IJBSCHS
[Original Article]

Biomedical Soft Computing and Human Sciences, Vol.19, No.1, pp.69-73

(Accepted on 2014.1.23)

A Saliency Detection Technique Considering Self- and Mutual-Information

Ahmed Boudissa[1)], Joo Kooi Tan[2)], Hyoungseop Kim[3)], Takashi Shinomiya[4)], Seiji Ishikawa[5)].

1,2,3,5) Kyushu Institute of Technology, Kitakyushu, Japan
4) Japan University of Economics, Tokyo, Japan

Abstract: *In this paper, we present a novel approach to saliency detection. We define a visually salient region in an image with following two properties; global spatial redundancy, i.e., mutual-information, and local saliency, i.e., self-information or simply the region complexity. The former is its probability of occurrence within the image, whereas the latter defines how much information is contained within a region, and it is quantified by the entropy. By combining the global spatial redundancy measure and local entropy, we can achieve a simple, yet robust saliency detector. We evaluate it quantitatively and qualitatively. The comparison to Itti et al. [6], the spectral residual approach by Hou and Zhang [5], Achanta et al. [13] as well as to Zhai and Shah [14], on publicly available data shows a significant improvement.*

Keywords *Visual attention, Saliency, Entropy, Segmentation, Human Visual System, Statistical redundancy*

1. Introduction

Recent advances in hardware architecture and processing power gave rise to the industrial interest in computer vision applications and algorithms. More specifically, there has been a significant progress in the field of object detection, which typically consists of a number of computationally intensive stages. It is believed that the human biological visual system has the early stage of attention [2], in which the background or clutter is discarded without the need of an exhaustive analysis of the scene as illustrated in **Fig. 1.** This motivates many researchers to model the visual attention and propose saliency detection approaches that can provide an input to the object detection approaches.

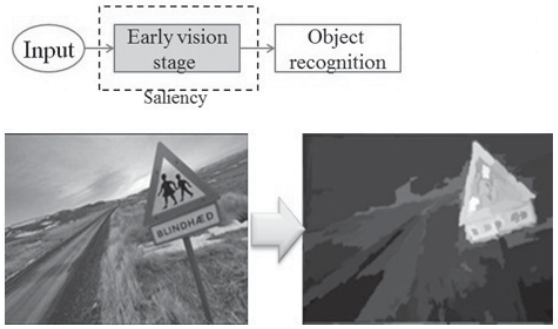

Fig.1. Illustration of saliency detection.

2. Related work

This work focuses on methods for automatic salient region extraction. The goal is to detect the salient regions in an image efficiently to produce promising candidates for subsequent stages of computationally demanding algorithms. Limiting the search can greatly speed up the process [10] and allows for real-time applications.

The existing approaches can be divided into two major categories, local and global. Bottom-up or local approaches are feature based, concerned with the local appearance of images. For example, Itti et al. [6] model the biological early attention mechanism using several feature maps (color, orientation, luminance, etc.), which have also shown promise in image compression applications. Within the same category, Hou and Zhang [5] propose a frequency based model

using the spectral representation of images. In [8], [9], [6] and [11], saliency is defined in terms of local Shannon's information maximization principles. In [8], it is proposed to use entropy based locally salient features to estimate the global transformation between two images. Global approaches attempt to detect image regions that exhibit certain properties: For instance, Seo and Milanfar [12] use regions self-similarity based on a regression kernel. Goferman et al. [4] present a multi-scale salient object detector using a distance computation between different image patches to find the region of interest, which proved useful in image retargeting and image collage creation. A theoretical formulation of top-down visual saliency, related to the recognition problem, is proposed in [3].

3. Saliency model

In this paper we propose a simple yet robust approach to detect salient regions in an image by measuring the following properties:

- Spatial redundancy, the probability of occurrence within an image which quantifies the uniqueness of the concerned region.
- Local information content which defines how much information is contained within the region, and it is quantified using entropy.

3.1 Spatial redundancy

The spatial redundancy, or the mutual-information, of a patch within an image is estimated by measuring how similar it is to the other patches. This measure is also referred to as the global saliency measure since it consider the saliency aspect on the global scale of an image in contrast with what will follow, which is the local aspect. Typically it is assumed that a patch is salient if it is dissimilar to the nearby patches. However, in practice, due to noise, change of a view point or illumination conditions, direct comparison of patches results in an inaccurate estimation.

