

Visual saliency detection based on multi-scale and multi-channel mean

Lang Sun · Yan Tang · Hong Zhang

Received: 18 January 2014 / Revised: 18 August 2014 / Accepted: 6 October 2014 /

Published online: 24 October 2014

© Springer Science+Business Media New York 2014

Abstract This paper proposes an effective method for visual saliency detection based on multi-scale and multi-channel mean. 2-D wavelet transform is used to decompose and reconstruct image. Bicubic interpolation algorithm is applied to narrow the filtered image in multi-scale. We take the distances between the narrowed images and the means of their channels as saliency values, and we only reserve part values which are not less than the mean saliency of the given image. Bicubic interpolation algorithm is applied again to amplify the images in multi-scale, and then the saliency map is calculated by adding the amplified images. Finally, linear normalization is employed to obtain the final saliency map. Experimental results show that the proposed method outperforms 9 state-of-the-art methods both on the definition and accuracy of salient detection.

Keywords Visual saliency · Saliency map · 2-D wavelet transform · Bicubic interpolation · Multi-scale · Multi-channel

1 Introduction

Images have become a major source for information. Humans routinely and effortlessly judge the importance of image regions, and focus attention on important parts. This ability, viewed as visual attention, plays a significant role in analyzing and extracting importance visual information in engineering applications. Let the computer has the ability to capture salient information is a hot research topic. A profound challenge in computer vision is to make the computer understand the surrounding scene via image, while three vital tasks of this field are focusing on perceiving the key object, detecting the shapes and contours, and capturing the

L. Sun · Y. Tang (✉)

School of Computer and Information Science, Southwest University, NO.2 Rd Tiansheng, Beibei District, Chongqing, China
e-mail: ytang@swu.edu.cn

L. Sun

e-mail: sylensun@swu.edu.cn

H. Zhang

Department of Computing Science, Alberta University, Edmonton, Alberta, Canada
e-mail: zhang@cs.ualberta.ca

context information. To achieve this goal, obtaining the visual saliency information from the image is the most basic and important step. Correctly extracting these salient regions can improve the efficiency and accuracy in image analysis and processing, reduce the complexity of calculations, allow preferential allocation of computational resources, break down the barrier between the content understanding and underlying characteristics, and make a higher level understanding of the image possible. Saliency maps are widely used in many computer vision and pattern recognition applications, such as image retrieval [7, 27], image segmentation [5, 19], object detection [6], image compression [21] and image fusion [3]. Exhaustive study of saliency detection has a far-reaching significance in improving the performance of image understanding and image analysis systems, as well as in enhancing the level of application of image processing technologies.

Human visual system scans the scene both in a rapid, bottom-up, task-independent and data driven saliency extraction as well as in a slower, top-down, task-dependent and goal driven saliency extraction. So far, most research papers estimate visual saliency in a bottom-up way for its task-independent characteristic. Over the past decades, many excellent methods have been proposed to detection visual saliency. Based on the highly influential biological inspired early representation model introduced by Koch and Ullman [12], Itti et al. [9] proposed one of the earliest works in visual saliency detection. They applied central-surrounded differences across multi-scale and multi-features visual space to obtain saliency map, and found the salient regions from the saliency map by the saliency values from strong to weak, and then represented the salient regions by a fixed radius circle. Ma and Zhang [15] incorporated a fuzzy growth technique in the saliency method for detecting different levels of saliency. Zhai and Shah [24] proposed a visual saliency method based on single-channel histogram contrast. Combined with Fourier spectral residual, Hou and Zhang [8] detected the image salient regions based on image spectral. Based on Bayesian framework, Zhang et al. [25] analyzed the statistical information of natural image to detect the image saliency. Combined with information theory, Bruce et al. [4] calculated the image saliency by using information maximization. Achanta et al. [1] proposed a purely computational method. In order to adapt large salient objects regions, they proposed an improved method by adjusting the image spatial frequency domain [2]. Based on mathematical and statistical principles, Murray et al. [18] applied inverse wavelet transform to process image, and then to extract saliency map by using nonlinear weighted scale fusion. Xie et al. [23] proposed a novel method for bottom-up saliency within the Bayesian framework by exploiting low and mid level cues. These methods can fleetly find the objects regions of human interest. But most of these methods center on human eye fixation prediction or certain task of salient object detection, and do not have a strong expansibility, for example, these methods have a good performance on human eye fixation prediction, while the results of salient object detection may be not satisfying, and vice versa. Furthermore, by defining visual saliency at each location as the dissimilarity between itself and its local neighborhood or global counterparts, many state-of-the-art methods have to segment image into block or region for visual saliency detection. The quality of image segmentation directly affects saliency extraction and method efficiency.

We propose an effective visual saliency detection method based on multi-scale and multi-channel mean to improve the definition and accuracy of the salient detection. 2-D wavelet transform is used to decompose and reconstruct image, where can effectively filter the background information of image and highlight salient regions without image segmentation. The bicubic interpolation algorithm is used to narrow the filtered image in multi-scale. We take the distances between the narrowed images and the means of their channels as saliency values, where can avoid tending to produce higher saliency values near edges instead of uniformly highlighting salient objects. In order to filter the background noises of saliency maps, we only

reserve part values which are not less than the mean saliency of a given image. Bicubic interpolation algorithm is used again to amplify the images in multi-scale, and then the saliency map is calculated by adding the amplified images. Finally, linear normalization is employed to obtain the final saliency map. We provide an objective comparison of the saliency maps against 9 state-of-the-art methods. Our method outperforms all of these methods in terms of definition and accuracy.

2 Relevant theories

2.1 Color space conversion

With the high correlation between the three components of RGB color space, and there is no direct interrelation with intuitional color concepts, such as hue, saturation and brightness, we would better not process these components directly. As a color model for the human eyes to distinguish compatibly, the model of HSV color space is with good perception characteristic and the ability to easily convert to the model of RGB color space, while it well reflects the human feelings for color and makes against image processing [22, 26]. As the widely used in image saliency detection of HSV color space model, we detect image saliency in HSV color space. Using the following formulas to transform RGB color space to HSV color space:

$$H = \begin{cases} \theta, G \geq B \\ 2\pi - \theta, G < B \end{cases} \quad (1)$$

$$S = 1 - \frac{3}{(R + G + B)} \min(R, G, B) \quad (2)$$

$$V = \frac{1}{\sqrt{3}}(R + G + B) \quad (3)$$

Where hue value H is basic pure color, saturation S is the ratio of white light doped into color, brightness value V is the ratio of black light doped into color, $R, G, B \in [0, 1]$ and $\theta = \arccos^{-1}((R-G) \frac{+(R-B)}{2\sqrt{(R-G)}}^2 + (R-B)(G-B))$.

