Volume 4, Issue 8, August 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Compendium of Salient Object Detection

[1]Arvinda Sharma, [2]Dr. Rajiv Mahajan
[1]M.Tech Scholar :Computer Science & Engineering Dept, Global Institutes of Management& Emerging Technology, India
[2]Professor: Computer Science & Engineering Dept. Global Institutes of Management &Emerging Technology, India

Abstract: - Object detection is a skill associated to computer vision and image processing that deals with detecting occurrences of meaningful objects of a category such as humans, buildings, or cars in digital images and videos. Paper here describes the different salient object detection techniques. Salient object detection is an image segmentation technique where we separate the salient object from the image background. It has applications in many areas of computer vision, including image retrieval and video surveillance. This paper explores the range of methods for resourcefully extracting the objects from digital images and in last explains the short comings of the existing techniques.

***Keywords- saliency, random forest, frequency, Efficient Maximum Appearance*.**

I. INTRODUCTION

Saliency detection [2] has been an area of extensive research in the recent vision literature. Salient objects have the features. Each feature describes either the whole image or the part of it. Earlier some work is done to envisage the locations of human eye fixations and so induce the fundamental principles of saliency detection. Some methods of detecting saliency are bottom-up and some are top-down saliency. Top to down type uses high -level knowledge for saliency computation, calculating saliency to discover the stimuli from certain object classes. While bottom-up saliency predicts eye fixations, so these methods are useful for high-level vision tasks, such as object recognition and localization. Conciliation is to either consider the responses of object detectors as features for bottom-up saliency, or invent saliency as a mid-level vision task. In this area, significant attention has been devoted to the problem of object saliency.

Detecting [1] such entities in digital images is preprocessing tread of segmentation. There are large numbers of methods that segment the multiple objects the scene. Each method has its own advantages and disadvantages. Object detection will discount [1] the effect of background from image and then detect the objects of interest. Saliency denotes the most protruding part in the images and Saliency maps compute saliency value for each pixel digitally. Saliency initiates from visual inimitability, randomness, infrequency, or revelation, and is often ascribed to variations in image attributes. The salient object is unique entity which is totally different from its surroundings and always have well established boundary.

Visual saliency [1], being closely related to how we perceive and process visual stimuli, is investigated by multiple disciplines including cognitive psychology, neurobiology, and computer vision. Computationally detecting such salient image regions remains a significant goal, as it allows preferential allocation of computational resources in subsequent image analysis and synthesis. There are large no of methods [1] exist for extracting object from image. Traditional methods often compute saliency based on contrasts, either local or global analysis of the contrast. Local methods are more sensitive to high-contrast edges and noise, and attenuate smooth regions in objects, which makes them more appropriate for detecting small objects. For global methods, the patch-based approach also tends to highlight the object boundaries rather than the whole object area.

Problems in object detection [1] arise due to abrupt object motion, changing look patterns of an object with respect to its background, non-rigid object structures.

Various Salient Object detection techniques that are reviewed here are as follows:

- Salient Object detection via random forest
- Efficient Maximum Appearance Search for Large-Scale Object Detection
- Global contrast based salient region detection
- Frequency tuned saliency region detection

Fig1: Salient Object Detection [4]

II. TECHNIQUES OF SALIENT OBJECT DETECTION

(a) Salient Object detection via random forest

Salient object detection [3] is a chief inclination in modelling saliency, which aims to situate the most fascinated object in a scene. In recent times, it attracts more attention as it serves as a pre-processing step for many applications. Conventional methods figure out saliency based on contrasts. Local methods are receptive to high-contrast edges and noise, and assuage flat regions in objects, which makes them suitable for detecting small objects. For global methods, the patch-based approach also tends to highlight the object boundaries rather than the whole object area. Though segmentation-based methods lessen the objects attenuation problem effectively, they still have difficulties of highlighting the entire object.

