Patch-based Salient Region Detection Using Statistical Modeling in the Non-subsampled Contourlet Domain

M. Rezaie Abkenar, *Student Member, IEEE,* H. Sadreazami, *Member, IEEE,* and M. Omair Ahmad, *Fellow, IEEE*
Department of Electrical and Computer Engineering
Concordia University, Montreal, Quebec, Canada H3G 1M8
E-mail: {mas_reza, h_sadrea, omair}@ece.concordia.ca

***Abstract*—A salient region is part of the image that captures the greatest attention by the human visual system. In this paper, we propose a novel salient region detection technique in the non-subsampled contourlet domain. The image is first decomposed into non-overlapping patches in order to fully exploit the repetitive patterns in the image. It is known that the non-subsampled contourlet transform provides an efficient multi-resolution, multi-directional, localized and shift invariant decomposition of images. In view of this, by using the statistical properties of non-subsampled contourlet coefficients of image patches, a set of feature descriptors are extracted to construct the feature map for each color channel. An entropy-based criterion is proposed to combine the channel feature maps into a saliency map. Simulations are conducted on a dataset of natural images to evaluate the performance of the proposed method and to compare it with that of the other existing methods. The results show that the proposed salient region detection method provides higher precision, recall, and F-measure and lower mean absolute error values as compared to the other existing methods.**

***Index Terms*—Salient region detection, Non-subsampled contourlet transform, Normal distribution, Non-overlapping patches.**

I. Introduction

Saliency detection is one of the main characteristics of the human visual system (HVS) which enables it to detect the most significant region of the image. This region which is called the salient region, is processed with a higher priority compared to the rest of the image. Developing computational methods to detect the salient regions has attracted a great deal of interest in various computer vision and image processing applications such as adaptive image or video compression [1], object-based image retrieval [2], image retargeting [3], [4], and medical imaging [5].

Several saliency detection techniques have been developed to detect the entire salient object from the background [6]-[12]. Some techniques have been established in the frequency domain such as those in the wavelet domain [9], [11]. The wavelet transform offers a multiscale and spatial-frequency localized image representation. It has been shown that the separable wavelets can represent point discontinuities in a two-dimensional (2D) signal efficiently. However, they are not optimal in capturing line discontinuities correspond to directional information in the image [13]-[15]. To improve the limited directional selectivity of the wavelet, other multiscale representations, such as the non-subsampled contourlet transform (NSCT) has been developed [16]. A contourlet domain saliency detection method has been proposed in [12]. However, the proposed method in [12] cannot fully takes into consideration all the image features. In addition, since it is a pixel-level method, it has some limitations in preserving the structure of the salient object and detecting the entire salient object uniformly. It is known that HVS is attracted to distinctive regions rather than individual pixels and thus, the pixel-level methods are not able to perform efficiently especially when the image has a messy background with small sudden intensity variations. There have been only few attempts made to detect salient objects by evaluating dissimilarities between regions rather than pixels [11], [17].

Extracting visual features which are effective in detecting the distinctive part of the image is one of the main challenges in saliency detection. In this paper, a new salient region detection method is proposed by generating local feature descriptors using the statistical properties of the NSCT coefficients of the three color channels. Also, since the entropy is a statistical measure of randomness, an entropy criterion is applied to combine the obtained channel feature maps. Due to the fact that meaningful image regions can preserve the general structure of the objects, in the proposed method the saliency map is finally abstracted into meaningful regions utilizing a mean-shift clustering method.

II. Proposed Patch-based Saliency Detection Method

The proposed salient region detection method is based on the NSCT decomposition of local image patches. To this end, the input image is first converted to the commonly-used CIELAB color space, having luminance, red/green, and blue/yellow channels, denoted by L, a, and b, respectively. Each of the three color channels are then divided into n non-overlapping patches $\{z_i\}_{i=1}^{n}$ of size $m \times m$ pixels. The NSCT

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

Fig. 1. From left to right: L, a b channel feature maps of the given image, respectively.

is applied to each patch and decomposes it into a number of subbands as

$$\left\{ \boldsymbol{A}_l, \boldsymbol{D}_{l,d}^c \right\} = \mathrm{NCST}\left(\boldsymbol{z_i} \right) \tag{1}$$

where $\boldsymbol{A}_l$ denotes the lowpass subband, $\boldsymbol{D}_{s,d}$ denotes the directional bandpass subbands corresponding to the level $l \in \{1, ..., L\}$ and the direction $d \in \{1, ..., D\}$.

