



A novel approach for salient object detection using double-density dual-tree complex wavelet transform in conjunction with superpixel segmentation

Rinki Arya¹ · R. K. Agrawal¹ · Navjot Singh²

Received: 14 January 2017 / Revised: 11 March 2018 / Accepted: 9 June 2018 / Published online: 9 July 2018
© Springer-Verlag London Ltd., part of Springer Nature 2018

Abstract

Salient object detection in wavelet domain has recently begun to attract researchers' effort due to its desired ability to provide multi-scale analysis of an image simultaneously in both frequency and spatial domains. The proposed algorithm exploits the inherent multi-scale structure of the double-density dual-tree complex-oriented wavelet transform (DDDTCWT) to decompose each input image into four approximate sub-band images and 32 high-pass detailed sub-band images at each scale. These 32 detailed high-pass sub-bands at each scale are adequate to represent singularities of any geometric object with high precision and to mimic zooming-in and zooming-out process of human vision system. In the proposed model, we first compute a rough segmented saliency map (RSSM) by fusing multi-scale edge-to-texture features generated from DDDTCWT with segmentation results obtained from bipartite graph partitioning-based segmentation approach. Then, each pixel in RSSM is categorized into either background region or salient region based on a threshold. Finally, the pixels of the two regions are considered as samples to be drawn from a multivariate kernel function whose parameters are estimated using expectation maximization algorithm, to generate a saliency map. The performance of the proposed model is evaluated in terms of precision, recall, F -measure, area under the ROC curve and computation time using six publicly available image datasets. Extensive experimental results on six benchmark datasets demonstrate that the proposed model outperformed the existing 29 state-of-the-art methods in terms of F -measure on all five datasets, recall on four datasets and area under ROC curve on two datasets. In terms of mean recall value, mean F -measure value and mean AUC value on all six datasets, the proposed method outperforms all state-of-the-art methods. The proposed method also takes comparatively less computation time in comparison with many existing spatial domain methods.

Keywords Salient object detection · Visual saliency · 2D double-density discrete wavelet transform (DDDWT) · 2D dual-tree complex wavelet transform (DTCWT) · 2D dual-tree real wavelet transform (DTRWT) · Double-density dual-tree complex wavelet transform (DDDTCWT) · Inverse double-density dual-tree complex wavelet transform (IDDDTCWT) · Gaussian mixture model (GMM) · Expectation maximization (EM) · Saliency map

✉ Rinki Arya
rinki.arya89@gmail.com

Extended author information available on the last page of the article

1 Introduction

The human visual system (HVS) can accurately detect the most salient object in an image, but the goal of developing a computational model for salient object detection [39, 40, 62, 69, 93] with comparable capabilities still remains an open challenge for computer vision and pattern recognition. The task of salient object detection is to locate the most salient object or region in an image. In recent years, salient object detection [10, 46, 49, 89] has gained increasing attentions for many computer vision and graphics applications, such as image and video compression [34], image retargeting [88], image thumb nailing [67], image segmentation [14, 53], object recognition [31, 70, 75], content-aware image editing [3], image classification/retrieval [4, 45], video surveillance systems [38], photograph rearrangement [72], image quality assessment [71], remote sensing [58], automatic image cropping [76], displaying items on small portable screens [17], automatic target detection [43, 44], robotics [17, 43, 76], medical imaging [50], advertising a design [43], image collection browsing [74] and image enhancement [29]. Motivated by these applications, salient object detection emphasizes on highlighting foreground objects.

In general, multi-scale analysis using wavelet helps in representing the images with different resolutions and in implementing the zooming-in and zooming-out process of human vision system (HVS). The ability of wavelet transform to represent singularities of images plays a key role in designing wavelet-based saliency detection algorithms as human eyes are very sensitive to orientation features. It has been verified that the multi-scale edge-to-texture features computed using discrete wavelet transform (DWT) play a significant role in the field of salient object detection because of its ability to provide multi-scale analysis of an image simultaneously in both frequency and spatial domains [41, 52]. Unfortunately, DWT-based saliency detection techniques have a limited ability to reveal singularities in different directions as it has only three directional sub-bands, oriented at 0° , 45° and 90° . But as natural images are comprised of smooth regions that are punctuated with edges at several orientations, DWT may fail to represent the geometric regularity along the singularities, which requires higher directional selectivity. In 2004, Selesnick [77–79] proposed 2D double-density dual-tree complex wavelet transform (DDDCWT) which gives rise to four approximate sub-band images and 32 high-pass detailed sub-band images at each scale which are adequate for the representation of any geometric object with high precision. This motivates us to use double-density dual-tree complex wavelet transform (DDDCWT) to detect salient object. In this paper, we first compute detailed multi-scale edge-to-texture feature maps using inverse double-density dual-tree complex wavelet transform (IDDCWT) to capture band-pass local information with different frequency bandwidths which helps in detecting irregularities at different bandwidths. Then, these feature maps are combined to generate a saliency map. This DDCWT-based saliency map is integrated with the segmentation results obtained from bipartite graph partitioning-based approach to generating an initial rough segmented saliency map (RSSM). Each pixel of RSSM is assigned to be part of salient region or the background region based on its value relative to threshold value. Finally, the pixels of the two regions are considered as samples to be drawn from a multivariate kernel function whose parameters are estimated using expectation maximization algorithm, to yield a final saliency map. The performance of the proposed model is evaluated in terms of precision, recall, F -measure, area under curve and computation time on six publicly available image datasets. Performance of the proposed model is also evaluated in terms of mean precision value, mean recall value, mean F -measure values and mean AUC values on all six datasets.

Both qualitative and quantitative evaluations on six publicly available benchmark datasets demonstrate the robustness and efficacy of the proposed method against 29 state-of-the-art methods.

The remainder of this paper is organized as follows. Section 2 reviews related state-of-the-art methods to detect salient object. Section 3 describes the proposed model (DDDTCWT-SS). Section 4 presents the experimental results and comparisons with several state-of-the-art salient region detection methods. Section 5 concludes the proposed model with discussions.

2 Related work

Recently, a plethora of computational models have been proposed for salient object detection [18, 19, 28, 55–57, 83–85], which can be roughly categorized into bottom-up and top-down approaches [11–13]. Bottom-up approaches are fast, stimulus driven and task independent. They extract certain low-level features from the image and combine them into a saliency map. However, the top-down approaches consist of the prior knowledge of the human visual system (HVS) and high-level data processing to support the task of salient object detection. Therefore, top-down approaches are slow and task dependant. Such approaches are integrated with the bottom-up approaches in order to detect the salient object. Both of the computational models focus on producing saliency maps to detect salient object in an input image. For a comprehensive review of related work, we refer readers to recent survey papers for detailed discussion of 256 focused researches in computer vision [12], as well as taxonomy and critical comparison of 65 models [11], and qualitative and quantitative analysis of 41 different state-of-the-art [13] methods in the two major research areas such as fixation prediction [11, 48] and salient object detection [13].

Here, we focus on different bottom-up approaches by which salient object detection has been designed. Inspired by the biologically plausible architecture proposed by Koch and Ullman [54], Itti et al. [44] (IT) determined image saliency by utilizing centre-surround differences across multi-scale image features using a difference of Gaussians (DoG) approach. Later, Bruce and Tsotsos [15] proposed a computational model (AIM) based on information maximization to implement saliency using joint likelihood, Shannon's self-information and features learned from input images using independent component analysis (ICA). Harel et al. [35] utilized Itti et al.'s [44] method to create low-level feature maps but performed normalization using a graph-based approach (GBVS). Liu et al. [64, 65] (SLRG) integrated saliency cues like centre-surround histogram contrast, multi-scale contrast and colour spatial distribution in a conditional random field to segment the salient object. Zhang et al. (SUN) [94] used Bayesian framework to locate the salient object in an image. Achanta et al. (FT) [3] computed saliency of each pixel in the image as the contrast of its colour feature to the mean colour information of the whole image. The research work of Achanta and Susstrunk (ASS) [2] relied on the maximum symmetric surround difference to compute saliency map. Goferman et al. [30] (CASD) used contrast of the patch to the K nearest patches in the image to compute saliency. Shen and Wu [80] solved saliency detection problem as a low-rank matrix (LRK) recovery problem, where salient objects are represented by a sparse matrix (noise), while background is indicated by a low-rank matrix. However, this sparse and low-rank assumption may not be satisfied in complex scenes, leading to unsatisfactory results. The work of Liu et al. (2014) (STREE) locates salient objects by exploiting the concept of saliency tree. Zhu et al. [96] (MSA) used multivariate normal distribution estimation to extract salient regions in an image. Xie et al. [90] (BSM) pro-

posed a Bayesian saliency method by utilizing the low- and mid-level cues. Sun et al. [86] (MCA) used the concept of Markov chain absorption to detect salient object present in the image. Jiang et al. [47] performed pre-segmentation on an input image and extracted a bunch of discriminative features from each segmented region. Then, a random forest regressor is adopted to map multiple features to a region saliency score. Singh et al. (SOD-C-PSO) [82] suggested linearly weighted combination of different feature maps and estimated the weights using constrained particle swarm optimization. Cheng et al. [19] proposed an unsupervised saliency cut (Grab cut)-based image segmentation approach to automatically segmenting the most salient object. They also proposed histogram-based contrast (HC) and spatial information-enhanced region-based contrast (RC) methods for salient object detection. Kim et al. [51] introduced a high-dimensional colour transform (HDCT)-based saliency detection approach. The main idea of their approach is to map a low-dimensional RGB colour to a feature vector in a high-dimensional colour space in order to separate the salient object from the background by finding an optimal linear combination of colour coefficients.

Recent studies [32, 36, 37] have tried to detect image saliency in transform domain [16]. Frequency domain approaches for salient object detection have been popular due to their fast computational speed. The first spectral domain approach for detecting saliency is due to Hou and Zhang [36], who computed image saliency in frequency domain by comparing the dissimilarity of the characteristic spectrum with the perceived spectrum of greyscale images. Hou et al. [37] proposed an image signature (IS) descriptor to spatially approximate the sparse foreground position concealed in a spectrally sparse background. In 2008, Guo et al. [33] observed that phase spectrum of the Fourier transform (PFT) in comparison with amplitude spectrum contributes more to locate the position of salient object. Guo et al. [32, 33] extended PFT model to a phase quaternion Fourier transform (PQFT) model in case of multiple channels to represent the multi-dimensional data at each pixel as a quaternion. Yu et al. [92] used lateral surround inhibition behaviour of neurons to compute saliency in an image. They captured this behaviour of neurons by utilizing the pulsed discrete cosine transform (PCT). In 2008, biological prediction and comparison with spatial biological models were verified by Bian and Zhang [7–9] in their frequency divisive normalization (FDN) model. Bian and Zhang [7–9] proposed a saliency detection approach that integrates the speed of frequency domain models with the topology of biologically based methods under the assistance of frequency domain divisive normalization (FDN). But, this model takes global surround into consideration. In order to relax the global surround constraint, Bian and Zhang [8, 9] extended FDN model into piecewise frequency domain divisive normalization (PFDN) [8, 9] by separating the input image into overlapping local patches and conducting FDN on every patch in order to provide better biological plausibility. Amplitude spectrum of quaternion Fourier transform (AQFT) [24] and modelling from bitstream (BS) [25] were proposed by Fang et al. In 2013, Li et al. [60] proposed hypercomplex Fourier transform (HFT)-based saliency detection approach which takes advantage of scale-space analysis of the amplitude spectrum. Li et al. [61] (SDS) designed a saliency detector by exploiting the phase of intermediate frequencies. Arya et al. [5] (HLGM) suggested a salient region detection approach by integrating both global saliency and local saliency in the frequency domain by using fast Walsh–Hadamard transform (FWHT) and PFDN, respectively. Arya et al. [6] (BHGT) developed a grey-level co-occurrence matrix (GLCM)-based saliency framework in both spatial domain and frequency domain.

Recently, wavelet transform (WT) has been found to be useful in the field of salient object detection. In 2001, Tian et al. [87] proposed a salient point detector based on wavelet transform. As this WT-based approach detects salient points in an image rather than salient

objects, therefore it is difficult for us to compare our proposed salient object detection model with this approach. Murray et al. [68] (SIM) computed saliency based on a nonparametric low-level vision approach, where the scale information is integrated through a simple inverse wavelet transform over the set of extended contrast sensitivity function (ECSF) responses for each colour sub-band. ECSF is a function of scale and centre-surround contrast energy which takes care of the human sensitivity to local contrast and energy ratio of the centre-surround regions. Moreover, they also introduced training steps on both colour appearance and eye-fixation psychophysical data to reduce ad hoc parameters. İmamoğlu et al. [41] (WT) proposed a salient object detection model by utilizing low-level features obtained from the discrete wavelet transform (DWT) domain to create multi-scale feature maps, which can represent different features from edges to texture. These multi-scale feature maps with increasing frequency bandwidths are obtained using inverse wavelet transform with the band-pass filtered regions of the input image at various scales. Using these features, local saliency at a location is modulated with its global saliency calculated based on the likelihood of the features to generate final saliency map. However, DWT cannot be an optimal choice to create feature maps as it gives weak line (curve) singularities because of being limited to few directional sub-bands.

3 Proposed model

In this paper, we propose a double-density dual-tree complex-oriented wavelet transform (DDDCWT)-based salient object detection model (DDDCWT-SS), as illustrated in Fig. 1. In this framework, an input image is first segmented using a computationally efficient bipartite graph partitioning-based segmentation approach to capturing intrinsic structural information of the image. For each segment, saliency is computed based on multi-scale low-level edge-to-texture features extracted from two-dimensional double-density dual-tree complex wavelet transform (2D DDDTCWT). Figure 1a, b shows the respective results of segmentation procedure and saliency map calculated using DDDTCWT. As humans are sensitive to orientation features [66], an initial rough segmented saliency map (RSSM) (as shown in Fig. 1c) is generated by assigning saliency to each segment on the basis of DDDTCWT-based saliency map which comprises specific information from 32 detailed sub-band images as shown in Fig. 3d.

