Region of Interest Extraction based on Hybrid Salient Detection for Remote Sensing Image

J. Maheswarreddy, S.A.K. Jilani

Abstract: Remote sensing images have huge amount of information in it due to use of high resolution cameras and sensors. Region of interest (ROI) is defined as the regions which draw the attention of viewer at first sight and they are the focal point of the image. ROI selection in remote sensing images allows the viewer to search for specific objects in the region. Traditional approaches for ROI selection are computationally complex and inaccurate. In this work, a hybrid approach which combines the best of frequency domain analysis and Super pixel based spatially weighted intensity contrasting is proposed for selecting the ROI in remote sensing images. Compared to previous methods the proposed hybrid ROI selection is able to extract the ROI accurately.

Keywords: ROI, Saliency Map, Gaussian Pyramid, Frequency Domain Analysis, Quaternion.

I. INTRODUCTION

Remote sensing image processing has gained increased attention due to many applications in areas of urban monitoring, fire detection or flood prediction, deforestation, crop monitoring, weather prediction, land use mapping and land cover mapping. With technological advancements in high resolution satellite cameras and sensors, remote sensing images with high resolution are common. The remote sensing images have huge depth of information in it. Lot of computing resources is needed to process those images and extract knowledge from it. ROI aids in this processing by selecting the focus points in the image, so that computing can be focused on those ROI regions. Typical ROI in remote sensing images include residential areas, airports, airplanes, wharfs, roads and ships. Compared with the background, these ROI regions have salient features that immediately grab human attention. The salient characteristics of ROI area is as follows

1. Vast amount of complex structural information , edge and texture information (typical residential areas)
2. Unique shape information like ships, aero planes.
3. Information about orientation.
4. Distinct spectra compared to surrounding environment.

ROI regions posses this characteristics when compared to background. Due to this high contrast stimuli are generated in the receptive fields of human visual system. Due to this property of saliency, saliency based ROI extraction has become research hotspot in recent years.

Saliency based ROI segmentation was originally designed for natural scene images. It worked by utilizing the intensity, color, orientation, texture, and other low-level features to determine contrast for saliency computation. The saliency methods worked by obtaining saliency map based on intensity, color and orientation. From saliency map, thresholding is done to get the binary image and fusion is done on original image to get the ROI.

Existing ROI extraction methods based on saliency map can be broadly classified to biological based, pure computation based and combination based. Biology based methods determine the center-surround contrast using a Gaussian pyramid on three image features: color of the pixel, intensity level of the pixel, and orientations in the image. These methods are sensitive to the center of images and fail to some extent when the ROIs are close to the image boundary. Pure computation based methods extract ROIs by transforming the original image into frequency domain and making spectrum analysis on them. These methods cannot provide a well defined ROI boundary and the ROI obtained by these methods have attenuated interior. Combination based methods use mix of biological and computation methods.

In this work a hybrid technique combining the best of frequency domain analysis in spatial domain and spatially weighted intensity contrast in pixel domain is used to generate individual saliency maps. Using QWT fusion, final saliency map is generated. On the fused salient map, OSTU thresholding is applied to get the binary mask. Binary mask is applied on the input color image to get the ROI.

II. RELATED WORK

The existing solutions for ROI extraction model using saliency analysis is analyzed in this section.

In [1] author proposed a saliency computation model. In this model saliency maps are obtained based on the intensity, color and orientation channels. From these individual maps, final master saliency map is obtained by combining these three conspicuity maps based on center surround differences. ROIs extracted by this model only have the general outline of objects but lose the well-defined boundary.

In [2] authors proposed approach based on statistical signature of targets to detect and classify targets in high-resolution satellite images. Signature of target is created using biologically-inspired low-level visual features.

Manuscript received January 25, 2019.

J. Maheswarreddy, Research Scholar, Department of ECE, Rayalaseema University, Kurnool, A.P, India.
Dr.S.A.K. Jilani, Professor, Department of ECE, Madanapalli Institute of Technology & Science, A.P. India.

The image is cut into small image patches and verified for presence of salient targets using Support Vector Machine (SVM) classifier. The approach works only for a trained target and it is not generic. Also computation complexity is high in this approach.

