SALIENCY MAP BASED IMPROVED SEGMENTATION

Prerana Mukherjee, Brejesh Lall, Archit Shah

Indian Institute of Technology, Delhi, India.
mukherjee.prerana@gmail.com, brejesh.lall@gmail.com, archit145@gmail.com

ABSTRACT

In this paper we present a novel approach for refining segmentation using saliency map. To achieve this, we first develop a new saliency detection method based on cues at various levels. Initially preprocessing step is done using non-linear anisotropic diffusion filtering in order to preserve the edge information in the foreground salient objects and smoothen the background. Then we apply grab cut segmentation using saliency map as the input to get improved segmentation. Repeated application of the scheme is used for multi-object segmentation. The experimental results for the saliency technique show high precision and recall rates against the state-of-the-art methods.

***Index Terms*—** Saliency, segmentation, anisotropic non-linear diffusion, PQFT

1. INTRODUCTION

Although a lot of efficient segmentation approaches have been developed, it still remains as a challenging and fundamental problem in image analysis domain. This motivates our study on finding features which aid in obtaining neat segmentation. In order to achieve this, we exploit the salient features of the image to find the focus of attention. Based on the properties of the image, segmentation can be done based on the discontinuities or the similarities in the image [1, 2]. Region based segmentation [3, 4] involves the segregation of distinct regions based on some threshold. Clustering [5] helps in partitioning based on some similarity feature into some cluster of pixels. Structural approaches [6] use region information for segmentation. Most segmentation schemes are not accurate especially if performed autonomously without human input. To circumvent the issues in order to obtain good segmentation salient features can be potential candidate.

In the last few decades, attentional selection [7] has gained a lot of attention. These mechanisms involve cognitive processes for parsing and organizing the visual stimuli. Attentional selection can be essentially viewed as saliency detection which is based on certain properties or cues of the sensory information. These visual attention mechanisms can be driven by top down (memory-driven) or bottom up (memory-free) influences. Usually it is difficult for humans to identify multiple salient objects simultaneously so often a combination of both approaches is used depending on the application. The selection of the salient object is also very subjective. So it is very tricky and cumbersome for any machine to learn and implement the saliency technique effectively. Animals also use the visual perception to track their prey, detect their predators or mates in a cluttered visual environment. Thus saliency plays a very important role in humans as well as animals in analyzing stimuli they *see* (analyzing implies segmentation e.g. segmenting prey as object and rest of the scene as background).

There are various algorithms which explore various features to define the saliency measure. Each of the methods has its own pros and cons and a major shortcoming of these is high computational complexity. There are two kinds of models which primarily mimic the saliency mechanism: a) spatial contrast analysis [8, 9, 10] and b) frequency analysis [11, 12, 13]. Most of the saliency techniques are based on bottom up approaches because of less computational complexity. It is difficult to implement and model the top down mechanisms as they employ different states of the observer like one's memories and experiences. The basic aim in any attentional model is to be able to capture the viewpoint of any observer without bias towards the scene context. In this work we propose novel saliency detection approach using cues at various levels viz. local, global and rarity for refining segmentation results. We combine the information contained in *Saliency Maps* to improve segmentation results. Preprocessing step involves the use of non-linear anisotropic diffusion filtering to preserve the edge and smoothen the irrelevant details thus acting as a good precursor for segmentation. Our proposed approach employs both spatial and frequency information and is able to capture maximum details. We target multi-object segmentation using the iterative approach for segmentation.

In Section 2 we describe the proposed approach, Section 3 elaborates the results and discussions and Section 4 gives the conclusion.

2. PROPOSED APPROACH

In this section we present the proposed segmentation scheme. The novelty of the scheme lies in (a) an enhanced saliency map scheme and (b) use of saliency as an input to improve segmentation performance. The improvement in saliency map is made possible by fusion of various saliency cues at different levels. Further, to increase the discriminability of the proposed saliency map, we apply a non-linear scale space preprocessor unit to the image. Finally the saliency map is used as initialization for segmentation. This technique is applied successively to perform multi-object segmentation. We now describe the proposed saliency map generation technique.

