IMAGE SALIENCY DETECTION VIA MULTI-SCALE STATISTICAL NON-REDUNDANCY MODELING

Christian Scharfenberger, Aanchal Jain, Alexander Wong, and Paul Fieguth

Department of Systems Design Engineering, University of Waterloo, Canada
{cscharfe, a27jain, a28wong, pfieguth} at uwaterloo.ca

ABSTRACT

A multi-scale statistical non-redundancy modeling approach is introduced for saliency detection in images. The statistical non-redundancy of pixels at different wavelet sub-bands is characterized using a multi-dimensional lattice of non-parametric statistical models, thus taking into account image saliency at multiple scales. This identifies saliency in image attributes at multiple scales, and makes saliency detection strongly robust against noisy input images. Results based on images from a public database show that the proposed approach outperforms existing single and multi-scale approaches, particularly when dealing with noisy images.

***Index Terms*—** multi-scale analysis, non-parametric methods, wavelet transforms, saliency detection

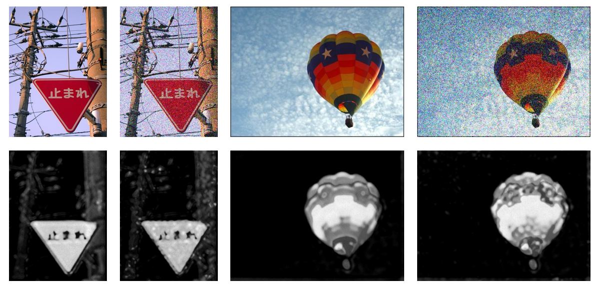


Fig. 1: Based on noiseless (columns 1 and 3) and noisy (columns 2 and 4) images (top), we analyze the statistical non-redundancy of pixels at different wavelet sub-bands to produce high-resolution saliency maps. Taking into account the image saliency at multiple scales, image noise barely influences the resulting saliency maps (bottom).

1. INTRODUCTION

Saliency detection in images involves highlighting those regions in the image which are unique in attributes – like color, texture, etc. – relative to other regions in the image. This forms an important first step in various image processing algorithms such as object detection [1], existence detection [2], object recognition [3], image segmentation [4, 5, 6], etc. Until very recently, the majority of traditional methods in saliency detection detect saliency at a single scale. Since different features in the image are highlighted at different scales, and natural images are commonly affected by noise, it is often desirable to capture saliency information of an image at multiple scales.

Achanta *et al.* [7] introduced a frequency-tuned saliency detection approach, which uses band pass filtering to obtain full-resolution saliency maps, and preserves boundary information of original images. The concept of frequency-tuned saliency detection has been extended in [8] with a maximum symmetric surround approach to overcome effects of very large salient regions or complex backgrounds. In addition, many local contrast based saliency detection methods have been proposed that evaluate saliency of an image with respect to small neighborhoods [9, 10, 11, 12, 13, 14]. These local contrast schemes focus on edges in images, rather than highlighting the entire salient region. To overcome the limitations of local contrast based concepts, Cheng *et al.* [15] suggested two saliency approaches that take the global contrast of images into account. The histogram contrast based approach relies on the color statistics of an image to compute saliency maps. The region contrast based approach segments input images first, and uses spatial relationships of image pixels within segmented regions along with the contrast. A similar global contrast based approach is discussed in [16] where the luminance contrast is used to compute the saliency values of the pixels in the image. Hou and Zhang [17] introduced a spectral residual approach that analyzes the log-spectrum of input images, and extracts the residuals of input images in the spectral domain. The latter is used to construct the saliency map in the spatial domain. The approach of [18] makes use of a learned neighborhood-based texture feature model, sparse texture modeling, and saliency computation based on the concept of textural distinctiveness for detecting salient regions in images. A recent non-parametric approach discusses saliency based on statistical non-redundancy of image pixel neighborhoods [19], and does not rely on training or prior information about the image.

The authors would like to thank the Natural Sciences and Engineering Research Council (NSERC) of Canada, the Ontario Ministry of Economic Development and Innovation, and the Canada Research Chairs Program for supporting this research project.