In [3] author proposed a novel saliency detection model by using wavelet transforms.. Wavelet transform is applied on image to create the feature maps representing different features from edge to texture. From the features computational model is proposed to generate the saliency map. This approach cannot provide full-resolution saliency maps, resulting in the loss of a well-defined ROI boundary.

In [4] author proposed a multiple detection framework for locating saliency. In each step, blocks of pixels are selection and manifold analysis is done for saliency score. Probable anomalous pixels are filtered and the remaining pixels are sent for next iteration. Final detection result is then obtained by fusion of all intermediate results. The ROI boundary is not well defined in this procedure.

In [5] author proposed a new bottom up visual saliency model called Graph-Based Visual Saliency. It has two steps – first a activation map is formed based on certain feature channels and normalizing them in a way which highlights conspicuity. The computational complexity is high in this approach and ROI accuracy is not high in this approach.

In [6] author proposed saliency detection based on Shannon's entropy measure. Image is split to patches and Shannon measure is calculated on each patch to select the likelihood of patch being salient. The product of such likelihoods yields the joint likelihood of the entire salient image. The computation complexity is very high.

In [7] authors proposed a Hierarchical Saliency Detection model to solve the problems due to small-scale high-contrast patterns affecting the accuracy of Saliency detection. First, three image layers of different scales are extracted from the input. Saliency cues are computed for each layer. They are finally fused into one single map using a graphical model. Though the approach works fine for natural images, it does not give accurate results for remote sensing images.

In [8] authors proposed ROI detection algorithm based on an adaptive visual attention model. The original image is filtered and then sub sampled to a specific level of the Gaussian pyramid in accordance with the original resolution. Features from image extracted using discrete moment transform (DMT) are used to get the edge description. Region growing strategy is employed based on the edge description to get the ROI. The solution is a over fit and has higher false positives in ROI extraction.

In [9] author proposes a faster and more efficient ROI detection algorithm based on fusion of two channel information from intensity and orientation.

Intensity saliency is obtained using multi scale spectrum residuals. Orientation saliency map is obtained by using interpolating bi-orthogonal integer wavelet transform (IB-IWT) with thresholding and filtering. Fusion using scale fusion at different scale is proposed to get the final saliency map. Holes in ROI cannot be avoided in this approach.

In [10] author proposed ROI extraction using frequency domain analysis based salient region detection. RSB image is converted to HSI space and quaternion transform is done to covert to image in spatial domain. After filtering on spatial domain, inverse transform is done to generate the image and Gaussian pyramid is got for that image using successive convolution. From the Gaussian pyramid adaptive thresholding and fusion is done to get the ROI. This approach is computation wise fast when compared to previous solutions but ROI holes appears for certain images where ROI boundaries are not clear.

In [11] author proposed saliency detection based on nonparametric regression framework. It is designed for noisy images. It calculates saliency based on dissimilarities in pixel on center path in patches. The accuracy is not high in this approach and computation complexity is high.

In [12] author proved that aggregating various saliency analysis methods provides better results than individual methods. Authors used Conditional Random Field model to aggregate individual saliency maps and interaction between neighboring pixels. The dependence of individual maps on aggregation was modeled as machine learning part, so the approach could work for all images.

In [13] author proposed a regression based solution for saliency map computation. Supervised learning algorithm is used to map the region features to saliency score. For images, saliency score is computed for all region features and the multi layer score are fused to get the saliency map. The solution is noise intolerant and the computation time is high in this approach.

In [14] authors proposed a unsupervised approach to detect built up areas in the remote sensing images. Corners are located using Harris corner locator and using likelihood estimation of region from corners built up area ROI is extracted. For images with multiple build up region, unsupervised grouping is done using spectrum clustering and graph cuts. The approach is computationally complex and can work for natural objects in remote sensing images.

In [15] author proposed an approach based on detection of statistical signature of target for segmenting the targets in remote sensing image. Low level visual features are extracted from training images of target and a SVM is trained to classify the targets. Input image is split to patches, and visual features extracted from the patches are given to SVM to decide the feasibility of patch to contain the target. The approach cannot work if target is shadowed or partly visible. The approach is not generic and target specific.