2.2 2-D wavelet transform

Multi-scale analysis was introduced into wavelet analysis by Mallat in 1989. He proposed the concept of multi-resolution analysis, and gave a general method for constructing orthogonal wavelet basis and a fast wavelet algorithm, viewed as Mallat algorithm, which relatives to Fast Fourier Transform [16, 17]. Wavelet transform can be well matched human visual system. Low frequency coefficients of wavelet blocks correspond to the average luminance of image blocks, where large coefficients represent the high average luminance of the image regions, and small coefficients represent the small average luminance of the image regions. High frequency coefficients represent the texture and the edge portion of the image, where large absolute values of the coefficients represent the complex texture and edge portion of the image, small absolute values of the coefficients represent the smooth part of the image. Therefore, we separate the low and the high frequency of the image by 2-D wavelet decomposition, highlight the low frequency components and attenuate the high frequency components, and then we

reuse 2-D wavelet transform for image reconstruction, after these, the salient region can be enhanced and the noises of the background are filtered. We consider $\{c_{k,m}^{j+1}\}$ as a 2-D image, where j is resolution ratio, k and m are row index and column index respectively. Then the decomposition and reconstruction algorithm for 2-D wavelet transform are as follows:

$$\begin{cases} c_{k,m}^j = \sum_{l,n} \tilde{h}_{l-2k} \tilde{h}_{m-2n} c_{l,n}^{j+1} \\ d_{k,m}^{j,1} = \sum_{l,n} \tilde{h}_{l-2k} \tilde{g}_{m-2n} c_{l,n}^{j+1} \\ d_{k,m}^{j,2} = \sum_{l,n} \tilde{g}_{l-2k} \tilde{h}_{m-2n} c_{l,n}^{j+1} \\ d_{k,m}^{j,3} = \sum_{l,n} \tilde{g}_{l-2k} \tilde{g}_{m-2n} c_{l,n}^{j+1} \end{cases} \tag{4}$$

$$\begin{aligned} c_{k,m}^{j+1} &= \sum_{l,n} h_{k-2l} h_{m-2n} c_{l,n}^j + \sum_{l,n} h_{k-2l} g_{m-2n} d_{l,n}^{j,1} + \\ &\sum_{l,n} g_{k-2l} h_{m-2n} d_{l,n}^{j,2} + \sum_{l,n} g_{k-2l} g_{m-2n} d_{l,n}^{j,3} \end{aligned} \tag{5}$$

Where \tilde{h}, \tilde{g}, h and g are biorthogonal filter banks, $c_{k,m}^j, d_{k,m}^{j,1}, d_{k,m}^{j,2}$ and $d_{k,m}^{j,3}$ are low frequency coefficient, horizontal high frequency coefficient, vertical high frequency coefficient and diagonal high frequency coefficient respectively. Filter banks of 2-D wavelet transform is illustrated in Fig. 1.

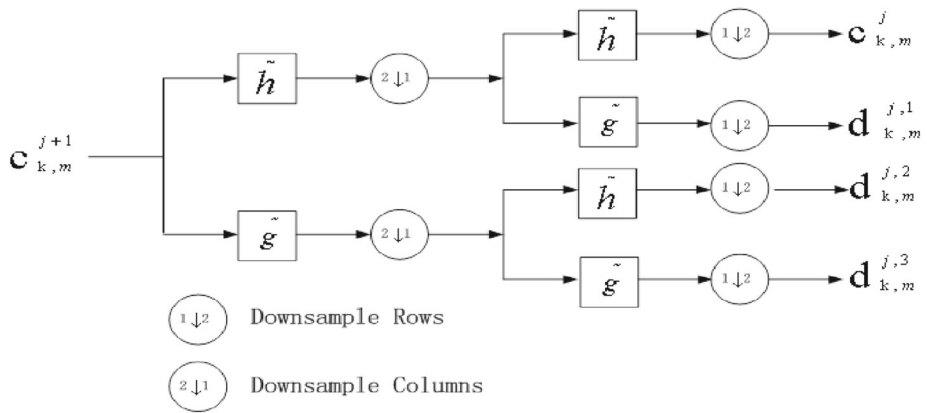
2.3 Bicubic interpolation

Multi-scale analysis is conventional and useful for visual saliency detection, and it is widely used in many literatures [9, 13, 14]. For a same object, attention on small scale image focuses on a whole object, while attention on large scale image cares more about the local details. Small scale image makes an integrate object more continuous and conspicuous. Bicubic interpolation [11] is for narrowing and amplifying images at different scales. Bicubic interpolation algorithm retains more image detail and is able to get a relatively clear picture quality in image scaling.

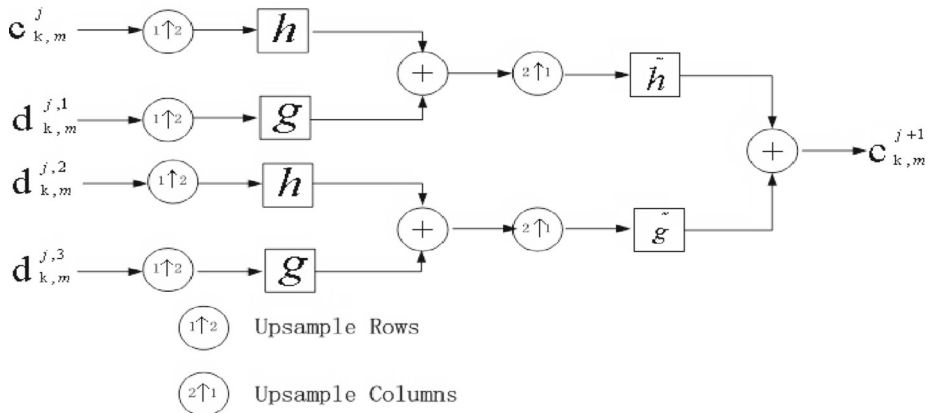
The process of bicubic interpolation is to create a continuous and simple analytical model based on the known observation pixel points. Bicubic interpolation is cubic interpolation in two dimensions. It obtains target pixel value $f(i + m, j + n)$ by calculating the weighted mean of the neighbor 4×4 matrix of the float coordinate $(i + m, j + n)$, where i and j represent the column and row of pixels, m and n represent integer floating points between -1 and 2. Interpolation basic function is the foundation for bicubic interpolation. The definition of the basic function $u(s)$ is as follows:

$$u(s) = \begin{cases} \frac{3}{2} |s|^3 - \frac{5}{2} |s|^2 + 1, & |s| < 1 \\ -\frac{1}{2} |s|^3 + \frac{5}{2} |s|^2 - 4|s| + 2, & 1 \leq |s| < 2 \\ 0, & 2 \leq |s|. \end{cases} \tag{6}$$

As illustrated in Fig. 2, $u(s)$ is approaching cubic interpolation curve $\sin(s * \pi) / (s * \pi)$.



