IET Image Processing

*Research Article*

IET Journals
The Institution of Engineering and Technology

# Salient region detection using feature extraction in the non-subsampled contourlet domain

ISSN 1751-9659
Received on 10th April 2017
Revised 11th May 2018
Accepted on 28th August 2018
E-First on 15th October 2018
doi: 10.1049/iet-ipr.2018.5479
www.ietdl.org

*Masoumeh Rezaei Abkenar[1], Hamidreza Sadreazami[1], M. Omair Ahmad[1]* ✉

[1]*Department of Electrical and Computer Engineering, Concordia University, Montreal, Quebec, Canada H3G 1M8*

✉ *E-mail: omair@ece.concordia.ca*

**Abstract:** The human visual system is attracted to the most dominant part of the image which is called salient region. There has been a surge of interest in the past few years to efficiently detect the salient regions of images. In this study, a new salient region detection method is proposed using the non-subsampled contourlet transform. It is known that this transform is capable of providing a multiscale, multi-directional and translation invariant decomposition of images. The proposed saliency detection method is realised by extracting various local and global features from the non-subsampled contourlet coefficients of the colour channels. A saliency map is obtained based on a linear combination of the local features and the distribution of the global features. In order to provide a better preservation of the structure and boundary of the objects and to obtain a more uniformly highlighted salient region, the saliency map is abstracted using an optimisation framework. Several experiments are conducted on sets of natural images to evaluate the performance of the proposed method. The results show that the performance of the proposed method is superior to that of the other existing methods in terms of precision-recall performance, *F*-measure, and mean absolute error values.

## 1 Introduction

It is known that the human visual system (HVS) is not capable of analysing all the visual information due to the enormous amount of received information. Thus, it is well accepted that the HVS has a saliency detection property to rapidly detect the most important part of a scene. In other words, the salient region of each image is detected and processed, while non-salient regions are ignored. In view of this, saliency detection has become an important research area in the past decade and plays a significant role in various image processing applications such as object of interest image segmentation [1], adaptive image and video compression [2], object-based image retrieval [3], image retargeting [4, 5], and medical imaging [6], to name a few. A comprehensive review of the works in saliency detection can be found in [7–9]. It should be mentioned that since 2015, a few supervised or unsupervised salient object detection methods have been proposed by utilising deep convolutional neural networks [10–19], autoencoders [20], and recurrent neural networks [21–23]. In general, a training-based method has the potential to provide accurate results. However, due to a tremendous amount of training required, our focus in this study is on approaches that do not require training.

In the past few years, there has been a surge of interest in developing saliency detection techniques both in the spatial [24–30] and frequency domains [31–38]. In order to detect the entire salient region uniformly, it is essential to retain an adequate range of frequency components. In most of the spatial-domain methods, the entire salient region may not be uniformly highlighted, since the high-frequency components of the image are indiscriminately discarded. On the other hand, the frequency-domain methods have been developed to address this limitation. In [32], the quaternion Fourier transform has been applied to the colour and intensity features of the image. In [31], a wavelet-based salient point detector for image retrieval has been proposed in which the global variation of the pixels are measured by using the absolute wavelet coefficients at the courser scales, and the pixels with higher global variations are traced back in the finer scales. In [33], the wavelet transform has been applied to the image colour channels and after each decomposition level, a feature map is generated by setting the low-pass coefficients to zero and applying the inverse transform. Then, a local saliency map is computed by linearly combining the feature maps of the colour channels at each level, while a global saliency map is computed based on the likelihood of the features. However, in a saliency detection method that uses the wavelet transform, the wavelets are optimal only in representing point discontinuities, but not effective in capturing line discontinuities, which correspond to directional information in the image [39]. In order to circumvent the lack of directional selectivity of the wavelet, other multi-scale and multi-resolution representations, such as the non-subsampled contourlet transform (NSCT) [40], have been developed. This transform is not only able to capture the smooth contours effectively in a flexible number of directions but is also invariant under translations. Despite its superior capability in representing the directional information of images, only a few attempts have been made to address the problem of saliency detection in the non-subsampled contourlet domain. For instance, a saliency detection method has been proposed in [36] by fusing local and global information. However, it has not fully taken into account the effective visual features, such as texture and structure, since these attributes provide significant information towards the local and global characteristics of the image and are known to be of great importance in saliency detection.

In view of this and the fact that developing accurate methods to obtain saliency maps is an on-going research area, in this study, a new salient region detection method is proposed by extracting local and global features from the non-subsampled contourlet coefficients of the three colour channels of *CIELAB* colour space at different scales. The ability to do a localised multi-resolution spatial and frequency analysis in the non-subsampled contourlet domain makes it a superior tool to extract image details at different scales. The local features are extracted from the local variations of the low-pass coefficients whereas the global features are obtained based on the distribution of the directional subband coefficients. To extract the image visual features effectively, the saliency map thus obtained is finally abstracted into meaningful regions by applying an optimisation framework. Simulations are carried out to evaluate the effectiveness of the proposed saliency method and to compare it with other existing works.

The paper is organised as follows. Section 2 presents a brief introduction of NSCT. Section 3 presents the proposed saliency detection method in the NSCT domain. Simulation results are provided in Section 4. Finally, Section 5 concludes the paper.


