Global Journal of Researches in Engineering: J
General Engineering
Volume 14 Issue 6 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4596 & Print ISSN: 0975-5861

A Novel Approach for Saliency Detection by using Stationary Wavelet Transform Low Level Features

By Mulagundla Mahalaxmi, Mr K. Durga Prasad, Dr. K. Manjunathachari & Dr. M.N. Giri Prasad
Vardhaman College of Engineering, India

Abstract- The ability of the Human Visual System (HVS) to detect an object in an image is extremely fast and reliable but how can a machine vision system detects the salient regions? many algorithms have been proposed to solve this problem by extracting features in either spatial or spectral domain, in this paper, A novel saliency detection model is introduced by utilizing low level features obtained from Stationary Wavelet Transform domain. Here Stationary Wavelet Transform (SWT) is preferred as the wavelet transform than Discrete Wavelet Transform (DWT), Since DWT is not a time-invariant transform. So to make it translation invariant SWT is introduced. And also unlike the other wavelet transforms SWT does not require down sampling, So image size is same as original even after decomposition, thus there is no information loss in respective sub bands. Experimental results demonstrate that proposed model produces better performance by using SWT than by using DWT with the overall F-Measure value being high.

Keywords: human visual system (HVS), saliency detection, stationary wavelet transform, feature map, saliency map.

GJRE-J Classification : FOR Code: 091599

ANOVELAPPROACHFORSALIENCYDETECTIONBYUSINGSTATIONARYWAVELETTRANSFORMLOWLEVELFEATURES

A Novel Approach for Saliency Detection by using Stationary Wavelet Transform Low Level Features

Mulagundla Mahalaxmi [α], Mr K. Durga Prasad [σ], Dr. K. Manjunathachari [ρ] & Dr. M.N. Giri Prasad [ω]

Abstract- The ability of the Human Visual System (HVS) to detect an object in an image is extremely fast and reliable but how can a machine vision system detects the salient regions? many algorithms have been proposed to solve this problem by extracting features in either spatial or spectral domain, in this paper, A novel saliency detection model is introduced by utilizing low level features obtained from Stationary Wavelet Transform domain. Here Stationary Wavelet Transform (SWT) is preferred as the wavelet transform than Discrete Wavelet Transform (DWT), Since DWT is not a time-invariant transform. So to make it translation invariant SWT is introduced. And also unlike the other wavelet transforms SWT does not require down sampling, So image size is same as original even after decomposition, thus there is no information loss in respective sub bands. Experimental results demonstrate that proposed model produces better performance by using SWT than by using DWT with the overall F-Measure value being high.

Keywords: human visual system (HVS), saliency detection, stationary wavelet transform, feature map, saliency map.

I. Introduction

The first step in recognition of an object is object detection. Object detection is nothing but extracting an object from its background. Human visual system (HVS) can easily identifies the important and compact information from the natural scenes [1]. But for the machine vision systems it is a challenging task. Many traditional models have been introduced to detect the salient regions by utilizing the low level features such as intensity, color, contrast.

There are two types of visual attention mechanism: top-down and bottom-up approaches. The top-down approach is goal-driven, consists of high level data processing and requires prior knowledge to support the tasks such as target detection, object recognition etc [2],[3]. Bottom-up approach is task independent and obtained from early features [4],[5]. Both the computational models are used to generate the salient regions for the images. In this paper bottom-up visual attention mechanism is used.

Most visual attention models computes the saliency maps by extracting low level features from the image. These models consists of the following three steps. First step is feature extraction in which multiple low level features such as such as color, intensity, orientation, texture and motion are extracted from the image at various scales[6],[7].