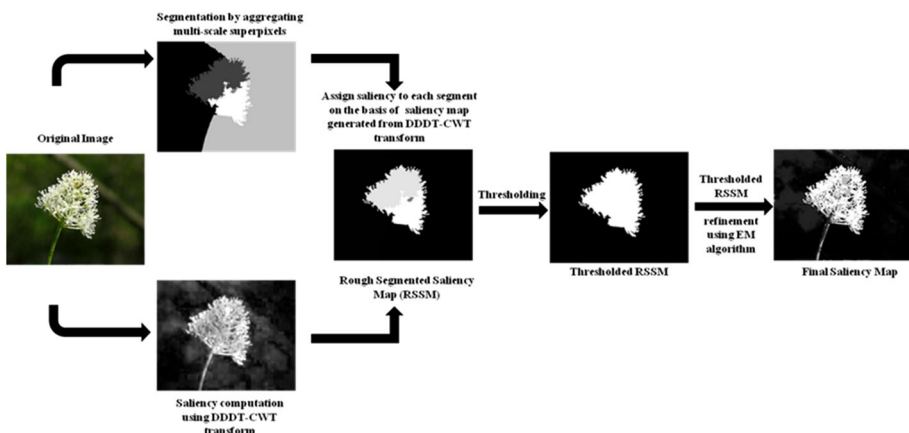


Fig. 1 Proposed framework for detecting the most salient object in an image

Then, a mean intensity value of DDDTCWT-based saliency map is used as a threshold to get a rough estimation of salient region and background region in RSSM. In the initial RSSM, a pixel with intensity value greater than or equal to the threshold value is considered to be a part of salient object and is assigned a value '1', and pixel with intensity value less than the threshold value is considered to be a part of background region and is assigned a value '0'. In this way, we get a thresholded RSSM which is shown in Fig. 1d. To improve the accuracy of the thresholded initial RSSM, the Gaussian mixture model (GMM) is built over it for saliency re-estimation of each region. The GMM parameters are updated using expectation maximization (EM) method to obtain the final saliency map (shown in Fig. 1e).

The main contribution of the proposed method is to exploit sensitiveness of human eyes to orientation features using multi-scale structure of the 2D DDDTCWT as natural images are comprised of smooth regions that are punctuated with edges at several orientations. The proposed algorithm performs a multi-scale frequency analysis of the image by representing it at different resolutions to exploit zooming-in and zooming-out process of HVS and sets a trade-off between the detection accuracy and computational time for achieving better detection accuracy in less computation time. The pseudo-code of the proposed algorithm is given as follows.

Algorithm: DDDTCWT-SS

Input: An input image I and the number of segments ($k=5$).

Output: A saliency map

- 1: Segment the input image by aggregating super-pixels using a bipartite graph.
 - 2: Compute pixel-wise saliency in the original image using multi-scale edge-to-texture features extracted from 2-D DDDTCWT.
 - 3: For each segment computed in Step 1, saliency values (computed in Step 2) of all the pixels belonging to that segment are averaged out to construct initial rough segmented saliency map (RSSM).
 - 4: In order to get a rough estimation of salient region and background region in RSSM, a threshold is calculated by taking mean of saliency values of all pixels in the image calculated in Step 2.
 - 5: Refinement of initial rough segmented saliency map (RSSM) using Gaussian Mixture Model (GMM) and Expectation Maximization (EM) algorithm to generate final saliency map.
-

3.1 Segmentation of an input image by aggregating superpixels using a bipartite graph

As demonstrated in recent studies [47, 51, 73, 80, 91], features from superpixels [27, 28] are effective and efficient for salient object detection. Superpixels group pixels into perceptually meaningful atomic regions which can be used to replace the rigid structure of the pixel grid. They capture image redundancy, provide a convenient primitive to compute image features, and greatly reduce the complexity of subsequent image processing tasks. These are key building blocks for many computer vision algorithms like image segmentation. There are many approaches to generating superpixels, each with its own advantages and drawbacks that may be better suited to a particular application. For example, graph-based method of Felzenszwalb and Huttenlocher [26] may be an ideal choice [1] to accurately capture image boundaries. Recently, Li et al. [59] proposed an improved image segmentation algorithm which takes advantage of different and complementary information from various popular

segmentation algorithms [20, 26, 81]. In order to fuse the complementary information, Li et al. collected a variety of superpixels generated by different segmentation algorithms with varying parameters. Superpixels generated in this way help in capturing diverse and multi-scale visual patterns in the input image. To effectively aggregate these multi-layer superpixels, Li et al. proposed a bipartite graph partitioning-based segmentation framework which is constructed over both pixels and superpixels. These pixels and superpixels work as the vertices of the bipartite graph and edges between these vertices are established on the basis of superpixel cues and smoothness cues. To enforce superpixel cues, a pixel is connected to the superpixel if pixel is a part of that superpixel while smoothness cues are enforced by connecting each superpixel to itself and its nearest neighbour in the feature space among its spatially adjacent superpixels. This bipartite graph segmentation framework is efficiently solved computationally by Li et al. [59] using a linear-time spectral algorithm.

To map the relationship between pixels and superpixels, a bipartite graph is built which consists of two parts describing the pixel–superpixel and superpixel–superpixel relationships. In particular, taking into account the demand of sparsity for a good-quality graph, a pixel is connected to the superpixel containing it and a superpixel is connected to neighbouring superpixel close in feature space. For a given image \mathbf{I} , a set of pixels and superpixels (multi-layer) are denoted by P and S , respectively. More precisely, let $\mathbf{G} = \{\chi, \gamma, \mathbf{B}\}$ denote the bipartite graph, where $\chi = P \cup S = \{\mathbf{x}_i\}_{i=1}^{N_\chi}$, $\gamma = S = \{\mathbf{y}_j\}_{j=1}^{N_\gamma}$ with $N_\chi = |P| + |S|$ and $N_\gamma = |S|$ and the number of nodes in χ and γ , respectively.

The across-affinity matrix $\mathbf{B} = (b_{ij})_{N_\chi \times N_\gamma}$ is defined as:

$$b_{ij} = \begin{cases} \alpha, & \mathbf{x}_i \in \mathbf{y}_j, \mathbf{x}_i \in P, \mathbf{y}_j \in S \\ e^{-\beta d_{ij}}, & \text{if } \mathbf{x}_i \sim \mathbf{y}_j, \mathbf{x}_i \in S, \mathbf{y}_j \in S, \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where d_{ij} signifies the distance between the features of superpixels \mathbf{x}_i and \mathbf{y}_j . We use $d_{ij} = \mathbf{c}_i - \mathbf{c}_j$, where \mathbf{c}_i and \mathbf{c}_j represent the average colour of the pixels within the superpixels \mathbf{x}_i and \mathbf{y}_j , respectively, on the basis of colour space. \sim signifies a certain neighbourhood between superpixels. $\mathbf{x} \sim \mathbf{y}, \mathbf{x} \in S, \mathbf{y} \in S$, if $\mathbf{x} = \mathbf{y}$, or \mathbf{y} is adjacent to \mathbf{x} and is most similar to \mathbf{x} in terms of average colour. α and β are set to greater than 0 to balance superpixel and smoothness cues, respectively. With the help of this construction, a pixel and a superpixel that it belongs to are likely to be grouped together due to the connections between them. Two superpixels close in feature space are also more likely to be grouped together. Bipartite graph, constructed in this manner, also enforces the smoothness over superpixels. Using bipartite graph \mathbf{G} , input image \mathbf{I} is segmented into k segments by accumulating same label nodes into a segment with the help of spectral clustering algorithm. To segment an image into k groups, k bottom eigenvectors of generalized eigenvalue problem are computed as:

$$\mathbf{L}f = \lambda \mathbf{D}f, \tag{2}$$

where \mathbf{L} and \mathbf{D} represent graph Laplacian and degree matrix, respectively. \mathbf{D} is calculated as: $\mathbf{D} = \text{diag}(\mathbf{B}\mathbf{1})$. Instead of solving this eigenvalue problem using Lanczos method and singular value decomposition method, which take $O(k(N_\chi + N_\gamma)^{3/2})$ running time [59], Li et al. utilized unbalanced structure of the graph to solve it efficiently. The number of columns in affinity matrix \mathbf{B} is much larger than the number of rows ($N_\chi = N_\gamma + |\mathbf{I}|$, and $\mathbf{I} \gg N_\gamma$), so we have $N_\chi \gg N_\gamma$. This large variation between the number of rows and number of columns clearly demonstrates the unbalanced structure of the bipartite graph. To exploit

the unbalanced structure, Li et al. proposed a transfer cut method to compute bottom k eigenvectors in reduced time as it transforms the eigenvalue problem into the following:

$$\mathbf{L}_\gamma \mathbf{v} = \eta \mathbf{D}_\gamma \mathbf{v}, \quad (3)$$

where $\mathbf{L}_\gamma = \mathbf{D}_\gamma - \mathbf{W}_\gamma$, $\mathbf{D}_\gamma = \text{diag}(\mathbf{B}^T \mathbf{1})$ and $\mathbf{W}_\gamma = \mathbf{B}^T \mathbf{D}_\chi^{-1} \mathbf{B}$, where $\mathbf{D}_\chi = \text{diag}(\mathbf{B} \mathbf{1})$. \mathbf{L}_γ is exactly the Laplacian of the bipartite graph $\mathbf{G}_\gamma = \{\gamma, \mathbf{W}_\gamma\}$ as $\mathbf{D}_\gamma = \text{diag}(\mathbf{B}^T \mathbf{1}) = \text{diag}(\mathbf{W}_\gamma \mathbf{1})$, where $\mathbf{1}$ is the vector of ones of appropriate size. The task of partitioning graph \mathbf{G} into k groups takes $O\left(2k(1+d_\chi)N_\chi + kN_\gamma^{\frac{3}{2}}\right)$ time where d_χ is the average number of edges connected to each node in χ . Our work belongs to salient object detection, for which a comprehensive discussion about segmentation approaches is beyond the scope of this paper. We refer readers to the research article of Li et al. [59] for a detailed discussion of this segmentation approach.

To choose the most salient region among these k segmented regions $\mathbf{H}_p, p = 1, \dots, k$, the saliency value of each region, \mathbf{H}_p , needs to be computed. To find the saliency value of each region, we utilize double-density dual-tree complex wavelet transform (DDDTCWT), which is discussed in Sect. 3.2.

3.2 Double-density dual-tree complex wavelet transform (DDDTCWT)-based saliency detection model

Psychophysical investigation has shown that the HVS performs a multi-scale frequency analysis when we observe an image [21, 87]. This mechanism is similar to zooming-in and zooming-out process of HVS. As natural images exhibit smooth regions that are punctuated with edges at several orientations, discrete wavelet transform (DWT) may fail to represent the geometric regularity along the singularities selectivity. In order to overcome the limitations of directional selectivity of traditional 2D discrete wavelet transform (DWT), we utilize double-density dual-tree complex wavelet transform (DDDTCWT) for salient object detection for the first time in the literature. DDDTCWT possesses the properties of the dual-tree complex wavelet transform (DTCWT) and double-density DWT (DDDWT). Both the double-density DWT and the dual-tree complex DWT are similar in several properties. (Both are nearly shift invariant; both are over-complete by a factor of 2; both are based on FIR perfect reconstruction filter banks.) However, the two wavelets used in dual-tree DWT form an approximate Hilbert transform pair, while the two wavelets used in the double-density DWT are offset by one half. We briefly describe DTCWT and DDDWT in the following sections.

3.2.1 Dual-tree complex wavelet transform (DTCWT)

The dual-tree complex wavelet transform (DTCWT) [79] has more directional sub-bands in comparison with 2D DWT, which has only three directional sub-bands oriented at 0° , 45° and 90° . 1D DTCWT is implemented using two real discrete wavelet transforms $\Psi h(t)$ and $\Psi_g(t)$ which are employed in parallel to generate the real and imaginary parts of complex wavelet $\Psi(t) = \Psi h(t) + j\Psi_g(t)$. Here, $\Psi h(t)$ is approximately analytic and $\Psi_g(t)$ is approximately the Hilbert transform of $\Psi h(t)$ i.e. $\Psi_g(t) \approx H(\Psi h(t))$.

2D DTCWT is realized by filtering an image separately row and column wise: two trees are used for the rows of the image and two trees for the columns. This process computes 12 sub-bands for each scale in six main directions $\pm 15^\circ$, $\pm 45^\circ$ and $\pm 75^\circ$, but there are two wavelets in each direction. One of the two wavelets can be interpreted as the real part of a

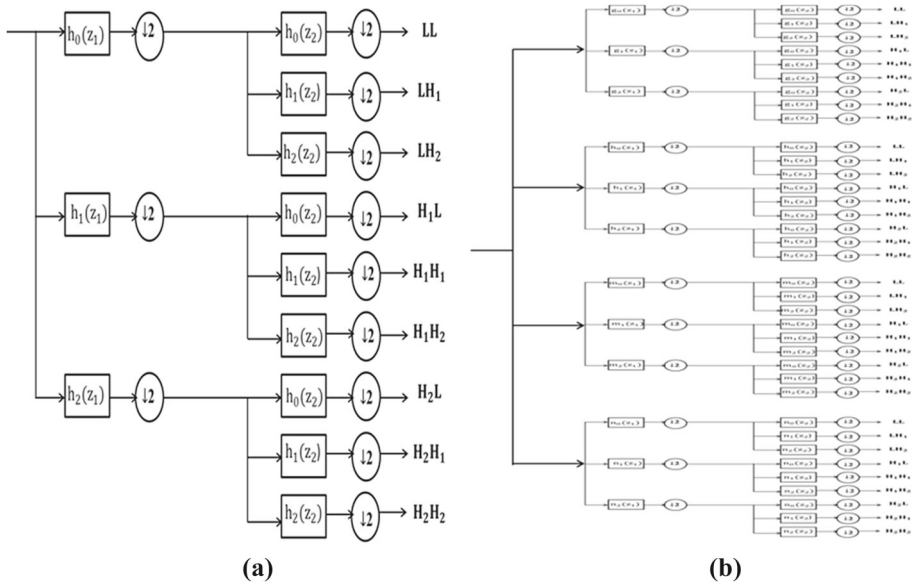


Fig. 2 **a** Filter bank structures for two-dimensional DDDWT, **b** filter bank for two-dimensional DDDTCWT

complex-valued 2D wavelet, while the other wavelet can be interpreted as the imaginary part of a complex-valued 2D wavelet. Unlike 2D DWT, all of the wavelets are free of draughtboard artefact. But six directions are also not sufficient to represent any geometric object with high precision. To capture more information from more directions, Selesnick proposed double-density discrete wavelet transform [77] which is explained in Sect. 3.2.2.