In [16] author proposed a bio inspired model based on sparse coding to extract information about low level saliency in remote sensing images. Discrete Cosine Transform (DCT) and wavelets needed to be executed on training images to create the dictionary for spare coding. Saliency can be obtained in a faster way using this approach but the ROI extracted using the saliency map lacks well defined boundaries.

In [17] author proposed a new saliency model in terms of colorimetric attributes in spectral domain of image. The spectral images is subjected to multiple feature extraction methods like PCA, Gabor filter and after sub sampling, center surround comparison is done to identify salient region based on each features and results fused to a saliency map.

The computational complexity in high in this approach and ROI extraction accuracy is low.

In [18] author proposed content aware saliency map detection based on the principle salient areas have distinctive colors or patterns. The approach needs support from context dependent detection algorithms like face detection etc to extract the ROI. Without the supporting detection algorithms the ROI extracted does not have clear defined boundaries.

III. ROI EXTRACTION BASED ON HYBRID SALIENCY DETECTION

The framework of the proposed hybrid solution for ROI extraction is given below

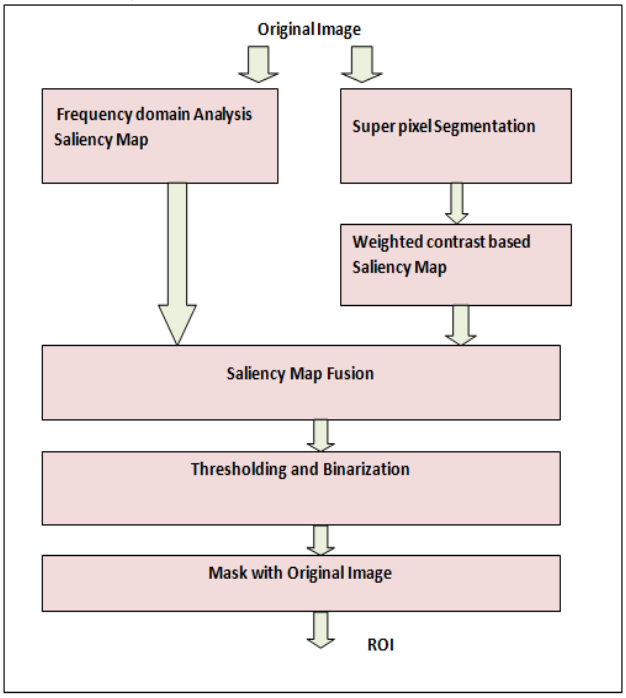


Figure 1: Hybrid Saliency Detection

From the original image, saliency maps are generated from Frequency domain analysis and weighted contrast. Both the saliency maps are fused to get the final saliency map. Thresholding is done on the final saliency map to get the binary mask. The mask is applied onto the original image to get the ROI.

A. Frequency Domain Analysis

The input image is sub sampled to half its original size as processing on original size would increase the computational cost. HSI transform is applied on the sub sampled RGB image to get Hue, Saturation and Intensity images (HSI). Due to its consistency with human visual system than RGB color space, HIS is used for frequency domain analysis.

The RGB values are normalized in range of 0 to 1 before converting to HSI.

Total = (R+G+B)
*R’ = R/total * 255*
G’ = G/total 255*
B’ = B/total 255*

Where R’, G’.B’ are the normalized values for the pixel.

HSI transform on RGB color image is done as

$$H = \begin{cases} \theta & B \le G \\ 360 - \theta & B > G \end{cases}$$

$$S = 1 - \frac{3}{(R+G+B)}[\min(R,G,B)].$$

$$I = \frac{1}{3}(R+G+B).$$

Where θ is given as

$$\theta = \arccos\{((1/2)[(R-G)+(R-B)])/([(R-G)^2+(R-B)(G-B)]^{1/2})\}.$$

Where R, G, B are the pixel values for Red, Green and Blue component of the image.

The HSI image can be represented in quaternion form as

$$f(n,m) = H(n,m)\mu_1 + S(n,m)\mu_2 + I(n,m)\mu_3$$

Where (n,m) is the location of the pixel. H(n,m),S(n,m),I(n,m) are Hue, Saturation , Intensity of pixel at (n,m) . The value of μ is selected as per the condition.

$\mu 3 = \mu 1\mu 2$, $\mu 1 \perp \mu 2$, $\mu 2 \perp \mu 3$, $\mu 3 \perp \mu 1$.

