Open access Journal International Journal of Emerging Trends in Science and Technology

A Saliency Detection Model Based on Wavelet Transform Through Fusion of Color Spaces

Authors

A.Srilakshmi[1], Mr. D.Prabhakar[2]

[1]M.Tech Student, Dept of DECS,DVR & Dr. HS MIC College of Technology, Kanchikacherla, Krishna District, AP, India

[2]Associate Professor, Dept of ECE, DVR & Dr. HS. MIC College of Technology, Kanchikacherla, Krishna District, AP, India

Abstract

Visual attention is studied by detecting a salient object in an input image. Visual attention is used in various image processing applications such as image segmentation, patch rarities, pattern recognition etc. In this paper, the saliency measurement is performed by using wavelet transform. In the proposed model we introduce a saliency measurement model based on two color spaces. One is the RGB and second is LAB. Both color spaces gives six different channels and those channels generates different feature maps using wavelet transform. Next, the measures of saliency (local and global) are calculated and fused to indicate saliency of each patch. Local saliency is distinctiveness of a patch from its surrounding patches. Global saliency is the inverse of a patch's probability of happening over the entire image. The final saliency map is built by normalizing and fusing local and global saliency maps of all channels from both color systems. Experimental evaluation gives the better results from the proposed model.

Key Words: *Saliency Map, Wavelet Transform, local saliency, global saliency*

1. Introduction

Visual attention is one of the most important features of the human visual system. Visual attention is a mechanism which filters out redundant visual information and detects the most relevant parts of our visual field [2]. Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought [3]. Nevertheless, our environment presents far more perceptual information than can be effectively processed. In order to keep the essential visual information, humans have developed a particular strategy, first outlined by James. This strategy, confirmed during the last two decades, involves two mechanisms. The first refers to the sensory attention driven by environmental events, Commonly called bottom-up or stimulus-driven. The second one is the volutational attention to both external and internal stimuli, commonly called top-down or goal-driven. Both of the computational models aim at generating saliency maps to detect the salient regions for images. We are to explore the bottom-up visual attention modeling in this work. In visual attention models, the Wavelet transform (WT) is more attracted and plays an important role [1]. The advantage of the proposed model is by applying the inverse wavelet transform on channels (R, G, B, L, A, B) in various decomposition levels. We create the more detailed feature maps from edge to texture and it is helps to observe the irregularities with different bandwidths. For two reasons we create these two saliency maps (local and global) to all channels: i) to avoid the normalization of each feature map separately which is not efficient for considering

the statistical relation among the feature maps for the saliency in a global perspective; ii) to incorporate local and global saliency as two different maps to make sure of taking both local and global contrast into consideration sufficiently. Finally, the local and global maps of both the color spaces of all the channels are combined to yield the final saliency map. Experimental results demonstrate that the proposed algorithm produces better performance with respect to the relevant existing algorithm.

2. The Proposed Saliency Measurement Model

2.1 Wavelet Analysis

By Alfred Haar, in the early 20th century the wavelets were firstly introduced, most of the developments in this area have been progressed since the late 20th century [5]. Recently, the use of wavelets in signal analysis has been rapidly increasing in many engineering applications such as signal detection, compression, enhancement, and pattern classification, etc.

In wavelet analysis, a signal is split into n^{th} levels then first itself split into a second-level approximation and detail, and the process is repeated. For an n-level decomposition, there are n+1 possible ways to decompose or encode the signal see in Fig. 1.The philosophy behind the saliency generation is to create features and feature maps which represent the contrast or center-surround difference, taking both local and global factors into account. The wavelet decomposition has the advantage in extracting oriented details (horizontal, vertical and diagonal) in the multi-scale perspective, and enables high spatial resolution with higher frequency component and low spatial resolution with lower frequency components, without information loss in details during the decomposition process [6].

