Neural Process Lett (2013) 38:361–374
DOI 10.1007/s11063-012-9276-3

A New Framework for Multiscale Saliency Detection Based on Image Patches

Jingbo Zhou · Zhong Jin

Published online: 8 January 2013

Abstract In this paper, we propose a new multiscale saliency detection algorithm based on image patches. To measure saliency of pixels in a given image, we segment the image into patches by a fixed scale and then use principal component analysis to reduce the dimensions which are noises with respect to the saliency calculation. The dissimilarities between a patch and other patches, which indicate the patch's saliency, are computed based on the dissimilarity of colors and the spatial distance. Finally, we implement our algorithm through multiple scales that further decrease the saliency of background. Our method is compared with other saliency detection approaches on two public image datasets. Experimental results show that our method outperforms the state-of-the-art methods on predicting human fixations and salient object segmentation.

Keywords Saliency detection · Multiscale · Principle component analysis · Object segmentation · Human fixation

1 Introduction

Humans can identify salient areas in their visual fields with surprising speed and accuracy before performing actual recognition. Computationally detecting such salient image regions remains a significant goal, as it allows preferential allocation of computational resources in subsequent image analysis and synthesis. There are many applications for visual attention, for example, automatic image cropping [1], adaptive image display on small devices [2], image/video compression, advertising design [3], and image collection browsing. Recent studies [4,5] demonstrated that visual attention helps object recognition, tracking, and detection as well.

J. Zhou (✉)· Z. Jin
School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China
e-mail: zhoujingbo2006@yahoo.com.cn

Visual attention analysis has generally progressed on two fronts: bottom-up and top-down approaches. Bottom-up approach, which is data-driven and task-independent, is a perception processing for automatic salient region selection for images. On the other hand, top-down approach is related to the recognition processing influenced by the prior knowledge such as tasks to be performed, the feature distribution of the target, the context of the visual scene and so on [6–8].

In this paper, we also focus on the bottom-up approaches. There exist several computational models for simulating human visual attention based on the bottom-up approaches. A representative work by Itti et al., which use a Difference of Gaussians approach to evaluate those features, is presented in [9]. The resulting saliency maps are generally blurry, and often overemphasize small, purely local features, which renders this approach less useful for applications such as segmentation and detection, etc. In frequency domain, frequency space methods [10,11] determine saliency based on the amplitude or phase spectrum of the Fourier transform of an image. The resulting saliency maps better preserve the high level structure of an image than [9], but exhibit undesirable blurriness and tend to highlight object boundaries rather than its entire area. For color space techniques one can distinguish between approaches using local or global analysis of contrast. Local methods estimate the saliency of a particular image region based on immediate image neighborhoods, e.g., based on dissimilarities at the pixel-level [12], using multiscale Difference of Gaussians [13] or histogram analysis [14]. While such approaches are able to produce less blurry saliency maps, they are agnostic of global relations and structures, and they may also be more sensitive to high frequency content like image edges and noise. Global methods take contrast relations over the complete image into account. For example, there are different variants of patch-based methods which estimate dissimilarity between image patches [14–16]. While these algorithms are more consistent in terms of global image structures, they suffer from the involved combinatorial complexity. The method of Achanta et al. [17] works on a per-pixel basis, but achieves globally more consistent results by computing color dissimilarities to the mean image color. They use Gaussian blur in order to decrease the influence of noise and high frequency patterns. However, their method does not account for any spatial relationship inside the image, and may highlight background regions as salient. Cheng et al. [18], who generate 3D histograms and compute dissimilarities between histogram bins, reported the best performing method among global contrast-based approaches so far. However, their method has problems handling images with cluttered and textured background.

Of all these works, the most related to ours is [19], since we also define the saliency in two elements: dissimilarity, spatial distance. However, we exploit multiple scales other than central bias, which was used in [19], to decrease the saliency of background patches and improve the contrast between salient and non-salient regions. In our proposed model, we first divide the input image into small image patches and measure the saliency value for each image patch through calculating the differences of color and spatial distance between this patch and all other patches in the image. But unlike [15], which define the pixel-level saliency by the contrast of different scales, we compute the saliency on patch-level which ease the computational burden. Similar to Cheng's method [18], we also compute the saliency map from global contrast perspective, which is more useful for applications to segmentation. The proposed method is different from them that we use PCA to extract the principle information of the image patches, which represent the background. Moreover, our algorithm calculates the saliency based on patches and don't need to segment original image before saliency detection.

The main contributions of our proposed model include the followings: (1) we propose to divide an image into small image patches for local information extraction and combine

information from different image patches in a global perspective; (2) we exploit PCA to reduce the dimensionality of each patch, which is important in our proposed algorithm to extract the meaningful spatial structure of the image; (3) we use a multiscale framework instead of center bias to compute the saliency map. The central bias, proposed by [20], is based on the principle that dominant objects often raise to the center of the image. This underlying hypothesis brings two problems. On one hand, background near the center of image may be more salient than the foreground which is far away from the center of the image. On the other hand, for a salient object, the part near the center of the input image is more salient than that far away from the center. The experiments justify the effectiveness of the proposed scheme with predicting human visual fixation and salient object segmentation.

The remainder of the paper is organized as follows: we state the framework of our saliency detection method in detail in Sect. 2. In Sect. 3, we demonstrate our experimental results based on two public image datasets and compare the results with other state-of-art saliency detection methods. The final section concludes the paper by summarizing our findings.