A. Local Feature Descriptor Using Statistical Modeling of the Non-subsampled Contourlet Coefficinets

The proposed saliency detection method is realized by defining a set of feature descriptors for each image patch. The features are obtained by taking into account the statistical properties of the NSCT coefficients in small patches. It is known that the distribution of the subbands having a small number of coefficients follow the Gaussian distribution [18]. Let $Y = \{y_1, ..., y_N\}$ be the independent and identically distributed (i.i.d) samples from one of the subbands. The probability density function (PDF) of Y is given by

$$p\left(Y|\mu, \sigma\right) = \frac{1}{\sqrt{2\pi\sigma}} \exp \frac{-\left(Y-\mu\right)^2}{2\sigma^2}, \tag{2}$$

where μ and σ are the mean and standard deviation of the distribution. These parameters can be estimated from the data samples using the maximum likelihood estimation as

$$\mu = \frac{\sum_{k=1}^{N} y_k}{N}, \tag{3}$$

and

$$\sigma = \sqrt{\frac{\sum_{k=1}^{N} \left(y_k - \mu\right)^2}{N}}, \tag{4}$$

where N is the number of samples. Having the estimated parameters for the subbands of each single patch, the corresponding feature descriptor for patch i is defined as

$$f_i = \left[\mu_1, \sigma_1, ..., \mu_S, \sigma_S\right], \tag{5}$$

where S denotes the number of subbands.

B. Channel Feature Maps based on the Feature Descriptor of the Patches

The salient region, is part of the image which is less similar to the rest of the image. Thus, a higher saliency value is assigned to a patch which is more different from all the other patches in the image. The distance between feature descriptors of two patches is computed as

$$dis(i,j) = \left\| f_i - f_j \right\|, \tag{6}$$

where f_i and f_j are feature descriptors of patches i and j, respectively, and $\|.\|$ denotes the Euclidean distance. The average distance value $mdis$, between feature descriptor of each patch and those of all the other patches is computed as,

$$mdis = \frac{1}{n} \sum_{j=1}^{n} \mathrm{dis}(i,j). \tag{7}$$

For each color channel, a channel feature map $\{\boldsymbol{map_c}; c \in \{L, a, b\}\}$, is constructed by replacing the patch pixels with its corresponding $mdis$. Fig. 1 depicts the channel feature maps for a given image.

C. Image Sailency Map

The image saliency map S_{map} is obtained by linearly combining the channel feature maps as

$$S_{map} = \sum_c w_c \boldsymbol{map_c}. \tag{8}$$

Taking into account the low entropy and center bias criteria, the weights w_c are assigned to each channel feature map. For a small entropy value and large intensity values around the center, a larger weight is assigned to the channel feature map. It should be noted that a low entropy value specifies having two distinctive regions for the channel feature map: salient and non-salient regions. On the other hand, the spatial information are also important and must be taken into account in computing the entropy. In view of this, the channel feature map is filtered by a lowpass Gaussian filter $\mathbf{I}$, and its entropy value α_c is given by