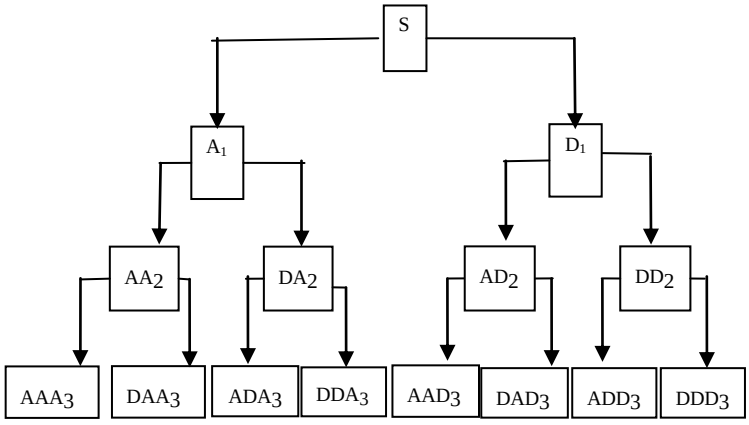

Fig-1: Illustration of frequency components for 3-level wavelet decomposition

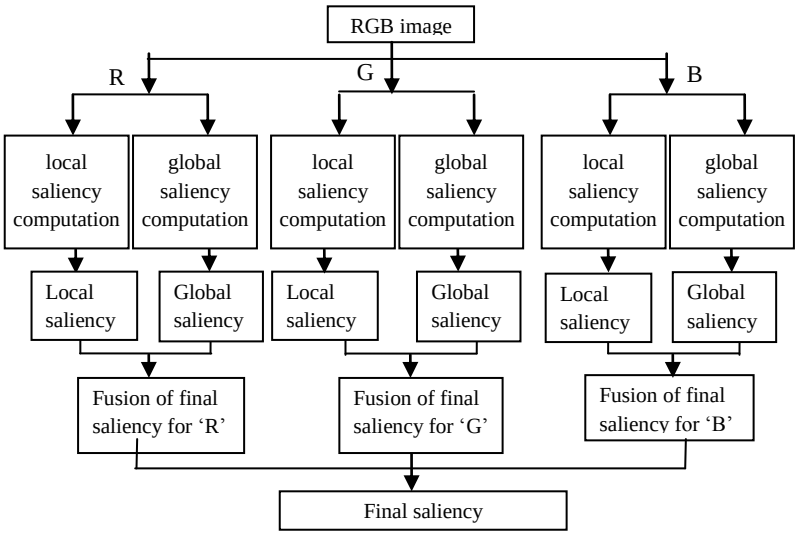

Fig-2: The framework of the RGB saliency in the proposed model

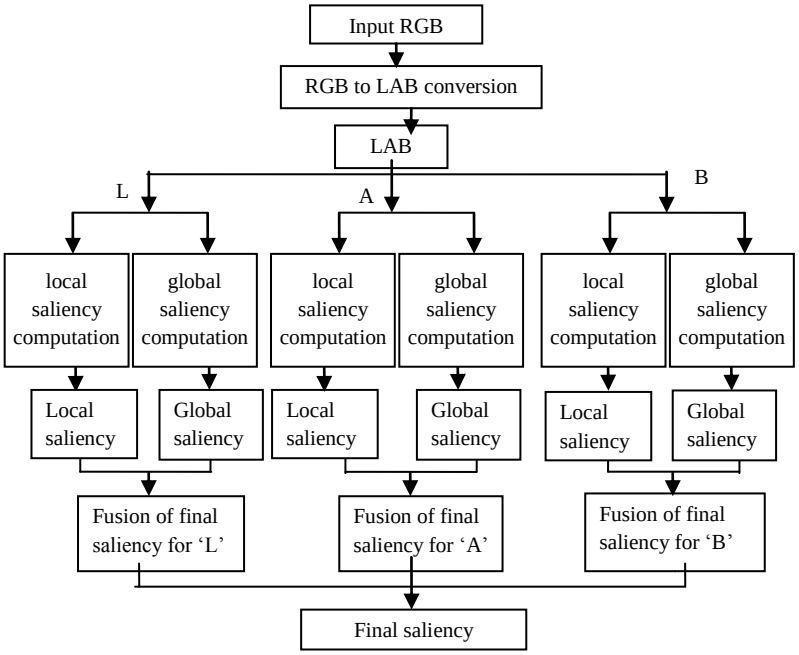

Fig-3: The framework of the LAB saliency in the proposed model

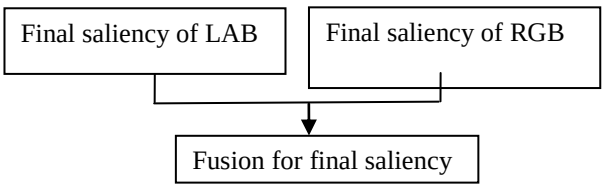

Fig-4: The framework of the final saliency in the proposed model

2.2 Overview of the Proposed Algorithm

The proposed framework is presented in Fig. 4. An input image in two formats (Lab and RGB) undergoes the same saliency detection and the resultant maps in each color system are normalized and summed. In each color format, two local and global saliency operations are applied to each color sub channel separately. While the first operation detects outliers in a local surrounding, the latter calculates the rarity of a feature or a region over the entire scene. Then, local and global rarities are combined to generate the output of each channel. Channel output maps are then normalized and summed once more to generate the saliency map. The whole process can be Performed over several scales.The proposed algorithm is to create the feature maps by increasing the bandwidths or the frequency components from higher to lower values. As shown in Fig. 4, it integrates both the color spaces of the channels of their local and global saliency maps .After combine the lab and the RGB saliencies. Both maps are attained by features of the same level, based upon the wavelet coefficients.

2.3 Feature Map Generation