2 Proposed Saliency Algorithm

In this section, we will state the framework of our saliency detection method in detail. The steps of our algorithm are fourfold: representing the image patches, using PCA to reduce dimensionality, computing each patch's saliency value and implementing our method to multiple scales. We will describe the details step by step in the following subsections.

2.1 Image Patches Representation

The first step of our algorithm is to divide each original input image into small image patches to gather local information. For simplicity, we take image patches from the original image without overlapping. Given an image I with dimension $H \times W$, non-overlapping patches with the size of $n \times n$ pixels are drawn from it. Generally speaking, the size of the patches located in the bottom and right boundary is smaller than the regular size. To make sure all patches have the same dimensions for feature extraction by PCA, we throw out the border regions for simplicity which don't have regular size. The total number of patches is $L = \lfloor H/n \rfloor \cdot \lfloor W/n \rfloor$. Denote the patch as $p_i,\ i = 1, 2, \ldots, L$. Then each patch is represented as a column vector x_i of pixel values. The length of the vector is $3n^2$ since the color space has three components. Finally, we get a sample matrix $X = [x_1, x_2, \ldots, x_L]$, L is the total number of patches as stated above.

Next, we extract features by PCA based on the matrix X of image patches. Another reason why we detect the saliency object is related to the efficiency. The previous works [10,21,22] resize the original image to a smaller size in order to ease the heavy computational burden. And in [18], they ease the complexity by computing saliency map based on regions which are generated by mean shift. Since the number of image patches or regions in an image is far smaller than the number of pixels, computing saliency at image patches or regions level can significantly reduce the computation. Therefore, like [18], the proposed algorithm can also produce full-resolution saliency map.

2.2 Dimensionality Reduction

In our method, we use principal component analysis (PCA [23]) to reduce the dimensionality of each image patch which is represented as a vector. Principal components (PCs) throw out

dimensions that are noises with respect to the saliency calculation. Our proposed model is based on the conclusion in [24] that while saliency implies uniqueness, the opposite might not always be true. Therefore, the number of patches divided by dominant object is smaller than that of background since salient object is unique when compared to background. The features extracted by PCs according to few largest eigenvalues represent the principal directions of the features that come from the patches of background. As described in [25], these few PCs contribute to saliency detection because of their meaningful spatial structure. In our proposed model, we exploit the above-mentioned features that discriminate the salient object from the background in the original image. We project the patches to these PCs which not only throw out the dimensions that are noises with respect to the saliency calculation, but also throw out the features that can represent the salient object.

Specifically, to effectively describe patches in a relatively low dimensional space, we reduce data dimension by an equivalent method to PCA. Each column in the matrix X subtracts the average along the columns. Then, we calculate the co-similarity matrix $A = (X^T X)/L^2$, so the size of the matrix A is $L \times L$. The eigenvalues and eigenvectors are calculated based on the matrix A selected with their eigenvector $U = [u_1, u_2, \ldots, u_d]^T$ according to the biggest d eigenvalues, where u_i is an eigenvector. The size of the matrix U is $d \times L$. In our method, we use matrix U to compute the saliency of each patch other than original patch vector.

It is worth noting that [21] and [26] applied independent component analysis (ICA) to a training set of natural images. The features are calculated as the responses to filters learned from natural images using ICA. In [25], they used PCA to image patches and analyzed the image features which are more suitable to salient detection. Unlike [25] which extracted the PCs by sampling the patches from a large number of images, we pay attention to the PCs only from the current image. We suppose that PCA over patches within each image emphasized the variability within the image, which is important to discriminate the salient object from background.

2.3 Detection of the Patch's Saliency

In the proposed algorithm, the saliency value of each image patch is determined by two factors: one is the dissimilarities of color between image patches in a reduced dimensional space; the other is the spatial distance between an image patch and all other patches.

A patch is salient if the color of its pixels is unique. We should not look at an isolated patch, but its surrounding patches, which is similar to the definition of center-surrounding contrast [27] method. Thus, a patch p_i is considered salient if the appearance of the patch p_i is distinctive with respect to all other image patches. Specifically, let $dist_{color}(p_i, p_j)$ be the distance between the patches p_i and p_j in the reduced dimensional space. Patch p_i is considered salient when $dist_{color}(p_i, p_j)$ is high for $\forall j$.

$$dist_{color}(p_i, p_j) = \sum_{n=1}^{d} |u_{ni} - u_{nj}| \tag{1}$$

The positional distance between patches is also an important factor. Generally speaking, background patches are likely to have many similar patches both near and far-away in the image. It is in contrast to salient patches that the latter tend to be grouped together. This implies that a patch p_i is salient when the patches similar to it are nearby, and it is less salient when the resembling patches are far away. Let $dist(p_i, p_j)$ be the Euclidean distance between the positions of patches p_i and p_j, which is represented by the two centers of patches