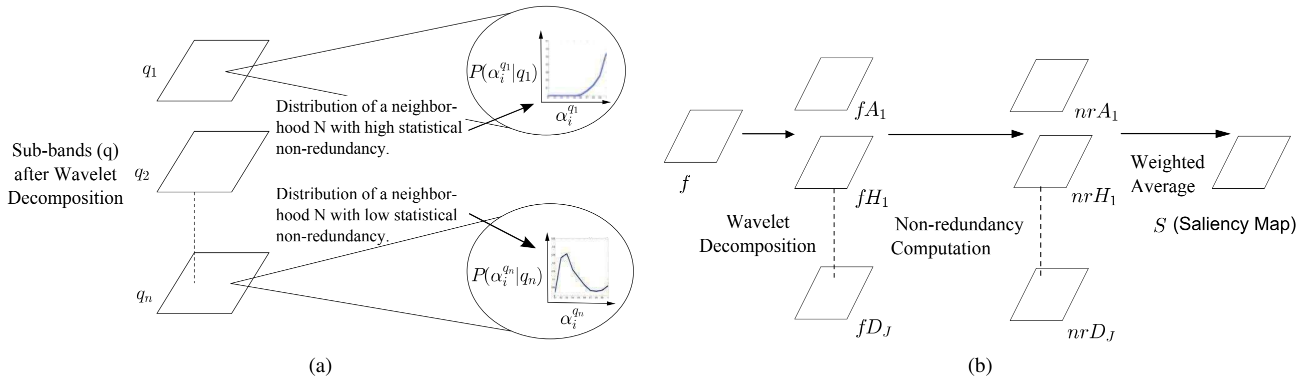


Fig. 2: The multi-scale statistical non-redundancy model (a) and the proposed saliency detection technique (b).

Existing multi-scale approaches consider local and global features at multiple scales for saliency computation. As a prominent example, Lui *et al.* [14] propose to use a linear combination of local color contrasts obtained at multiple scales in a Gaussian image pyramid as an important feature for saliency computation. The approach of Imamogilu *et al* [20] employs wavelet transformation on natural images to create saliency maps considering local edge and texture features at multiple scales in a center-surround manner. The final saliency map is then computed by fusing the individual saliency maps based on maximum computation.

In this paper, we extend the promising concept of statistical non-redundancy [19] to a multi-scale modeling framework based on wavelet decompositions, where a multi-dimensional lattice of nonparametric statistical models is used to characterize statistical non-redundancy of pixels at different wavelet sub-bands. Given the Gaussian characteristic of noise in natural images, our proposed framework involves the construction of statistical modeling of wavelet coefficients, and computes the saliency value of every pixel by analyzing its statistical non-redundancy with respect to the different wavelet sub-bands. This multi-scale approach does not only offer the advantage of identifying saliency in image attributes at multiple scales, but also makes saliency detection strongly robust against noise in natural images (see Fig. 1) because of its explicit modeling of noise in the statistical non-redundancy framework. This is an important difference to existing approaches such as [14] and [20] which do not explicitly model noise for saliency computation.

The rest of the paper is organized as follows. We present our multi-scale approach in Section 2, along with saliency detection in Section 3. We discuss experimental results in Section 4. Finally, conclusions are drawn and some directions for future work is presented in Section 5.

2. MULTI-SCALE STATISTICAL NON-REDUNDANCY MODELING

Salient regions of interest within natural images are regions that typically exhibit low content redundancy when compared to the rest of the image. Since such salient regions of interest may exist at different scales, our aim is to quantify and characterize such "non-redundancies" at different scales via a multi-scale statistical modeling framework (see Fig. 2) to better account for image saliency in a robust manner, particularly when dealing with noisy images. Suppose that an image f of size $X \times Y$ is decomposed into a series of Q sub-bands at different levels via wavelet decomposition [21]. The statistical non-redundancy between two pixel neighborhoods N_i and N_j (around the pixels i and j respectively) at a particular sub-band q is defined as

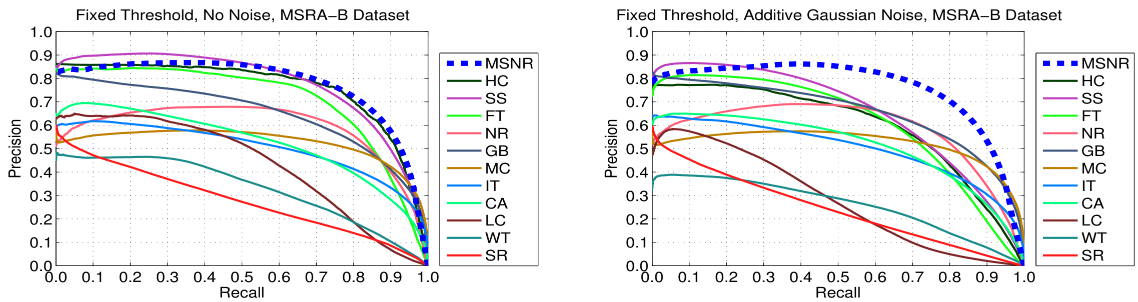
$$\alpha_{ij}^q = 1 - P(N_{i,q}|N_{j,q}) \tag{1}$$