Here the approach mentioned is patch-based approach [3] to detect the salient object through its approximate contour. First, an image is taken, and then it is divided into patches. Its patches are adjusted to left and right side of the tree , and hence the random forest is constructed .So we can explain here the concept of Random Forest[8] as ensemble classifier that is a collection of number of decision trees. Although the performance of ensemble classifier is highly related to the correlation among each model and the trees of random forest are constructed with some randomization of factors. So we can say randomization comes from two points: Sub sampling the training data and each tree are grown with different data and for each internal node, selecting some attributes for split and also each internal node contains a best split of training data.

This division of patches to right and left side are done by taking in account some of the features such as color, feature, dimension and some angular distance. Here splitting of patches is done on basis of this equation [4],

$$tr_n\left(s_n; h_{1,} h_2\right) = \begin{cases} p_i \epsilon s_l, & \text{if } d_i(h_{1,} h_2) \leq \theta_{h1,h2} \\ p_i \epsilon s_r, & \text{otherwise} \end{cases}$$

Where s_l and s_r $a$re [3] the patch sets contained in node's left and right child, $\theta_{h1,h2} = \frac{1}{|s_n|} \sum \quad \forall p_i \epsilon s_n \, d_i(h_{1,} h_2)$, $|s_n|$ is the cardinality of s_n .After the forest is built, we use it to measure the rarities of patches and compute similarities among them, equation [3]:

$$cS(p_i) = \frac{1}{\sum_{i=0}^{n} \quad |L_k|} . w(x_i, x_c),$$

The above equation is used to evaluate the rarity of p_i, where $|L_k|$ is the number of patches contained in$L_k . w(x_i, x_c)$.Through level set based Active Contour model, we will discover the salient object according to the detected contour C.

The basic idea of ACM [9] is to evolve a curve under some constraints to extract the desired object. According to the nature of constraints, the existing ACMs can be categorized into two types: edge-based models and region-based models. Some edge-based ACMs [9] introduce a balloon force term to shrink or expand the contour, yet it is difficult to design the balloon force. If the balloon force is large, the contour will pass through the weak edge of the object. On the other hand, if the balloon force is not large enough, the contour may not pass through; Very narrow part of the object. In addition, the edge-based models are prone to local minimum, failing to detect the exterior and interior boundaries when the initial contour is far from the desired object boundary.

Region-based ACMs have many advantages over edge-based ones. First, region-based models utilize the statistical information inside and outside the contour to control the evolution, which are less sensitive to noise and have better performance for images with weak edges or without edges. Second, they are significantly less sensitive to the location of initial contour and then can efficiently detect the exterior and interior boundaries simultaneously. One of the most popular region-based models is the C–V model, which is based on Mumford–Shah segmentation techniques and has been successfully applied to binary phase segmentation. We partition these patches into two subsets, inside and outside, respectively. The formal definition can be written as [3]:

$$\mathbf{p}_i \in \begin{cases} S_{in}, & \text{if } \frac{|\mathbf{p}_i \cap C|}{r^2} > \lambda, \\ S_{out}, & \text{otherwise,} \end{cases}$$

Where $|\mathbf{p}_i \cap C|$ denotes the number of pixels inside the extracted contour, λ is a constant, and its value varies between 0 and 1 according to the dataset. The p_i and p_j are similar patches if they fit in to sale leaf node and then we measure the contrasts between the inner patches and the outer patches, aiming to suppress the patches of the inner part similar to the outer patches while highlight the outer patches similar to the inner patches. Through the contrasts among patches, our methods highlight the whole object uniformly.This contour extraction method is simple and no edge detector is required. Finally, we do graph-cut based segmentation.

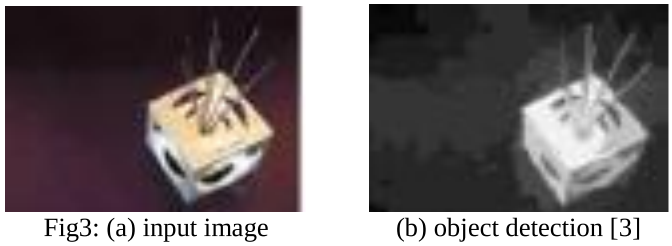

Fig3: (a) input image (b) object detection [3]

(b) Frequency tuned salient Region detection

The focal point of this paper is the habitual detection of visually salient regions in images, which is useful in applications such as adaptive content delivery [5], adaptive region-of-interest based image compression image segmentation, object recognition. Our algorithm finds low-level, primitive, bottom-up saliency. It is inspired by the genetic concept of centre-surround contrast, but is not based on any biological model.