3.2.2 Double-density discrete wavelet transform (DDDWT)

The double-density discrete wavelet transform [77] utilizes one scaling function and two distinct wavelets which are designed to be offset from one another by one half ($\Psi_2(t) = \Psi_1(t - 0.5)$). It satisfies the properties of approximate shift invariant and perfect reconstruction with limited redundancy. In 2D images, this transform outperforms the standard DWT and DTCWT as both have fewer degrees of freedom in comparison with the DDDWT. The procedure of 2D DDDWT is shown in Fig. 2a. 2D DDDWT is realized by alternatively applying the transform first to the rows and then to the columns of the image. After such process, one approximate sub-band image and eight detail sub-band images are attained to describe information in eight distinct directions as shown in Fig. 2a.

3.2.3 Double-density dual-tree complex wavelet transform (DDDCWT)

Although the 2D DDDWT utilizes more wavelets, some lack a dominant spatial orientation, which prevents them from being able to isolate those directions. To overcome this problem, Selesnick [78] suggested double-density dual-tree DWT (DDDCWT) which combines the characteristics of the double-density DWT (DDDWT) and dual-tree complex DWT (DTCWT). The DDDTCWT employs two different scaling functions $\Phi_h(t)$ and $\Phi_g(t)$ and

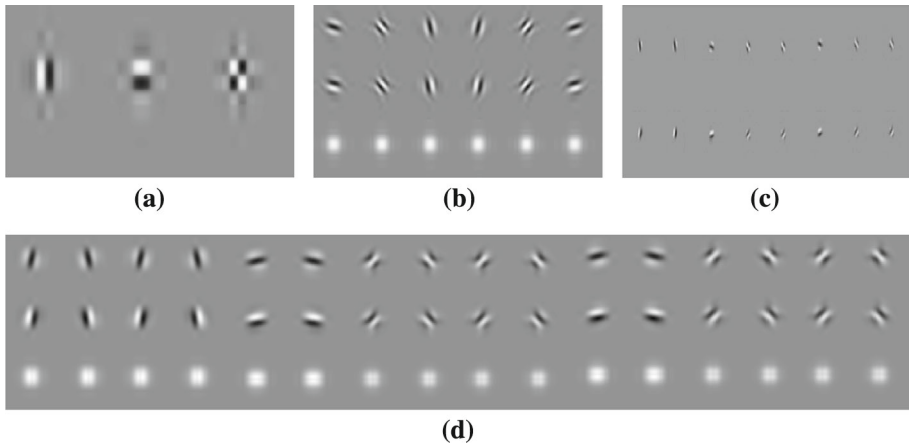


Fig. 3 Impulse responses of **a** DWT, **b** DTCWT, **c** DDDWT, **d** DDDTCWT

four distinct wavelet functions $\Psi_{h,j}(t), \Psi_{g,j}(t) (j = 1, 2)$, where the two wavelets $\Psi_{h,j}(t)$ and $\Psi_{g,j}(t)$ are offset from one another by one half:

$$\Psi_{h,1}(t) = \Psi_{h,2}(t - 0.5), \Psi_{g,1}(t) = \Psi_{g,2}(t - 0.5) \tag{4}$$

and the two wavelets $\Psi_{h,1}(t)$ and $\Psi_{g,1}(t)$ form an approximate Hilbert transform pair:

$$\Psi_{g,1}(t) \approx H\{\Psi_{h,1}(t)\}. \tag{5}$$

Similarly, the two wavelets $\Psi_{h,2}(t), \Psi_{g,2}(t)$ form an approximate Hilbert transform pair:

$$\Psi_{g,2}(t) \approx H\{\Psi_{h,2}(t)\}. \tag{6}$$

The properties satisfied by the four wavelet functions ensure that the DDDTCWT has improved directional selectivity. The 2D DDDTCWT is realized by employing four over-sampling 2D DDDWT in parallel to the same image with different filter sets for the rows and columns. We then take the sum and difference. This gives rise to 36 2D sub-band images as shown in Fig. 2b, four of which are the 2D low-pass sub-bands and the other 32 are 2D high-pass (detailed) sub-bands which describe more specific information in 16 directions. The procedure of two levels 2D DDDTCWT is shown in Fig. 2b.

The impulse responses of DWT, DTCWT, DDDWT and DDDTCWT are shown in Fig. 3a–d, respectively.

The filter banks are applied recursively to the low-pass sub-band, using the analysis filters for the forward transform and the synthesis filters for the inverse transform. The synthesis filters are the time-reversed versions of the analysis filters. The filter bank structure can be implemented using FIR (finite impulse response) perfect reconstruction filter banks. It is believed that 32 detailed sub-band images generated by 2D DDDTCWT are sufficient to represent any geometric object exactly with high precision [95].

In Fig. 4a, an image containing a curve is purposely designed to demonstrate the improved directionality property of DDDTCWT [95]. As shown in Fig. 4b, DWT reconstructions can only accurately represent vertical and horizontal lines. The reconstructed curve looks smoother with reduced artefacts in Fig. 4c, d due to more directional sub-bands of DTCWT and DDDWT, respectively. However, the reconstructed image shown in Fig. 4e is much closer

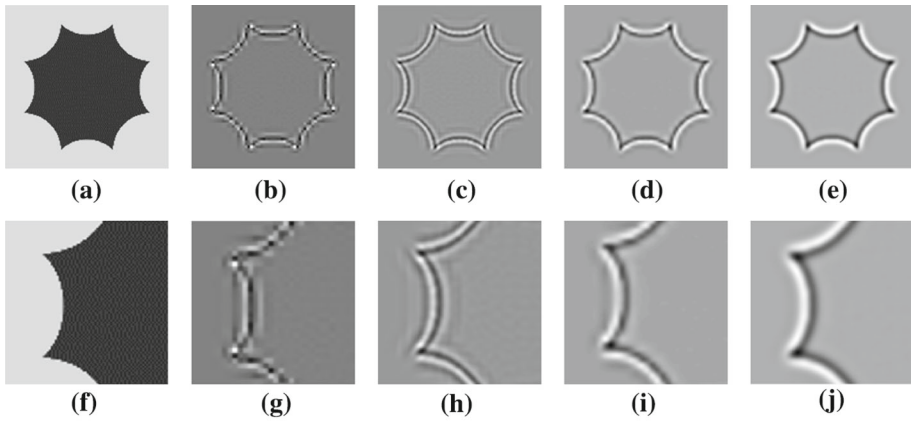


Fig. 4 Improved directionality of double-density dual-tree complex wavelet transform: **a** original test image; reconstructed image using only the lowest level coefficients of **b** DWT, **c** DTCWT, **d** DDDWT, **e** DDDTCWT, **f** zoomed-in original image; zoomed-in reconstructed image using only the lowest level coefficients of **g** DWT, **h** DTCWT, **i** DDDWT, **j** DDDTCWT. Grey level is normalized between [0, 1] for all images and 4-level transform is used

to the original image due to more number of orientations sub-bands in case of DDDTCWT. Therefore, in this paper, we utilize DDDTCWT to compute edge-to-texture feature maps.

3.2.4 Features extraction using double-density dual-tree complex wavelet transform (DDDCWT)

To generate feature maps, input RGB image **I** is first converted into CIE Lab colour space, which is much closer to human vision. Then, the input image **I** is filtered using a low-pass filter to remove the noise:

$$\mathbf{I}' = \mathbf{I} * \mathbf{W}, \tag{7}$$

where **W** is a 2D Gaussian low-pass filter of size 3×3 ; **I'** is the filtered image; * represents the convolution operation. Using Eq. 8, we compute the 2D DDDTCWT (·) of the given image **I'**. The wavelet coefficients *o* are stored in a cell array data structure $o\{s\}\{g\}\{t\}\{d\}$, for $s = 1, 2, \dots, N, t = 1, 2, g = 1 - 2d \in 1 \dots 8$, where *g* represents either the real or imaginary part (by 1 or 2, respectively), and (*t, d*) represents the orientation.

$$[A_N^c, (o\{s\}\{g\}\{t\}\{d\})_c] = DDDTCWT_N(\mathbf{I}'), \tag{8}$$

where 2D DDDTCWT is implemented using analysis filters (as explained in Sect. 3.2.3) for N level decomposition; N denotes maximum scaling number for DDDTCWT, i.e. the resolution index $s \in \{1, \dots, N\}$ and the Nth level refers to the coarsest resolution; *c* is the channel of **I'** as $c \in \{L, a, b\}$; A_N^c denotes the approximation output at Nth level for channel *c*. In total, DDDTCWT isolates edges by capturing information from 32 detailed high-pass sub-band images.

The detailed wavelet coefficients (neglecting approximation coefficients A_N^c) are utilized to compute several feature maps, signifying the contrast from edge to texture, by using inverse double-density dual-tree complex wavelet transform (IDDDTCWT), which is implemented

using synthesis filters of DDDTCWT as explained in Sect. 3.2.3. The feature map for a given pixel, channel c and s th-level decomposition is computed using IDDDTCWT(\cdot) as:

$$f_s^c(x, y) = \frac{((\text{IDDDTCWT}_s(\{o\}\{g\}\{t\}\{d\})_c)^2}{\eta}, \quad (9)$$

where η is the scaling factor. For a given input image, we obtained $3 \times N$ feature maps. Further, these feature maps are utilized to generate a saliency map, $S(x, y)$, as follows:

$$S(x, y) = \left(\sum_{s=1}^N \operatorname{argmax} \left(f_s^L(x, y), f_s^a(x, y), f_s^b(x, y) \right) \right) * \mathbf{P}, \quad (10)$$

where $f_s^L(x, y)$, $f_s^a(x, y)$ and $f_s^b(x, y)$ represent feature maps for L , a , and b channels, respectively, at scale s and \mathbf{P} is a 2D low-pass Gaussian filter.

3.3 Initial rough segmented saliency map (RSSM) generation

We use DDDTCWT coefficients to locate the salient object in an image, while superpixel segmentation is utilized to improve the object contours. Given a segmented region H_p , $p = 1, \dots, k$, where k is the number of segmented regions, the average intensity of each region H_p is computed based on the corresponding DDDTCWT coefficients of the region in the saliency map. Each pixel $x \in H_p$ is assigned the average intensity value v , which is computed as:

$$v = \sum_{i=1}^{|H_p|} v_i / |H_p|, \quad \forall x \in H_p, p = 1, \dots, k, \quad (11)$$

where v_i is the intensity value of the i th pixel x_i . $|H_p|$ is the number of pixels in region H_p . In this way, an initial rough segmented saliency map (RSSM) is obtained where each segment is assigned with a saliency value calculated from DDDTCWT coefficients. Then, we use an average intensity value of DDDTCWT coefficients as a threshold to get a rough estimation of salient and background regions in RSSM. If a pixel intensity value in initial RSSM is greater than or equal to the threshold, then the pixel is considered to be salient and assigned a value of '1' otherwise background by assigning it '0'. By examining RSSM, it is noted for some images that the some parts of salient objects are not highlighted. It might be because of the following reasons: (1) some pixels may be misclassified by the bipartite graph partitioning-based segmentation approach or (2) some pixels may be wrongly detected as a part of background object while actually being a part of the salient object (or vice versa) by the proposed DDDTCWT-based saliency detection method.

To further improve the accuracy of the initial RSSM, the Gaussian mixture model (GMM) is built over RSSM to re-estimate the saliency of each region. The parameters of GMM are updated using expectation maximization (EM) method to obtain the final saliency map, which is discussed in Sect. 3.4.

3.4 Refinement of initial rough segmented saliency map (RSSM) using Gaussian mixture model (GMM) and expectation maximization (EM) algorithm

A Gaussian mixture model (GMM) is useful for modelling data that come from one of several groups: The groups might be different from each other, but data points within the same group

can be well modelled by a Gaussian distribution. The parameters of the GMM include the strengths (weights), means and covariances of the Gaussian distributions. Initial RSSM, given in Sect. 3.3, has two regions: salient and background. The pixels of these two regions in initial RSSM are regarded as two different Gaussian kernels. Then, a GMM is constructed over these two regions with the help of the Gaussian signals' parameters (means, variances and strengths) estimated using expectation maximization (EM) for both the Gaussian signals. The EM algorithm estimates the parameters of the multivariate probability density function in the form of a Gaussian mixture distribution with a specified number of mixtures. Finally, the GMM initialization for EM algorithm is done in the following way. The strengths w_1 and w_2 of both the Gaussian signals are given as:

$$w_1^0 = \frac{n_1}{n_1 + n_2} \quad \text{and} \quad w_2^0 = \frac{n_2}{n_1 + n_2}, \tag{12}$$

where $n_1 = \sum_{p \in P} \text{RSSM}(p)$ and $n_2 = \sum_{p \in P} (1 - \text{RSSM}(p))$ and P denotes the set of image pixels. Since RSSM segments the two regions spatially, their initial spatial means μ_1^0 and μ_2^0 are given as $\mathbf{I}(U_1^0)$ and $\mathbf{I}(U_2^0)$, respectively, where

$$U_1^0 = \frac{1}{n_1} \sum_{\mathbf{U} \in \text{SC}_{xy}} \mathbf{U} \cdot \text{RSSM}(\mathbf{U}) \quad \text{and} \quad U_2^0 = \frac{1}{n_2} \sum_{\mathbf{U} \in \text{SC}_{xy}} \mathbf{U} \cdot (1 - \text{RSSM}(\mathbf{U})), \tag{13}$$

where SC_{xy} are image spatial coordinates. Their covariances Σ_1 and Σ_2 are given as:

$$\begin{aligned} \Sigma_1^0 &= \frac{1}{n_1 - 1} \sum_{p \in P} (\mathbf{I}(p) - \mu_1^0) \cdot (\mathbf{I}(p) - \mu_1^0)^T \cdot \text{RSSM}(p), \\ \Sigma_2^0 &= \frac{1}{n_2 - 1} \sum_{p \in P} (\mathbf{I}(p) - \mu_2^0) \cdot (\mathbf{I}(p) - \mu_2^0)^T \cdot (1 - \text{RSSM}(p)). \end{aligned} \tag{14}$$

After initialization step, GMM parameters are updated using EM algorithm until convergence is reached. The probability of a pixel p to be a part of either of the cluster by utilizing the current parameters of the l th iteration is calculated as:

$$\begin{aligned} \text{Prob}^l(1|\mathbf{I}(p)) &= \frac{w_1^l \mathcal{N}(\mathbf{I}(p) | \mu_1^l, \Sigma_1^l)}{w_1^l \mathcal{N}(\mathbf{I}(p) | \mu_1^l, \Sigma_1^l) + w_2^l \mathcal{N}(\mathbf{I}(p) | \mu_2^l, \Sigma_2^l)}, \\ \text{Prob}^l(2|\mathbf{I}(p)) &= \frac{w_2^l \mathcal{N}(\mathbf{I}(p) | \mu_2^l, \Sigma_2^l)}{w_1^l \mathcal{N}(\mathbf{I}(p) | \mu_1^l, \Sigma_1^l) + w_2^l \mathcal{N}(\mathbf{I}(p) | \mu_2^l, \Sigma_2^l)}. \end{aligned} \tag{15}$$