2.1 Saliency Map Generation

2.1.1 Preprocessing

Initially we use non-linear anisotropic diffusion filtering to enhance the edge and line details in the foreground and suppress the non relevant details in the background [14]. An example of a filtered image is shown in Fig. 3(b).

2.1.2 Local Features Map

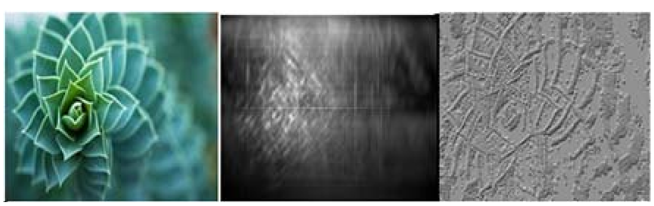
First we calculate the local features from the image. The local features are color, intensity, orientation, motion and depth. The color and intensity channels are calculated using centre surround scheme [8]. It helps in identifying the locations which locally stand out in the image. The difference between the fine centre and coarser surrounding scales is computed. The centre is calculated at scale $c\epsilon\{2,3,4\}$ and surrounding, $s = c + \delta$ where $\delta\epsilon\{3,4\}$. The scale difference between the two maps is computed. The intensity channel is computed as,

$$I(c,s) = I(c) - I(s) \tag{1}$$

The color channels are computed as,

$$R = r - \frac{(g+b)}{2}, G = g - \frac{(r+b)}{2}, B = b - \frac{(r+g)}{2} \tag{2}$$

$$Y = \frac{(r+g)}{2} - \frac{|r-g|}{2} - b \tag{3}$$



a) b) c)

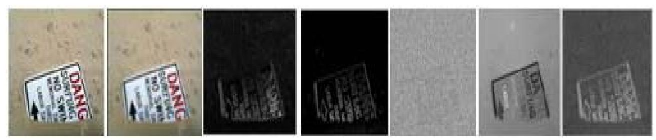
Fig. 1. a) Original image b) Normalized orientation map (using Gabor filter) c) Normalized orientation map (using HOG)

The RG and BY color channels are computed by taking the difference between center and surrounding at different scales. 6 intensity maps and 12 color maps are normalized into one color and one intensity map.

The normalized orientation map is obtained by Histogram of Oriented Gradients (HOG) [15]. HOG gives the object's appearance and shape based on the intensity gradients and edge directions. HOG provided better results than Gabor due to better edge locations as shown in Fig. 1. The optical flow is calculated using Proesman's method [16] which is a gradient based method as given in Fig. 2. The depth map is calculated using monocular cues like blur and defocus [17]. We compute the defocus map which gives a good notion of depth. Higher intensity regions correspond to higher amount of depth. The defocus blur amount is estimated at the edge locations by re-blurring the image and then taking the ratio of input to the blurred image. This blur amount obtained is used to get the defocus map. The local feature maps are shown in Fig. 3.

a) b) c)

Fig. 2. a) Frame 1 b) Frame 2 (with motion in horizontal direction) c) Motion map



a) b) c) d) e) f) g)

Fig. 3. a) Original image b) Non-linear anisotropic diffusion filtered image c) Normalized color map d) Normalized intensity map e) Normalized orientation map f) Depth map g) Local features map

2.1.3 Global Features Map

The original image is divided into a set of regions using mean shift segmentation. The global features like global contrast and spatial sparsity [18] are then computed using these primitive regions. We assume that salient objects have more contrasting features with respect to the background. The background colors are denser as compared to the salient colors as the salient object is located in a small region. The histogram is calculated using the quantized colors of the pixels in the region followed by normalization. The region similarity between each pair of regions is calculated as,

$$Sim(R_i, R_j) = Color\_Sim(R_i, R_j).Spatial\_Sim(R_i, R_j)$$

The color similarity was calculated using chi square distance measure between the histograms of the regions. Color similarity indicates the similarity of two regions with respect to their color histograms and spatial similarity gives the measure for the spatial distance between the regions given as,