© The Institution of Engineering and Technology 2018

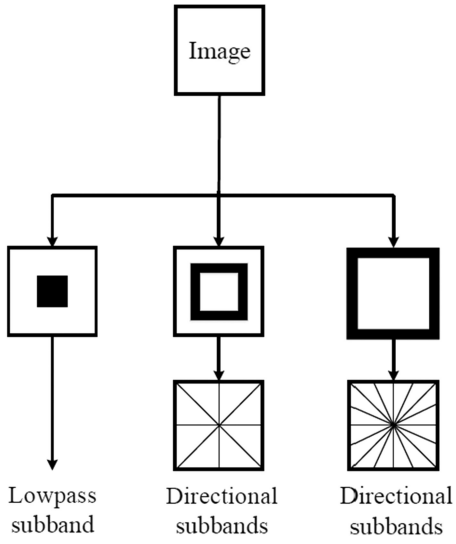


**Fig. 1** *Filter bank structure of the NSCT*

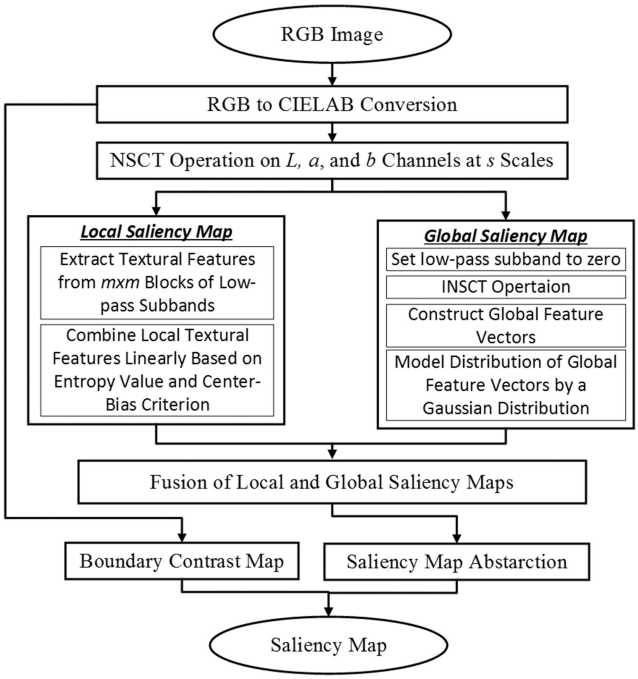


**Fig. 2** *Block diagram of the proposed saliency detection method*

## 2 Non-subsampled contourlet transform

The NSCT has been proposed as a multi-scale and multi-directional decomposition framework for images [40]. NSCT is composed of two filtering stages including non-subsampled pyramid (NSP) and non-subsampled directional filter bank (NSDFB). The multi-scale property is realised by using a two-channel non-subsampled filter bank resulting in low- and high-frequency images at each NSP decomposition level. Subsequently, to capture the singularities in the image, NSP iteratively decomposes the low-frequency component. As a result, the NSP provides $J+1$ sub-images comprising one low- and $J$ high-frequency image components having the same size as that of the original image, where $J$ denotes the number of decomposition levels. Finally, multi-directional decomposition is achieved by applying NSDFB to high-frequency images at each scale. NSDFB is constructed by combining the directional fan filter banks and then used to produce flexible directional sub-images with the same size as that of the original image. Fig. 1 shows a schematic of the NSCT decomposition process. In the following section, we introduce the proposed saliency detection method in the NSCT domain.

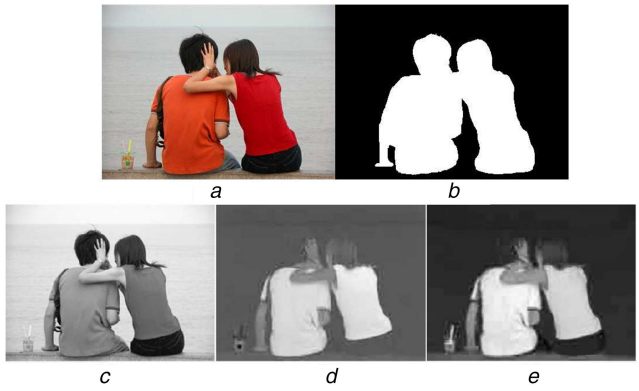


**Fig. 3** *A sample image and the corresponding color channels and ground* ***(a)*** Sample Image, ***(b)*** Ground truth, ***(c)*** Colour channel *L*, ***(d)*** Colour channel *a*, ***(e)*** Colour channel *b*

## 3 Proposed non-subsampled contourlet-based saliency detection method

In this section, the various steps of the proposed saliency detection method are presented. Fig. 2 shows the overall framework of the proposed saliency detection method.

The proposed method involves generating local and global feature maps. The local feature maps are generated in order to detect pixels that are salient in a limited neighbourhood, while the global feature maps detect pixels that are salient in the entire image. To this end, the input image is first converted to *CIELAB* colour space, having luminance, red/green, and blue/yellow channels, denoted by $L$, $a$, and $b$, respectively. This colour space is known to be perceptually more uniform than the *red, green, blue* colour space and also more similar to the HVS perception. A sample image and its $L, a, b$ channels and the corresponding ground truth are shown in Fig. 3. The NSCT decomposes each colour channel into a number of levels, $s = 1, \ldots, S$, in order to extract the local and global features from the non-subsampled contourlet coefficients. At each decomposition level, each channel is decomposed into a low-pass subband and a number of directional band-pass subbands as

$$\{\boldsymbol{A}_s^c, \boldsymbol{D}_{s,d}^c\} = \mathrm{NCST}_s(\boldsymbol{I}^c), \tag{1}$$

where $\boldsymbol{I}^c$, $c \in \{L, a, b\}$, is the matrix consisting of the pixel values of the channel $c$ component of the image, $\boldsymbol{A}_s^c$ denotes the low-pass subband, $\boldsymbol{D}_{s,d}^c$ denotes the directional subbands corresponding to the level $s$ and the direction $d$ of the channel $c$. By taking advantage of such multi-scale and directional decomposition, in the proposed method, two feature maps are generated, namely, a local feature map based on the statistical parameters of the low-pass coefficients and a global feature map by utilising the directional subband coefficients.

### 3.1 Local saliency map using textural features

In this subsection, the local saliency map generation method used in this study is described. Our approach to this end is first applying $s$ level decomposition to each colour channel. Then, the low-pass subband is divided into blocks of $m \times m$ pixels after each decomposition level, and the local variance of each block is considered as the local feature map value of the block as given by

$$\mathrm{Lmap}_s^c(x, y) = \mathrm{Var}\{A_s^c(x+i, y+j)\}_{\{i,j = \pm 3, \pm 2, \pm 1, 0\}}. \tag{2}$$

Fig. 4*a* shows the local feature maps for the colour channels corresponding to the image in Fig. 3 after the $s$th decomposition level. It is seen from this figure that the local feature maps represent various textural details of the image at different levels.


© The Institution of Engineering and Technology 2018

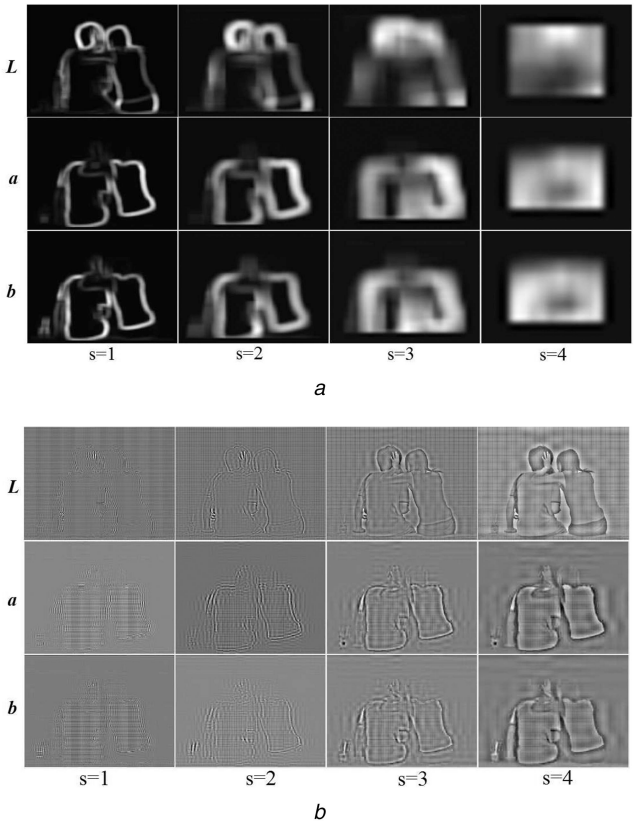


**Fig. 4** *Feature maps obtained after the first, second, third, and forth levels of NSCT decompositions for the L, a and b colour channels*
***(a)*** Local feature maps, ***(b)*** Global feature maps