The first step of the proposed algorithm is first converts the RGB color space to LAB color space for saliency detection purpose, an image is converted to the *CIE Lab* color space (CIE illumination *D65* model is selected for conversion as the *white-point* parameter in *Matlab® rgb* to CIE *Lab* converter). Almost all saliency approaches utilize a color channel. Some have used RGB while others have employed Lab inspired by the finding that it better approximates human color perception. In particular, Lab aspires to perceptual uniformity, and its L component closely matches human perception of lightness, while the a and b channels approximate the human chromatic opponent system [9] RGB, on the other hand, is often the default choice for scene representation and storage due to this reason the conversion is needed.. To remove noise, we apply an nxn 2D Gaussian low-pass filter to the input color image $G^{C:}$ [1]

$$G^{IC} = G^{C} * I_{n\times n} \quad (1)$$

Where I is the nxn 2-D filter; G^{IC} is noise-removed version of G^{C}, $*$ denotes the convolution operation. For noise reduction, a small filter size is selected ($n=3$ in this work) to filter very High frequency noise. Then, each channel is normalized to the range of [0,255].The sub bands will be formed by wavelet transform based on the number of levels, the respective sub bands are shown in (1). Here we choose the daubechies wavelets (Daub.5), because its filter size is appropriate for pixel neighborhoods. so by this the computation time for generating Output will be less and we get the result accurately.

$$\left[A_{N,}^{C} H_{S,}^{C} V_{S}^{C}, D_{S}^{C}\right] = WT_{N}(G^{IC}) \quad (2)$$

