21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

An Effective Regional Saliency Model Based on Extended Site Entropy Rate

Yan Huang, Wei Wang, Liang Wang, Tieniu Tan
National Laboratory of Pattern Recognition, Institute of Automation
Chinese Academy of Sciences, Beijing 100190, China
{*yhuang, wangwei, wangliang, tnt*}*@nlpr.ia.ac.cn*

Abstract

In this paper, we propose a new region-based saliency model to simulate the human visual attention. First, we construct a pixel-level fully-connected graph representation for an image, and perform normalized cut to segment the image based on the proximity and similarity principles. After obtaining image regions, we reconstruct a region-based fully-connected graph. Based on the saliency principle "center-surround contrast", we define new dissimilarity functions in terms of several visual features. Finally we run a random walk on the region graph and apply site entropy rate to measure the region saliency. We evaluate the proposed model on a public dataset consisting of 120 images. Experimental results demonstrate that our model predicts eye fixations more accurately than the other four state-of-the-art methods. We also apply our saliency model to improve the performance of image retargeting.

1. Introduction

Although human beings have limited computing and storage capacities, we can easily process visual information from the outside world. As we know, human visual attention plays an important role in this process, which is able to select the most valuable visual information from a large amount of the sensory data to interpret complex scenes real time. So it is very necessary to propose the computational models of visual attention, which has a range of potential applications, e.g. object recognition [9] and image retargeting [2].

Recently, various computational models based on visual saliency have been proposed to simulate human visual attention. From the perspective of saliency processing units, these models can be classified into two major categories. (1) The first category computes the saliency value for each *pixel* in an image. Through modeling the center-surround mechanism of primary visual cortex, Itti and Koch [7] propose a saliency model based on center-surround differencing. Considering that human visual system tends to focus on the most informative points in an image, Bruce et al. [3] utilize the self-information of visual features to measure saliency, and Wang et al. [11] use the random walk on a fully-connected graph to simulate the information transmission among the interconnected neurons, and propose the Site Entropy Rate (SER) of each graph node as the saliency measure. Based on the frequency analysis, Hou et al. [6] extract the spectral residual of an image in the spectral domain and compute the saliency map in the spatial domain. (2) The second category computes visual saliency for each *region* in an image. Cheng et al. [4] compute local region contrast to measure regional saliency. Achanta et al. [1] propose a frequency tuned approach to compute saliency. Zhai et al. [12] use color histograms to compute image region saliency.

In this paper, we explore the relationship between normalized cut [10] and site entropy rate [11], and propose a new region-based saliency model which measures regional saliency by the *extended site entropy rate*. First, we define a similarity function and a dissimilarity function for every two pixels in an image. The similarity function is used in normalized cut to segment the image based on the proximity and similarity principles in Gestalt law, and the dissimilarity function is exploited to compute the transition probability between every two regions. Then, we apply the random walk on the obtained image regions to simulate the signal transition among neurons. Finally, in accordance with Wang et al. [11], we use the site entropy rate to measure the saliency for image regions. The main contribution of this paper is that we principally extend site entropy rate to the image region saliency measure, and the resulting regional saliency model obtains the state-of-the-art performance.

2. The Regional Saliency Model

A probabilistic interpretation of normalized cut as a random walk has been revealed by Maila et al. [8] which considers feature similarities as edge flows in the

