 0 to 255; and (f) the selected Peak 47. The rectangle is the search window; (g) the selected Peak 101; and (h) the selected Peak 235.

Then, the image space is separated into several non-overlapping regions by K-means clustering [37] based on the saliency values $[S_1^{cb}, \ldots, S_N^{cb}]$. For K-means clustering, the number of clusters K and the initial positions of centroids are determined automatically. We use the Hill Climbing algorithm [38] to analyze the histogram of S^{cb}. The computed peaks of histogram are selected as the starting points for clustering and K is set as the number of peaks. The procedure for saliency driven clustering is described below.

1. We first build the histogram ranging [0,255] of combined saliency map S^{cb}.
2. We construct a search window of size 30 and the center of the window will move from 0 to 255 to search for the peak of histogram. The pixel number of current bin (The bin which lies in the center of the search window is selected as the current bin.) is compared with the neighboring bins' pixel numbers, and the current bin will be selected as a peak if its number is the largest in the search window. For the bins which are in the range [0,15], the size of left half search window will be less than 15 and for the bins in the range of [240, 255], the size of right half search window will be less than 15. We will ensure that there exists at least one half search window whose size is 15 in the searching process. The number of computed peaks is K and the set of peak bins is Pb.
3. We set the clusters number as K and take Pb as the starting centroids for K-means clustering. Then K saliency driven clusters $RS_1, \ldots, RS_K$ are generated.

Fig. 2 displays the process of histogram analysis by moving search window. In Fig. 3, a clustering example is displayed. The cluster number $K=9$ is first determined by analyzing the histogram of combined saliency map. Then, the image displayed is separated into nine non-overlapping regions by K-means. It is observed from the clustering map (Fig. 3(l)) that pixels in clusters (Fig. 3(i)–(k)) only belong to the salient object. The cluster (Fig. 3(c)) contains only the background pixels. The clusters (Fig. 3(d)–(h)) contain both object and background pixels. It is clear that the visual patterns reflected by clusters are distinct.

In the next section, we will introduce the method for region selection by analyzing the property of each cluster, so as to obtain more reliable information about "what the objects look like and what the backgrounds look like".

4. Regional saliency computation for region selection

For the generated K clusters, we compute three kinds of regional saliency values, i.e., regional color contrast saliency, regional boundary prior saliency and color spatial distribution. Different from the definition of distribution in [39], we model the color distribution based on the statistic of generated clusters. The statistics of each region is represented using a GMM and the color distribution is modeled as how widely the color contained in a cluster is separated in the whole image region.

4.1. Representation of color statistics using GMM

To represent color statistics in all the regions, we choose RGB colors as the local features to describe color information for each region $l \in [RS_1, \ldots, RS_K]$ and they are modeled using a Gaussian mixture model (GMM). Let the color models be represented by GMM $\{\alpha_c, \mu_c, \Sigma_c\}_{c=1}^{C}$ in the color space, where $\{\alpha_c, \mu_c, \Sigma_c\}$ contains the weight, the mean color and the covariance matrix of the c-th component. For pixels in each region, a set of GMM parameters are learned. The Gaussian mixture distribution can be written as

$$V(I_x|l) = \sum_c \alpha_{cl} N(I_x|\mu_{cl}, \Sigma_{cl}), \quad l \in [RS_1, \ldots, RS_K], \tag{7}$$

where α_{cl}, μ_{cl} and Σ_{cl} represent the weight, the mean color and the covariance matrix of the c-th component learned from pixels in region l respectively. The parameters of a GMM can be obtained by maximizing the log likelihood function for a GMM using techniques like gradient-based optimization or expectation–maximization algorithm. In our experiments, GMM with five components are used to represent the color statistics in each cluster and the EM algorithm [40] is applied to generate the parameters of GMMs.

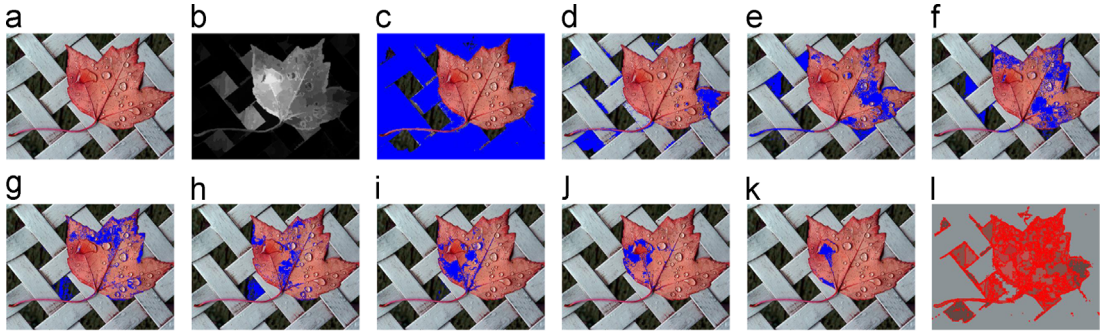

Fig. 3. Illustration of the process of clustering using saliency driven K-means. (a) Original image; (b) the combined saliency map; (c)–(k) are the separated nine regions (marked in blue); (l) clustering map, different clusters are labeled different colors. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

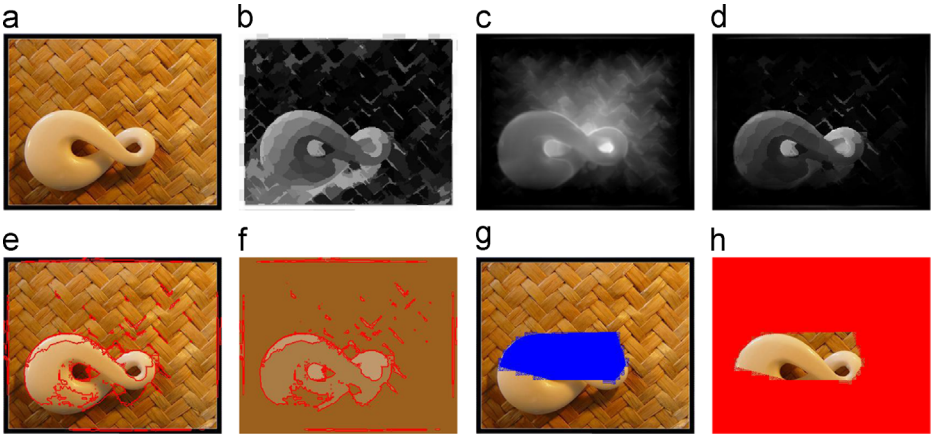

Fig. 4. Illustration of the process for region selection. Brighter pixels represent higher measurement values. (a) Original image; (b) color contrast saliency; (c) boundary prior saliency; (d) S^{cb}; (e) the boundary of regions; (f) to display regions with different colors; (j) the pixels in region *PSR* are labeled blue; and (h) the pixels in region *BK* are labeled red. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

4.2. Definition of regional saliency values

Regional color contrast saliency: The regional color contrast saliency is defined as

$$color(i) = \frac{\sum_{x \in RS_i} S_x^{con}}{|RS_i|}, \quad i \in \{1, \ldots, K\}. \tag{8}$$

Regions with higher average saliency values are more likely to be contained in a salient object.