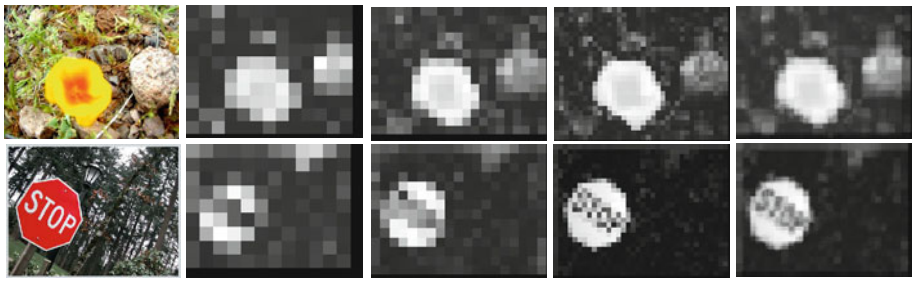

Fig. 1 The original images and its saliency maps with different patch sizes; **a** original images, **b** saliency maps with the image patch size 30×30; **c** saliency maps with the image patch size 20×20; **d** saliency maps with the image patch size 10×10; **e** final saliency maps which combines three results together

p_i and p_j in the original image, normalized by the larger image dimension. Based on the observations above we define a dissimilarity measure between a pair of patches p_i and p_j as:

$$dissimilarity(p_i, p_j) = \frac{dist_{color}(p_i, p_j)}{1 + dist(p_i, p_j)} \tag{2}$$

This dissimilarity measure is proportional to the difference in appearance and inverse proportional to the positional distance.

To evaluate a patch's uniqueness, we can compute dissimilarities between the patch and all other patches and take the sum of these dissimilarities as the saliency value of the related patch. The saliency value of patch p_i can be expressed as follows:

$$S_i = 1 - \exp\left\{-\frac{1}{L}\sum_{k=1}^{L} dissimilarity(p_i, q_k)\right\} \tag{3}$$

As described previously, the patch size would influence the calculation of saliency map. With a smaller patch size, the saliency map will become more distinguishable, as shown in Fig. 1 where the saliency map with the smallest image patch size (shown in Fig. 1d) is more distinguishable than the other two with larger patch size (shown in Fig. 1b, c). Of course, to obtain more accurate saliency map, we hope to divide image into smaller image patches. However, in this situation, the computational complexity will increase. The computational complexity of our algorithm includes two fold: the first is the computational complexity on preprocessing, such as dividing original images into patches and PCA; the other time consuming cost is computing dissimilarities between patches. Given an input image with size of $H \times W$ (where H is the height and W is the width) and the patch size of $n \times n$, the computational complexity of our algorithm is $O(L^3 + L^2)$, in which L^3 and L^2 correspond to the computational cost of preprocessing and dissimilarity calculation respectively, where $L = \lfloor H/n \rfloor \cdot \lfloor W/n \rfloor$. Therefore, with the smaller patch size, the computational complexity will be higher.

In addition, large patch size may lead to another problem. In the saliency map, the saliency values of all pixels in a patch are decided by the dissimilarity between this patch and all other patches. Therefore, the saliency values in a patch are the same. Our algorithm can not describe the boundary of small salient object when the patch size is larger than the salient object. We use Eq. (3) to compute saliency value of the original image with different patch sizes can obtain the saliency map with different scales. The saliency map with large scales is to detect

the whole information and that with small scales is to describe the salient object in detail. Therefore, we extend the saliency detection algorithm to multi-scale to strengthen the saliency value of the salient areas in an input image.

2.4 Extended by Multiple Scales

Based on the observation that patches in background are likely to have similar patches at multiple scales, which is in contrast to more salient patches that could have similar patches at a few scales but not at all of them. (It is equal to the principle proposed by [27,28] that salient object always smaller than the background.) Therefore, we wish to incorporate multiple scales to further decrease the saliency of background patches, improve the contrast between salient and non-salient regions.

For a patch p_i of scale r, the saliency value according to Eq. (3) is defined as:

$$S_i^r = 1 - \exp\left\{-\frac{1}{L}\sum_{k=1}^{L} dissimilarity\,(p_i^r, q_k^r)\right\} \tag{4}$$

Considering the scales $R_c = \{r_1, r_2, \ldots, r_M\}$, we use Eq. (4) to calculate the saliency of patch i as $\{S_i^{r1}, S_i^{r2}, \ldots, S_i^{rM}\}$. The final saliency is computed as:

$$S_i = \frac{1}{M}\sum_{r\in R_c} S_i^r \tag{5}$$

As aforementioned in last subsection, the computational complexity of the proposed model is decided by the patch size, which determines the scale of saliency map. Therefore, we choose the suitable patch size from our experiments to compute the saliency map based on the consideration of final saliency detection performance and computational complexity.

3 Experiments

We evaluate our method in two aspects: predicting human visual fixations and segmenting the salient object from natural images. In this section, we compare the proposed method with the state-of-art models and give the quantitative evaluation on the public database from the perspective of human visual fixation and salient object segmentation. For human visual fixation, our approach is compared with seven state-of-the-art saliency detection methods, including IT [9], AIM [21], SUN [26], GBVS [29], Duan's method [19], Hou' method [22] and RC [18]. In them, all methods except RC are more suitable to fixation. For salient object segmentation, we compared our results with AC [30], CA [15], FT [17] and RC, which can generate good segment results on the salient object segmentation database. In addition, we also give the comparison between our algorithm and the methods used in human visual fixation on this database.

3.1 Predicting Human Visual Fixations

In this subsection, we show several experimental results on detecting saliency in natural images. We used the image dataset and its fixation data collected by Bruce and Tsotsos [21] as a benchmark for comparison. This dataset contains eye fixation records from 20 subjects for a total of 120 images of size 681×511. To compare our results with [19], we choose 11 as the number of reduced dimensions which is the best value to maximize saliency predictions.