Existing methods of saliency detection produce region that have stumpy pledge, scantily defined limitations, or are expensive to compute. Some methods fabricate higher saliency ethics at object edges as an alternative of generating maps that uniformly cover the whole object, which results from failing to exploit all the spatial regularity content of the original image. We analyze the spatial frequencies in the original image that are retained by some of the methods, and visually exemplify that these techniques primarily operate using extremely low-frequency content in the image. We introduce a frequency-tuned approach to estimate centre-surround contrast using colour and luminance features efficiency. The saliency map generated can be more effectively used in many applications, and here we present results for Object segmentation.

Visual consideration marks jointly from fast, pre-attentive, bottom-up saliency of the retinal input, as well as from slower top to down reminiscence and craving based processing that is objective-dependent.The frequency tuned method [5] places of interest the complete salient regions. In it, let [5] W_{LC} the low frequency cut-off value and W_{HC} be the high frequency cut-off value.

To bring to light the salient objects, we necessitate by considering very low frequencies from the original Image, i.e. W_{LC} has to be low. This also helps to highlight salient objects unvaryingly order to have fully defined boundaries, we need to preserve soaring frequencies from the original image, i.e. W_{HC} has to be high to evade noise, coding artifacts, and surface pattern. Saliency map S as for an image I of width W and height H pixels can thus be formulated as [5]:

$$S(x,y) = |I_U - I_{Whc}(x,y)|$$

Where I_U the arithmetic mean pixel [5] is value of the image and $I_{Whc}(x,y)|$ is the Gaussian blurred version of the inventive image to eradicate all right texture details as well as noise and coding artifacts. The standard of the difference is used as we are interested only in the enormity of the difference. This is computationally fairly proficient .We activate on the original image without any down variety and obtain a full resolution saliency map. Our method of finding the saliency map S for an image I of width W and height H pixels can thus be formulated to use features of color and luminance; we rewrite it as [5]:

$$S(x,y) = ||I_U - I_{Whc}(x,y)||$$

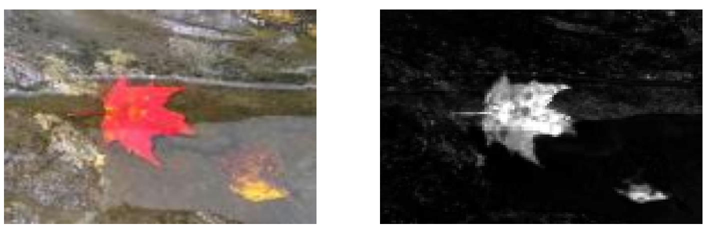

Fig4: (a) Original image (b) saliency map [5]

This analysis illustrated that the deficiencies of these techniques arise from the use of an inappropriate range of spatial frequencies. Based on this analysis, we presented a frequency tuned approach of computing saliency in images using low level features of colour and luminance, which is easy to implement, fast, and provides full resolution saliency maps. The resulting saliency maps are better suited to salient object segmentation, demonstrating both higher precision and better recall than other methods.

(c) Global contrast based salient region detection

In Global contrast based saliency [6] calculation methods, the HC method is fast and generates results with fine details, the RC method generates coherent high quality saliency maps at the cost of reduced computational efficiency. A regional contrast based saliency extraction method simultaneously evaluates global contrast differences and spatial coherence. Such method is efficient, and yields full resolution saliency maps.

i. Histogram based contrast is done on equation

$$S(I_{K)=\sum_{\forall I_i \in I} D(I_k,I_i})$$

Where $D(I_k, I_{i)}$ is the color distance metric between pixels I_k *and* I_i in the L*b*c color space.