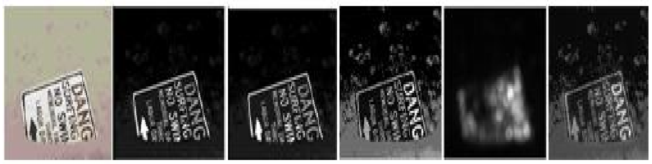
$$Spatial\_Sim(R_i, R_j) = 1 - \frac{eucl\_distance(\psi_i, \psi_j)}{d} \tag{4}$$

Here d is the diagonal length of the image and ψ is the spatial center of the region which is calculated using the spatial coordinates of the pixels in the region. Similarity between the regions is high when the regions are similar in color and the spatial distance between them is less. Global contrast for each region is computed in the following manner, $\sum_{j=1}^{n} |R_j|.Spatial\_Sim(R_i, R_j).eucl\_distance(m_i, m_j)$

where m_i, m_j are mean color of region R_i and R_j respectively. Spatial sparsity of the region can be computed as,

$$Spatial\_sparsity(R_i) = \frac{\sum_{j=1}^{n} Sim(R_i,R_j).eucl\_dist(\psi_j, c_0)}{\sum_{j=1}^{n} Sim(R_i,R_j)} \qquad (5)$$

ψ_j is the spatial center of region R_j and c_0 is the center position of the image. The global features are shown in Fig. 4.



a) b) c) d) e) f)

Fig. 4. a) Mean shift segmented image b) Global Contrast map c) Spatial Sparsity map d) Global features map e) PQFT rarity map f) Combined saliency map

2.1.4 Rarity Features Map

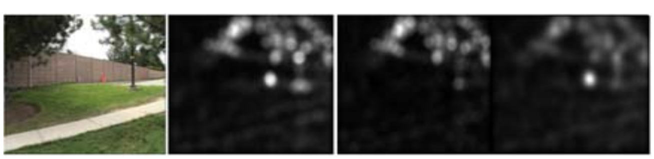
The rarity measure defines the peculiarity in any image. The Phase spectrum of Fourier Transform (PFT) of an image indicates the location of the sinusoidal components in the image and is a pivotal factor in determining saliency [19]. Quaternion representation of the image (intensity, color, motion features) gives multiresolution spatio-temporal representation (Phase Spectrum of Quaternion Fourier Transform PQFT). Phase spectrum gives local information about the image. Regions with less periodicity and homogeneity are the salient objects. It is also the fastest way of computation in Fourier domain. Quaternion image is represented as:

$$Q(t) = M(t) + RG(t)\alpha_1 + BY(t)\alpha_2 + I(t)\alpha_3 \qquad (6)$$

where t indicates the frame number. PQFT is the phase spectrum of the quaternion Fourier Transform. QFT is given as:

$$Q(t) = \|Q(t)\| e^{\alpha\varphi(t)} \qquad (7)$$

The properties of a quaternion image give better performance to PQFT over PFT. In the results as shown in Fig. 5 Pulse Cosine Transform (PCT) and PQFT are able to detect the red object and tree branches. PQFT gives more sharp results as compared to PCT. PFT missed the red object. Here, PQFT is chosen as the rarity feature.



a) b) c) d)

Fig. 5. a) Original image b) PQFT c) PFT d) PCT

a) b) c) d) e) f) g)

Fig. 6. a) Original image (ASD) b) Saliency map c) Binarised saliency map d) Initial segmented image e) Original image (SOD) f) Ground Truth g) Saliency map

2.2 Saliency Map based Segmentation

The proposed segmentation technique aims to exploit the information generated by the saliency map to perform an improved multi-object segmentation. It achieves this by applying the proposed (two-object segmentation algo.) successively to the intermediate segmented regions till all the objects are segmented.

The proposed scheme exploits the edge preservation property of non-linear scale space filtering. This technique is applied as a pre-processor to blur the object (remove intra-object detail) while retaining the boundary. This results in a saliency map that discriminates the foreground (or objects) and background better than if the preprocessor is not used.