where $P(N_{i,q}|N_{j,q})$ represents the probability of a pixel neighborhood N_i being a noisy observation of another pixel neighborhood N_j following a noise process at a particular sub-band q, where N_i and N_j differ only in an additive random component. We wish to model the noise process as an independent and identically distributed Gaussian random field with zero-mean and a variance σ and define the probability $P(N_{i,q}|N_{j,q})$ as

$$P(N_{i,q}|N_{j,q}) = \prod_k e^{\frac{-(N_{i,q}^k - N_{j,q}^k)^2}{\sigma^2}} \tag{2}$$

where $N_{i,q}^k$ denotes a pixel k within the neighborhood N_i of a pixel i at sub-band q. As such, a high value of α_{ij}^q denotes high statistical non-redundancy, which indicates low content redundancy between the two pixel neighborhoods.

Given this metric for quantifying the statistical non-redundancy between pixels at a particular sub-band, we can characterize the statistical non-redundancy of a pixel with respect to entire sub-bands within a nonparametric multi-scale modeling framework. Let $P(\alpha_i^q|q)$ denote the probability of statistical non-redundancy at pixel i (i.e., α_i^q), given the entire sub-band q. To obtain $P(\alpha_i^q|q)$ in a computationally efficient manner, we employ a stochastic nonparametric estimation strategy, where a set of n samples (which we will denote as χ) drawn from a uniform distribution across q is used to



(a) (b)

Fig. 3: Precision-recall curves for state-of-the-art multi (MC, WT) and single-scale saliency approaches (solid lines), and for our approach (MSRN, dashed line). We averaged the curves over the images from the MSRA-B dataset for noiseless (a) and noisy images (b) using additive Gaussian noise, with $\sigma = 24\%$ of the dynamic range.

approximate $P(\alpha_i^q|q)$ based on the following formulation:

$$P(\alpha_i^q|q) = \frac{\sum_{j \in \chi} \delta\left(\alpha_i^q - \alpha_{ij}^q\right)}{Z} \tag{3}$$

where Z is a normalization term such that $\sum P(\alpha_i^q|q) = 1$. Based on empirical results, it was found that $n = 300$ samples was sufficient for obtaining a reliable approximation of $P(\alpha_i^q|q)$ for the purpose of saliency detection. Given this nonparametric estimation strategy, we can then obtain an $X \times Y \times Q$ lattice of nonparametric statistical models that characterizes the statistical non-redundancy of different pixels at different scales.

3. SALIENCY MAPS

Given the proposed multi-scale statistical non-redundancy modeling framework as shown in Fig. 2, we define the overall saliency $S(i|q)$ at a pixel i at sub-band q as the expected value of α_i, given the entire sub-band q:

$$S(i|q) = \sum \alpha_i^q P(\alpha_i^q|q). \tag{4}$$

For the purpose of computing a final saliency map for the detection of salient regions within an image, irrespective of scale, one effective strategy is to compute the weighted average of the normalized saliency S, $S \in [0..1]$ across all sub-bands q to take salient regions identified at N decomposition levels into consideration. Given that the expected saliency $E(S|q)$ represents the overall non-redundancy of a sub-band q, weighting $w(i|q)$ is based on the difference between the saliency $S(i|q)$ and $E(S|q)$ and allows us to assign greater weights to sub-bands containing regions with significant, while de-emphasizing sub-bands with less or almost no significant non-redundancy characteristics. Thus, we define the final saliency as

$$S(i) = \frac{1}{N} \cdot \sum_{q}^{N} w(i|q) S(i|q), \tag{5}$$

$$w(i|q) = \begin{cases} S(i|q) - E(S|q), & \text{if } S(i|q) - E(S|q) > 0 \\ 0, & \text{otherwise} \end{cases}$$

In the experiments discussed in this paper, we have used $N = 3$ decomposition levels. In addition, an un-decimated wavelet transform [22] was employed for wavelet decomposition, as it was found to provide improved saliency detection performance when compared to saliency maps based on the decimated wavelet transform [21]. For computing the saliency maps of color images using the proposed method, we calculate the saliency maps for the CIELab channels L, a, b, and take a weighted average of these maps.