(a) 2-D wavelet decomposition



(b) 2-D wavelet reconstruction

Fig. 1 Filter banks of 2-D wavelet transform (a) 2-D wavelet decomposition (b) 2-D wavelet reconstruction

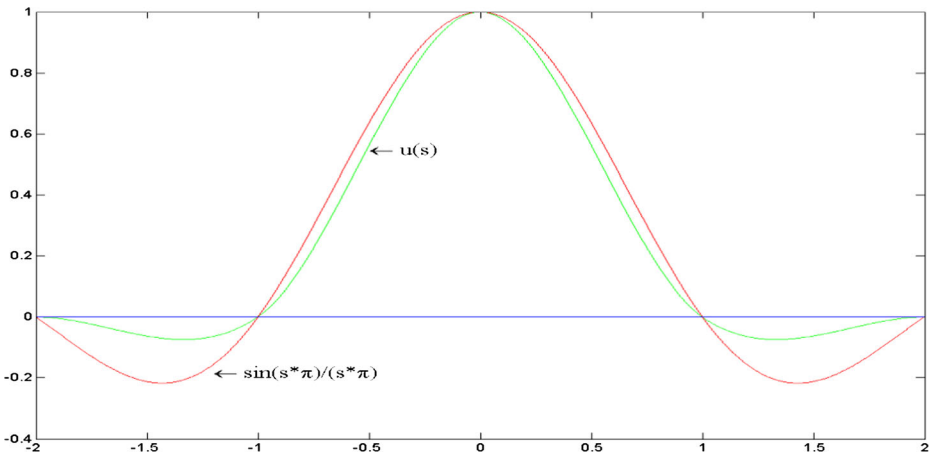


Fig. 2 Cubic interpolation curve and its approximation curve

$f(i + m, j + n)$ can be obtained by the following interpolation formula:

$$f(i + m, j + n) = \begin{bmatrix} u(m+1) \\ u(m) \\ u(m-1) \\ u(m-2) \end{bmatrix}^T \begin{bmatrix} f(i-1, j-1) & f(i-1, j) & f(i-1, j+1) & f(i-1, j+2) \\ f(i, j-1) & f(i, j) & f(i, j+1) & f(i, j+2) \\ f(i+1, j-1) & f(i+1, j) & f(i+1, j+1) & f(i+1, j+2) \\ f(i+2, j-1) & f(i+2, j) & f(i+2, j+1) & f(i+2, j+2) \end{bmatrix} \begin{bmatrix} u(n+1) \\ u(n) \\ u(n-1) \\ u(n-2) \end{bmatrix} \tag{7}$$

3 Visual saliency detection based on multi-scale and multi-channel mean

Many researchers have proposed methods of saliency, which always center on human eye fixation or certain task of salient object detection, and invariably require image segmentation. These often lead to a weak expansibility, and affect saliency extraction and algorithm efficiency. Aiming at these problems, we propose an effective visual saliency detection method based on multi-scale and multi-channel mean. This method includes four parts: image filtering, resize image, calculate saliency values and generate saliency map. The framework of the proposed method is shown in Fig. 3.

3.1 Image filtering

Human visual physiological characteristics determine the sensitivity of the low frequency signal is greater than the sensitivity of the high frequency signal of an image. The low frequency coefficients describe the main energy part of the image and the high frequency coefficients describe the details. Therefore, we highlight the low frequency components and attenuate the high frequency components by 2-D wavelet transform in HSV color space. For efficiency and simplicity, we consider three level 2-D wavelet transform for image decomposition and reconstruction. The tower structure of three-level 2-D wavelet transform is illustrated in Fig. 4.

Haar wavelet was proposed by Alfred Haar in 1909. It is the simplest transform in wavelet transform. In addition, it is the only orthogonal wavelet having symmetry and compact support [10, 20]. Therefore, we construct filter banks based on Haar wavelet. Filter function of Haar wavelet is as follows:

$$h_k = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } k = 0, 1 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Where h_k is a real filter. Combined with Section 2.2, we obtain $g_k = (-1)^k h_{1-k}$, $\tilde{h}_k = h_k$ and $\tilde{g}_k = g_k$. High frequency signal image and low frequency signal image after 2-D Haar wavelet decomposition and reconstruction are shown in Fig. 5(c) and (d) respectively. The filtered image is obtained by

$$I = \alpha I_H + \beta I_L \tag{9}$$

Where I_H and I_L are high frequency signal image and low frequency signal image respectively, α and β are weight parameters and we set them to be 0.9 and 1.1 respectively in our experiments. The filtered image I is the last picture in Fig. 5.

3.2 Resize image

Salient regions are associated with image scales. For a same object, attention on small scale image focuses on a whole object with same features, which raises the saliency in small scale