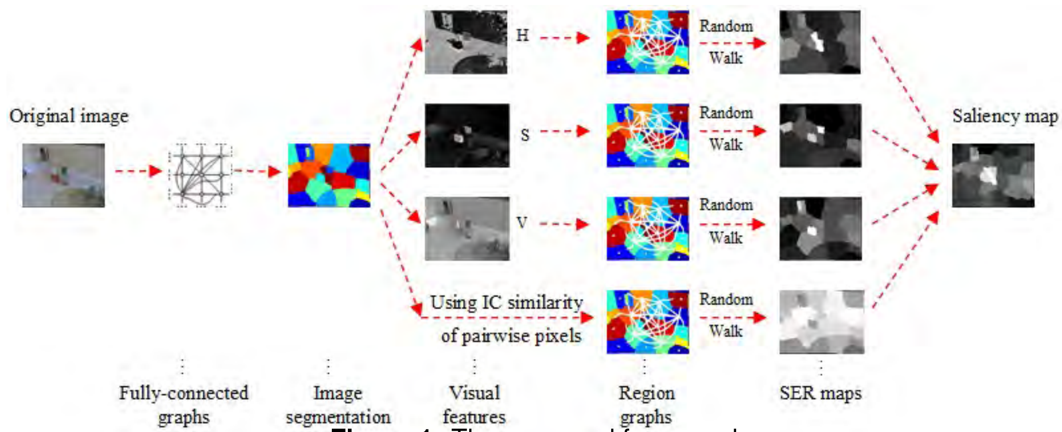

Figure 1: The proposed framework

random walk. The ratio term in normalized cut provides a natural definition for the transition probability from one region to another. In our model, image regions from the results of normalized cut are taken as nodes based on which we construct a fully-connected graph. As we know, image segmentation aims to group similar elements based on visual proximity and similarity principles while visual saliency aims to highlight some elements based on feature contrast. So we define a similarity function for pairwise pixels in normalized cut and defines a dissimilarity function for pairwise regions in extended site entropy rate. Fig. 1 illustrates the process of computing the saliency map of an image. First, we construct a fully-connected graph representation for the input image, and compute the intensity and contour affinity of pairwise pixels. Normalized cut is performed to obtain an image segmentation result. Then we build the region-based fully-connected graphs with image segments using several visual features, such as hue (H), saturation (S), value (V), intensity and contour (IC). As like Wang et al. [11], we run random walks on the region graphs to simulate the information transmission among neurons. The dissimilarity of two regions is the sum of the feature dissimilarities of all the pairwise pixels in these two regions. After computing the transition probability of every two regions, we use the site entropy rate to measure the saliency of each region.

2.1. Acquiring image regions

In our model, we first construct a pixel-based fully-connected graph representation for an image and then use normalized cut [10] to obtain image regions. The key of normalized cut is to define the similarity function for pairwise pixels as the edge weights of the graph. We use two widely used grouping features: intensity and contour, to compute the similarity $W(i,j)$ [5]:

$$W(i,j) = \sqrt{W_I(i,j) \times W_C(i,j)} + \alpha W_C(i,j) \quad (1)$$

where $W_I(i,j)$ and $W_C(i,j)$ are respectively the intensity affinity and the contour affinity of pixel i and j, and α is the relative weight of these two features. See [5] for more details. After computing the similarity of pairwise pixels, normalized cut is performed to obtain image regions.

2.2. Measuring region saliency

Based on the segmentation results of normalized cut, we construct the region-based fully-connected graphs in terms of several visual features, and compute the dissimilarity of pairwise regions as the edge weights of these graphs. Similar to Wang et al. [11], we measure the region saliency by site entropy rate.

2.2.1. Constructing region graphs. For four image features that we consider, namely intensity and contour (IC), hue (H), saturation (S), and value (V), we first define the dissimilarity of pairwise pixels, and then sum up the dissimilarities of all the pairwise pixels between two regions as the dissimilarity of these two regions.

For intensity and contour (IC), we compute the dissimilarity $D_{ic}(i,j)$ of pixel i and j by subtracting the corresponding similarity $\overline{W}(i,j)$ which is the normalization form of $W(i,j)$:

$$D_{ic}(i,j) = 1 - \overline{W}(i,j) \quad (2)$$

As to the other three features, we compute the dissimilarity in the same way. For simplicity, we only introduce the process of computing the hue dissimilarity here. The dissimilarity between pixel i and j is defined from two aspects: the feature dissimilarity denoted by ϕ_{ij} and the spatial distance denoted by d_{ij}:

$$D_h(i,j) = \phi_{ij} d_{ij} \quad (3)$$

where ϕ_{ij} and d_{ij} are defined as $|h_i - h_j|$ and $exp\{-\frac{(l_i - l_j)^2}{\lambda}\}$ respectively, h_i, h_j and l_i ,l_j are respectively the hue values and locations of pixel i and j, and λ is a tuning parameter.