The terms with the same color value Cj are grouped together; we get saliency value for each color as [6],

$$S(I_{K)=}S(c_l) = \sum_{j=1}^{n} f_j\, D(c_l, c_j)$$

Here the symbol of equation c_l is the color value of pixel I_K, n is the number of distinct pixel colors, and f_j is the probability of pixel color c_j in image I.

ii. Color Space Smoothing is done

$$S'(c) = \frac{1}{(m-1)T}\sum_{i=1}^{m}\big(T - D(c,c_i)\big)S(c_i)$$

Where the symbol [6] $T = \sum_{i=1}^{m}\big(D(c,c_i)\big)S(c_i)$, is the sum of distances between color c and its m nearest neighbors c_i and the normalization factor comes from $\sum_{i=1}^{m}\big(T - D(c,c_i)\big) = (m = 1)T$.

iii. Region based contrast

We first segment the input image into regions using a graph based image segmentation method. Then we build the color histogram for each region. For a region r_K we compute its saliency value by measuring its color contrast to all other regions in the image, [6]

$$S(r_{K)=}\sum_{r_k \neq r_i} w(r_i)D_r(r_i, r_k)$$

Where $w(r_i) is$ the weight of region r_i and $D_r(,)$ is the color distance metric between the two regions.
Finally it is seen that the above explained scheme is superior in terms of both precision and recall, while still being simple and efficient.

Fig5: (a) Input Image (b) Saliency maps (c) Region of interest [6]

(d) Efficient Maximum Appearance Search for Large-Scale Object Detection

Large-scale object detection [7] is an imperative vision predicament that is concerned with detecting a huge number of object categories and localizing them in a large number of images. Remarkable amount of research has fixed on developing new feature representations and classification algorithms to boost accuracy of scale object detection. A common thread that ties most of the different approaches together would be detection models that are measured to different object shape from background on densely sampled inner parts of window image. Among these approaches, the template-based approaches, such as the popular Deformable Part Model near models constructed from a number of part uses templates of image gradient features. Since templates are sensitive to sampling scale and the pose of objects, inference of such models often entails exhaustively searching for the best template configuration regarding pose, scale, rotation, etc.
Maximum of existing object detection methods spotlight on recuperating accurateness of large-scale object detection with competence being an afterthought. The technique presented here is faster than the existing approaches, while maintaining comparable accuracy. Such a model consists of representing an image as an ensemble of densely sampled feature points with the proposed Point wise Fisher Vector encoding method, so that the learnt discriminative scoring function can be applied locally.

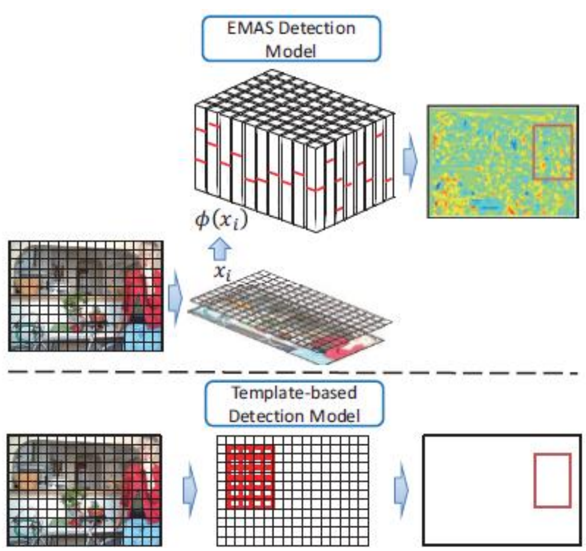

Fig [7]: Upper part is the proposed EMAS model; while the lower part is the template-based detection with exhaustive convolution over scales and positions.

Fig illustrates that we represent an image as an ensemble of densely sampled feature points with the proposed Point wise Fisher Vector encoding. This enriched local representation enables us to transform the object detection problem into searching for an image sub-window with maximum sum of object possibility, which can be performed extremely perfectly. The advantage of low computation complexity enables us to explore the large scale object detection problem with huge number of categories.