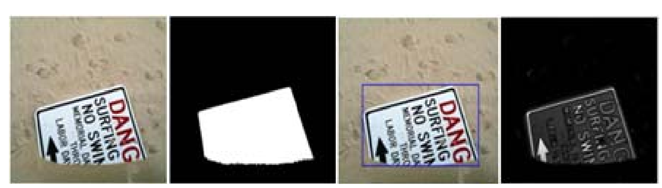
We now describe the segmentation scheme. The scheme consists of the following steps:

1. Preprocess image using non linear anisotropic scale space filtering.
2. Compute the saliency map of the preprocessed image using the modified saliency map.
3. Obtain global variance of the saliency map.
4. If var>Threshold Goto step 5 else ***Stop*** no further segmentation of the region is possible.
5. Binarize the saliency map and use the two regions as initial input for grab-cut segmentation.
6. Apply bounding box to the two regions and generate two images, each containing the pixels lying within the bounding box.
7. Repeat for the two images generated in Step 6.

The final saliency map of the pre-processed image is obtained as follows: After the computation of the 3 maps they are assigned weights $w_i = \frac{var_i}{\sum_{i=1}^{n} var_i}$ where i=1…n, n is number of maps and var indicates variance. The maps are namely local features map, global features map and rarity features map. They are linearly combined to give the saliency map as in Fig. 6. The block diagram of the proposed segmentation scheme is shown in Fig 9(a).

3. RESULTS AND DISCUSSIONS

ASD-1000 and SOD-300 database is used for experimentation. ASD dataset contains 1000 images from MSRA (Microsoft Research Asia) dataset. SOD dataset contains 300 images from BSD (Berkeley segmentation dataset).



a) b) c) d)

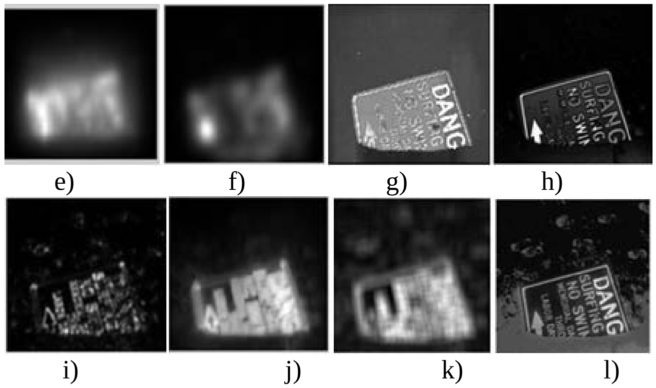


Fig. 7. a) Original image b) Binarised ground truth c) Ground truth label d) Maximum Symmetric Surround e) GBVS f) Itti-Koch g) Seg [20] h) Frequency Tuned [21] i) Hou j) Context Aware [22] k) SUN [23] l) Our Approach

3.1 Saliency Map Evaluation

The saliency map computed by various state of the art methods is given in Fig. 7. Comparative analysis of the performance measures is given in Fig. 8. Performance measures are given as:

$$Precision = |S_g \cap S_b| / |S_b| \quad (9)$$

$$Recall = |S_g \cap S_b| / |S_g| \quad (10)$$

$$F - measure = \frac{(1+\alpha).precision.recall}{\alpha.precision+recall} \quad (11)$$

Value of α is taken as 0.5. S_g is the ground truth and S_b is the binarised saliency map. The average of the precision and recall rates for the entire dataset is chosen for an adaptive threshold according to the image. The results show significantly high performance on the datasets using the proposed saliency method against the state-of-the-art methods. The time required for saliency map computation after initial filtering was greatly reduced from 171.23 sec to 83 sec. The precision rate for ASD database was improved to 90% (using filtering) from 87.37% (without filtering).