For different features, after obtaining the dissimilarity of pairwise pixels, we can compute the dissimilarity of region r_i and r_j denoted by e_{r_i,r_j} as:

$$e_{r_i,r_j} = \sum_{m \in r_i} \sum_{n \in r_j} D(m,n) \tag{4}$$

where $D(m,n)$ is the dissimilarity between pixel m and n. Regarding the image regions as graph nodes and the corresponding dissimilarities as edge weights, we build the region-based fully-connected graphs.

2.2.2. Extending site entropy rate to region saliency measure . For each feature, we simulate the information transition between regions as a random walk and define the transition probability of the random walk from region r_i to region r_j as:

$$P_{r_i,r_j} = \frac{e_{r_i,r_j}}{vol(r_i)} \tag{5}$$

where $vol(r_i) = \sum_{m \in r_i} \sum_{n \in R} D(m,n)$, and R is the whole image.

The stationary probability π_{r_i} of the random walk can be computed as $\pi_{r_i} = vol(r_i)/vol(R)$. Similar to Wang et al. [11], we compute site entropy rate (SER) to measure the saliency of region r_i as.

$$SER_{r_i} = \pi_{r_i} \sum_{r_j} -P_{r_i,r_j} \log P_{r_i,r_j} \tag{6}$$

According to the Feature-Integrated Theory [2], the final saliency map is the sum of all the SER maps corresponding to the used four features.

$$S_{r_i} = \sum_{k=1}^{4} SER^k_{r_i} \tag{7}$$

3. Experimental Results

In order to evaluate the performance of our model, we predict eye fixations on a public image dataset and compare our model with four state-of-the-art models qualitatively and quantitatively. We also apply our saliency model to improve the performance of image retargeting.

3.1. Experiments of predicting eye fixations

Our method is an extension of [11], which achieves the similar saliency detection performance to [11]. Here we compare our model with four state-of-the-art methods ([4], [6], [12], [1]) on a dataset provided by Bruce et al. This dataset contains 120 color images and their corresponding eye fixations. Due to space limitation, we just show the results of 5 example images in Fig. 2.

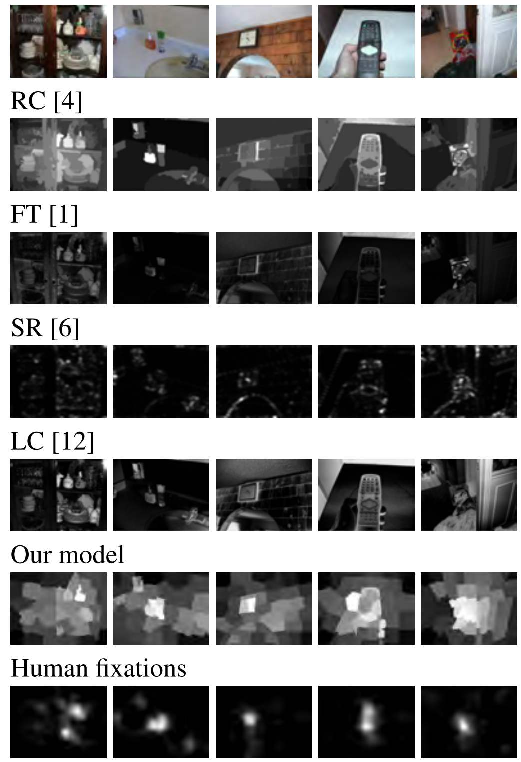

Figure 2: Results for a qualitative comparison between our model and other four approaches.

We can see that our saliency maps are much more similar to human eye fixations maps than those of the other methods.

To test our model quantitatively, we also classify the saliency maps with varying thresholds to fixations and non-fixations and compute the Hit Rate (HR) and the False Positive Rate (FPR). The average ROC curves of these compared methods on the used dataset are plotted in Fig. 3 and the corresponding ROC areas are listed in Table 1. The bigger the ROC area, the better the model. From the comparison results, we can again see that our model performs much better.