Regional boundary prior saliency: The Regional boundary saliency of region RS_i is defined as

$$bound(i) = \frac{\sum_{x \in RS_i} S_x^{Boundary}}{|RS_i|}, \quad i \in \{1, \ldots, K\}. \tag{9}$$

Color spatial distribution: The color spatial distribution $csd(i)$ of region RS_i describes the spatial distribution of the color information contained in region RS_i. It is computed as the spatial variance of RS_i's color distribution.

$$V_h(i) = \frac{1}{|X|_i} \sum_{x=1}^{N} V(I_x|l = RS_i)|x_h - S_h(i)|^2$$
$$S_h(i) = \frac{1}{|X|_i} \sum_{x=1}^{N} V(I_x|l = RS_i) \cdot x_h, \tag{10}$$

where x_h is the x-coordinate of pixel x and $|X|_i = \sum_{x=1}^{N} V(I_x|l = RS_i)$. The saliency value is interpreted as how widely the pixels are distributed. $V_h(i)$ is the horizonal variance of the spatial position of pixels in the image. The vertical variance $V_v(i)$ is similarly defined. $VP(i) = V_h(i) + V_v(i)$, and VP is normalized to [0,1] for all regions. Then, the color spatial distribution of region RS_i is defined as

$$csd(i) = 1 - VP(i). \tag{11}$$

The pixel level saliency values are defined according to the regional saliency values *color*, *bound* and *csd*

$$M_{color}(x) = color(j), \quad x \in RS_j, i = [1, 2, \ldots, N],$$
$$M_{bound}(x) = bound(j), \quad x \in RS_j, i = [1, 2, \ldots, N],$$
$$M_{csd}(x) = csd(j), \quad x \in RS_j, i = [1, 2, \ldots, N], \tag{12}$$

where N is the size of image. Then a pixel level weight map $W_{map} = M_{color} * M_{csd} * M_{bound}$ is constructed by combining three pixel level saliency maps nonlinearly and W_{map} is normalized into [0, 1], $W_{map} = W_{map} - \min(W_{map}) / \max(W_{map}) - \min(W_{map})$. We take W_{map} as a prior probability inferred by contrast prior, color distribution prior and boundary prior. We define the set of potential salient pixels as *PSR* and the set of background pixels as *BK*. P_{PSR} represents the prior probability with respect to potential salient regions and $P_{PSR}(I_p) = W_{map}(p)$. P_{BK} is defined as the prior probability with respect to background and $P_{BK}(I_p) = 1 - P_{PSR}(I_p)$.

For the image in Fig. 3, it is separated into nine regions and the computed three measures for nine regions are

$$color = [0.1597, 0.6226, 0.6662, 0.6673, 0.7313, 0.7376, \times 0.7566, 0.8759, 0.9456],$$
$$bound = [0.0947, 0.2579, 0.4273, 0.5568, 0.6299, 0.7549, \times 0.8618, 0.8660, 0.9412],$$
$$csd = [0.8080, 0.8007, 0.8385, 0.9062, 0.9154, 0.8930, \times 0.9324, 0.9518, 0.9540], \tag{13}$$

4.3. Region selection

The performance of the statistic information extraction is dependent on the correctness of the GMMs to model foreground and background objects. Then, we propose a method for selecting foreground and background regions. The image space is divided into background region and potential salient region using adaptive threshold selection. The adaptive threshold λ is determined using OTSU [41] and threshold λ controls the process of region selection. We define *SR* and *SB* as two sets of pixels' indices, $SB = \{i | W_{map} < \lambda\}$ consists of the indices of background pixels and $SR = \{i | W_{map} \geq \lambda\}$ consists of the indices of potential salient pixels. We compute a convex hull C to enclose all the potential salient pixels in *SR*. The initial set of salient pixels is computed as $PSR = \{i | i \in C\}$ and the related set of background pixels is $BK = \{i | i \notin C\}$. The procedure of region selection is displayed in Fig. 4.

5. Statistic saliency generation

In the process of our algorithm, two GMM models shown in Eq. (7) are trained over the pixels in set *PSR* and *BK* respectively. The pixels in *PSR* tend to be contained by a salient object while pixels in *BK* are more likely to be part of the background. The pixels' similarity with salient region *PSR* is defined as $P_{gmm,s}(I_x|PSR)=V(I_x|l=PSR)$. Similarly, the similarity with background region is $P_{gmm,s}(I_x|BK)=V(I_x|l=BK)$. For pixel p, the likelihood is determined by its similarity to the pixels in salient region and its difference from the pixels in background region. The normalized likelihood probability which expresses how probable that the observed data is salient on pixel p is

$$P_{likli}(I_p|PSR)=\frac{P_{gmm,s}(I_p|PSR)}{P_{gmm,s}(I_p|PSR)+P_{gmm,s}(I_p|BK)}. \tag{14}$$

Similarly, $P_{likli}(I_p|BK)=1-P_{likli}(I_p|PSR)$. From the Bayesian framework, the posterior probability on pixel p is

$$P_{posterior}(PSR|I_p)=\frac{P_{likli}(I_p|PSR)*P_{PSR}(I_p)}{Z}, \tag{15}$$

where Z is the normalization constant, which ensures that the posterior distribution on the left-hand side is a valid probability density and integrates to one. $Z=P_{likli}(I_p|PSR)*P_{PSR}(I_p)+P_{likli}(I_p|BK)*P_{BK}(I_p)$. Finally, the statistic saliency (also called GMM saliency) is defined as

$$S^{gmm}(I_p)=P_{posterior}(PSR|I_p). \tag{16}$$

6. Experiments

The empirical analysis is implemented on two popular saliency databases: MSRA-1000 [42] and Berkley-300 database [43]. For quantitative comparison, the precision and recall rates of various models are evaluated. For a given threshold T, the *precision* and *recall* rates of a certain saliency detection method are defined as

$$Precision(T)=\frac{1}{INUM}\sum_{i=1}^{INUM}\frac{|M_i(T)\cap G_i|}{|M_i(T)|},$$
$$Recall(T)=\frac{1}{INUM}\sum_{i=1}^{INUM}\frac{|M_i(T)\cap G_i|}{|G_i|}. \tag{17}$$