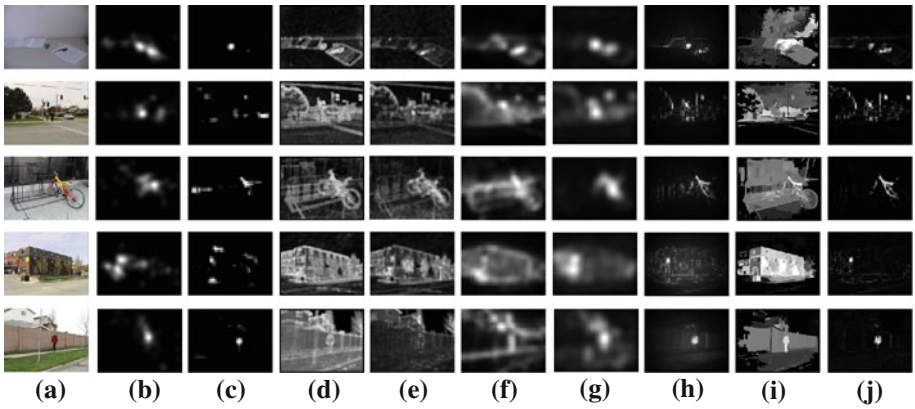

Fig. 2 Results on predicting human visual fixation data: **a** input images; **b** human fixations; **c** saliency map from Itti's model [9]; **d** saliency map from Bruce's model [21]; **e** saliency map from Zhang's model [26]; **f** saliency map from Harel's model [29]; **g** saliency map from Hou's model [22]; **h** saliency map from Duan's model [19]; **i** saliency map from Cheng's model [18]; **j** saliency map from the proposed model

For the patch size, we choose {30, 20, 10} because better results are easy to obtaining in these values [19]. We obtained an overall saliency map by using YCbCr color space in all experiments. Some visual results of our algorithm are compared with the state-of-art methods in Fig. 2.

The comparison results show that the most salient locations on our saliency maps are more consistent with the human fixation density maps. Note that our method is much less sensitive to background texture, which is different from AIM, GBVS, RC and SUN. Duan's method, which used the center bias mechanism indicating a strong bias to the center of the image, is easy to detect the salient object near the center of image. Our algorithm not only detects the dominant object in the center but also that located in the boundary of the original image. We also computed the area receiver operating characteristic (ROC) curve [21], i.e., the area under the curve to quantitatively evaluate the algorithm performance. To neutralize the effects of center bias during the computation of ROC area, we used the same procedure as in [26]. More specifically, we first compute true positives from the saliency maps based on the human eye fixation points. In order to calculate false positives from the saliency maps, we use the human fixation points from other images by permuting the order of images. This permutation of images is repeated 100 times. To calculate the area under the ROC curve, we compute detection rates and false alarm rates by thresholding histograms of true positives and false positives at each stage of shuffling. The final ROC area shown in Table 1 is the average value over 100 permutations. The mean and standard errors are also reported in Table 1. It is observed that our model outperforms all other methods in terms of ROC area.

3.2 Salient Object Segmentation Database

We have evaluated the results of our approach on the publicly available database provided by Achanta et al. [17]. To the best of our knowledge, the database is the largest of its kind, and has ground truth in the form of accurate human-marked labels for salient regions. In this subsection, we compare the proposed method with state-of-the-art saliency detection methods on the performance of segmenting salient object in nature images.

Table 1 Performance in predicting human visual fixation data

Attention model	ROC (SE)
Itti et al. [9]	0.6146 (0.0008)
AIM [21]	0.6727 (0.0008)
SUN [26]	0.6682 (0.0008)
GBVS [29]	0.6818 (0.0007)
Duan et al. [19]	0.6837 (0.0008)
RC [18]	0.6839 (0.0007)
Hou et al. [22]	0.6841 (0.0007)
Our method	0.6842 (0.0007)

SE Standard errors

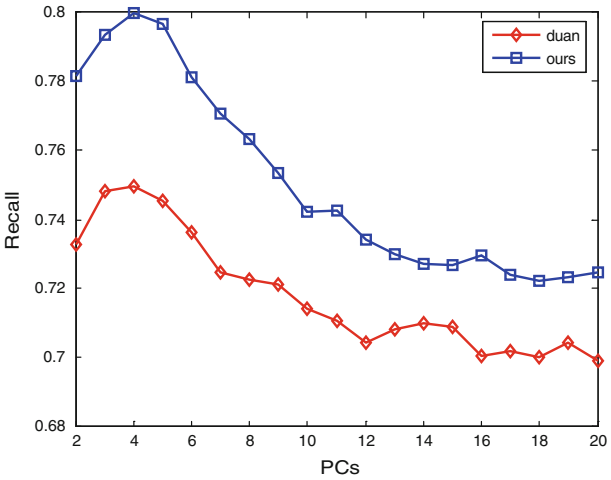

Fig. 3 The relationship between the Recall and the dimension when patch size set to 14

We use our method and the others to compute saliency maps for all the 1,000 images in the database. To reliably compare our method with Duan's algorithm on performance according to the selection of PCs number, we vary the number of PCs from 2 to 20. Figure 3 shows the comparison between the proposed algorithm and Duan's method on the index of recall when choosing different number of PCs. From Fig. 3, average recall of the proposed method approaches 0.8 corresponding to the number of PCs from 3 to 5. With the number of PCs increase, recall is decrease. That is to say, except the first few PCs, other features in the original image can not discriminate foreground and background because they do not have meaningful spatial structure. We choose 4 as PCs number which is also the best selection to Duan's method based on the low curve showed in Fig. 3. The patch size set to be the same to previous subsection, which is still the best parameter in this database.