4. RESULTS AND DISCUSSIONS

We wish to use quantitative measures to evaluate the performance of the proposed MSNR algorithm against the current state-of-the-art in saliency detection. We compared our approach with 11 state-of-the-art saliency detection methods which have been selected based on the following criteria [7, 15, 18]: number of citations (spectral-residual (SR) [17], visual attention (IT) [23]), recency (luminance-contrast (LC) [16], graph-based (GB) [24], frequency-tuned (FT) [7], Maximum Symmetric Surround (SS) [8], context-aware (CA) [25], histogram-contrast (HC) [15], non-redundancy (NR) [19]), and being related to our approach (multi-scale contrast (MC) [26], multi-scale wavelet transform (WT) [20]). Alternate saliency approaches suggest to incorporate high level priors, image segmentation, and training into saliency computation. Since these concepts differ fundamentally from our proposed saliency approach that relies on low-level features only, we excluded them in our evaluation to ensure a consistent and fair comparison.

The robustness of the MSNR approach is shown by means of precision-recall characteristics (see Fig. 3), obtained from noiseless and noisy images (contaminated by additive Gaussian noise, with $\sigma = 24\%$ of the dynamic range) from the MSRA-B database [7]. In addition, this dataset contains accurate human-marked labels as ground truth and is widely used as a benchmark for comparing saliency approaches.

Based on the precision-recall curves, we can see that the proposed MSNR method either outperforms or provides state-of-the-art performance when compared to existing ap-

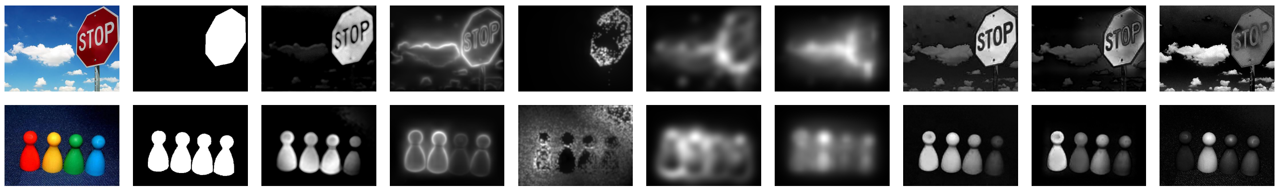


Fig. 4: Saliency maps for the state-of-the-art in multi-scale (*italic*) and single-scale saliency detection, and our MSNR scheme.

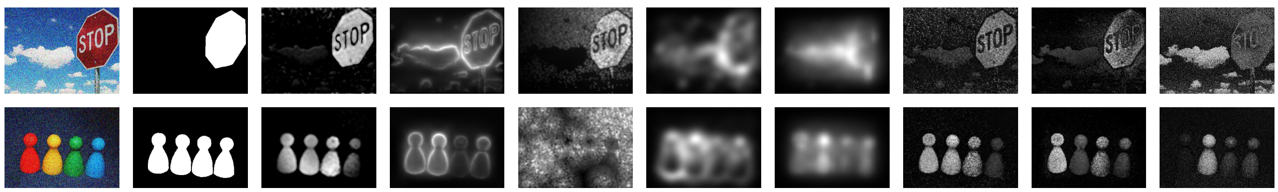


Fig. 5: Saliency maps for noisy images (additive white Gaussian noise, $\sigma = 24\%$ of the dynamic range). It can be seen that the salient regions in saliency maps produced by our MSNR approach are well defined, despite the image noise.

proaches using noiseless (see Fig. 3a) and noisy images (see Fig. 3b). From the precision-recall curves, we can see that noise barely influences the MSNR approach, but noticeably affects the single-scale saliency approaches. MSNR also outperforms the multi-scale contrast (MC) [26] and wavelet transform (WT) [20] for precision-recall values larger than 0.7 for noiseless and noisy images.

For a visual comparison, Fig. 4 illustrates the saliency maps obtained from different saliency detection algorithms. We can see that the saliency maps obtained from the proposed algorithm (MSNR) highlight the salient regions very clearly. In Fig. 5, we show the saliency maps produced for noisy images having additive white Gaussian noise. A direct comparison between the saliency maps in Fig. 4 shows a significant degradation of the quality of the saliency maps in the case of noise for all but those produced by the MSNR approach. This illustrates that, by taking into account the image saliency within the statistical non-redundancy framework at multiple scales, the influence of noise on the quality of saliency maps produced by the MSNR approach is weakened.