In Eq. (17), $M_i(T)$ is the binary mask obtained by directly thresholding the saliency map using threshold T on the i-th image. G_i is the ground truth. $|\cdot|$ denotes the mask's sum area. *INUM* is the amount of images in a database. In addition to precision–recall (PR) curves, for each image, we follow [42,15] to segment a saliency map by adaptive threshold

$$T_s=\min\left\{2\times\frac{\sum_i^N V_i}{N}, T_{max}\right\}, \tag{18}$$

where N denotes the number of pixels in the saliency map and i is the pixel index. V_i is the saliency value on pixel i. T_{max} is the upper bound for the saliency value. In the experiment, the saliency values are projected into the range of [0,255] and we set $T_{max}=255$ in the experiment. Then the precision, recall and *F*-Measurement values are computed over the ground truth maps, where *F*-Measurement is defined as

$$F_\beta(T)=\frac{(1+\beta^2)Precision(T)\times Recall(T)}{\beta^2\times Precision(T)+Recall(T)}, \tag{19}$$

where $\beta=0.3$. Besides, we evaluate the strategy of saliency driven clustering (introduced in Section 3) and the regional saliency measures (introduced in Section 4).

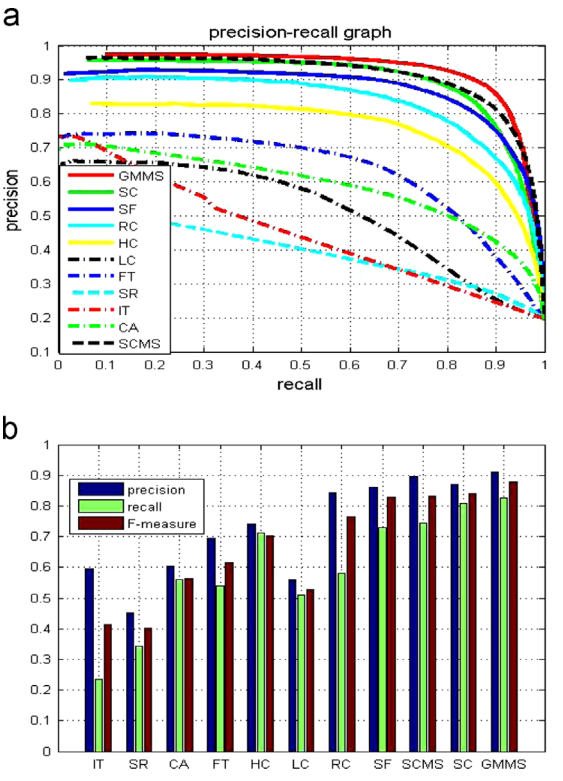

Fig. 5. Experimental results on MSRA-1000 database. (a) Average precision–recall curves using different approaches and (b) average precision, recall and *F*-measure using different approaches with adaptive thresholding.

6.1. Evaluation on MSRA-1000 dataset

First, we generate the saliency maps for all 1000 testing images using the proposed saliency model. The saliency detection performance of proposed saliency model (with abbreviation GMMS) is compared with nine state-of-the-art saliency models, that are IT (Ittis model)) [9], LC (Luminance Contrast) [44], SR (Spectral Residual) [23], FT (Frequency-tuned) [42], HC (Histogram Contrast) [16], RC (Region Contrast) [16], SF (Saliency Filters) [15], CA (Context-awear model) [22], SC (Superpixel-based Contrast) [17] and SCMS (spatial-color constraint and multi-scale segmentation) [45] for comparison.

The saliency maps of the state-of-the-art works excluding SC, SF and SCMS are provided in [16].[1] The SF [15] saliency maps are obtained from the author's webpage.[2] The SCMS [45] saliency maps are downloaded from the

[1] http://cg.cs.tsinghua.edu.cn/people/~cmm/Saliency/Index.htm
[2] http://www.fedeperazzi.com/saliency_filters/