The second step is saliency computation, it is computed by self information [6] and center-surround operation [7] and last step is few key locations on saliency map are identified by applying non linear operations. Recent studies have tried to obtain saliency map for images in different domains namely, Spatial, Fourier Transform, Wavelet Transform domains. One of the earliest computational models of visual attention is proposed by Itti et.al [3], [7]. He used low level features to calculate saliency map, however local information loss is unavoidable in this algorithm since the saliency map is calculated in coarser scales. problems. By using Fourier Transform, a signal in frequency domain can be decomposed in to amplitude spectrum and phase spectrum. Oliva et al [8], [9]. proposed an algorithm by using Fourier Transform (FT), amplitude spectrum gives the shape and position of the object and the phase spectrum gives global information of the image that contributes to overall scene. Hou et al. [10] extracted the spectral residual of an image by analyzing the log spectrum of an image in spectral domain and saliency map is computed by transforming the spectral domain to spatial domain by applying inverse Fourier Transform. However in this model global irregularities are more dominant than local irregularities and another disadvantage is it requires high down-sampling rate. And main disadvantage of Fourier Transform is, it cannot be applied to non- stationary signals and gives better results with only periodic or stationary signals.

Recently Wavelet Transform [11], [12] has begun to draw the much attention in visual attention modeling because of its advantage of being applicable to non-stationary signals. Wavelet Transforms are used to provide time-frequency representation [12]. It is capable of providing the time and frequency information simultaneously. In this paper Stationary Wavelet Transform (SWT) is used to decompose the image into sub-bands. Feature maps are created by applying Inverse SWT on the multi-level decomposition.

Rest of this paper is organized as follows: section II gives the brief view of existing mode land wavelet decomposition, section III presents the

Author α σ: Vardhaman College of Engineering, Hyderabad, India. e-mails: mahalaxmim.reddy@gmail.com, kalasa.durgaprasad@gmail.com

Global Journal of Researches in Engineering (J) Volume XIV Issue VI Version I Year 2014

proposed model in detail, experimental results are given in section IV and conclusions are given in final section.

II. Conventional Model

Nevrez Imamoglu et al. [13] proposed an algorithm by utilizing low-level features obtained from the discrete wavelet Transform domain. DWT decomposes the input image and generates four sub-bands of the image [12],[13] namely approximation, horizontal, vertical, diagonal representing details of the image are shown in fig.1. Then feature maps are generated by applying IDWT and then global and local saliency maps are computed separately from these features to form a final saliency map.

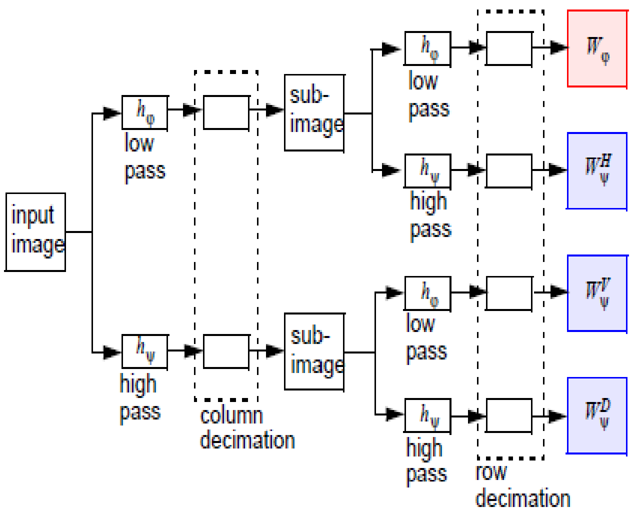

Fig. 1 : Two dimensional DWT decomposition scheme

This model is aimed to modulate local contrast at a location with its global saliency computed based on likelihood of the features and also considered local center-surround differences and global contrast in the final saliency map. But Discrete wavelet Transform (DWT)'s drawback is that it is not translation invariant and includes down sampling steps so image size will be reduced in the decomposition levels, in order to overcome this and to get more complete characteristic of the input image the un decimated WT i.e Stationary Wavelet Transform (SWT) is proposed.