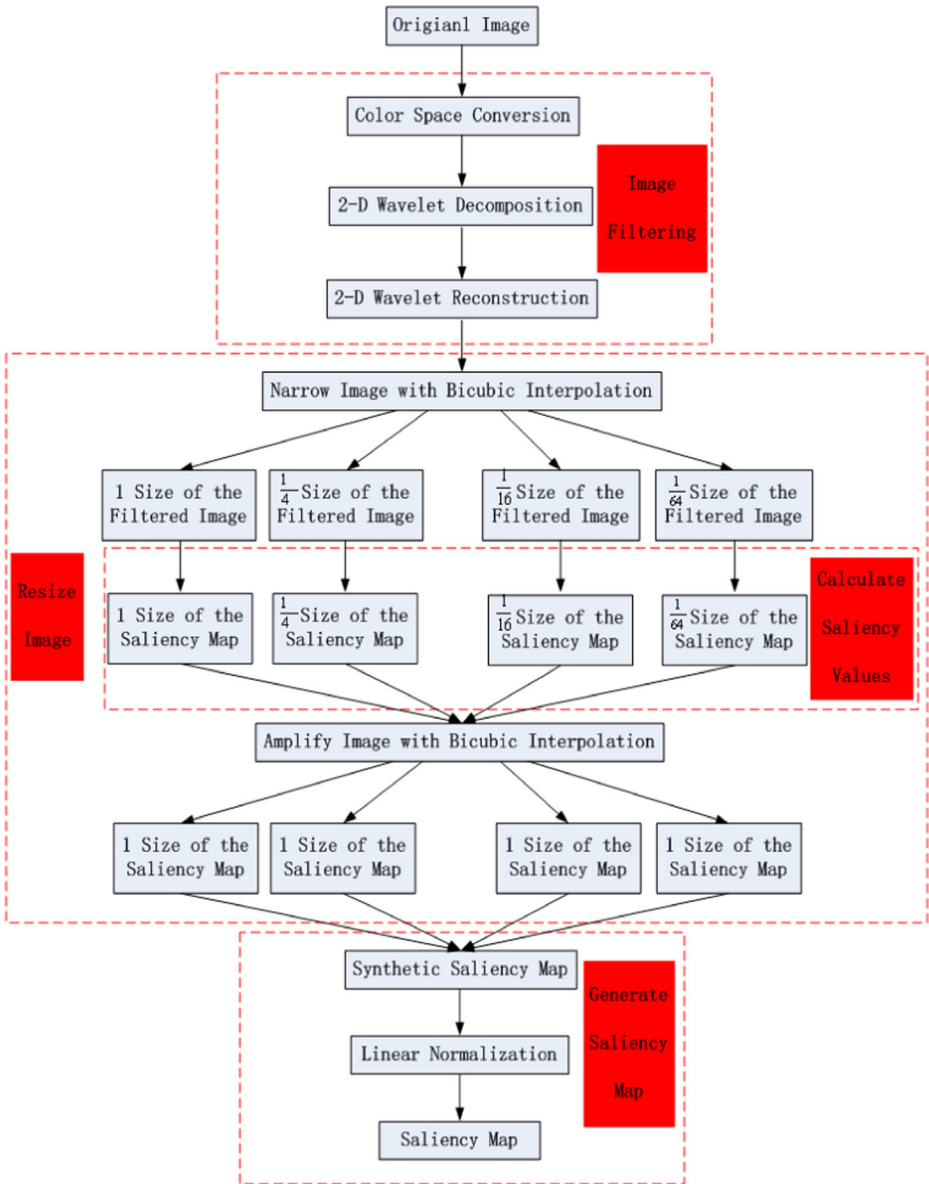


Fig. 3 Framework of the visual saliency detection based on multi-scale and multi-channel mean

regions, while attention on large scale image raises the saliency in large scale regions. Therefore, we resize image with bicubic interpolation, narrow image on filtered image to obtain images at different scales, and amplify saliency maps to obtain images at unified scale.

By doing wavelet decomposition, we obtain series sub-images of different resolutions, while each sub-image is 1/4 size of the filtered image. After doing three-level 2-D wavelet transform, we get 1/4, 1/16 and 1/64 size of the filtered image respectively. After many experiments, we found that a better experimental result can be obtained on four scales of the

LL_3	HL_3	HL_2	HL_1
LH_3	HH_3		
LH_2		HH_2	
LH_1		HH_1	

Fig. 4 Tower structure of three-level 2-D wavelet transform

filtered image I . We consider four different scales with scale factors $\beta_1=1, \beta_2=1/4, \beta_3=1/16$ and $\beta_4=1/64$, which means the same operations are implemented on the four scale image. The filtered image I is narrowed by bicubic interpolation algorithm on four scales, and the narrowed results are illustrated in Fig. 6.

3.3 Calculate saliency values

In order to avoid saliency computation process tending to produce higher saliency values near edges instead of uniformly highlighting salient objects, and ignoring spatial relationships across image parts, we calculate pixels means in each channel in HSV color space and take the distances between the images and the mean of their channels as saliency values. Our method of finding the saliency map SV_n for an image can be formulated as:

$$SV_n = (H_n - H_Mean_n)^2 + (S_n - S_Mean_n)^2 + (V_n - V_Mean_n)^2 \tag{10}$$

Where H_n, S_n and V_n are H, S, V channel of the filtered image in HSV color space respectively, H_Mean_n, S_Mean_n and V_Mean_n are the mean of each channel respectively. Multi-Scale saliency maps images are illustrated in Fig. 7.

From Fig. 7 we can clearly see that there are too many background noises in multi-scale saliency maps. In order to filter the background noises of saliency maps, highlight salient regions, and avoid poor image definition, we only reserve part values which are not less than

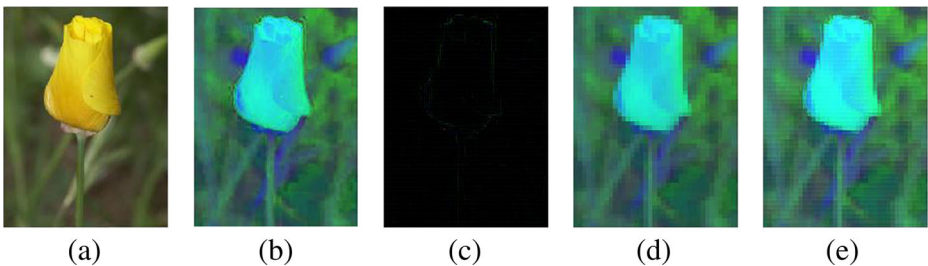


Fig. 5 **a** Original image **(b)** The original image in HSV color space **(c)** High frequency signal image **(d)** Low frequency signal image **(e)** Filtered image

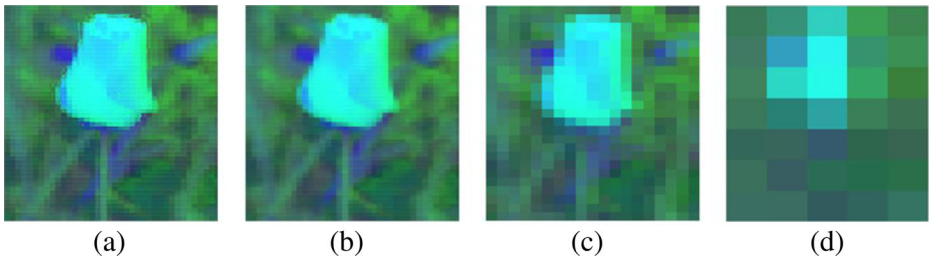


Fig. 6 Resized images. (a) 1 size of the filtered image (b) 1/4 size of the filtered image (c) 1/16 size of the filtered image (d) 1/64 size of the filtered image

the mean saliency of a given image. Combined with Section 2.3 and Section 3.2, amplify saliency maps with bicubic interpolation algorithm to the size of original image. Multi-Scale saliency maps and their corresponding amplified images are illustrated in Fig. 8.

3.4 Generate saliency map

We obtain the synthetic saliency map CS by a simple linear addition of 4 amplified saliency maps. The linear model is used to normalize CS to find the saliency map SM_1 . The final saliency map SM_1 is defined as,

$$SM_1 = 255 * \frac{CS - CS(min)}{CS(max) - CS(min)} \tag{11}$$

Where $CS(max)$ and $CS(min)$ are the maximum and the minimum values of pixels in CS .

Combined with Section 4.1, we can obtain different AUC scores by different values of δ based on detection framework in Fig. 3, where $\delta(\delta \in (0,1])$ is the multiple of the original image. As shown in Fig. 9, we can obtain higher AUC scores when $\delta = 0.6$ and $\delta = 1$.