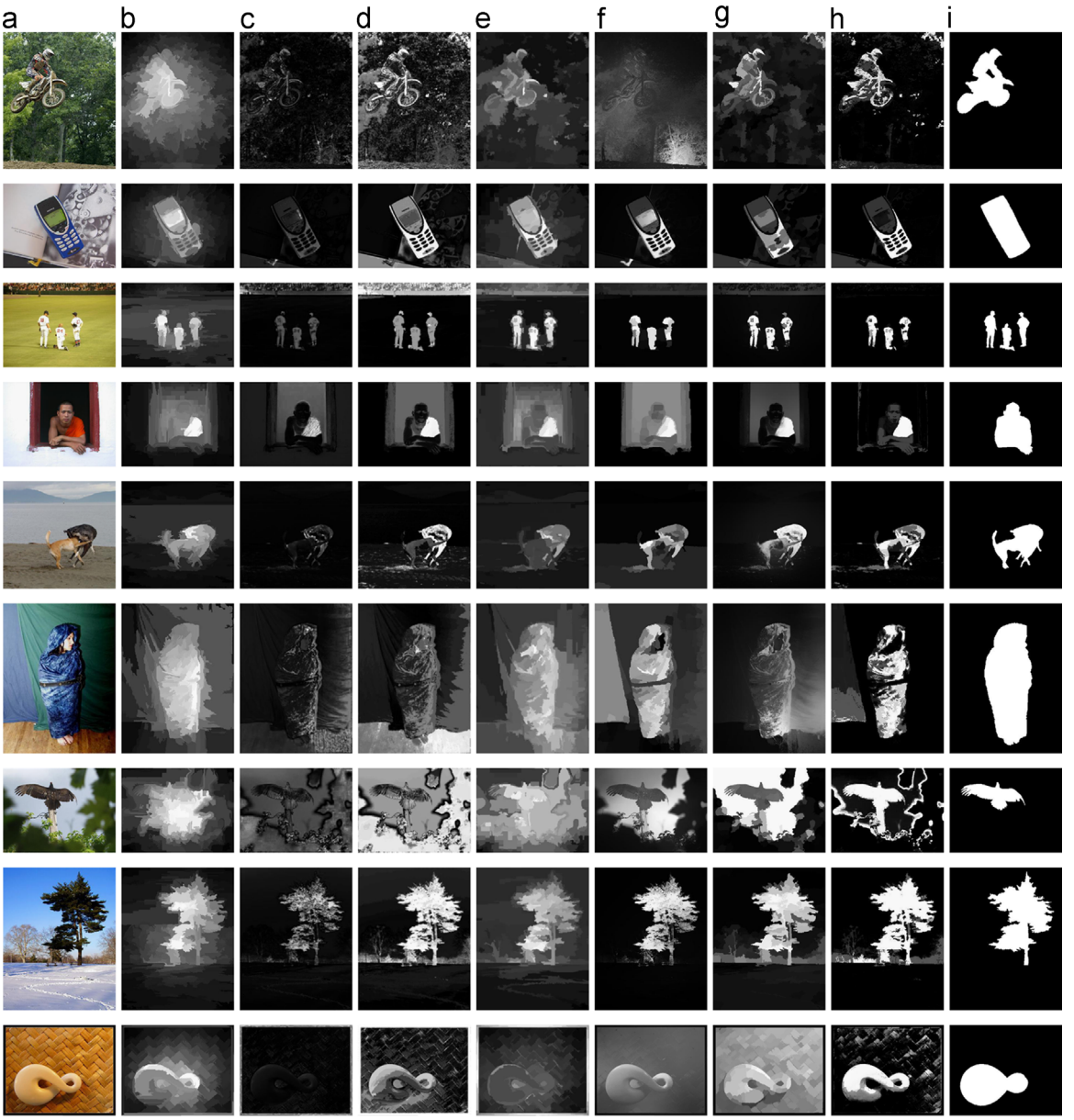

Fig. 6. Subjective comparison of our saliency map with six state-of-the-art methods on the MSRA database. (a) Original image; saliency maps of (b) SCMS; (c) FT; (d) HC; (e) RC; (f) SF; (g) SC; (h) GMMS and (I) ground truth. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

author's webpage.[3] The results of SC are generated using the codes provided by author. In Fig. 5(a), we compare our precision–recall curve with other methods. Compared with other approaches, the proposed GMMS demonstrates the highest precision level corresponding to all the recall rates ranging from 0 to 1. Fig. 5(b) displays the average precision, recall and *F*-measure values in the adaptive threshold experiment. Among all the approaches, GMMS achieves the highest precision, recall and *F*-measure values compared with other approaches. Besides the quantitative evaluation, our method is visually compared with six methods, i.e, SCMS, HC, RC, SR, SF and SC, and the results of some testing images are displayed in Fig. 6. Brighter pixels indicate higher saliency probabilities. Visually, it can be seen that our GMMS obtains relatively higher quality saliency maps compared with the compared state-of-the-art methods. The RC model sometimes highlights only some parts of an salient object.

The methods like HC and FT are sensitive to background noise and they often fail to identify background patches correctly. Compared with those methods, GMMS achieves better performance. For images with clear contrast between the salient object and the background (see the

[3] http://ivipc.uestc.edu.cn/lfxu/

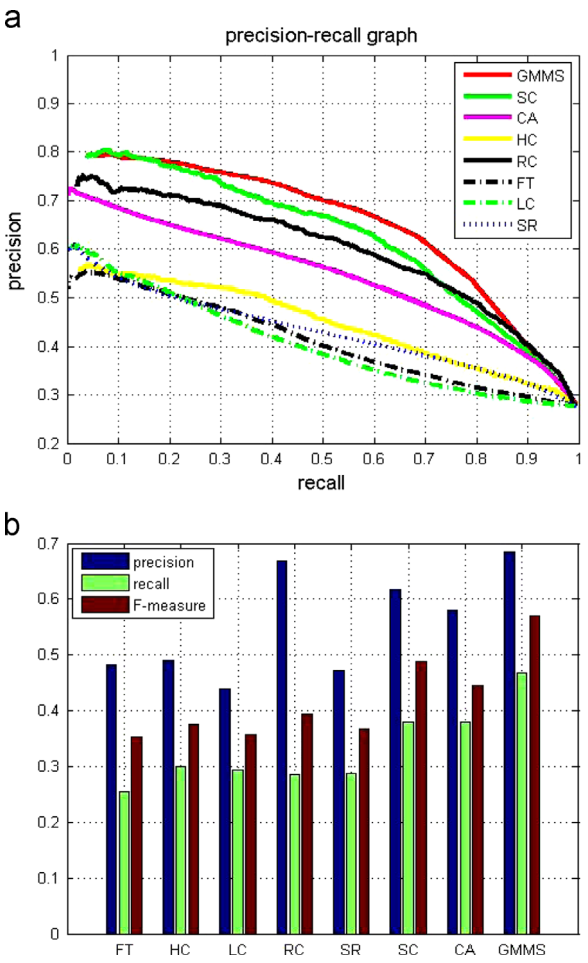

Fig. 7. Experimental results on Berkley-300 database. (a) Average precision–recall curves using different approaches and (b) average precision, recall and *F*-measure using different approaches with adaptive thresholding.

examples in the 3rd and 8th rows in Fig. 6), the salient objects can be located well in most of the saliency maps. For images with complex background scenes containing structure or texture pattern or images with relatively low contrast boundary between salient objects and the surrounding background regions (see the 1st, 6th 7th and 9th example in Fig. 6), the quality of color contrast based model (HC, RC, SCMS and SC) is obviously degraded. In contrast, GMMS can locate and highlight the object region more correctly, and suppress background noise correctly by utilizing the Bayesian framework integrating the region based saliency values as priority. For example, in the 6th picture listed in Fig. 6, our model can label the boundary correctly although both of the background and object have blue color. Because we integrate the boundary priority into process of clustering and the structural information (the generated clusters) may provide more reliable cues about "what the object looks like", even though the pixels near the boundary have similar color with the object. The methods (such as HC, RC) that use only color features (like color contrast) are sensitive to background noise in a complex scene (see the 1st and 3th example images in Fig. 6) and GMM saliency can generate more stable and reliable saliency information in cluttered background. However, some parts of the salient object may be assigned relatively low saliency probabilities and hence the salient objects are incompletely extracted (see the 2nd example, where saliency values in black regions of the cell phone are computed incorrectly because the black color is not the dominant color in the statistic model). The results can be further improved by refining the GMM saliency maps, so as to obtain more smoother saliency maps.

6.2. Evaluation on Berkley-300 database

The Berkley-300 database is a more challenging database which contained 300 images with more complex background or multiple objects of different sizes and positions. The foreground masks are provided by [18] as the ground truth. We compare the curve of our approach with LC [44], SR [23], FT [42], HC [16], RC [16], CA [22] and SC [17]. The PR curves are shown in Fig. 7(a) and the average precision, recall and *F*-measure values are displayed in Fig. 7(b). GMMS achieves the best performance both in the terms of PR curve and the adaptive segmentation experiments. It is observed from the visual comparison in Fig. 8 that GMMS performs better in highlighting salient objects and suppressing background clutter under various condition, such as for images with texture (the 6th and 7th examples), images with weak boundary (the 5th and 10th examples), image with small objects (the 9th example) or color salient objects which are similar to part of image's background (the 3rd and 4th examples) (Fig. 9).

