Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Salient region extraction based on Intensity Mapping for image retrieval

Lang Congyan, Xu De, Li Ning, Feng Songhe

Institute of Computer Science and Engineering,
Beijing Jiaotong University
Beijing, 100044, China
{cylang, dxu}@bjtu.edu.cn

Abstract

Salient Region Extraction provides an alternative methodology to image description in many applications such as adaptive content delivery and image retrieval. In this paper, we propose a robust approach to extracting the salient region based on bottom-up visual attention. The main contributions are twofold: 1) Instead of the feature parallel integration, the proposed saliencies are derived by serial processing between texture and color feature. 2) A constructive approach is proposed for rendering an image by a non-linear intensity mapping, which can efficiently eliminate high contrast noise regions in the image. And then the salient map can be robustly generated for a variety of nature images. Finally, the salient region extracted by our algorithm is used for image semantic retrieval. Experiments show that the proposed algorithm can characterize the human perception well and achieve satisfied retrieval performance.

I. Introduction

Visual attention is a mechanism of the human visual system, and has the ability to select and gate visual information based on saliency in the image itself (bottom-up or image driven) and on both external and internal stimuli about the scene (top-down or goal-driven). Many physiological experiments suggest that human vision system only processes part of incoming information in full detail. That is, instead of processing all the available information attention can implement an information processing bottleneck by seeking interesting areas in images.

Detecting salient region has become an important topic in recent years, and the major differences among computational attention models are image features and difference mechanisms of saliency measure. In particular, some statistical signal-based approaches have been proposed[1]. Most of them are luminance-based and exploit spatial contrast and spatial entropy of pixels as its saliency measure. One of the most important works related to visual attention is proposed by Itti et al [2], they proposed a biologically-plausible computational model by utilizing the contrasts in color, intensity and orientation of images. Here, the concept of saliency map is presented as an integration of different measurable and low-level image feature. In [3,4], the saliency map is weighted by a Gaussian function with the assumption that humans generally pay more attention to objects near to the center of an image. Another model proposed by Hu et al.[5] relies on subspace estimation and analysis. The most recent work is proposed by Yu et al.[6], they described a rule based approach for visual attention region extraction. Then, a set of hierarchical salient region are generated based on a confidence factor. .

In this paper, without needing the full semantic image understanding, we attempt to develop a robust approach to modeling bottom-up visual attention. Specially, contrast is an important parameter in assessing vision. However, contrast is very sensitive to local noise structure for modeling attention. In order to take the challenge, we proposed a non-linear intensity mapping, the result of intensity mapping is a new grey image generated by compressing the local contrast range of the image in similar texture regions, while enhancing it in dissimilar texture regions. Fig 1 shows an example of non-linear intensity mapping, where we can see that in the original image, the intensity contrast of circle region is larger than that of the triangle region. After non-linear intensity mapping, the intensity-mapped image is generated by compressing the contrast range of circle regions with global texture consistency. Meanwhile the triangle region has the highest intensity contrast in the intensity-mapped image so that it can be extracted as salient region more robustly using simple saliency measure.

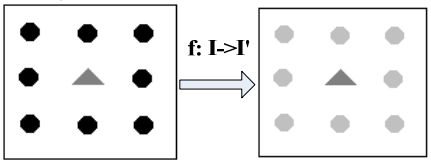


Fig 1 an example of non-linear intensity mapping

The remainder of this paper is organized as follows. The detail of the proposed computational attention model is presented in Section II. Experimental results are reported in

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

Section III. Finally, conclusions will be presented in Section IV.

II. The proposed computational attention model

A. Contextual Texture Extraction

In this section, a novel contextual texture feature is extracted to describe spatial consistency of a region globally. The main goal is to obtain global feature for constraining visual attention modeling so that outliers can be eliminated for salient map.

First, the JSEG algorithm[7] is used to obtain the segmented regions, which can segment the image into regions with homogeneous chrominance component. Here, the level of accuracy in this initial step is not important for the overall performance of the proposed salient region extraction, as the clear contours do not need for model salient map.

