World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:4, No:12, 2010

Performance Evaluation of ROI Extraction Models from Stationary Images

K.V. Sridhar, Varun Gunnala, K.S.R Krishna Prasad

***Abstract*—**In this paper three basic approaches and different methods under each of them for extracting region of interest (ROI) from stationary images are explored. The results obtained for each of the proposed methods are shown, and it is demonstrated where each method outperforms the other. Two main problems in ROI extraction: the channel selection problem and the saliency reversal problem are discussed and how best these two are addressed by various methods is also seen. The basic approaches are 1) Saliency based approach 2) Wavelet based approach 3) Clustering based approach. The saliency approach performs well on images containing objects of high saturation and brightness. The wavelet based approach performs well on natural scene images that contain regions of distinct textures. The mean shift clustering approach partitions the image into regions according to the density distribution of pixel intensities. The experimental results of various methodologies show that each technique performs at different acceptable levels for various types of images.

***Keywords*—**clustering, ROI, saliency, wavelets.

I. Introduction

IN today's information society, multimedia contents such as images and videos have become very important. However processing multimedia contents is very slow and its computational complexity is very high. So how to process the multimedia information fast is a key technique that needs to be solved. Researchers have found out that most information is only from some key regions in an image. If these key regions are extracted and processed, the computational speed can be highly improved. Therefore based on this idea regions of interest is developed. The applications of ROI extraction range from image compression wherein different image regions can be compressed at different rates, to content based image retrieval (CBIR), which is a technique that utilizes the visual content of an image to search for semantically similar images in a larger scale database of target images.

The objective of this work is to study some of the important approaches to ROI extraction, features and their limitations.

K.V. Sridhar is with the National Institute of Technology-Warangal, India and is currently working as an Associate Professor.
Varun Gunnala is with the National Institute of Technology-Warangal, India and is currently pursuing Master's in Communication Systems.
K.S.R Krishna Prasad is with National Institute of Technology-Warangal, India and is currently working as a Professor.

Each approach and the different methods under it are also discussed. The first approach is the saliency based, second is the wavelet based and third is the approach based on the mean shift algorithm. These three principle approaches have been tested on five types of images (1) Natural images of Type 1 (2) Natural images of Type 2 (3) Medical images (4) Camouflaged images (5) Noisy images (10 dB salt & pepper noise). Here all the simulations are done on MATLAB® (R2008 version).

The remainder of the paper is organized as follows. In section II, the saliency based approach and the different methods under it are explained. In section III, the wavelet based approach and the different methods under it are explained. In section IV, the mean shift clustering approach is outlined. In section V, the experimentation results are given. In section VI, a quantitative analysis of the results is given. In section VII, the results are summarized.

A. Types of images used

In this paper all the algorithms have been tested on five different types of images. They are: (1) Natural images of Type 1: These images have smooth background and cluttered salient region (Fig 1a). (2)Natural images of Type 2: These images have cluttered background and smooth salient region (Fig 1b). (3)Medical images: Color MRI scans of brain and mammogram images (with tumor) are used (Fig 1c). (4)Camouflaged Images: Highly camouflaged images in which the salient region and the background cannot be easily differentiated are used (Fig 1d). (5)Noisy images: 10dB salt and pepper noise has been added to the above images and is used to check noise robustness for each method.

Figure: 1a 1b 1c 1d

Fig. 1. Types of images used

II. Saliency Based Approach

This approach locates areas within the image that contain high saliency in color, intensity and spatial structure. Such regions are usually objects, parts of objects, or regions standing out from the image background. This approach can be viewed as a transformation from a color or gray scale image to a saliency field.

A. Itti-Koch Algorithm

This is the basic model for saliency based visual attention. Input is provided in the form of a static color image. The visual features are extracted by using Gaussian dyadic pyramids (low pass filtering). Considering red, green and blue channels of the input image, the intensity image I and four color channels(R,G,B and Y) are created. From these, their respective Gaussian pyramids are also constructed. Similarly the orientation information is obtained by Gabor pyramids. Centre surround difference (cross scale difference between 2 maps) is performed to get the feature maps. These maps cannot be directly combined as they represent apriori non comparable modalities. These feature maps are normalized, which globally promotes maps in which a small number of strong peaks of activity is present while suppressing maps which contain numerous comparable peak responses. After normalization the feature maps are combined into three conspicuity maps which are obtained by cross scale addition. The three conspicuity maps are normalized and summed into a final saliency map. To select the most salient image location (Focus Of Attention), the saliency map is modeled as a 2D layer of leaky integrate and fire neurons. The saliency map feeds to another 2D 'Winner Take All' WTA neural network. The neurons in the saliency map receive excitatory inputs from different points in the saliency map. Each saliency map neuron excites its corresponding WTA neuron, until one (the winner) first reaches threshold and fires. This in turn triggers 3 simultaneous mechanisms (1) The FOA is shifted to the location of the winner neuron (2) Global inhibition of WTA is triggered which prevents simultaneous triggering of other WTA neurons (3) Local inhibition in the saliency map is triggered which prevents the triggering of the same WTA neuron, when the FOA is shifted to other point [1].