The local saliency map $\boldsymbol{S}_{\text{Local}}$ is then obtained by linearly combining the local feature maps $\mathbf{Lmap}_s^c$ based on the following two criteria: low entropy and border avoidance, as given by

$$\boldsymbol{S}_{\text{Local}} = \sum_{c,\,s} w_s^c \mathbf{Lmap}_s^c, \tag{3}$$

where $w_s^c$ denotes the weights. The weights are assigned according to the entropy values of the local feature map and pixel intensity values around the centre of the map, in such a way that a larger weight is assigned to a local feature map having a small value of entropy and large values around the centre of the map. An optimum feature map is composed of two distinctive parts: the salient region with large intensity values, and the non-salient region with low intensity values; thus, it has a small entropy value. In addition, in saliency detection, it is desirable to find a map with a localised salient region. Therefore, the spatial information must be considered in computing the entropy value. To this end, the local feature map is first filtered by a Gaussian kernel $\boldsymbol{G}$ [32] and its entropy is computed. It is to be noted that in most of the natural images the salient region is located close to the image centre rather than its borders. In view of this, in this work, a border-avoidance criterion [32] is utilised to represent the strength of the local feature map around the centre $\lambda_s^c$, as given by

$$\lambda_s^c = \sum_{x,\,y} \boldsymbol{K} N(\mathbf{Lmap}_s^c), \tag{4}$$

where $\boldsymbol{K}$ is a 2D centred Gaussian mask with its entries having the maximum value of 1 and its size the same as that of the channel feature map, and $N(.)$ is a normalisation function so that the summation of all the pixel values of $\mathbf{Lmap}_s^c$ is 1. The weights in (4) are calculated as

$$w_s^c = \frac{\lambda_s^c}{H(\mathbf{Lmap}_s^c * \boldsymbol{G})}, \tag{5}$$

where $H(.)$ denotes the entropy value of the smoothened feature map.

### *3.2 Global saliency map using NSCT directional subband coefficients*

In this subsection, the global saliency map generation method used in this study is described. In order to generate the global feature map, after each decomposition level, we set the low-pass coefficients to zero (6) for each colour channel, and apply the inverse NSCT to the bandpass coefficients as

$$\mathbf{Gmap}_s^c = \text{INSCT}_s(\boldsymbol{A}_s^c = \mathbf{0}, \boldsymbol{D}_{s,d}^c). \tag{6}$$

The resulting inverse transform is the global feature maps. The global feature maps corresponding to the different levels of decomposition and each channel for the image in Fig. 3 are shown in Fig. 4*b*.

The elements of the global feature maps are then used to form global feature vectors, each corresponding to one pixel of the image. For each pixel, a global feature vector $\boldsymbol{g}(x, y)$ with a length of $3S$ ($S$ being the number of the global feature maps for each channel) is constructed. If a particular vector is less similar to the others, the corresponding pixel is different from the other pixels and thus it can be considered to be more salient. In view of this, the distribution of the global feature vectors are modelled by a Gaussian distribution [41] and the global saliency of each pixel is defined as the likelihood of finding its global feature vector amongst all the vectors, as (see (7)) where $\mu$ is a vector containing the means of the global feature maps, $\boldsymbol{\Sigma}$ is a $n \times n$ covariance matrix, T is the transpose operator, and $|.|$ denotes the determinant of the covariance matrix. Using (7), the global saliency map $\boldsymbol{S}_{\text{Global}}$ is computed as

$$S_{\text{Global}}(x, y) = \log\left(p(\boldsymbol{g}(x, y))^{-1}\right)^{0.5}. \tag{8}$$

The local and global saliency maps obtained for the sample image in Fig. 3 are shown in Figs. 5*a* and *b*. It is seen from this figure that


© The Institution of Engineering and Technology 2018

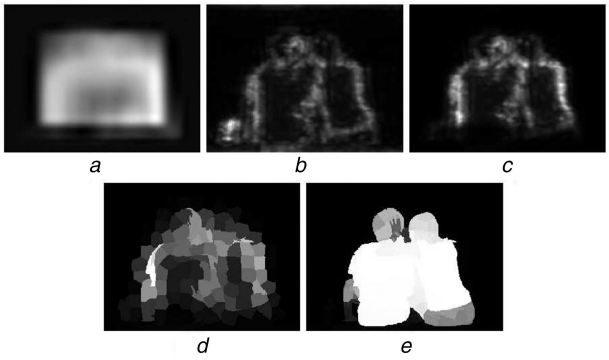


**Fig. 5** *Maps obtained after applying the different steps of the proposed method*

***(a)*** Local map, ***(b)*** Global saliency maps, ***(c)*** Map obtained by fusion of local and global saliency maps, ***(d)*** Saliency map averaged over superpixels, ***(e)*** Saliency map after abstraction

the global saliency map provides some meaningful information, especially around the edges of the salient object, which cannot be detected well only by using the local saliency map. This is due to the fact that the global saliency map consists of the statistical relation among the global feature maps.

An image region is salient if it is different from its surrounding regions and also it stands out in the entire image. In view of this, after computing the local and global saliency maps, the pixels which are both locally and globally salient are considered as the final salient pixels. Thus, the local and global saliency maps are merged using a fusion algorithm that involves a Hadamard product followed by a process that normalises the pixels in this last map in the range [0,1]. The propose of the fusion algorithm is to give prominence to the pixels that have large values both in the local and global saliency maps, whereas to assign small values to pixels that have small values in either of the local and global maps or in both. The fused map $\boldsymbol{S}_M$ is given by

$$\boldsymbol{S}_M = Nr(\boldsymbol{S}_{\text{Local}} \circ \boldsymbol{S}_{\text{Global}}), \tag{9}$$

where $Nr(.)$ represents the normalisation process of the pixels of the associated map in the range [0,1] and $\circ$ denotes the Hadamard matrix product. The fused map obtained is shown in Fig. 5*c*.

### *3.3 Image saliency map abstraction*

It is known that a salient object detection technique aims at detecting the most distinctive regions or objects, rather than individual pixels [42]. In view of this, the saliency map is further abstracted into perceptually meaningful regions. More specifically, the input image is segmented into $k$ superpixels using simple linear iterative clustering (SLIC) algorithm [43], and the saliency value of each superpixel, is replaced by the average of the saliency values of all the pixels belonging to that superpixel. Fig. 5*d* shows the saliency map averaged over superpixel for the sample image in Fig. 3. The saliency map abstraction is implemented using the optimisation technique of [26] in which the saliency values of superpixels are used as the foreground weights. In this optimisation problem, the cost function given by

$$\sum_{i=1}^{k} wf_i S_i^2 + \sum_{i=1}^{k} wb_i (S_i - 1)^2 + \sum_{i,j} ws_{ij} (S_i - S_j)^2, \tag{10}$$

where $S_i$ is the optimised saliency value of the superpixel $i$, and $wf_i$, $wb_i$, and $ws_{ij}$ denote the foreground, background and smoothness weights, respectively, are minimised. The cost function contains three terms corresponding to the pixels in the foreground and background, and the smoothness of the pixels within the two regions. The foreground term tries to assign the saliency value of 1 to the superpixels with a larger value in the saliency map. On the other hand, the background term aims at assigning a saliency value of 0 to the superpixels with a strong boundary connectivity. Boundary connectivity is measured in terms of the spatial distance and the CIELAB colour distance between a superpixel and the superpixels located around the image boundary. The smoothness term tries to assign the same saliency values to the superpixels within the foreground region or the background region that are spatially close to each other.

Since in most of the images, image boundary pixels belong to the background, the degree of contrast of a pixel from the boundary pixels, provides a measure of its belonging to the salient region. Therefore, we generate a boundary contrast map by computing each pixel's distance to the mean colour and covariance matrix of the boundary pixels. The abstracted saliency map and the boundary contrast map are then pixel-wise added.

The saliency maps are finally refined employing the post-processing algorithm used in [29]. First, to smooth the saliency map while keeping the details of object boundaries, a morphological smoothing step composed of a reconstruction-by-dilation operation followed by a reconstruction-by-erosion [44] is employed. The contrast between the salient and non-salient regions is then enhanced using a sigmoid function.

Fig. 5*e* shows the saliency map after the abstraction for the sample image in Fig. 3. It is seen from this figure that the salient region is detected more precisely and more uniformly after the abstraction.

## 4 Experimental results

In order to evaluate the performance of the proposed saliency detection method, several experiments are conducted on five datasets of images, namely, MSRA-1000 [45] (1000 images), MSRA-10K [28] (10,000 images), HKU-IS [13] (4447 images), PASCAL-S [46] (850 images), and DUT-OMRON [25] (5172 images) datasets. The image is decomposed into four scales and four directional subbands in each scale by using NSCT. The block size in local feature extraction is set to 7.

To compare the results with the ground truth, the grey level saliency maps $\boldsymbol{S}$ are converted to binary maps $\boldsymbol{S}_b$ using two different thresholds, fixed and adaptive. In case of a fixed threshold, all possible thresholds in the range [0, 255] are applied to the saliency maps and different binary maps are generated and compared to the ground truth. Next, the adaptive image-dependent threshold is applied to the grey level saliency maps to obtain the binary maps. This threshold is set as twice the mean of the saliency values of the saliency map as in [45]

$$T_{adp} = \frac{2}{L \times W} \sum_{x} \sum_{y} S(x, y), \tag{11}$$

where $L$ and $W$ are the number of rows and columns in the saliency map.

There exist different evaluation metrics for quantitative comparison. In this work, since we use datasets with accurate object masks, we need to measure whether the salient objects can be detected as a whole. Saliency detection can be regarded as the binary classification of salient and non-salient regions. In such a problem, precision (also known as positive predictive value) is the fraction of the retrieved instances that are relevant, while recall (also known as sensitivity) is the fraction of relevant instances that are retrieved. The performance of the proposed method is evaluated based on the metrics precision, recall, and $F$-measure as defined as

$$p(\boldsymbol{g}(x, y)) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} \exp\left\{\frac{-1}{2} (\boldsymbol{g}(x, y) - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{g}(x, y) - \boldsymbol{\mu})\right\}, \tag{7}$$


© The Institution of Engineering and Technology 2018

$$P = \frac{\sum_x \sum_y S_b(x,y) GT(x,y)}{\sum_x \sum_y S_b(x,y)},$$
$$R = \frac{\sum_x \sum_y S_b(x,y) GT(x,y)}{\sum_x \sum_y GT(x,y)}, \quad (12)$$
$$F = \frac{(1+\xi)PR}{\xi P + R},$$

where $S_b(x,y)$ is the saliency map, $GT(x,y)$ is the binary ground truth and $\xi = 0.3$ is a parameter to specify the relative importance of $P$ and $R$.

In order to further evaluate the performance of the proposed method the mean absolute error (MAE) between the binary saliency maps using the adaptive threshold and the ground truth is obtained as

$$\text{MAE} = \frac{1}{L \times W} \sum_x \sum_y |S_b(x,y) - GT(x,y)|. \quad (13)$$

The performance of the proposed method is compared to six existing schemes that have been originally proposed for salient object detection, namely, saliency filters (SF) [24], manifold ranking (MR) [25], saliency optimisation (SO) [26], region-based contrast (RC) [28], minimum barrier distance (MBD+) [27], and minimum spanning tree (MST) [29]. The saliency maps obtained using the proposed method as well as that of the other methods for two test images from each of the datasets are shown in Fig. 6. It can be seen from this figure that the saliency maps obtained by the proposed method are more similar to the ground truth as compared to that provided by the other methods.

Figs. 7–11 depict the average precision-recall curves obtained by applying the proposed saliency detection method and the other methods to the images in the MSRA-1000, MSRA-10K, HKU-IS, PASCAL-S, and DUT-OMRON datasets when the threshold is fixed. It is seen from this figure that the proposed method generally outperforms all the other methods in terms of precision-recall performance when applied to the images in MSRA, MSRA10 K, DUT-OMRON, and HKU-IS datasets. For images in the PASCAL-S dataset, the MBD+ method provides generally a better performance, while the proposed method still outperforms the other methods except for MBD+.

The average precision, recall, and $F$-measure values obtained using the proposed method with the adaptive threshold as well as

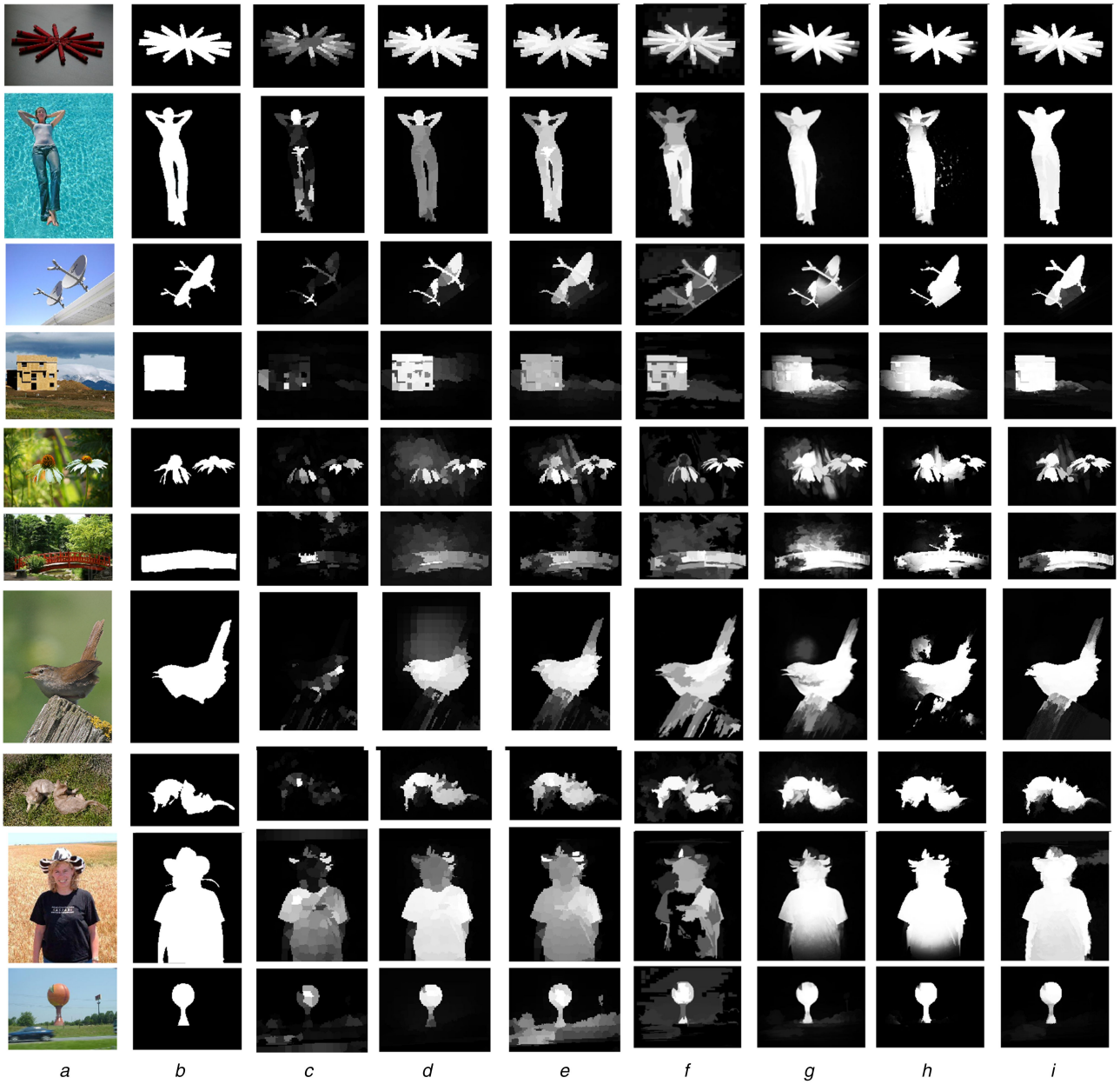


**Fig. 6** *Saliency maps obtained by applying the proposed method and the other methods on two images from MSRA-1000, MSRA-10K, HKU-IS, PASCAL-S, and DUT-OMRON datasets*

***(a)*** Original image, ***(b)*** Ground truth, ***(c)*** SF [24], ***(d)*** MR [25], ***(e)*** SO [26], ***(f)*** RC [28], ***(g)*** MBD+ [27], ***(h)*** MST [29], ***(i)*** Proposed method


© The Institution of Engineering and Technology 2018

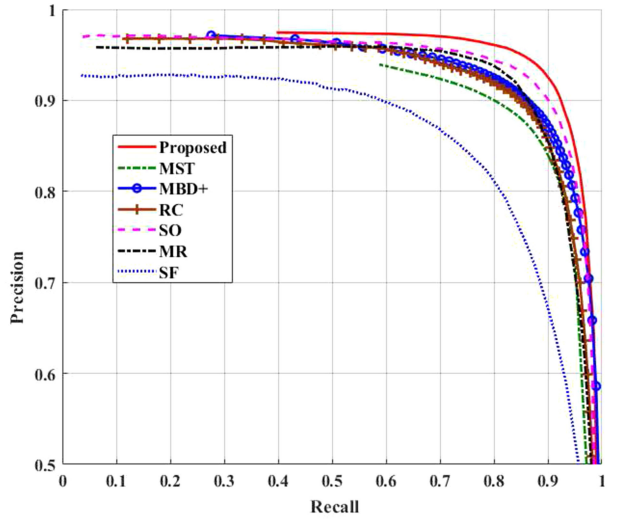


**Fig. 7** *Precision-recall curves obtained by applying the proposed saliency detection method and the other methods, when the threshold is fixed for images in the MSRA-1000 dataset*

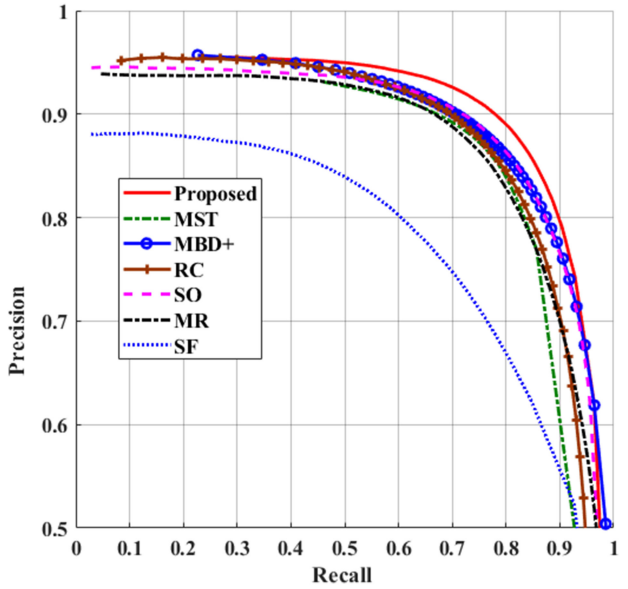


**Fig. 8** *Precision-recall curves obtained by applying the proposed saliency detection method and the other methods, when the threshold is fixed for images in the MSRA-10K dataset*

that of the other methods are calculated and given in Tables 1–5. It is seen from these tables that the proposed method provides the largest values of precision and *F*-measure for all the datasets except for PASCAL-S, where its *F*-measure value is lower by 1.24% in comparison with that of the best value provided by MBD+. The recall values provided by the proposed method are the largest for MSRA-1000 and MSRA-10K datasets and competitive to the largest values for the other three datasets.

The MAE values obtained using the proposed method as well as that of the other existing methods are also presented in Tables 1–5. It is seen from these tables that the proposed method yields the smallest MAE values when applied to images in the MSRA, MSRA-10K, HKU-IS, and DUT-OMRON datasets. For images in the PASCAL-S dataset, MBD+ is the leading method while the proposed method provides MAE value that is 0.31% larger than that provided by MBD+ for this dataset. Thus, overall the proposed scheme provides the best performance and the lowest MAE values for most of the datasets on which the various schemes have experimented.

The proposed method is implemented in MATLAB on an Intel Core i7 3.4 GHz personal computer with 16 GB RAM. We have obtained the average processing time of the proposed method to be 3.9, 4.1, 4.0, 3.8s, and 3.8 s per image for images in the

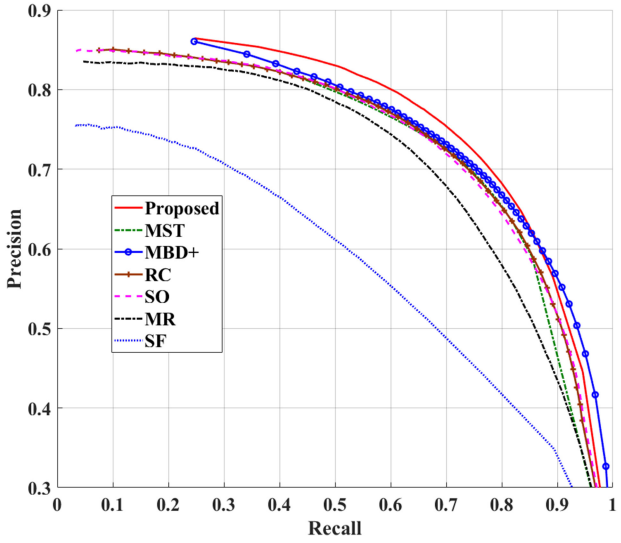


**Fig. 9** *Precision-recall curves obtained by applying the proposed saliency detection method and the other methods, when the threshold is fixed for images in the HKU-IS dataset*

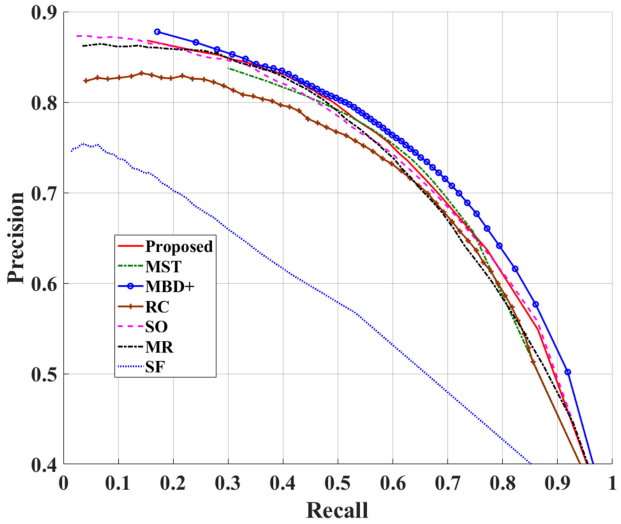


**Fig. 10** *Precision-recall curves obtained by applying the proposed saliency detection method and the other methods, when the threshold is fixed for images in the PASCAL-S dataset*

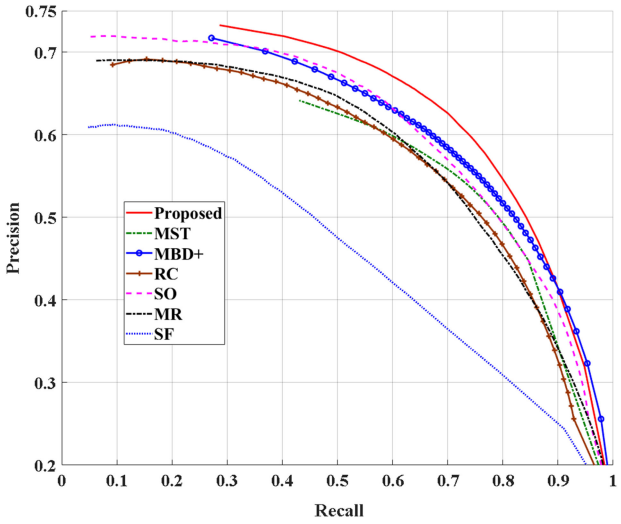


**Fig. 11** *Precision-recall curves obtained by applying the proposed saliency detection method and the other methods, when the threshold is fixed for images in the DUT-OMRON dataset*




© The Institution of Engineering and Technology 2018

**Table 1** Precision, recall, *F*-measure and MAE values obtained by applying the proposed method and the other methods on images in MSRA-1000 dataset

| Method | Precision | Recall | *F*-measure | MAE |
|---|---|---|---|---|
| proposed | **0.9219** | **0.9059** | **0.9182** | **0.0332** |
| MST | 0.8671 | 0.8639 | 0.8664 | 0.0562 |
| MBD+ | 0.8732 | 0.8808 | 0.8750 | 0.0517 |
| RC | 0.8740 | 0.8668 | 0.8723 | 0.0517 |
| SO | 0.8900 | 0.9031 | 0.8930 | 0.0415 |
| MR | 0.9141 | 0.8519 | 0.8990 | 0.0450 |
| SF | 0.8667 | 0.7141 | 0.8260 | 0.0822 |

The numerical values indicating the best performance are shown in bold.

**Table 2** Precision, recall, *F*-measure and MAE values obtained by applying the proposed method and the other methods on images in MSRA-10K dataset

| Method | Precision | Recall | *F*-measure | MAE |
|---|---|---|---|---|
| proposed | **0.9005** | **0.7921** | **0.8729** | **0.0718** |
| MST | 0.8798 | 0.7299 | 0.8400 | 0.0893 |
| MBD+ | 0.8630 | 0.7885 | 0.8446 | 0.0828 |
| RC | 0.8568 | 0.7751 | 0.8365 | 0.0845 |
| SO | 0.8711 | 0.7883 | 0.8505 | 0.0775 |
| MR | 0.8856 | 0.7342 | 0.8454 | 0.0835 |
| SF | 0.8159 | 0.5562 | 0.7365 | 0.1322 |

The numerical values indicating the best performance are shown in bold.

**Table 3** Precision, recall, *F*-measure and MAE values obtained by applying the proposed method and the other methods on images in HKU-IS dataset

| Method | Precision | Recall | *F*-measure | MAE |
|---|---|---|---|---|
| proposed | **0.7429** | 0.7112 | **0.7354** | **0.1117** |
| MST | 0.7127 | 0.7085 | 0.7117 | 0.1216 |
| MBD+ | 0.6779 | **0.7628** | 0.6957 | 0.1235 |
| RC | 0.6865 | 0.7168 | 0.6933 | 0.1233 |
| SO | 0.7028 | 0.7109 | 0.7046 | 0.1168 |
| MR | 0.7209 | 0.6267 | 0.6967 | 0.1238 |
| SF | 0.6366 | 0.4218 | 0.5697 | 0.1618 |

The numerical values indicating the best performance are shown in bold.

**Table 4** Precision, recall, *F*-measure and MAE values obtained by applying the proposed method and the other methods on images in PASCAL-S dataset

| Method | Precision | Recall | *F*-measure | MAE |
|---|---|---|---|---|
| proposed | **0.7717** | 0.5294 | 0.6976 | 0.1872 |
| MST | 0.7675 | 0.5553 | 0.7053 | 0.1858 |
| MBD+ | 0.7617 | **0.5789** | **0.7100** | **0.1842** |
| RC | 0.7306 | 0.5180 | 0.6674 | 0.2000 |
| SO | 0.7510 | 0.5484 | 0.6920 | 0.1895 |
| MR | 0.7643 | 0.5178 | 0.6886 | 0.1868 |
| SF | 0.6390 | 0.3017 | 0.5080 | 0.2442 |

The numerical values indicating the best performance are shown in bold.

MSRA-1000, MSRA-10K, HKU-IS, PASCAL-S, and DUT-OMRON datasets, respectively.

## 5 Conclusion

In this study, a new salient region detection method has been proposed using multi-scale and directional selectivity properties of the NSCT. The local features have been extracted from the local variations of the low-pass coefficients whereas the global features have been obtained based on the distribution of the directional subband coefficients. The final saliency map has been obtained by combining the local and global features and shown to efficiently represent the pixels that are locally and globally distinctive. In addition, the structure and boundary of the image objects have been preserved by abstracting the saliency maps into meaningful image regions. The proposed method has been compared with a number of existing methods on various datasets. The results have shown that the performance of the proposed method is generally superior to that of the other methods by providing the higher precision, recall, and *F*-measure values and lower MAE values.

**Table 5** Precision, recall, *F*-measure and MAE values obtained by applying the proposed method and the other methods on images in DUT-OMRON dataset

| Method | Precision | Recall | *F*-measure | MAE |
|---|---|---|---|---|
| proposed | **0.5807** | 0.7218 | **0.6082** | **0.1218** |
| MST | 0.5477 | 0.6952 | 0.5759 | 0.1435 |
| MBD+ | 0.5181 | **0.7507** | 0.5580 | 0.1449 |
| RC | 0.5147 | 0.6484 | 0.5404 | 0.1463 |
| SO | 0.5282 | 0.7160 | 0.5619 | 0.1345 |
| MR | 0.5520 | 0.6230 | 0.5670 | 0.1314 |
| SF | 0.4798 | 0.4529 | 0.4733 | 0.1537 |

The numerical values indicating the best performance are shown in bold.

## 6 Acknowledgments

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada and in part by the Regroupement Strategique en Microelectronique du Quebec (ReSMiQ).

## 7 References

[1] Ko, B., Nam, J.-Y.: 'Object-of-interest image segmentation based on human attention and semantic region clustering', *J. Opt. Soc. Am. A*, 2006, **23**, (10), pp. 2462–2470

[2] Itti, L.: 'Automatic foveation for video compression using a neurobiological model of visual attention', *IEEE Trans. Image Process.*, 2004, **13**, (10), pp. 1304–1318

[3] Ma, Y.-F., Zhang, H.-J.: 'Contrast-based image attention analysis by using fuzzy growing'. Proc. ACM Int. Conf. on Multimedia, Berkeley, USA, 2003, pp. 374–381

[4] Avidan, S., Shamir, A.: 'Seam carving for content-aware image resizing', *ACM Trans. Graph.*, 2007, **26**, (3), pp. 10-1–10-9

[5] Zhang, G., Cheng, M., Hu, S., *et al.*: 'A shape preserving approach to image resizing', *Comput. Graph. Forum*, 2009, **28**, (7), pp. 1897–1906

[6] Yuan, Y., Wang, J., Li, B., *et al.*: 'Saliency based ulcer detection for wireless capsule endoscopy diagnosis', *IEEE Trans. Med. Imaging*, 2015, **34**, (10), pp. 2046–2057

[7] Borji, A., Itti, L.: 'State-of-the-art in visual attention modeling', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2013, **35**, (1), pp. 185–207

[8] Borji, A., Cheng, M., Jiang, H., *et al.*: 'Salient object detection: a survey', CoRR, 2014. [Online]. Available at: http://arxiv.org/abs/1411.5878

[9] Borji, A., Cheng, M.M., Jiang, H., *et al.*: 'Salient object detection: a benchmark', *IEEE Trans. Image Process.*, 2015, **24**, (12), pp. 5706–5722

[10] Wang, L., Lu, H., Ruan, X., *et al.*: 'Deep networks for saliency detection via local estimation and global search'. Proc. IEEE Conf. on Computer Vision Pattern Recognition (CVPR), Boston, USA, 2015, pp. 3183–3192

[11] Zhao, R., Ouyang, W., Li, H., *et al.*: 'Saliency detection by multi-context deep learning'. Proc. IEEE Conf. on Computer Vision Pattern Recognition (CVPR), Boston, USA, 2015, pp. 1265–1274

[12] Zou, W., Komodakis, N.: 'HARF: hierarchy-associated rich features for salient object detection'. Proc. IEEE Int. Conf. on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 406–414

[13] Li, G., Yu, Y.: 'Visual saliency based on multiscale deep features'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, USA, 2015, pp. 5455–5463