The quaternion Fourier transform is done for pixel at (u,v) is done as

$$F(u,v) = F_1(u,v) + F_2(u,v)\mu_2$$

$$F_i(u,v) = \frac{1}{\sqrt{MN}}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} e^{-\mu_1 2\pi((mv/M)+(nu/N))} f_i(n,m).$$

The edges and other sharp transitions in the pixels of an image contribute significantly to the high frequency content of its Fourier transform. Hence, smoothing of a image is achieved in the frequency domain by attenuating the specified range of high frequency components in the quaternion Fourier transform of the image. Gaussian Quaternion high pass filter is applied on the quaternion Fourier transformed results. The two dimensional Gaussian high pass filter transfer function with the cut-off frequency at a distance D_0 is given by

$$H_q(u,v) = 1 - e^{-\frac{D^2(u,v)}{2\sigma^2}}$$

where σ is a measure of Gaussian spread and D(u,v) is the distance of point (u,v) to the origin of frequency rectangle (M/2,N/2). It is calculated as

$$D(u,v) = \left[\left(u - \frac{M}{2}\right)^2 + \left(v - \frac{N}{2}\right)^2\right]^{\frac{1}{2}}$$

The results obtained with Gaussian high pass filtering are smoother than with ideal high pass filters. After Gaussian high pass filtering, inverse quaternion transform is done to generate a output image. The inverse transform is done as

$$f_i(n,m) = \frac{1}{\sqrt{MN}}\sum_{v=0}^{M-1}\sum_{u=0}^{N-1} e^{\mu_1 2\pi((mv/M)+(nu/N))} F_i(u,v).$$

The output image obtained after inverse transform is the frequency domain saliency map.

B. Weighted Contrast based Saliency

Super pixels are the clusters of spatially connected pixel with similar properties in terms of color or intensity. Super pixel segmentation segments images to portions with natural boundaries preserved. For this work, super pixel segmentation is done using the most popular method simple linear iterative clustering (SLIC). SLIC has been choosed due to following factors

1. It is more efficient for segmentation
2. Enforces compactness and regularity in super pixel shapes
3. Seamless working for both gray scale and color images.

Super pixel segments of salient regions show noticeable color contrast with other segments and the spatial distribution of the salient region segments are sparser compared to other segments. Based on these two observation weighted contrast for measuring the saliency of the super pixel segments is formulated as

$$WC = \sum_{j=1}^{n} W(i,j)\cdot \left\| mc_i - mc_j \right\|$$

The weight W is calculated based on the area of super pixel segment and spatial similarity as

$$W(i,j) = \left| SP_j \right| \cdot Sim_d(i,j)$$

Super pixel segments are clustered based on the WC values to get the weighted contrast saliency map.

C. Saliency Map Fusion

Image fusion of salient map generated using frequency domain analysis and weighted contrast is done using Quaternion Wavelet Transform (QWT). QWT is done on each saliency images to obtain the low frequency coefficients and high frequency coefficients. Average fusion is proposed to fuse the low frequency sub bands. Fusion rule based on maximum value of energy of coefficients is proposed to fuse the high frequency sub bands. Inverse QWT is done to get the final fused image.

QWT of image f(x,y) can be defined as

$$f(x,y) = A_n^q f(x,y) + \sum_{s=1}^{n} \left[D_{s,1}^q f(x,y) + D_{s,2}^q f(x,y) + D_{s,3}^q f(x,y) \right]$$

Where $A_s^q f(x,y)$ is the low frequency is band and $D_{s,p}^q f(x,y)$ is the high frequency band of the image.

After QWT is applied on the image a low frequency part and n groups of high frequency parts are obtained.

