Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 89 (2016) 607 – 613

Procedia Computer Science

Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016)

Finding Visual Attention in an Image: Learning through Hierarchal Inheritance Approach

Thoorpu Karnakar Reddy*, Mangalraj and Anupam Agrawal

Indian Institute of Information Technology, Allahabad

Abstract

Finding visually attractive region in an image is one of the interesting research area due to its real time applications. Generally we take the images to capture objects (visually attractive region). In this paper the proposed approach regards the important object detection in an image. The proposed algorithm contains multiple phases: preparing the super-pixel using ensemble decision trees to estimate the adjacency probability between the regions, and forming the saliency vector (visually attractive region) using the learned weights through hierarchal inheritance approach. The proposed approach is tested on several popular benchmark data-sets and performed well among the present approaches. The results were compared and shown in section 6.

Peer-review under responsibility of organizing committee of the Organizing Committee of IMCIP-2016

Keywords: Saliency Vector; Salient Object; Superpixel; Visually Attractive Object; Visually Attractive Region.

1. Introduction

Eye fixation has been a basic problem in cognitive sciences, neuroscience, visual saliency and computer vision and image processing for a long time[1,2]. Eye fixation has been identified as the problem of identifying the important region[3] (visually attractive region) in the given image. There are variant real time applications of salient object detection for instance object recognition[4,5], image compression[6], image cropping[7], photo collage[8], dominant color detection[9] and so on.

According to the study of cognitive science the visually attractive region is the uniqueness, rarity and surprise of an image. The characteristics of the visual attention is based on the variety of features for instance position, color, appearance, contrast[10,11] (local and global perspective), and so on.

Detecting the important object in an image by humans is a trivial task but it is not for the systems. In this paper the proposed algorithm deals how to find out the visual attention in an image by using learned weights (generated by hierarchal segmentation approach). The visually attractive region detection is the regression problem and the supervised (random forest regressor) learning technique is used to assign the saliency values to the regions of segmented image and these values are used to find the saliency vector (visually attractive region).

The key assets of this paper are 1) formation of super-pixels used for primary segmentation (acts as regions) in a given image 2) choosing the appropriate features for finding out the visually attractive region 3) providing the

*Corresponding author. Tel.: +919494979696;
E-mail address: thoorpukarnakar@gmail.com

1877-0509
Peer-review under responsibility of organizing committee of the Organizing Committee of IMCIP-2016
doi:10.1016/j.procs.2016.06.021

values to the regions based on the ensemble (random forest regressor) technique[1,33] 4) weights (W), generated by hierarchal inheritance approach, provide the visually attractive object without regressive search on complete image. In the proposed approach the feature vectors are compared between the regions locally and globally[2,12]. The background removal also plays a vital role in finding out the visually attractive object, the proposed approach handles it effectively.

2. Related Work

The following gives review of how the research is going on towards the salient object detection. A comprehensive study on salient object detection be available[13]. The review on visual saliency, eye fixation and also included analysis on salient object detection.

Most of the important object detection algorithms are based on the feature combining theory[14] which integrates the several features for the prediction of important object detection. The biologically-plausible architecture was proposed[15] which is certainly depends on the studies concerning the detection, localization, and recognition of objects in visual fields.

In the recent past very intensive research is going on, to specify what are the effective feature vectors to find out the salient object detection. The variant centre-surround hypothesis is analyzed[16]. Centre-surround differences are generally calculated based on the color histograms. A strong mathematical formulation was proposed[17].

Centre-surround difference acts as one of the feature vector in salient object detection. The difference between the neighbouring regions will be useful to calculate the saliency values of the regions[18]. The global and regional contrast differences are computed to calculate the salient features. The global contrast differences are computed between the regions and the regional contrast differences are calculated within the region[19].

Different approaches have been proposed to find out the salient object detection. Usually the salient object will be present in the middle of the image, by the majority cases, this is known as centre-bias[18]. The related concepts such as auto-context cue[20]. A graphical based approach was used to generate the salient object[3]. A matrix formation scheme is proposed to detect the important objects[21].