III. Proposed Saliency Detection Model

a) Stationary wavelet Transform (SWT)

It is also called Un-decimated Wavelet Transform or the Invariant Wavelet Transform or the redundant Wavelet Transform [14]. The key point is that it gives better approximation than the DWT since it is redundant, linear and shift invariant. SWT is very useful algorithm for analyzing a linear system. Stationary Wavelet Transform is preferred as the Wavelet Transform since unlike the other Wavelet Transforms, the SWT procedure does not include any down sampling steps as it can be seen in fig.2. Normally Discrete Wavelet Transform (DWT) is used to decompose an input image into different sub band images i.e approximation, horizontal, vertical and diagonal. Three high frequency sub bands (LH,HL,HH) contain the high frequency components of an input image and LL band, which is a low frequency band gives the approximation of the input image. Down sampling in each of the DWT sub bands causes information loss in respective sub bands that is why SWT is used to minimize this loss. As no decimation steps are involved in SWT, it produces more precise information for the frequency localization [15],[16],[17].

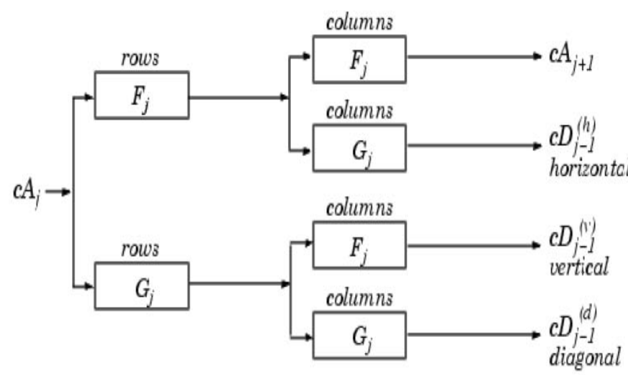

Fig. 2 : Stationary Wavelet Transform decomposition scheme

b) Over view of the proposed model

Proposed model is same as saliency detection model using wavelet transform proposed by Nevrez Imamoglu et al [13]. except one difference here SWT is used instead of DWT because in SWT no decimation steps are there, therefore sub bands of image size is same as original image even after decomposition for number of levels thus there is no loss of Information in the sub bands, Thus SWT gives better saliency detection than DWT. The framework of the proposed model is shown in fig. 3, first of all, rgb image is converted into CIE Lab color space since lab color space is similar to human perception ,with a luminance and two chromatic channels (RG and BY) and another advantage is it is device independent . To remove noise an mxm 2D Gaussian low-pass filter is applied to the input color image g^c.

$$g^{IC} = g^c * I_{m\times m} \tag{1}$$

Where I is the mxm 2-D filter, g^{IC} is noise removed version of g^c, here a small filter size m=3 is selected for noise reduction.

SWT is based on the idea of no decimation. It does not include down sampling in the forward and up sampling in the inverse transform. The sub-bands of the image is formed by applying SWT for number of levels representing approximation, horizontal, vertical and diagonal coefficients of the image. H, V, D coefficients are saved and approximation is used for next level, thus the decomposition is done until the coarsest resolution is possible. Here as no decimation and interpolation process are involved the size of the sub-bands of the image do not diminish from level to level, thus there will be no information loss, thus the number of pixels involved in computing a given coefficients grows slower and so the relation between the frequency and spatial information is more precise.

$$[A_N^c, H_s^c, V_s^c, D_s^c] = SWT_N(g^{IC}) \tag{1}$$

Where N is the maximum number of scaling for decomposition process. i.e resolution.