The parameters of both Gaussian signals are updated in the following manner:

$$\begin{aligned} w_1^{l+1} &= \frac{1}{W \times H} \sum_{p \in P} \text{Prob}^l(1|\mathbf{I}(p)) \quad \text{and} \quad w_2^{l+1} = \frac{1}{W \times H} \sum_{p \in P} \text{Prob}^l(2|\mathbf{I}(p)), \\ \mu_1^{l+1} &= \frac{\sum_{p \in P} \text{Prob}^l(1|\mathbf{I}(p)) \cdot \mathbf{I}(p)}{\sum_{p \in P} \text{Prob}^l(1|\mathbf{I}(p))} \quad \text{and} \quad \mu_2^{l+1} = \frac{\sum_{p \in P} \text{Prob}^l(2|\mathbf{I}(p)) \cdot \mathbf{I}(p)}{\sum_{p \in P} \text{Prob}^l(2|\mathbf{I}(p))}, \\ \Sigma_1^{l+1} &= \frac{\sum_{p \in P} \text{Prob}^l(1|\mathbf{I}(p)) \cdot (\mathbf{I}(p) - \mu_1^l) \cdot (\mathbf{I}(p) - \mu_1^l)^T}{\sum_{p \in P} \text{Prob}^l(1|\mathbf{I}(p))} \end{aligned}$$

and

$$\Sigma_2^{l+1} = \frac{\sum_{p \in P} Prob^l(2|\mathbf{I}(p)) \cdot (\mathbf{I}(p) - \mu_2^l) \cdot (\mathbf{I}(p) - \mu_2^l)^T}{\sum_{p \in P} Prob^l(2|\mathbf{I}(p))}, \tag{16}$$

where W denotes the width and H denotes the height of the image. The log-likelihood for $l + 1$ iteration is computed as:

$$loglikelihood^{l+1} = \sum_{p \in P} \left(\log \left(w_1^{l+1} \cdot \mathcal{N}(\mathbf{I}(p) | \mu_1^{l+1}, \Sigma_1^{l+1}) + w_2^{l+1} \cdot \mathcal{N}(\mathbf{I}(p) | \mu_2^{l+1}, \Sigma_2^{l+1}) \right) \right). \tag{17}$$

The inequality for the convergence condition is given as:

$$abs(loglikelihood^{l+1} - loglikelihood^l) < 1.0e - 3. \tag{18}$$

On completion of the updating procedure, the final parameter values of the GMM are used to assign each and every pixel p to the two regions (salient or background) with a probability, which is given as:

$$Prob^{final}(1|\mathbf{I}(p)) = \frac{w_1 \mathcal{N}(\mathbf{I}(p) | \mu_1, \Sigma_1)}{w_1 \mathcal{N}(\mathbf{I}(p) | \mu_1, \Sigma_1) + w_2 \mathcal{N}(\mathbf{I}(p) | \mu_2, \Sigma_2)},$$

$$Prob^{final}(2|\mathbf{I}(p)) = \frac{w_2 \mathcal{N}(\mathbf{I}(p) | \mu_2, \Sigma_2)}{w_1 \mathcal{N}(\mathbf{I}(p) | \mu_1, \Sigma_1) + w_2 \mathcal{N}(\mathbf{I}(p) | \mu_2, \Sigma_2)}. \tag{19}$$

The two segments are weighted based on a centre prior computed as:

$$Center_prior(i) = \sum_{p \in P} Prob^{final}(i|\mathbf{I}(p)) \cdot sdist_p; i = 1, 2, \tag{20}$$

where $sdist_p$ is the spatial distance of the p th pixel with the image centre. $Center_prior(i)$ is normalized between $[0, 1)$, which is computed as

$$Center_prior(i) = \frac{Center_prior(i) - \min\{Center_prior(j)\}_{j=1}^2}{\max\{Center_prior(j)\}_{j=1}^2}. \tag{21}$$

The saliency value of a pixel p is given as:

$$SM(p) = \sum_{i=1}^2 Prob^{final}(i|\mathbf{I}(p)) \cdot (1 - Center_prior(i)), \tag{22}$$

where SM represents the final saliency map generated by the proposed algorithm, which is normalized in the range 0–1.

4 Experimental set-up and results

In this section, we evaluate and compare the performances of our method (DDDTCWT-SS) against 29 state-of-the-art algorithms on six representative benchmark datasets.

Salient object datasets and state-of-the-art methods

Six benchmark datasets for evaluation include commonly used Microsoft Research Asia Salient Object Database¹ (MSRA SOD) image set B (5000 images), Achanta Saliency Database (ASD²) (1000 images), SAA_GT³ (5000 images), SOD⁴ (500 images), SED which consists of two parts, i.e. SED1⁵ (one object set) and SED2⁶ (two object sets) each containing 100 images. All the images are of size 400×300 or 300×400 having intensity values in [0255]. We compare the proposed method (DDDCWT-SS) with 29 state-of-the-art salient object detection methods: IT [44], AIM [15], GBVS [35], SR [36], SLRG [64, 65], SUN [94], FT [3], ASS [2], CASD [30], LRK [80], SIM [68], WT [41], PFT [32], STREE [63], MSA [96], PCT [92], FDN [8, 9], PFDN [8, 9], PQFT [33], AQFT [24], BS [25], HFT [60], SDS [61], HLGm [5], IS [37], BSM [90], MCA [86], HDCT [51] and BHGT [6].

Experimental Set-up All the experiments are carried out using Windows 7 environment over Intel(R) Xeon(R) processor with a speed of 2.27 GHz and 4 GB RAM.

4.1 Qualitative performance

The qualitative analysis of our algorithm (DDDCWT-SS) with 29 other state-of-the-art recently proposed saliency models is presented in Fig. 5.

We include these models based on relevance to our work, recency and availability of their saliency maps. We randomly choose three images from MSRA dataset, two images from SOD, two images from SED1 and two images from SED2 datasets for qualitative comparison. Figure 5 clearly shows that the better saliency maps are achieved by the proposed model (DDDCWT-SS) in comparison with 29 state-of-the-art methods (Fig. 6).

4.2 Quantitative performance

We evaluate the quantitative performance of the proposed model (DDDCWT-SS) against 29 other state-of-the-art models in terms of precision, recall, F -measure, area under curve (AUC) and computation time. Performance of the proposed model (DDDCWT-SS) is also evaluated in terms of mean precision value, mean recall value, mean F -measure values and mean AUC values on all six datasets. The outcome of the salient object detection procedure is a saliency map. To compare the quality of saliency maps in terms of precision, recall, F -measure and area under curve for the task of segmenting salient objects, we rely on a ground truth database. Therefore, a suitable threshold t [5, 82] is first applied to the saliency map of an image \mathbf{I} to generate an attention mask \mathbf{R} (also called ‘predicted condition’ or detection result) and then \mathbf{R} is compared with the ground truth \mathbf{G} (also called ‘true condition’) associated with \mathbf{I} . Both \mathbf{R} and \mathbf{G} consist of pixels with only two values 0 or 1. Based on the values of \mathbf{R} and \mathbf{G} , the following terms are defined:

¹ http://www.research.microsoft.com/enus/um/people/jiansun/salientobject/salient_object.htm.

² http://ivrgwww.epfl.ch/supplementary_material/RK_CVPR09/GroundTruth/binarymasks.zip.

³ E-mail at “rinki.arya89@gmail.com” or “navjot.singh.09@gmail.com”.

⁴ <http://elderlab.yorku.ca/~vida/SOD/>.

⁵ http://www.wisdom.weizmann.ac.il/~vision/Seg_Evaluation_DB.

⁶ http://www.wisdom.weizmann.ac.il/~vision/Seg_Evaluation_DB.

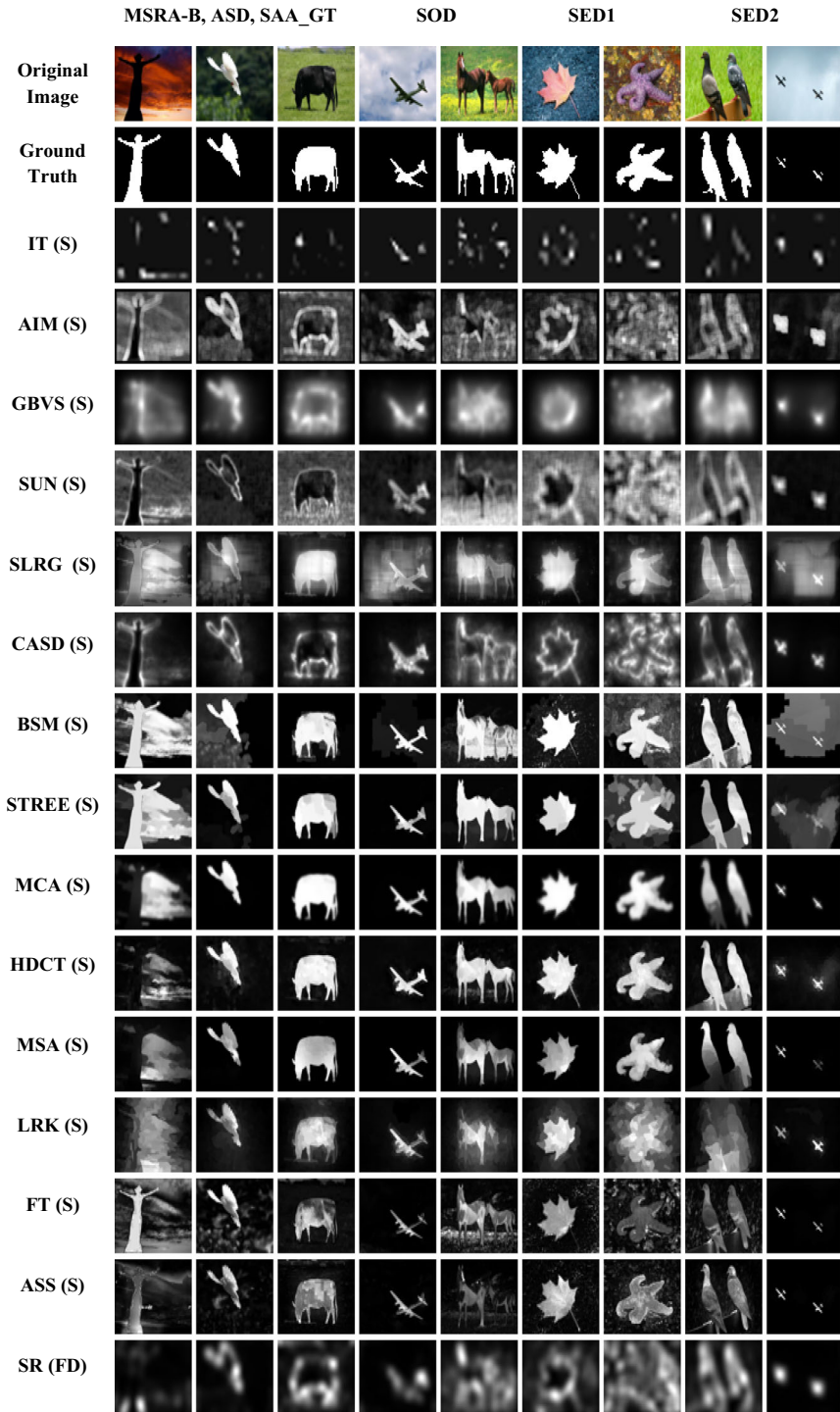


Fig. 5 Qualitative comparison of the proposed model with existing 29 models. *S* spatial domain, *FD* frequency domain, *W* wavelet domain

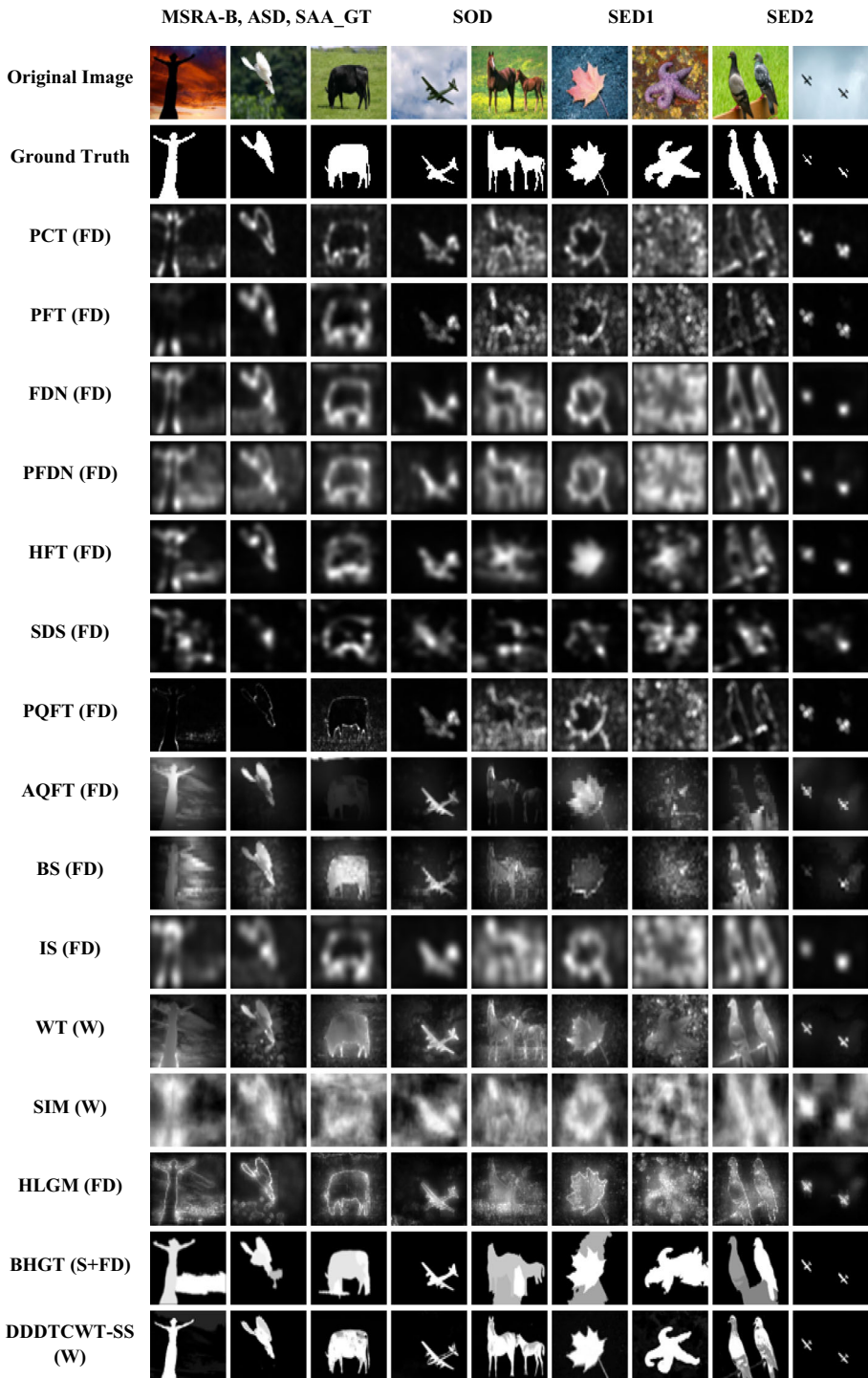


Fig. 5 continued

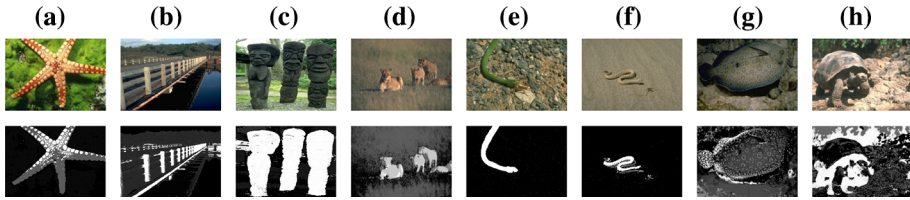


Fig. 6 Saliency maps generated by the proposed algorithm for some challenging and complex images