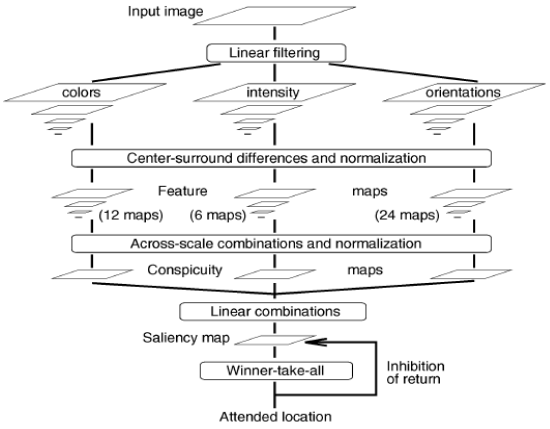

Fig. 2. Architecture of the Itti-Koch model

This method can be slightly altered to get the Modified Itti-Koch algorithm in which the initial feature extraction process is carried out using wavelets by which the interference of the background information on the salient regions can be decreased. Wavelets are used to create 9 different spatial scales and each level is decomposed into red, green, blue, yellow, intensity and local orientation. If the red, green and the blue channels are normalized by image intensity, local orientation information is got by applying wavelets to the intensity. From these channels centre surrounded feature maps are constructed. Over the obtained feature maps, a combination scheme is made to normalize each feature map to a fixed dynamic range. Finally the summation of these maps gives us the desired saliency map.

Features and limitations: 1) Works for natural images of Type 1. 2) Does not work for natural images of Type 2 (does not address the saliency reversal problem (see sec: 2.3)). 3) Does not work for medical and camouflaged images. 4) Channel selection problem (see sec: 2.3) addressed as all channel features (low level) are used. 5) Multi scale analysis performed. 6) Robust to noise. 7) Computational time very low. 8) Size and shape of the detected salient regions is fixed (circle). 9) Its performance depends on the existence of specific neural detectors.10) It cannot reproduce phenomena like contour completion and closure. 11) Suffers with the problem of weighting features from different channels.

B. Spectral Residue Method

In this method concentration is laid on the innovation part by removing the redundant part. Log spectrum representation of the image is used. Consider an image I(x), for which the Fourier Transform is found from which the amplitude and the phase spectrum are calculated. Take the logarithm of the amplitude spectrum, which denotes the spectrum of the complete image (i.e innovation + prior knowledge). The spectral residual is calculated by subtracting the averaged log spectrum from the actual log spectrum of the image (this represents only the innovation part). Finally the saliency map is obtained by taking the IFFT of the spectral residue and the phase spectrum. The entire process can be summarized by the below equations:

$$\mathrm{A(f)} = \Re(\mathrm{FT(I(x))}) \tag{1}$$

$$P(f) = \Im(FT(I(x))) \tag{2}$$

$$L(f) = \log(A(f)) \tag{3}$$

$$R(f) = L(f) - h_n(f) * L(f) \tag{4}$$

$$\mathrm{S(x)} = \mathrm{g(x)} * \mathrm{FT}^{-1}(\exp(R(f) + P(f))) \tag{5}$$

Where $h_n(f)$ is an n^{th} order averaging filter and g(x) is a Gaussian filter [2].

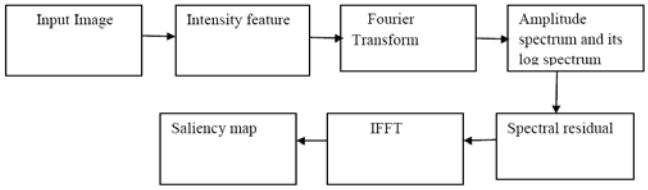

Fig. 3. Architecture of the spectral residue model

Features and limitations: 1) Works for natural images of Type 1. 2) Does not work for natural images of Type 2 (i.e does not address the saliency reversal problem). 3) It uses only one feature (intensity) and one scale (64 x 64) to

compute the saliency map. Therefore it is scale variant and prone to channel selection problem. 4) Does not work for medical and camouflaged images. 5) Robust to noise (as amplitude spectrum is less affected by noise).