Where N is the maximum number of the scaling for wavelet transform decomposition process, i.e., the resolution index S $\in$ { 1,2,…….N } and the N^{th} level corresponds to the coarsest resolution; and S is the channels of G^{IC} as C $\in$ {L,A,B} if it is lab color space, otherwise it may be C $\in$ {R,G,B} if it is rgb color space.A_N^C (to represent scaling coefficients) is the approximation output at the coarsest resolution for each channel; H_S^C,V_S^C and D_S^C are the wavelet coefficients of horizontal, vertical and diagonal details for the given C and S, respectively. The wavelet coefficients representing the details of the image at various scales are used to create several feature maps with increasing frequency bandwidths. The feature maps can be calculated by IWT. Since we already apply the Gaussian filter, we can create feature maps from the details of WT by neglecting approximation data during the IWT process. Hence, several feature maps can be obtained while representing the contrast from edge to texture. The approximation data of the selected decomposition level S is not used during IWT operation as in (3)

below, to detect the global saliency:

$$f_S^C(x,y) = \frac{\left(\mathrm{IWT_S}\left(\mathrm{H_S^C}, \mathrm{V_S^C}, \mathrm{D_S^C}\right)\right)^2}{\eta} \quad (3)$$

Where $f_s^c(x,y)$ is the feature map generated for the s^{th} level for each images (both the LAB and RGB) sub-band c, η is the scaling factor (since the range of the images for each channel is [0,255], there is a large range for feature values in (3); therefore, an appropriate value of η is the scaling factor to limit the feature maps, and $\eta = 10^4$ in (3) where this scaling is necessary to avoid the huge variations in the covariance matrix among the feature maps during the computation of global Saliency map in (4)).,IWTs(.) is the reconstruction function referring to the IWT of H_s^c, V_s^c and D_s^c by neglecting the A_N^c. Thus for $s \in \{1,\ldots,N\}$ and $c \in \{L,A,B,R,G,B\}$, equation (3) creates 6xN feature maps for an input color images i.e., for rgb image and for lab image and each feature map's resolution is equal to the size of the input image.

2.4 Global Distribution of Features

After obtaining the feature maps, the next step is to calculate the global distribution of the local features to obtain the global saliency map for all channels. From $f_s^c(x,y)$ in (3), a location (x,y) can be represented as a feature vector f(x,y) with a length of 6×N (6 channels L, a, b, R,g,b and N-level wavelet-based features for each channel) from all feature maps. Regarding the feature maps, the likelihood of the features at a given location can be defined by the probability density function (PDF) with a normal distribution [9]. Therefore, the Gaussian PDF in multi-dimensional space can be written as [7], [8]:

$$\mathrm{P}\left(\mathrm{f(x,y)}\right) = \frac{1}{(2\pi)^{l/2}|\mathrm{S}|^{1/2}} \times \mathrm{e}^{\left(-1/2(\mathrm{f(x,y)}-\mathrm{m})^{\mathrm{T}}\mathrm{S}^{-1}(\mathrm{f(x,y)}-\mathrm{m})\right)} \quad (4)$$

With

$$S = E[(f(x,y)-m)(f(x,y)-m)^T] \quad (5)$$

Where m is the mean vector containing the mean of each feature map, i.e., m =E[f]; T is transpose operation; S in (5) is the nxn covariance matrix; n=6xN, the number of the feature vector referring to the dimension of the feature space including 6 color channels and N feature maps for each color channel, and |S| is the determinant of the covariance matrix [8]. Using the PDF in (4), the global saliency map can be computed as (6) below. By using the (6) we calculate the global saliency for all channels without combining. As can be seen in (6), the result is filtered with a KxK 2-D Gaussian low-pass filter to obtain a smooth map where K=5, which is a commonly used filter size in many saliency applications.

$$S_G(x,y) = \left(log(P(f(x,y))^{-1})\right)^{1/2} * I_{K\times K} \quad (6)$$

Where S_G includes the information of both locally and globally for all channels. Since it is computed from the local features in (3). However, it can be seen as a global saliency map because its effect on global distribution on the saliency map is much higher, and may become dominant (i.e., overestimated) due to the content or structure of the scene. Also, it includes the statistical relation among the feature maps so it may give some important information which cannot be detected well enough by the local contrast. Moreover, the result from (6) may generate a saliency map with small salient regions, and thus causes some loss in local saliency information. This is the result when the distribution of the local features for the salient regions is balanced with the local contrast for the given features. However, there are also some cases where global saliency can suppress the local contrast too much. On the other hand, it may yield important salient regions which are less salient locally or the locally salient regions may not be as attractive as globally salient regions. However, the global distribution of the local features gives more attention. The different saliency maps can be beneficial to adjust the saliency for local features or to alleviate overestimation for global information for the better saliency map.

