MVA2013 IAPR International Conference on Machine Vision Applications, May 20-23, 2013, Kyoto, JAPAN

A novel saliency measure using combined spatial redundancy and local appearance

Ahmed Boudissa[1], Joo Kooi Tan[2], Hyoungseop Kim[3], Seiji Ishikawa[4], Takashi Shinomiya[5]
1,2,3,4: Kyushu Institute of Technology. Kitakyushu, Japan
5: Japan University of Economics, Tokyo, Japan

abstract

In this paper, we present a novel approach to saliency detection. The method we propose here aims at synthesizing common knowledge of saliency in an image. We define a visually salient region with following two parameters; the spatial redundancy and its local appearance. The former is its probability of occurrence within the image, which is a quantification of the "rarity" of the concerned region, whereas the latter defines how much information is contained within the region, and it can be quantified using the entropy. By combining the global spatial redundancy measure and local entropy, we can achieve a simple, yet robust measure. We evaluated and compared it to Itti's and the spectral residual methods, and it has shown a significant improvement of performance.

1. Introduction

In the last decade, the increasing rate of computational abilities of processors gave rise to the interest in computer vision applications and algorithms. More specifically, there has been a considerable amount of researches on object recognition.

It is believed that the human biological vision architecture has the early stage of "attention". In which the background or clutter is discarded without the need to an exhaustive scan of the scene, which is the motivation driving many researchers to model the visual attention.

2. Related works

This work consists of an automatic salient region extraction algorithm. The goal is to detect the most salient regions efficiently to produce a minimal data input to an object recognition algorithm. The latter ones suffering usually from time constraints for real-time applications, and a high level of false positives, limiting the search region, will help in solving those issues [1].

The existing frameworks can be mainly sorted into two major categories: local and global.

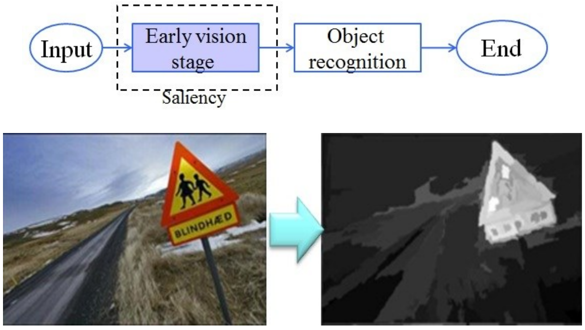


Fig.1. Simple diagram for the biological human vision flow

Bottom-up or local approaches, are feature based, concerned with the local appearance of images. *Itti and Koch* [4] tried to recreate the biological early attention mechanism using several feature maps (color, orientation, luminance…etc.). It has shown promise in image compression applications. Within the same category, Hou and Zhang[6] presented a purely frequency based model using the spectral representation of images. In [2][3][5] and [7], saliency is defined in terms of local Shannon's information maximization principles. In [2], it is proposed that using locally salient features at salient scales from a pair of images would make it possible to estimate the global transform between them.

Global approaches, attempt to locate the "special" patches: Seo [8] used regions self-similarity based on a regression kernel. Goferman et al. [9] presented a multi-scale salient object detector using distance computation between different image patches to extract the surrounding region of the object of interest. It has shown good prospects in image retargeting and image collage creation.

3. Saliency Model

The method we propose here aims at synthesizing the common knowledge about what can be defined as salient in an image. For this, we define a visually salient region with the two following parameters: (1) its probability of occurrence within the image (i.e.

redundancy) which is a quantification of the "rarity" of the concerned region within the image. (2) its local appearance, which defines how much information is contained within the region, and it can be quantified using the entropy. A salient region should contain a high level of information concentration.

3.1. Spatial redundancy

In the first part, we quantify the rarity of a patch within an image by finding how close it is to the other patches in the image. In theory, it is assumed that a patch in an image can be reproduced in another part of the image, which makes it less salient. But in practice, due to noise, view and lighting condition changes, direct comparison of the patches results in a not very accurate estimation of that difference.