Visual comparison of saliency maps obtained by the proposed method and other algorithms can be seen in Figs. 4 and 5. The comparison results show that the saliency map generated by IT [9] always detect the location of salient object approximately rather than a whole area. CA [15], which focuses on the boundary of salient area, is failing to segment the salient object from natural images. An excellent model, termed FT [17], generally detect the foreground from input images. However, it is easy to influence by the background that the salient area contains not only salient object but also clutter background. The results

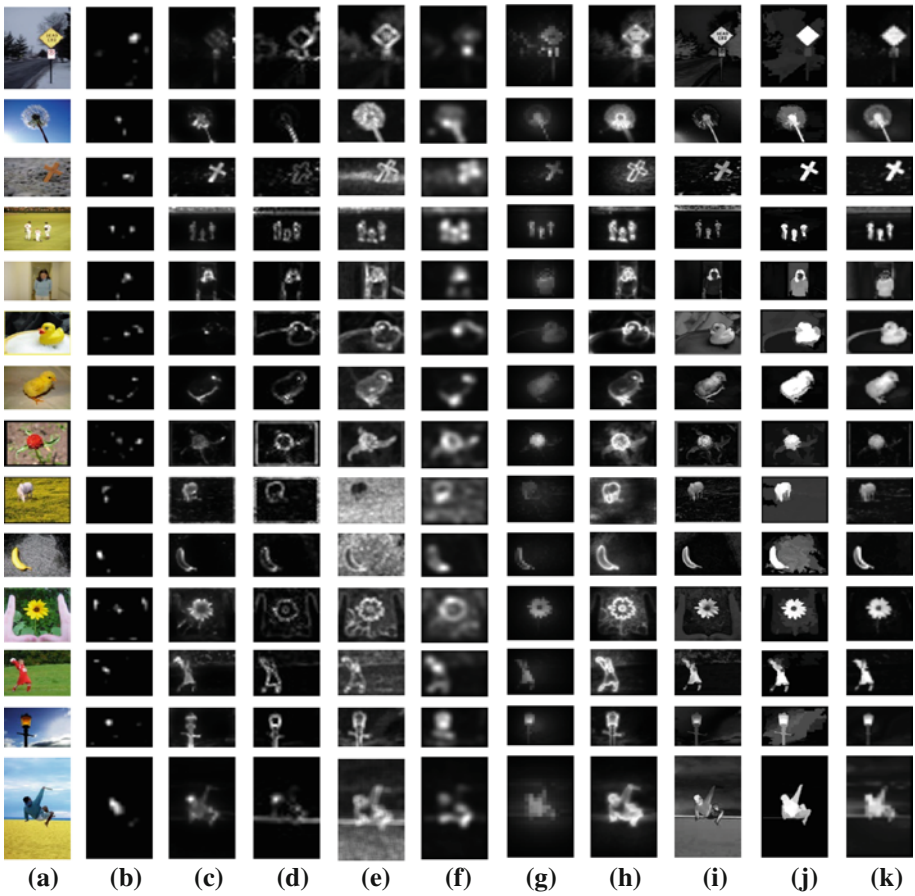

Fig. 4 Saliency maps from different saliency detection models on segmentation image database. **a** Original image, **b**–**k** are the results generated by IT [9], GBVS [29], MZ [12], SUN [26], Hou et al. [22], Duan et al. [19], CA [15], FT [17], RC [18] and the proposed algorithm, respectively

generated by RC [18] showed in Figs. 4 and 5g are outstanding in the models mentioned before. But similar to FT, it is also influenced by the clutter background. The reason might be that RC compute color contrast based on color histogram to measure the difference between two regions. It failed when the regions located in salient object and in background have the same color histogram. From the ninth to eleventh rows in Fig. 5, RC detects the contour of the salient object. But the salient regions contain not only the salient object, but also the clutter background. On the contrary, such favorable saliency maps can be achieved since our algorithm robustly works in cluttered scenes, which fits with the human visual perception.

Note that Duan's method can detect the foreground in the image. However, the salient area just focuses on the center of the image. This characteristic might be lead to two problems: first, the foreground near the center of the input image is more salient than that far from the center; second, the background near the center is more salient than the foreground located in the boundary of the input image. For instance, in Fig. 5c, the head of the horse and the background which located in the center of the input image is more salient than