By analyzing large mount natural images, we found that the more similar spatial structure of regions is distributed in the image, the less possible a salient object contains this texture. Hence, global feature can be extracted based on contextual texture information. In this paper, we use Weibull distribution to extract region-level contextual texture feature. The parameters of the Weibull distribution can characterize the spatial structure of uniform stochastic texture of image and have been successfully applied[8] . In particular, the distribution of edge response of a region can be modeled by a Weibull distribution as following:

$$f(x)=\frac{\gamma}{\beta}\left(\frac{x}{\beta}\right)^{\gamma-1} e^{-(\frac{x}{\beta})^{\gamma}}, x \in R_i \qquad (1)$$

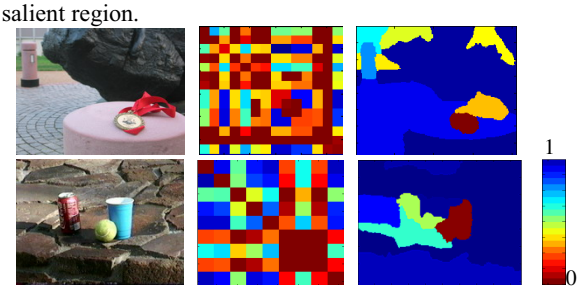
where the β (the width of the distribution) indicates the contrast of the region, and the shape of the distribution is given by γ(the peakedness of the distribution), the higher is γ, the smaller is more fine textures. In order to smooth small regions and reduce computational cost, small regions (region size les T_s , $T_s = 0.05 \times N_I$) of the image I (N_I denotes the total pixel number) is combined into the color similar regions. Similar to the method[8], Weibull parameter are derived from a histogram of edge responses in x and y-direction. To describe the contextual texture information, for each segmented region R_i in the image I , the distance of region texture distribution is defined as:

$$C(R_{pq})=\sum_{i=p,\ q}\frac{N_i}{N_{p\cap q}}\| \gamma_i-\gamma_{p\cap q} \| \qquad (2)$$

Where N_i is the pixel number of region R_i , γ_i is the peakedness parameter of Weibull distribution of the region R_i. Based on experiment for larger image database, we found the parameter γ can be well used to describe contextual texture information. Then a contextual texture consistency for the region R_i is used as following:

$$\xi(R_p)=1-\sum_{j\in N_R, j\neq p} C(R_{pj})\cdot e^{-D\|R^c{}_p-R^c{}_j\|^2 / 2*\sigma^2} \qquad (3)$$

where $R^c{}_p$ is the chrominance vector (v_2 , v_3) of homogenous region. And $D\| R^c{}_p - R^c{}_j \|$ denotes the color similarity of the region R_p and R_j , $\sigma = 3$ is used. According to (3), texture consistency of a region is derived as a weighted sum of texture similarity. As shown in Figure 2, texture consistency shows a better prediction of the saliency of each region. Specially, the smaller a consistency value is, the high probability the region belongs to the salient region.



(a) (b) (c)

Fig.2 (a) Original image (b) The distance map of region texture distribution (c) Texture consistency map

B. Adaptive Non-linear Intensity Mapping Algorithm

Contrast is one of the major visual feature attractors and can efficiently guide the attention to the most salient areas of our visual field. Most of the existing attention models should take advantage of this visual dimension. Motivated by the fact, our motivation is that intensity mapping has properties that intrinsically lead to a different contrast distribution. In this way, we use a non-linear intensity mapping to achieve a new intensity-mapped image, which can provide good contrast and detail preservation in global texture distinctive areas while achieve to compress the contrast of the texture similar areas.

In order to preserve more details and contrast in rich texture areas and compress luminance in weak texture areas, we propose an adaptive intensity-mapped method, which adjusts the intensity value according to each pixel's contextual texture consistency. This function, here defined as:

$$\psi_i=\frac{\bar{I}_R+f(I_i)}{I_i+f(I_i)}\bullet\alpha I_i$$
$$f(I_i)=\beta\bullet I_i^{\xi(R_i)} \qquad (4)$$

where ψ_i and I_i denote intensity value for the pixel p_i, α and β are global and local parameters for intensity contrast respectively. The adaptation factor $f(I_i)$ is determined by the contextual texture consistency ε. In our method, $f(I_i)$ varies for each pixel, and it is a local variable given by the contextual texture consistency in the homogenous region of one pixel.