C. Improved Spectral Residue Method

This method is an extension of the spectral residual method in which the coarse salient regions are extracted by using the spectral residue model. However here an attempt has been made to resolve channel selection and saliency reversal problems.

Channel Selection Problem: If a region in an image is considered as a salient region then one of its visual channels should be different from the rest. Here the HSV color space is considered. If only one particular channel is selected and the actual contrast mainly resides in other channels, then the algorithm would fail. To overcome this an automatic technique to select the most effective channel is used in which first the saliency maps for all the three channels (H,S,V) are computed and then the k-means clustering for binary clustering is used and finally the saliency map with the largest distance between the two centroids is selected.

$$\text{Effective Channel} = \arg\max_{x} (|\, centroid\,1_x - \text{centroid}\,2_x \,|) \quad (6)$$

Saliency Reversal problem: Saliency is distinguished by contrast of visual properties. There are two basic cases of contrast patterns: smooth background with cluttered salient region (Natural images of Type 1) and smooth salient region with cluttered background (Natural images of Type 2). Spectral residue method is applicable for Type 1 images only and not for Type 2 images. To deal with this, the decision is reversed based on the spatial distribution of salient pixels in the raw saliency map. The spatial variance is calculated as follows:

$$\text{var}(R) = \frac{\sum_{i \in R} \sqrt{(r_i - \overline{r_i})^2 + (c_i - \overline{c_i})^2}}{size\,(R)} \quad (7)$$

$$Inverse = \begin{cases} \text{Yes, var(backg round)} < \beta \text{ x var(raw salient region)} \\ \text{No, otherwise} \end{cases} \quad (8)$$

Here R can be the raw salient region or the background, 'i' is a pixel in R, r_i and c_i are row and column coordinates, ŕ and ć are the average row and column coordinates of all pixels in R, size(R) denotes the total pixel number in R and β is a threshold constant in (0 1] [3].

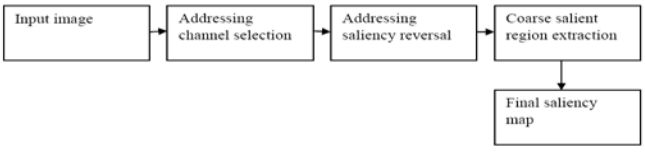

Fig. 4. Architecture of the improved spectral residue model

Features and limitations: 1) Works for natural images of Type 1. 2) Works for natural images of Type 2(addresses the saliency reversal problem). 3) Moderate performance on medical and camouflaged images. 4) Addresses the channel selection problem. 5) Multi scale analysis not performed. 6) Robust to noise 7) Computational time is low (< 10 seconds).

D. Phase Spectrum Method

In this method only the phase spectrum of the feature maps are analyzed and the saliency maps are generated. Morphological operations are applied on the segmented binary image to extract the ROI (Amplitude spectrum represents the frequency of value change. Phase spectrum represents the location where the value changes). The RGB input image is converted into the HSI space from which the feature maps (H, S, I) are automatically selected. This selection is based on the assumption that the salient regions are sparse in the image. Only those features are selected, in whose feature maps the size of the salient regions is less than 70% of the feature map area. On the selected feature maps multi scale phase spectrum analysis is performed to get the saliency maps in each scale. The saliency map calculation is as shown below:

$$f(x, y) = \text{FFT(Feature map)} \quad (9)$$

$$p(x, y) = \text{phase}(f(x, y)) \quad (10)$$

$$\text{Saliency map}(x, y) = \text{IFFT}\,(j*p(x, y)) \quad (11)$$

The multi scale saliency maps of all the selected features are combined to get the integrated saliency map. The integrated saliency map is threshold segmented and the binary image is improved by applying morphological operations. Extracting the edges of the white region in the binary image the edge map (contour of the ROI) is got. Finally the extraction result is got by adding the edge map to the input image [4][10].