6.3. Evaluation of saliency driven clusters

From the aspect of saliency computation, good clusters should generate satisfactory separation of salient objects and background. To quantitatively evaluate the cluster results, we report our method's scores related to two criteria (1) Variation of Information (VoI) [46], computing the information of one results not contained in the other; (2) Global Consistency Error (GCE) [47], measuring the extent to which on segment is a refinement of the other. The results are obtained on Berkley Segmentation Database [47]. We compare the average scores of our cluster strategy (with abbreviation SDC) with Ncut [48], Meanshift [49], Normalized Tree Partitioning (NTP) [50] and JSEG. The scores are listed in Table 1 and SDC obtains the lowest VoI with a value of 1.8360 and the second lowest GCE with a value of 0.1955 across the test dataset, which means that SDC can separate pixels into different clusters with high precision according to their saliency level more correctly. Moreover, it costs less than 1 s for SDC to process a typical 400×300 image.

6.4. The role of measures for regional saliency computation

Three regional saliency measures based on the generated clusters, the regional color contrast saliency Eq. (8), regional boundary prior saliency Eq. (9) and color spatial distribution Eq. (10) are computed. We have explored the effectiveness of each individual measure and the

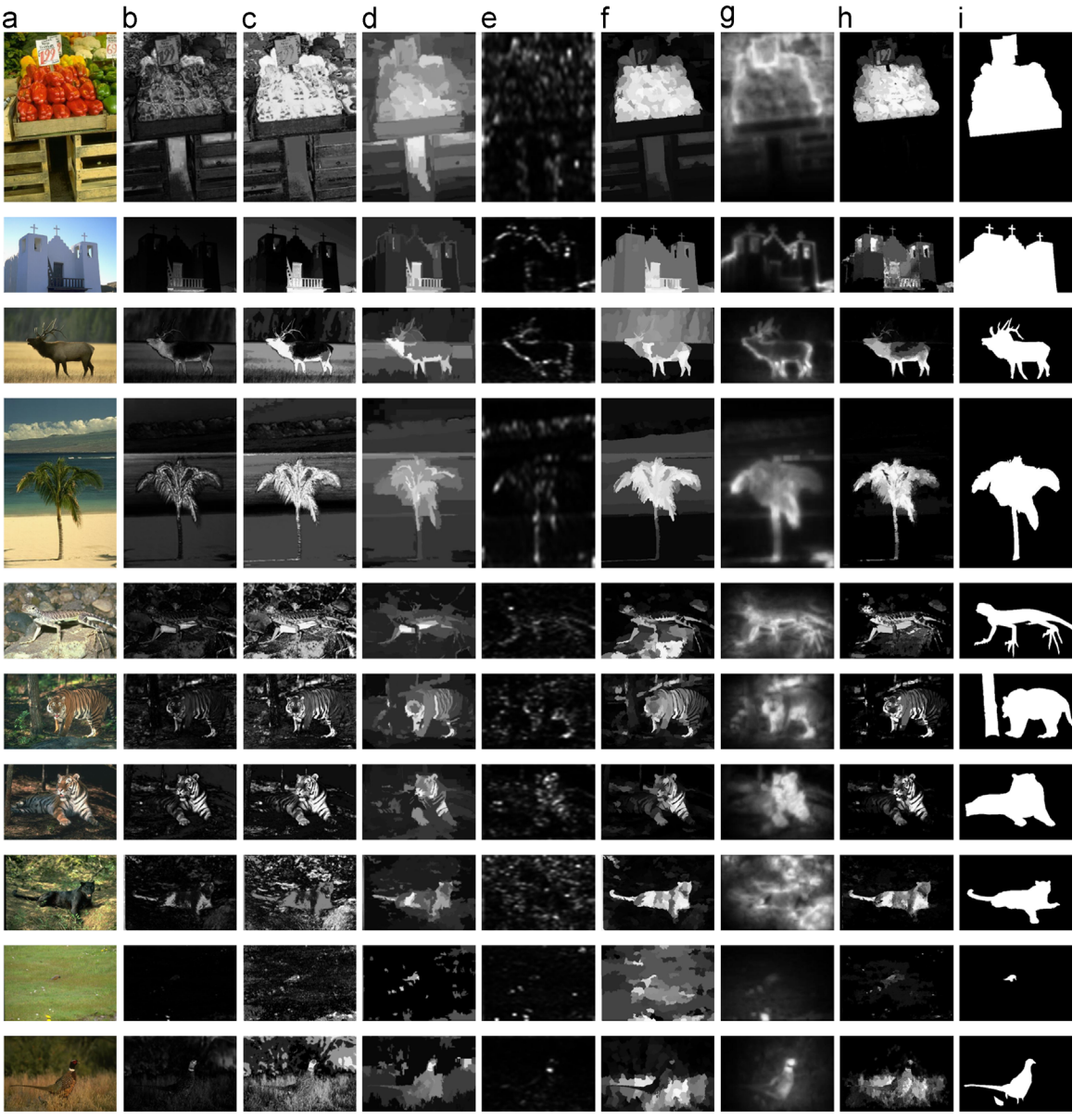

Fig. 8. Subjective comparison of our saliency map with six state-of-the-art methods on the Berkley-300 database. (a) Original images; (b) FT; (c) HC; (d) RC; (e) SR; (f) SC; (g) CA; (h) GMMS; and (i) the ground truth.

Fig. 9. Illustration of the clusters generated. The images with cluster boundaries are listed in the first row. In the second row, clusters are labeled with different colors. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

combination strategies on MASA-1000 dataset. The PR curves for comparison are displayed in Fig. 10, and some meaningful conclusions can be drawn. On one hand, the regional computation over clusters can lead to performance improvement for color contrast saliency and boundary prior saliency Fig. 10(a). The improvement mainly lies in the integration of visual patterns of images.

On the other hand, the non-linear combinations between $M_{col}, M_{bound}, M_{csd}$ are evaluated as well. The performance of four kinds of combinations , $M_{col}*M_{bound}$, $M_{col}*M_{csd}$, $M_{bound}*M_{csd}$ and $M_{col}*M_{bound}*M_{csd}$ are evaluated on MSRA-1000 dataset and the results are listed in Fig. 10(b). The combination of $M_{bound}* M_{csd}$ can achieve significant improvement over the individual component. Even if the combinations $M_{col}* M_{bound}$, $M_{col}* M_{csd}$ and $M_{col}* M_{bound}* M_{csd}$ only improve the performance slightly over that of M_{col} by analyzing the ROC curves, the complementary strengths of each saliency map would contribute to generate better performance. Fig. 11 shows some combination results visually. The combination can improve the performance over any of the individual measure. In summary, the improvement mainly lies in that the saliency measures can either suppress background clutters or highlight salient objects, which could be explained by the examples listed in Fig. 11. First, the color spatial distribution and boundary prior will contribute to suppress the widely distributed background clutters. As shown in images 2, 5, 6 and 8 of Fig. 11, the widely distributed background noise tends to have lower *csd* value (Eq. (11)). Moreover, the background clutters are mostly likely to be near the boundary and the

Table 1
The VOI and GCE scores for evaluating the generated clusters.