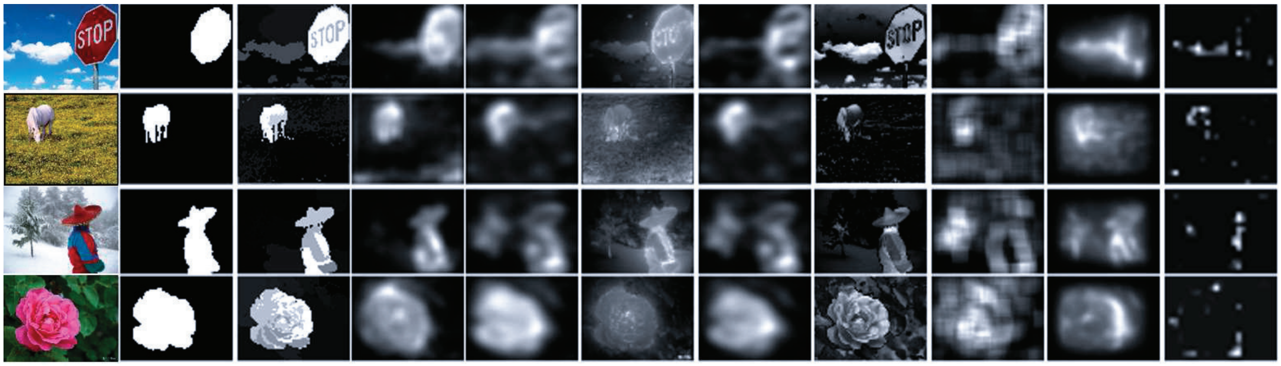
$$\alpha_c = H\left(\boldsymbol{map_c} * \mathbf{I}\right), \tag{9}$$

where $H(.)$ denotes the entropy.

Since most of the natural images are center-biased, a larger weight is assigned to the channel feature map for which the pixels around the center have larger intensity values. The strength of channel feature maps around the center β_c, is computed by multiplying the map by a 2D centered Gaussian mask $\mathbf{Z}$ as [8]

$$\beta_c = \sum_x \sum_y \mathbf{Z}.Nr\left(\boldsymbol{map_c}\right), \tag{10}$$

where $Nr(.)$ is a normalization function so that sum of all the pixel values of $\boldsymbol{map_c}$ is 1. Using (9) and (10), the weights are obtained as $w_c = \frac{\beta_c}{\alpha_c}$.



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k)

Fig. 2. Saliency maps obtained using the proposed method and the other methods. (a) Original image. (b) Ground truth. (c) Proposed method. (d) SW [11]. (e) WQ [10]. (f) WAV [9]. (g) HFT [8]. (h) FTU [7]. (i) SR [6]. (j) GBVS [21]. (k) IT [20].

D. Abstracting the Image Saliency Map

In view of the fact that a saliency detection method aims to detect the most distinctive region from the image background, we take into account meaningful image regions to improve the performance of the proposed method. To this end, color and spatial information of pixels are vectorized and the mean-shift clustering algorithm [19] is employed to cluster the image into k regions. The saliency map is superimposed on the obtained partitioned image and finally, the saliency value of each of the k regions S_k, is replaced by the average of the saliency values of all the M pixels belonging to that region as

$$S_k = \frac{1}{M} \sum_{\acute{i}=1}^{M} S_{map}(x_{\acute{i}}, y_{\acute{i}}) \tag{11}$$

where $(x_{\acute{i}}, y_{\acute{i}}) \in$ region k.

III. Experimental Results

We have conducted experiments on a set of publicly available dataset [7], which contains 1000 natural images of different sizes and their associated human-labeled ground truth, to evaluate the performance of the proposed saliency detection method and to compare it with that of the other existing methods. The parameter m is set to 16. The image is decomposed into one scale and eight directional subbands using NSCT with *pkva* filters. The evaluation metrics for quantitative comparison are: precision P, recall R and F-measure F, given by

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

where $S(x,y)$ is the saliency map, $GT(x,y)$ is the binary ground truth and $\xi = 0.3$ is a parameter to specify the relative importance of P and R, and is compared with that of a number of state-of-the-art methods, namely, SR [6], FTU [7], HFT [8], WAV [9], WQ [10], SW [11], IT [20] and GBVS [21]. The saliency maps obtained using the proposed method as well as that of the other methods for a few of the test images are shown in Fig. 2. It is seen from this figure that the saliency maps obtained by the proposed method is more similar to the ground truth as compared to that provided by the other methods. Besides, due to utilizing patches rather than individual pixels and also considering meaningful image regions, the proposed method can detect the entire salient object more uniformly.