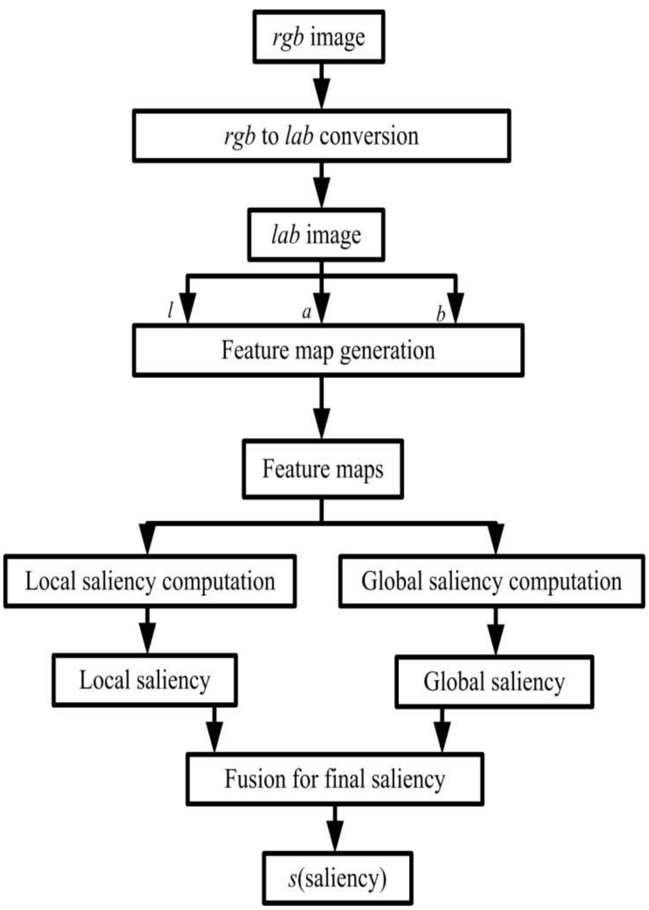

Fig. 3 : The frame work of proposed saliency detection model

index s $\epsilon\{1,....N\}$. C is the channels of g^{IC} as C $\epsilon\{L,a,b\}$. A_N^c is the approximation output at the coarsest resolution for each channel, H_s^c, V_s^c, D_s^c are wavelet coefficients of Horizontal, Vertical and diagonal details of the given image.

Several feature maps are created by applying inverse SWT to the horizontal, vertical and diagonal details of the image by neglecting approximation during IWT process

$$f_s^c(x,y) = \frac{(ISWT_s(H_s^c, V_s^c, D_s^c))^2}{\eta} \tag{3}$$

here $f_s^c(x,y)$ is the feature map generated for the s^{th} level decomposition for each image sub-band c, η is the scaling factor to limit the feature maps. Since the range of Lab input image for each channel is [0, 255], there will be higher range for feature values, here $\eta = 10^4$ is chosen. Thus eqn (3) creates 3xN feature maps for an input color image and each feature maps resolution is equal to the size of the input image. Once the feature maps are generated, the next step is to calculate the global distribution of the local features to obtain the global saliency map. From $f_s^c(x,y)$ in (3), a location (x,y) is selected which represents the feature vector f(x,y) with a length of 3xN from all feature maps. The likelihood of the features of a given location can be defined by applying probability density function (PDF) [8], [18]. By using PDF global saliency map can be computed as shown below.

$$s_G(x,y) = (\log(p(f(x,y))^{-1}))^{1/2} * I_{k\times k} \tag{4}$$

where kxk 2-D Gaussian low pass filter is used to obtain a smooth map, the global saliency map is generated with small salient regions and causes some loss in local saliency information thus it can be seen that global distribution on the saliency map is much higher due to the content or structure of the scene. After obtaining global saliency map, the next step is to compute local saliency, local saliency is created by adding the feature maps at each level linearly without any normalization operation in [7], as the formula to be given in (5) below. This new map is computed by taking the maximum value between channels of the input image at each level

$$s_L(x,y) = \left(\sum_{s=1}^{N} \arg\max\left(f_s^L(x,y), f_s^a(x,y), f_s^b(x,y)\right)\right) * I_{k\times k} \tag{5}$$

where $f_s^L(x,y), f_s^a(x,y), f_s^b(x,y)$ are the feature maps for L, a and b channels respectively. The final saliency map is computed by combining global and local saliency maps