- The true positive (TP) is the number of pixels correctly detected in **R** as belonging to the salient object in ground truth image **G**.
- False positive (FP) is the number of pixels wrongly detected as salient in **R** as belonging to the background in the ground truth image **G**.
- False negative (FN) is the number of pixels incorrectly detected as background in **R** as belonging to the salient object in the ground truth image **G**.
- True negatives (TN) are the pixels correctly detected as background in **R** as belonging to the background in the ground truth image **G**.

The obtained attention mask **R** and ground truth map **G** are used to compute the precision, recall and *F*-measure as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad F_\beta = \frac{(1 + \beta^2) \times \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}}, \quad (23)$$

where $\text{TP} = \sum_{G(x,y)=1} R(x,y)$ $\text{FP} = \sum_{G(x,y)=0} R(x,y)$ $\text{FN} = \sum_{R(x,y)=0} G(x,y)$.

β is chosen to be 1 to give equal importance to both precision and recall. ROC curve is used for measuring the similarity between the saliency map and ground truth, and the area under curve (AUC) is used for quantitative comparison between different models. The ROC curve is generated by plotting the true positive rate (TPR) on the y-axis against false positive rate (FPR) values on the x-axis, respectively [82]. TPR and FPR are given by

$$\text{TPR} = \frac{\text{TP}}{\sum_{(x,y)} G(x,y)} \quad \text{FPR} = \frac{\text{FP}}{W \times H - \sum_{(x,y)} G(x,y)}, \quad (24)$$

where *W* and *H* represent the width and height of the image, respectively. A model is considered to be good if it achieves high values for precision, recall, *F*-measure and AUC. Tables 1, 2, 3, 4 and 5 show the quantitative performance analysis of the proposed method in comparison with other 29 state-of-the-art methods on all the six datasets in terms of precision, recall, *F*-measure, AUC and average computation time per image, respectively. The mean value for each quantitative measure on all six datasets is also shown in the respective table for each model.

In Fig. 7, we show the ROC curves of the state-of-the-art algorithms, including the proposed method (DDDCWT-SS) corresponding to the six datasets. On the basis of Fig. 7 and Tables 1, 2, 3, 4 and 5 (the best results are shown in bold), we make the following observation:

1. *Quantitative evaluation on MSRA dataset* The proposed model (DDDCWT-SS) achieves highest *F*-measure for MSRA dataset. The proposed model DDDTCWT-SS outperforms all state-of-the-art methods except HFT, SDS and MCA in terms of precision. DDDTCWT-SS outperforms all state-of-the-art methods except SIM and PFDN in

Table 1 Quantitative comparison on six datasets in terms of precision

Model	MSRA-B	ASD	SAA_GT	SOD	SED1	SED2	Mean
IT (S)	0.672	0.550	0.545	0.685	0.720	0.676	0.6083
AIM (S)	0.728	0.535	0.527	0.412	0.562	0.470	0.6085
GBVS (S)	0.800	0.666	0.658	0.520	0.695	0.542	0.7128
SUN (S)	0.598	0.542	0.668	0.379	0.561	0.417	0.6119
SLRG (S)	0.674	0.700	0.763	0.423	0.589	0.431	0.7007
CASD (S)	0.712	0.697	0.679	0.492	0.659	0.551	0.6854
BSM (S)	0.724	0.798	0.768	0.621	0.732	0.525	0.7431
STREE (S)	0.831	0.861	0.830	0.627	0.849	0.773	0.8241
MCA (S)	0.893	0.856	0.922	0.897	0.851	0.624	0.8997
HDCT (S)	0.844	0.870	0.924	0.648	0.857	0.805	0.8718
MSA (S)	0.833	0.859	0.811	0.689	0.934	0.893	0.8210
LRK (S)	0.703	0.716	0.680	0.476	0.658	0.590	0.6832
FT (S)	0.717	0.599	0.800	0.608	0.735	0.830	0.7388
ASS (S)	0.786	0.635	0.801	0.655	0.817	0.757	0.7739
SR (FD)	0.761	0.502	0.588	0.479	0.614	0.504	0.6494
PCT (FD)	0.797	0.626	0.617	0.483	0.644	0.538	0.6885
PFT (FD)	0.764	0.578	0.583	0.505	0.655	0.567	0.6571
FDN (FD)	0.663	0.573	0.550	0.398	0.579	0.503	0.5936
PFDN (FD)	0.593	0.508	0.495	0.371	0.586	0.439	0.5330
HFT (FD)	0.892	0.776	0.755	0.579	0.784	0.573	0.8065
SDS (FD)	0.872	0.725	0.722	0.560	0.758	0.494	0.7778
PQFT (FD)	0.815	0.694	0.677	0.502	0.664	0.550	0.7288
AQFT (FD)	0.671	0.638	0.593	0.446	0.579	0.578	0.6236
BS (FD)	0.767	0.746	0.703	0.509	0.719	0.571	0.7247
IS (FD)	0.741	0.596	0.567	0.422	0.616	0.496	0.6375
WT (W)	0.662	0.606	0.612	0.451	0.622	0.575	0.6257
SIM (W)	0.486	0.395	0.408	0.338	0.447	0.331	0.4369
HLGM (FD)	0.686	0.657	0.624	0.446	0.609	0.584	0.6452
BHGT (S + FD)	0.769	0.780	0.796	0.500	0.808	0.817	0.7707
DDTCWT-SS (W)	0.857	0.845	0.861	0.804	0.822	0.783	0.8545

S spatial domain, *FD* frequency domain, *W* wavelet domain

terms of recall. DDDTCWT-SS outperforms all state-of-the-art methods except BHGT in terms of AUC.

2. *Quantitative evaluation on ASD dataset* In terms of precision, DDDTCWT-SS outperforms all state-of-the-art methods except MSA, HDCT, MCA and STREE. In terms of recall, DDDTCWT-SS outperforms all state-of-the-art methods. In terms of *F*-measure, DDDTCWT-SS outperforms all state-of-the-art methods. In terms of AUC value DDDTCWT-SS outperforms all state-of-the-art methods except MCA and HDCT.
3. *Quantitative evaluation on SAA dataset* In terms of precision, DDDTCWT-SS outperforms all state-of-the-art methods except HDCT and MCA. In terms of recall and

Table 2 Quantitative comparison on six datasets in terms of recall

MODEL	MSRA-B	ASD	SAA_GT	SOD	SED1	SED2	Mean
IT (S)	0.614	0.695	0.609	0.154	0.150	0.297	0.5925
AIM (S)	0.762	0.859	0.777	0.625	0.790	0.816	0.7715
GBVS (S)	0.692	0.634	0.612	0.584	0.597	0.600	0.6466
SUN (S)	0.857	0.848	0.764	0.431	0.611	0.659	0.7945
SLRG (S)	0.889	0.921	0.895	0.737	0.806	0.803	0.8864
CASD (S)	0.763	0.782	0.726	0.518	0.496	0.559	0.7343
BSM (S)	0.816	0.822	0.711	0.724	0.816	0.759	0.7672
STREE (S)	0.888	0.913	0.854	0.699	0.807	0.802	0.8661
MCA (S)	0.780	0.744	0.839	0.744	0.720	0.639	0.7989
HDCT (S)	0.885	0.901	0.661	0.626	0.738	0.740	0.7771
MSA (S)	0.661	0.674	0.585	0.477	0.590	0.614	0.6208
LRK (S)	0.907	0.903	0.841	0.693	0.771	0.790	0.8671
FT (S)	0.575	0.606	0.517	0.300	0.347	0.533	0.5388
ASS (S)	0.704	0.670	0.524	0.366	0.452	0.589	0.6066
SR (FD)	0.526	0.440	0.372	0.336	0.360	0.450	0.4426
PCT (FD)	0.579	0.403	0.382	0.360	0.376	0.422	0.4673
PFT (FD)	0.528	0.379	0.371	0.287	0.311	0.397	0.4349
FDN (FD)	0.899	0.831	0.774	0.714	0.733	0.758	0.8292
PFDN (FD)	0.944	0.887	0.857	0.831	0.849	0.792	0.8950
HFT (FD)	0.580	0.535	0.490	0.453	0.461	0.477	0.5304
SDS (FD)	0.366	0.319	0.272	0.274	0.343	0.256	0.3167
PQFT (FD)	0.245	0.117	0.114	0.329	0.328	0.416	0.1838
AQFT (FD)	0.675	0.636	0.544	0.446	0.465	0.600	0.6035
BS (FD)	0.783	0.757	0.696	0.623	0.603	0.617	0.7338
IS (FD)	0.692	0.585	0.556	0.585	0.515	0.559	0.6175
WT (W)	0.84	0.801	0.702	0.564	0.608	0.720	0.7629
SIM (W)	0.947	0.825	0.793	0.733	0.765	0.794	0.8588
HLGM (FD)	0.877	0.734	0.669	0.598	0.630	0.741	0.7607
BHGT (S + FD)	0.873	0.889	0.882	0.710	0.780	0.857	0.8703
DDTCWT-SS (W)	0.913	0.948	0.928	0.832	0.871	0.824	0.9178

S spatial domain, *FD* frequency domain, *W* wavelet domain

F-measure, DDDTCWT-SS outperforms all state-of-the-art methods. In terms of AUC value, DDDTCWT-SS outperforms all state-of-the-art methods except HDCT.

4. *Quantitative evaluation on SOD dataset* In terms of precision, DDDTCWT-SS outperforms all state-of-the-art methods except MCA. In terms of recall, *F*-measure and AUC, DDDTCWT-SS outperforms all state-of-the-art methods.
5. *Quantitative evaluation on SEDI dataset* In terms of precision, DDDTCWT-SS outperforms all state-of-the-art methods except MSA, HDCT, MCA and STREE. In terms of recall and *F*-measure, DDDTCWT-SS outperforms all state-of-the-art methods. In terms of AUC value, DDDTCWT-SS outperforms all state-of-the-art methods except BHGT, HDCT, MCA, STREE and BSM.

Table 3 Quantitative comparison on six datasets in terms of F -measure

MODEL	MSRA-B	ASD	SAA_GT	SOD	SED1	SED2	MEAN
IT (S)	0.642	0.614	0.575	0.252	0.248	0.413	0.5890
AIM (S)	0.745	0.659	0.628	0.500	0.657	0.597	0.6752
GBVS (S)	0.742	0.650	0.634	0.550	0.642	0.570	0.6775
SUN (S)	0.704	0.661	0.713	0.403	0.585	0.511	0.6886
SLRG (S)	0.767	0.795	0.824	0.538	0.681	0.561	0.7815
CASD (S)	0.737	0.737	0.702	0.505	0.566	0.555	0.7091
BSM (S)	0.767	0.810	0.738	0.669	0.772	0.621	0.7529
STREE (S)	0.859	0.886	0.842	0.661	0.828	0.787	0.8447
MCA (S)	0.833	0.796	0.879	0.813	0.780	0.631	0.8465
HDCT (S)	0.864	0.885	0.771	0.637	0.793	0.771	0.8149
MSA (S)	0.737	0.755	0.680	0.564	0.723	0.728	0.7066
LRK (S)	0.792	0.799	0.752	0.564	0.710	0.676	0.7641
FT (S)	0.638	0.603	0.628	0.402	0.471	0.649	0.6193
ASS (S)	0.743	0.652	0.634	0.470	0.580	0.663	0.6749
SR (FD)	0.622	0.469	0.456	0.395	0.454	0.476	0.5256
PCT (FD)	0.671	0.489	0.472	0.413	0.475	0.473	0.5560
PFT (FD)	0.625	0.458	0.454	0.366	0.422	0.467	0.5235
FDN (FD)	0.763	0.678	0.643	0.511	0.647	0.605	0.6913
PFDN (FD)	0.728	0.646	0.628	0.513	0.693	0.565	0.6674
HFT (FD)	0.702	0.633	0.594	0.508	0.581	0.521	0.6391
SDS (FD)	0.516	0.443	0.400	0.368	0.472	0.337	0.4520
PQFT (FD)	0.377	0.201	0.195	0.397	0.439	0.477	0.2864
AQFT (FD)	0.673	0.637	0.568	0.446	0.516	0.589	0.6133
BS (FD)	0.775	0.752	0.699	0.560	0.656	0.593	0.7288
IS (FD)	0.716	0.590	0.561	0.490	0.561	0.526	0.6264
WT (W)	0.741	0.690	0.654	0.501	0.615	0.639	0.6873
SIM (W)	0.642	0.534	0.539	0.463	0.564	0.467	0.5789
HLGM (FD)	0.770	0.693	0.646	0.511	0.619	0.653	0.6971
BHGT (S+FD)	0.818	0.830	0.837	0.587	0.794	0.837	0.8172
DDTCWT-SS (W)	0.884	0.895	0.893	0.818	0.846	0.803	0.8849

S spatial domain, FD frequency domain, W wavelet domain

6. *Quantitative evaluation on SED2 dataset* In terms of precision, DDDTCWT-SS outperforms all state-of-the-art methods except BHGT, FT, MSA and HDCT. In terms of recall and F -measure, DDDTCWT-SS outperforms all state-of-the-art methods except BHGT. In terms of AUC, DDDTCWT-SS outperforms all state-of-the-art methods.
7. Some models are better in terms of precision and others in terms of recall. A model is considered to be good if both precision and recall are higher. But that is difficult to achieve. It is suggested that a model should have a higher F -measure value which is the weighted harmonic mean of precision and recall. The proposed model furnishes the highest F -measure value on all the six datasets.

Table 4 Quantitative comparison on six datasets in terms of AUC

MODEL	MSRA-B	ASD	SAA_GT	SOD	SED1	SED2	Mean
IT (S)	0.663	0.529	0.590	0.550	0.623	0.601	0.6146
AIM (S)	0.705	0.631	0.673	0.796	0.880	0.861	0.6917
GBVS (S)	0.698	0.579	0.636	0.813	0.868	0.821	0.6688
SUN (S)	0.681	0.602	0.641	0.716	0.851	0.776	0.6609
SLRG (S)	0.802	0.733	0.767	0.796	0.868	0.812	0.7815
CASD (S)	0.776	0.705	0.741	0.791	0.833	0.813	0.7564
BSM (S)	0.756	0.635	0.686	0.817	0.898	0.828	0.7202
STREE (S)	0.818	0.831	0.853	0.806	0.889	0.865	0.8346
MCA (S)	0.819	0.972	0.886	0.818	0.888	0.809	0.8612
HDCT (S)	0.842	0.975	0.929	0.819	0.920	0.831	0.8901
MSA (S)	0.782	0.849	0.826	0.729	0.869	0.709	0.8044
LRK (S)	0.783	0.713	0.753	0.794	0.860	0.814	0.7656
FT (S)	0.669	0.625	0.648	0.595	0.650	0.676	0.6530
ASS (S)	0.698	0.630	0.664	0.790	0.840	0.797	0.6836
SR (FD)	0.658	0.505	0.581	0.732	0.780	0.796	0.6174
PCT (FD)	0.737	0.777	0.746	0.736	0.807	0.781	0.7452
PFT (FD)	0.706	0.731	0.723	0.714	0.753	0.794	0.7169
FDN (FD)	0.782	0.826	0.795	0.775	0.838	0.795	0.7916
PFDN (FD)	0.83	0.856	0.833	0.801	0.873	0.767	0.8321
HFT (FD)	0.858	0.827	0.873	0.803	0.879	0.731	0.8585
SDS (FD)	0.734	0.779	0.753	0.725	0.817	0.683	0.7459
PQFT (FD)	0.540	0.538	0.530	0.729	0.781	0.771	0.5477
AQFT (FD)	0.728	0.807	0.764	0.702	0.733	0.729	0.7491
BS (FD)	0.849	0.853	0.834	0.799	0.829	0.766	0.8399
IS (FD)	0.778	0.849	0.833	0.766	0.832	0.743	0.8072
WT (W)	0.743	0.693	0.718	0.785	0.824	0.817	0.7312
SIM (W)	0.768	0.837	0.835	0.783	0.822	0.806	0.8040
HLGM (FD)	0.860	0.848	0.818	0.773	0.829	0.838	0.8369
BHGT (S + FD)	0.873	0.892	0.878	0.815	0.893	0.853	0.8743
DDDTCWT-SS (W)	0.867	0.930	0.913	0.895	0.886	0.877	0.8935

S spatial domain, *FD* frequency domain, *W* wavelet domain

8. Our model covers the maximum area under the ROC curve in comparison with all state-of-the-art methods for SOD and SED2 datasets and hence gives the highest AUC value on these two datasets.
9. PFT [32] takes the least computation time.
10. Spatial domain models provide good detection accuracy at the cost of high computational time while frequency domain models offer fast computational speed to meet real-time requirements at the cost of poor detection accuracy. In order to induce a trade-off between computational time and accuracy, our model provides high detection accuracy and takes less time in comparison with most of the existing methods in spatial domain, which is given in Table 5.

Table 5 Quantitative comparison in terms of computation time

Model	Average time (in sec) per image
IT (S)	1.70
AIM (S)	50.8
GBVS (S)	59.8
SUN (S)	3.64
SLRG (S)	25.7
CASD (S)	124.0
BSM (S)	183.50
STREE (S)	104.9
MCA (S)	5.25
HDCT (S)	14.85
MSA (S)	3.51
LRK (S)	71.9
FT (S)	0.17
ASS (S)	0.31
SR (FD)	0.02
PCT (FD)	0.06
PFT (FD)	0.018
FDN (FD)	0.29
PFDN (FD)	0.25
HFT (FD)	0.69
SDS (FD)	0.17
PQFT (FD)	0.18
AQFT (FD)	6.4
BS (FD)	7.57
IS (FD)	0.04
WT (W)	6.55
SIM (W)	4.06
HLGM (FD)	0.26
BHGT (S+FD)	5.98
DDDCWT-SS (W)	9.34

S spatial domain, *FD* frequency domain, *W* wavelet domain

11. Bipartite graph-based image segmentation and DDDTCWT-based saliency map computation steps are independent from each other. Therefore, these steps are carried out in parallel during experiments. This helps us in reducing overall execution time of the proposed model.
12. In terms of mean precision value, DDDTCWT-SS outperforms all state-of-the-art methods except HDCT and MCA. In terms of mean recall value, mean *F*-measure value and mean AUC value, DDDTCWT-SS outperforms all state-of-the-art methods.
13. We have used a threshold which is equivalent to average intensity value of all pixels computed using multi-scale edge-to-texture features extracted from 2D-DDDCWT coefficients to get a rough estimation of salient region and background region in initial rough segmented saliency map (RSSM). We need to choose the threshold more carefully as we can lose too much of the salient pixels and sometimes get too many extraneous background pixels while thresholding. To assign appropriate salient and background

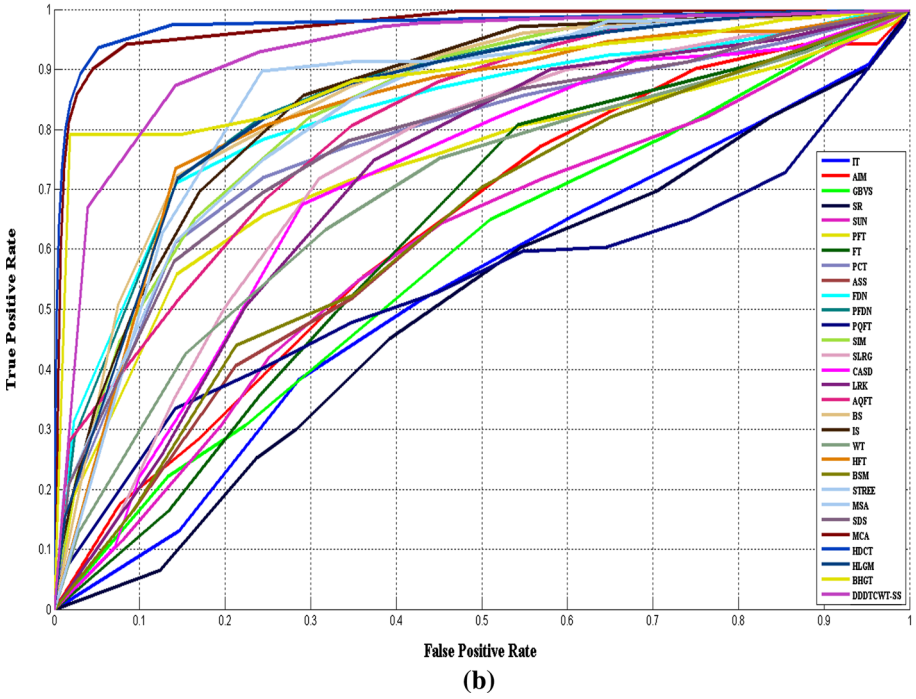
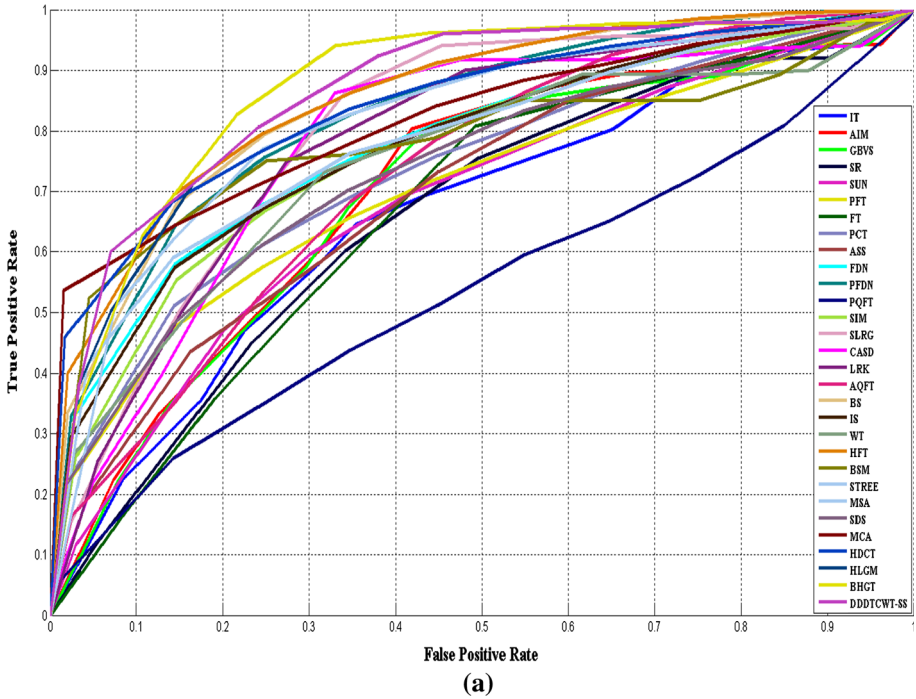
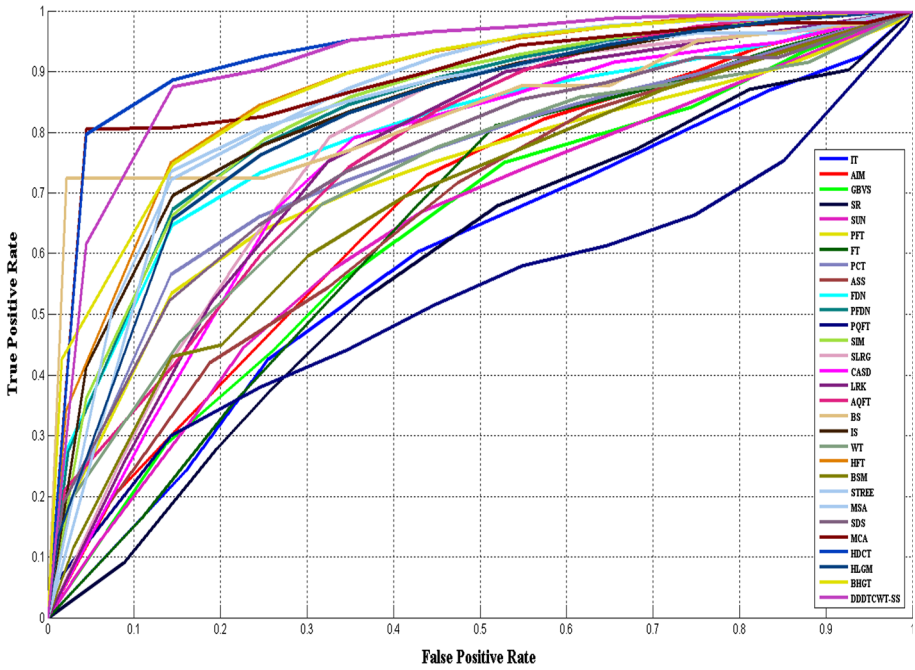
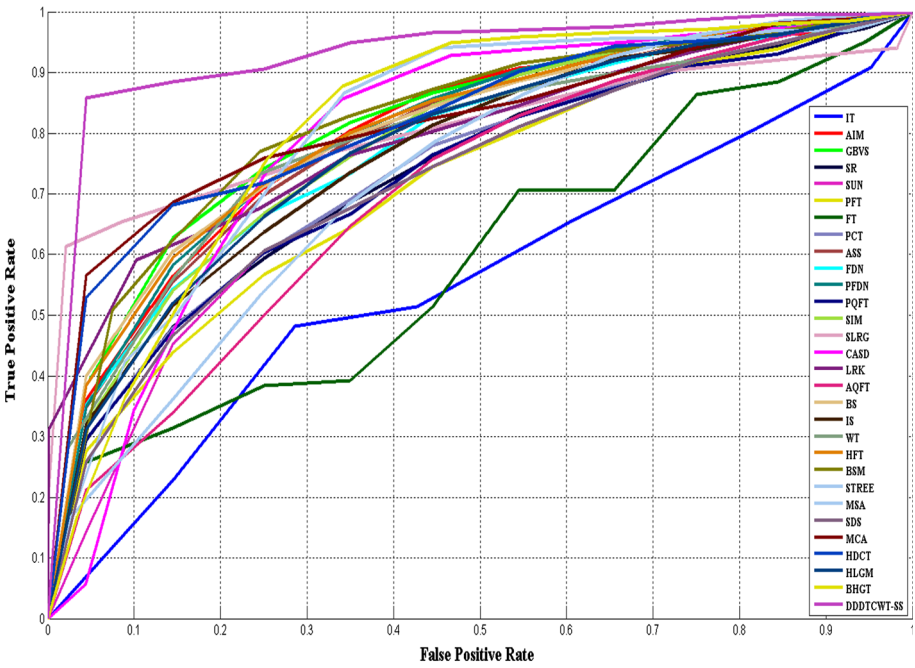


Fig. 7 ROC for the six datasets a MSRA-B, b ASD, c SAA_GT, d SOD, e SED1, f SED2



(c)



(d)

Fig. 7 continued

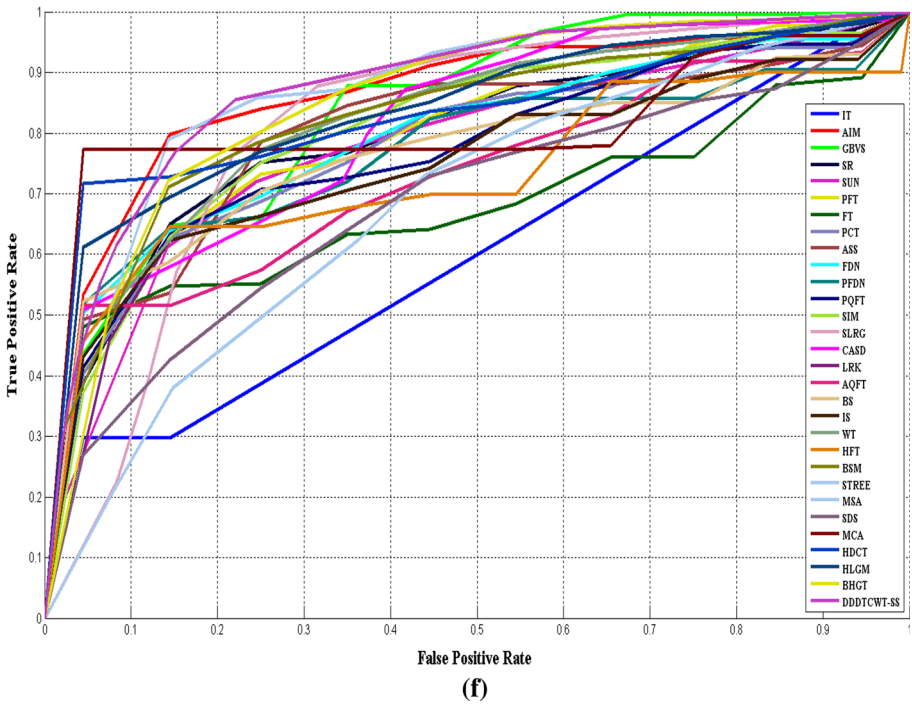
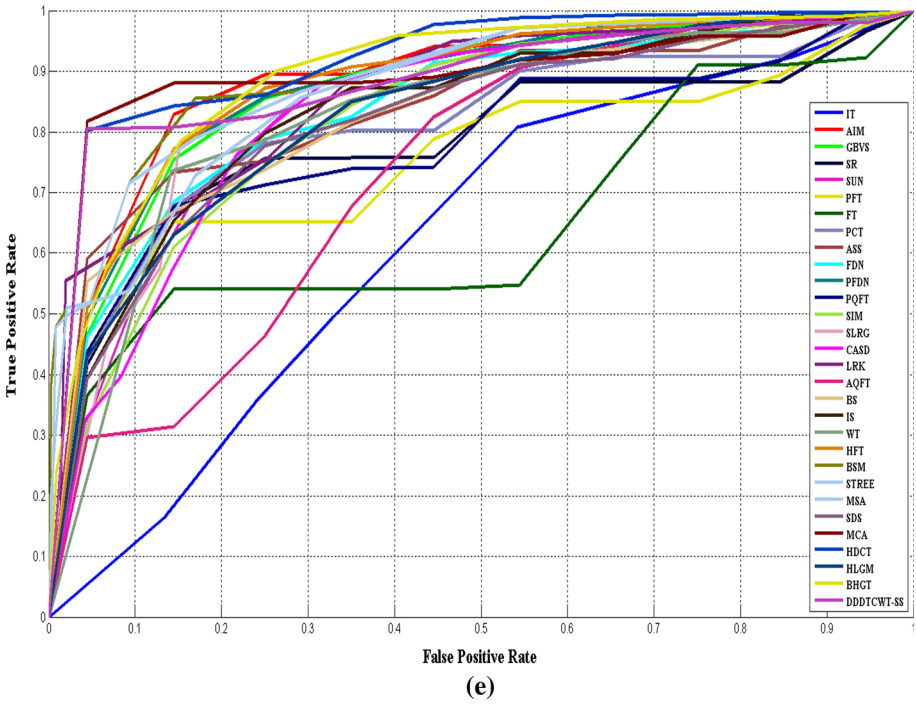


Fig. 7 continued

Table 6 Average ranking of the algorithms corresponding to F -measure and AUC using Friedman statistic

Algorithm	Ranking for F -measure	Ranking for AUC
IT (S)	26.25	29.17
AIM (S)	15.00	15.08
GBVS (S)	14.92	17.33
SUN (S)	18.00	23.00
SLRG (S)	10.42	12.67
CASD (S)	15.58	16.00
BSM (S)	7.25	12.33
STREE (S)	2.83	6.50
MCA (S)	5.17	5.67
HDCT (S)	4.00	2.83
MSA (S)	9.50	15.25
LRK (S)	6.75	13.75
FT (S)	21.33	27.33
ASS (S)	15.42	19.58
SR (FD)	26.83	25.33
PCT (FD)	24.67	19.50
PFT (FD)	28.08	22.00
FDN (FD)	13.25	14.92
PFDN (FD)	15.83	10.25
HFT (FD)	19.67	10.33
SDS (FD)	28.33	21.25
PQFT (FD)	28.00	26.25
AQFT (FD)	21.00	22.00
BS (FD)	10.67	11.92
IS (FD)	21.50	15.50
WT (W)	14.00	17.67
SIM (W)	24.00	14.33
HLGM (FD)	11.75	11.25
BHGT (S+FD)	3.83	3.50
DDDTTCWT-SS (W)	1.17	2.50

pixels to two different Gaussian kernels for refinement process, we have chosen more reliable way based on DDDTCWT coefficients.

14. To show the statistical significance of performance results, Friedman statistical test is performed on F -measure and AUC performance measures, which is based on the research work of Demšar [22] and Derrac et al. [23]. The null hypothesis assumes that each of the models is equivalent in terms of their performance. A comparison of multiple models can be accomplished after ranking them according to their F -measure and AUC values. For each case, rank ranging from 1 to k is associated with every model. Rank values 1 and k denote the best and worst result, respectively. Let this rank be denoted by r_i^j ($1 \leq i \leq N$, $1 \leq j \leq k$). For each model, j , let R^j denote the average of ranks over the N experimental observations. The ranks computed are given in Table 6 for 30 models for both performance measures. Table 6 shows that the best performing

Table 7 Quantitative performance of the proposed algorithm before and after the refinement step

Dataset	Precision		Recall		<i>F</i> -measure		AUC	
	Before Refinement	After Refinement	Before Refinement	After Refinement	Before Refinement	After Refinement	Before Refinement	After Refinement
MSRA	0.685	0.857	0.895	0.913	0.776	0.884	0.7796	0.867
ASD	0.718	0.845	0.928	0.948	0.809	0.895	0.8920	0.930
SAA	0.609	0.861	0.917	0.928	0.732	0.893	0.8225	0.913
SOD	0.477	0.804	0.743	0.832	0.581	0.818	0.8132	0.895
SEDI	0.586	0.822	0.759	0.871	0.661	0.846	0.8271	0.886
SED2	0.662	0.783	0.823	0.824	0.759	0.803	0.7963	0.877

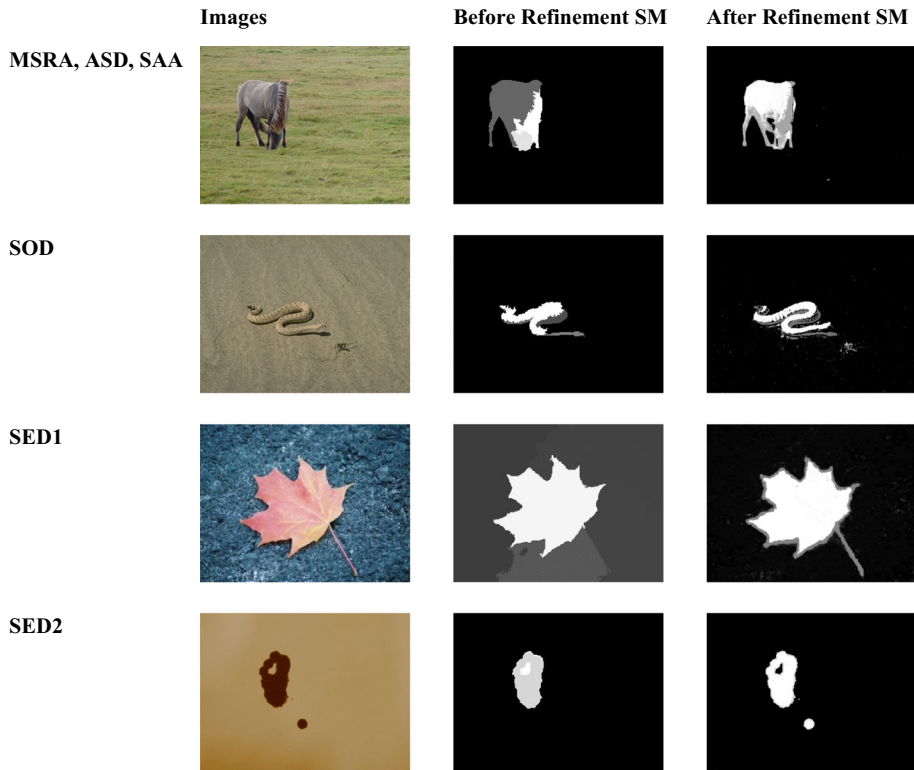


Fig. 8 Qualitative performance of the proposed algorithm before and after the refinement step

DDDCWT-SS algorithm has the least rank value for both F -measure (1.17) and AUC (2.50). The p values computed by Iman and Davenport [42] statistic for F -measure and AUC are $4.0622267404084276E-55$ and $-2.2204454167762003E-16$, respectively, which suggest the statistical difference among all models considered in our quantitative comparison, hence rejecting the null hypothesis.

15. We have used six datasets containing images with wide range of shapes, scales and appearances. The proposed method produces more accurate saliency maps in various challenging cases, e.g. salient object touching the image border (columns 6(a) and 6(b)), multiple disconnected salient objects (columns 6(c) and 6(d)), low contrast between salient object and background (column 6(e)) and Camouflage condition (column 6(f)), while the proposed algorithm may not work properly for some images containing multiple objects on highly cluttered backgrounds (columns 6(g) and 6(h)). It also may not provide better performance on images with salient objects under partial occlusion.
16. To show contribution of the refinement step using Gaussian mixture model (GMM) and expectation maximization (EM) algorithm in the proposed algorithm, quantitative and qualitative performance of the algorithm before and after the refinement step is shown in Table 7 and Fig. 8, respectively. Table 7 (shown in bold) and Fig. 8 show that the performance after refinement step is improved.

5 Conclusion and future work

Performance assessment of the proposed model is done in terms of precision, recall, F -measure, AUC and computation time on six publicly available image datasets. Performance of the proposed model (DDDCWT-SS) is also evaluated in terms of mean precision value, mean recall value, mean F -measure values and mean AUC values on all six datasets. Experimental results exhibited that the proposed model (DDDCWT-SS) outperforms several existing competitors in terms of F -measure on five datasets, recall on four datasets and AUC on two datasets. The proposed method outperforms all state-of-the-art methods in terms of mean recall value, mean F -measure value and mean AUC value on all six datasets. However, the proposed model demands less computation time in comparison with most of the existing methods in spatial domain. Although the proposed method (DDDCWT-SS) is simple, still there are several important issues which require further investigation like incorporation of more sophisticated visual features to further improve the performance. The research work can also be reached out to make the framework powerful by handling some difficulties like partial occlusion, articulation, background clutter and real-time requirements. We hope to encourage more future work along this direction. A challenging dataset related to specific challenges like partial occlusion and background clutter with accurate annotation and appropriate evaluation methodology would be desirable. In future, we will also be focussing on solving an application-oriented problem like image or video compression and video summarization using visual saliency.

Acknowledgements The authors express their gratitude to the University Grant Commission (UGC), India, and DST-Purse, India, for the obtained financial support in performing this research work.

References

1. Achanta R (2012) SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans Pattern Anal Mach Intell* 34(11):2274–2282
2. Achanta R, Susstrunk S (2010) Saliency detection using maximum symmetric surround. In: *Proceedings of 17th IEEE international conference on image processing (ICIP)*, pp 2653–2656
3. Achanta R, Hemami S, Estrada F, Susstrunk S (2009) Frequency-tuned salient region detection. In: *Proceedings of IEEE conference on Computer vision and pattern recognition*, pp 1597–1604
4. Alpert S, Galun M, Brandt A, Basri R (2012) Image segmentation by probabilistic bottom-up aggregation and cue integration. *IEEE Trans Pattern Anal Mach Intell* 34(2):315–327
5. Arya R, Singh N, Agrawal R (2015) A novel hybrid approach for salient object detection using local and global saliency in frequency domain. *Multimed Tools Appl* 75(14):8267–8287
6. Arya R, Singh N, Agrawal R (2017) A novel combination of second-order statistical features and segmentation using multilayer superpixels for salient object detection. *Appl Intell* 46(2):254–271
7. Bian P, Zhang L (2008) Biological plausibility of spectral domain approach for spatiotemporal visual saliency. In: *Proceedings of the international conference on neural information processing*, pp 251–258
8. Bian P, Zhang L (2010a) Piecewise frequency domain visual saliency detection. In: *Proceedings of IEEE third international conference on information and computing (ICIC)*, pp 269–272
9. Bian P, Zhang L (2010) Visual saliency: a biologically plausible contourlet-like frequency domain approach. *Cogn Neurodyn* 4(3):189–198
10. Borji A, Itti L (2013) State-of-the-art in visual attention modeling. *IEEE Trans Pattern Anal Mach Intell* 35(1):185–207
11. Borji A, Sihite DN, Itti L (2013) Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study. *IEEE Trans Image Process* 22(1):55–69
12. Borji A, Cheng M-M, Jiang H, Li J (2014) Salient object detection: a survey. In: *arXiv preprint arXiv:1411.5878*
13. Borji A, Cheng M-M, Jiang H, Li J (2015) Salient object detection: a benchmark. *IEEE Trans Image Process* 24(12):5706–5722

14. Boykov YY, Jolly M-P (2001) Interactive graph cuts for optimal boundary and region segmentation of objects in ND images. In: Proceedings of the eighth IEEE international conference on computer vision, pp 105–112
15. Bruce N, Tsotsos J (2006) Saliency based on information maximization. In: Advances in neural information processing systems, pp 155–162
16. Castleman Kenneth R (1996) Digital image processing. Prentice Hall Press, Upper Saddle River
17. Chen L-Q, Xie X, Fan X, Ma W-Y, Zhang H-J, Zhou H-Q (2003) A visual attention model for adapting images on small displays. *Multimed Syst* 9(4):353–364
18. Cheng M-M, Warrell J, Lin W-Y, Zheng S, Vineet V, Crook N (2013) Efficient salient region detection with soft image abstraction. In: Proceedings of IEEE international conference on computer vision, pp 1529–1536
19. Cheng M, Mitra NJ, Huang X, Torr PH, Hu S (2015) Global contrast based salient region detection. *IEEE Trans Pattern Anal Mach Intell* 37(3):569–582
20. Comaniciu D, Meer P (2002) Mean shift: a robust approach toward feature space analysis. *IEEE Trans Pattern Anal Mach Intell* 24(5):603–619
21. Daubechies I (1988) Orthonormal bases of compactly supported wavelets. *Commun Pure Appl Math* 41(7):909–996
22. Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. *J Mach Learn Res* 7(Jan):1–30
23. Derrac J, García S, Molina D, Herrera F (2011) A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm Evolut Comput* 1(1):3–18
24. Fang Y, Lin W, Lee B-S, Lau C-T, Chen Z, Lin C-W (2012) Bottom-up saliency detection model based on human visual sensitivity and amplitude spectrum. *IEEE Trans Multimed* 14(1):187–198
25. Fang Y, Chen Z, Lin W, Lin C-W (2012) Saliency detection in the compressed domain for adaptive image retargeting. *IEEE Trans Image Process* 21(9):3888–3901
26. Felzenszwalb PF, Huttenlocher DP (2004) Efficient graph-based image segmentation. *Int J Comput Vis* 59(2):167–181
27. Fu K, Gong C, Yang J, Zhou Y, Gu IY-H (2013) Superpixel based color contrast and color distribution driven salient object detection. *Sig Process Image Commun* 28(10):1448–1463
28. Fu K, Gong C, Gu IY-H, Yang J (2015) Normalized cut-based saliency detection by adaptive multi-level region merging. *IEEE Trans Image Process* 24(12):5671–5683
29. Gasparini F, Corchs S, Schettini R (2007) Low-quality image enhancement using visual attention. *Opt Eng* 46(4):040502
30. Goferman S, Zelnik-Manor L, Tal A (2012) Context-aware saliency detection. *IEEE Trans Pattern Anal Mach Intell* 34(10):1915–1926
31. Gopalakrishnan V, Hu Y, Rajan D (2010) Random walks on graphs for salient object detection in images. *IEEE Trans Image Process* 19(12):3232–3242
32. Guo C, Zhang L (2010) A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. *IEEE Trans Image Process* 19(1):185–198
33. Guo C, Ma Q, Zhang L (2008) Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In: Proceedings of IEEE conference on computer vision and pattern recognition (CVPR), pp 1–8
34. Hadizadeh H, Bajic IV (2014) Saliency-aware video compression. *IEEE Trans Image Process* 23(1):19–33
35. Harel J, Koch C, Perona P (2006) Graph-based visual saliency. In: Proceedings of the advances in neural information processing systems, pp 545–552
36. Hou X, Zhang L (2007) Saliency detection: A spectral residual approach. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1–8
37. Hou X, Harel J, Koch C (2012) Image signature: highlighting sparse salient regions. *IEEE Trans Pattern Anal Mach Intell* 34(1):194–201
38. Huang K, Tao D, Yuan Y, Li X, Tan T (2011) Biologically inspired features for scene classification in video surveillance. *IEEE Trans Syst Man Cybern B Cybern* 41(1):307–313
39. Huang X, Su Y, Liu Y (2016) Iteratively parsing contour fragments for object detection. *Neurocomputing* 175:585–598
40. Huo L, Jiao L, Wang S, Yang S (2016) Object-level saliency detection with color attributes. *Pattern Recogn* 49:162–173
41. Imamoglu N, Lin W, Fang Y (2013) A saliency detection model using low-level features based on wavelet transform. *IEEE Trans Multimed* 15(1):96–105
42. Iman RL, Davenport JM (1980) Approximations of the critical region of the fbietkan statistic. *Commun Stat Theory Methods* 9(6):571–595

43. Itti L (2000) Models of bottom-up and top-down visual attention. In: Doctoral dissertation California Institute of Technology
44. 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
45. Jian M, Dong J, Ma J (2011) Image retrieval using wavelet-based salient regions. *Imaging Sci J* 59(4):219–231
46. Jian M, Lam K-M, Dong J, Shen L (2015) Visual-patch-attention-aware saliency detection. *IEEE Trans Cybern* 45(8):1575–1586
47. Jiang H, Wang J, Yuan Z, Wu Y, Zheng N, Li S (2013) Salient object detection: a discriminative regional feature integration approach. In: Proceedings of IEEE conference on computer vision and pattern recognition (CVPR), pp 2083–2090
48. Judd T, Durand F, Torralba A (2012) A benchmark of computational models of saliency to predict human fixations. In: MIT technical report technical report
49. Kannan R, Ghinea G, Swaminathan S (2015) Salient region detection using patch level and region level image abstractions. *IEEE Signal Process Lett* 22(6):686–690
50. Karssemeijer N, te Brake GM (1996) Detection of stellate distortions in mammograms. *IEEE Trans Med Imaging* 15(5):611–619
51. Kim J, Han D, Tai Y-W, Kim J (2016) Salient region detection via high-dimensional color transform and local spatial support. *IEEE Trans Image Process* 25(1):9–23
52. Kingsbury N (1999) Image processing with complex wavelets. *Philos Trans R Soc Lond A Math Phys Eng Sci* 357(1760):2543–2560
53. Ko BC, Nam J-Y (2006) Object-of-interest image segmentation based on human attention and semantic region clustering. *JOSA A* 23(10):2462–2470
54. Koch C, Ullman S (1987) Shifts in selective visual attention: towards the underlying neural circuitry. In: Proceedings of the matters of intelligence, pp 115–141
55. Kumar K (2017) An efficient SOM technique for event summarization in multi-view surveillance videos. In: Proceedings of 5th international conference on advanced computing networking and informatics (ICACNI-17), pp 1–6
56. Kumar K, Shrimankar DD, Singh N (2016) Equal partition based clustering approach for event summarization in videos. In: Proceedings of IEEE conference on signal-image technology and internet-based systems (SITIS), pp 119–126
57. Kumar K, Shrimankar DD, Singh N (2017) Eratosthenes sieve based key-frame extraction technique for event summarization in videos. *Multimed Tools Appl* 77(6):7383–7404
58. Li Z, Itti L (2011) Saliency and gist features for target detection in satellite images. *IEEE Trans Image Process* 20(7):2017–2029
59. Li Z, Wu X-M, Chang S-F (2012) Segmentation using superpixels: a bipartite graph partitioning approach. In: Proceedings of IEEE conference on computer vision and pattern recognition (CVPR), pp 789–796
60. Li J, Levine MD, An X, Xu X, He H (2013) Visual saliency based on scale-space analysis in the frequency domain. *IEEE Trans Pattern Anal Mach Intell* 35(4):996–1010
61. Li J, Duan L-Y, Chen X, Huang T, Tian Y (2015) Finding the secret of image saliency in the frequency domain. *IEEE Trans Pattern Anal Mach Intell* 37(12):2428–2440
62. Liang J, Zhou J, Tong L, Bai X, Wang B (2018) Material based salient object detection from hyperspectral images. *Pattern Recogn* 76:476–490
63. Liu Z (2014) Saliency tree: a novel saliency detection framework. *IEEE Trans Image Process* 23(5):1937–1952
64. Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X et al (2007) Learning to detect a salient object. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1–8
65. Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X et al (2011) Learning to detect a salient object. *IEEE Trans Pattern Anal Mach Intell* 33(2):353–367
66. Ma Y-F, Zhang H-J (2003) Contrast-based image attention analysis by using fuzzy growing. In: Proceedings of ACM international conference on Multimedia, pp 374–381
67. Marchesotti L, Cifarelli C, Csürka G (2009) A framework for visual saliency detection with applications to image thumbnailing. In: Proceedings of 12th IEEE international conference on computer vision, pp 2232–2239
68. Murray N, Vanrell M, Otazu X, Parraga CA (2011) Saliency estimation using a non-parametric low-level vision model. In: Proceedings of IEEE conference on computer vision and pattern recognition (CVPR), pp 433–440
69. Naqvi SS, Browne WN, Hollitt C (2016) Salient object detection via spectral matting. *Pattern Recogn* 51:209–224

70. Navalpakkam V, Itti L (2006) An integrated model of top-down and bottom-up attention for optimizing detection speed. In: Proceedings of IEEE computer society conference on computer vision and pattern recognition, pp 2049–2056
71. Ninassi A, Meur OL, Callet PL, Barba D (2007) Does where you gaze on an image affect your perception of quality? Applying visual attention to image quality metric. In: Proceedings of IEEE international conference on image processing, pp II-169
72. Park J, Lee J-Y, Tai Y-W, Kweon IS (2012) Modeling photo composition and its application to photo re-arrangement. In: Proceedings of IEEE international conference on image processing (ICIP), pp 2741–2744
73. Perazzi F, Krahenbuhl P, Pritch Y, Hornung A (2012) Saliency filters: contrast based filtering for salient region detection. In: Proceedings of IEEE conference on computer vision and pattern recognition (CVPR), pp 733–740
74. Rother C, Bordeaux L, Hamadi Y, Blake A (2006) Autocollage. *ACM Trans Graph (TOG)* 25(3):847–852
75. Rutishauser U (2004) Is bottom-up attention useful for object recognition? In: Proceedings of IEEE computer society conference on computer vision and pattern recognition, pp II-37
76. Santella A, Agrawala M, DeCarlo D, Salesin D, Cohen M (2006) Gaze-based interaction for semi-automatic photo cropping. In: Proceedings of SIGCHI conference on human factors in computing systems, pp 771–780
77. Selesnick IW (2001) The double density DWT. In: Proceedings of wavelets in signal and image analysis, pp 39–66
78. Selesnick IW (2004) The double-density dual-tree DWT. *IEEE Trans Signal Process* 52(5):1304–1314
79. Selesnick IW, Baraniuk RG, Kingsbury NG (2005) The dual-tree complex wavelet transform. *IEEE Signal Process Mag* 22(6):123–151
80. Shen X, Wu Y (2012) A unified approach to salient object detection via low rank matrix recovery. In: Proceedings of IEEE conference on computer vision and pattern recognition (CVPR), pp 853–860
81. Shi J, Malik J (2000) Normalized cuts and image segmentation. *IEEE Trans Pattern Mach Intell* 22(8):888–905
82. Singh N, Arya R, Agrawal R (2014) A novel approach to combine features for salient object detection using constrained particle swarm optimization. *Pattern Recogn* 47(4):1731–1739
83. Singh N, Arya R, Agrawal R (2016) A convex hull approach in conjunction with Gaussian mixture model for salient object detection. *Digit Signal Process* 55:22–31
84. Singh N, Arya R, Agrawal R (2016b) A novel position prior using fusion of rule of thirds and image center for salient object detection. *Multimed Tools Appl* 76:1–18
85. Singh N, Arya R, Agrawal RK (2017) Performance enhancement of salient object detection using super-pixel based Gaussian mixture model. *Multimed Tools Appl* 77:1–19
86. Sun J, Lu H, Liu X (2015) Saliency region detection based on Markov absorption probabilities. *IEEE Trans Image Process* 24(5):1639–1649
87. Tian Q, Sebe N, Lew MS, Loupiaz E, Huang TS (2001) Content-based image retrieval using wavelet-based salient points. In: *Photonics west 2001-electronic imaging*, pp 425–436
88. Wang Y-S, Tai C-L, Sorkine O, Lee T-Y (2008) Optimized scale-and-stretch for image resizing. *ACM Trans Graph (TOG)* 27(5):118
89. Wang K, Lin L, Lu J, Li C, Shi K (2015) PISA: pixelwise image saliency by aggregating complementary appearance contrast measures with edge-preserving coherence. *IEEE Trans Image Process* 24(10):3019–3033
90. Xie Y, Lu H, Yang M-H (2013) Bayesian saliency via low and mid level cues. *IEEE Trans Image Process* 22(5):1689–1698
91. Yang C, Zhang L, Lu H, Ruan X, Yang M-H (2013) Saliency detection via graph-based manifold ranking. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 3166–3173
92. Yu Y, Wang B, Zhang L (2009) Pulse discrete cosine transform for saliency-based visual attention. In: Proceedings of IEEE 8th international conference on development and learning, pp 1–6
93. Yu J-G, Xia G-S, Samal A, Tian J (2016) Globally consistent correspondence of multiple feature sets using proximal Gauss-Seidel relaxation. *Pattern Recogn* 51:255–267
94. Zhang L, Tong MH, Marks TK, Shan H, Cottrell GW (2008) SUN: a Bayesian framework for saliency using natural statistics. *J Vis* 8(7):32
95. Zhu Z, Wahid K, Babyn P, Yang R (2013) Compressed sensing-based MRI reconstruction using complex double-density dual-tree DWT. *Int J Biomed Imaging* 2013:907501. <https://doi.org/10.1155/2013/907501>

96. Zhu L, Klein DA, Frintrop S, Cao Z, Cremers AB (2014) A multisize superpixel approach for salient object detection based on multivariate normal distribution estimation. *IEEE Trans Image Process* 23(12):5094–5107



Rinki Arya is working as an Artificial Intelligence Specialist in Accenture Labs, India. She obtained M.Tech. (Computer Science and Technology) and Ph.D. from Jawaharlal Nehru University, New Delhi. Her current research areas are computer vision, machine learning, image processing, object detection and pattern recognition.



R. K. Agrawal obtained M.Tech. (Computer Application) from Indian Institute of Technology Delhi, New Delhi, and Ph.D. (Computational Physics) from University of Delhi, Delhi. Presently he is working as a Professor in the School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi. His current research areas are classification, feature extraction and selection for pattern recognition problems in domains of image processing, security and bioinformatics.



Navjot Singh is working as an Assistant Professor in Motilal Nehru National Institute of Technology Allahabad, Allahabad, India. He obtained M.Tech. and Ph.D. (Computer Science and Technology) from Jawaharlal Nehru University, New Delhi, India. His current research areas are computer vision, image processing, object detection, pattern recognition, feature extraction and classification.

Affiliations

Rinki Arya¹ · R. K. Agrawal¹ · Navjot Singh²

R. K. Agrawal
rkajnu@gmail.com

Navjot Singh
navjot.singh.09@gmail.com

¹ School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi 110067, India

² Motilal Nehru National Institute of Technology Allahabad, Allahabad 211004, India