3.2 Segmentation Evaluation

After the computation of the saliency map, it is used as an input for the iterative grab-cut segmentation to get the initial segmentation which is used to again re-compute the saliency map for the rest of the background (other than the proto-object). The next saliency map is used to segment other objects subsequently. Once the segmented images are obtained, we perform k-means clustering to find the regions for different objects. The results and comparisons are given in Fig. 9. We have shown the comparative analysis by different machine learning based state-of-the-art techniques for segmentation. All these segmentation techniques require prior knowledge about the initialization of parameters to perform segmentation. In GMM-HMRF (Gaussian mixture models based hidden Markov Random Field), EM-GMMbic (Expectation-Maximization Gaussian Mixture Model + Bayesian information criteria) and HMRF-EM, MRF we require knowledge about initial cluster numbers, number of GMM components, EM algorithm and MAP (Maximum Aposteriori) iterations which is difficult to predict. Active contour model gave poor segmentation results. In the proposed approach we do not need to consider the parameter evaluation and it is an autonomous process.

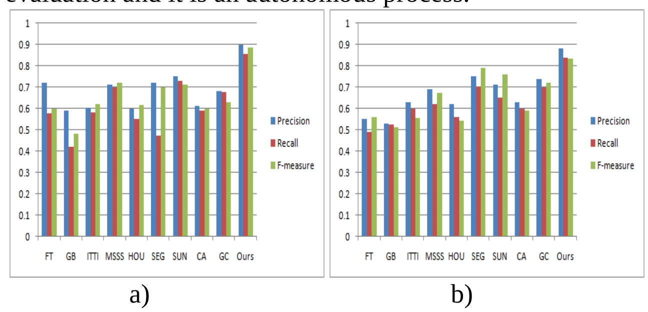


Fig. 8. Performance measures for a) ASD-1000 [21] database b) SOD-300 [24] database

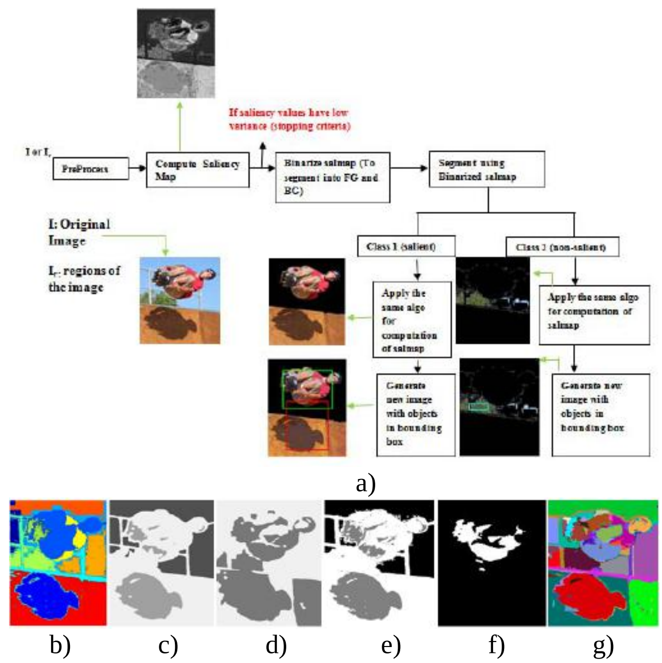


Fig. 9. Segmentation results using a) Proposed iterative grab-cut segmentation b) EM-GMMbic c) GMM-HMRF [25] d) HMRF-EM e) MRF f) Active contour models (snakes algorithm) g) Graph based segmentation [26]

4. CONCLUSION

To achieve better segmentation results we use non-linear anisotropic diffusion filtering which substantially reduces the time complexity for computation of saliency map and improves the performance rates. Secondly we target multi-object segmentation using different levels of cues of saliency.

The proposed saliency map technique gives relatively high results compared to the prior state-of-the-art methods. The segmentation technique is better in the sense that it does not involve manual intervention for providing the bounding box and foreground-background markers; instead it uses the image properties from the saliency map to identify the most conspicuous parts of the image.

5. REFERENCES

[1] J. Canny, "A computational approach to edge detection," *In IEEE Trans. On PAMI,* vol.8, no. 6, November 1986.

[2] W. Gao, L. Yang, X. Zhang and H. Liu, "An improved Sobel edge detection," *In ICICT,* vol. 5, pp. 67-71, July 2010.