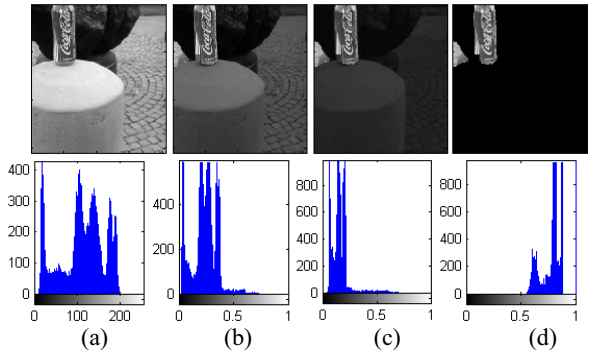


Fig.3 an example of the adaptive intensity mapping and the intensity histogram with different local contrast parameter (a) original intensity image, (b,c,d) β =0.5, 0.7, 0.9, orderly.

Fig.3 shows the example of the adaptive non-linear intensity mapping, which include the intensity histogram distribution of original image and three intensity-mapped images with different local contrast parameter β (β=0.5, 0.7, and 0.9), respectively. From the figure, it is clear that the intensity distribution of the mapped images has a more distinct peak than the distribution of the original image. And the lager local contrast parameter β is, the more regions of texture consistency are suppressed.

III. Experimental Results

We implement the proposed algorithm using MATLAB 6.5 on a PC with 3.0G Pentium IV CPU and 1G memory. To demonstrate the effectiveness of the propose attention model, we have extensively applied the method on three image databases: the Corel Dataset, SIVAL data set (from http://www.cse.wustl.edu/_sg/accio), and the STIMautobahn,STIMCoke database (from http://ilab.usc.edu/imgdbs/).

For comparison, we also show the results of Itti's model (IM). In Fig.4, the left column (a) contains the original images and the right column includes the extracted feature maps with (b) IM and (c) our proposed method orderly, the fourth column (d) is the salient region extraction using these two method (green rectangle indicates the attention region by Itti's method and red rectangle is the one by our method).

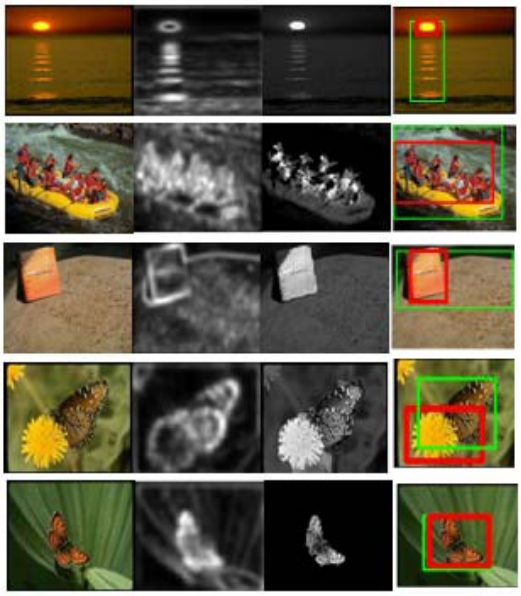


Fig.4 some results of attention analysis by IM and the proposed model

In the first and second image, the salient map using IM includes many background regions and salient region can not correspond to the main object, while our method successfully detects the salient objects: the sun in the first image and the man in the second image. Specially, for the second image, our method correctly encapsulates most of people, while that of IM includes a lager portion of the water region. In the third image, the salient object is close to one of the edges of the image, one drawback of IM is these points often gather on edges where the change of image feature is significant. In contrast, our method may discard these areas with the help of global texture consistency. In the salient map of butterfly images, there is a clearly separate dominant object in an image, both attention models focus on finding the areas where there is significant change with respect to the image feature. As a result, our approach is robust for the detection of the visual salient regions in many different situations. It is noted that, the traditional salient map only includes the information of attention importance, the larger the pixel intensity is, and the more likely the pixel attracts the observer's attention. While the salient map based on our method has more detail information, the intensity contrast for the salient object can be preserved well.