$$s'(x,y) = M(s'_L(x,y) \times e^{S'_G(x,y)}) * I_{k\times k} \tag{6}$$

where $s'(x,y)$ is the final saliency map, $s'_L(x,y)$ and S'_G(x,y) are the local and global saliency maps scaled to the range [0,1]. Final saliency map is enhanced with a similar fashion in [25]. As stated by Goferman et al. [18] the locations around the focus of attention (FoA) have to be more attentive than those away from the (FoA). Therefore saliency values are increased to enhance the performance of the saliency map.

Where s'(x',y') is the salient value of the most salient points at the location (x',y') extracted from the

Global Journal of Researches in Engineering (J) Volume XIV Issue VI Version I Year 2014

saliency map in (6), s(x,y) is the saliency value at point (x,y) and $d_{cFoA}(x, y)$ is the distance between location (x,y) and its closest FoA at the location (x',y'). The proposed final saliency map is better than the global an local saliency maps.

IV. Experimental Results

For experimental results, the quantitative performance of the proposed model and conventional model is evaluated based on overall precision P, recall R, and F-Measure Fα [19] as defined below.

$$precision(p) = \frac{\sum_x \sum_y (t(x,y) \times s(x,y))}{\sum_x \sum_y s(x,y)}$$

$$recall(R) = \frac{\sum_x \sum_y (t(x,y) \times s(x,y))}{\sum_x \sum_y t(x,y)}$$

$$F - measure(F_\alpha) = \frac{(1+\alpha) \times P \times R}{\alpha \times P + R}$$

where s(x,y) is the saliency map from the computational model , t(x,y) is the ground truth map and α is a positive parameter to decide the importance of the precision over the recall. Precision is related to the saliency detection performance of the proposed model. Recall is the ratio of correct detection of salient regions to the ground truth map. F-measure is the harmonic mean of precision and recall [19]. Precision(P), recall(R), F-Measure (F_α) results for the proposed and conventional model are shown in table 1.

Table 1 : Performance comparison of conventional and proposed model

Parameters	Conventional model	Proposed model
Precision	0.0085	0.0943
Recall	0.3203	0.1140
F-measure	0.0166	0.1032

It can be seen that precision and F-Measure values of the computational model using DWT [13] are very low due to high recall value which causes more irrelevant salient regions occur and the proposed model has the better Precision and F-Measure values compared to conventional model. Therefore the overall performance of the proposed model is reliable and yields better results with respect to the relevant conventional model.

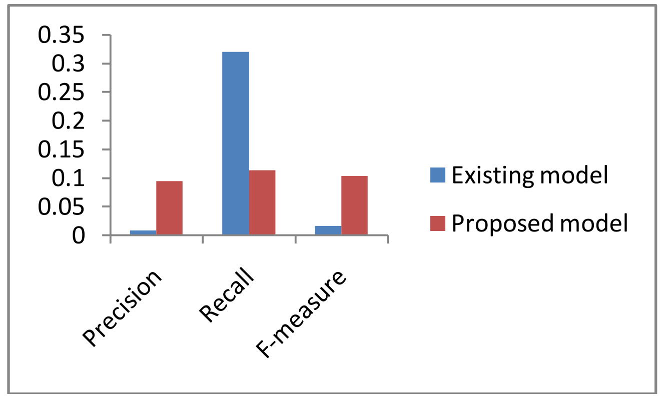

Fig. 4 : Comparison of conventional and proposed model

Microsoft public data base is used to evaluate the performance of the proposed model and is implemented by using MATLAB. The final saliency maps along with local and global saliency maps of DWT and SWT are shown in figs 5 and 6.