For a better estimate of the similarity between two regions R_i and R_k , we model it with the Gaussian function with zero-mean and standard deviation σ. The global saliency measure for region R_i compared to N other regions denoted by R_k is defined by the following;

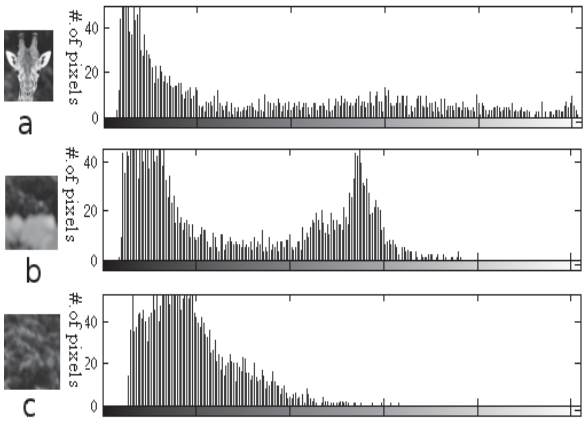

Fig.2. Different histogram distributions resulting in different entropy for (a) high, (b) medium and (c) low complexity patches.

$$S_{R_i} = \frac{1}{N}\sum_{k=1}^{N}(1-e^{\frac{-D_E(R_i,R_k)}{\sigma^2}})/D_S(R_i,R_k) \qquad (1)$$

where $D_E(R_i,R_k)$ is the Euclidean distance between region attributes such as color distribution. In addition, we incorporate the spatial distance between patches normalized by the image size i.e. $D_S(R_i,R_k)$ in Eq. (1) in order to down-weight the importance of distant patches. This allows to maintain the integrity of an object that typically consists of nearby patches. The proposed similarity measure combines the appearance and the spatial distance and it normalizes distances computed for different attributes of image patches, or in different color spaces (Gray, RGB, L*a*b or HSV), unlike the Euclidean distance directly used in [4] [7].

3.2 Local information content

According to Shannon's information theory, the information content can be measured with the entropy of the signal at a certain location in time or space. Maximizing the self-information by including the entropy in our saliency measure eliminates the homogeneous backgrounds and unique regions of low complexity. In **Fig. 2**, we show images of different complexity which is reflected by their histograms resulting in different entropy. The entropy is characterized by the spread of the histogram which reflects the complexity of the region. The wider spread is the histogram (Fig.2-a), the higher is the entropy.

We compute the entropy using the following equation:

$$H_{R_i} = -\sum_{v=1}^{256} p_{v,R_i} \ln p_{v,R_i} \quad (2)$$

where p_{v,R_i} represents the probability of pixel intensity v within a region R_i.

p_{v,R_i} is computed using the probability density function of R_i simply by normalizing the grayscale histogram by the number of pixels constituting the region.

3.3 Global-local full saliency model

In contrast to the candidate selection process from [9] we combine the local entropy of a region defined in section 3.2 with the global saliency in section 3.1. The spatial redundancy and the local entropy measures expressed in Eqs. (1) and (2) are combined into a full saliency model S_i which is used to generate a saliency map:

$$S_i = H_{R_i} \times S_{R_i} \quad (3)$$

The map is generated by computing the saliency for every region in the image implemented as a scanning window of size W. In the following section, we provide a qualitative and quantitative evaluation of this saliency detection measure.

4. Experimental results

The dataset we use to evaluate the proposed method has also been used in [7] and [5]. It consists of 62 natural images with salient and non-salient regions manually labeled by 4 different annotators.

Instead of selecting a specific threshold for the saliency measure as in [9] and [4], which gives a limited insight into the overall performance, we evaluate our method and compare it to other approaches using precision-recall curves. Given the ground-truth segmentation maps, we perform pixel-wise comparison to the thresholded saliency map and calculate precision-recall values by varying the saliency threshold. The method proposed in this paper is parameterized by the region size as well as the size of the Gaussian in color space for patch similarity function. By varying each parameter, we assess their mutual correlation and the optimal operating configuration.

Fig. 3 (a, b and c) show the performance for different sizes of the scanning window W= {5; 10; 15} and standard deviation sigma = [1;...; 8] of the Gaussian in Eq. (1). This is done to assess the performance of the system in accordance with the different parameters. The higher the curve is, the better is the performance.

By analyzing the graphs in Fig.3 (a, b and c), we find that the performance is relatively independent of the size of selected windows, but it varies slightly depending on the size of the Gaussian. The smaller the scanning window, the faster the performance peaks, and remains stable. For a large scanning window, i.e. W = 20, the complexity, thus the uniqueness of the region increases, resulting in high saliency scores for the cluttered background.

Fig.3 (d) shows the performance comparison between the proposed method and other popular methods; Itti et al. [6], Achanta et al. [13], Zhai and Shah[14] as well as Hou and Zhang [5]. We see that our approach outperforms the other methods due to the combination of local and global characteristics.