Considering the close correlation between saliency map and its multi-scale maps, we take 0.6 sizes of the original image and the original image into account, generate saliency maps based on detection framework in Fig. 3, denoted by $SM_{0.6}$ and SM_1 shown in Fig. 10(a) and (b) respectively. The final integrated saliency map SM (Fig. 10(c)) is obtained by

$$SM = \gamma SM_{0.6} + (1-\gamma)SM_1 \tag{12}$$

Where γ is a weight parameter, $\gamma \in [0,1]$. AUC scores for γ values are illustrated in Fig. 11. We set γ to be 0.45 in our experiments.

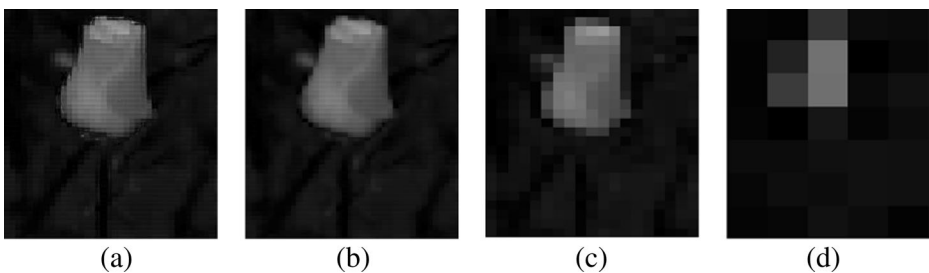


Fig. 7 Saliency maps in multi-scales. (a) 1 size of the saliency map (b) 1/4 size of the saliency map (c) 1/16 size of the saliency map (d) 1/64 size of the saliency map

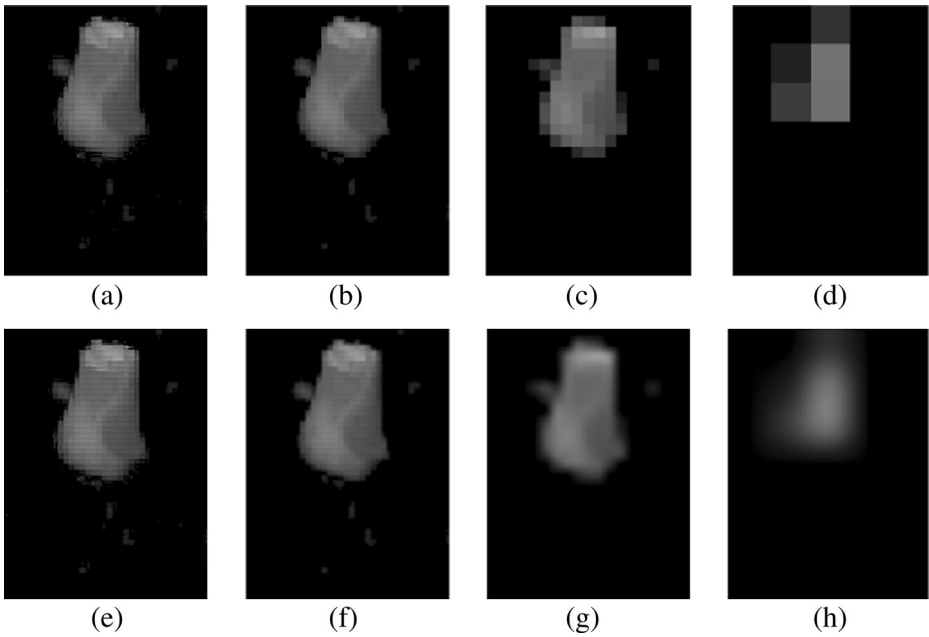


Fig. 8 Saliency maps in multi-scales and their corresponding amplified images. **(a)** 1 size of the saliency map **(b)** 1/4 size of the saliency map **(c)** 1/16 size of the saliency map **(d)** 1/64 size of the saliency map **(e)** Amplified image corresponding to 1 size of the saliency map **(f)** Amplified image corresponding to 1/4 size of the saliency map **(g)** Amplified image corresponding to 1/16 size of the saliency map **(h)** Amplified image corresponding to 1/64 size of the saliency map

4 Experiments

We present empirical evaluation and analyze of the proposed method against the 9 state-of-the-art methods on the MSRA salient object database with the labeled ground truths [1], which contains 1000 color images with accurate pixel-wise object-contour segmentations. All the programs are operated by Windows 7, AMD Athlon(tm) X2 Dual-Core QL-64 2.1 GHz and MATLAB R2011b.

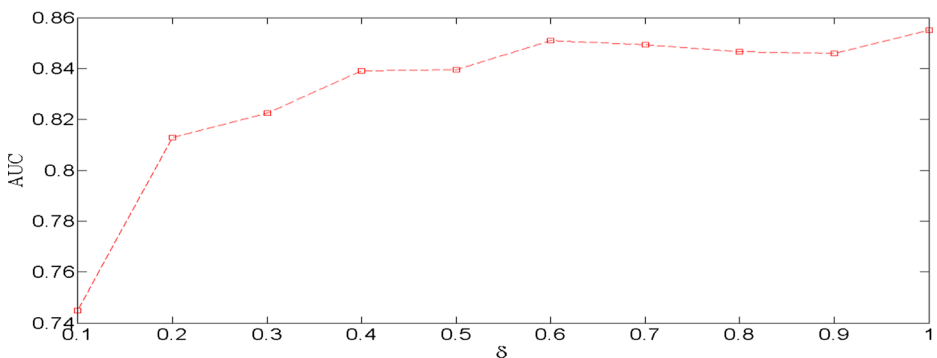


Fig. 9 AUC scores for δ values

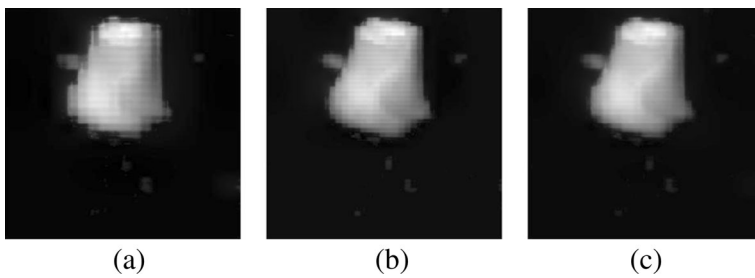


Fig. 10 Generated saliency maps. **a** Generated saliency map corresponding to 0.6 sizes of the original image **(b)** Generated saliency map corresponding to 1 size of the original image **(c)** Final integrated saliency map

4.1 Evaluation standards

For comparing the quality of different saliency maps, we utilize a widely used method, the receiver operating characteristics curve (ROC Curve) [4, 18, 25]. Meanwhile, average values of Precision, Recall, and F-Measure [1, 2, 23] are obtained to measure the performance of different saliency methods.