Besides to the above approaches, several approaches have been proposed to detect the important regions in a given image. By combining the image cues are used to find out generic objectness measurement[23]. Checks whether by decomposing the neighbouring regions will form the salient object[24].

Visual saliency is another important area used for the salient object detection. In the recent past developments includesthe isocentric curvedness and colorness[25], image histograms[26], utilization of depth cues[27], top-down and bottom-up feature specification[28], and etc. There are many other saliency definitions aiming to detect the visual attention in a given image.

The proposed approach differs from the existing approaches in feature integration and learning to combine the segmented regions in a given image. The advantage of learned through hierarchal inheritance approach is that the regressive search on complete image for finding the visually attractive regions is not required. The proposed approach learns to automatically assign the values to the regions[2]. In most of the existing algorithms the saliency vector generation is done by combining the different types of feature vectors[3,23]. In the proposed approach the random forest regressor (Ensemble learning) is used to provide the appropriate saliency values to the segmented regions rather than simply combining or integrating the features of the given image. In the proposed approach both regional and global feature differences are used. The recent learning approach[3] aims to find out the eye fixation.

3. Visually Attractive Region Computation

The process of finding the visually attractive region in the given image contains various phases: segmenting the image in to different regions based on the features (color, background, contrast, properties and so on), providing the values to the regions based on the learned ensemble (random forest regressor) technique, and weights (W), learned by hierarchal inheritance approach, and mapping the saliency vector (svector).

(a) (b) (c)

Fig. 1. (a) Input (b) Segmented Results after Super-pixel Operation (c) Final Result of the Proposed Approach.

3.1 Segmentation

In the present work, segmentation is performed based on graph based methodology[29] besides to it by combining the pixels using the variant feature vectors for instance color, contrast, and geometrical orientations called super-pixel (P). The ensemble decision tree (supervised learning technique) is used to estimate adjacency probability between the regions, which performed well in almost all the standard benchmark datasets. The regions are segmented based on two factors that is internal (inside of a region) and external (outside of the region) histogram feature differences of every region of the given image Fig. 1(a) depicts the input for segmentation and Fig. 1(b) depicts the segmented image using super-pixel approach.

$$P(I) = h(I) \Big/ \sum h(I)$$

$h(I)$: Intensity calculation of given image (I) based on different color models.

3.2 Computing saliency values

The segmented image I with M regions and each region is awarded with the value V. The learned regressor calculates the saliency values to every region in segmented image. In the proposed approach the random forest regressor computes the saliency values to the regions based on combined features. The feature vector z is the composition of different dimensionalities (color, height, width, contrast etc.). To quantize the color feature vector we used three different types of color models RGB, LAB, HSV. The random forest regressor (supervised learning technique) is trained to yield the saliency values to the regions based on the above mentioned features.

3.3 Computing the visually attractive region

The novel mechanism based on weights is used to form saliency vector. The weights (learned by hierarchal inheritance mechanism) are multiplied with every pixel to elevate the visually attractive object. The hierarchal inheritance approach is explained in learning section. Figure 1(c) depicts the result of the proposed approach.

4. Feature Specification

Feature specification is an important aspect in differentiating the regions from its neighbouring regions in the given image. Generally the region is visually attractive if it is different from its neighbouring regions. The feature specification in the proposed approach is unlike to most of the existing approaches for computing the salient object. The proposed approach computes the regional and global feature descriptor, which will be pumped into the regressor to assign the values to every region.

All the features are described in a feature vector z, includes contrast, color, background, similarities between the regions (which is a pre-processing for feature extraction), and properties (object orientation, texture characteristics . . .) and so on. For the contrast feature and histogram feature the algorithm computes the differences between the

Table 1. Features used for Building Color and Texture Feature Vector.