To evaluate the retrieval performance of the proposed algorithm, we choose about 5000 images of 40 categories from the COREL collection as our test image database. These categories include "beach", "sunset", "flowers", "Horses", etc. Each semantic category consists of 100 images. We use the standard measures, retrieval precision to evaluate the results. The precision is the fraction of the returned images that is indeed relevant for the query. Alike

[9], every image in each category is used as the query for retrieval in our experiment. We compared the performances of the following methods: (1)The global feature (32-bin HSV histogram) based Euclidean metric, noted as "ER"; (2) The well known SIMPLIcity system using Integrated region matching [10], noted as "IRM"; (3) The proposed method based on salient map, noted as "SR". We extract the 64-d HSV histogram as region low feature, and use Euclidean metric as region matching measure. Table 1 shows the experimental results, we can see that the IRM and our SR method are comparable, in which IRM achieves 0.62 average precision. Our SR method achieves better results than IRM and ER, i.e., 0.70 average precision. Note that for the "sunset" retrieval, IRM was fail to improvement, while our method can achieve the good performance due to better representation of mapped intensity image.

Table 1. Average Precision on Top 20 Returned images

Category	ER	IRM	SR
Horses	0.62	0.73	0.80
Car	0.56	0.61	0.69
Butterfly	0.43	0.65	0.73
Tiger	0.51	0.74	0.79
Flowers	0.51	0.65	0.73
Mountain	0.53	0.57	0.56
Sunset	0.49	0.36	0.61
AVG.	0.52	0.62	0.70

IV. Conclusions

This paper presents a computational model of visual attention to construct salient map. In particular, a novel contextual texture feature is extracted to describe spatial consistency of a region globally, which guarantees a better prediction of the saliency of each region. Then, a modeling visual attention modeling is presented based on globally contextual texture information, which provides an efficient way of constructing a unique salient map. We conducted extensive experiments to evaluate the performance of image retrieval based on extracted salient map. The promising results show that our algorithms are simple but quite effective in data representation for image retrieval.

Acknowledgments

This work was supported by the National Nature Science Foundation of China (60803072), Science Foundation of Beijing JiaoTong University (Grant No. 2007XM008) and National High Technology Research and Development Program of China (2007AA01Z168).

References

[1] T. Kadir and M. Brady, "Scale saliency: A novel approach to salient feature and scale selection," in Proc. Int. Conf. Visual Information Engineering, Surrey, U.K., pp. 25–28,Nov.2000.

[2] L. Itti, C. Koch, E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", IEEE Trans on Pattern Analysis and Machine Intelligence, vol.20(11), pp. 1254-1259. 1998.

[3] J.W. Han, King N. Ngan, et al. "Unsupervised Extraction of Visual Attention Objects in Color Images". IEEE Trans. On Circuits And Systems for Video Tech., vol.16(1), pp. 141-145,2006.

[4] Olivier Le Meur, Patrick Le Callet, et al. "A Coherent Computational Approach to Model Bottom-Up Visual Attention". IEEE Trans. On Pattern Analysis and Machine Imtelligence , VOL. 28(5), pp.802-817,2006.

[5] Y.Hu, D.Rajan, and L.T.Chia, Robust subspace analysis for detecting visual attention regions in images, ACM Multimedia, pp. 716-724. 2005.

[6] Z.W. Yu, H.S. Wong, A Rule Based Technique for Extraction of Visual Attention Regions Based on Real-Time Clustering, IEEE Trans. On Multimedia, Vol.9(4), pp.766-784, 2007.

[7] Deng YN, Manjunath BS. "Unsupervised segmentation of color-texture regions in images and video". IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23(8), pp:800~810, 2001.

[8] Arjan Gijsenji, and Theo Gevers, "Color Constancy using Natural Image Statistics", in Proc. of the Int. Conference on Computer Vision and Pattern Recognition (CVPR'07), pp. 1-8, 2007.

[9] Steven C.H.Hoi, Wei Liu, et al. "Learning Distance Metrics with Contextual Constraints for Image Retrieval ", in Proc. of the Int. Conference on Computer Vision and Pattern Recognition (CVPR'06), pp. 2072-2078, 2006.

[10] James Z. Wang, Jia Li and Gio Wiederhold, "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries" IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 9, pp. 947-963, 2001.