Given a saliency map and a labeled ground truth data, the true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) can be calculated as follows:

$$\begin{cases} TP = \sum_i^{N_{is}} f(SM_i, t)BM_i \\ FP = \sum_i^{N_{is}} f(SM_i, t)(1-BM_i) \\ TN = \sum_i^{N_{is}} f(t, SM_i)(1-BM_i) \\ FN = \sum_i^{N_{is}} f(t, SM_i)BM_i \end{cases} \quad (13)$$

Where N_{is} is the total number of pixels in SM , SM_i is the pixel in SM , t is the threshold for binarization, $t \in [0, 255]$, BM_i is the binary mask, and function $f(n_1, n_2)$ is defined as,

$$f(n_1, n_2) = \begin{cases} 1, n_1 \geq n_2 \\ 0, n_1 < n_2 \end{cases} \quad (14)$$

Correspondingly, the true positive rate (TPR) and the false positive rate (FPR) are calculated as,

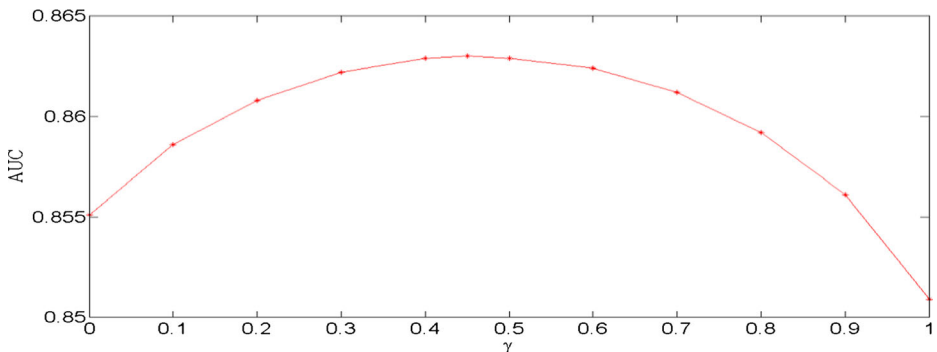


Fig. 11 AUC scores for γ values

$$\begin{cases} TPR = \frac{TP}{TP + FN} \\ FPR = \frac{FP}{FP + TN} \end{cases} \quad (15)$$

By varying the threshold t from 0 to 255, furthermore, the ROC Curve for the saliency model is plotted as the mean FPR versus mean TPR. ROC Curve is a composite indicator reflects the sensitivity (corresponding TPR) and 1-specificity (corresponding FPR) of continuous variables, and it is the most prevalent criteria for evaluating the performance of visual saliency detection methods. The area under the ROC Curve (AUC) can be used as statistical standard of quantitative experiment results. The higher AUC is, the higher the accuracy of the method is.

Analogously, the Precision and the Recall are defined as,

$$\begin{cases} Precision = \frac{TP}{TP + FP} \\ Recall = \frac{TP}{TP + FN} \end{cases} \quad (16)$$

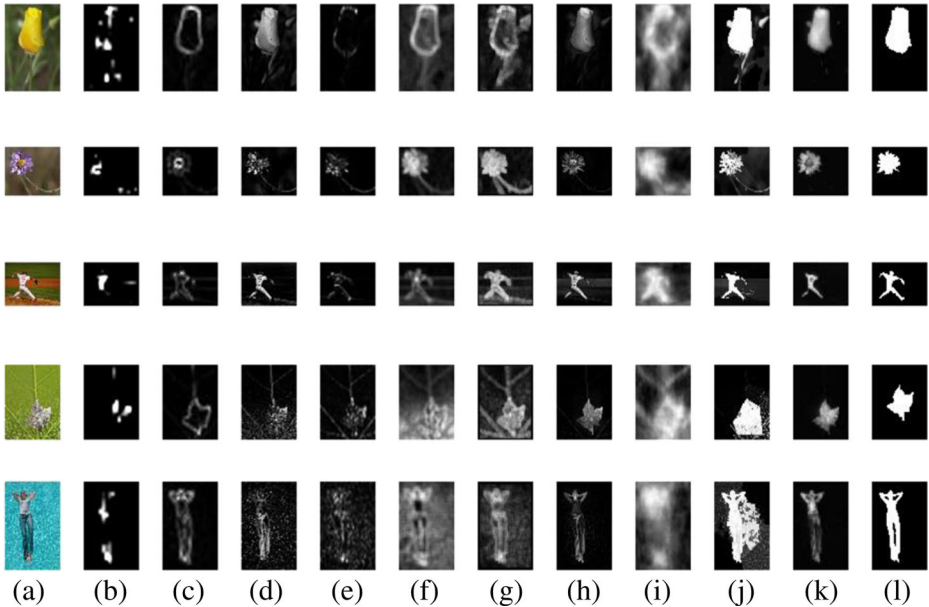


Fig. 12 Visual comparison of our saliency maps with 9 state-of-the-art methods. **a** Original images **(b)** IK [9] **(c)** MZ [15] **(d)** ZS [24] **(e)** HZ [8] **(f)** ZT [25] **(g)** BT [4] **(h)** AS [2] **(i)** MV [18] **(j)** XL [23] **(k)** Proposed **(l)** Labeled ground truths

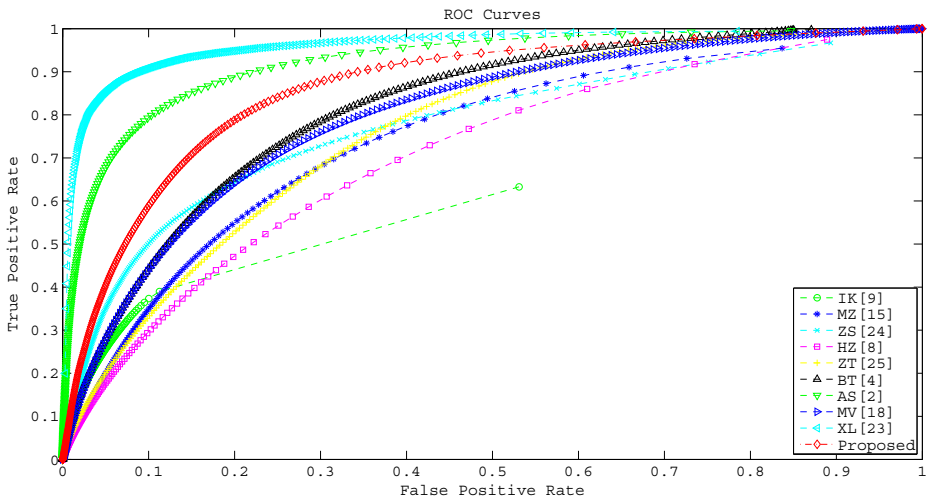


Fig. 13 ROC curves

We vary the threshold from 0 to 255 on a given saliency map with saliency values in the range [0,255], and calculate Precision and Recall at each value of the threshold, and then compute the average values of Precision and Recall. Average value of F-Measure is obtained over the same labeled ground truths.

$$F_{\beta} = \frac{(1 + \beta^2) \times P \times R}{\beta^2 \times P + R} \tag{17}$$

Where P is the average value of Precision, R is the average value of Recall. We use $\beta^2 = 0.3$ to weigh Precision more than Recall as suggested in [1, 2, 23]. The higher Precision, Recall and F-Measure are, the better the performance of the method is.