Fig. 5 : Local, Global and Final saliency map for a given image by using DWT

Fig. 6 : Local, Global and Final saliency map for a given image by using SWT

V. Conclutions

In this paper, a bottom up computational model of visual attention system based on wavelet coefficients is proposed to obtain the saliency map for images. Various feature maps are generated by applying ISWT to the band pass regions of images at various scales. Using these features, the local and global saliency maps are computed, by combining these maps final saliency map is formed. The performance of the proposed model based on precision, recall an F-Measure is evaluated. Experimental evaluation shows that performance of the proposed model is better than the conventional model.

References

1. Treisman and G. Gelade, (1980) "A feature-integration theory of attention," *Cognit. Psychol.*, vol. 12, no. 1, pp. 97–136.
2. S. Frintrop, (2005) "VOCUS: A visual attention system for object detection and goal directed search," Ph.D. dissertation, Rheinische Friedrich-Wilhelms-Universi¨tat Bonn, Bonn, Germany.
3. L. Itti, (2000). "Models of bottom-up and top-down visual attention," Ph.D. dissertation, Dept. Computat. Neur. Syst., California Inst. Technol, Pasadena.
4. J.M. Wolfe, et. al, (2003) "Changing your mind: On the contributions of top-down and bottom-up guidance in visual search for feature singletons," *J. Exp. Psychol. Human Percept Perform.*, vol. 29, pp. 483–502.
5. O.L. Meur, et. al, (2006) "A coherent computational approach to model the bottom-up visual attention," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 5, pp. 802–817.
6. N. Bruce and J. Tsotsos (2005), Saliency based on information maximization. In *NIPS*, pages 155–162.
7. L. Itti, et. al, (1998) "Model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11 pp.1254–1259, Nov.
8. A. Oliva, et. al, (2003) "Topdown control of visual attention in object detection," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, pp. 253–256.
9. A. Oliva and A. Torralba, (2001) "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vision*, vol. 42, no. 3, pp. 145–175.
10. X. Hou and L. Zhang, (2007) "Saliency detection: A spectral residual approach," in *Proc. IEEE Int. Conf. Comput. Vision and Pattern Recognition*, pp. 1–8.
11. R. J. E. Merry, (2005) *"Wavelet Theory and Application: A Literature Study", DCT 2005.53.* Eindhoven, The Netherlands: Eindhoven Univ. Technol.
12. N. Murray, et. al, (2011) "Saliency estimation using a non-parametric low-level vision model," in Proc. IEEE Int. Conf. Comput. Vision and Pattern Recognition.
13. Nevrez İmamoğlu, Weisi Lin, (2013)*" A Saliency Detection Model Using Low-Level features based on wavelet transform" in* IEEE transactions on multimedia, vol. 15, no. 1, pp. 96-105, January.
14. J. E. Fowler, (2004) "The redundant discrete wavelet transform and additive noise," Mississippi State ERC, Mississippi State University, Tech. Rep. MSSU-COE-ERC-04-04, Mar.
15. Nason G.P. Silverman B.W. (1995), "The Stationary Wavelet Transform and Some Statistical Applications." Tech. Rep. BS8 1Tw, University of Bristol.
16. Y. Kocyigit andM. Korurek, (2005), "EMG signal classification using wavelet transform and fuzzy logic classifier," *ITU dergisi/d mühendislik*, vol. 4, no. 3.
17. John and L. Semmlow, (2004), *Biosignal and Biomedical Image Processing: MATLAB-Based Applications*. New York: Marcel Decker,.
18. S. Theodoridis and K. Koutroumbas, (2009), Pattern Recognition, 4th ed. London, U.K.: Academic/Elsevier, pp. 20–24.
19. S. Goferman, L. Zelnik-Manor, and A. Tal, (2010), "Context-aware saliency detection," in *Proc. IEEE Int. Conf. Comput. Vision and Pattern Recognition*, pp. 2376–2383.
20. T. Liu, et.al, (2007), "Learning to detect a salient object," in *Proc. IEEE Int. Conf. Comput. Vision and Pattern Recognition*, pp. 1–8.

This page is intentionally left blank