Table 1: The ROC area comparison

Methods	ROC areas
FT [1]	0.5186
LC [12]	0.5391
SR [6]	0.6193
RC [4]	0.7017
Our model	**0.7585**

3.2. Application of image retargeting

Seam carving [2], as a popular image retargeting method, needs an energy map to measure the importance of image pixels, and then removes the relative low

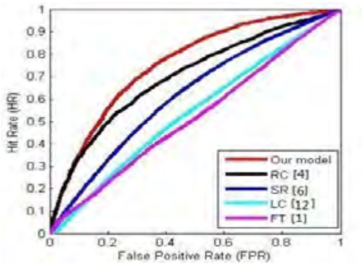

Figure 3: The ROC curves of our model and other four state-of-the-art approaches on the image dataset in [3].

energy pixels for reducing the image size. Image gradients are often used as an energy function to guide image retargeting, but this does not always work well when the objects of interest have a smooth visual appearance. The second table in Fig. 4 shows the retargeting results only using gradient maps. In this table, The first row shows the gradient maps, the second row shows the process of seam carving and the third row shows the reduced images. We see that the objects appear to have varying degrees of distortion.

Using saliency map individually to guide image retargeting does not work well naturally, because pixels of non-salient regions share the same low energy values, and thus the entire regions will be removed when reducing the image size, which makes the image lose too many details. In this paper, we use the weighted sum of the saliency map and the gradient map as one energy map to improve image retargeting. The first row in the third table in Fig. 4 shows the combining energy maps. We can see that the retargeting results with the new energy maps in the third row preserve the objects of interest well, which shows that our model which correctly highlights the objects of interest in the saliency map plays an important role in this application.

4. Conclusions

This paper has proposed a new region-based computational model for visual saliency by extending original pixel-level saliency measure (i.e., site entropy rate) to image regions. Experiments have demonstrated that the proposed model achieves better fixation prediction accuracy than several state-of-the-art methods.

5. Acknowledgments

This work is jointly supported by National Natural Science Foundation of China (61175003), Hundred Talents Program of CAS, The strategic Priority Research Program of CAS (XDA06030300), and National Basic Research Program of China (2012CB316300).

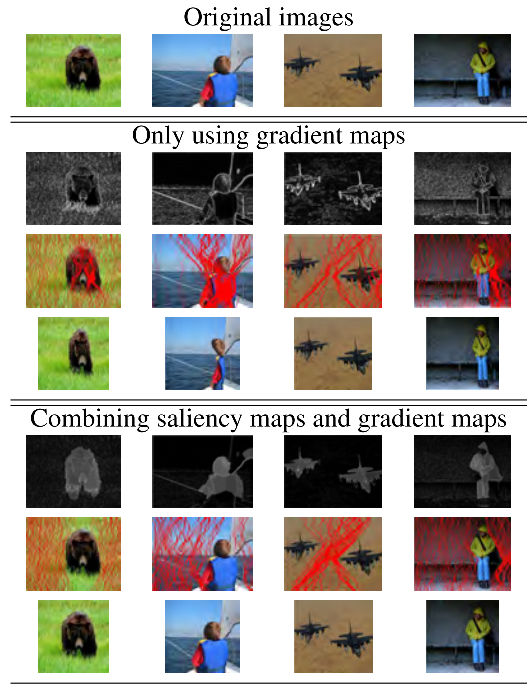

Figure 4: Results of using different energy maps to guide seam carving.

References

[1] R. Achanta, S. Hemami, F. Estrada, and S. Susstunk. Frequency-tuned salient region detection. *CVPR*, 2009.

[2] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. *ICIP*, 2009.

[3] N. Bruce and J. K. Tsotsos. Saliency based on information maximization. *NIPS*, June 2006.

[4] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu. Global contrst based salient region detection. *CVPR*, 2011.

[5] T. Cour, F. Benezit, and J. Shi. Spectral segmentation with multiscale graph decomposition. *CVPR*, 2005.

[6] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. *CVPR*, June 2007.

[7] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE TPAMI*, Nov 1998.

[8] M. Maila and J. Shi. Learning segmentation with random walks. *NIPS*, 2001.

[9] U. Rutishauser, D. Walther, C. Koch, and P. Perona. Is bottom-up attention useful for object recognition. *CVPR*, 2004.

[10] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE TPAMI*, 22, 2000.

[11] W. Wang, Y. Wang, Q. Huang, and W. Gao. Measuring visual saliency by site entropy rate. *CVPR*, 2010.

[12] Y. Zhai and M. Shah. Visual attention detection in video sequences using spatiotemporal cues. *In ACM Multimedia*, 2006.