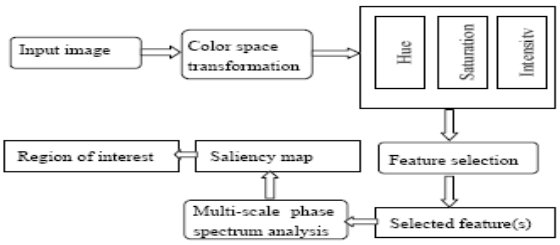

Fig. 5.Architecture of the phase spectrum method

Features and limitations: 1) Works well for natural images of Type 1. 2) Works well for camouflaged images. 3) Partially works for medical images (i.e only the edges of the tumor are extracted). 4) Does not work for natural images of Type 2 (i.e does not address the saliency reversal problem). 5) Addresses the channel selection problem by an automatic channel selection method. 6) Scale invariant as multi-scale analysis is performed. 7) Sensitive to noise. 8) Computational time is low (< 50 sec).

E. Contrast Based Method

In this approach the global contrast of each pixel is computed and is used to construct the saliency map. In the first step the image is resized to reduce the computational time and then the image is transformed from the RGB space to the HSI space. Saliency map is obtained by calculating the contrast between two pixels x and y for I, H and S features

individually and summed up to get the saliency map. The following equations explain the procedure.

The Intensity contrast is $\Delta I(x,y) = |I(x) - I(y)|$ (12)

The Hue contrast is $\Delta H(x,y) = |H(x) - H(y)|$ (13)

The saturation contrast is $\Delta S(x,y) = |S(x) - S(y)|$ (14)

Therefore the saliency map S(x) of pixel x is calculated by

$$S(x) = \sum_{y=1}^{m*n} (\Delta I(x, y) + \Delta H(x, y) + \Delta S(x, y)) \quad (15)$$

where m x n are number of pixels of the image. Finally the saliency map is segmented using a dynamic threshold. Then the edge map of the segmented image is obtained which is added to the original image to extract the ROI [5].

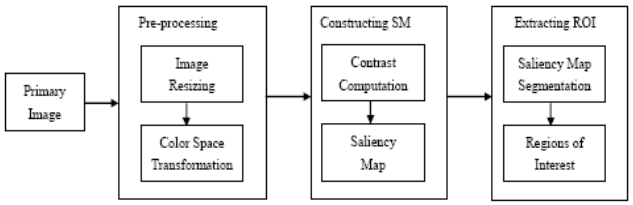

Fig. 6. Architecture of the contrast based model

Features and limitations: 1) Works best for medical images. 2) Works well for natural images of Type 1. 3) Does not work for camouflaged images as contrast is uniform throughout. 4) Conditionally addresses the saliency reversal problem of Type 2 images (i.e if the smooth foreground has high contrast, it is extracted). 5) Channel selection is avoided by computing saliency maps using all 3 feature maps 6) Multi scale analysis is not done as the effect of multi-scales on contrast is negligible.7) Robust to noise. 8) Computational time is very high (if original image size is considered). 9) Works only for color images

F. Graph Based Visual Saliency (GBVS) Method

This method is an example of a bottom-up visual saliency model. It has two steps, first forming activation maps on certain feature channels and then normalizing them in order to highlight the conspicuity points. First the feature maps are extracted from the input image. These feature maps are used to construct activation maps by using a Markovian approach. The dissimilarity between two nodes is defined as the distance between them. A fully connected directed graph is now considered. The directed edge from each node to the other is assigned a weight which is proportional to their dissimilarity. A Markov chain is defined on the graph by normalizing the weights of the outbound edges of each node to 1 and drawing equivalence between nodes, states, edge weights and transition probabilities. This process will accumulate mass (here saliency) at nodes that have high dissimilarity with the surrounding nodes thus creating the activation map. These activation maps must be normalized prior to additive combination for which again a graph is constructed for the activation map and the weights of the outbound edges of each node are normalized to unity by treating the resulting graph as a Markov chain and then the equilibrium distribution over the nodes is computed. Therefore mass will flow to nodes with high activation. Finally the normalized activation maps are combined to get the saliency map [6].

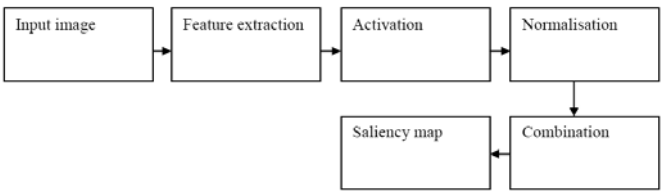

Fig. 7 Architecture of the GBVS model

Features and limitations: 1) Here the computational power, topographical structure and parallel nature of graph algorithms are exploited. 2) Works for natural Images of Type 1. 3) Does not work for natural images of Type 2 (does not address the saliency reversal problem). 4) Does not work for medical images. 5) Moderate performance for camouflaged images. 6) Robust to noise. 7) Computational time is low (< 10 seconds). 8) Channel selection problem is addressed by combining activation maps of each feature map. 9) Multi-scale linear filtering not performed.