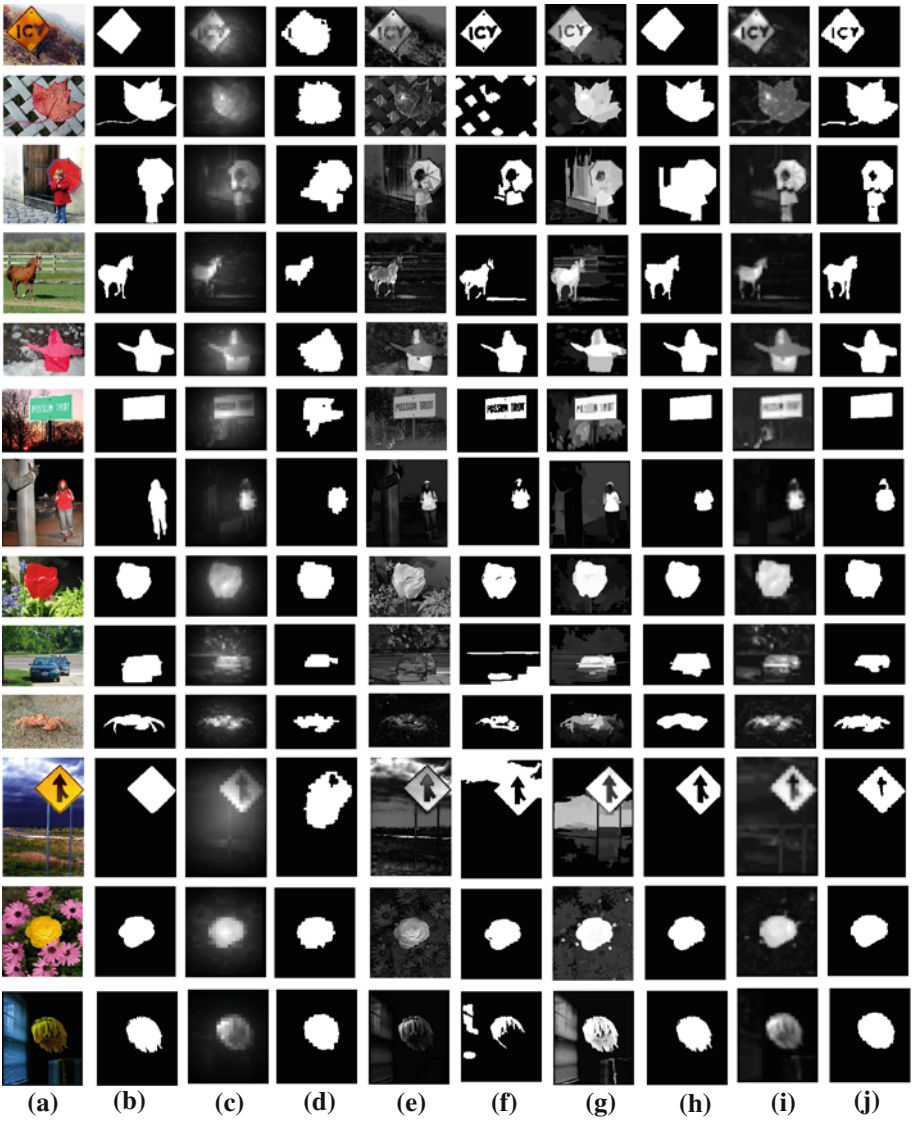

Fig. 5 Saliency maps from different saliency detection models and corresponding binary masks. **a** original images, **b** ground truth, **c**, **e**, **g** and **i** are the saliency maps from Duan's method, FT, RC and the proposed method, respectively. **d**, **f**, **h** and **j** are the binary masks results using Grabcut according to (**c**), (**e**), (**g**) and (**i**) respectively

the legs which are located in the boundary of the original image. Our method, which can detect the salient area in the whole image, overcomes these problems. In addition, comparing with Duan's method, our algorithm can detect the salient object with precision boundary.

Thus, the quantitative evaluation for a saliency detection algorithm is to see how much the saliency map from the algorithm overlaps with the ground-truth saliency map. Here,

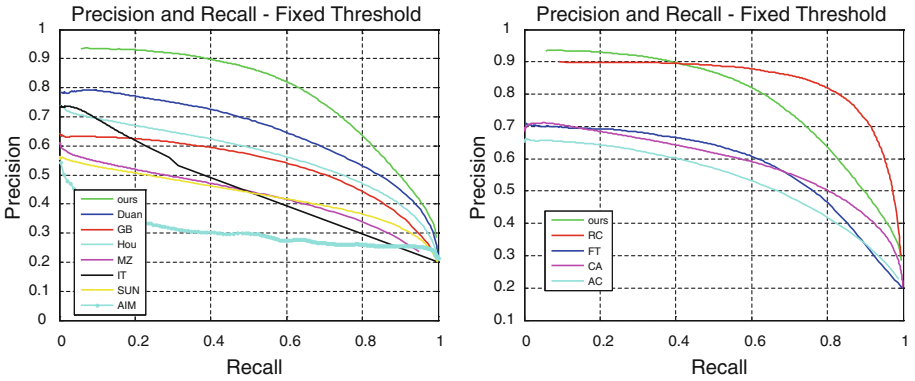

Fig. 6 Precision and recall rates for all algorithms

we exploit the precision, recall, and F-measure to evaluate the performance of our proposed model. Precision is computed as the ratio of correctly detected saliency region to the detected salient region from the saliency detection algorithm. Recall is calculated as the ratio of correctly detected salient region to the ground-truth salient region. Given a ground-truth saliency map $G = [g_1, g_2, \ldots, g_n]$ and the detected saliency map $S = [s_1, s_2, \ldots, s_n]$ for an image, we have:

$$presicion = \sum_x g_x s_x / \sum_x s_x \tag{6}$$

$$recall = \sum_x g_x s_x / \sum_x g_x \tag{7}$$

F-measure, a harmonic mean of precision and recall, is a measure that combines precision and recall. It is calculated as follows:

$$F_\beta = \frac{(1+\beta) precision \times recall}{\beta \times precision + recall} \tag{8}$$

where β is a positive parameter to decide the importance of precision over recall in computing the F-measure.