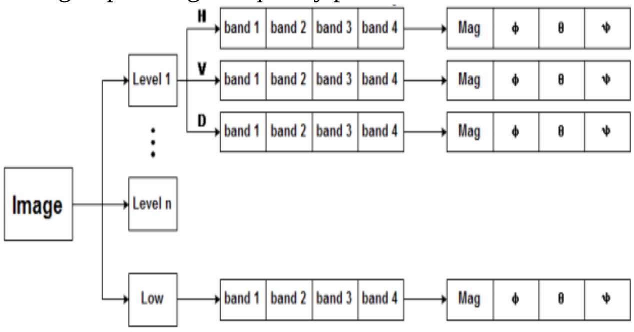


Figure 2: QWT Coefficients of Image

The average fusion rule for fusing the low frequency bands is given as average of the coefficients pair wise between the Low frequency coefficients of two saliency images. The fusion rule for fusing the high frequency sub bands is given as selecting the maximum value of coefficient between the pair wise high frequency sub bands. Final fused image is obtained by inverse QWT on the fused low frequency and high frequency coefficients.

D. Thresholding and Binarization

OSTU Thresholding is done on the fused saliency map to identify the threshold for generating the binary image. OSTU method chooses the threshold to minimize the intra class variance of the thresholded black and white pixels. The intra class variance is given as

$$\sigma_w^2(t) = q_1(t)\sigma_1^2(t) + q_2(t)\sigma_2^2(t)$$

The class probabilities q1 and q2 is evaluated as

$$q_1(t) = \sum_{i=1}^{t} P(i) \qquad q_2(t) = \sum_{i=t+1}^{I} P(i)$$

The individual class variance is given as

$$\sigma_1^2(t) = \sum_{i=1}^{t} [i - \mu_1(t)]^2 \frac{P(i)}{q_1(t)}$$

$$\sigma_2^2(t) = \sum_{i=t+1}^{I} [i - \mu_2(t)]^2 \frac{P(i)}{q_2(t)}$$

Where the class mean is calculated as

$$\mu_1(t) = \sum_{i=1}^{t} \frac{iP(i)}{q_1(t)} \qquad \mu_2(t) = \sum_{i=t+1}^{I} \frac{iP(i)}{q_2(t)}$$

OSTU method iterates the t value from 1 to 256 and chooses the t value which minimizes the intra class variance. Once threshold T is selected, binary mask is generated using rule.

$$g(x,y) = \begin{cases} 1 & if\ f(x,y) >= T \\ 0 & otherwise \end{cases}$$

E. Masking with Original Image

The binary mask is multiplied to the original image to generate the ROI image. There is no need to do any region growing procedure as post processing as the obtained ROI image is free from the fragments and intra regional black spots.

IV. RESULTS

Performance evaluation of the proposed hybrid solution is done using select remote sensing images. The performance of proposed solution is compared to FDA-SRD model [10].

The methods are tested against following high resolution remote sensing images.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Figure 3: Test-1

Figure 4: Test-2

Figure 5: Test-3

Test-1and Test2 were taken by the SPOT5 satellite which offers a resolution of 2.5 m.

Test-3 was taken by the GeoEye-1 which provides a resolution of 1 m. Rural residential areas are defined as ROI in the test images and this must be detected by the methods. The ROI regions in these test images have one or more following characteristics.

1. rich edge and texture features
2. the brightness area
3. Color highlighting area.

For each of the test images ground truth images are set based on the ROI to given as result.

The performance of the methods is compared in terms of accuracy and structural similarity(SSIM) to the ground truth image.