4.2 Experimental results and analyses

We evaluate the proposed method with several state-of-the-art methods: IK [9], MZ [15], ZS [24], HZ [8], ZT [25], BT [4], AS [2], MV [18] and XL [23]. Some sample results where brighter pixels indicate higher saliency probabilities are illustrated in Fig. 12. The IT [9] method is generated in low resolution and tends to highlight the boundaries and assign relatively low probabilities to pixels inside the objects, and extracts only small parts of salient objects. The MZ [15], HZ [8] and ZT [25] methods care more about local abrupt changes so they only can capture edges of objects. The ZS [24], AS [2] and MV [18] methods pay attention to the whole regions of salient objects, however, they either miss large parts of salient objects, or produce unreasonable or diffuse maps. The BT [4] and XL [23] methods are able to locate the whole salient objects, but their results involve

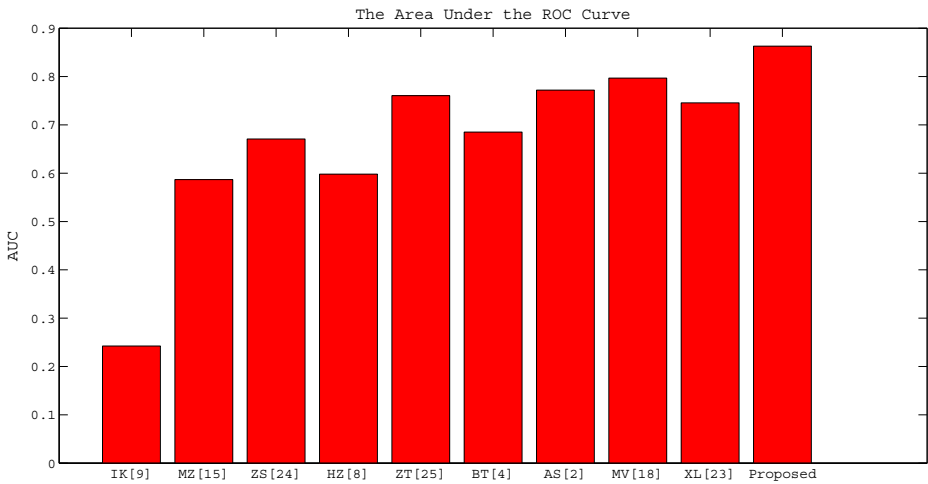


Fig. 14 AUC bars

a lot of background details. Our method not only considers global and local saliency, but also remains edges of salient objects, and we filter a lot of background details and make higher image definition. Overall, the saliency maps of our method are much more closely similar to the labeled ground truths.

The ROC Curves of various saliency methods and corresponding AUC bars are shown in Figs. 13 and 14 respectively. The maximum sensitivities, the maximum 1-specificities, and corresponding AUC scores are given in Table 1. As shown in Figs. 13, 14, and Table 1, we achieve the clearest relationship between sensitivities and 1-specificities of saliency map and the highest AUC score 0.8630, maximum 1-specificity and maximum sensitivity being to 1

Table 1 Comparison the proposed method with 9 state-of-the-art methods

Method	Maximum sensitivity	Maximum 1-specificity	AUC score	Precision	Recall	F-Measure
IK [9]	0.6630	0.5309	0.2423	0.5922	0.1715	0.3781
MZ [15]	0.9546	0.8369	0.5867	0.5123	0.2006	0.3771
ZS [24]	0.9671	0.8934	0.6708	0.5804	0.3626	0.5098
HZ [8]	0.9747	0.8892	0.5980	0.5211	0.1096	0.2792
ZT [25]	0.9999	0.9985	0.7605	0.4554	0.4400	0.4518
BT [4]	0.9975	0.8709	0.6851	0.4712	0.5036	0.4783
AS [2]	0.9953	0.8472	0.7720	0.8035	0.3518	0.6198
MV [18]	1	0.9994	0.7968	0.3507	0.6591	0.3932
XL [23]	0.9942	0.7844	0.7456	0.7204	0.8647	0.7493
Proposed	1	1	0.8630	0.8030	0.7145	0.7807

The values in bold shows the result of our proposed method

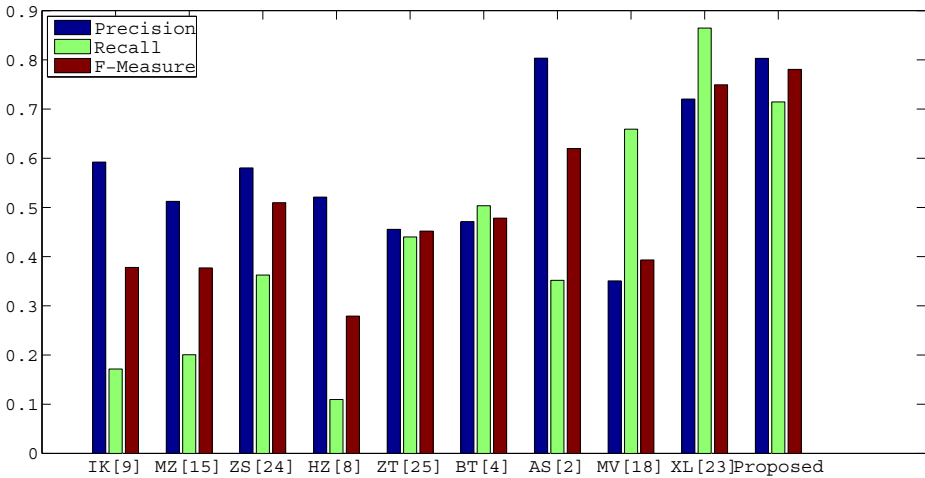


Fig. 15 Mean Precision, Recall and F-Measure values of the evaluation standards

respectively. The proposed method performs better than the other 9 state-of-the-art methods, which indicates our saliency maps are more precision and effective to the salient regions. Although the sensitivities of the AS [2] and XL [23] methods in low 1-specificities are higher than the sensitivity of the proposed method, their maximum 1-specificities are 0.8472 and 0.7844 respectively, obviously, lower than the proposed method. The maximum sensitivity and 1-specificity of the ZT [25] and MV [18] methods are approaching the results of the proposed method, however, their ROC Curve curvature are obviously lower than the curvature of the proposed method. For these reasons, their AUC scores are lower than the AUC scores of the proposed method.