III. WAVELET BASED APPROACH

This approach locates areas within the image that usually contain regions of different textures. This method can be viewed as a transformation from a image to the wavelet domain, where the variances for the sub-bands are calculated and are used along with a clustering algorithm or a saliency method to extract the region of interest in an image.

A. Block DWT-ROI Extraction Method

The DWT at the first level of resolution, applied to an image x $[n_1, n_2]$, produces the discrete wavelet coefficients of 4 frequency sub-bands: LL, LH, HL and HH, where LL is the low frequency sub-band in both the horizontal and vertical frequencies respectively. Applying the same transformation again to the LL sub-band further subdivides the LL sub-band into LL, LH, HL and HH sub-bands, hence obtaining the transform coefficients at a second level of resolution. For an M-level decomposition, the lowest frequency sub-band LL_M represents the original image (low frequency) approximation. The higher frequency sub-bands mainly represent edge and texture information. The procedure is as follows:

Consider the M x N color image matrix I as in equation (16)

$$I = \{I(i, j); 0 \le i \le M - 1, \ 0 \le j \le N - 1\} \quad (16)$$

I(i, j) is a 3-D vector representing the pixel color and intensity components at location (i, j). The image is transformed to the HSV color space since the HSV space provides a better correspondence with human visual perception of color than the RGB space. As the hue component varies from 0 to 1, the corresponding colors vary from red through yellow, green, cyan, blue, magenta, and back. As saturation varies from 0 to 1, the corresponding colors (hue) vary from unsaturated (shades of gray) to fully saturated (no white component). As value, or brightness, varies from 0 to 1, the corresponding colors become increasingly brighter. The ROI are extracted

only from the brightness component of the image, since the human visual system is most sensitive to the V component. The region extraction proceeds according to the following steps: The *V* component of the M x N image *I* is divided into non-overlapping blocks of size L x W in the spatial domain. A matrix *U* of DWT coefficients for each block is obtained. If a block in *I* at location (i, j) is represented as shown in equation (17), where $0 \le i < M/L$, and $0 \le j < N/W$, then the corresponding LH, HL and HH blocks of DWT coefficients in *U* are computed using equation (18) where λ is either LH, HL or HH.

$$V_{i,j} = \{v(m \times i, n \times j), 0 \le m < L, 0 \le n < W\} \quad (17)$$

$$U_{i,j}^{\lambda} = \{u^{\lambda}(m\,x\,i, n \times j), 0 \le m < \frac{L}{2}, 0 \le n < \frac{W}{2}\} \quad (18)$$

The second level DWT is applied to the image to obtain a multi-scale representation of the visual content. The variances of the LH, HL and HH sub-band coefficients for each block DWT in U are then computed as in equations (19) and (20).

$$\mu_{i,j}^{\lambda} = \frac{1}{L/2 \times W/2} \sum_{m=0}^{L/2-1} \sum_{n=0}^{W/2-1} | \mu^{\lambda}(mi, nj) | \quad (19)$$

$$\sigma_{i,j}^{\lambda} = \frac{1}{L/2 \times W/2} \sum_{m=0}^{L/2-1} \sum_{n=0}^{W//2-1} (| \mu^{\lambda}(mi, nj) | - \mu_{i,j}^{\lambda})^2 \quad (20)$$

Next, the sub-band variances for each block are averaged to obtain $\sigma_{i,j}$ which represents the image segmentation Eigen value, or the image segmentation feature. The image segmentation Eigen values are then, clustered by the fuzzy c-means clustering algorithm. Each cluster represents a possible region of interest. The higher the Eigen value of a block, the more likely it is to belong to a region of interest. Each block is then, labeled by its cluster number. A threshold on the size of the cluster can be chosen to avoid region sizes below a certain threshold [7].

Features and limitations: 1) Moderate performance for natural image of Type 1. 2) Works for natural images of Type 2 (addresses the saliency reversal problem). 3) Moderate performance for medical and camouflaged images. 4) Multi scale analysis is not performed. 5) Channel selection problem is not addressed as only intensity feature is selected. 6) Computational time is low(< 30 seconds).7) Robust to noise.