2.5 Linear Combination of the Local Features

In this work, local saliency is created by fusing the feature maps at each level linearly without any normalization operation in [9], as the formula to be given in (7) below. Hence, this new map will be computed based on the local features computed in (3) in which the value of the image with their respective channel is taken into account at each level. The feature maps obtained in (3) are used for calculating the local saliency map as:

$$S_L(x,y) = \left(\sum_{S=1}^{N} arg\left(f_s^c(x,y)\right)\right) * I_{K\times K} \qquad (7)$$

By using the above equation (7) we calculate the local saliency for each channel individually. The $f_S^C(x,y)$ are the feature maps at scale s for L, a and b and R,g,b channels(in $f_S^C(x,y)$,C represents the respective channel) respectively; $S_L(x,y)$ is the local saliency map.

2.6 Combination of the Global and Local Saliency Maps

Based on (6) and (7), we can create the global and local saliency maps. The final saliency is the result of combining these two maps. The integration is performed to modulate the local saliency map with its corresponding global saliency map de-fined as:

$$S'(x,y) = M\left(S_L'(x,y) \times e^{S_G'(x,y)}\right) * I_{K\times K} \qquad (8)$$

where S'(x,y) is the final saliency map for one channel. By using the above formula we calculate the final saliency for all channels which we have like l,a,b and r,g,b,S_L'(x,y) and S_G'(x,y) are the local and global saliency maps for single channel and are linearly scaled to the range [0,1] . Since the modulation is applied by the multiplication of local saliency and the exponential value of the global saliency, M (.) as $M(.)=(.)^{\ln\sqrt{2}/\sqrt{2}}$ is used as the non-linear normalization function to diminish the effect of amplification on the map. The possible values of the output in (8) will be between 0 and 1 due to the parameter selection. In (8), we obtain a saliency map which considers local features at a location with its respective global feature distribution. Therefore, the global relation between local feature maps is established without the need of any complex feature map normalization process for enhancing each feature map as in [4]. In addition, we enhance the final saliency map with a similar fashion in [8]. As stated by Goferman *et al.* [8], based on Gestalt laws, saliency values to describe the regions of interest can be re-evaluated around the regions that are the most salient points of the scene. The idea is: the locations around the focus of attention (FOA) have to be more attentive than those away from the FOA. Therefore, saliency values around the most salient points are increased to enhance the performance of the saliency map as:

$$S(x,y) = S'(x',y')\left(1 - d_{CFOA}(x,y)\right) \qquad (9)$$

where s(x,y) is the saliency value at point (x,y),s'(x',y') is the salient value of the most salient points at the location (x',y') extracted from the saliency map in (8) with a threshold of 0.8 ,$d_{cFoA}(x,y)$ is the distance between the location(x,y) and its closest FOA at the location(x',y'). Obviously the saliency values around the salient regions will increase in the final saliency map; on the other hand, the saliency values of the points that are distant to the attention regions will decrease or re-main largely unchanged.

3. Experimental Results

In this work, the Microsoft public database [9] including 5000 color images is used to evaluate the performance of the proposed model quantitatively. Besides the color images, there are also ground-truths for images in the database: the human-labeled attention regions highlighted with a bounding box created by 9 subjects [9]. These bounding boxes represent the attention regions, i.e., the object/region of interest in scenes perceived by the subjects. As a result of averaging the human responses, the performance evaluation of a saliency detection model can be achieved

quantitatively by checking the consistency between the human-labeled ground-truth and the saliency map from any computational model.