Itti et al. [6] use only local features to describe salient regions. As for Zhai and Shah[14] and Hou and Zhang [5], they base their methods solely on frequencies, which is a global definition of saliency. The method by Achanta et al.[13] is more suited to videos, since it uses temporal cues, and this is why it performs poorly on still images. On the other side, we can see that, within our framework, the entropy measure participates greatly in eliminating low complexity regions. The uniform areas have very compact histograms, leading to low entropy, which balances the saliency score in case of an erroneously low spatial redundancy value.

In **Fig.4,** we display a qualitative comparison of the saliency maps generated by the proposed method as well as by the approaches from [6], [13], [14] and [5]. Since the objective is to evaluate the saliency of a whole object, the final or the optimal goal is to create masks that converge to the ground truth. According to this principle, we can see that our method gives sharper saliency maps compared with other methods in which the objects as a whole are salient. Also, the use of the entropy down weights the uniform areas, but the spatial redundancy preserves the object's integrity even if the entropy within the region is low.

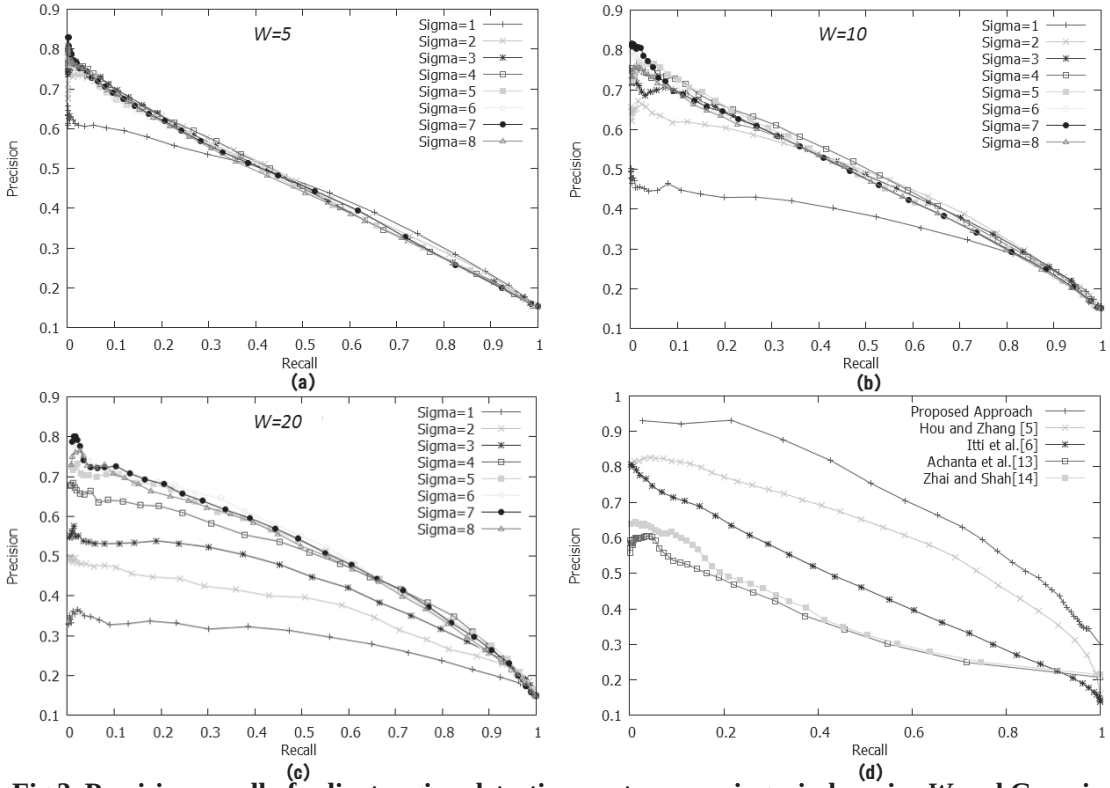

Fig.3. Precision-recall of salient region detection w.r.t. a scanning window size *W* and Gaussian weighting = Sigma (cf. Eq. 1) in color similarity. Comparison to Itti et al. [6] , Achanta et al.[13], Zhai and Shah[14] as well as Hou and Zhang [5] methods is in the bottom-right