Method	VoI	GCE
Ncut	2.9061	0.2232
Mean-shift	1.9725	**0.1888**
NTP	2.4954	0.2373
JSEG	2.3217	0.1989
Proposed SDC	**1.8360**	0.1955

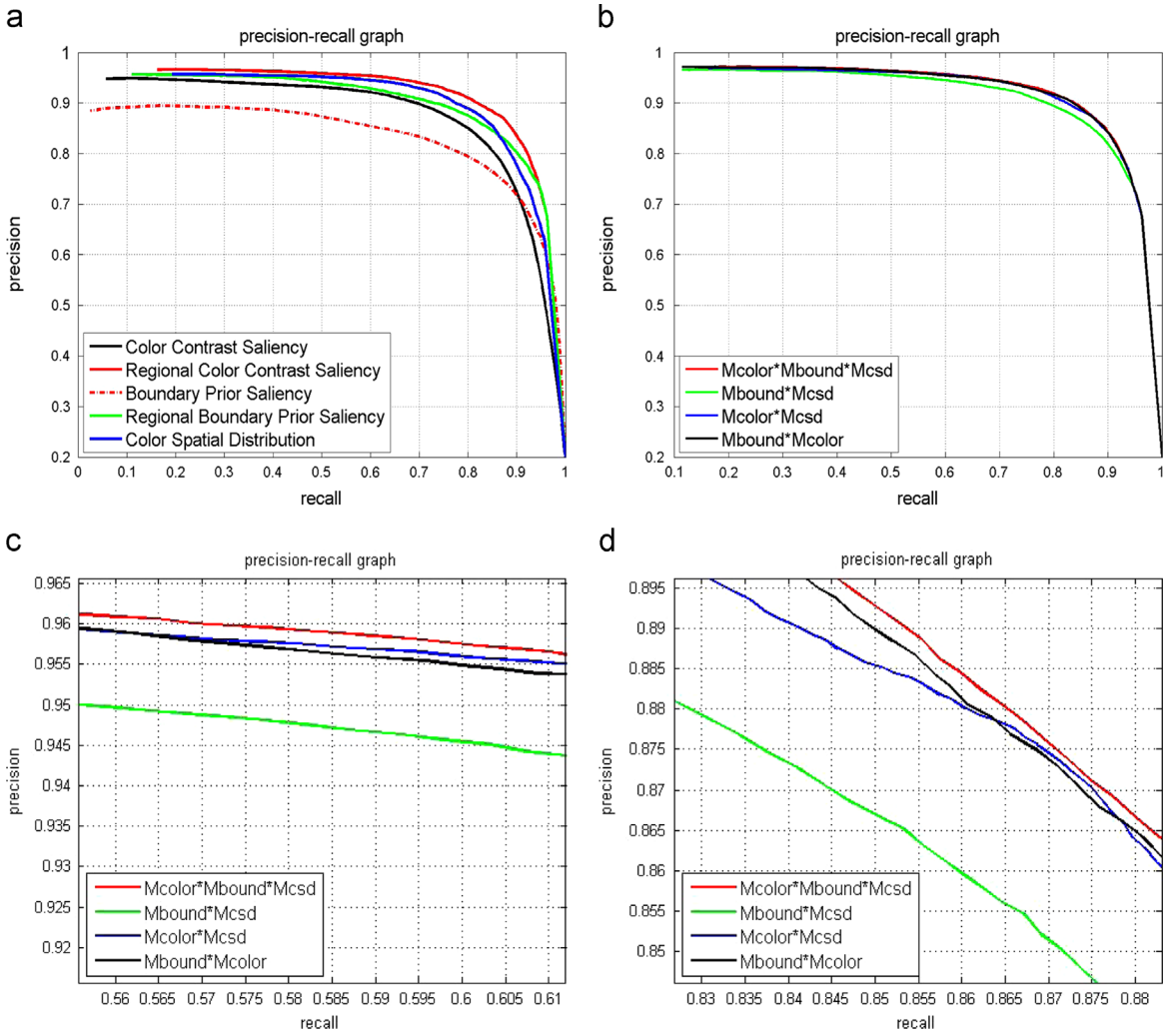

Fig. 10. Comparison of the individual component on MSRA dataset. (a) Comparison of three measurements for regional saliency computation. (b) Comparison of the combination strategy. (c) Enlarged image of subfigure (b) in low recall region. (d) Enlarged image of subfigure (b) in high recall region.