The results for Test-1 is below

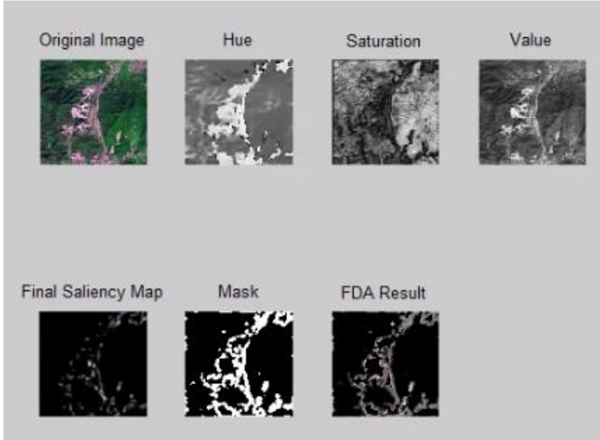


Figure 6: Frequency analysis saliency result

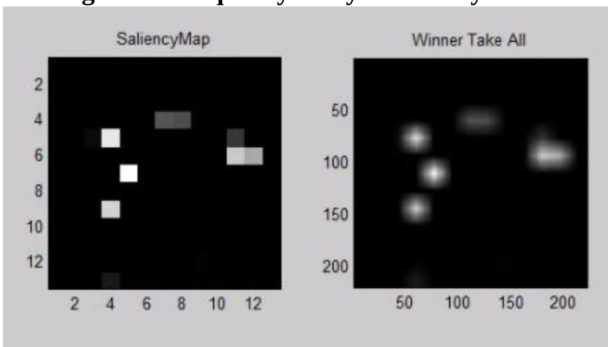


Figure 7: Weighted contrast saliency result

The ROI extracted from the integrated solution

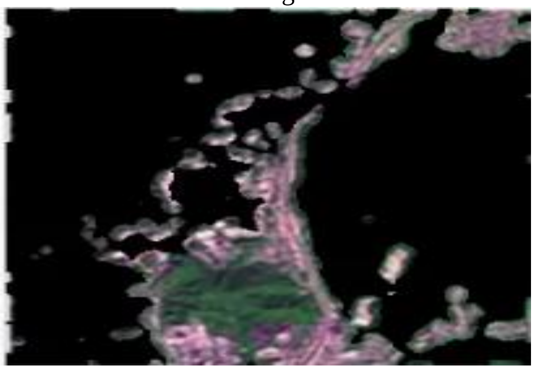


Figure 8: ROI from hybrid

The results for Test-2 is below

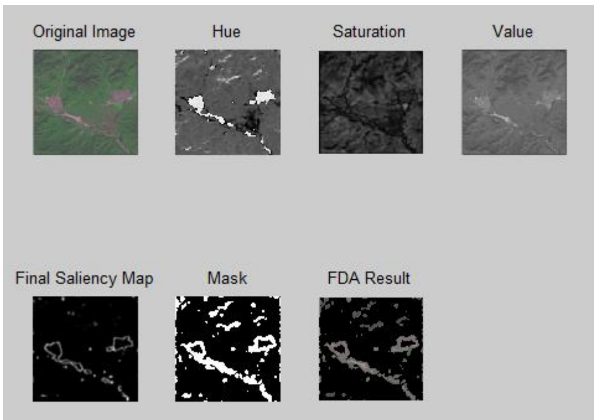


Figure 9: Frequency analysis saliency result

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

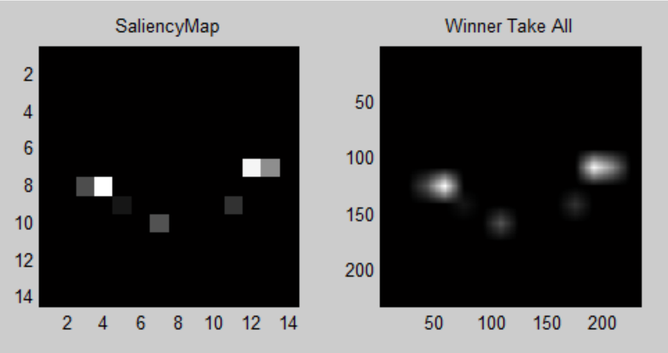


Figure 10: Weighted contrast saliency result

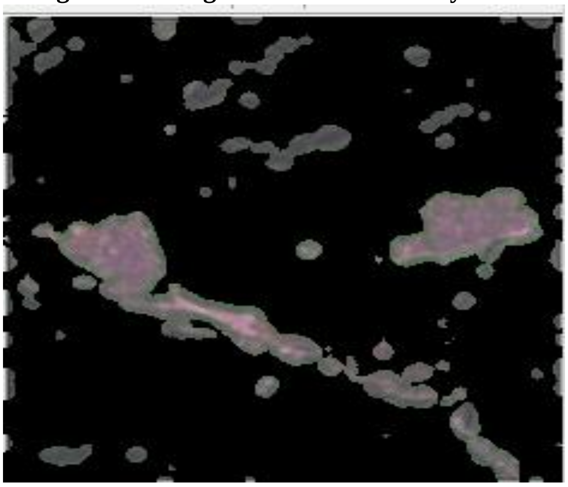


Figure 11: ROI from hybrid

The results for Test-3 is below

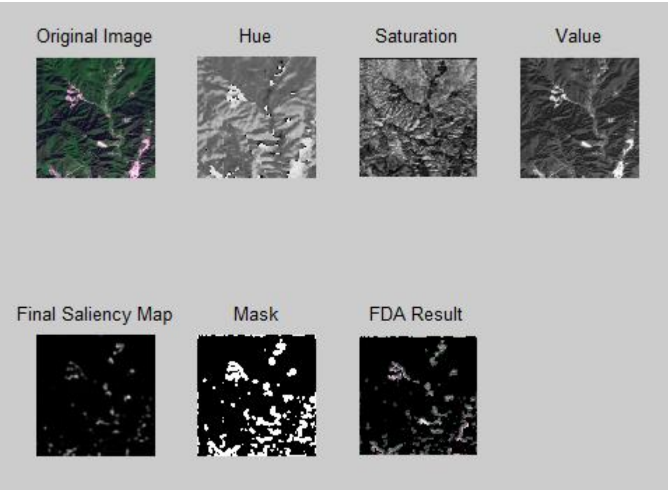


Figure 12: Frequency analysis saliency result

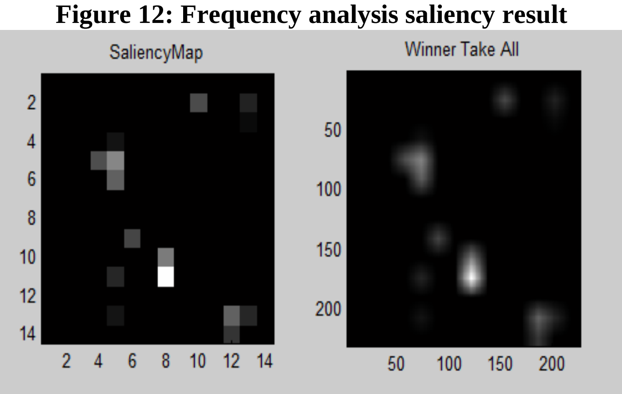


Figure 13: Weighted contrast saliency result

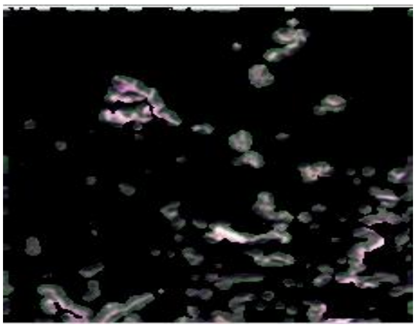


Figure 14: ROI from hybrid