For a better estimation of the color or appearance distance between two – different- regions R_1 and R_2 - supposedly similar- , we model the noise as an independent Gaussian process with zero-mean and a variance σ . The global measure, when region R_i is compared to N other regions, becomes the following;

$$SS_{R_i} = \frac{\sum_{k=1}^{N} (1 - e^{\frac{-D(R_i,R_k)}{\sigma^2}})}{N} \quad (1)$$

SS_{R_i} is the spatial saliency of the region R_i .

$D(R_i, R_k)$: A distance function that depends on what space our image is being considered. If we consider, like some other researchers did, [9][10], direct Euclidean distance between patches, the inaccuracy occurs due to the shift between patches that might be similar but will give raise to a high distance. In our case, we will compare the probability density functions of each patch, in different spaces (Gray-RGB-L*a*b or HSV).

We also include the spatial distance d_{i-k} between R_k and the original region R_i (See Eq.(3)). This is to penalize the effect of far patches even if they might be similar. The following table summarizes that effect:

Color distance	Spatial distance	Saliency
High	Near	High
Low	Near	Low
High	Far	X
Low	Far	Low

Table.1. Saliency proportionality to spatial and color distance (X : no effect)

3.2. Local appearance

According to Shannon's information theory, "Information" in a signal can be computed using the entropy of the signal in a certain time or space location. As mentioned previously, the local complexity of a region determines how much it "pops-out" within the image. By local complexity, we refer to complexity of the patterns present in the image. We can quantify this complexity based on the spread of the histogram of the concerned region. Including the entropy measure eliminates the homogenous backgrounds and areas even if they are "rare". In **Fig. 2**, we can see that the more complex is the area, the more spread is the histogram which leads to a higher entropy.

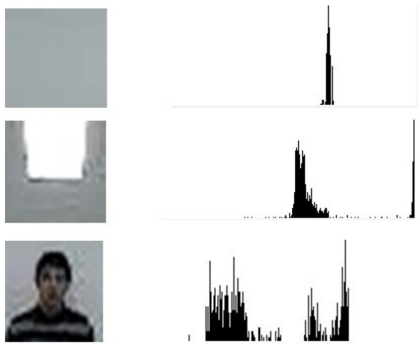


Fig.2. Different histogram distributions

We use the following equation to estimate the entropy of each region:

$$H_{D,R_i} = -\sum_{v=1}^{255} p_{v,R_i} Log_2(p_{v,R_i}) \quad (2)$$

H : Entropy.

R_i : Local neighborhood around pixel *i*.

v: the gray values of pixel *i* taken within R_i .

p_{v,R_i} : Probability Density Function of the area R_i .

In our framework, the value of the entropy of a region is used as another parameter in the whole formula, as opposed to a candidate selection process as stated in [3].

Finally, we combine both the spatial redundancy and the local appearance measures expressed in equations (1) and (2) into a single saliency value S_i with which we generate a saliency map:

$$S_i = H_{D,R_i} \frac{\sum_{k=1}^{N} \frac{(1 - e^{\frac{-D(R_i,R_k)}{\sigma^2}})}{d_{i-k}}}{N} \quad (3)$$

d_{i-k} : Normalized Spatial distance between regions *i* and *k*.

4. Experimental results and evaluation

The dataset used to evaluate the proposed method is a popular dataset used in [10] and [6]. It consists of 62 natural images labeled by 4 different humans as salient or non-salient regions.

Also, instead of selecting a specific threshold like in [3] and [9], we evaluate our method using a precision versus recall curve. This measure shows how discriminatory each method is, rather than randomly setting a cut-off threshold which can be over fitting the test data without giving a general overview of the performance.

The method that we propose is parametric; it depends on the region size and shape, on the choice of the color space, and also, the Gaussian process's standard deviation. By varying each parameter, we try to assess their mutual correlation and the optimal operating configuration.