[3] D. Comaniciu and P. Meer, "Mean shift analysis and applications," *In Proc.7th Intrnl. Conf. Computer Vision,* pp. 1197-1203, September 1999.

[4] A. Bleau and L. Leon, "Watershed based Segmentation and Region Merging," *In Computer Vision and Image Understanding,* vol. 77, pp. 317-370, March 2000.

[5] K. Chuang, H. Tzeng, S. Chen, J. Wu and T. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *In Computerized Medical Imaging and Graphics,* vol. 30, pp. 9-15, January 2006.

[6] O. Nempont, J. Atif and I. Bloch, "A constraint propagation approach to structural model based image segmentation and recognition," *In Information Sciences, Elsevier,* vol. 246, pp. 1-27, October 2013.

[7] J.R. Mounts, "Attentional selection: A salience-based competition for representation," *In Perception and Psychophysics*, vol. 67, pp. 1190-1198, October 2005.

[8] L. Itti, C. Koch and E. Niebur, "A Model of saliency based visual attention for rapid scene analysis," *In IEEE Trans. On Pattern Analysis and Machine Intelligence,* vol. 20, no.11, pp. 1254-1259, November 1998.

[9] R. Achanta, F. Estrada, P. Wils and S. Süsstrunk, "Salient Region Detection and Segmentation," *In International Conference on Computer Vision Systems (ICVS '08)*, Springer Lecture Notes in Computer Science, vol. 5008, pp. 66-75, 2008.

[10] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *In Advances in neural information processing systems*, vol. 19, pp. 545-552, 2007.

[11] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," *In IEEE Conference on Computer Vision and Pattern Recognition,* pp. 1-8, 2007.

[12] R. Achanta and S. Susstrunk, "Saliency detection using maximum symmetric surround," *In Proceedings of 17th IEEE International Conference on Image Processing (ICIP)*, 2010.

[13] C. Guo, "A novel multiresolution spatiotemporal saliency detection model and its application in image and video compression," *In IEEE Trans. On Image Processing,* vol. 19, no. 1, January 2010.

[14] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 7, pp. 629–639, 1990.

[15] N. Dalal and B. Triggs, "Histogram of oriented gradients for human detection," *In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* pp. 886-893, 2005.

[16] M. Proesmans, L. Gool, E. Pauwels and A. Oosterlinck, "Determination of optical flow and its discontinuities using non linear diffusion," *In Third European Conference on Computer Vision,* vol. 801, 1994.

[17] S. Zhuo and T. Sim, "Defocus map estimation from a single image," *In Pattern Recognition, Elsevier,* vol. 44, pp. 1852-1858, September 2011.

[18] Z. Liu, W. Zou and O. Meur, "Saliency tree: A novel saliency detection framework," *In IEEE Trans. On Image Processing,* vol. 23, no. 5, pp. 1937-1952, May 2014.

[19] C. Guo, Q. Ma, and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion Fourier Transform," *In Computer vision and Pattern Recognition, 2008.*

[20] E. Rahtu, J. Kannala, M. Salo and J. Heikkila, "Segmenting salient objects from images and videos," *In ECCV*, pp. 366-379, 2010.

[21] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," *In CVPR*, pp. 1597–1604, 2009.

[22] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *In IEEE Trans. On Pattern Analysis and Machine Intelligence,* vol. 34, no. 10, October 2012.

[23] L. Zhang, M. Tong, T. Marks, H. Shan and G. Cottrell, "SUN: A bayesian framework for saliency using natural statistics," *In Journal of Vision,* vol. 8, no.7, pp. 32:1-20, 2008.

[24] V. Movahedi, and J. H. Elder, ''Design and perceptual validation of performance measures for salient object segmentation,'' *In CVPRW,* pp. 49–56, June 2010.

[25] Q. Wang, ''GMM-based Hidden Markov Random Field for color image and 3D volume segmentation,'' *In arXiv preprint* arXiv:1212.4527, January 2012.

[26] P. Felzenszwalb and D. Huttenlocher, ''Efficient Graph- based image segmentation," *In IJCV,* September 2004.