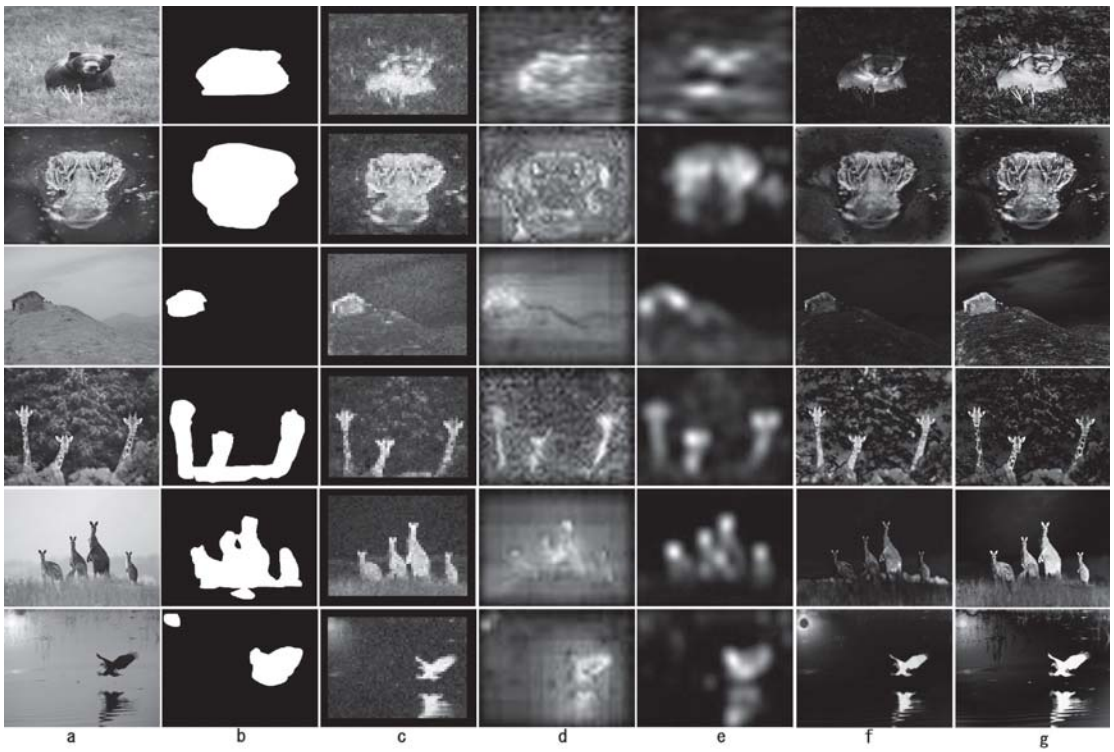

Fig. 4. Saliency maps comparison. (a) Original image, (b) ground truth, (c) our approach, (d) Itti et al. [6], (e) Hou and Zhang [5], (f) Achanta et al. [13], (g) Zhai and Shah[14]. The use of the entropy measure creates sharper saliency maps and highlights the body of the objects in addition to the boundaries.

5. Conclusion and future work

In this paper, we introduced a novel combined global-local saliency measure. It is based on local entropy and global redundancy of the region.
We evaluated the approach on available benchmarks and compared to state-of-the-art approaches. The proposed method showed that it outperforms the classical methods, and owing it to its simplicity, it has the potential of improving speed and accuracy of various object detection algorithms. As future prospects, an automatic way to select the parameters of the method should be explored as well as more efficient implementations. The challenge remains to define a more robust saliency measure and collect diverse benchmark data. We also intend to evaluate the saliency algorithms as a first step in more complex computer vision applications.

Acknowledgments

This study was supported by JSPS KAKENHI under Grant-in-Aid for scientific Research (22510177)

References

[1] N. Bruce and J. Tsotsos. Saliency based on information maximization. Advances in Neural Information Processing Systems 18, 155-162 (2006)

[2] M. Carandini, J. Demb, V. Mante, D. Tolhurst, Y. Dan, B. Olshausen, et al. Do we know what the early visual system does? Journal of Neuroscience, 25, 10577-10597, (2005)

[3] D. Gao, S. Han, and N. Vasconcelos, Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, (2009)

[4] S. Goferman, L. Zelnik-Manor, and A. Tal, Context aware saliency detection. IEEE Conference on Computer Vision and Pattern Recognition, 2376-2383 (2010)

[5] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. IEEE Conference on Computer Vision and Pattern Recognition, 1-8 (2007)

[6] 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 20, 1254-1259 (1998)

[7] A. Jain, A. Wong, and P. Fieguth. Saliency detection via statistical non-redundancy. IEEE International Conference on Image Processing, 1073-1076 (2012)

[8] T. Kadir. Scale, saliency and scene description. University of Oxford (2001)

[9] T. Kadir and M. Brady. Scale saliency and image description. International Journal of Computer Vision. 45, 83-105 (2001)

[10] 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 33, 353-367 (2011)

[11] W. Wang, Y. Wang, Q. Huang, W. Gao. Measuring visual saliency by site entropy rate. IEEE Conference on Computer Vision and Pattern Recognition (2010)

[12] H.J. Seo and P. Milanfar. Static and space-time visual saliency detection by self-resemblance. Journal of Vision, 1-27 (2009)

[13] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, Frequency-tuned salient region detection, IEEE CVPR, 1597-1604 (2009)

[14] Y. Zhai and M. Shah, Visual attention detection in video sequences using spatiotemporal cues, ACM Multimedia, 815-824 (2006)