Fig. 3 shows the performance comparison when varying the window size and the standard deviation. We can see that the performance does not depend on the size of selected windows, but there is a correlation between the size of the patches and the standard deviation. The smallest the window is, the faster the performance peaks, and remains almost stable. This is due to the fact that, when we choose a Gaussian distribution approaching the size of the window, the standard deviation value acts as a normalization factor, and the distance (color distance) between different patches becomes normalized throughout the whole image.

Fig. 4 shows the performance comparison between the proposed method and two other popular methods: Itti's method [4] and Hou [6] method.

We can see that our approach outperforms those classical methods due to the combination of both local and global characteristics. Also, the use of probability density functions to compare the patches makes up for the shift and illumination differences that may occur when comparing two different patches.

Also, the entropy measure participates greatly in eliminating false positives occurring in natural images. The uniform areas have very sharp histograms, leading to low entropy, but in some cases, the patch random selection when computing the color distance might result erroneously in a high spatial non-redundancy value, but the low entropy will balance the whole saliency formula.

Fig. 5 shows examples of extracted objects using our method. We can see that even though we are using a global threshold, the regions inside and on the border of the objects remain salient.

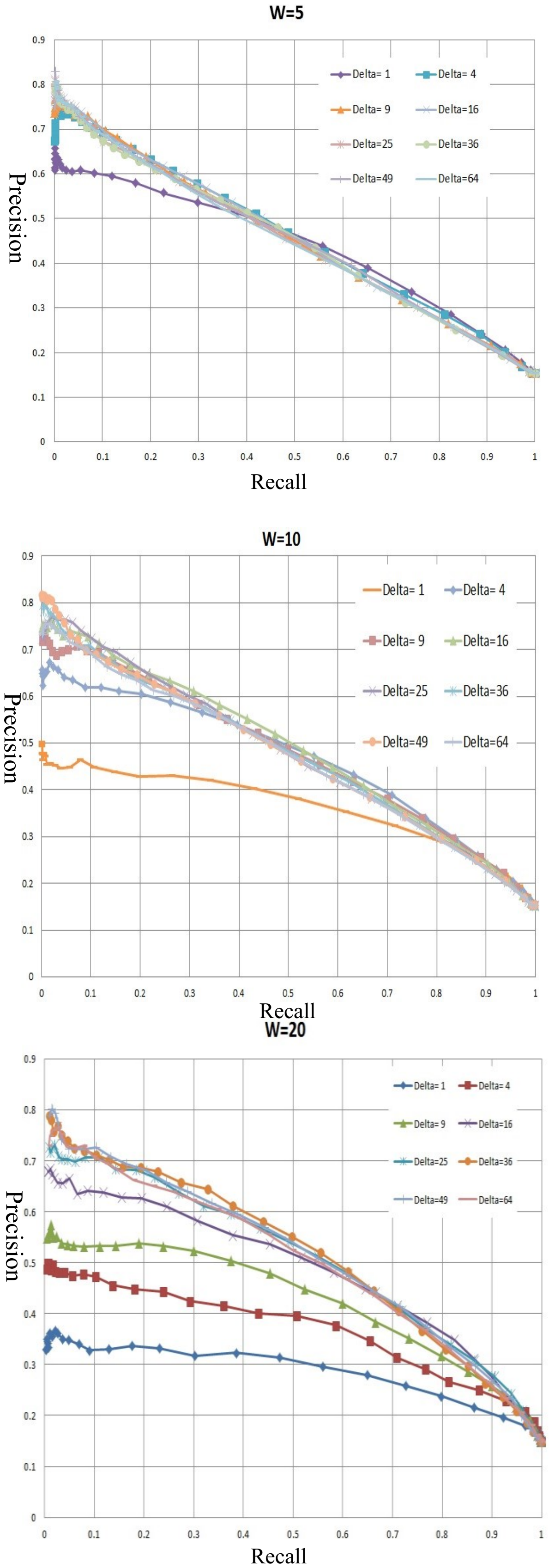
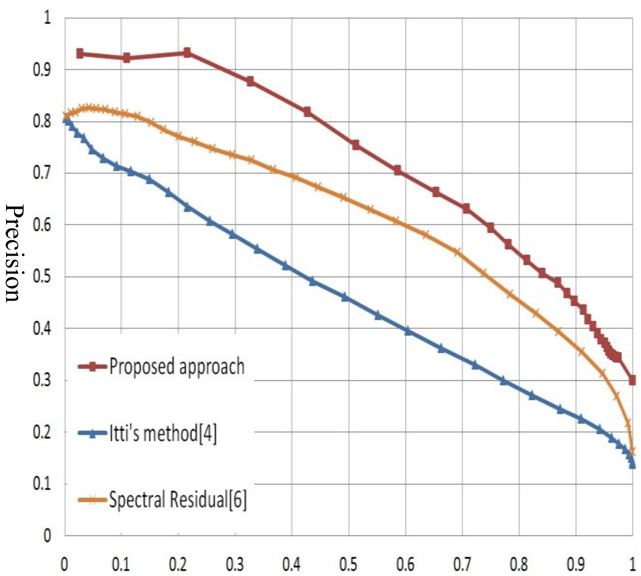