Minor modifications were made to 3 test images and 10 images are generated. Performance in terms of accuracy and SSIM to the ground truth is measured and given below

Table 1: SSIM Comparison

Sample	SSIM in Hybrid	SSIM in FDA-SRD
1	77.97	76
2	72.94	67.73
3	78.31	76.2
4	87.95	86
5	86.30	85
6	85.22	84.11
7	76.93	75.92
8	73.71	72.95
9	85.95	84.92
10	86.44	85.26

From the results, it can be seen that the Hybrid solution results were much closer to ground truth compared to FDA-SRD.

The accuracy is measured in terms of three parameters

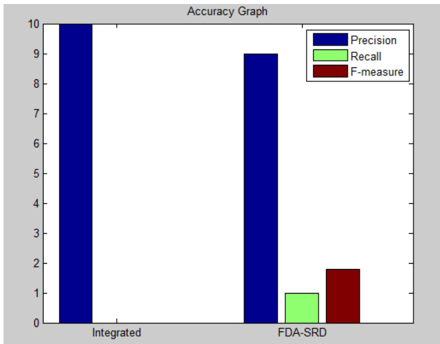
1. Precision
2. Recall
3. F-Measure

Precision is defined as more than 90% closeness to ground truth.

Recall is defined as less than 90% closeness to ground truth.

F-Measure is calculated as

$$F = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$



For the testing with 10 images, all 10 images were found to have more than 90% match to ground truth in the proposed solution making precision as 10 and recall as 0. In FDA-SRD the prevision is 9.