Next, in order to quantitatively evaluate the proposed saliency detection method and compare the results with the ground truth, the gray level saliency maps are converted to binary maps using two different thresholds, fixed and adaptive. In fixed threshold binarization, the threshold is varied from 0 to 255, while in the adaptive thresholding, This threshold is set as twice the mean of the saliency values of the saliency map as defined in [7]. Therefore, the adaptive threshold value depends on the saliency values of each map.

Fig. 3 illustrates averaged precision-recall curves obtained using various methods over a set of images taken from MSRA-1000 dataset. It can be seen from this figure that the proposed method provides higher precision-recall values for a wide range of thresholds as compared to the other methods. Fig. 4 shows the precision, recall and F-measure values obtained using various methods using adaptive thresholding. It is seen from this figure that the proposed salient region detection method is superior to the other existing methods, as indicated by the higher precision, recall and F-measure values.

In some applications such as [3], the quality of the gray level saliency maps are of higher importance than the binary saliency maps. In order to evaluate the performance of the proposed method in terms of the detection of the salient as well as non-salient pixels in an image, the mean absolute error (MAE) between the gray level saliency map and the ground truth is obtained as in [22]. The MAE values obtained using the proposed method as well as that of the other existing methods are depicted in Fig. 5. It is seen from this figure that the proposed salient region detection method provides lower value of MAE as compared to the other methods.

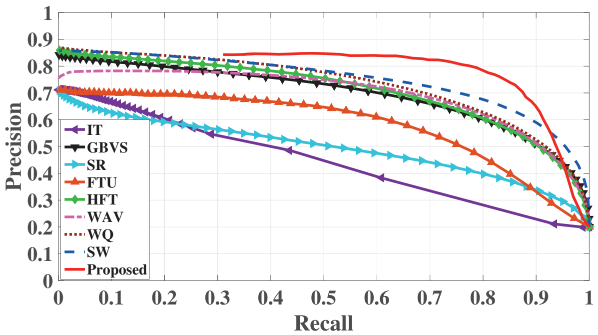


Fig. 3. Precision-recall curves obtained using the proposed method as well as that of the other methods, when the threshold is fixed.

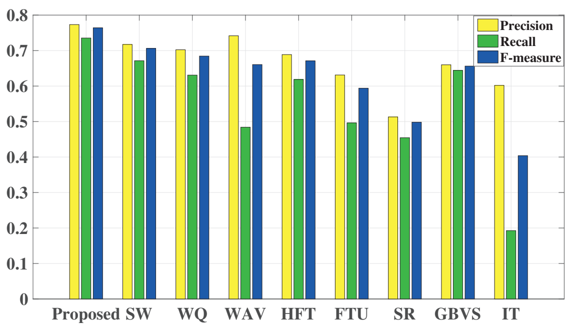


Fig. 4. Precision, recall, and F-measure values obtained using the proposed method as well as that yielded by using other methods, when the adaptive threshold is deployed.

IV. Conclusions

In this paper, a novel salient region detection method has been proposed by using the statistical modeling of non-subsampled contourlet coefficients. To this end, a set of statistical feature descriptors have been extracted from the non-subsampled contourlet coefficients of image patches and used to detect the salient regions in the color channels. The saliency map is constructed by combining the channel feature maps based on low entropy and center bias criteria. In addition, the saliency map is abstracted into regions using the mean-shift clustering method. Experimental results have shown that the proposed method outperforms the other methods in terms of precision, recall, F-measure and mean absolute error metrics. The improved performance of the proposed method is attributed to utilizing effective statistical-based local feature descriptors and preserving the structure of the objects by considering image regions rather than individual pixels.

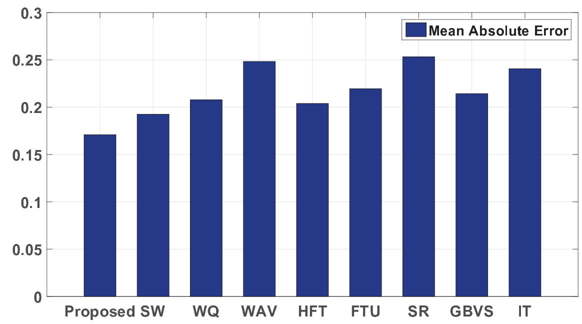


Fig. 5. MAE values obtained using the proposed method as well as that yielded by the other methods.