Color and texture features	
Features	No. of dim's
The mean RGB values	3
The mean LAB values	3
The absolute responses of filters	49
The response of filters	1
The LAB histogram	8*16*16
The HSV histogram	8
The saturation histogram	8
The texture histogram	65

neighboring regions (regionally and globally) of the image, which will be helpful in differentiating (boundary tracking) the regions of the given image.

The visual characteristics of every region are described by the color and texture features as a result almost 26 feature dimensionalities are used for the feature specification. To check the independency between the regions chi squared test is used $\chi^2 = \sum (2(a_{1i} - a_{2i})^2/(a_{2i}))$. Color and texture features are classified into different features (shown in Table 1).

The dimensions of geometric features are height, width, and position of the objects in a given image. In several existing algorithms[30] the background is calculated heuristically based on the characteristics but in the proposed approach the background of the given image is calculated from the corners of the image, i.e., around 10–15 pixels from borders of an image are considered as the pixels belongs to background.

5. Learning

Learning the object, visually attractive regressor is one of the challenging task in important object detection of an image. The learned regressor will provide the saliency values to every region through which the important region will be detected. The training set includes the set of images, whose regions C are confident $C = \{c_1, c_2, \dots\}$ and the respective saliency values are $S = \{s_1, s_2, \dots\}$. The confident region in an image belongs to the important region of that image[29].

As aforementioned we use an ensemble learning technique to assign the saliency values to the regions. In learning process each object is provided by the feature matrix z, defined by the different dimensionalities such as color, appearance, height, width, contrast, backgroundness, and so on. Learned regressor is helpful in predicting the saliency values to every segmented region.

The proposed algorithm uses the learned weights, which are built by the hierarchal segmentation (Inheritance) approach. The learned weights are multiplied by the each pixel value to obtain the saliency vector. As aforementioned the weights are obtained by the hierarchal segmentation (Inheritance) approach, in this approach image is segmented into multiple levels such as I (given image) $= \{i_1, i_2, \dots i_k\}$. On every level of the image the learned regressor maps the saliency values to all the segmented regions. Every level of the image (i_k) is obtained from previous level of the image (i_{k-1}) by combining the sorted regions (based on the saliency values) up to some threshold value[29] $(<= t$ (say)). The weights (W) are obtained by hierarchal inheritance approach. Below equations serves the purpose. Equation (1) used to build the matrix $H(n, n)$, where n is the number of segmentation levels, based on summing up the segmented images. Equation (2) used to build the matrix $f(n, 1)$ based on summing up the given image and segmented images. Equation (3) is the quadratic function (quadprog), I, (I_i, I_j) are the given Image, segmented images respectively.

$$H = \sum_{i=1}^{n} \left(\sum_{j=1}^{k} H(i, j) + \text{sum}\,(I_i * I_j) \right) \tag{1}$$

$$f = \sum_{i=1}^{n} \left(\sum_{j=1}^{k} f(i, j) - 2 * \text{sum}\,(I * I_j) \right) \tag{2}$$

$$W = \frac{1}{2}\, x^T\, Hx + f^T\, x \tag{3}$$

6. Experimental Results

6.1 Performance analysis

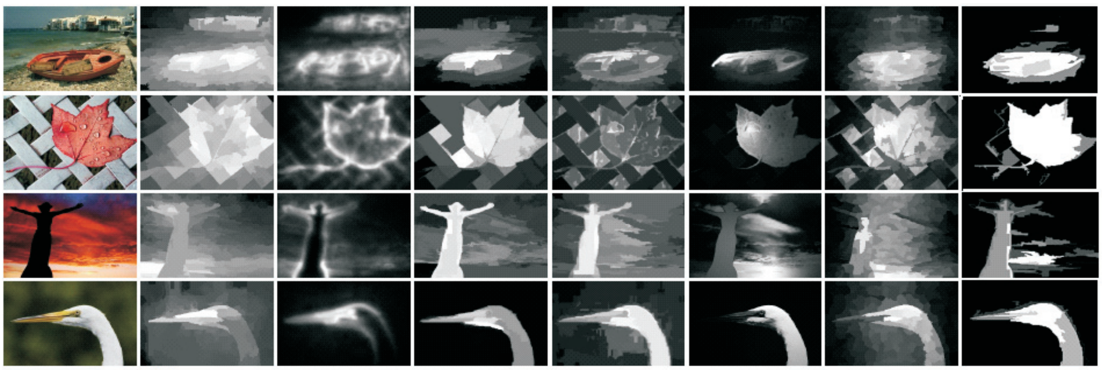
The proposed approach for executing the visual attention in the given image is evaluated on three data sets (MSRA-B, SED) that are used in previous approaches. MSRA-B (includes 5000 images), SED (includes two subsets). The performance evaluation is done using the measures[13].