To obtain the quantitative evaluation, we perform two different experiments. In the first experiment, similar to [17], we use the simplest way to get a binary segmentation of salient objects by thresholding the saliency map with a threshold from 0 to 255. Figure 6 shows the resulting precision versus recall curves. From Fig. 6, the minimum recall values of our methods are higher than those of the other methods, because the saliency maps computed by our methods are smoother and contain more pixels with the saliency value 255.

In second experiment, we use the iterative GrabCut [18] to obtain a binary mask for a given saliency map. To automatically initialize GrabCut, we use a segmentation obtained by binarizing the saliency map using a fixed threshold. We set the threshold to 0.3 empirically. Once initialized, we iteratively run GrabCut 4 times to improve the saliency cut result. Final saliency cut result generated by this way is as our binary mask to obtain the quantitative evaluation (see Fig. 5).

We use $\beta = 0.3$ in our work for fair comparison. The comparison results are shown in Fig. 7. Generally speaking, the precision indicates the performance of the saliency detection

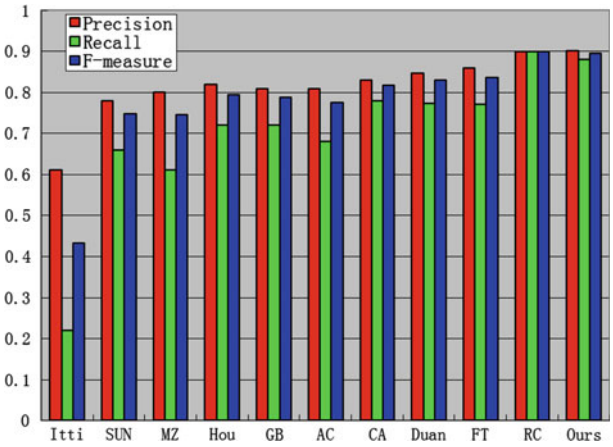

Fig. 7 The experiment results for the comparison between our proposed model and other state-of-art methods

Table 2 Average time taken to compute a saliency map for images in the database by [17]

Method	IT	Hou et al. [22]	GB	CA	SUN	MZ	FT	RC	Duan	Our method
Time (s)	0.725	0.086	2.14	53.67	8.97	0.084	0.48	0.317	3.587	4.43
Code	Matlab	Matlab	Matlab	Matlab	Matlab	C++	Matlab	C++	Matlab	Matlab

Algorithms were tested using a Dual Core 1.8 GHz machine with 1 GB RAM

algorithms compared with ground-truth saliency map. To compare the proposed model with others, we always see the precision value for different algorithms, for the precision value is the ratio of the correctly detected region over the whole detected region. In Fig. 7, we find out that the performance of our algorithm is close to that of RC, but be better than others. In addition, compared with Duan's method on this database (precision = 85 %, recall = 77 %), we achieved better accuracy (precision = 90 %, recall = 88 %).

To better understand the proposed model, we compare the average time taken by each method which summarized in Table 2. For the methods mentioned in Table 2 except Duan's method, we used the author's implementations. While for Duan's method, we implemented the algorithm in Matlab since we could not find the author's implementation. Our method is slower than others as it requires feature extraction using PCA which is rather time-consuming in this processing. Furthermore, our method is also slower than Duan's method since our algorithm request computing saliency maps three times, but produces superior quality results. Similar to Duan's method, the computation times of the proposed algorithm are decided by the speed of PCA process which is used for feature extraction. Moreover, the time-consuming of each step and corresponding time cost by PCA are compared as shown in Fig. 8. It is observed that more than half of the computation times spend in the process of PCA (2.62 s) in our finally saliency calculation (4.43 s). On the other hand, the computation time rises significantly with the lower of the scale of patches which is important for the accuracy of our finally salient object segmentation. In future, we would speed up the PCA process or find another efficient feature extraction method to make our algorithm maintain a reasonable computation time.

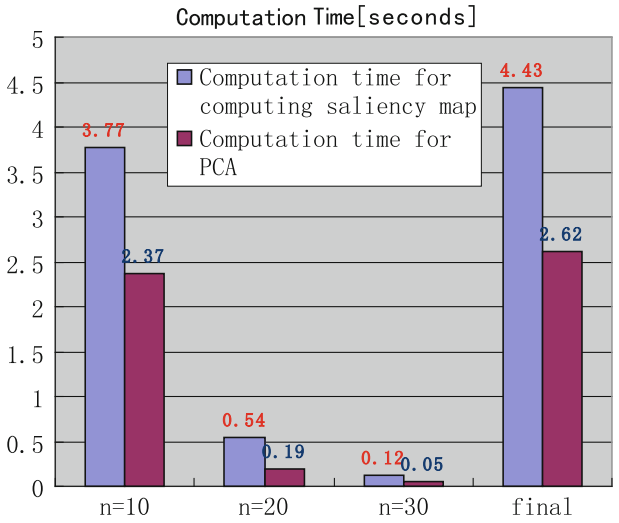

Fig. 8 A study of the efficiency with different scales of patch on Achanta's image database

4 Conclusions

We present a multiscale saliency detection algorithm based on image patches to detect the saliency object in color image. In this algorithm, an input image is segmented into patches and exploits PCA to reduce the dimensions which are noises with respect to the saliency calculation. Our saliency algorithm is based on three elements: color dissimilarity of patches, spatial distance and multiple scales. In according with the inadequacy of the methods used center-bias, the proposed algorithm used multiple scales to solve two problems. First, background near the center of image may be more salient than the foreground which is far away from the center of the image. Second, for a salient object, the part near the center of the input image is more salient than that far away from the center. In addition, using multiple scales can also to decrease the saliency of background patches and to improve the contrast between salient and non-salient regions. We evaluate our method on two publicly available data sets and compare our scheme with the state-of-art models. The resulting saliency maps are much less sensitive to background texture for predicting human visual fixations. Furthermore, it is also suitable to salient object segmentation.