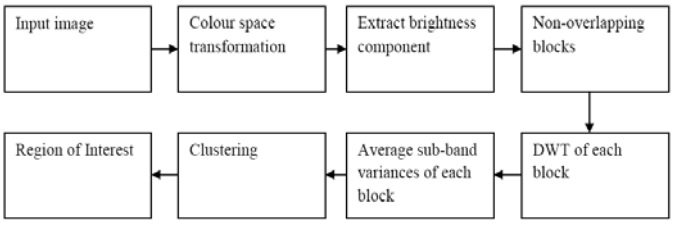

Fig. 8. Architecture of the block DWT-ROI method

B. Wavelet Coefficient Variance Saliency (WCVS) Method

This method can be viewed as a transformation from a color image to saliency field based on wavelet coefficients variance. The wavelet coefficients variance field is defined as the 3-dimensional sigmoid function shown in equation (21) where HLvar (x,y), LHvar (x,y) and HHvar (x,y) are the local variances for the HL, LH, HH channels of the wavelet transform of the original gray scale version of the image at location (x,y). The weights of the different channels are controlled by the coefficients p, q, and r which are all set to 0.5 in the experiments. $HLvar_{ave}$, $LHvar_{ave}$, and $HHvar_{ave}$ are the mean values of the local variances corresponding to the wavelet channels. The resulting saliency is normalized to be in the range [0, 1] [8].

$$f_w(x,y) = \frac{1}{p + \exp(-\frac{hlvar(xy)}{hlvar_{ave}})} \cdot \frac{1}{q + \exp(-\frac{lhvar(xy)}{lhvar_{ave}})} \cdot \frac{1}{r + \exp(-\frac{hhvar(xy)}{hhvar_{ave}})} \quad (21)$$

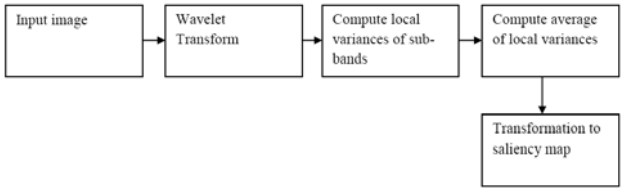

Fig. 9. Architecture of the WCVS method

Features and limitations: 1) Works well for natural images of Type 1. 2) Does not work for natural images of Type 2 (i.e does not address saliency reversal problem). 3) Does not work effectively for medical images. 4) Works well for camouflaged images. 5) Channel selection problem is not addressed as only intensity feature is selected. 6) Scale variant as multi scale analysis is not performed. 7) Interference of background on foreground is very less. 8) Computational time is low (< 20 seconds). 9) Not robust to noise (can withstand up to 1.423 dB of noise).

IV. CLUSTERING BASED APPROACH

This approach groups pixels based on a certain criteria. Each group represents a salient region in the image. Clustering approaches are of two types: 1.Parametric 2.Non Parametric. Here a Non Parametric clustering approach is discussed which clusters image pixels based on pixel intensities.

B. Mean Shift Clustering (MSC) Approach

In this approach to ROI extraction, the mean shift algorithm is used to cluster image pixels. Every pixel in the image is represented as a five dimensional vector of two spatial coordinates i and j, and three color space components H (hue), S (saturation) and V (value). The clusters are located by applying a search window in the feature space, which shifts towards the cluster centre. The magnitude and the direction of the shift in the feature space are based on the difference between the window centre spatial coordinates and the cluster centroid coordinates. The centre of the search window eventually converges to the cluster centroid .When applying the mean shift algorithm to color segmentation of an image, several randomly chosen locations in feature space are considered and the one with the highest density of feature vectors is selected. This is done to make sure that the search starts in a high-density region and thus reduces the number of shifts needed to reach convergence. The mean shift algorithm

can be summarized as follows: Choose the radius r of the search window. Choose the initial location of the window. Compute the mean shift vector and translate the search window by that amount. Repeat till convergence. The outline of the general procedure is given as follows: Map the image domain into the feature space. Define an adequate number of search windows at random locations in the space. Find the high-density region centers by applying the mean shift algorithm to each window [9].

Features and limitations: 1) It partitions the image into regions according to the density distribution of pixel intensities. 2) Problem of optimal thresholds for segmenting is implicitly resolved. 3) Moderate performance for natural images of Type 1. 4) Works well for natural images of type 2 (i.e addresses the saliency reversal problem). 5) Moderate performance for medical images. 6) Does not work for camouflaged images. 7) Multi scale analysis is not performed 8) Channel selection problem is not addressed as image is partitioned into regions according to the density distribution of pixel *intensities.* 9) Robust to noise.