6.2 Performance comparison

The visual comparison and performance comparison is shown on Fig. 2 and Table 2 respectively. As can be seen the proposed approach achieved the better results among some of the existing approaches, the performance on the MSRA-B and SED1 data sets shows the uniqueness of the proposed approach. The PR and ROC curves of the proposed approach are showing best on these two data sets. Table 2 (3), corresponding to the SED2 data set, shows that the performance is not up to the mark due to the dataset contains multiple important objects. Intuitively, the proposed approach has limited ability when finding out the multiple attractive objects within an image. The reason might be the orientations and size of the two or more objects in SED2, which are very different from the training set of SED1, where most of the images contain only one important object. The comparisons are shown in Fig. 2 and Table 2. The visual comparison depicts that our approach outperforms for the purpose.

6.3 Evaluation

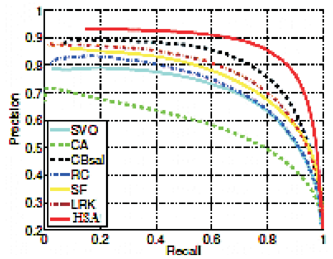
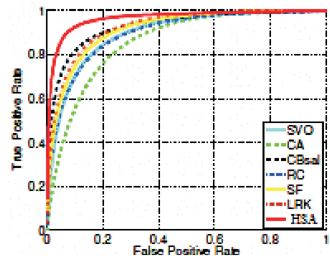
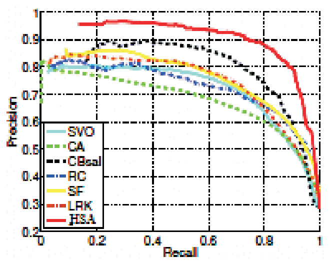
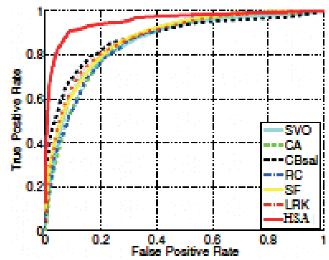
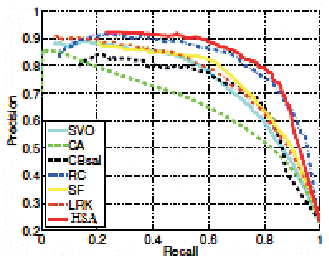
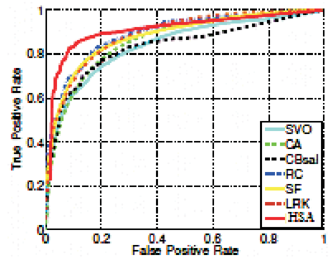
The proposed approach takes five hours for building the weights (W) using hierarchal inheritance approach and four sec's to execute (finding important object) the image (400×300) with an Intel i3 CPU. Matlab code is the optimized one and it can be accelerated on GPU computing systems.



(1) Input (2) SVO[3] (3) CA[31] (4) CBsal[18] (5) RC[12] (6) SF[32] (7) LRK[21] (8) HSA

Fig. 2. Visual Analysis of Different Approaches with Proposed Approach (Hierarchal Segmentation Approach (HSA)).