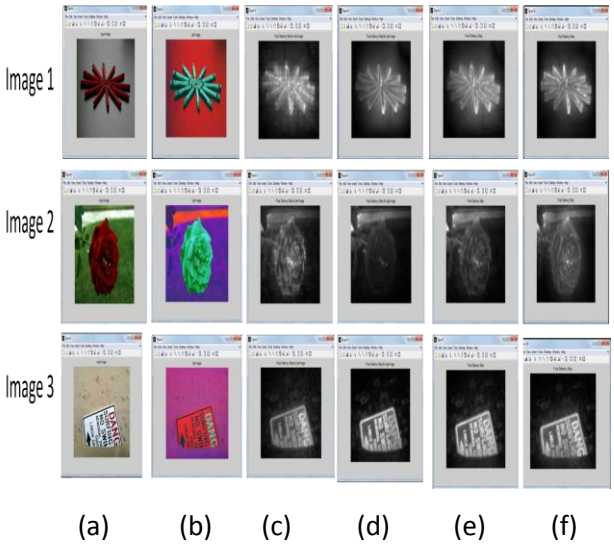

Fig.5. The performance evaluation with different images for the existing and for the proposed models. (a) RGB image (b) LAB image (c) final saliency map for LAB image (d)final saliency map for RGB image (e) final saliency map for proposed method (f)final saliency map for existing method.

Table-1:

Comparisons of Existing and Proposed Method - Precision, Recall and F– measure values.

Images	**Existing Method Values**			**Proposed Method Values**		
	Precision	**Recall**	**F-Measure**	**Precision**	**Recall**	**F-Measure**
Image1	0.498	0.325	0.443	0.498	0.651	0.5266
Image2	0.498	0.207	0.376	0.498	0.607	0.5197
Image3	0.498	0.266	0.415	0.498	0.611	0.5204

The quantitative performance for the database is evaluated based on overall *precision P, recall R,* and *F-Measure F*. $F\beta^2$, as defined below respectively [9]:

$$P = \frac{\Sigma_X \Sigma_Y (g(x,y) * S(x,y))}{\Sigma_X \Sigma_Y s(x,y)} \quad (10)$$

$$R = \frac{\Sigma_X \Sigma_y (g(x,y) * S(x,y))}{\Sigma_X \Sigma_Y g(x,y)} \quad (11)$$

$$F_{\beta^2} = \frac{(1+\beta^2) \times P \times R}{\beta^2 \times P + R} \quad (12)$$

where g(x,y) is the ground truth map, s(x,y) is the saliency map from the computational model, and β^2 in (12) is a positive parameter to decide the relative importance of the precision over the recall in evaluating the precision (a greater value for β^2 indicates the higher importance of recall over precision: β^2 is chosen as 0.3 in this work).

For the experimental results, P is related to the saliency detection performance of the computational model; R is the ratio of salient regions from correct detection and ground truth; F-measure is a performance measure as being the harmonic mean of P and R. In the evaluation the generated saliency map is converted to binary image with an appropriate threshold for performance comparison, and g(x,y) and s(x,y) are the binary maps to calculate P,R, and F-measure values. For this analysis, Otsu's automatic threshold algorithm [10] is selected for the binary map generation since it makes the test less independent than the user defined threshold values. We argue that employing just one Color system does not always lead to successful outlier detection.

In Fig.5, we show that interesting objects in some images are more salient in Lab color space, while, for some others, saliency detection works better in RGB. Hence, in the proposed strategy, the next contribution is combining saliency maps from both color spaces Thus, the proposed model can be used for images with different object sizes and extents of uniformity. It also generates the saliency map with the same resolution of the input

image. In figure (5)
we observe the final saliency map is better in our model compared to previous model. This will be observed based on the performance parameters like precision, recall, and the f-measure values regarding to the Otsu's threshold method and comparisons are observed in Table -1 and in Fig. 6.