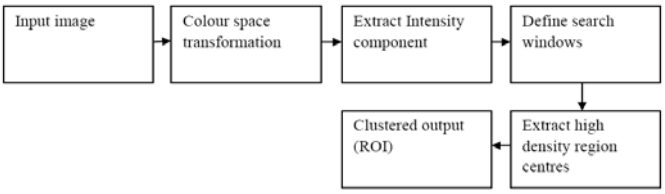

Fig. 10. Architecture of the MSC approach

V. Experimental Results

The three approaches are implemented on MATLAB (R2008b) on five types of images as mentioned previously, taken from the web. Saliency based approaches perform well on images having objects of high saturation and brightness. The Itti-Koch algorithm is computationally very complex as it uses cross scale subtractions and additions and works only for natural images of Type 1 and does not address the saliency reversal problem. The GBVS method is an extension of the Itti-Koch algorithm in which the activation of the feature map and normalization of the activation maps are done by using graph algorithms. Its performance is better than Itti-Koch algorithm. It works well for Type 1 images and moderate performance is seen for camouflaged images. It also does not address the saliency reversal problem. The spectral residual approach is a very simple one in which the conspicuous points are highlighted. It uses only the intensity feature map and a single scale for its saliency map. It works for natural images of Type1 and does not address any of the discussed problems. This method is extended in the improved spectral residue method, in which the channel selection and saliency reversal problems are resolved. The phase spectrum method uses only the phase spectrum and not the amplitude spectrum for the saliency map. The saliency map obtained by this method is much sharper when compared to other methods. It addresses only the channel selection problem and works for natural images of Type 1 and camouflaged images. For medical images only the contour of the tumor is extracted without highlighting the tumor as the salient region. In the contrast based method the saliency map is constructed by computing the global contrast of each pixel. It extracts regions of high contrast of background or foreground. It works the best for medical images and natural images of Type 1 and 2. Wavelet based methods extract regions of different textures. The WCVS method constructs the saliency map based on the variance of sub bands. It does not address channel selection and saliency reversal problems. It works best for natural images of Type 1 and camouflaged images. The block DWT method works best on Type 2 images and moderate performance is seen on other types. This method requires the selection of several design parameters. These include the size of the image blocks and the number of clusters as an input argument to the fuzzy c-means clustering algorithm. The MSC approach partitions the image into regions according to the density distribution of pixel intensities. It works well for natural images of Type 2 and moderate performance is seen for Type 1 and medical images. Here the number of clusters is selected automatically but the
size of the search window and the number of search windows is selected manually. Finally for all methods for ROI extraction, a threshold needs to be selected between 0 and 1. Table I tabulates the extraction results.