Table 2. Performance Analysis of Different Approaches with Proposed Approach on Popular Data sets.

Datasets	Precision and Recall Curves	Receiver Operating Characteristic Curves
1) MSRA-B		
2) SED1		
3) SED2		

7. Conclusions

The proposed approach addresses, finding the visual attention in a given image in an efficient way among the present approaches. The major challenges dealt in the present work are preparing the super-pixel using ensemble decision trees, and hierarchal Inheritance approach for generating the weights (W). The performance results on various popular benchmark data sets proves that our approach is best for the purpose. The proposed approach is executing in four seconds, which is 60 percentages faster than the present approach (DRFI[2]). The MATLAB implementation is available online.

References

[1] Dietterich and G. Thomas, Ensemble Methods in Machine Learning, In *Multiple Classifier Systems*, Springer Berlin Heidelberg, pp. 1–15, (2000).

[2] Jiang, Huaizu, Jingdong Wang, Zejian Yuan, Yang Wu, Nanning Zheng and Shipeng Li, Salient Object Detection: A Discriminative Regional Feature Integration Approach, In *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE*, pp. 2083–2090, (2013).

[3] Chang, Kai-Yueh, Tyng-Luh Liu, Hwann-Tzong Chen and Shang-Hong Lai, Fusing Generic Objectness and Visual Saliency for Salient Object Detection, In *2011 IEEE International Conference on Computer Vision (ICCV), IEEE*, pp. 914–921, (2011).

[4] Kanan, Christopher and Garrison Cottrell, Robust Classification of Objects, Faces, and Flowers using Natural Image Statistics, In *2010 IEEE Conference on IEEE Computer Vision and Pattern Recognition (CVPR)*, pp. 2472–2479, (2010).