In the future, we plan to investigate the practicability of the proposed saliency maps can be used for efficient object detection, reliable image classification, robust image scene analysis, leading to improved image retrieval. In addition, the proposed algorithm has high time complexity since the process of PCA is time-consuming. How to make our algorithm more efficient is also our future work.

Acknowledgments This research is partially sponsored by Natural Science Foundation of China (No. 90820306 and No. KT06015), Natural Science Research Project of Jiang Su Provincial Colleges and Universities (No. 11KJD520003). The authors would like to thank the technical assistance from the PhD student Dongyan Guo.

References

1. Santella A, Agrawala M, Decarlo D, Salesin D, Cohen M (2006) Gaze-based interaction for semi-automatic photo cropping. In: ACM human factors in computing systems (CHI). ACM, Montreal, pp 771–780
2. Chen L, Xie X, Fan X, Ma W, Shang H, Zhou H (2002) H A visual attention mode for adapting images on small displays. Technical report, Microsoft Research, Redmond, WA
3. Itti L (2000) Models of bottom-up and top-down visual attention. PhD thesis, California Institute of Technology, Pasadena
4. Navalpakkam V, Itti L (2006) An integrated model of top-down and bottom-up attention for optimizing detection speed. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 2049–2056
5. Rutishauser U, Walther D, Koch C, Perona P (2004) Is bottom-up attention useful for object recognition? In: IEEE conference on computer vision and pattern recognition (CVPR), pp 37–44
6. Jong Seo H, Milanfar P (2009) Nonparametric bottom-up saliency detection by self-resemblance. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 45–52
7. Baldi P, Itti L (2010) Of bits and wows: a Bayesian theory of surprise with applications to attention. Neural Netw 23(5):649–666
8. Torralba A, Oliva A, Castelhano M, Henderson J (2006) Contextual guidance of eye movements and attention in real-world scenes: the role of global features on object search. Psychol Rev 113(4):766–786
9. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20(11):1254–1259
10. Hou X, Zhang L (2008) Saliency detection: a spectral residual approach. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 1–8
11. Guo C, Ma Q, Zhang L (2008) Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 1–8
12. Ma Y-F, Zhang H (2003) Contrast-based image attention analysis by using fuzzy growing. In: ACM multimedia, pp 374–381
13. Itti L, Baldi P (2005) Bayesian surprise attracts human attention. In: NIPS, Cambridge
14. Liu T, Sun J, Zheng N, Tang X, Shum H-Y (2007) Learning to detect a salient object. In: CVPR
15. Goferman S, Zelnik-Manor L, Tal A (2010) Context-aware saliency detection. In: CVPR, pp 2376–2383
16. Jiang H et al (2011) Automatic salient object segmentation based on context and shape prior. In: BMVC
17. Achanta R, Hemami S, Estrada F, Süsstrunk S (2009) Frequency-tuned salient region detection. IEEE conference on computer vision and pattern recognition (CVPR), pp 1597–1604
18. Cheng M et al (2011) Global contrast based salient region detection. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 409–416
19. Duan L, Wu C, Miao J, Qing L, Fu Y (2011) Visual saliency detection by spatially weighted dissimilarity. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 21–23
20. Tatler BW (2007) The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions. J Vis 7(14):4.1–4.17
21. Bruce N, Tsotsos J (2006) Saliency based on information maximization. In: Advances in neural information processing systems, pp 155–162
22. Hou X, Harel J, Koch C (2012) Image signature: highlighting sparse salient regions. IEEE Trans Pattern Anal Mach Intell 34(1):194–201
23. Yang J, Zhang D, Frangi AF, Yang JY (2004) Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Trans Pattern Anal Mach Intell 26(1):131–137
24. Kadir T, Brady M (2001) Saliency, scale and image description. Int J Comput Vis 45(2):83–105
25. Hyvärinen A, Hurri J, Hoyer PO (2009) Natural image statistics: a probabilistic approach to early computational vision. Springer, London
26. Zhang L, Tong M, Marks T, Shan H, Cottrell G (2008) SUN: a Bayesian framework for saliency using natural statistics. J Vis 8(7):1–20
27. Gao D, Mahadevan V, Vasconcelos N (2008) On the plausibility of the discriminant center-surround hypothesis for visual saliency. J Vis 8(7):1–18
28. Gao D, Vasconcelos N (2004) Discriminant saliency for visual recognition from cluttered scenes. In: Advances in neural information processing systems, pp 481–488
29. Harel J, Koch C, Perona P (2007) Graph-based visual saliency. In: Advances in neural information processing systems, pp 545–555
30. Achanta R, Estrada FJ, Wils P, Süsstrunk S (2008) Salient region detection and segmentation. In: ICVS, pp 66–75