[14] Lee, G., Tai, Y., Kim, J.: 'Deep saliency with encoded low level distance map and high level features'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016, pp. 660–668

[15] Li, G., Yu, Y.: 'Deep contrast learning for salient object detection'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016, pp. 660–668

[16] Hu, P., Shuai, B., Liu, J., *et al.*: 'Deep level sets for salient object detection'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 2017, pp. 540–549

[17] Wang, L., Lu, H., Wang, Y., *et al.*: 'Learning to detect salient objects with image-level supervision'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 2017, pp. 3796–3805

[18] Chen, X., Zheng, A., Li, J., *et al.*: 'Look, perceive and segment: finding the salient objects in images via two-stream fixation-semantic CNNs'. Proc. IEEE Int. Conf. on Computer Vision (ICCV), Venice, Italy, 2017, pp. 1050–1058


© The Institution of Engineering and Technology 2018

[19] He, S., Jiao, J., Zhang, X., *et al.*: 'Delving into salient object subitizing and detection'. Proc. IEEE Int. Conf. on Computer Vision (ICCV), Venice, Italy, 2017, pp. 1059–1067
[20] Han, J., Zhang, D., Hu, X., *et al.*: 'Background prior-based salient object detection via deep reconstruction residual', *IEEE Trans. Circuits Syst. Video Technol.*, 2015, **25**, (8), pp. 1309–1321
[21] Kuen, J., Wang, Z., Wang, G.: 'Recurrent attentional networks for saliency detection'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016, pp. 3668–3677
[22] Liu, N., Han, J.: 'DHSNet: deep hierarchical saliency network for salient object detection'. Proc. IEEE Conf. Computer Vision Pattern Recognition (CVPR), Las Vegas, USA, 2016, pp. 678–686
[23] Wang, L., Lu, H., Wang, L., *et al.*: 'Saliency detection with recurrent fully convolutional networks'. Proc. European Conf. on Computer Vision (ECCV), Amsterdam, The Netherlands, 2016, pp. 825–841
[24] Perazzi, F., Krahenbuhl, P., Pritch, Y., *et al.*: 'Saliency filters: contrast based filtering for salient region detection'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Providence, USA, 2012, pp. 733–740
[25] Yang, C., Zhang, L., Lu, H., *et al.*: 'Saliency detection via graph-based manifold ranking'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Portland, USA, 2013, pp. 3166–3173
[26] Zhu, W., Liang, S., Wei, Y., *et al.*: 'Saliency optimization from robust background detection'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Columbus, USA, 2014, pp. 2814–2821
[27] Zhang, J., Sclaroff, S., Lin, Z., *et al.*: 'Minimum barrier salient object detection at 80 fps'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, USA, 2015, pp. 1404–1412
[28] Cheng, M.M., Mitra, N.J., Huang, X., *et al.*: 'Global contrast based salient region detection', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2015, **37**, (3), pp. 569–582
[29] Tu, W.C., He, S., Yang, Q., *et al.*: 'Real-time salient object detection with a minimum spanning tree'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016, pp. 2334–2342
[30] Xia, C., Li, J., Chen, X., *et al.*: 'What is and what is not a salient object? Learning salient object detector by ensembling linear exemplar regressors'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 2017, pp. 4399–4407
[31] Tian, Q., Sebe, N., Lew, M.S., *et al.*: 'Image retrieval using wavelet-based salient points', *J. Electron. Imaging*, 2001, **10**, (4), pp. 835–849
[32] Li, J., Levine, M.D., An, X., *et al.*: 'Visual saliency based on scale-space analysis in the frequency domain', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2013, **35**, (4), pp. 996–1010
[33] Imamoglu, N., Lin, W., Fang, Y.: 'A saliency detection model using low-level features based on wavelet transform', *IEEE Trans. Multimed.*, 2013, **15**, (1), pp. 96–105
[34] Rezaei Abkenar, M., Ahmad, M.O.: 'Quaternion-based salient region detection using scale space analysis'. Proc. Signal Processing and Intelligent Systems Conf. (SPIS), Tehran, Iran, 2015, pp. 78–82
[35] Rezaei Abkenar, M., Ahmad, M.O.: 'Superpixel-based salient region detection using the wavelet transform'. Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS), Montreal, Canada, 2016, pp. 2719–2722
[36] Liu, D., Chang, F., Liu, C.: 'Salient object detection fusing global and local information based on non-subsampled contourlet transform', *J. Opt. Soc. Am. A*, 2016, **33**, (8), pp. 1430–1441
[37] Rezaei Abkenar, M., Ahmad, M.O.: 'Salient region detection using efficient wavelet-based textural feature maps', *Multimedia Tools Appl.*, 2018, **77**, (13), pp. 16291–16317
[38] Rezaei Abkenar, M., Sadreazami, H., Ahmad, M.O.: 'Patch-based salient region detection using statistical modeling in the non-subsampled contourlet domain'. Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS), Baltimore, USA, 2017, pp. 1–4
[39] Do, M.N.: 'Directional multi-resolution image representations', PhD dissertation, School Comput. Commun. Sci., Swiss Fed. Inst. Technol, 2001
[40] Da Cunha, A.L., Zhou, J., Do, M.N.: 'The non-subsampled contourlet transform: theory, design and applications', *IEEE Trans. Image Process.*, 2006, **15**, (10), pp. 3089–3101
[41] Oliva, A., Torralba, A., Castelhano, M.S., *et al.*: 'Top down control of visual attention in object detection'. Proc. IEEE Int. Conf. on Image Processing (ICIP), Barcelona, Spain, 2003, pp. 253–256
[42] Ren, Z., Gao, S., Chia, L.T., *et al.*: 'Region-based saliency detection and its application in object recognition', *IEEE Trans. Circuits Syst. Video Technol.*, 2014, **24**, (5), pp. 769–779
[43] Achanta, R., Shaji, A., Smith, K., *et al.*: 'SLIC superpixels compared to state-of-the-art superpixel methods', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012, **34**, (11), pp. 2274–2282
[44] Vincent, L.: 'Morphological grayscale reconstruction in image analysis: applications and efficient algorithms', *IEEE Trans. Image Process.*, 1993, **2**, (2), pp. 176–201
[45] Achanta, R., Hemami, S.S., Estrada, F.J., *et al.*: 'Frequency tuned salient region detection'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Miami, USA, 2009, pp. 1597–1604
[46] Li, Y., Hou, X., Koch, C., *et al.*: 'The secrets of salient object segmentation'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Columbus, USA, 2014, pp. 4321–4328


© The Institution of Engineering and Technology 2018