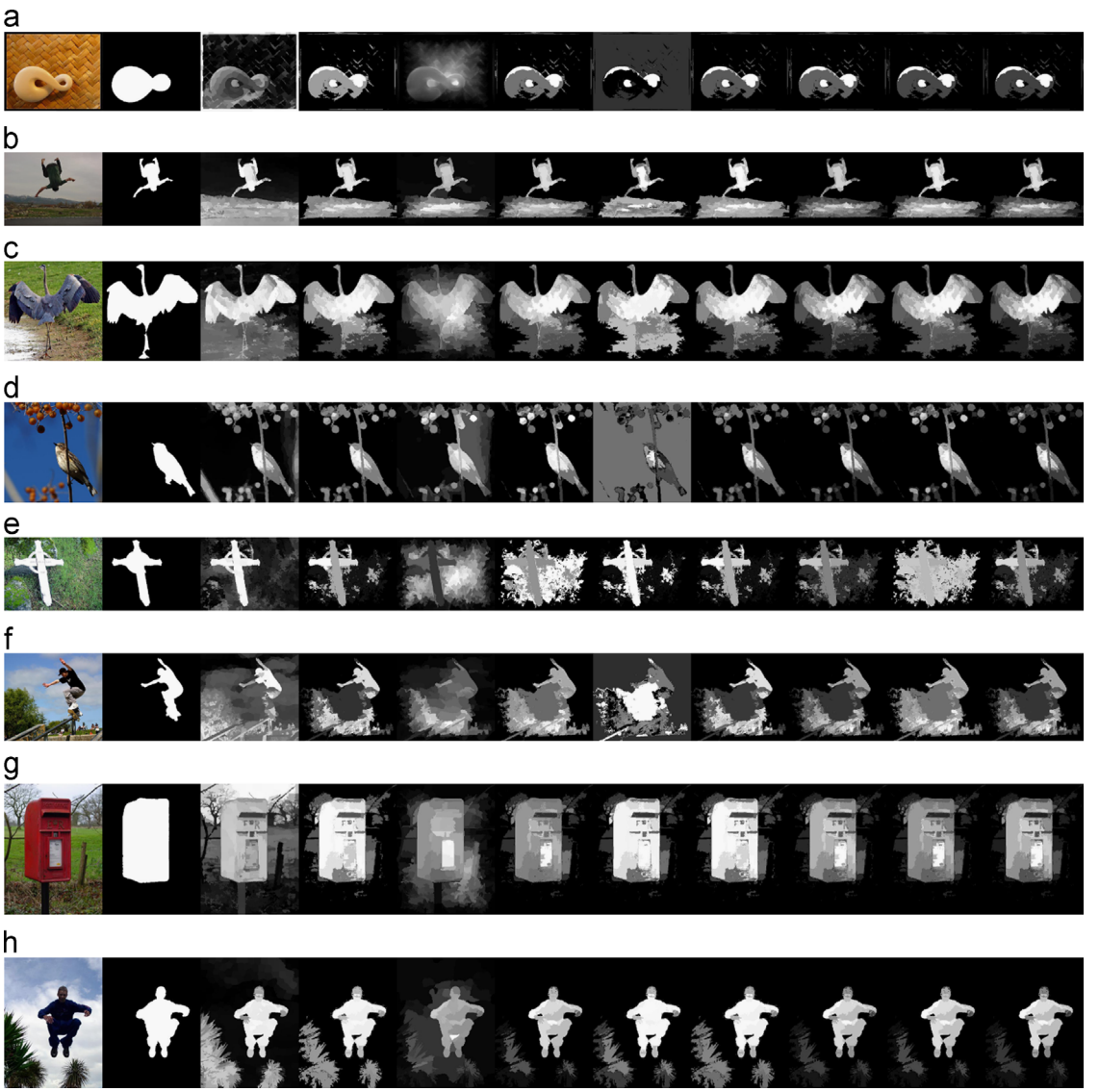

Fig. 11. Illustration of three measures for regional saliency computation. From left to right: original image; ground truth; color contrast saliency; regional color saliency; boundary prior saliency; regional boundary saliency; color spatial distribution; combination of M_{color}*M_{csd}, M_{bound}*M_{color}, M_{bound}*M_{csd}, and the combined of three measures. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

boundary prior we use can contribute to suppress the background noise. The results of Image 2 and Image 6 show that the background clutters can be suppressing by combining boundary prior. In image 5, the boundary prior saliency fails to locate the object, however the nonlinear combination of three measures can still generate satisfactory result. Second, even if the performance of the computed color spatial distribution is not robust for images with complex background scenes, it still owns some characteristics like highlighting the salient objects or enhancing the saliency difference between salient objects and backgrounds (image 1, image 7 and image 8 in Fig. 11).

The contributions of extracted visual patterns for saliency performance improvement can be summarized in two aspects. On one hand, the regional computation of color contrast saliency and boundary prior saliency can achieve considerable performance improvement over individual measures (Fig. 10.(a)). On the other hand, the color spatial distribution which is calculated over the statistic of clusters provides meaningful complementary cues for salient object detection and background clutters suppression (visual examples listed in Fig. 11).

6.5. Computation cost

It takes about 9.3 s for our method to process a typical 400×300 image on our 2.7 GHZ Pentium Dual-Core machine with 2 G RAM. The computation of color contrast saliency costs 2.2 s, which respectively takes 1 s for SLIC segmentation and 1.2 s for superpixel-based saliency computation Eq. (1). Computation of the boundary prior saliency and combination takes about 0.8 s. The saliency driven clustering takes about 1.3 s. It costs 4.5 s to compute the regional color contrast saliency, color spatial distribution and regional boundary prior saliency. The most time-consuming step is the computation of

GMM parameters Eq. (7) for color spatial distribution, it costs about 0.4 s to estimate the parameters for a cluster. Generally, it takes about 4 s for an image separated into 10 clusters. The final step of GMM saliency computation Eq. (16) only costs 0.5 s.

7. Conclusion

We have presented a saliency detection framework by modeling "what the salient objects look like" and "what the background should look like" using statistic of images. To exhibit diverse and meaningful visual patterns information of natural images, we propose a saliency driven clustering method based on the combination of contrast prior saliency and boundary prior saliency. To incorporate the visual pattern information of images into saliency model, the clusters are used for computing the color spatial distribution, region based color contrast saliency values and region based boundary prior saliency values. Then, the salient region GMM and background GMM are constructed based on the separated salient regions and background region using adaptive threshold which is computed over the combined regional saliency values. Finally, a Bayesian model is applied for generating high quality full resolution saliency maps. Experimental results on the most popular datasets indicate the advantages of our method against other state-of-the-art approaches on highlighting salient objects and suppressing the cluttered background. The comparison experiments also indicate the advantages of building a saliency model based on visual patterns information of images, such as clusters. Since a simple clustering method based on color contrast and boundary prior saliency map is used for saliency detection, we will exploit a more effective clustering strategy for generating semantic regions in our future work.

Acknowledgments

This research is partly supported by NSFC, China (No: 61273258, 61375048), Ph.D. Programs Foundation of Ministry of Education of China (No. 20120073110018).