[5] Walther, Dirk and Christof Koch, Modeling Attention to Salient Proto-objects, *Neural Networks*, vol. 19, no. 9, pp. 1395–1407, (2006).
[6] Itti and Laurent, Automatic Foveation for Video Compression using a Neurobiological Model of Visual Attention, *IEEE Transactions on Image Processing*, vol. 13, no. 10 pp. 1304–1318, (2004).
[7] Marchesotti, Luca, Claudio Cifarelli and Gabriela Csurka, A Framework for Visual Saliency Detection with Applications to Image Thumbnailing, In *2009 IEEE 12th International Conference on Computer Vision, IEEE*, pp. 2232–2239, (2009).
[8] Goferman, Stas, Ayellet Tal and Lihi Zelnik-Manor, Puzzle-like Collage, In *Computer Graphics Forum*, Blackwell Publishing Ltd., vol. 29, no. 2, pp. 459–468, (2010).
[9] Rui, Yong, Thomas S. Huang and Shih-Fu Chang, Image Retrieval: Current Techniques, Promising Directions, and Open Issues, *Journal of Visual Communication and Image Representation*, vol. 10, no. 1, pp. 39–62, (1999).
[10] Lowe and G. David, Object Recognition from Local Scale-invariant Features, In *1999 The Proceedings of the Seventh IEEE International Conference on Computer Vision, IEEE*, vol. 2, pp. 1150–1157, (1999).
[11] Achanta, Ravi, Sheila Hemami, Francisco Estrada and Sabine Susstrunk, Frequency-tuned Salient Region Detection, In *2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE*, pp. 1597–1604, (2009).
[12] Cheng, Ming, Niloy J. Mitra, Xumin Huang, Philip HS Torr and Song Hu, Global Contrast Based Salient Region Detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3 pp. 569–582, (2015).
[13] Borji, Ali, Dicky N. Sihite and Laurent Itti, Salient Object Detection: A Benchmark, In *Computer Vision–ECCV 2012*, Springer Berlin Heidelberg, pp. 414–429, (2012).
[14] Treisman, M. Anne and Garry Gelade, A Feature Integration Theory of Attention, *Cognitive Psychology*, vol. 12, no. 1 pp. 97–136, (1980).
[15] Koch, Christof and Shimon Ullman, Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry, In *Matter of Intelligence*, Springer Netherlands, pp. 115–141, (1987).
[16] Gao, Dashan, Vijay Mahadevan and Nuno Vasconcelos, The Discriminant Center-surround Hypothesis for Bottom-up Saliency, In *Advances in Neural Information Processing Systems*, pp. 497–504, (2008).
[17] Klein, Dominik and Simone Frintrop, Center-surround Divergence of Feature Statistics for Salient Object Detection, In *2011 IEEE International Conference on Computer Vision (ICCV), IEEE*, pp. 2214–2219, (2011).
[18] Jiang, Huaizu, Jingdong Wang, Zejian Yuan, Tie Liu, Nanning Zheng and Shipeng Li, Automatic Salient Object Segmentation Based on Context and Shape Prior, In *BMVC*, vol. 6, no. 7, pp. 9, (2011).
[19] Niu, Yuzhen, Yujie Geng, Xueqing Li and Feng Liu, Leveraging Stereopsis for Saliency Analysis, In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE*, pp. 454–461, (2012).
[20] Xue, Jianru, Le Wang, Nanning Zheng and Gang Hua, Automatic Salient Object Extraction with Contextual Cue and its Applications to Recognition and Alpha Matting, *Pattern Recognition*, vol. 46, no. 11 pp. 2874–2889, (2013).
[21] Shen, Xiaohui and Ying Wu, A Unified Approach to Salient Object Detection via Low Rank Matrix Recovery, In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE*, pp. 853–860, (2012).
[22] Alexe, Bogdan, Thomas Deselaers and Vittorio Ferrari, What is an Object? In *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE*, pp. 73–80, (2010).
[23] Wei, Yichen, Jie Feng, Litian Tao and Jian Sun, Salient Object Detection by Composition, U.S. Patent 8,989,437, Issued March 24, (2015).
[24] Valenti, Roberto, Nicu Sebe and Theo Gevers, Image Saliency by Isocentric Curvedness and Color, In *2009 IEEE 12th International Conference on Computer Vision, IEEE*, pp. 2185–2192, (2009).
[25] Lu, Shijian and Joo-Hwee Lim, Saliency Modeling from Image Histograms, In *Computer Vision–ECCV 2012*, Springer Berlin Heidelberg, pp. 321–332, (2012).
[26] Lang, Congyan, Tam V. Nguyen, Harish Katti, Karthik Yadati, Mohan Kankanhalli and Shuicheng Yan, Depth Matters: Influence of Depth Cues on Visual Saliency, In *Computer Vision–ECCV 2012*, Springer Berlin Heidelberg, pp. 101–115, (2012).
[27] Borji and Ali, Boosting Bottom-up and Top-down Visual Features for Saliency Estimation, In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE*, pp. 438–445, (2012).
[28] Felzenszwalb, F. Pedro and Daniel P. Huttenlocher, Efficient Graph-based Image Segmentation, *International Journal of Computer Vision*, vol 59, no. 2, pp. 167–181, (2004).
[29] Wei, Yichen, Fang Wen, Wangjiang Zhu and Jian Sun, Geodesic Saliency using Background Priors, In *Computer Vision–ECCV 2012*, Springer Berlin Heidelberg, pp. 29–42, (2012).
[30] Goferman, Stas, Lihi Zelnik-Manor and Ayellet Tal, Context-aware Saliency Detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1915–1926 (2012).
[31] Perazzi, Federico, Philipp Krähenbühl, Yael Pritch and Alexander Hornung, Saliency Filters: Contrast Based Filtering for Salient Region Detection, In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE*, pp. 733–740, (2012).
[32] Liaw, Andy and Matthew Wiener, Classification and Regression by Random Forest, *R news*, vol. 2, no. 3 pp. 18–22, (2002).