The Precision, Recall and F-Measure of a saliency map are averaged over 1000 images, and the results are shown in Fig. 15 and Table 1. AS [2] shows a high Precision but very poor Recall, indicating that it is better suited for gaze tracking, but perhaps not well suited for salient regions segmentation. XL [23] shows a high Recall but low Precision. Among all the methods, the proposed method achieves one of the best Precision with higher recall and the best F-Measure values. Overall, the proposed method not only enhances the definition of salient regions, but also improves the accuracy of visual saliency detection.

5 Summary

We propose an effective visual saliency detection method based on multi-scale and multi-channel mean. This method neither centers on human eye fixation prediction nor centers on certain task of salient object detection, nor requires segmenting image. We analyze frequency signals and color channels, detect salient in multi-scales. Based on MSRA image database and several evaluation standards, we demonstrate that the proposed method outperforms 9 state-of-the-art saliency methods. However, the accuracy of salient detection for complex textured

background is not very high. Future work may be beneficial to incorporate high level factors like symmetry and semantic into saliency maps, while try to find out more effective physiological, psychological and computer vision models for salient detection.

Acknowledgments This work has been supported by the foundation of Chunhui Program from the Ministry of Education of China (GrantNo.z2011149).

References

1. Achanta R, Hemami S, Estrada F, Süsstrunk S (2009) Frequency-tuned salient region detection. *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, pp 1597–1604
2. Achanta R, Süsstrunk S (2010) Saliency detection using maximum symmetric surround. *IEEE International Conference on Image Processing*, Hong Kong, pp 2653–2656
3. Bhatnagar G, Wu QMJ (2011) An image fusion framework based on human visual system in framelet domain. *Int J Wavelets Multiresolution Inf Process* 10(1):1–30
4. Bruce ND, Tsotsos JK (2009) Saliency, attention, and visual search: an information theoretic approach. *J Vis* 9(3):1–24
5. Chang KY, Liu TL, Lai SH (2011) From co-saliency to co-segmentation: an efficient and fully unsupervised energy minimization model. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Providence, pp 2129–2136
6. Ding ZH, Yu Y, Wang B, Zhang LM (2012) An approach for visual attention based on biquaternion and its application for ship detection in multispectral imagery. *Neurocomputing* 76(1):9–17
7. Furuya T, Ohbuchi R (2014) Visual saliency weighting and cross-domain manifold ranking for sketch-based image retrieval. *Proc. Multi-Media Modeling*, Dublin, pp 37–49
8. Hou XD, Zhang LQ (2007) Saliency detection: a spectral residual approach. *IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, pp 1–8
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. Kekre HB, Thepade SD, Chaturvedi RN (2013) Block based information hiding using Cosine, Hartley, Walsh and Haar Wavelets. *Int J Adv Comput Res* 3(1):1–6
11. Keys RG (1981) Cubic convolution interpolation for digital image processing. *IEEE Trans Acoust, Speech, Signal Process* 29(6):1153–1160
12. Koch C, Ullman S (1985) Shifts in selective visual attention: towards the underlying neural circuitry. *Hum Neurobiol* 4:219–227
13. Lin YW, Tang YY, Fang B et al (2013) A visual-attention model using earth mover’s distance-based saliency measurement and nonlinear feature combination. *IEEE Trans Pattern Anal Mach Intell* 35(2):314–328
14. Liu T, Yuan ZJ, Sun J, Wang JD, Zhang NN (2011) Learning to detect a salient object. *IEEE Trans Pattern Anal Mach Intell* 33(2):353–367
15. Ma YF, Zhang HJ (2003) Contrast-based image attention analysis by using fuzzy growing. *ACM International Conference on Multimedia*, New York, pp 374–381
16. Mallat SG (1989) Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Trans Am Math Soc* 315(1):69–87
17. Mallat SG (1989) A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans Pattern Anal Mach Intell* 11(7):674–693
18. Murray N, Vanrell M, Otazu X, Parraga CA (2011) Saliency estimation using a non-parametric low-level vision model. *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, pp 433–440
19. Qin CC, Zhang GP, Zhou YC, Tao WB, Cao ZG (2014) Integration of the saliency-based seed extraction and random walks for image segmentation. *Neurocomputing* 129(7):378–391
20. Ray SS (2012) On Haar wavelet operational matrix of general order and its application for the numerical solution of fractional Bagley Torvik equation. *Appl Math Comput* 218(9):5239–5248
21. Wang R, Yu ZX, Du LF, Lee TY (2013) Saliency-based adaptive block compressive sampling for image signals. *J Image Graph* 18(10):1255–1260
22. Wang Q, Yuan Y, Yan PK, Li XL (2013) Saliency detection by multiple-instance learning. *IEEE Trans Cybern* 43(2):660–672

23. Xie YL, Lu HC, Yang MH (2013) Bayesian saliency via low and mid level cues. *IEEE Trans Image Process* 22(5):1689–1698
24. Zhai Y, Shah M (2006) Visual attention detection in video sequences using spatiotemporal cues. *ACM International Conference on Multimedia*, New York, pp 815–824
25. Zhang LY, Tong MH, Marks TK, Shan HH, Cottrell GW (2008) SUN: a Bayesian framework for saliency using natural statistics. *J Vis* 8(7):1–20
26. Zhao GP, Yin MF, Chen Y (2013) Image salient region detection based on histogram. Xi'an CN: Proceedings of the 32nd Chinese Control Conference, pp 3570–3574
27. Zhu XQ, Huang JC, Shao ZF, Cheng GQ (2013) A new approach for interesting local saliency features definition and its application to remote sensing imagery retrieval. *Geomatics Inform Sci Wuhan Univ* 38(6):652–655



Lang Sun received the B.S. degree in information and computing sciences from Xi'an University of Science and Technology, Xi'an, China in July 2012. He is currently working towards his M.S. degree in computer software and theory at Southwest University, China. His current research interests include image processing, pattern recognition.



Yan Tang received the B.S. degree in department of computer science from Southwest University, Chongqing, China in July 1985, the M.S. degree in department of computer science from Southwest University, Chongqing, China in July 1991. She is the professor at the school of computer and information science, Southwest University, Chongqing, China. She has published more than 80 technical papers. Her research interests include artificial intelligence, image processing and pattern recognition.



Hong Zhang received the B.S. degree in electronic engineering from Northeastern University, Boston, USA in 1982, the Ph.D. degree in electronic engineering from Purdue University, Lafayette, USA in 1986. He is the tenured professor at the department of computing science, Alberta University, Edmonton, Canada. He has published more than 100 technical papers. His research interests include span robotics, computer vision, image processing.