Performance with Threshold of Otsu's Method

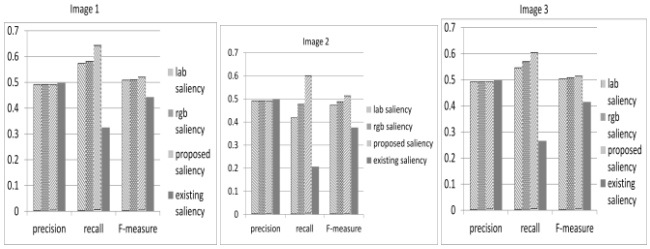

Fig- 6: Performance Comparison between the proposed model and other existing ones for images (1-3)

4. Conclusion

In this paper, a novel bottom-up computational model of visual attention is proposed to obtain the saliency map for images based on wavelet coefficients. Various feature maps are generated by IWT in various scales. The feature map includes components from the edge to the texture. We enhance the state-of-the-art in saliency modelling by proposing an accurate and easy to implement model that utilizes image representations in both RGB and LAB color spaces. Furthermore, we introduce one local and one global saliency operator each representing a class of previous models to some extent. We conclude that employing just one color system channels does not always lead to successful outlier detection.

Integration of both the local and global saliency operators of both the color systems channels works better than just using either one. Similarly, combining both color systems strongly benefits saliency detection and eye fixation prediction.

References

1. Nevrez İmamoğlu, Weisi Lin,Yuming Fang, "A Saliency Detection Model Using Low-Level Features Based on Wavelet Transform", IEEE transactions on multimedia, Vol. 15, No. 1, January 2013, pp-96-105.
2. A. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognit. Psychol.*, vol. 12, no. 1, pp. 97–136, 1980.
3. C. Koch and S. Ullman, "Shifts in selective visual attention: To-wards the underlying neural circuitry," *Human Neurobiol.*, vol. 4, pp. 219–227, 1985.
4. L. Itti, "Models of bottom-up and top-down visual attention," Ph.D. dissertation, Dept. Computat. Neur. Syst., California Inst. Technol, Pasadena, 2000.
5. R. J. E. Merry, *Wavelet Theory and Application: A Literature Study*, *DCT 2005.53*. Eindhoven, The Netherlands: Eindhoven Univ. Technol., 2005.
6. L. Itti, C. Koch, and E. Niebur, "Model of saliency-based visual atten-tion for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
7. S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. London, U.K.: Academic/Elsevier, 2009, pp. 20–24.
8. A. Oliva, A. Torralba, M. S. Castelhano, and J. M. Henderson, "Top-down control of visual attention in object detection," in *Proc. IEEE Int. Conf. Image Processing*, 2003, vol. 1, pp. 253–256.
9. T. Liu, J. Sun, N.-N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," in *Proc. IEEE Int. Conf. Comput. Vision and Pattern Recognition*, 2007, pp. 1–8.
10. R.C. Gonzalez, R. E. Woods, and S. L.Eddins,*Digital Signal Procesing Using Matlab®*. Englewood Cliffs, NJ: Prentice Hall, 2004.

Biographies

A.Srilakshmi received B.Tech degree in Electronics And Communication Engineering from Nova College of Engineering and Technology, Jangareddygudem, A.P, India in 2011.She is currently pursuing M.tech in Digital Electronics and Communication Systems (DECS) in Devineni Venkata Ramana & Dr.HimaSekhar MIC College of Technology, Kanchikacherla, A.P, India.

Mr. D. Prabhakar, working as Associate professor in DVR & Dr HS MIC College of Technology, Kanchikacherla, Krishna (Dt).He has 10years of teaching experience He has completed B.Tech (ECE) from Acharya nagarjuna University, Guntur (AP), India in 2001 and M.Tech from Andhra University ,Visakhapatnam (AP), India in 2003. Since 2003 he is working as faculty member in different Engineering colleges in AP in different capacities. He is doing research work on Antennas at ECE Department, AU College of Engg, Andhra University, and Visakhapatnam. His fields of interest are Antennas, Electromagnetics, Microwaves, Radar, and EMI/EMC.