By the time, the object detection problem is transformed into searching an image sub-area for maximum local appearance probability, thereby making Efficient Maximum Appearance Search an order of magnitude faster than the traditional detection methods. In addition; the proposed model is also suitable for incorporating global context at a negligible extra computational cost. EMAS can also incorporate fusion of multiple features, which greatly improves its performance in detecting multiple object categories. We have designed an efficient large-scale object detection approach by extending Fischer Vector encoding to the point-level. This enabled us to transform the object detection problem into a problem of searching for a sub-window with the maximum sum leading to an order of magnitude of speed-up over the other some of the methods. The proposed approach could further integrate global object contextual information into the detection model with some extra info to make object detection a bit effective.

III. CONCLUSION

Object Detection [1] is a technology of image processing that deals with discovering instances of objects such as humans, cars, buildings in digital images and videos. In this paper we review the existing techniques of salient object detection. Frequency tuned salient region detection that offers three advantages over existing Methods: uniformly highlighted salient regions with well defined boundaries, full resolution, and computational efficiency. The saliency map generated can be more effectively used in many applications. Their [5] analysis illustrated that the deficiencies of these techniques arise from the use of an inappropriate range of spatial frequencies.

The salient object detection through use of random forest have many of the disadvantages that if no object will be found then level set based method will be very poor to operate and also normal thresholding will add more pixels or delete the existing due to presence of noise.Some other of the disadvantages is complexity of background part, no object found, few techniques are applicable only for some natural scenes, many methods have overhead of reinitializing the energy functions, some have reduce the large number of trees[8], often results in poor performance i.e. will leave the most important one's from the view of whole system, some may output the object on basis of contrast which may undergo error if noise is present in the images. So these methods [1] still have difficulties of highlighting the entire object when the inner region of the object is inhomogeneous. To reduce the problems of existing literature a new technique will be proposed in near future that will consider all the weaknesses of various methods and will produce better results.

ACKNOWLEDGMENT

I would like to express my truthful thanks to Dr. Rajiv Mahajan (H.O.D.), Computer Science, Global Institutes of Management and Emerging Technology, Amritsar, Punjab for his guidance, motivation and productive suggestions that help us in the preparation of this paper.

REFERENCES

[1] Arvinda Sharma"A scrutiny of detecting salient object in digital images". International Journal of advanced research in computer science and software engineering, IJARCSSE in Volume 4, Issue 5, 2014.

[2] Lu, Song, Vijay Mahadevan, and Nuno Vasconcelos. "Learning Optimal Seeds for Diffusion-based Salient Object Detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013.

[3] Du, Shuze, and Shifeng Chen. "Salient Object Detection via Random Forest." (2013): 1-1

[4] Jiang, Huaizu, et al. "Automatic salient object segmentation based on context and shape prior." BMVC. Vol. 3. No. 4. 2011

[5] Achanta, Radhakrishna, et al. "Frequency-tuned salient region detection."Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.

[6] Cheng, Ming-Ming, et al. "Global contrast based salient region detection."Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011

[7] Chen, Qiang, et al. "Efficient maximum appearance search for large-scale object detection." Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE, 2013.

[8] Chiu, Tzu-Hsuan, et al. "Sharing the trees among random forests for effective and efficient concept detection." Multimedia Signal Processing (MMSP), 2012 IEEE 14th International Workshop on. IEEE, 2012.

[9] Zhang, Kailua, et al. "Active contours with selective local or global segmentation: A new formulation and level set method." Image and Vision Computing 28.4 (2010): 668-676.

[10] Gall, Juergen, Nima Razavi, and Luc Van Gool. "An introduction to random forests for multi-class object detection." Outdoor and Large-Scale Real-World Scene Analysis. Springer Berlin Heidelberg, 2012. 243-263.

[11] Prasad, Dilip K. "Survey of the problem of object detection in real images."International Journal of Image Processing (IJIP) 6.6 (2012): 441.

[12] Zhang, Chao, et al. "Contour detection via random forest." Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012

[13] Ksantini, Riadh, BoubakeurBoufama, and Sara Memar. "A new efficient active contour model without local initializations for salient object detection."EURASIP Journal on Image and Video Processing 2013.1 (2013): 1-13.

[14] Feng, Jie, et al. "Salient object detection by composition." Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.