In terms of approaches relying on image segmentation or training, the quality of their saliency maps strongly depends on carefully selected pre-processing stages, whereas our proposed algorithm does not. Hence, the proposed MSNR approach may be more suitable for real-life images, e.g., those containing strongly textured backgrounds which may lead image segmentation to fail, or those for which a priori knowledge for training may not be available.

5. CONCLUSIONS AND FUTURE WORK

A saliency detection algorithm based on a multi-scale statistical non-redundancy modeling approach has been proposed. The proposed method has also been shown to provide reliable saliency maps that are particularly robust to noise. Furthermore, the proposed approach was shown to provide promising performance when compared to other saliency detection methods, particularly in situations characterized by noise. Our future work involves investigating the following areas for improving saliency detection performance. First, rather than employ a uniform distribution for the stochastic nonparametric estimation strategy in building the multi-scale statistical non-redundancy model, it would be worth investigating more intelligent guided sampling approaches, to better capture the distribution of salient and non-salient regions within input images. Second, it would be worth investigating different weighting schemes for the individual sub-bands, when computing the final saliency map based on the individual saliency maps corresponding to the different sub-bands. Third, different weighting schemes for the individual color channels could be considered when computing the final saliency maps for the color images based on the importance of the individual channels with respect to image saliency. In addition, when dealing with color-based weighting schemes, a database such as [27] combining eye tracking and segmentation results might be more suitable to study the effects of color on saliency computation than other datasets such as MSRA-B [7] ignoring the impact of color on saliency maps at all.

6. REFERENCES

[1] A. Das, M. Dui, C. Scharfenberger, J. Servos, A. Wong, J. S. Zelek, D. A. Clausi, and S. Waslander, "Mapping, planning, and sample detection strategies for autonomous exploration," *Journal of Field Robotics*, vol. 31, no. 1, pp. 75–106, 2014.

[2] C. Scharfenberger, S. L. Waslander, J. S. Zelek, and D. A. Clausi, "Existence detection of objects in images for robot vision using saliency histogram features," in *IEEE Conf. on Computer and Robot Vision*, 2013.

[3] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition?," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2004, vol. 2.

[4] J. Han, K.N. Ngan, M. Li, and H.J. Zhang, "Unsupervised extraction of visual attention objects in color images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 1, pp. 141–145, 2006.

[5] D. Lui, C. Scharfenberger, K. Fergani, A. Wong, and D. A. Clausi, "Enhanced decoupled active contour using structural and textural variation energy functionals," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 855–869, 14.02.01 2014.

[6] K. Fergani, D. Lui, C. Scharfenberger, A. Wong, and D. A. Clausi, "Hybrid structural and texture distinctiveness vector field convolution for region segmentation," *Computer Vision and Image Understanding*, vol. 125, pp. 85–96, 2014.

[7] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2009.

[8] Radhakrishna Achanta and Sabine Suesstrunk, "Saliency Detection using Maximum Symmetric Surround," in *IEEE International Conference on Image Processing*, 2010.

[9] 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.

[10] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2010.

[11] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry.," *Human Neurobiology*, vol. 4, no. 4, pp. 219–27, 1985.

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

[13] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Advances in neural information processing systems*, vol. 19, pp. 545, 2007.

[14] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.Y. Shum, "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353–367, 2011.

[15] M.M. Cheng, G.X. Zhang, N.J. Mitra, X. Huang, and S.M. Hu, "Global contrast based salient region detection," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2011.

[16] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *ACM International Conference on Multimedia*, 2006.

[17] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2007.

[18] C. Scharfenberger, A. Wong, K. Fergani, J. S. Zelek, and D. A. Clausi, "Statistical textural distinctiveness for salient region detection in natural images," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2013.

[19] A. Jain, A. Wong, and P. Fieguth, "Saliency detection via statistical non-redundancy," in *IEEE International Conference on Image Processing*, 2012.

[20] N. Imamogilu, W. Lin, and Y. Fang, "A saliency detection model using low-level features based on wavelet transform," *IEEE Transactions on Multimedia*, vol. 15, pp. 96–105, 2013.

[21] S.G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[22] XH Wang, R.S.H. Istepanian, and Y.H. Song, "Microarray image enhancement by denoising using stationary wavelet transform," *IEEE Transactions on NanoBioscience*, vol. 2, no. 4, pp. 184–189, 2003.

[23] 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.

[24] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in Neural Information Processing Systems 19*, 2007.

[25] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2010.

[26] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Shum, "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353–367, 2011.

[27] T. Judd, F. Durrant, and A. Torralba, "A benchmark of computational models of saliency to predict human fixations," in *CSAIL Technical Reports, MIT*, 2012.