References

[1] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Transactions on Image Processing,* vol. 13, no. 10, pp. 1304-1318, 2004.

[2] Y. -F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proc. ACM International Conference on Multimedia,* pp. 374-381, 2003.

[3] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Transactions on graphics (TOG),* vol. 26, no. 3, 2007.

[4] G. Zhang, M. Cheng, Sh. Hu and R. Martin, "A shapepreserving approach to image resizing," *Computer Graphics Forum,* vol. 28, no. 7, 2009.

[5] Y. Yuan, J. Wang, B. Li and M.-H Meng, "Saliency based ulcer detection for wireless capsule endoscopy diagnosis," *IEEE Transactions on Medical Imaging,* vol. 34, no. 10, pp. 2046-2057, 2015.

[6] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 1-8, 2007.

[7] R. Achanta, S. Hemami, F. Estrada and S. Susstrunk, "Frequency tuned salient region detection," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 1597-1604, 2009.

[8] J. Li, M. D. Levine, X. An, X. Xu and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 35, no. 4, pp. 996-1010, 2013.

[9] N. Imamoglu, W. Lin and Y. Fang, "A saliency detection model using low-level features based on wavelet transform," *IEEE Transactions on Multimedia,* vol. 15, no. 1, pp. 96-105, 2013.

[10] M. Rezaie Abkenar and M. O. Ahmad, "Quaternion-based salient region detection using scale space analysis," in *Proc. Signal Processing and Intelligent Systems Conference (SPIS),* pp. 78-82, 2015.

[11] M. Rezaie Abkenar and M. O. Ahmad, "Superpixel-based salient region detection using the wavelet transform," in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS),* pp. 2719-2722, 2016.

[12] D. Liu, F. Chang and Ch. Liu, "Salient object detection fusing global and local information based on nonsubsampled contourlet transform," *Journal of the Optical Society of America A,* vol. 33, no. 8, pp. 1430-1441, 2016.

[13] H. Sadreazami, M. Omair Ahmad, M. N. S. Swamy, "Despeckling of synthetic aperture radar images in the contourlet domain using the alpha-stable distribution," *IEEE International Symposium on Circuits & Systems (ISCAS)*, 121-124, 2015.

[14] H. Sadreazami, M. Omair Ahmad, M. N. S. Swamy, "Image denoising utilizing the scale-dependency in the contourlet domain," *IEEE International Symposium on Circuits & Systems (ISCAS)*, 2149-2152, 2015.

[15] M. N. Do, *Directional multiresolution image representations,* Ph.D. diss., School Comput. Commun. Sci., Swiss Fed. Inst. Technol, 2001.

[16] A. L. Da Cunha, J. Zhou and M. N. Do, "The nonsubsampled contourlet transform: theory, design and applications," *IEEE Transactions on Image Processing,* vol. 15, no. 10, pp. 3089-3101, 2006.

[17] R. Kannan, G. Ghinea and S. Swaminathan, "Salient region detection using patch level and region level image abstractions," *IEEE Signal Processing Letters,* vol. 22, no. 6, pp. 686-690, 2015.

[18] H. Sadreazami, M. O. Ahmad and M. N. S. Swamy, "Contourlet domain image modeling by using the alpha-stable family of distributions," in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS),* pp. 1288-1291, 2014.

[19] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Transactions on Information Theory,* vol. 21, no. 1, pp. 32-40, 1975.

[20] L. Itti, C. Koch and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 20, no. 11, pp. 1254-1259, 1998.

[21] J. Harel, C. Koch and P. Perona, "Graph-based visual saliency," *Advances in neural information processing systems,* pp. 545-552, 2006.

[22] F. Perazzi, P. Krahenbuhl, Y. Pritch and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE Conf. on Comp. Vis. and Pattern Recognition (CVPR),* pp. 733-740, 2012.