Noname manuscript No.
(will be inserted by the editor)

How Important is Location Information in Saliency Detection of Natural Images

Tongwei Ren · Yan Liu · Ran Ju · Gangshan Wu

Received: date / Accepted: date

Abstract Location information, i.e., the position of content in image plane, is considered as an important supplement in saliency detection. The effect of location information is usually evaluated by integrating it with the selected saliency detection methods and measuring the improvement, which is highly influenced by the selection of saliency methods. In this paper, we provide direct and quantitative analysis of the importance of location information for saliency detection in natural images. We firstly analyze the relationship between content location and saliency distribution on four public image datasets, and validate the distribution by simply treating location based Gaussian distribution as saliency map. To further validate the effectiveness of location information, we propose a location based saliency detection approach, which completely initializes saliency maps with location information and propagate saliency among patches based on color similarity, and discuss the robustness of location information's effect. The experimental results show that location information plays a positive role in saliency detection, and the proposed method can outperform most state-of-the-art saliency detection methods and handle natural images with different object positions and multiple salient objects.

Tongwei Ren
State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
E-mail: rentw@nju.edu.cn

Yan Liu
Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China
E-mail: csyliu@comp.polyu.edu.hk

Ran Ju
State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
E-mail: juran@smail.nju.edu.cn

Gangshan Wu
State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
E-mail: gswu@nju.edu.cn

Keywords Saliency detection · location information · patch representation · saliency propagation

1 Introduction

Saliency detection, i.e., detecting the regions attracting human attention from image content, is used as a fundamental of many real world multimedia applications [27]. In the past years, numerous saliency detection methods have been proposed [8], which can be roughly classified into two directions: fixation prediction [9] and salient object detection [7]. Fixation prediction aims to simulate the attention mechanism of human visual system by highlighting a few salient points [9]. Different to fixation prediction, salient object detection aims to extract the entire salient objects [12], which is more suitable to multimedia applications, such as object segmentation [13, 23], image classification [46, 47], object tracking [40, 49], image and video annotation [4, 37, 43], information retrieval [6, 45, 48] and content-aware editing [38, 39, 44].

Human visual system can quickly and accurately identify the salient objects in different scenes, but automatic identification of such salient objects is very challenging [12]. So it is necessary to explore the potential of different cues to improve the effectiveness of salient object detection. The existing methods mainly pay attention to low-level image content cues, such as color orientation histogram [26], texture and shape [33], and some high-level visual information, such as object detection [19] and human perception criteria [21]. Location information, i.e., the position of content in image plane, is also usually used as an important supplement for saliency detection in natural images, for salient objects are usually placed near to the center or golden section ratio of image in photography [3] and the objects placed at the center of image plane attract more attention than other locations [14]. Jiang *et al.* [20] proposed three characteristics of a salient object, including that it is most probably placed near the center of the image. Borji *et al.* [7] showed that most of the existing image datasets for saliency detection had center-bias. Some salient object detection methods also used this characteristic to improve their performance [24]. However, these works didn't provide further quantitative evaluation of the effect of location information in saliency detection. Schauerte *et al.* [41] analyzed the salient region distribution on 1,000 images of a salient object dataset [30] and the effect of location by combining explicit center bias on some existing methods, but the direct and quantitative analysis for the effectiveness of location information is still lacking.

In this paper, we focus on directly and quantitatively analyzing the importance of location information in saliency detection. Based on the analysis of the relationship between content location and saliency distribution on four public datasets, we validate the statistic distribution of salient objects by simply treating location based Gaussian distribution as saliency map. Moreover, we propose a location based saliency detection approach to further validate the effectiveness of location information, and compare it with the state-of-the-art saliency detection methods. The experimental results show that location information plays a positive role in saliency detection, and the proposed approach obtains better performance than most of the compared methods. We further analyze the robustness of location information's effect for saliency detection with various object positions and different salient object numbers. The experiments show that location information also benefits to solve these problems.

The remaining of the paper is organized as follows. In Section 2, we briefly survey the existing saliency detection methods. Then, in Section 3, we show the analysis of the relationship between content location and saliency distribution. After this, we present the details of the proposed location based saliency detection approach in Section 4 and the experiment results and discussion in Section 5. Finally, the paper is concluded in Section 6.

2 Related Work

Here we provide a brief summary of the existing saliency detection methods. More details can be found in the recent survey papers [7,8].

The early saliency detection started from simulating the attention mechanism of human visual system. These methods usually work in bottom-up manners by integrating visual features across multiple scales, which aims at predicting the fixation points. Itti *et al.* [18] first proposed a computational model for gaze prediction, which used a center-surround operator across different scales. Ma *et al.* [33] proposed a local contrast method based on fuzzy model and using color, texture and shape as features. Tatler *et al.* revealed the correlation between human fixations and the basic features of images. Goferman *et al.* [15] combined global considerations and visual organization rules with low-level clues to detect salient object. More methods about fixation precision can be found in [9].

Compared to fixation prediction, salient object detection focuses on detecting the entire salient objects. According to whether using extrinsic cues, current salient object detection methods can be classified into intrinsic cues based methods and extrinsic cues based methods [11]. Intrinsic cues based methods only extract the cues from the input image itself for salient object detection. Hou *et al.* [17] extracted image spectral residual in spectral domain and rapidly constructed saliency map in spatial domain. Achanta *et al.* [1] propose to centering the intensities of whole images to highlight the salient objects. Gopalakrishnan *et al.* [16] formulated salient region detection problem as Markov random walks performed on images represented as graphs. Cheng *et al.* [12] employed global color contrast for salient object detection. Margolin *et al.* [35] tried to measure the pattern differences of patches using PCA and achieves encouraging results in five open datasets. Jia *et al.* [19] proposed to combine high level saliency priors by objectness measurement with low level appearance models. Jiang *et al.* [21] explored uniqueness, focusness and objectness for salient region detection. Bao *et al.* [5] generated multi-scale feature maps to obtain the global saliency based on global probability density distribution and measure local saliency based on local area entropy. Liu *et al.* [32] partitioned an image into a set of primitive regions using adaptive color quantization and further merged the salient regions with dynamic scale control scheme. Extrinsic cues based methods consider the intrinsic cues insufficient and bring in extrinsic cues to distinguish salient objects. Marchesotti *et al.* [34] detected the salient objects by leveraging the similar images from an annotated image dataset. Liu *et al.* [30] made a combination of local and global color features to infer the salient object. Liu *et al.* [31] proposed a two step framework by segmenting image into several regions and further generating the saliency value of each region by SVM. Li *et al.* [28] utilized light field as the additional information, in which the regions generated by Mean-shift with high foreground likelihood scores and background likelihood scores were selected as foreground and background, respectively.

According to the usage of location information, current salient object detection methods can be classified into three categories, including using no location, implicit location, and explicit location. Only a few methods completely ignore location information, such as calculating the distance between the color of each pixel and the mean color of the whole image as saliency value [1]. Most existing methods implicitly use location information, such as considering spatial relationships between regions [12], or explicitly use location information, such as combining center-bias to saliency detection results [23], to improve their performance. However, quantitative analysis and evaluation of the effectiveness of location information in saliency detection is still lacking. A similar work to this paper was proposed in [41], but it only analyzed the effect of location information on a small dataset by adding center-bias to some existing methods. Different to [41], we directly validate the effectiveness of location information by quantitative evaluating the performance of location based Gaussian and a proposed location based saliency detection approach on four datasets for saliency detection, and further discuss the robustness of location information's effect in handling different object positions and multiple salient objects.

3 Relationship between Location and Saliency Distribution

3.1 Statistics of Salient Object Distribution

To discover the distribution of salient objects, we make statistics on four public image datasets, including *SED1* [2], *NJU400* [24], *ASD* [1] and *MSRA10K* [12]. These datasets contain 100, 400, 1,000 and 10,000 source images with the pixel-level manually labeled saliency maps as ground truths, respectively. To $NJU400$ dataset, we only use the left-view source images and their corresponding ground truths. Figure 1 shows some examples of source images and corresponding ground truths. We resize all the ground truths to square and calculate the mean value and variance of the resized ground truths. As

Fig. 1 Examples of images in *MSRA10K* dataset. (top row) Source images. (bottom row) Corresponding manually labeled ground truths.

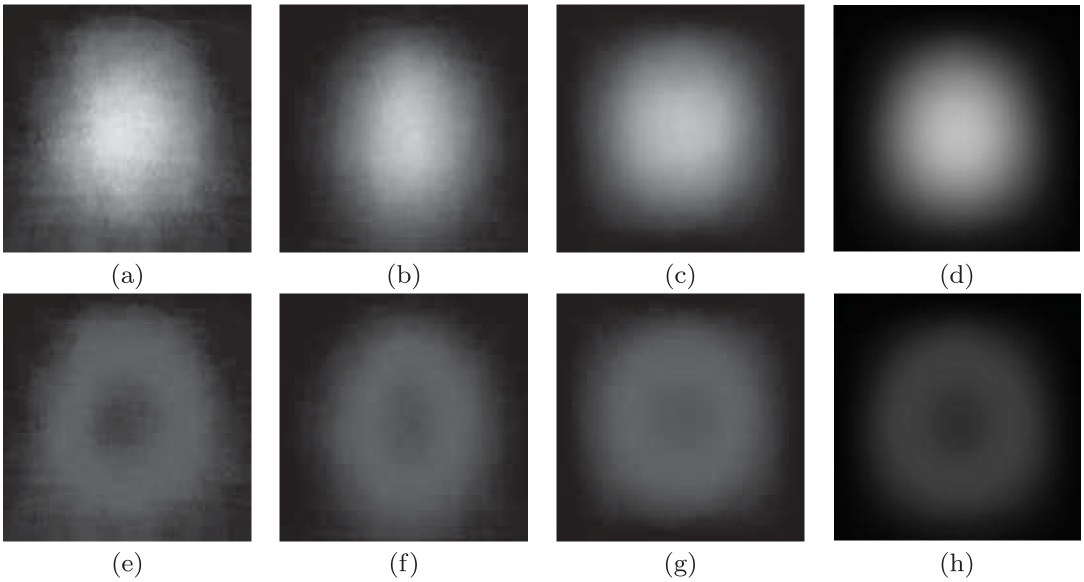


Fig. 2 Saliency distribution statistics on *SED1*, *NJU400*, *ASD* and *MSRA10K* datasets. (a)-(d) Mean values of all ground truths on *SED1*, *NJU400*, *ASD* and *MSRA10K* datasets, respectively. (e)-(h) Variances of all ground truths on *SED1*, *NJU400*, *ASD* and *MSRA10K* datasets, respectively.

shown in Figure 2, we can find that the regions near to image centers have high mean values and low variance values, which means the centric regions of most images are more likely to be salient. On the contrary, the regions far from image centers have low mean values and variance values, which means these regions are seldom treated as salient in detection.

3.2 Location based Gaussian Distribution

Inspired by Figure 2(a)-(d), we validate the effect of location information in saliency detection by simply treating *location based Gaussian distribution* (LGD) as saliency map. We decompose each source image into $M \times N$ patches, i.e., the number of patches is MN, and assign the saliency value to each patch

$p_{m,n}$ based on its normalized distance to the center of image:

$$s_{m,n} = \frac{1}{2\pi\sigma^2} e^{-\frac{(m'-1)^2+(n'-1)^2}{2\sigma^2}}, \tag{1}$$

where $m' = \frac{2m}{M+1}$ and $n' = \frac{2n}{N+1}$ compose the normalized coordinate (m', n') of patch $p_{m,n}$ to the center $(\frac{M+1}{2}, \frac{N+1}{2})$; σ is a parameter to adjust saliency distribution, and $\sigma^2 = 0.4$ in our experiments. Based on Eq. (1), we obtain the initial saliency map, and further normalize its value to the range of [0, 1] and resize it to source image size.

Figure 3 shows some examples of saliency maps generated by LGD using different patch decomposition. To a specific patch decomposition, the saliency maps for different source images are generated from the same initialization and resized to the same sizes of source images. So the generated saliency maps with same patch decomposition have similar appearances and only have different resolutions and aspect ratios.

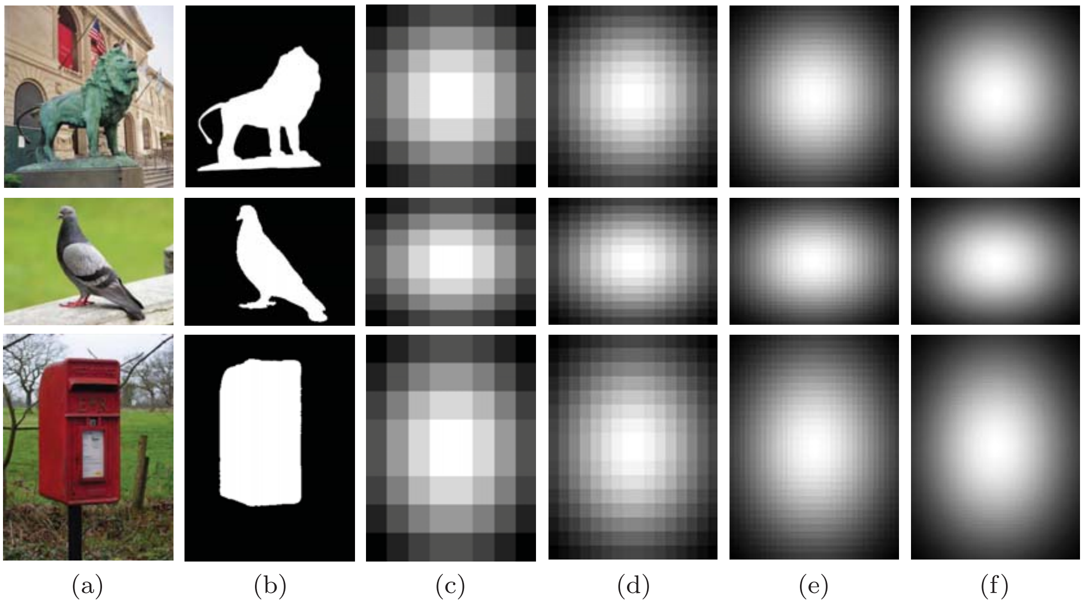
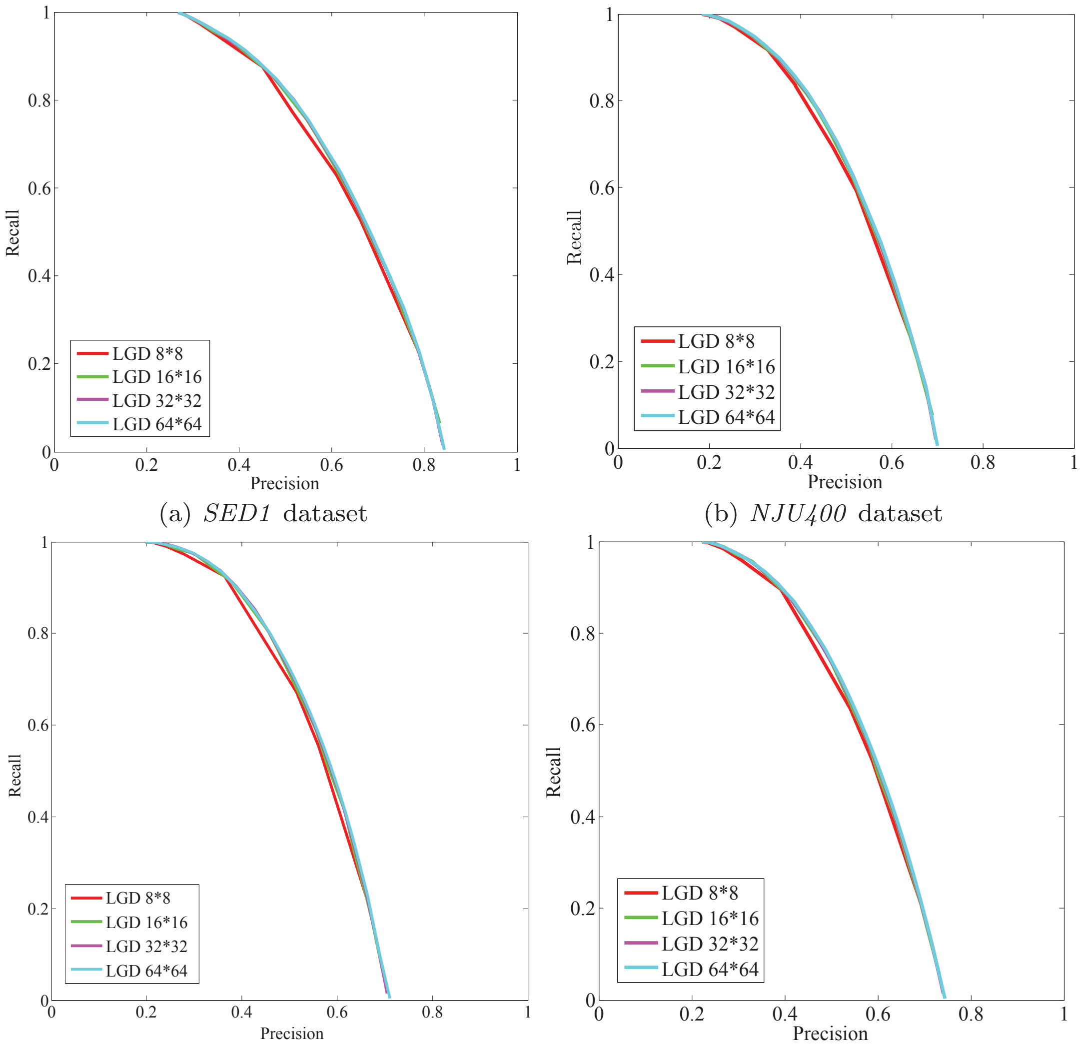


Fig. 3 Examples of saliency maps generated by LGD using different patch decomposition. (a) Source images. (b) Ground truths. (c)-(f) Saliency maps generated by LGD using patch decomposition in 8×8, 16×16, 32×32 and 64×64, respectively.

Figure 4-5 show the performance evaluation of LGD using different patch decomposition on the four datasets with precision-recall (PR) curves and receiver operating characteristic (ROC) curves, respectively. For different patch composition only influences the smooth of saliency maps, the performance of LGD with different patch decomposition is similar on all the datasets. Meanwhile, it is obvious that the performance of LGD is much better than random on all the datasets, that means location information plays a positive role in saliency detection.



(a) *SED1* dataset (b) *NJU400* dataset

(c) *ASD* dataset (d) *MSRA10K* dataset

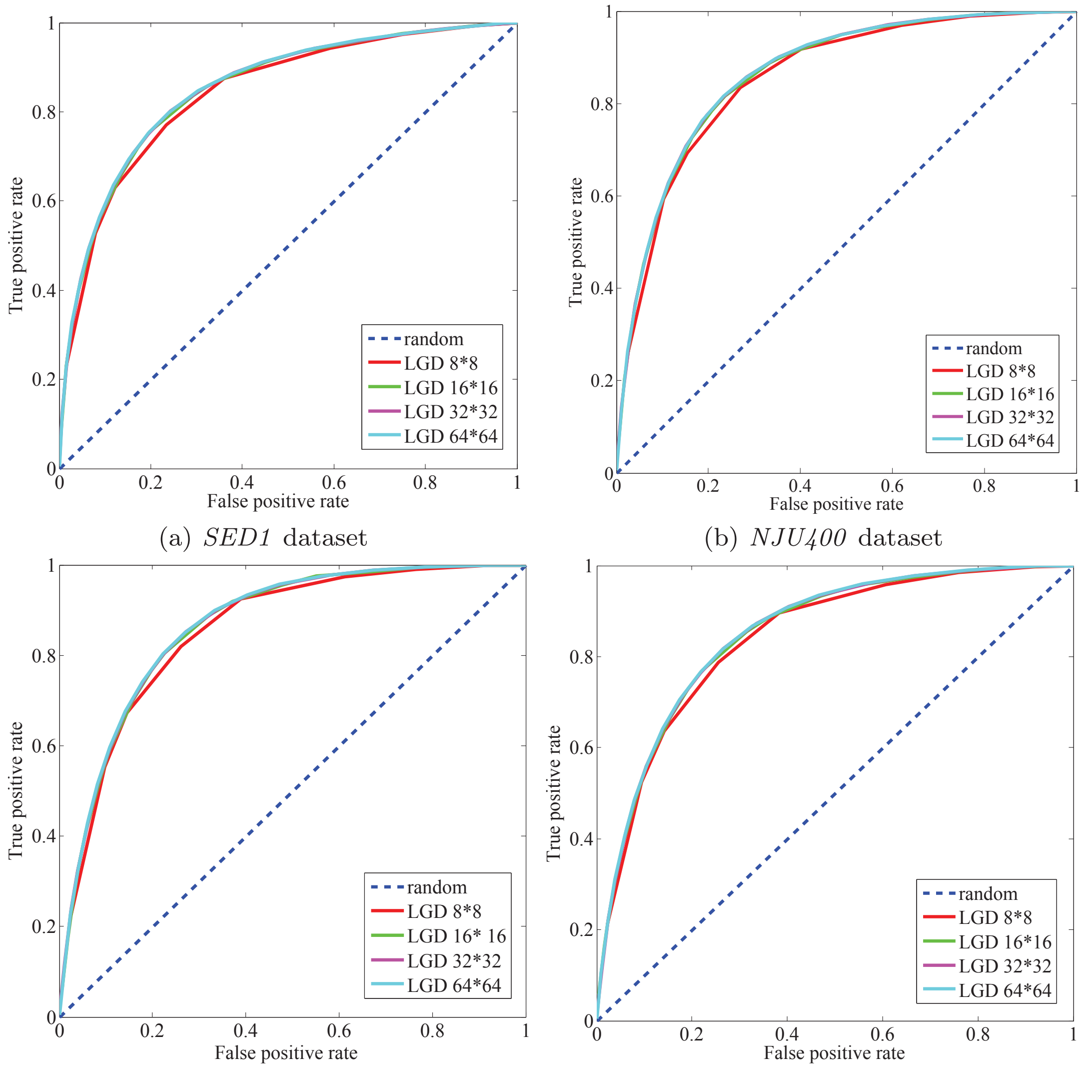
Fig. 4 PR curves for LGD using different patch decomposition on *SED1*, *NJU400*, *ASD*, *MSRA10K* datasets, respectively.

4 Location based Saliency Detection Approach

To further validate the effectiveness of location information in salient object detection, we propose a location based saliency detection approach, known as *location based saliency propagation* (LSP). To keep the dominant role of location information in saliency detection, we initialize saliency maps in LSP completely depending on location information, and only use color information in saliency propagation. In this way, we can minimize the influence of other cues when evaluating the effect of location information.

Similar to LGD, we decompose a source image into patches and initialize the saliency value of each patch $p_{i,j}$ with Eq. (1). Then we propagate the saliency from all the patches to $p_{i,j}$, which is similar to VisualRank [22,50]:

$$s^{'}_{i,j} = \boldsymbol{\omega}(:, p_{i,j})^T \mathbf{s}, \tag{2}$$



(a) *SED1* dataset (b) *NJU400* dataset

(c) *ASD* dataset (d) *MSRA10K* dataset

Fig. 5 ROC curves for LGD using different patch decomposition on *SED1*, *NJU400*, *ASD*, *MSRA10K* datasets, respectively.

where $\boldsymbol{\omega}(:, p_{i,j}) = [\omega(p_{1,1}, p_{i,j}), \ldots, \omega(p_{M,N}, p_{i,j})]^T$ is the propagation weight vector of all the patches to patch $p_{m,n}$, and $\mathbf{s} = [s_{1,1}, \ldots, s_{M,N}]^T$ is the saliency vector composed of the saliency values of all the patches.

We define the propagation weight based on the normalized spatial distance and mean color similarity between two patches:

$$\omega(p_{m,n}, p_{i,j}) = \omega_s(p_{m,n}, p_{i,j}) \cdot \omega_c(p_{m,n}, p_{i,j}). \tag{3}$$

Here, $\omega_s(p_{m,n}, p_{i,j})$ is the weight based on spatial distance between patch $p_{m,n}$ and $p_{i,j}$:

$$\omega_s(p_{m,n}, p_{i,j}) = e^{-\frac{(m'-i')^2+(n'-j')^2}{\sigma^2}}, \tag{4}$$

where σ is a parameter to adjust saliency distribution, and (m', n') and (i', j') are the normalized coordinates of patch $p_{m,n}$ and $p_{i,j}$ as same as Eq. (1).

And $\omega_c(p_{m,n}, p_{i,j})$ is the weight based on the distance between the mean color values of patch $p_{m,n}$ and $p_{i,j}$:

$$\omega_c(p_{m,n}, p_{i,j}) = 1 - ||\mathbf{c}_{m,n} - \mathbf{c}_{i,j}||_2, \tag{5}$$

where $\mathbf{c}_{m,n}$ and $\mathbf{c}_{i,j}$ are the mean color values of patch $p_{m,n}$ and $p_{i,j}$ in $L^*a^*b^*$ color space.

To make the saliency propagation convergence, we normalize the propagation weights of each patch to other patches, and conserve the saliency in propagation:

$$\omega^*(p_{m,n}, p_{i,j}) = \frac{\omega(p_{m,n}, p_{i,j})}{\sum_{i=1}^{M} \sum_{j=1}^{N} \omega(p_{m,n}, p_{i,j})}. \tag{6}$$

Based on Eq. (2)-(6), saliency map is iteratively updated by propagating saliency among patches till the change of saliency map is less than a pre-defined threshold, which equals $\frac{1}{MN}$ in our experiments. The saliency propagation is usually terminated after about four times iteration, and we assign ten to the maximum iteration number in our experiments. Similar to LGD, the generated saliency map by LSP is also normalized to the value range of [0, 1] and resized to source image size.

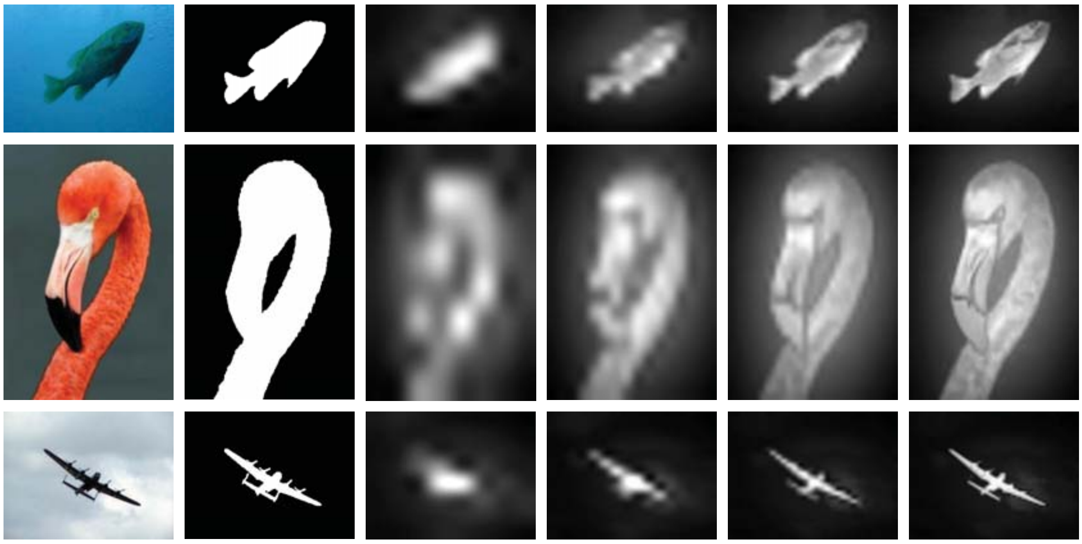
5 Experiments

5.1 Dataset and Experiment Setting

We evaluate LSP approach on the same datasets in LGD evaluation, including *SED1*[2], *NJU400*[24], *ASD*[1] and *MSRA10K*[12]. We evaluate its performance under four kinds of patch settings, including the patch numbers of 8×8, 16×16, 32×32 and 64×64. All the experiments were implemented in Matlab on a computer with 3.4GHz CPU and 8GB memory.

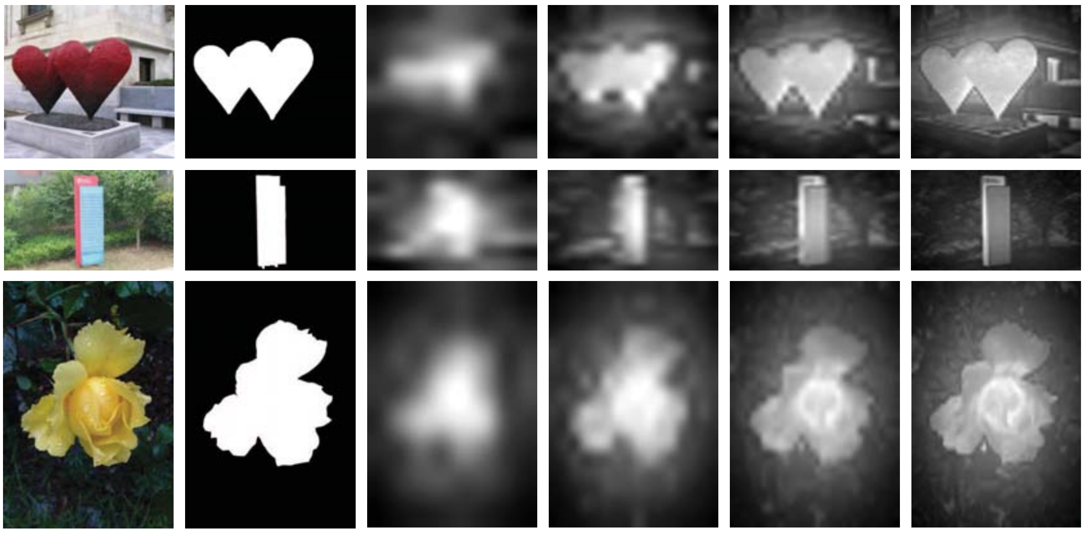
5.2 Experimental Results

Figure 6-9 show some examples of saliency maps generated by LSP using different patch decomposition on the four datasets, respectively. And Figure 10-11 show the performance evaluation of LSP using different patch decomposition on the four datasets with PR curves and ROC curves, respectively. Comparing Figure 4-5 and Figure 10-11, LSP retains the effectiveness of location information and has better performance than LGD for bringing in the assistance of color information. And different to LGD, the number of patches obviously influences the saliency detection performance in LSP. When increasing patch number from 8×8 to 32×32, PR curves are improved on three datasets except *NJU400* and ROC curves are improved on all the datasets. The reason is that the increasing of patch number makes the patch size smaller, which benefits to obtain more accurate salient region boundaries, such as the



(a) (b) (c) (d) (e) (f)

Fig. 6 Examples of saliency maps generated by LSP using different patch decomposition on *SED1* dataset. (a) Source images. (b) Ground truths. (c)-(f) Saliency maps generated by LSP using patch decomposition in 8×8, 16×16, 32×32 and 64×64, respectively.



(a) (b) (c) (d) (e) (f)

Fig. 7 Examples of saliency maps generated by LSP using different patch decomposition on *NJUDS400* dataset. (a) Source images. (b) Ground truths. (c)-(f) Saliency maps generated by LSP using patch decomposition in 8×8, 16×16, 32×32 and 64×64, respectively.

top and middle rows in Figure 9. But when increasing patch number from 32×32 to 64×64, the improvement becomes imperceptible. It berceuses that too small patches make the mean color values of patches more distinct and prevent the saliency propagation among patches, such as the bottom row in Figure 9, which may lead to more low saliency patches within salient object regions. In the following experiments, we use LSP 32×32 as an example.

For LSP uses color similarity together with location information, it is interesting to validate whether location information plays the primary role

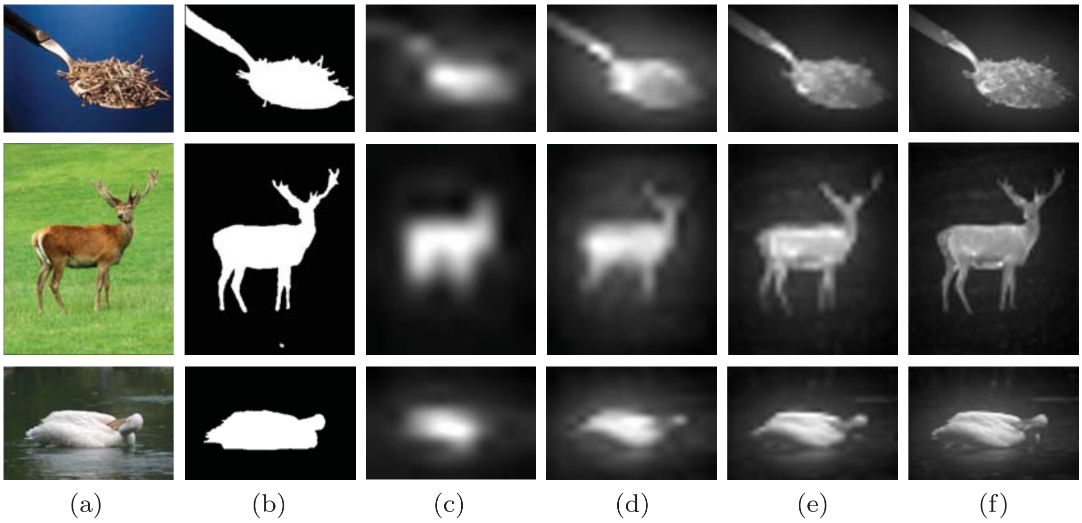


Fig. 8 Examples of saliency maps generated by LSP using different patch decomposition on *ASD* dataset. (a) Source images. (b) Ground truths. (c)-(f) Saliency maps generated by LSP using patch decomposition in 8×8, 16×16, 32×32 and 64×64, respectively.

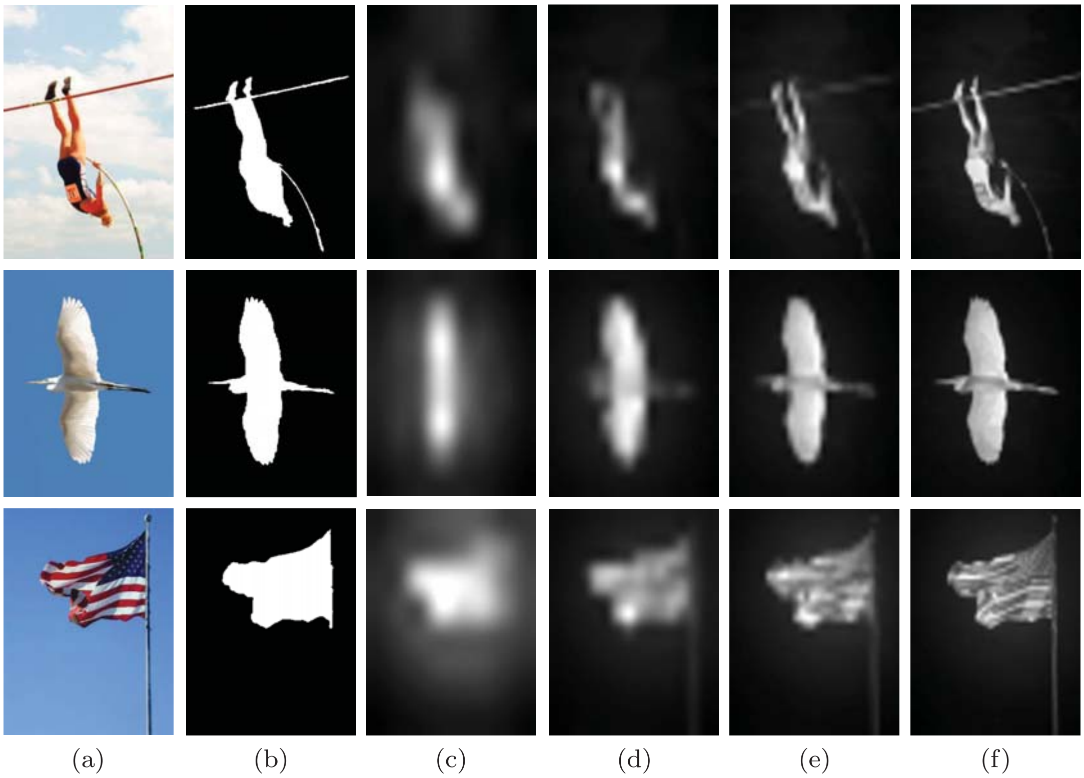
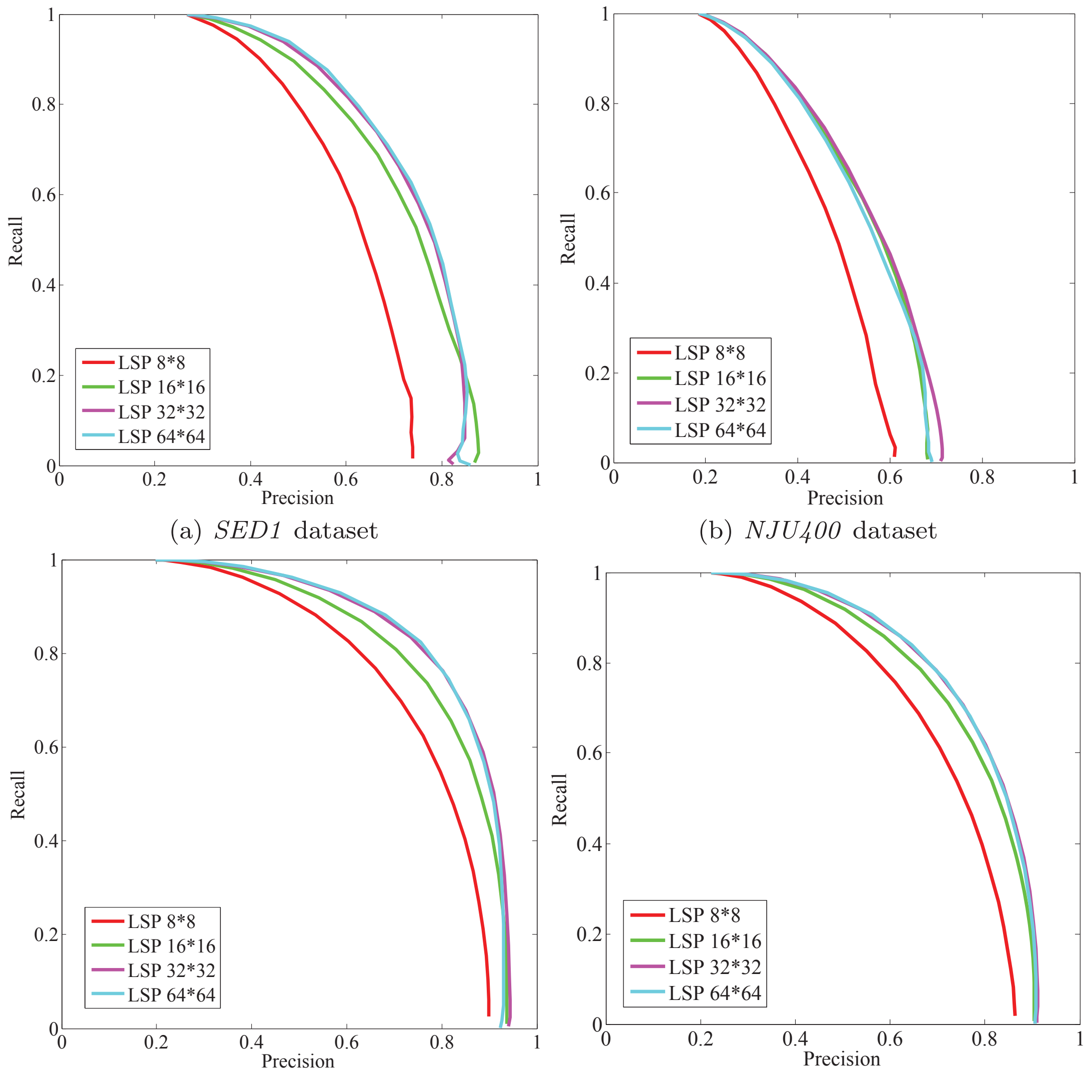


Fig. 9 Examples of saliency maps generated by LSP using different patch decomposition on *MSRA10K* dataset. (a) Source images. (b) Ground truths. (c)-(f) Saliency maps generated by LSP using patch decomposition in 8×8, 16×16, 32×32 and 64×64, respectively.

in LSP's performance. We use FT as the basis and propagate saliency among patches with ω_c in Eq. (3). Similar to FT, the generated FT+SP method only uses color information in saliency detection. Figure 12 shows the performance comparison of LGD, LSP, FT and FT+SP on *MSRA10K*



(a) *SED1* dataset

(b) *NJU400* dataset

(c) *ASD* dataset

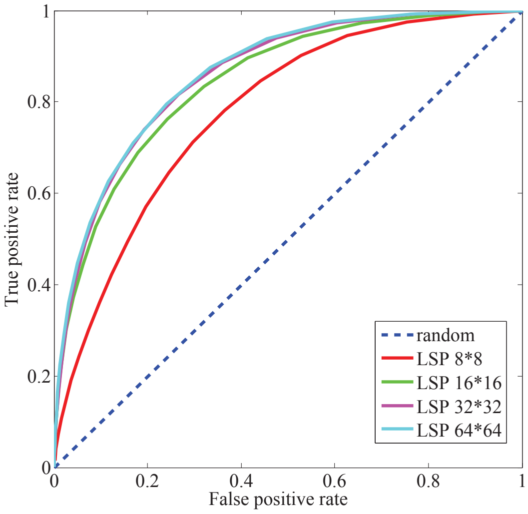
(d) *MSRA10K* dataset

Fig. 10 PR curves for LSP using different patch decomposition on *SED1*, *NJU400*, *ASD*, *MSRA10K* datasets, respectively.

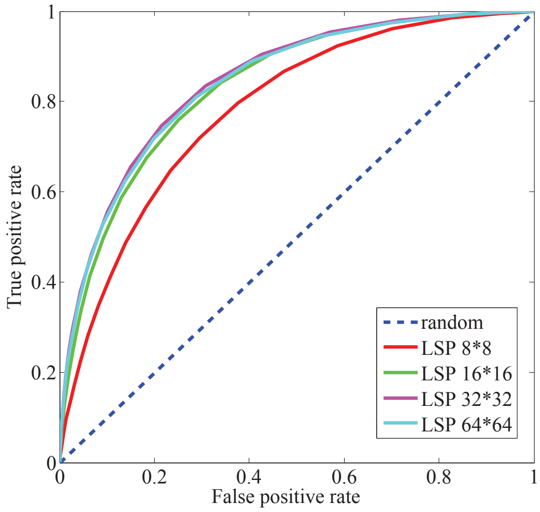
dataset with PR curve and ROC curve, respectively. It shows that FT+SP has better performance than FT but worse performance than LSP. It validates the importance of location information in salient object detection.

5.3 Comparison

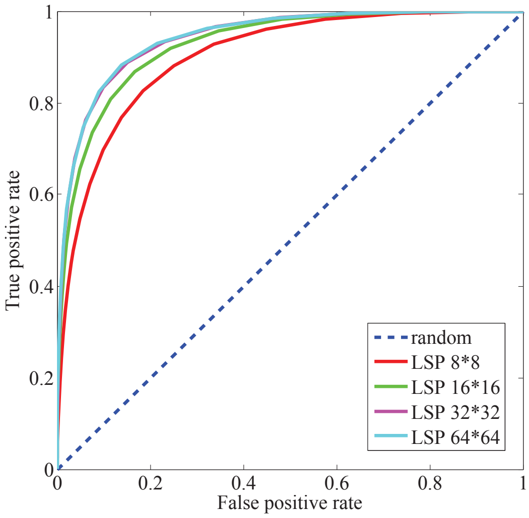
We also compared the location based saliency detection approach with the state-of-the-art saliency detection methods. All the saliency detection results of 17 compared methods are provided by *MSRA10K* dataset [12]. Figure 13-14 show the comparison results of PR curves and ROC curves for different saliency detection methods. It shows that the performance of LSP is better than most state-of-the-art saliency detection methods, though it is completely initialized by location information and it only uses color information in saliency



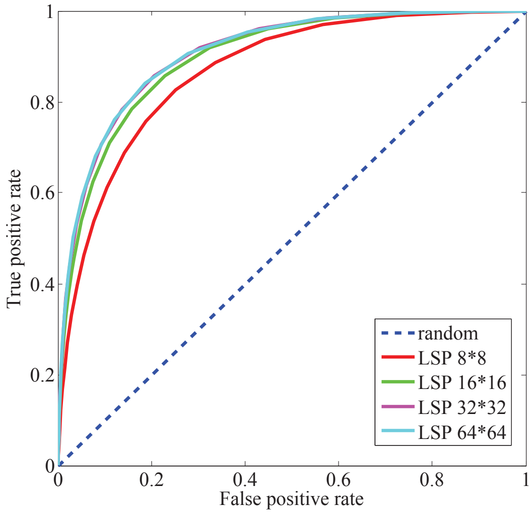
(a) *SED1* dataset



(b) *NJU400* dataset



(c) *ASD* dataset

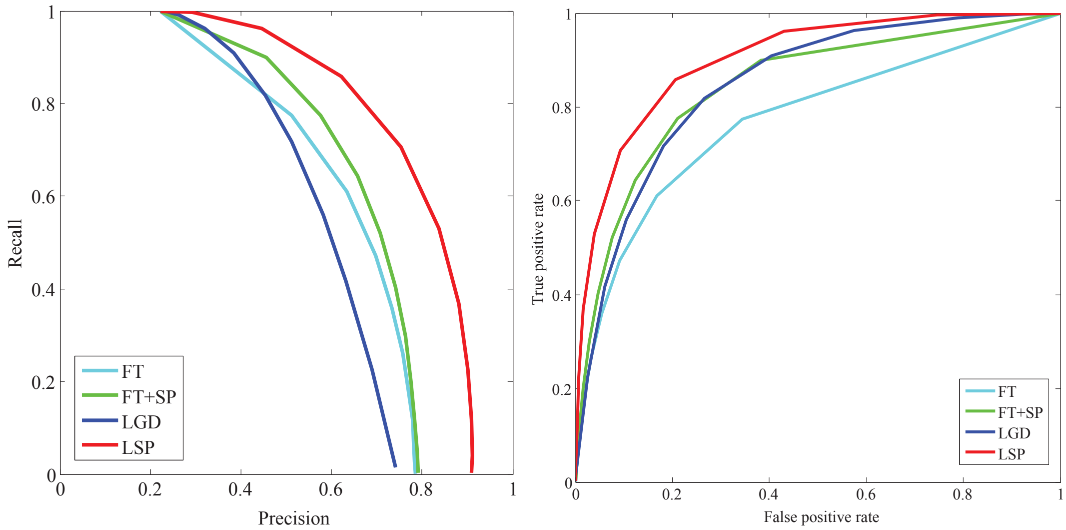


(d) *MSRA10K* dataset

Fig. 11 ROC curves for LSP using different patch decomposition on *SED1*, *NJU400*, *ASD*, *MSRA10K* datasets, respectively.

Table 1 Comparison of running time for different saliency detection methods on *MSRA10K* dataset.

Method	Code	Time	Method	Code	Time
AC	Matlab	0.109	LC	C++	0.018
AIM	Matlab	4.288	MSS	Matlab	0.106
CA	Matlab	53.1	RC	C++	0.254
CB	M&C	5.568	SEG	Matlab	4.921
FT	C++	0.102	SeR	Matlab	1.019
GB	Matlab	1.614	SR	Matlab	0.064
HC	C++	0.019	SUN	Matlab	1.116
IM	Matlab	0.991	SWD	Matlab	0.100
IT	Matlab	0.611	LSP 32×32	Matlab	0.343



(a) PR curves (b) ROC curves

Fig. 12 PR curves and ROC curves for comparison of location based methods and color based methods on *MSRA10K* dataset.

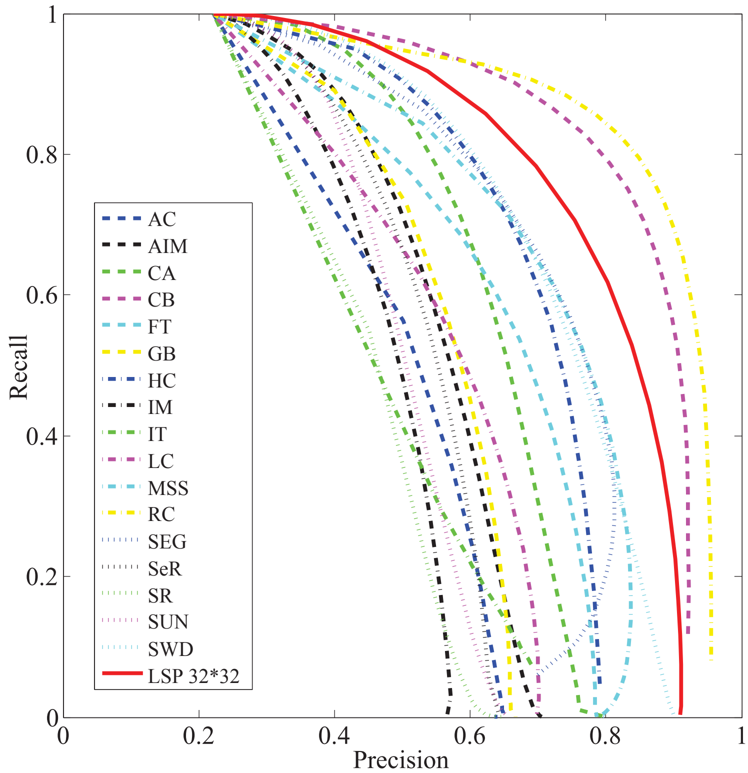


Fig. 13 Comparison of PR curves for different saliency detection methods on *MSRA10K* dataset.

propagation. And the two methods outperforming LSP, i.e., CB [20] and RC [12], both utilize center-bias in saliency detection together with some complex image content cues, such as shape prior and global color contrast. Table 1 shows the running time of different methods, here the running time

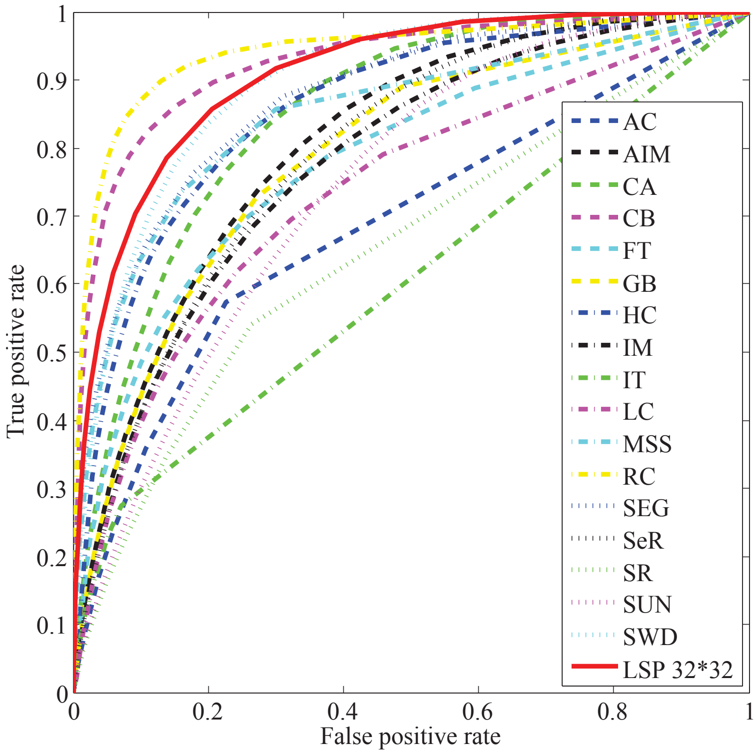


Fig. 14 Comparison of ROC curves for different saliency detection methods on *MSRA10K* dataset.

of the compared method is from [12] and the running time of our approaches is evaluated under the same condition with Dual Core 2.6 GHz CPU and 2GB RAM. It shows that the running time LSP is acceptable with Matlab implementation.

5.4 Discussion

5.4.1 Object location

One critical problem is whether location based saliency detection approach keeps effective when the salient objects are not located near to image centers.

We make statistics on the distribution of salient objects' center positions in *MSRA10K* dataset, i.e., normalizing the distance between salient object center and image center by dividing half length of image diagonal for each image. Figure 15 shows a histogram of salient object center distribution, in which the horizontal axis denotes the normalized salient object center position and the vertical axis denotes the number of images in *MSRA10K* dataset. We select the bins including more than 50 images, i.e. selecting 11 bins from 0.05 to 0.55, and calculate the average F_β with the adaptive threshold strategy for binarization [1], here $\beta^2 = 0.3$. Figure 16 shows the comparison results of different saliency detection methods when salient object center positions change, here horizontal axis denotes the normalized salient object center position. It shows that LSP

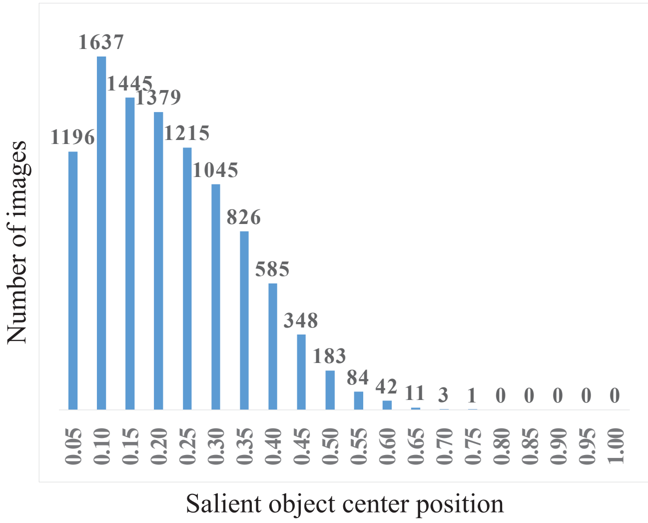


Fig. 15 Distribution of salient object center positions on *MSRA10K* dataset.

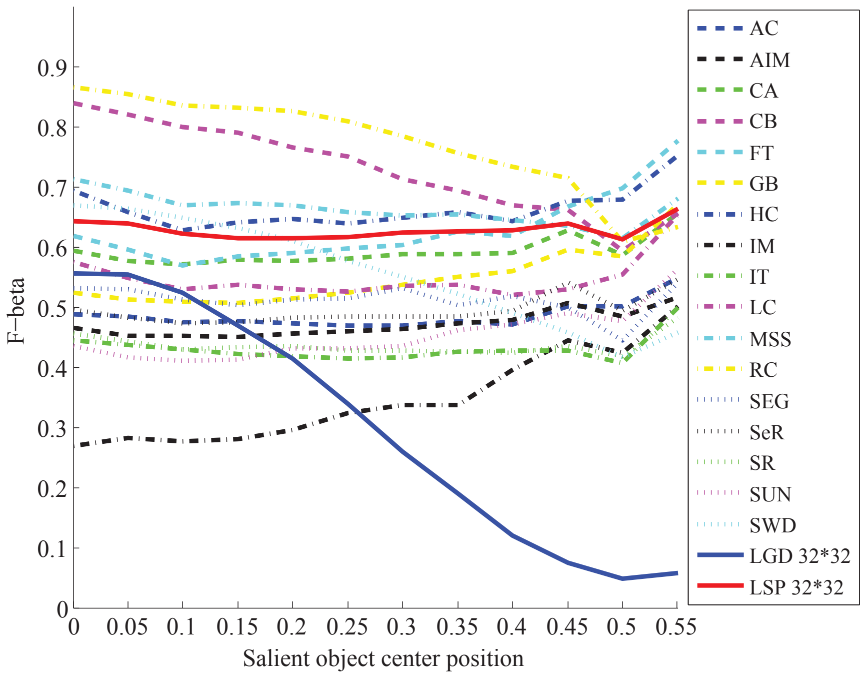
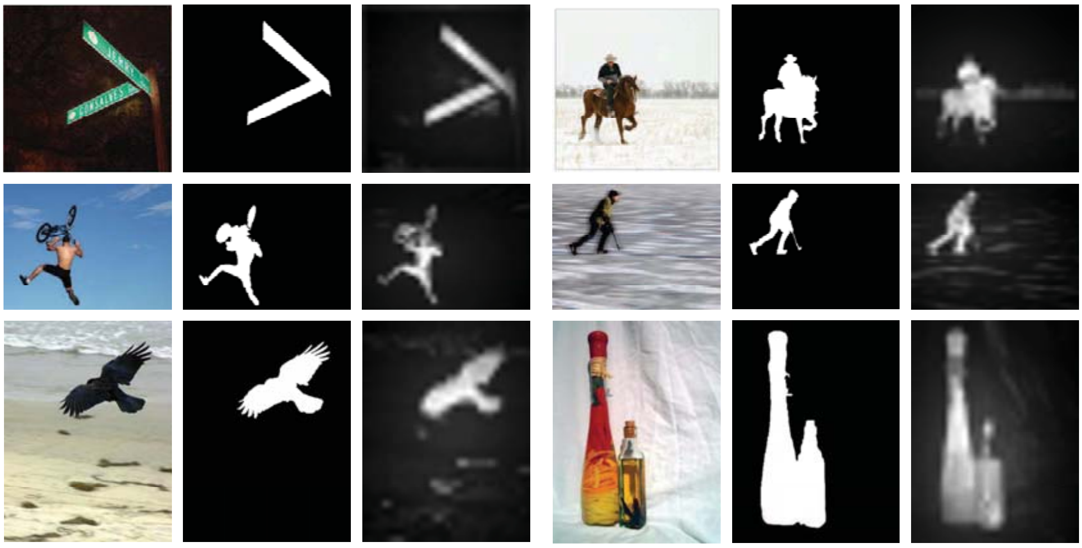


Fig. 16 Comparison of salient object detection performance with different salient object center positions.

obtains stable performance when the salient object center becomes far from image center. Figure 17 shows some examples from the above four datasets, in which the salient objects are located near to image boundaries or in the corners. It shows that LSP keeps effective when the salient objects are located in different positions.



(a) (b) (c) (d) (e) (f)

Fig. 17 Examples of saliency maps generated by LSP when the salient objects are not located near to image centers. (a) and (d) Source images. (b) and (e) Ground truths. (c) and (f) Saliency maps generated by LSP 32×32.

5.4.2 Multiple salient objects

Another interesting problem is whether location based saliency detection approach can handle multiple salient objects detection. We validated LSP on *SED2* dataset [2], which contains 100 images with two salient objects. Figure 19 shows the performance of LSP using different patch decomposition on $SED2$ dataset. We can find that LSP has good performance on $SED2$ dataset, which shows location based saliency detection approach can handle multiple salient objects detection. Figure 18 shows some examples of saliency maps generated by LSP 32×32 on $SED2$ dataset.

5.4.3 Limitation

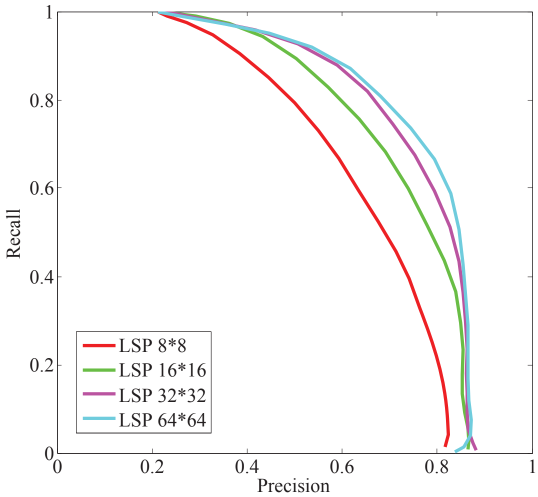
In the experiments, we also find some limitations of location based saliency detection approach. For only color similarity is used in saliency propagation, the performance of LSP will unstable when the objects and background are complex in color composition or they cannot be distinguished (Figure 20).

Another limitation is LSP may be sensitive to patch decomposition in rare cases. As shown in Figure 21, with some patch decomposition ((c) and (e) in the top row and (e) in the bottom row), LSP may confuse salient objects and background.

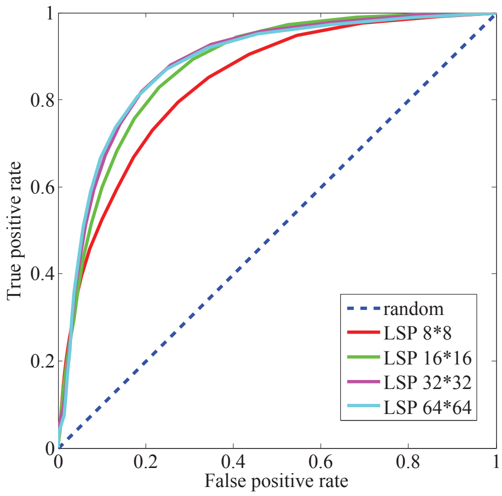
6 Conclusions

In this paper, we provide direct and quantitative analysis of the importance of location information in saliency detection. Based on the analysis of

Fig. 18 Examples of saliency maps generated by LSP using different patch decomposition on *SED2* dataset. (top row) Source images. (middle row) Ground truths. (bottom row) Saliency maps generated by LSP 32×32.



(a) PR curves



(b) ROC curves

Fig. 19 PR curves and ROC curves for LSP using different patch decomposition on *SED2* dataset.

relationship between image content location and salient object distribution, we find that location information has obvious influence to saliency detection. Furthermore, we propose a location based saliency detection approach, which is completely initialized by location information and only uses color information in saliency propagation. It shows that location based saliency detection approach outperforms most state-of-the-art saliency detection methods, and it can handle natural images with different object positions and multiple salient objects.

Our future work will focus on discovering the interaction relationships between location information and other cues in saliency detection. We will also consider the possibility to combine location information in video saliency detection.

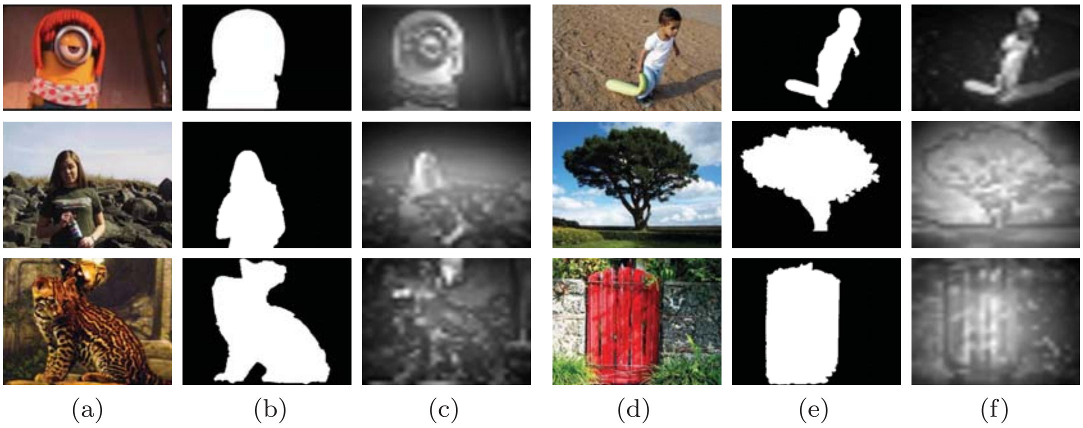


Fig. 20 Examples of our drawback when handling complex salient objects and background. (a) and (d) Source images. (b) and (e) Ground truths. (c) and (f) Saliency maps generated by LSP 32×32.

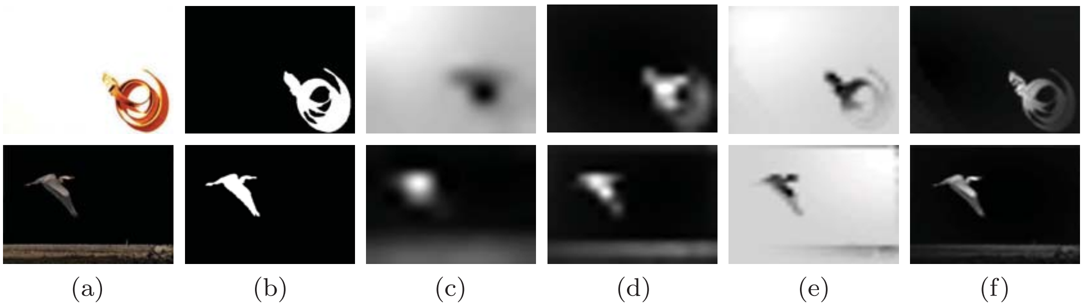


Fig. 21 Examples of our drawback when sensitive to patch numbers. (a) Source images. (b) Ground truths. (c)-(f) Saliency maps generated by LSP using patch decomposition in 8×8, 16×16, 32×32 and 64×64, respectively.

Acknowledgements The authors would like to thank the anonymous reviewers for the associate editor for their valuable comments, which have greatly helped us to make improvements, and Jingfan Guo for his contribution in experiment. This paper is supported by the National Science Foundation of China (No. 61321491, 61202320), Research Project of Excellent State Key Laboratory (No.61223003), Natural Science Foundation of Jiangsu Province (No.BK2012304), and National Special Fund (No.2011ZX05035-004-004HZ).

References

1. Achanta, Radhakrishna, Hemami, Sheila, Estrada, Francisco, and Susstrunk, Sabine. Frequency-tuned salient region detection. IEEE International Conference on Computer Vision and Pattern Recognition, 1597–1604. IEEE (2009)
2. Alpert, Sharon, Galun, Meirav, Basri, Ronen, and Brandt, Achi. Image segmentation by probabilistic bottom-up aggregation and cue integration. IEEE International Conference on Computer Vision and Pattern Recognition, 1–8. IEEE (2007)
3. Barnbaum, Bruce. The art of photography: An approach to personal expression. Rocky Nook Press, 2010.
4. Bao, Bing-Kun, Li, Teng, and Yan, Shuicheng. Hidden-concept driven multilabel image annotation and label ranking. IEEE Transactions on Multimedia, 14(1):199–210. IEEE (2012)
5. Bao, Lei, Lu, Jianjiang, Li, Yang, and Shi, Yanwei. A saliency detection model using shearlet transform. Multimedia Tools and Applications, 1–14. Springer (2014)

6. Bimbo, Alberto Del, and Pala, Pietro. Visual image retrieval by elastic matching of user sketches. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(2):121–132. IEEE (1997)
7. Borji, Ali, Sihite, Dicky N., and Itti, Laurent. Salient object detection: A benchmark. European Conference on Computer Vision, 414–429. Springer (2012)
8. Borji, Ali, and Itti, Laurent. State-of-the-art in visual attention modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(1):185–207. IEEE (2013)
9. Borji, Ali, Sihite, Dicky N, and Itti, Laurent. Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. IEEE Transactions on Image Processing, 22(1):55–69. IEEE (2013)
10. Borji, Ali, Tavakoli, Hamed R., Sihite, Dicky N., and Itti, Laurent. Analysis of scores, datasets, and models in visual saliency prediction. IEEE International Conference on Computer Vision, 921–928. IEEE (2013)
11. Borji, Ali, Cheng, Ming-Ming, Jiang, Huaizu, and Li, Jia. Salient object detection: A survey. arXiv:1411.5878. arXiv (2014)
12. Cheng, Ming-Ming, Mitra, Niloy J., Huang, Xiaolei, Torr, Philip H.S., and Hu, Shi-Min. Global contrast based salient region detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(3):409–416. IEEE (2015)
13. Cheng, Ming-Ming, Mitra, Niloy J., Huang, Xiaolei, and Hu, Shi-Min. Salientshape: Group saliency in image collections. The Visual Computer, 30(4):443–453. Springer (2014)
14. Desingh, Karthik, Krishna, K. Madhava, Rajan, Deepu, and Jawahar, C.V. Depth really matters: Improving visual salient region detection with depth. British Machine Vision Conference, 98:1–11 (2013)
15. Goferman, Stas, Zelnik-Manor, Lihi, and Tal, Ayellet. Context-aware saliency detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(10):1915–1926. IEEE (2012)
16. Gopalakrishnan, Viswanath, Hu, Yiqun, and Rajan, Deepu. Random walks on graphs to model saliency in images. IEEE Conference on Computer Vision and Pattern Recognition, 1698–1705. IEEE (2009)
17. Hou, Xiaodi, and Zhang, Liqing. Saliency detection: A spectral residual approach. IEEE Conference on Computer Vision and Pattern Recognition, 1–8. IEEE (2007)
18. Itti, Laurent, Koch, Christof, and Niebur, Ernst. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(11):1254–1259. IEEE (1998)
19. Jia, Yangqing and Han, Mei. Category-independent object-level saliency detection. IEEE International Conference on Computer Vision, 1761–1768. IEEE (2013)
20. Jiang, Huaizu, Wang, Jingdong, Yuan, Zejian, Liu, Tie, Zheng, Nanning, and Li, Shipeng. Automatic salient object segmentation based on context and shape prior. British Machine Vision Conference, 3(4):1–7 (2011)
21. Jiang, Peng, Ling, Haibin, Yu, Jingyi, and Peng, Jingliang. Salient region detection by UFO: Uniqueness, focusness and objectness. IEEE International Conference on Computer Vision, 1976–1983. IEEE (2013)
22. Jing, Yushi, and Baluja, Shumeet. Visualrank: Applying pagerank to large-scale image search. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(11):1877–1890. IEEE (2008)
23. Ju, Ran, Xu, Xiangyang, Yang, Yang, and Wu, Gangshan. Stereo GrabCut: Interactive and consistent object extraction for stereo images. Pacific-Rim Conference on Multimedia, 418–429. Springer (2013)
24. Ju, Ran, Ge, Ling, Geng, Wenjing, Ren, Tongwei, and Wu, Gangshan. Depth saliency based on anisotropic center-surround difference. IEEE International Conference on Image Processing. IEEE (2014)
25. Judd, Tilke, Ehinger, Krista, Durand, Frédo, and Torralba, Antonio. Learning to predict where humans look. IEEE International Conference on Computer Vision, 2106–2113. IEEE (2009)
26. Kim, Wonjun, Jung, Chanho, and Kim, Changick. Spatiotemporal saliency detection and its applications in static and dynamic scenes. IEEE Transactions on Circuits and Systems for Video Technology, 21(4):446–456. IEEE (2011)

27. Lang, Congyan, Liu, Guangcan, Yu, Jian, and Yan, Shuicheng. Saliency detection by multitask sparsity pursuit. IEEE Transactions on Image Processing, 21(3):1327–1338. IEEE (2012)
28. Li, Nianyi, Ye, Jinwei, Ji, Yu, Ling, Haibin, and Yu, Jingyi. Saliency detection on light field. IEEE International Conference on Computer Vision and Pattern Recognition, 2806–2813. IEEE (2014)
29. Liu, Taosheng, Slotnick, Scott D., Serences, John T., and Yantis, Steven. Cortical mechanisms of feature-based attentional control. Cerebral Cortex, 13(12):1334–1343. Oxford University Press (2003)
30. Liu, Tie, Yuan, Zejian, Sun, Jian, Wang, Jingdong, Zheng, Nanning, Tang, Xiaoou, and Shum, Heung-Yeung. Learning to detect a salient object. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(2):353–367. IEEE (2011)
31. Liu, Qiang, Han, Tao, Sun, Yantao, Chu, Zhong, and Shen, Bingwen. A two step salient objects extraction framework based on image segmentation and saliency detection. Multimedia Tools and Applications, 67(1):231–247. Springer (2012)
32. Liu, Zhi, Zou, Wenbin, and Le Meur, Olivier. Saliency tree: A novel saliency detection framework. IEEE Transactions on Image Processing, 23(5):1937–1952. IEEE (2014)
33. Ma, Yu-Fei, and Zhang, Hong-Jiang. Contrast-based image attention analysis by using fuzzy growing. ACM International Conference on Multimedia, 374–381. ACM (2003)
34. Marchesotti, Luca, Cifarelli, Claudio, and Csurka, Gabriela. A framework for visual saliency detection with applications to image thumbnailing. IEEE International Conference on Computer Vision, 2232–2239. IEEE (2009)
35. Margolin, Ran, Tal, Ayellet, and Zelnik-Manor, Lihi. What makes a patch distinct? IEEE International Conference on Computer Vision and Pattern Recognition, 1139–1146. IEEE (2013)
36. Movahedi, Vida, and Elder, James H. Design and perceptual validation of performance measures for salient object segmentation. IEEE International Conference on Computer Vision and Pattern Recognition Workshops, 49–56. IEEE (2010)
37. Qi, Guo-Jun, Hua, Xian-Sheng, Rui, Yong, Tang, Jinhui, Mei, Tao, and Zhang, Hong-Jiang. Correlative multi-label video annotation. ACM International Conference on Multimedia, 17–26. ACM (2007)
38. Qiu, Zhongyan, Ren, Tongwei, Liu, Yan, Bei, Jia, and Song, Muyang. Image retargeting by combining region warping and occlusion. Pacific-Rim Conference on Multimedia, 200–210. Springer (2013)
39. Ren, Tongwei, Liu, Yan, and Wu, Gangshan. Image retargeting based on global energy optimization. IEEE International Conference on Multimedia and Expo, 406–409. IEEE (2009)
40. Ren, Tongwei, Qiu, Zhongyan, Liu, Yan, Yu, Tong, and Bei, Jia. Soft-assigned bag of features for object tracking. Multimedia Systems, 1–17. Springer (2014)
41. Schauerte, Boris, and Stiefelhagen, Rainer. How the distribution of salient objects in images influences salient object detection. IEEE International Conference on Image Processing, 74–78. IEEE (2013)
42. Tatler, Benjamin W., Baddeley, Roland J., and Gilchrist, Iain D. Visual correlates of fixation selection: Effects of scale and time. Vision Research, 45(5):643–659. Elsevier (2005)
43. Wang, Meng, Hua, Xian-Sheng, Hong, Richang, Tang, Jinhui, Qi, G.J., and Song, Yan. Unified video annotation via multigraph learning. IEEE Transactions on Circuits and Systems for Video Technology, 19(5):733–746. IEEE (2009)
44. Wang, Jing, Lu, Ke, Pan, Daru, He, Ning, and Bao, Bing-kun. Robust object removal with an exemplar-based image inpainting approach. Neurocomputing, 123:150–155. Elsevier (2014)
45. Xu, Xiangyang, Geng, Wenjing, Ju, Ran, Yang, Yang, Ren, Tongwei and Wu, Gangshan. OBSIR: Object-based stereo image retrieval. IEEE International Conference on Multimedia and Expo. IEEE (2014)
46. Yang, Xiaoshan, Zhang, Tianzhu, and Xu, Changsheng. Locality discriminative coding for image classification. International Conference on Internet Multimedia Computing and Service, 52–55. ACM (2013)
47. Zha, Zheng-Jun, Hua, Xian-Sheng, Mei, Tao, Wang, Jingdong, Qi, G.J., and Wang, Zengfu. Joint multi-label multi-instance learning for image classification. IEEE Conference on Computer Vision and Pattern Recognition, 1–8. IEEE (2008)

48. Zhang, Shiliang, Tian, Qi, Huang, Qingming, Gao, Wen, and Rui, Yong. Multi-order visual phrase for scalable image search. International Conference on Internet Multimedia Computing and Service, 145–149. ACM (2013)
49. Zhong, Wei, Lu, Huchuan, and Yang, Ming-Hsuan. Robust object tracking via sparsity-based collaborative model. IEEE Conference on Computer Vision and Pattern Recognition, 1838–1845. IEEE (2012)
50. Zhou, Wengang, Tian, Qi, Lu, Yijuan, Yang, Linjun, and Li, Houqiang. Latent visual context learning for web image applications. Pattern Recognition, 44(10):2263–2273. Elsevier (2011)
51. Zhu, Shiai, Wang, Gang, Ngo, Chong-Wah, and Jiang, Yu-Gang. On the sampling of web images for learning visual concept classifiers. ACM International Conference on Image and Video Retrieval, 50–57. ACM (2010)