Fig.3. Precision versus recall curves when varying W the patch size and σ

5. Conclusion and future work

In this paper, we introduced a new global-local saliency measure. The results outperform



Recall

Fig.4. performance comparison with [4] and [6]

classical methods, and also, the simplicity of method gives it a good prospect to be introduced in the already existing object detection algorithms, either by directly eliminating non salient regions according to a certain pre-set threshold, or, depending on the time constraint and the frame rate of the application, directly using the saliency map as a probability map.

As future prospects, an automatic way to select the threshold and window selection methods should be explored. Other noise models can be tested. But also, the challenge remains to define saliency formally and benchmark databases thoroughly, and to include the saliency algorithms as a first vision step in other computer vision application.

Acknowledgment

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

11. References

[1] Tie Liu, Zejian Yuan, Jian Sun, Jingdong Wang, NanningZheng, Xiaoou Tang, and Heung-Yeung Shum. Learning to Detect a Salient Object. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 2, pp. 353–67, March, 2011.

[2] T. Kadir. Scale, Saliency and Scene Description. PhD thesis, University of Oxford, Deptartment of Engineering Science, Robotics Research Group, University of Oxford, Oxford, UK, 2001.

[3] Timor Kadir and Michael Brady. Scale, Saliency and Image Description. International Journal of Computer Vision. 45 (2):83-105, November 2001.

Fig.5. Extracted objects (first row) and their respective input and ground truth images (second and last row).

[4] 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,pp. 1254–1259, November 1998.

[5] N. Bruce and J. Tsotsos. Saliency based on information maximization. Advances in Neural Information Processing Systems, vol. 18, pp. 155–162, 2006.

[6] X. Hou and L. Zhang. Saliency detection: A spectralresidual approach. IEEE Conference on ComputerVision and Pattern Recognition, 2007, pp. 1–8.

[7] Wei Wang, Yizhou Wang, Qingming Huang, Wen Gao. Measuring Visual Saliency by Site Entropy Rate. IEEE Computer Vision and Pattern Recognition, CVPR, June 2010.

[8] H.J. Seo and P. Milanfar. Static and space-time visual saliency detection by self-resemblance. Journal of Vision, vol. 9, no. 12, pp. 1–27, 2009.

[9] S. Goferman, L. Zelnik-Manor, and A. Tal . Contextaware saliency detection. IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 2376–2383.

[10] Aanchal Jain, Alexander Wong, and Paul Fieguth, Saliency detection via statistical non-redundancy. IEEE ICIP, 2012, pp. 1073- 1076