V. CONCLUSION

Hybrid saliency detection based ROI extraction method is proposed and validated in this work. The input image is subjected to two independent models for saliency detection. First model is based on spectral characteristics. Frequency domain analysis is done of down sampled image in HSI space. By applying Quaternion transform on HSI image and Gaussian Quaternion high pass filters noises removed and edge preservation done. The inverse quaternion resulted in frequency domain salient map. Second model is based on analyzing contrast in the super pixel segmented image based on color. The second model generated weighted contrast saliency map. Quaternion based fusion is done for both saliency maps in spectral domain. The fused image is thresholded to get the binary mask. The binary mask is multiplied to original image to get the ROI. The ROI results were compared to ground truth image to measure accuracy and structural similarity. The results proved that the proposed hybrid solution produced more accurate and satisfying results.

REFERENCES

1. L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
2. Z. Li and L. Itti, “Saliency and gist features for target detection in satellite images,” IEEE Trans. Image Process., vol. 20, no. 7, pp. 2017–2029, Jul. 2011.
3. N. Imamoglu, W. Lin, and Y. Fang, “A saliency detection model using low-level features based on wavelet transform,” IEEE Trans. Multimedia, vol. 15, no. 1, pp. 96–105, Jan. 2013.
4. B. Du and L. Zhang, “Target detection based on a dynamic subspace,” Pattern Rec fognit., vol. 47, no. 1, pp. 344–358, 2014.
5. J. Harel, C. Koch, and P. Perona, “Graph-based visual saliency,” Adv. Neural Inf. Process. Syst., vol. 19, pp. 545–552, 2007.
6. N. Bruce and J. Tsotsos, “Saliency based on information maximization,” Adv. Neural Inf. Process. Syst., vol. 18, pp. 155–162, 2006.
7. Q. Yan, L. Xu, J Shi, and J Jia, “Hierarchical saliency detection,” in Proc. IEEE Conf. Comput. Vision Pattern Recognit., 2013, pp. 1155–1162.
8. L. Zhang, H. Li, P. Wang, and X. Yu, “Detection of regions of interest in a high-spatial-resolution remote sensing image based on an adaptive spatial subsampling visual attention model,” GISci. Remote Sens., vol. 50, no. 1, pp. 112–132, 2013.
9. L. Zhang, K. Yang, and H. Li, “Regions of interest detection in panchromatic remote sensing images based on multiscale feature fusion,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 7, no. 12, pp. 4704– 4716, Dec. 2014.
10. L. Zhang and K. Yang, “Region-of-interest extraction based on frequency domain analysis and salient region detection for remote sensing image,” IEEE Geosci. Remote Sens. Lett., vol. 11, no. 5, pp. 916–920, May 2014.
11. C. Kim and P. Milanfar, “Visual saliency in noisy images,” J. Vision, vol. 13, no. 4, pp. 1–14, 2013.
12. L. Mai, Y. Niu, and F. Liu, “Saliency aggregation: A data-driven approach,” in Proc. IEEE Conf. Comput. Vision Pattern Recognit., 2013, pp. 1131–1138.
13. H. Jiang et al., “Salient object detection: A discriminative regional feature integration approach,” in Proc. IEEE Conf. Comput. Vision Pattern Recognit., 2013, pp. 2083–2090.
14. C. Tao, Y. Tan, Z. R. Zou, and J. Tian, “Unsupervised detection of builtup areas from multiple high-resolution remote sensing images,” IEEE Geosci. Remote Sens. Lett., vol. 10, no. 6, pp. 1389–1393, Nov. 2013.
15. Z. Li and L. Itti, “Saliency and gist features for target detection in satellite images,” IEEE Trans. Image Process., vol. 20, no. 7, pp. 2017–2029, Jul. 2011.
16. I. Rigas, G. Economou, and S. Fotopoulos, “Low-level visual saliency with application on aerial imagery,” IEEE Geosci. Remote Sens. Lett., vol. 10, no. 6, pp. 1389–1393, Nov. 2013.
17. S. L. Moan, A. Mansouri, J. Y. Hardeberg, and Y. Voisin, “Saliency for spectral image analysis,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 6, no. 6, pp. 2472–2479, Dec. 2013.
18. S. Goferman, L. Zelnik-Manor, and A. Tal, “Context-aware saliency detection,” in Proc. IEEE Conf. Comput. Vision Pattern Recognit., 2010, pp. 2376–2383.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication