Neurocomputing 179 (2016) 186–201

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Region of interest extraction in remote sensing images by saliency analysis with the normal directional lifting wavelet transform

Libao Zhang *, Jie Chen, Bingchang Qiu

The College of Information Science and Technology, Beijing Normal University, Beijing, China

ARTICLE INFO

Article history:
Received 10 June 2015
Received in revised form
29 November 2015
Accepted 30 November 2015
Communicated by Y Gu.
Available online 15 December 2015

Keywords:
Image processing
Region of interest
Saliency analysis
Normal directional lifting wavelet transform

ABSTRACT

Region of interest (ROI) extraction techniques based on saliency comprise an important branch of remote sensing image analysis. In this study, we propose a novel ROI extraction method for high spatial resolution remote sensing images. High spatial resolution remote sensing images contain complex spatial information, clear details, and well-defined geographical objects, where the structure, edge, and texture information has important roles. To fully exploit these features, we construct a novel normal directional lifting wavelet transform to preserve local detail features in the wavelet domain, which is beneficial for the generation of edge and texture saliency maps. We also improve the extraction results by calculating the amount of self-information contained in the spectra to obtain a spectral saliency map. The final saliency map is a weighted fusion of the two maps. Our experimental results demonstrate that the proposed extraction algorithm can eliminate background information effectively as well as highlighting the ROIs with well-defined boundaries and shapes, thereby facilitating more accurate ROI extraction.

1. Introduction

Increases in the ability to acquire high spatial resolution remote sensing images by various satellites and sensors have led to great challenges in the detection of valuable targets from high spatial resolution remote sensing images [1–3]. Region of interest (ROI) extraction techniques based on saliency have been introduced into the remote sensing image analysis field and they have become a research hotspot in recent years [4–9]. In addition, these techniques are employed as an efficient information processing method to handle the rapidly growing volume of remote sensing images. After providing a potential ROI, the viewer can search for specific objects in the region and computing resources can be allocated in a reasonable manner to enhance the operating efficiency of an image processing system [10].

In remote sensing images, typical ROIs include residential areas, airports, airplanes, wharfs, and ships. Compared with the background, they have salient features that immediately grab human attention; hence, it is suitable to extract ROIs via saliency models. In particular, the salient characteristics are as follows.

a) Abundant and complex structure, edge, and texture information, which is typical of the interior of a residential area.
b) Unique shapes, particularly for airplanes, which are not as texture rich as residential areas, but their unique shape makes them stand out.
c) Orientation information, e.g., the ships usually head in the same direction because of the similar ocean currents and weather patterns in nearby waters.
d) Their distinct spectra compared with the surrounding environment.

The ROIs possess these characteristics, whereas the background does not, so high contrast stimuli are generated in receptive fields of the human visual system and human cortical cells may be hardwired to respond preferentially to these stimuli [11]. Visual saliency refers to distinctive parts of a scene that immediately attract significant attention without any prior information, thus it is flexible in adapting to different ROI extraction tasks, in which retraining is unnecessary.

Saliency-based methods were originally designed for natural scene images [12–15] by utilizing the intensity, color, orientation, texture, and other low-level features to determine contrast for saliency computation. One of the earliest computational models, which was built on a biologically plausible architecture [16], was proposed by Itti et al. (IT) [17]. This model obtains saliency maps based on the intensity, color, and orientation channels, and computes the final master saliency map by combining these three conspicuity maps based on center-surround differences.

Various computational models have been inspired by the biological concept of center-surround contrast in the IT model. These

* Corresponding author. Tel.: +86 13366069027.
E-mail address: libaozhang@bnu.edu.cn (L. Zhang).

http://dx.doi.org/10.1016/j.neucom.2015.11.093
0925-2312/

estimation models can be broadly classified as biologically-based, purely computational, and a combination. Harel's graph-based visual saliency method (GB) [18] is a combination model that employs an idea from graph theory to concentrate mass in activation maps and to obtain activation maps from raw features.

Among the purely computational models, Achanta et al. [14,15] attempted to build a saliency model (FT) using color contrast information. The feature vector was acquired in the CIE Lab color space and the absolute difference between the Gaussian-blurred image and the arithmetic mean vector was then calculated to obtain the saliency map. Goferman et al. [19] proposed a novel algorithm called context-aware saliency detection (CA). Supported by psychological evidence, CA uses a detection algorithm that relies on four basic principles reported in the psychological literature.

Cheng et al. [20] proposed a histogram-based contrast (HC) method to measure saliency for image pixels using color statistics determined for an input image. They also presented a regional contrast-based saliency extraction algorithm (RC), which simultaneously evaluates the global contrast differences and spatial coherence. In RC, the input image is first segmented into regions, before estimating saliency for each region as the weighted sum of the region's contrasts compared with all of the other regions in the image. The weights are set according to the spatial distance, where more distant regions are assigned smaller weights. RC obtains high precision and recall rates with natural images. In addition to these models, saliency models have been proposed in the spatial domain, such as an information theory-based computational model [21] and contrast-based filtering for salient region detection [22].

Recently, researchers have also tried to obtain solutions in the transform domain. The Fourier transform can be expressed in polar form using two different components: phase and amplitude spectra. By analyzing the log amplitude spectrum, Hou et al. [23] defined the spectral residuals (SR) algorithm, where the saliency map is derived by applying the inverse Fourier transform to an exponential function that combines spectral residual and phase spectrum information. In addition, Guo et al. [24] proposed a computational model based on the quaternion Fourier transform. Compared with the Fourier transform, the wavelet transform can perform multi-scale spatial and frequency analyses simultaneously, and thus it has begun to attract more attention from researchers. Murray et al. [25] computed weight maps from the high-pass wavelet coefficients of each level and the saliency map was obtained by the inverse wavelet transform of the weight maps. To improve this model, Imamoglu et al. [26] proposed a wavelet transform-based computational model (WT) that uses low-level features, which considers both local center-surround differences and the global contrast, thereby obtaining better results than the method of Murray et al. [25].

It should be noted that these standard saliency detection methods were not designed specifically for remote sensing images and differences exist between ROI extraction from remote sensing image and natural scenes. The ROIs in natural scenes have less complex textures and their distinct colors make them instantly recognizable from the surroundings. In addition, when shooting a picture, photographers manually set the lens to blur the background and focus on the ROI, which helps to highlight the ROIs. Furthermore, there is a strong center bias because human photographers tend to place one or two objects of interest in the center of photographs [27], which significantly narrows the search when locating ROIs. By contrast, ROIs such as airplanes and ships are scattered in the background of remote sensing images, where their positions and number are unpredictable. Moreover, the structure, shape, and texture information is abundant and complex in a high spatial resolution remote sensing image. To ensure that the input is accurate for subsequent applications, such as object recognition, image compression, and image retrieval, the principles followed to achieve good ROI extraction are stricter in remote sensing images. For example, the ROIs should be uniformly highlighted with well-defined boundaries to ensure the integrity of ROIs. In addition, the final object maps should retain full resolution without any loss of detail to preserve the fineness of remote sensing images. Thus, there is a difference between our method and standard saliency methods for objective evaluation, as described in Section 3.2.

In general, standard saliency methods may ignore the fact that high spatial resolution remote sensing images contain complex spatial information, clear details, and well-defined geography objects. Thus, they are likely to extract these complex structure, edge, and texture features in a coarse manner. For example, the IT model [17] obtains the saliency map based on the intensity, color, and orientation channels. In addition, the orientation information is obtained using oriented Gabor pyramids $O(\sigma, \theta)$ with $\theta \in \{0°, 45°, 90°, 135°\}$. GB [18] is derived from IT, so it obtains orientation information in the same manner. However, only four directions are used, which makes this method less accurate than ours. Moreover, the saliency map is 1/256 of the original image size, which inevitably leads to the loss of texture details. The same problem affects SR [23], which down-samples the input image to 64×64 pixels. Other methods such as FT [14], CA [19], HC [20], and RC [20] neglect texture features and they focus preferentially on the color and luminance features during saliency calculations. WT [26] represents different features that range from edges to textures by wavelets, but it employs Daubechies wavelets (Daub.5), which are not suitable for approximating image features with an arbitrary orientation that is not vertical or horizontal.

In recent years, saliency computational models have also been introduced into the remote sensing image processing field and they have become a research hotspot. A saliency computation approach (RS) was introduced [4] to select perceptually salient and highly informative regions that represent the main contents of high-resolution remote sensing images. Zhang et al. [7] used high-frequency filters for frequency domain analysis (FDA) to detect ROIs in high spatial resolution remote sensing images, but the model produces an attenuated interior in ROIs, thereby yielding incomplete ROI extraction. Zhang et al. [8] then proposed another model based on multi-scale feature fusion (MFF) of the intensity saliency result and the orientation saliency result to obtain one saliency map for ROI extraction, where the orientation saliency is based on the conventional lifting wavelet transform (LWT). Furthermore, Wang et al. [28] successfully applied the saliency technique to airport detection. Ding et al. [29] also attempted to implement ship detection using a saliency technique. Another efficient ROI extraction method based on spectral analysis was introduced for saliency testing in remote sensing images [6].

Based on the four characteristics of ROIs mentioned above, we propose a novel directional wavelet called normal directional LWT (ND-LWT) to fully exploit the information contained in texture, shape, and orientation. This method is designed specifically for high spatial resolution remote sensing images and it also makes use of spectral information to facilitate accurate extraction. The proposed method can preserve the local details of features in the wavelet domain, which is beneficial for generating the edge and texture saliency map. Our experimental results demonstrate that the proposed extraction algorithm can eliminate the background information in an effective manner as well as highlighting the ROIs with well-defined boundaries and shapes, thereby allowing more accurate ROI extraction.

The remainder of this paper is organized as follows. The proposed ROI extraction algorithm is introduced in Section 2. Section 3 presents the experimental results and discussion. In Section 4, we give our conclusions.

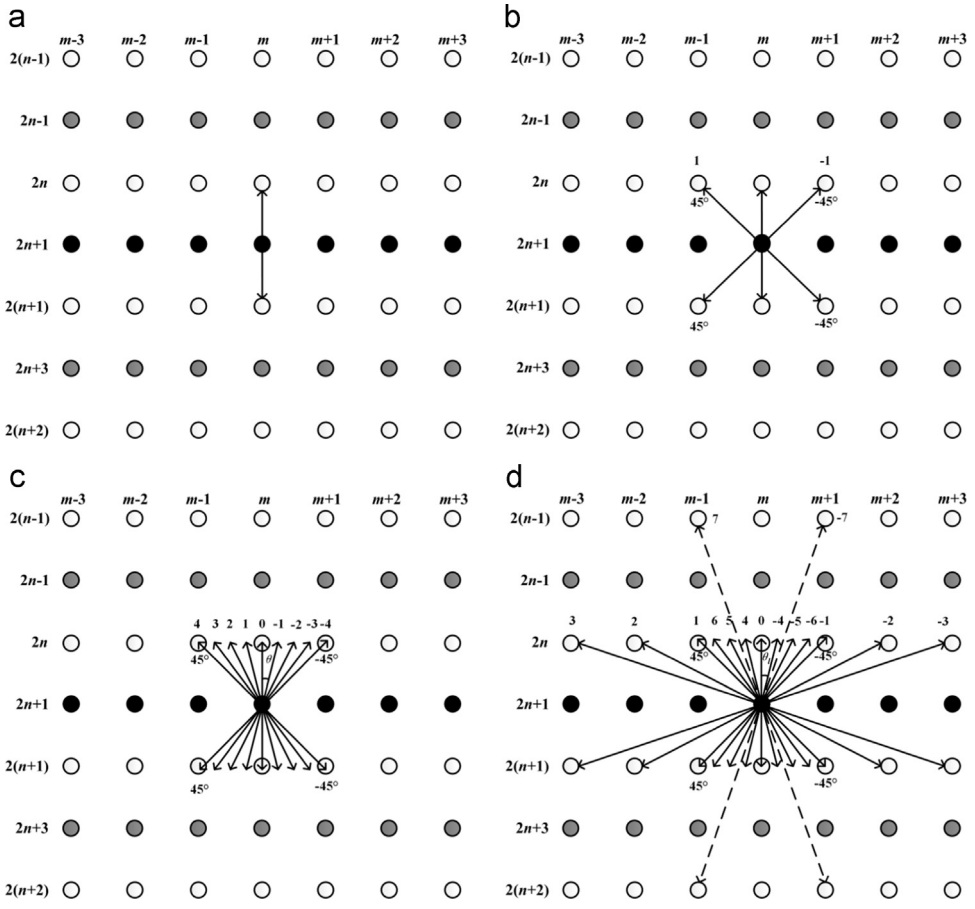


Fig. 1. Directions for prediction and updating in the vertical processing.

2. Methodology

In this section, we describe the architecture of our method in detail. There are three main steps. First, we generate an edge and texture saliency map based on ND-LWT. In this subsection, we briefly outline the evolution from LWT to directional LWT (DLWT), and then to ND-LWT. Next, the spectral saliency map is obtained through self-information computation. Finally, we produce the final saliency map by weighted fusion.

2.1. Edge and texture saliency map

2.1.1. Traditional lifting scheme

The traditional lifting scheme [30,31] can be considered as an alternative implementation of the first generation classical discrete wavelet transform, where it comprises four steps: *Split*, *Predict*, *Update*, and *Normalize*. The traditional lifting scheme can decompose a two-dimensional (2D) lifting wavelet using two one-dimensional (1D) lifting wavelets. Without any loss of generality, we assume that it is first decomposed by a 1D lifting transform in the vertical direction and then by a 1D lifting transform in the horizontal direction. Let $x[m, n]$ be a 2D signal. The classical lifting scheme is performed as follows.

2.1.1.1. Split:. The input signal is split into two parts: the even subset $x_e[m, n]$ and the odd subset $x_o[m, n]$.

$$\begin{cases} x_e[m,n] = x[m,2n] \\ x_o[m,n] = x[m,2n+1] \end{cases} \tag{1}$$

2.1.1.2. Predict:. The odd subset $x_o[m, n]$ located at an integer position is predicted from the neighboring even subset $x_e[m, n]$. The resulting prediction residuals or high-frequency subband coefficients are

$$d[m,n] = x_o[m,n] - P_e[m,n], \tag{2}$$

where the prediction value $P_e[m, n]$ is a linear combination of the neighboring even subset:

$$P_e[m,n] = \sum_i \alpha_i x_e[m, n+i], \tag{3}$$

where α_i is the high-pass filter coefficient given by the filter taps.

2.1.1.3. Update:. The even subset $x_e[m, n]$ is updated with the neighboring high-frequency subband coefficients $d[m, n]$. The coarse approximation values or low-frequency subband coefficients are

$$c[m,n] = x_e[m,n] + U_d[m,n], \tag{4}$$

where $U_d[m, n]$ is a linear combination of neighboring prediction residual values $d[m, n]$:

$$U_d[m,n] = \sum_j \beta_j d[m, n+j], \tag{5}$$

where β_j is the low-pass filter coefficient, which is also given by the filter taps.

2.1.1.4. Normalize:. The outputs are weighted by K_e and K_o. These values are used to normalize the energy of the underlying scaling and the wavelet functions.

After completing the 1D lifting-based vertical transform, the 1D lifting-based horizontal transform is performed in the same manner. Each lifting step is always invertible. If the same $_{pi}$ and u_i are selected in the prediction and updating stages, the lifting scheme guarantees perfect reconstruction.

2.1.2. Directional lifting wavelet transform

To relax the condition of LWT that the predictor should use samples from the current row (or column during columnwise processing) that is being processed (see Fig. 1(a)), Gerek et al. [32] proposed a 2D orientation-adaptive lifting structure that introduces two more diagonal directions ($\pm$ 45°) from the upper and lower rows in the prediction and updating stages of the lifting structure (see Fig. 1(b)). Instead of always making the predictions in a horizontal or vertical direction, they proposed a rule that allows the optimal direction for prediction to be selected, where the prediction error is minimal in the chosen direction. As a result, the detailed images obtained by the directionally adaptive 5/3 wavelet generally contain less signal energy at several decomposition levels. High-band signal energy reduction at the diagonal edge locations yields better compression, which preserves the sharp edges in the original image better compared with the ordinary 5/3 wavelet decomposition. Subsequently, Ding et al. [33] proposed another 2D wavelet transform scheme for adaptive directional lifting (ADL) in image coding to further enhance the spatial resolution with higher accuracy. Instead of only applying integer pixel positions, as found in the previous method, ADL achieves high angular resolution in the prediction and update operations by using fractional pixels, which can be calculated with any existing interpolation method (see Fig. 1(c)). Therefore, the multi-orientation attribute of DLWT is highly suitable for capturing the rich texture information found in remote sensing images. In particular, when the edges are not horizontal or vertical, a better description can be obtained by aligning the direction of wavelet transform to the direction of the edges.

2.1.3. Normal direction lifting wavelet transform

We propose ND-LWT to better measure the saliency for ROI extraction, which differs from the aforementioned DLWT method in four mainly respects, as follows.

a) The ROIs in remote sensing images are usually texture rich with complicated structures in the form of edges, so ND-LWT aims to capture this characteristic. In contrast to the well-studied directional LWT, which selects the prediction direction to significantly reduce the signal energy in the high-pass subbands, ND-LWT uses the normal in the DLWT to highlight the edges, which obtains large-magnitude high-frequency coefficients on the edges.
b) As shown in Fig. 1(d), we refine the directional lifting approaches by improving the choice of directions. In total, 15 discrete directions labeled from −7 to 7, including samples from both the neighboring and the second most distant odd rows, are defined as vertical transform direction templates to achieve higher angular resolution during prediction, which allows more accurate edge detection and enhancement.
c) Traditional DLWT is used for data compression, which means that it is necessary to save the optimal transform direction based on the angle for each pixel in the odd subset to recover the original data. ND-LWT is used for feature extraction during saliency analysis, but there is no need to implement an inverse wavelet transform to reconstruct the original image. Therefore, the storage requirements can be reduced by abandoning directional angle saving.
d) We also propose a novel high-pass coefficient filtering process to reduce the background disturbance when generating the binary mask. It should be noted that not all of the coefficients in the high-pass subband correspond to the pixels of ROIs, such as non-edge coefficients. They are typically smaller than edge coefficients after the application of ND-LWT. Inspired by the wavelet threshold denoising scheme, we successfully eliminate these unwanted non-edge coefficients by setting an appropriate threshold. Hence, the successfully retained coefficients are mainly those that preserve the edge and texture information.

Next, we describe the ND-LWT algorithm explicitly. The prediction of the odd subset $x_o[m, n]$ involves a linear combination of neighboring even coefficients with strong or weak correlations. The prediction values are shown in Eq. (6), where sign(x) is 1 for $x \geq 0$ but -1 otherwise, and θ_l is the optimal transform direction.

$$P_e[m,n] = \begin{cases} \sum_i \alpha_i x_e[m+\text{sign}(i-1)\tan\ \theta_l, n+i] & l \in [-6\cdots 6] \\ \sum_i \alpha_i x_e[m+\text{sign}(i-1)\text{sign}(\tan\ \theta_l), n+i+\text{sign}(i-1)] & l \in [-7,7] \end{cases} \quad (6)$$

Similar to the prediction stage, during the updating stage for ND-LWT, the updated value of the even subset $x_e[m, n]$ is a linear combination of the neighboring prediction residual values with strong or weak correlations, which is computed using the following equation.

$$U_d[m,n] = \begin{cases} \sum_j \beta_j d[m+\text{sign}(j)\tan\ \theta_l, n+j] & l \in [-6\cdots 6] \\ \sum_j \beta_j d[m+\text{sign}(j)\text{sign}(\tan\ \theta_l), n+j+\text{sign}(j)] & l \in [-7,7] \end{cases} \quad (7)$$

To perform the ND-LWT at angle θ_l, the intensity values are required at the fractional pixel locations. In other words, tan θ_l may not be an integer in Eqs. (6) and (7). The sinc interpolation technique, which was applied successfully in a previous study [33], is employed to calculate the fractional pixel values in the proposed method.

It should be noted that the principle followed for selecting the optimal transform direction is critical in ND-LWT. After analyzing the local spatial correlations in all directions, ND-LWT selects a direction to maximize the high-frequency energy, which is typically the direction of the edge normal. Hence, ND-LWT can preserve the local detail features in the wavelet domain. An example using the popular 5/3-tap biorthogonal wavelet filter is shown in Fig. 2, which indicates that the ND-LWT preserves the local features better in the wavelet domain.

The edge and texture saliency map is generated according to two key concepts. First, the ND-LWT preserves the local features. Second, self-similarity is exploited across different scales of the wavelet transform. Fig. 3 shows a three-step frame diagram to illustrate the edge and texture saliency generation process. The first step involves performing the ND-LWT, which is followed by wavelet coefficient integration and high-pass coefficient filtering. Subsequently, the wavelet coefficients that belong to ROIs are selected according to their self-similarity across different scales. Finally, the saliency map is computed in the third step.

2.1.3.1. Step 1:. The input panchromatic remote image X is first decomposed using an N-level ND-LWT:

$$[A_N, H_s, V_s, D_s] = \text{ND}_N(X), \quad (8)$$

where $\text{ND}_N(\bullet)$ denotes the N-level ND-LWT decomposition. The maximum number of scaling for the ND-LWT decomposition process is $N = \lfloor \log_2 h/2 \rfloor$, where, h is the shorter border of the input remote sensing image, the resolution index $s \in \{1, 2, \ldots, N\}$, and the Nth level corresponds to the coarsest resolution. A_N is the approximation output at the coarsest resolution, and H_s, V_s, and D_s

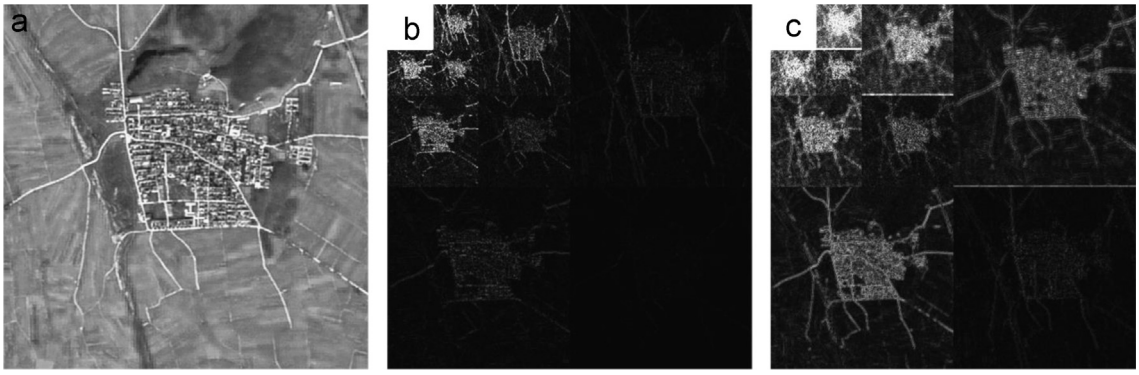


Fig. 2. Comparison of wavelet transforms: (a) intensity image, (b) conventional LWT, and (c) ND-LWT.

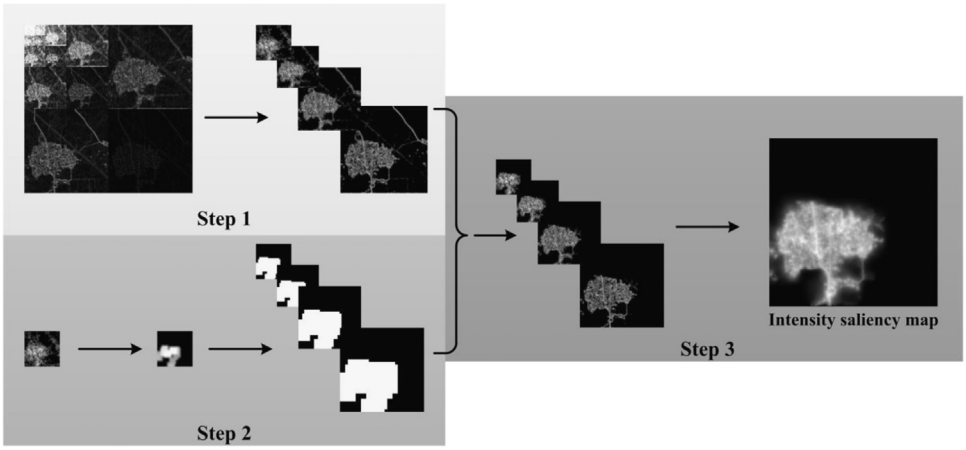


Fig. 3. Visual frame diagram of the edge and texture saliency map generation.

are the horizontal, vertical, and diagonal details for the given scale s, respectively.

To perform wavelet coefficient integration and high-pass coefficient filtering, we implement the following stages to modify the ND-LWT coefficients.

a) For every scale s, the coefficients in H_s, V_s, and D_s are replaced by their absolute values, and then normalized to [0, 1].
b) The coefficients in H_s, V_s, and D_s are further integrated using one pyramid.

$$L_s = H_s + V_s + D_s, \quad s \in \{1, \cdots, N\} \tag{9}$$

c) It should be noted that not all of the coefficients in the high-pass subband correspond to pixels in the ROIs, such as non-edge coefficients. Thus, a high-pass coefficient filtering process is necessary to eliminate these unwanted non-edge coefficients by using an appropriate threshold.

$$\bar{L}_s = \begin{cases} 0 & \text{if } L_s < t \\ L_s & \text{otherwise} \end{cases} \tag{10}$$

Inspired by wavelet shrinkage and the wavelet threshold denoising theory [34], t is computed as

$$t = \rho\sqrt{2 \ln z}, \tag{11}$$

$$\rho = \frac{Median(L_s)}{0.6745}, \tag{12}$$

where ρ is the noise variance, z is the size of the input signal, and Media(L_s) denotes the median value of L_s.

An example showing the results of N-level ND-LWT decomposition and the $\bar{L}_s$ coefficients is presented in Step 1 in Fig. 3.

2.1.3.2. Step 2:. In addition to the high-pass small coefficients, other non-ROIs (e.g., linear road and mountain ridge) are also undesirable forms of interference during ROI extraction. First, we select the wavelet coefficients that belong to ROIs at scale N. Next, the coefficients of the other scales are determined in an efficient manner based on the self-similarity principle across different scales [35]. The specific operation is described as follows.

a) Morphological opening is generally used to smooth the contours of an object and to break narrow isthmuses. This method works well when selecting ROI coefficients at scale N, and thus the background coefficients are eliminated in an effective manner:

$$R_N = \bar{L}_N \circ f_m, \tag{13}$$

where $\circ$ denotes the morphological opening operation and f_m is the $m \times m$ 2D unit matrix, where $m = 5$ works well in our method.

b) Filter R_N using a 3×3 Gaussian template g:

$$R_N \leftarrow R_N * g, \tag{14}$$

where * denotes the convolution operation.
c) Determine the root $\bar{R}_N$ as follows.

$$\bar{R}_N = \begin{cases} 1 & \text{if } R_N > 0 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

As shown in Fig. 4, one parent coefficient corresponds to four child coefficients. If a parent coefficient belongs to the ROIs, its child coefficients also belong to the ROIs.

$$\overline{R}_s = \begin{cases} 1 & \text{if } \overline{R}_{s+1} = 1 \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

An example showing the results for $\overline{L}_N$, R_N, and $\overline{R}_s$ is presented in Fig. 3, Step 2.

2.1.3.3. Step 3:. The edge and texture saliency map is created by linearly fusing the feature maps at each scale without any normalization operation [16]:

$$I = \oplus_{s=1}^{N} \{(\overline{L}_s * g) \cdot \overline{R}_s\}, \tag{17}$$

where • is the dot multiplication operation between two matrices, and $\oplus$ denotes interpolation of the map to level 1 and point-to-point addition.

Fig. 5 shows some examples of the edge and texture saliency maps produced by the proposed algorithm.

2.2. Spectral saliency map

Self-information is a valid method for measuring saliency based on the probability of occurrence. First, a 1D intensity histogram is constructed for different channels. Spectral feature maps are then produced by calculating the amount of self-information in the spectra. Finally, we construct the spectral saliency map by fusing these spectral feature maps.

Let X_l denote the multi-spectral channel images for the remote sensing image, where l is the label number for each spectral channel. Then, $X_l(i,j)$ represents the intensity value of X_l located at (i, j). The 1D intensity histogram for the range [0, 255] is first constructed according to Eq. (18), where k is the kth intensity value and n_k is the pixel count for the intensity k in the image.

$$h(k) = n_k, \quad k = 0, 1, \cdots, 255 \tag{18}$$

The normalized histogram is given by (19), where $p(k)$ is an estimate of the probability of occurrence for intensity k in X_l and X_l measures $M \times N$.

$$p(k) = n_k/(M \times N), \quad k = 0, 1, \cdots, 255 \tag{19}$$

The amount of self-information at each intensity level is computed as,

$$L(k) = -\log(p(k)). \tag{20}$$

Next, we substitute k with $X_l(i, j)$ to generate the spectral feature map E_l of channel l,

$$E_l(i,j) = L(X_l(i,j)). \tag{21}$$

The final spectral conspicuity map $\tilde{E}$ is

$$\tilde{E} = \sum_{l=1}^{4} w_l E_l, \tag{22}$$

where the weight w_l is obtained by

$$w_l = -\log\left(\frac{h_l}{h_1 + h_2 + h_3 + h_4}\right), \tag{23}$$

$$h_l = \frac{\sum_i \sum_j X_l(i,j)}{\sum_{l=1}^{4} \sum_i \sum_j X_l(i,j)}. \tag{24}$$

Fig. 6 shows a visual frame diagram to illustrate the spectral information content analysis in spectral saliency map computing.

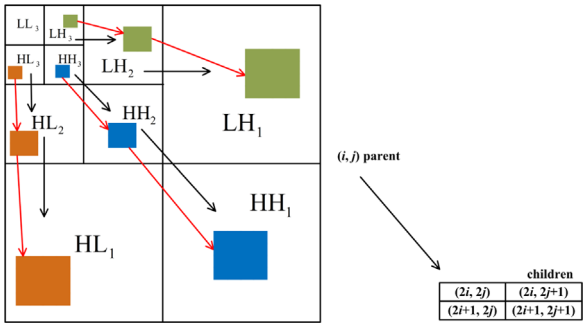


Fig. 4. Schematic diagram of the self-similarity across different scales.

2.3. Final saliency map

We enhance the final saliency maps by combining the edge and texture saliency maps with spectral saliency maps. We propose a weighted fusion method, which is denoted as $W(\bullet)$ and it includes the following steps.

a) Normalize the maps to [0, 255].
b) Compute the average intensity $\overline{m}$ of each map.
c) Multiply the map by $(255 - \overline{m})^2$.

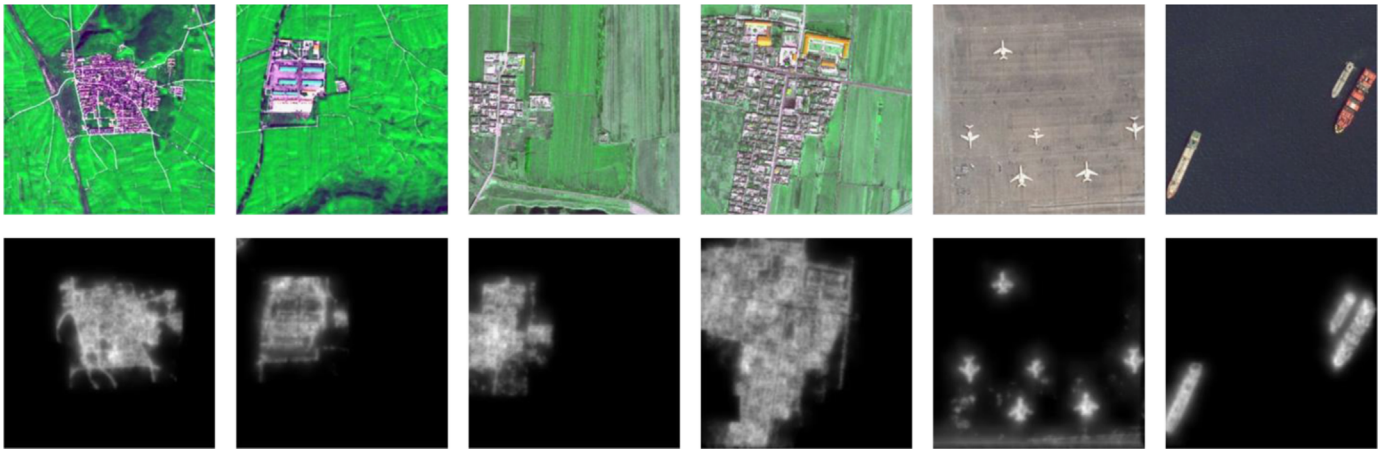


Fig. 5. Top: original remote sensing images. Bottom: edge and texture saliency maps.

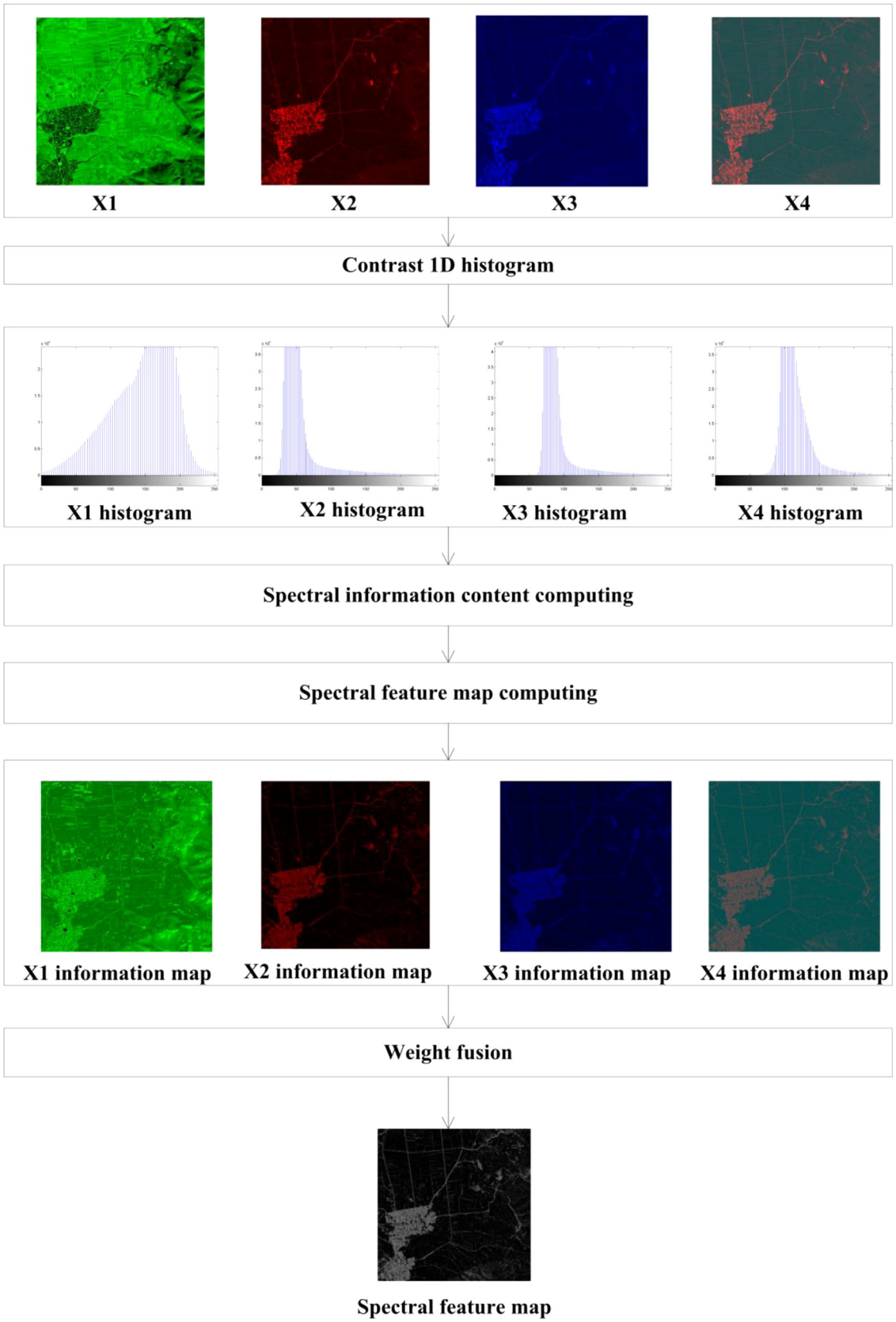


Fig. 6. Spectral self-information analysis for spectral saliency map computation.

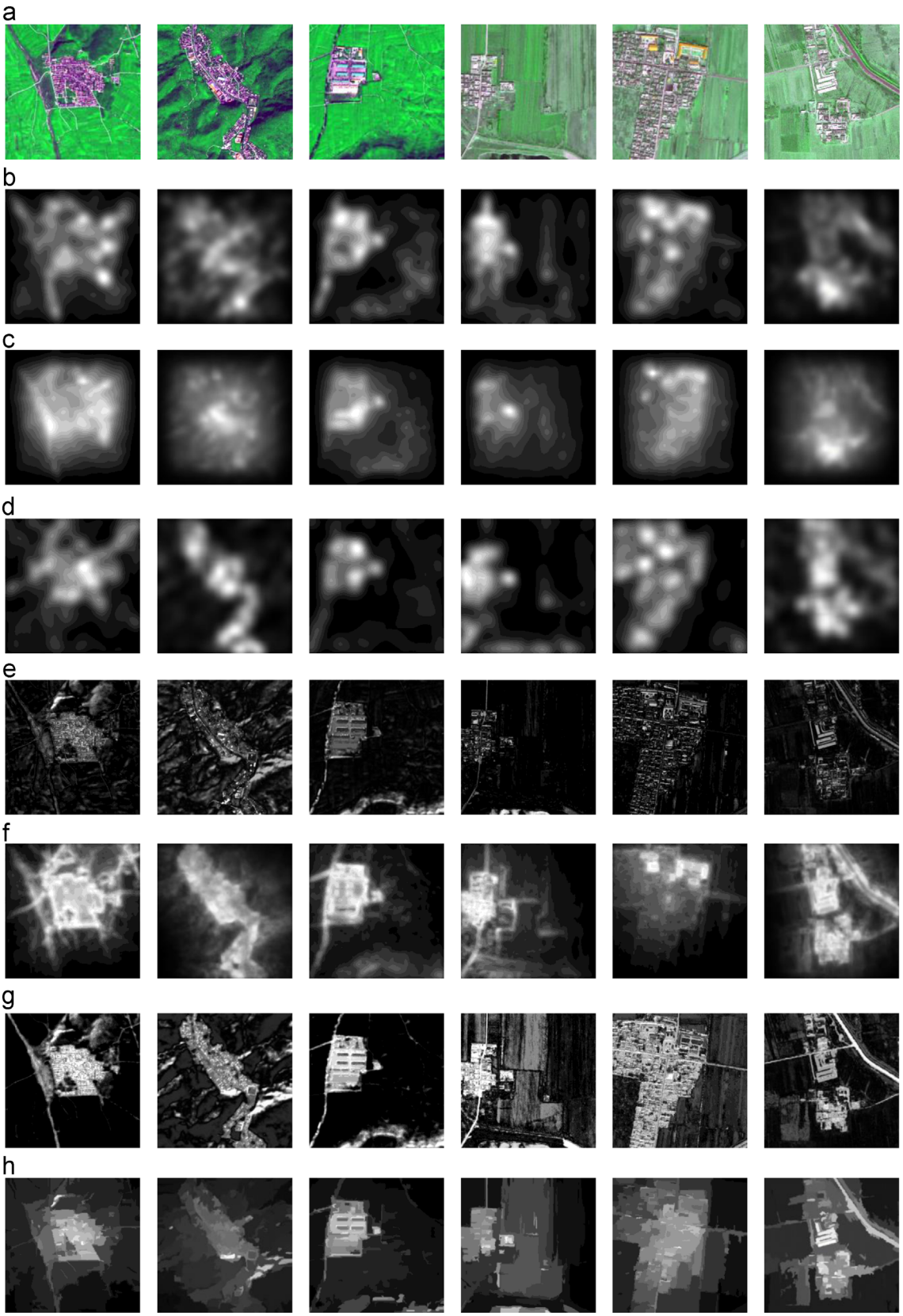


Fig. 7. Visual comparison of saliency maps on the satellite set (column 1–3: SPOT5, column 4–6: GeoEye-1). From top to bottom are (a) original images, saliency maps obtained by (b) IT, (c) GB, (d) SR, (e) FT, (f) CA, (g) HC, (h) RC, (i) WT, (j) RS, (k) FDA, (l)MFF, and (m) ours.

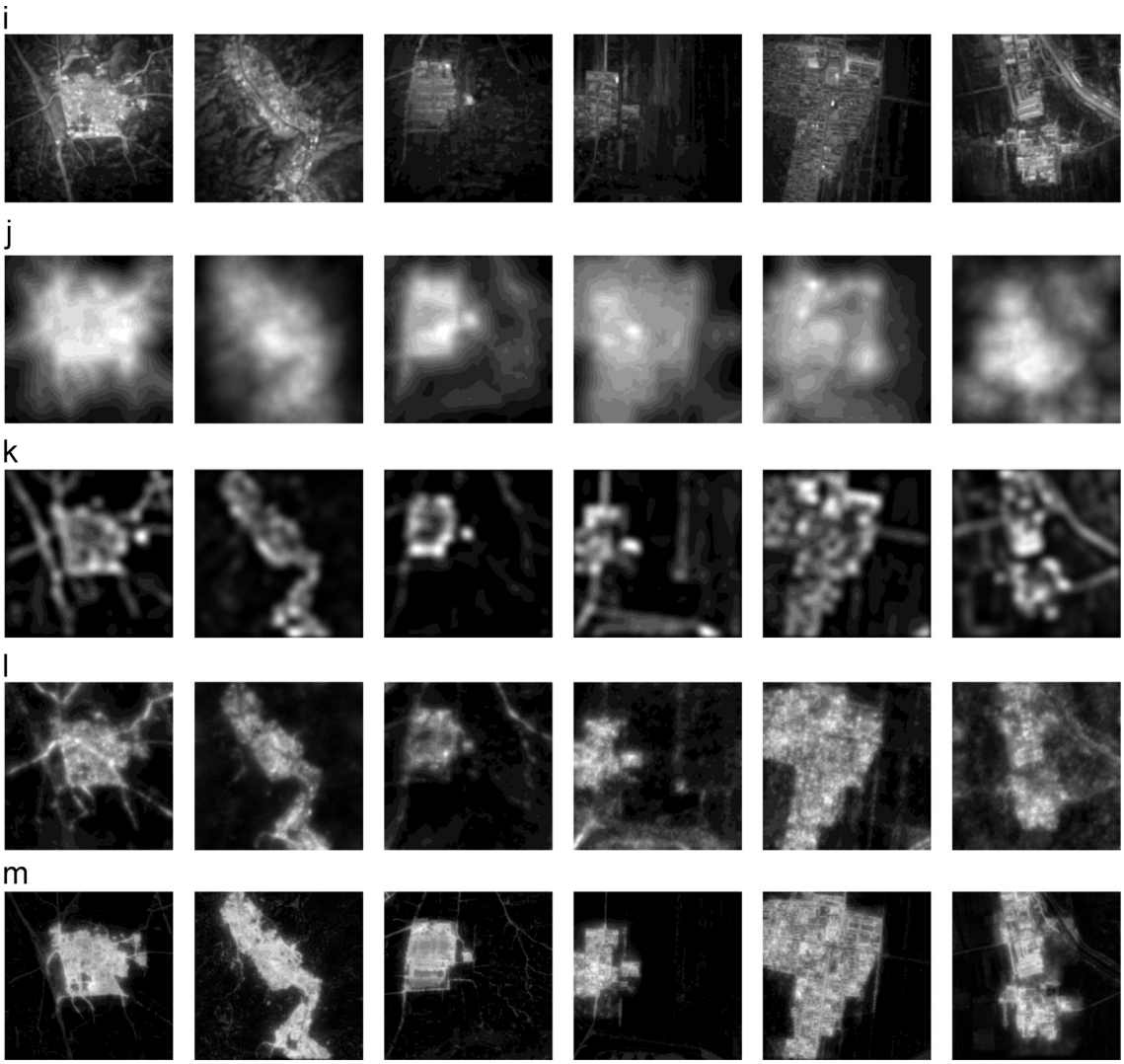


Fig. 7. (*continued*)

Using the weighted operator $W(\bullet)$, the final saliency map is calculated as follows.

$$S = N\left(W(I) + W(\tilde{E})\right) \tag{25}$$

After generating the final saliency map, an optimal threshold is computed as described by Otsu [36] to transform the saliency map into a binary image. Finally, we multiply the binary image by the original remote sensing image for ROI extraction.

3. Experiments and discussion

To evaluate the model performance in both qualitative and quantitative terms, we compared our ROI extraction model with 11 other state-of-the-art methods, i.e., IT [17], GB [18], SR [23], FT [14], CA [19], HC [20], RC [20], WT [26], RS [4], FDA [7], and MFF [8]. These models were selected for the following reasons: IT is biologically motivated; FT is a purely computational model in the frequency domain; SR estimates saliency in the Fourier transform domain; GB is a combination approach; CA, HC, and RC are all implemented in the spatial domain; WT estimates saliency in the Wavelet Transform domain; RS, FDA, and MFF were all developed for remote sensing images, and they are implemented in the spatial, frequency, and wavelet domains, respectively.

We applied our model to 80 high spatial resolution remote sensing images from two sets. The images in the first set were acquired by two satellites: the SPOT5 satellite with a resolution of 2.5 m and the GeoEye-1 satellite with a resolution of 1 m. The images in the other set were obtained from Google Earth with a resolution of 0.5 m and 1.0 m. However, to keep this study reasonably concise, we only present the visual results for 12 images in the qualitative comparison. The objective quality evaluation described in Section 3.2 was actually based on the whole set of 80 images. In addition, the same computer configuration was used for all of the experiments: a Windows platform with an Intel (R) Pentium(R) G630 (2.70 GHz) CPU and 4 GB RAM.

3.1. Qualitative experiment

Visual comparisons of the saliency maps and the ROI results are shown in Figs. 7–10. The resolution of the saliency maps produced by the IT, GB, and SR models was low. Thus, when these down-sampled saliency maps were used to extract ROIs, interpolation was required to enlarge the maps to full resolution. Therefore, these models sacrificed some precision when detecting the general outline of the ROIs, and thus many background regions were falsely recognized and some ROIs were lost. The remaining models produced saliency maps at full resolution, where more details and well-defined borders were visible, although they still had their

own limitations. FT failed to highlight the entire salient area, which resulted in incomplete descriptions of the salient area interior and the common occurrence of scattered background noise, such as green space and shadows, in the corresponding ROIs. CA, as indicated by its full name of "context-aware," aims to find the ROI and its near ambience. However, the implicated

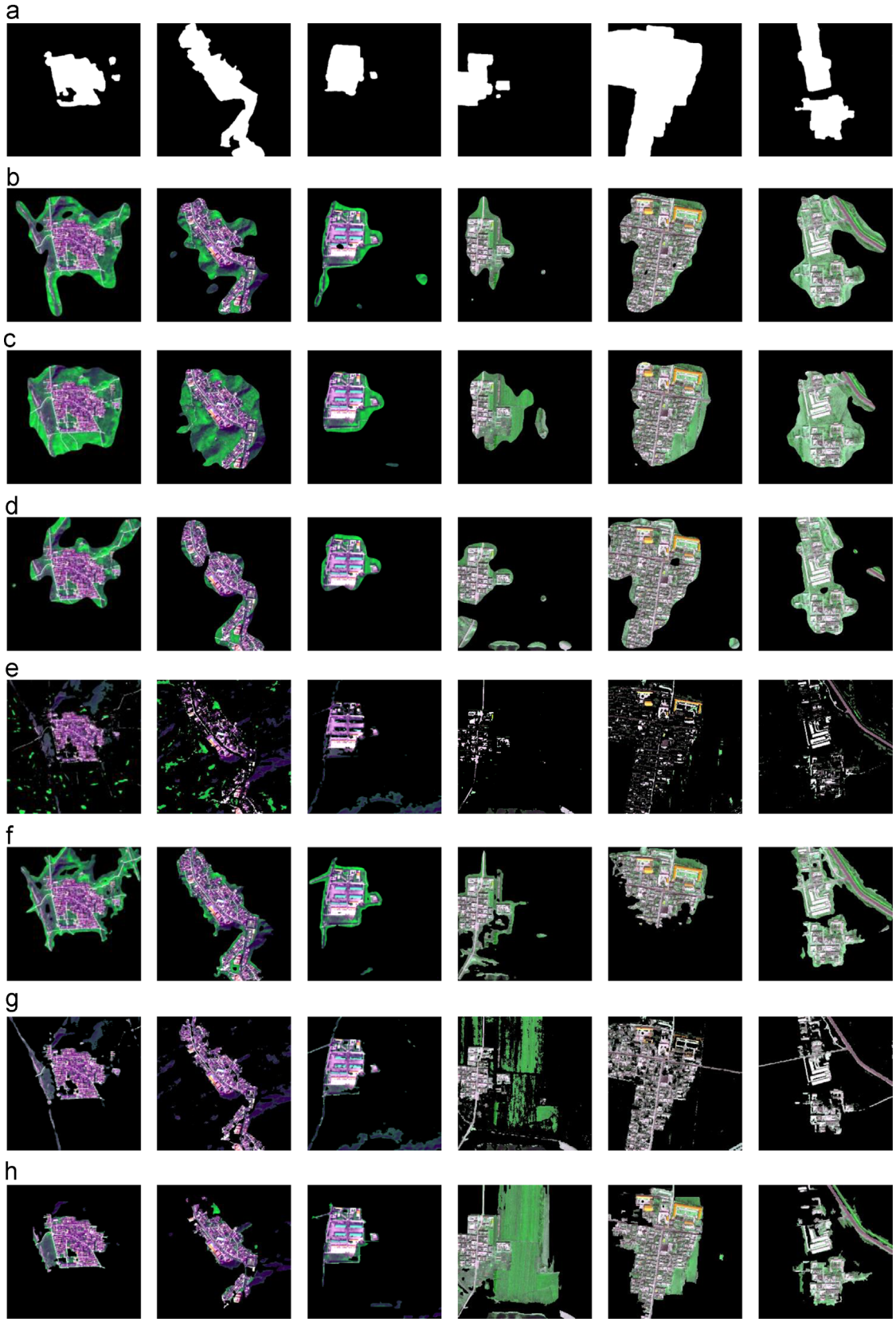


Fig. 8. Visual comparison of ROIs on the satellite set (columns 1–3: SPOT5, columns 4–6: GeoEye-1). From top to bottom are (a) groundtruth masks, ROIs obtained by (b) IT, (c) GB, (d) SR, (e) FT, (f) CA, (g) HC, (h) RC, (i) WT, (j) RS, (k) FDA, (l) MFF, and (m) ours.

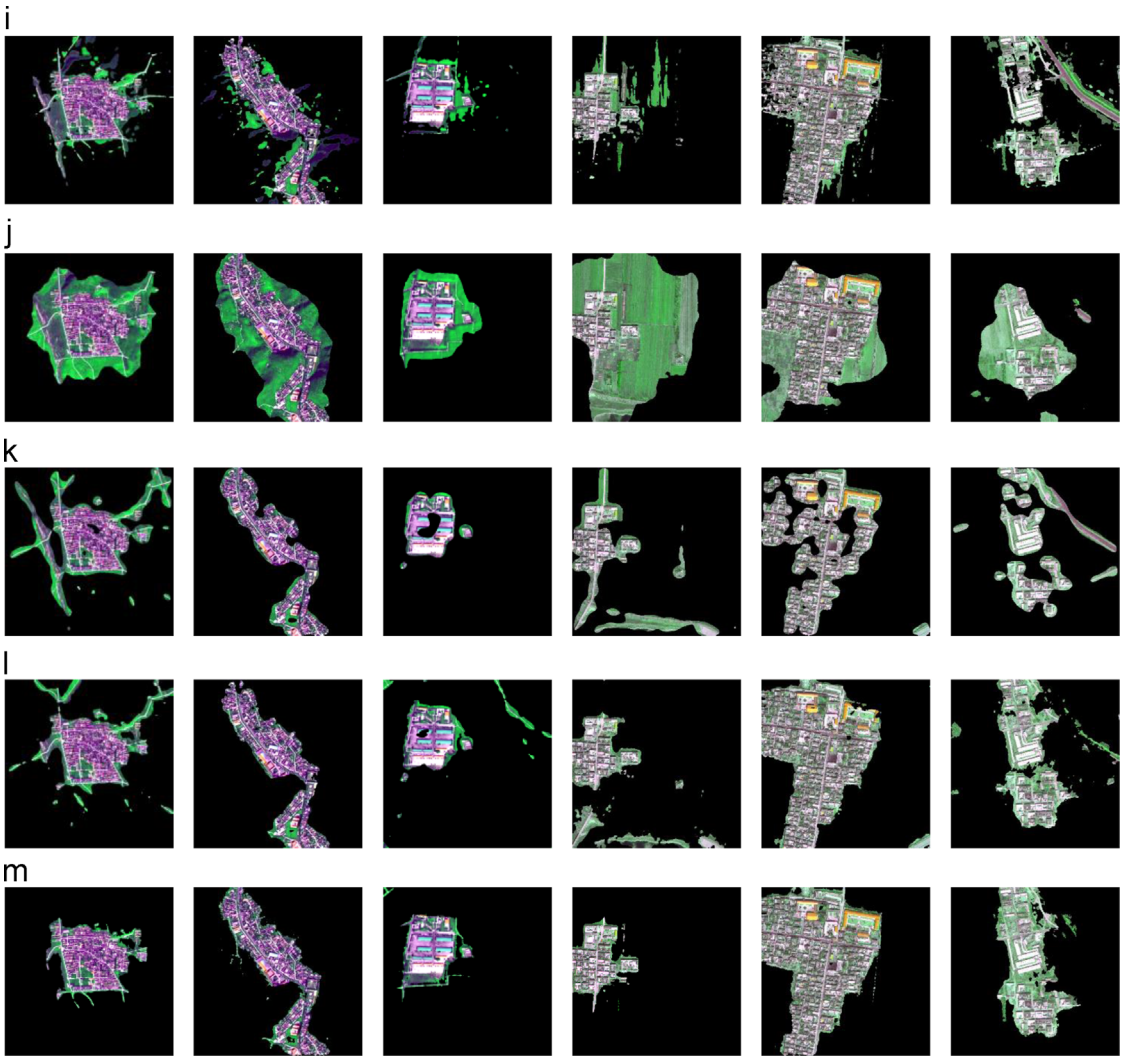


Fig. 8. (*continued*)

background was not applicable to our goal of identifying residential regions as ROIs with high precision. The HC and RC models obtained some better results, but they still failed to eliminate non-ROIs such as isolated linear roads and mountain ridges. The results obtained by the RC models included some fragmented background regions inside the ROIs. The WT model yielded well-defined boundaries, but the high-frequency details extracted using the wavelet transform contained some redundant background areas in the ROIs. Furthermore, the last three methods used in the comparison were specifically designed for ROI extraction in remote sensing images, where RS is a modified version of GB and the ROIs could be detected effectively, but the detected regions contained some background information. FDA could segment the images precisely but incompletely, especially within the interior of the residential area, which contained some undesirable holes. The MFF model also obtained good performance with remote sensing images, but it still failed to eliminate the interruption by non-ROIs. By contrast, in the saliency maps produced by our model, the most salient areas were clear and they could be separated easily from the surroundings. Therefore, our method is advantageous compared with other methods for ROI extraction because it highlights the ROIs with well-defined boundaries as well as effectively eliminating the non-ROIs.

3.2. Quantitative experiment

We used three different evaluation methods in this experiment, i.e., the precision, recall, and F-measure values, as well as the receiver operator characteristic (ROC) curve/area and a set of geometric error indices to ensure a relatively comprehensive comparison.

3.2.1. Precision (P), recall (R), and F-measure (F) values

The quantitative performance evaluation metrics in terms of the overall precision (P), recall (R), and F-measure (F) are defined as follows [26,37],

$$P = \frac{\sum_i \sum_j (g(i,j) \times s(i,j))}{\sum_i \sum_j s(i,j),} \tag{26}$$

$$R = \frac{\sum_i \sum_j (g(i,j) \times s(i,j))}{\sum_i \sum_j g(i,j),} \tag{27}$$

$$F = (1+\alpha)\frac{P \times R}{\alpha \times P + R}, \tag{28}$$

where g is the ground truth, s denotes the binary mask obtained

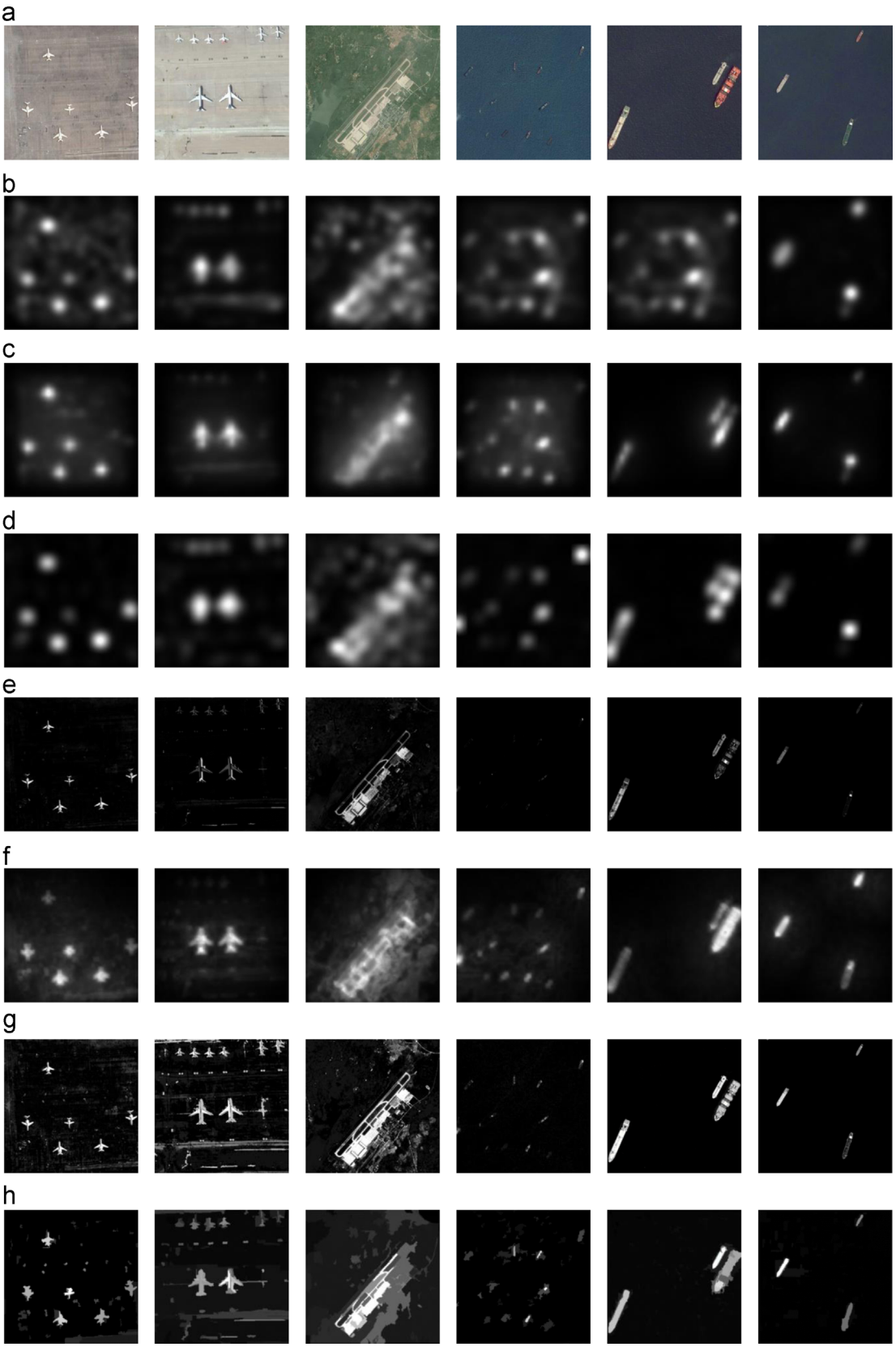


Fig. 9. Visual comparison of saliency maps on the Google Earth set. From top to bottom are (a) original images, saliency maps obtained by (b) IT, (c) GB, (d) SR, (e) FT, (f) CA, (g) HC, (h) RC, (i) WT, (j) RS, (k) FDA, (l)MFF, and (m) ours.

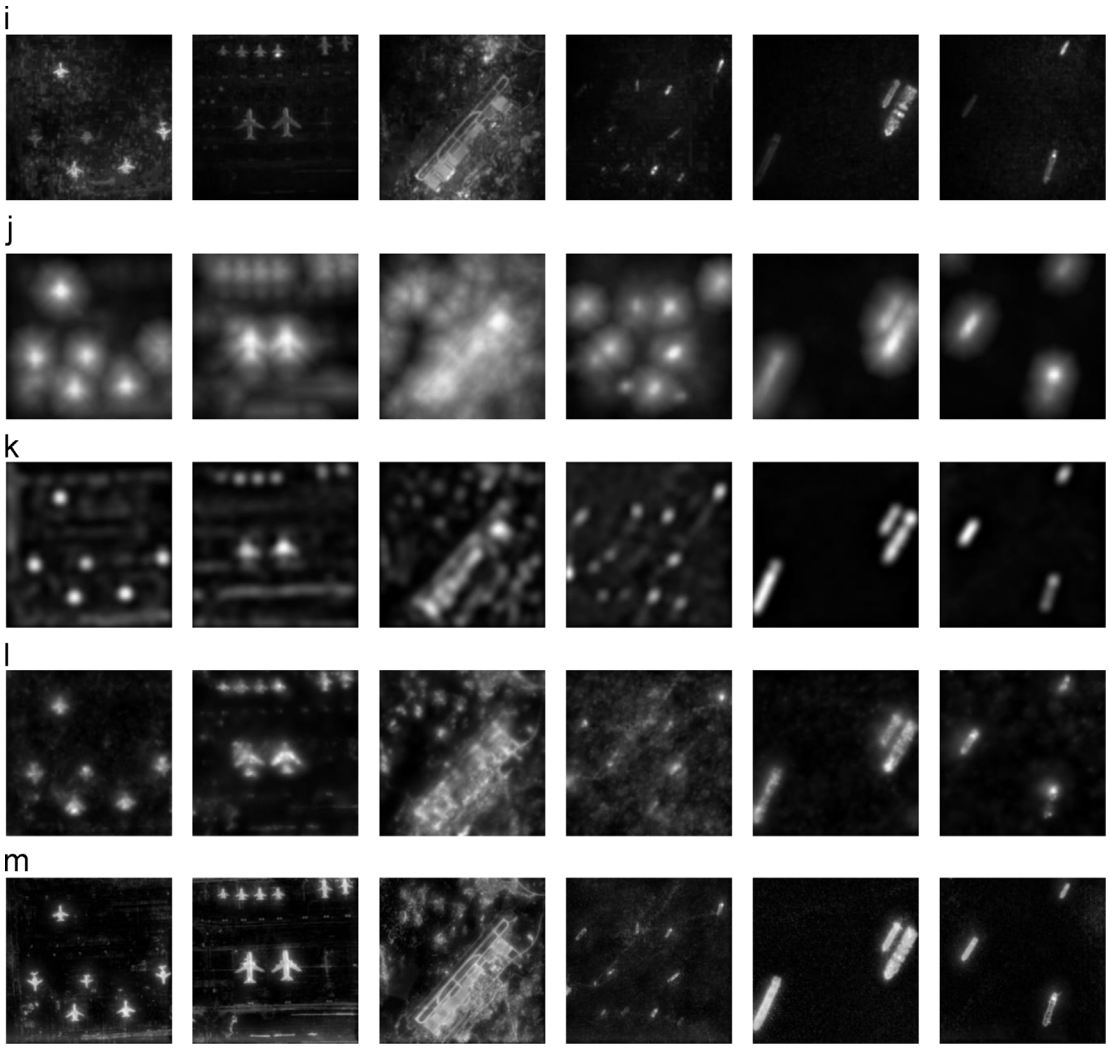


Fig. 9. (*continued*)

by segmenting the saliency map, and α is a positive parameter used to determine the relative importance of the precision compared with the recall when evaluating the F value. In our experiment, α was set to 0.3, as suggested previously [26].

Fig. 11 compares the performance of the existing models and the proposed method based on the Otsu automatic threshold segmentation method. The proposed method obtained the highest P, R, and F values, thereby providing a quantitative demonstration of its reliability.

3.2.2. ROC curve/area

We also objectively evaluated the extraction quality by using the ROC curves. The ordinate and abscissa of a ROC curve represent the true positive rate (TPR) and the false positive rate (FPR), which are defined as below,

$$T' = R, \tag{29}$$

$$F' = \frac{\sum_i \sum_j ([1 - g(i,j)] \times s(i,j))}{\sum_i \sum_j [1 - g(i,j)],} \tag{30}$$

where R is defined by Eq. (27), T' is the TPR value, and F' is the TPR value.

At the same FPR value, a higher TPR value indicates better performance. By contrast, at the same TPR value, a smaller FPR value indicates better performance. The first column in Table 1 lists the ROC areas for the different saliency detection models, where a larger ROC area indicates better performance. As shown in Table 1 and Fig. 12, the proposed model achieved the best performance among all of the approaches.

3.2.3. Geometric error indices

To quantify the geometric accuracy of the ROIs, we employed a set of object-based indices [38,39] to evaluate different geometric properties of the ROIs represented in a saliency map compared with the ground truth. In particular, the edge location, fragmentation, and shape errors are shown in Table 1. Overall, our method obtained the minimum values for the edge location, fragmentation, and shape errors, which suggests that it performed better in terms of these geometric properties.

In summary, in the subjective quality comparison, we first evaluated the proposed method based on the saliency maps and the extracted ROIs. Next, we conducted an objective evaluation based on the precision, recall, F-measure values, ROC curve/area, and a set of geometric error indices for the quantitative experiment. According to these comparisons, the proposed model consistently yielded the most reliable ROI extraction results for high spatial resolution remote sensing images.

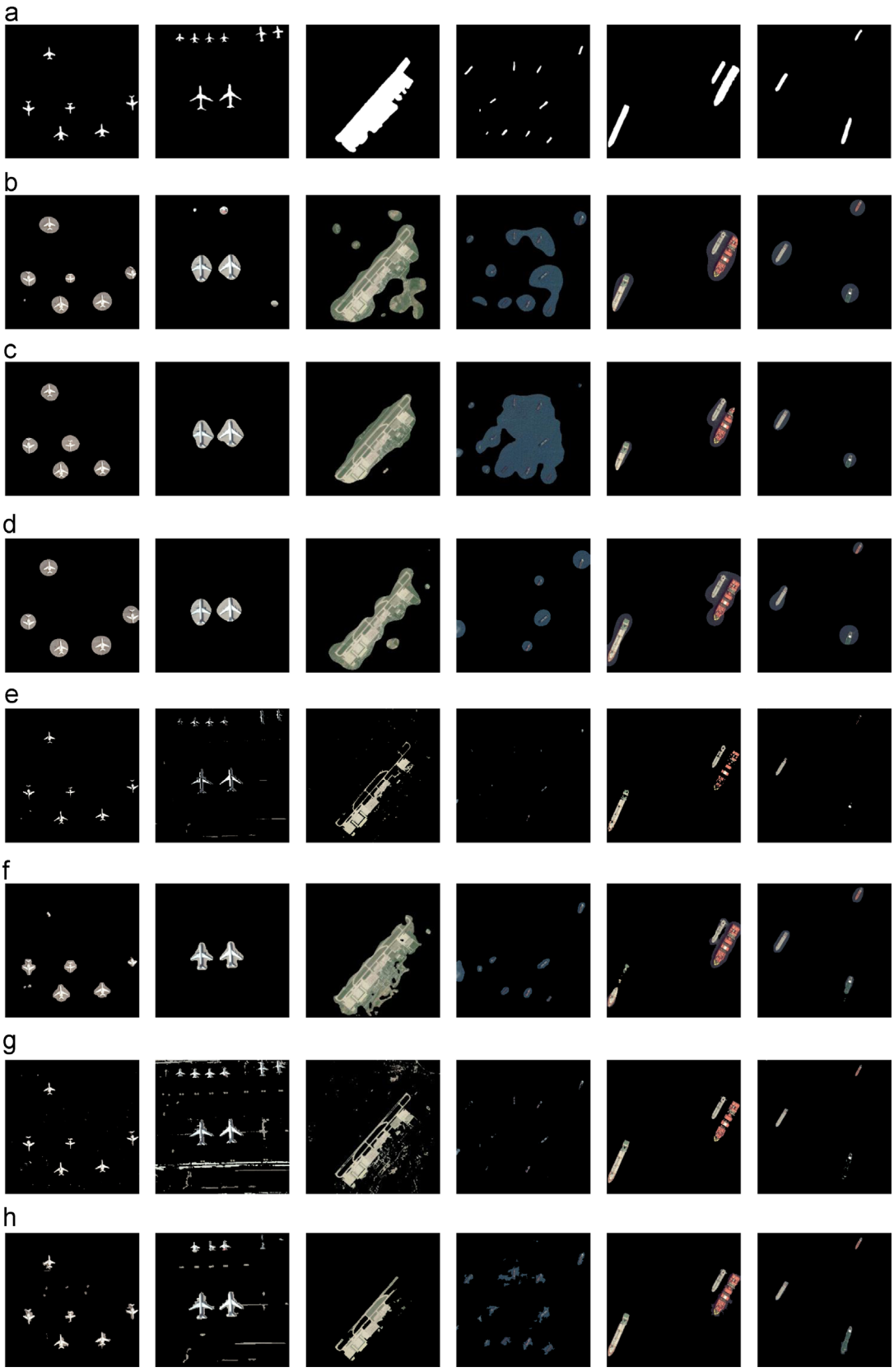


Fig. 10. Visual comparison of ROIs on the Google Earth set. From top to bottom are (a) groundtruth masks, ROIs obtained by (b) IT, (c) GB, (d) SR, (e) FT, (f) CA, (g) HC, (h) RC, (i) WT, (j) RS, (k) FDA, (l)MFF, and (m) ours.

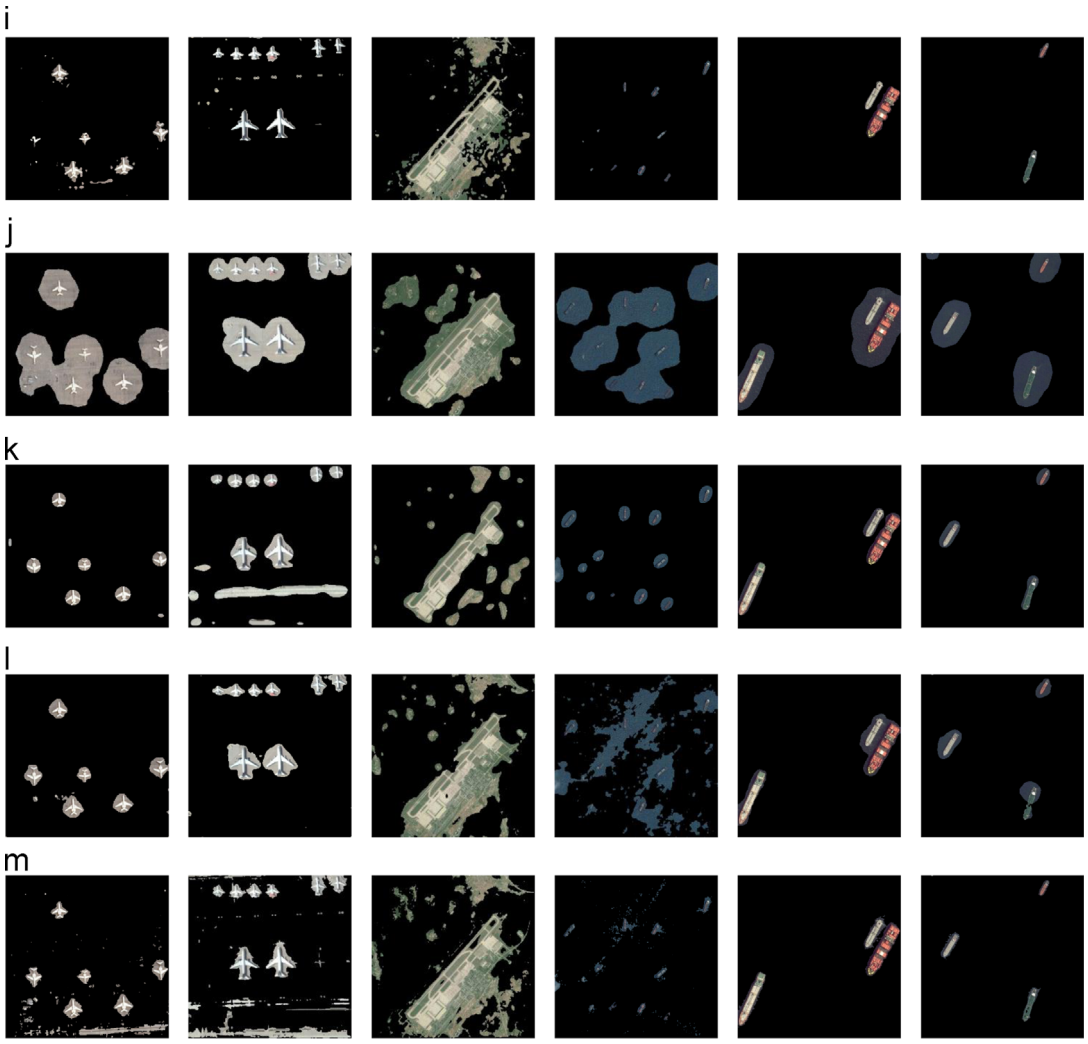


Fig. 10. (*continued*)

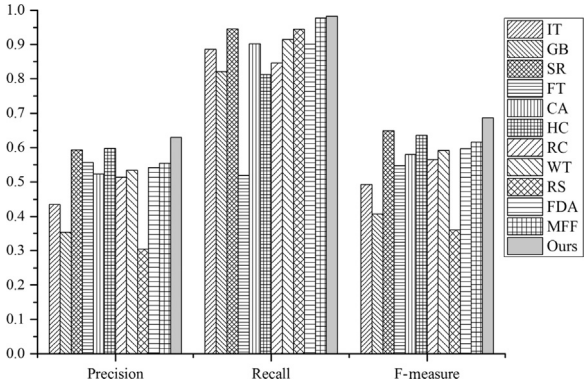


Fig. 11. Precision, recall, and *F*-measure.

Table 1
ROC area results and geometric error indices for different ROI extraction models.

Method	ROC area (%)	Edge-location error (ED) (%)	Fragmentation error (FG) (%)	Shape error (SH) (%)
IT	93.21	100	0	38.05
GB	89.53	100	0	37.44
SR	98.62	100	0	38.01
FT	95.75	92.28	0	28.99
CA	98.11	100	0	38.05
HC	98.28	99.16	0	28.99
RC	98.81	99.31	0	35.53
WT	98.54	100	0	38.05
RS	97.45	100	0	38.05
FDA	98.32	100	0	38.05
MFF	99.11	100	0	38.05
Ours	99.67	89.66	0	19.79

4. Conclusion

In this study, we introduced a novel ROI extraction method for high spatial resolution remote sensing images, which employs the ND-LWT and spectral self-information. The proposed method achieves automatic ROI extraction with high accuracy. High spatial resolution remote sensing images contain complex spatial information, clear details, and well-defined geographic objects, where the structure, edge, and texture information have significant importance. Hence, to make full use of these features, we proposed the novel ND-LWT method to preserve local detail features in the wavelet domain, which is beneficial for generating the edge and texture saliency map. In addition, we utilize the spectral information found in high spatial resolution remote sensing images to improve ROI extraction. Compared with 11 other state-of-the-art models, the proposed extraction algorithm performed better at effectively eliminating the background information and highlighting the ROIs with well-defined boundaries and shapes, thus achieving more accurate ROI extraction.

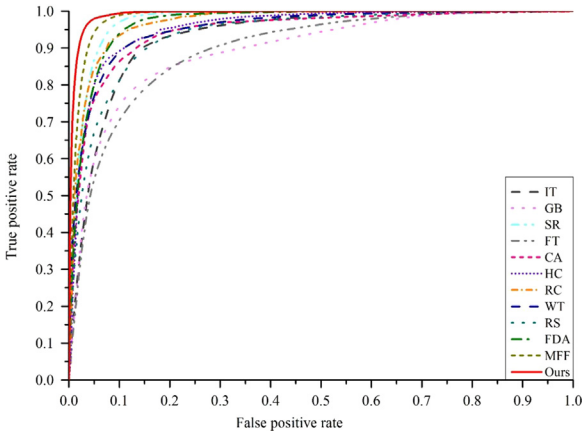


Fig. 12. ROC curves of the proposed model and eleven competing models.

Acknowledgments

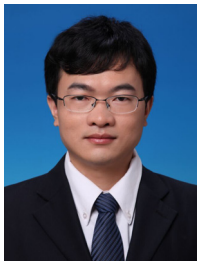
This work was sponsored by the National Natural Science Foundation of China (Nos. 61571050 and 61071103) and by the Fundamental Research Funds for the Central Universities (2012LYB50).

References

[1] B. Chalmond, B. Francesconi, S. Herbin, Using hidden scale for salient object detection, Image Process. IEEE Trans. 15 (2006) 2644–2656.
[2] D. Faur, I. Gavat, M. Datcu, Salient remote sensing image segmentation based on rate-distortion measure, Geosci. Remote Sens. Lett. IEEE 6 (2009) 855–859.
[3] T. Chao, T. Yihua, Z. Zheng-rong, T. Jinwen, Unsupervised detection of built-up areas from multiple high-resolution remote sensing images, Geosci. Remote Sens. Lett. IEEE 10 (2013) 1300–1304.
[4] J. Sun Y. Wang Z. Zhang Y. Wang. Salient region detection in high resolution remote sensing image, in: Proceedings of the Wireless and Optical Communications Conference (WOCC), 2010, pp. 1–4.
[5] L. Mai Y. Niu F. Liu Saliency aggregation: a data-driven approach, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (CVPR), 2013, pp. 1131–1138.
[6] S. Le Moan, A. Mansouri, J.Y. Hardeberg, Y. Voisin, Saliency for spectral image analysis, selected topics in applied earth observations and remote sensing, IEEE J. 6 (2013) 2472–2479.
[7] L. Zhang, K. Yang, Region-of-interest extraction based on frequency domain analysis and salient region detection for remote sensing image, Geosci. Remote Sens. Lett. IEEE 11 (2014) 916–920.
[8] L. Zhang, K. Yang, H. Li, Regions of interest detection in panchromatic remote sensing images based on multiscale feature fusion, selected topics in applied earth observations and remote sensing, IEEE J. 7 (2014) 4704–4716.
[9] L. Zhang, B. Qiu, X. Yu, J. Xu, Multi-scale hybrid saliency analysis for region of interest detection in very high resolution remote sensing images, Image Vis. Comput. 35 (2015) 1–13.
[10] Z. Li, L. Itti, Saliency and gist features for target detection in satellite images, Image Process. IEEE Trans. 20 (2011) 2017–2029.
[11] J.H. Reynolds, R. Desimone, Interacting roles of attention and visual salience in V4, Neuron 37 (2003) 853–863.
[12] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1998) 1254–1259.
[13] X.D. Hou, L.Q. Zhang, Saliency detection: a spectral residual approach, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007, pp. 2280–2287.
[14] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2009, pp. 1597–1604.
[15] R. Achanta, S. Susstrunk, Saliency detection using maximum symmetric surround, in: Proceedings of the IEEE International Conference on Image Processing, ICIP, 2010, pp. 2653–2656.
[16] C. Koch, S. Ullman, Shifts in selective visual attention: towards the underlying neural circuitry, in: L. Vaina (Ed.) Matters of Intelligence, Netherlands, 1987, pp. 115–141.
[17] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, Pattern Anal. Mach. Intell. IEEE Trans. 20 (1998) 1254–1259.
[18] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, Adv. Neural Inf. Process. Syst. (2006) 545–552.
[19] S. Goferman, L. Zelnik-Manor, A. Tal, Context-aware saliency detection, Pattern Anal. Mach. Intell. IEEE Trans. 34 (2012) 1915–1926.
[20] M.-M. Cheng, G.-X. Zhang, N.J. Mitra, X. Huang, S.-M. Hu., Global contrast based salient region detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2011, pp. 409–416.
[21] N. Bruce, J. Tsotsos, Saliency based on information maximization, Adv. Neural Inf. Process. Syst. 18 (2006) 155.
[22] F. Perazzi, P. Krahenbuhl, Y. Pritch, A. Hornung, Saliency filters: contrast based filtering for salient region detection, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2012, pp. 733–740.
[23] X. Hou, L. Zhang, Saliency detection: a spectral residual approach, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[24] C. Guo, Q. Ma, L. Zhang, Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.
[25] N. Murray, M. Vanrell, X. Otazu, C.A. Parraga, Saliency estimation using a non-parametric low-level vision model, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2011, pp. 433–440.
[26] N. Imamoglu, W. Lin, Y. Fang, A saliency detection model using low-level features based on wavelet transform, Multimedia IEEE Trans. 15 (2013) 96–105.
[27] L. Zhang, M.H. Tong, T.K. Marks, H. Shan, G.W. Cottrell, SUN: a Bayesian framework for saliency using natural statistics, J. Vis. 8 (2008) 1–20 32 3.
[28] X. Wang, B. Wang, L. Zhang, Airport detection in remote sensing images based on visual attention, Neural Inf. Process. (2011) 475–484.
[29] Z. Ding, Y. Yu, B. Wang, L. Zhang, An approach for visual attention based on biquaternion and its application for ship detection in multispectral imagery, Neurocomputing 76 (2012) 9–17.
[30] W. Sweldens, The lifting scheme: a custom-design construction of biorthogonal wavelets, Appl. Comput. Harmon. Anal. 3 (1996) 186–200.
[31] I. Daubechies, W. Sweldens, Factoring wavelet transforms into lifting steps, J. Fourier Anal. Appl. 4 (1998) 247–269.
[32] O.N. Gerek, A.E. Cetin, A 2-D orientation-adaptive prediction filter in lifting structures for image coding, Image Processing IEEE Trans. 15 (2006) 106–111.
[33] W. Ding, F. Wu, X. Wu, S. Li, H. Li, Adaptive directional lifting-based wavelet transform for image coding, Image Process. IEEE Trans. 16 (2007) 416–427.
[34] D.L. Donoho, J.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 81 (1994) 425–455.
[35] J.M. Shapiro, Embedded image coding using zerotrees of wavelet coefficients, Signal Process. IEEE Trans. 41 (1993) 3445–3462.
[36] N. Otsu, A. Threshold, Selection method from gray-level histograms, Syst., Man Cybernet. IEEE Trans. 9 (1979) 62–66.
[37] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, H.-Y. Shum, Learning to detect a salient object, Pattern Anal. Mach. Intell. IEEE Trans. 33 (2011) 353–367.
[38] L. Bruzzone, C. Persello, A novel protocol for accuracy assessment in classification of very high resolution multispectral and SAR images, in: Proceedings of the IEEE International Symposium on Geoscience and Remote Sensing, 2008, pp. II-265–II-268.
[39] C. Persello, L. Bruzzone, A novel protocol for accuracy assessment in classification of very high resolution images, Geosci. Remote Sens. IEEE Trans. 48 (2010) 1232–1244.

Libao Zhang: He received the B.S. degree in electronic and information engineering, the M.S. degree in signal and information processing, and the Ph.D. degree in communication and information system from Jilin University, Changchun, China, in 1999, 2002, and 2005, respectively. Since August 2005, he has been with the College of Information Science and Technology, Beijing Normal University, Beijing, China, where he is currently an Associate Professor. He has taken charge of three National Natural Science Foundations of China. His research interests include remote sensing image processing, image compression, saliency analysis and object recognition.

Jie Chen: She received the B.S. degree in electronic science and technology from Beijing Normal University, Beijing, China, in 2014. She is now persuing the M.S. degree in communication and information system at Beijing Normal University. Her research interest lies in modeling biologically-plausible computational visual attention and object detection.



Bingchang Qiu: He received the B.S. degree in electronic science and technology and the M.S. degree in communication and information system from Beijing Normal University, Beijing, China, in 2012 and 2015 respectively. His primary research interest is in the field of remote sensing image processing and directional lifting wavelet transform.