References

[1] H. Fu, Z. Chi, D. Feng, Attention-driven image interpretation with application to image retrieval, Pattern Recognit. 39 (9) (2006) 1604–1621.
[2] L. Itti, Automatic foveation for video compression using a neurobiological model of visual attention, IEEE Trans. Image Process. 13 (10) (2004) 1304–1318.
[3] D. Walther, C. Koch, Modeling attention to salient proto-objects, Neural Netw. 19 (9) (2006) 1395–1407.
[4] A. Oliva, A. Torralba, et al., Trends Cogn. Sci. 11 (12) (2007) 520–527.
[5] P.L. Rosin, A simple method for detecting salient regions, Pattern Recognit. 42 (11) (2009) 2363–2371.
[6] J. Qin, N.H. Yung, Scene categorization via contextual visual words, Pattern Recognit. 43 (5) (2010) 1874–1888.
[7] Y.T. Wu, F.Y. Shih, J. Shi, Y.T. Wu, A top-down region dividing approach for image segmentation, Pattern Recognit. 41 (6) (2008) 1948–1960.
[8] J. Xue, L. Wang, N. Zheng, G. Hua, Automatic salient object extraction with contextual cue and its applications to recognition and alpha matting. 〈http://www.sciencedirect.com/science/article/pii/S0031320313001581〉, 2013.
[9] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259.
[10] W. Einhäuser, P. Koènig, Does luminance-contrast contribute to a saliency map for overt visual attention? Eur. J. Neurosci. 17 (5) (2003) 1089–1097.
[11] D. Parkhurst, K. Law, E. Niebur, et al., Modeling the role of salience in the allocation of overt visual attention, Vis. Res. 42 (1) (2002) 107–124.
[12] L. Itti, N. Dhavale, F. Pighin, Realistic avatar eye and head animation using a neurobiological model of visual attention, in: SPIE's 48th Annual Meeting, International Society for Optics and Photonics, Optical Science and Technology, 2004, pp. 64–78.
[13] Le.O. Meur, Le.P. Callet, D. Barba, D. Thoreau, A coherent computational approach to model bottom-up visual attention, IEEE Trans. Pattern Anal. Mach. Intell. 28 (5) (2006) 802–817.
[14] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, in: 2006 Conference on Neural Information Processing Systems (NIPS), 2006.
[15] F. Perazzi, P. Krahenbuhl, Y. Pritch, A. Hornung, Saliency filters: contrast based filtering for salient region detection, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2012, pp. 733–740.
[16] M.M. Cheng, G.X. Zhang, N.J. Mitra, X. Huang, S.M. Hu, Global contrast based salient region detection, in: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2011, pp. 409–416.
[17] K. Fu, C. Gong, J. Yang, Y. Zhou, Salient object detection via color contrast and color distribution, in: 2012 IEEE 11th Asian Conference on Computer Vision (ACCV), IEEE, 2012.
[18] Y. Wei, F. Wen, W. Zhu, J. Sun, geodesic saliency using background priors, in: Computer Vision–ECCV 2012, Springer, 2012, pp. 29–42.
[19] W. Zhang, Q. Wu, G. Wang, H. Yin, An adaptive computational model for salient object detection, IEEE Trans. Multimed. 12 (4) (2010) 300–316.
[20] Y. Xie, H. Lu, M. Yang, Bayesian saliency via low and mid level cues, IEEE Trans. Image Process. 22 (5) (2012) 1689–1698.
[21] Z. Liang, Z. Chi, H. Fu, D. Feng, Salient object detection using content-sensitive hypergraph representation and partitioning, Pattern Recognit. 45 (11) (2012) 3886–3901.
[22] S. Goferman, L. Zelnik-Manor, A. Tal, Context-aware saliency detection, IEEE Trans. Pattern Anal. Mach. Intell. 34 (10) (2012) 1915–1926.
[23] X. Hou, L. Zhang, Saliency detection: a spectral residual approach, in: 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2007, pp. 1–8.
[24] C. Guo, L. Zhang, A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression, IEEE Trans. Image Process. 19 (1) (2010) 185–198.
[25] N.D. Bruce, Features that draw visual attention: an information theoretic perspective, Neurocomputing 65–66 (2005) 125–133.
[26] L. Itti, P. Baldi, Bayesian surprise attracts human attention, Vis. Res. 49 (10) (2009) 1295–1306.
[27] D.A. Klein, S. Frintrop, Center-surround divergence of feature statistics for salient object detection, in: 2011 IEEE International Conference on Computer Vision (ICCV), IEEE, 2011, pp. 2214–2219.
[28] W. Hou, X. Gao, D. Tao, X. Li, Visual saliency detection using information divergence, Pattern Recognit. 46 (10) (2013) 2658–2669.
[29] V. Gopalakrishnan, Y. Hu, D. Rajan, Salient region detection by modeling distributions of color and orientation, IEEE Trans. Multimed. 11 (5) (2009) 892–905.
[30] Z. Liu, R. Shi, L. Shen, Y. Xue, K.N. Ngan, Z. Zhang, Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut, IEEE Trans. Multimed. 14 (4) (2012) 1275–1289.
[31] L. Zhang, M.H. Tong, T.K. Marks, H. Shan, G.W. Cottrell, Sun: a Bayesian framework for saliency using natural statistics, J. Vis. 8 (7) (2008) 1–7.
[32] X.P. Hu, L. Dempere-Marco, E.R. Davies, Bayesian feature evaluation for visual saliency estimation, Pattern Recognit. 41 (11) (2008) 3302–3312.
[33] W.F. Noh, P. Woodward, Slic (simple line interface calculation), in: Proceedings of the Fifth International Conference on Numerical Methods in Fluid Dynamics, June 28–July 2, Twente University, Enschede, Springer, 1976, pp. 330–340.
[34] Y. Wei, F. Wen, W. Zhu, J. Sun, Geodesic saliency using background priors, in: 2012 Europe Conference on Computer Vision, Springer, 2012, pp. 29–42.
[35] V. Gulshan, C. Rother, A. Criminisi, A. Blake, A. Zisserman, Geodesic star convexity for interactive image segmentation, in: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2010, pp. 3129–3136.

[36] L. Yatziv, A. Bartesaghi, G. Sapiro, O(n) implementation of the fast marching algorithm, J. Comput. Phys. 212 (2) (2006) 393–399.
[37] J.A. Hartigan, M.A. Wong, Algorithm as 136: a k-means clustering algorithm, J. R. Stat. Soc. Ser. C (Appl. Stat.) 28 (1) (1979) 100–108.
[38] T. Ohashi, Z. Aghbari, A. Makinouchi, Hill-climbing algorithm for efficient color-based image segmentation, in: IASTED International Conference on Signal Processing, Pattern Recognition, and Applications, 2003, pp. 17–22.
[39] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, H.Y. Shum, Learning to detect a salient object, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2) (2011) 353–367.
[40] T.K. Moon, The expectation-maximization algorithm, IEEE Signal Process. Mag. 13 (6) (1996) 47–60.
[41] N. Otsu, A threshold selection method from gray-level histograms, Automatica 11 (285–296) (1975) 23–27.
[42] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2009, pp. 1597–1604.
[43] V. Movahedi, J.H. Elder, Design and perceptual validation of performance measures for salient object segmentation, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2010, pp. 49–56.
[44] Y. Zhai, M. Shah, Visual attention detection in video sequences using spatiotemporal cues, in: Proceedings of the 14th Annual ACM International Conference on Multimedia, ACM, 2006, pp. 815–824.
[45] L. Xu, H. Li, L. Zeng, K.N. Ngan, Saliency detection using joint spatial-color constraint and multi-scale segmentation, J. Vis. Commun. Image Represent 24 (4) (2014) 465–476.
[46] M. Meila, Comparing clusterings: an axiomatic view, in: Proceedings of the 22nd International Conference on Machine Learning, ACM, 2005, pp. 577–584.
[47] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proceedings of Eighth IEEE International Conference on Computer Vision, ICCV 2001, vol. 2, IEEE, 2001, pp. 416–423.
[48] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (8) (2000) 888–905.
[49] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, IEEE Trans. Pattern Anal. Mach. Intell. 24 (5) (2002) 603–619.
[50] J. Wang, Y. Jia, X.S. Hua, C. Zhang, L. Quan, Normalized tree partitioning for image segmentation, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008, CVPR 2008, IEEE, 2008, pp. 1–8.