TABLE I COMPARISON OF EXTRACTION RESULTS OF VARIOUS METHODS

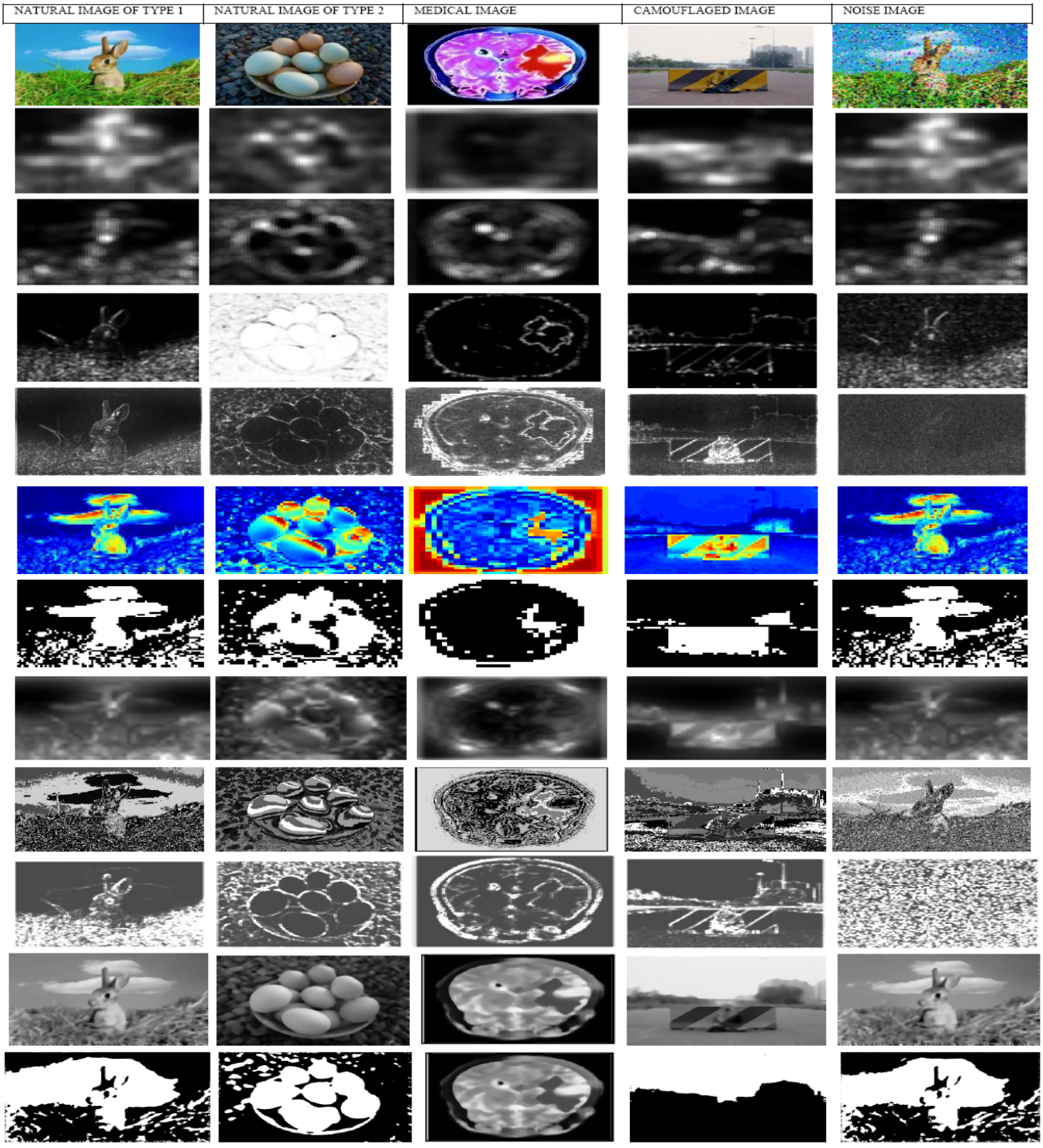

Table I shows the saliency maps of all the methods discussed in this paper. Row 1: Original images. Row 2 : Saliency maps of Itti-Koch algorithm. Row 3: Saliency maps of spectral residue method. Row 4: Sm's of Improved spectral residue method. Row 5: Sm's of Phase spectrum method. Row 6 and 7: Contrast sm's and threshold sm's respectively of Contrast based method. Row 8: Sm's of GBVS method. Row 9: Output image of Block DWT method. Row 10: Sm's of WCVS method. Row 11 and 12: Mean shift clustered and threshold images respectively of MSC algorithm. (sm: saliency map).

VI. Quantitative Analysis Of The Results

To give a quantitative analysis, the extraction results are evaluated by computing the correlation between the extracted results and the human labeled results using equation (22).

$$r = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(bm_{ij} - \overline{bm})(hl_{ij} - \overline{hl})}{\sqrt{\left(\sum_{i=1}^{m}\sum_{j=1}^{n}(bm_{ij} - \overline{bm})^2\right)\left(\sum_{i=1}^{m}\sum_{j=1}^{n}(hl_{ij} - \overline{hl})^2\right)}} \quad (22)$$

where 'bm' is the segmented binary image and 'hl' is the human labeled result. Here the segmented binary image is got by taking the threshold as mean of the final output of the corresponding approach. This analysis is performed only for natural images of Type 1. Below shown are the original images and their corresponding human labeled images. In the latter image the white region represents the salient region which was accepted by all the viewers and the gray region represents the salient region which was accepted by few and rejected by others.

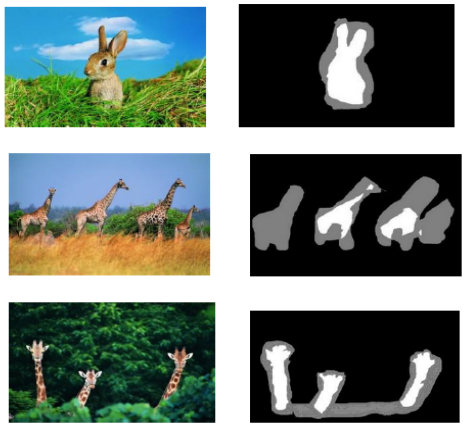

Figure 11. Original image Human labeled image

Table II Comparison Of Correlation Coefficients Of Various Methods For Natural Images Of Type 1

METHOD	FIGURE A	FIGURE B	FIGURE C
Itti-Koch	0.4632	0.6239	0.5992
Spectral Residue	0.2426	0.6970	0.5020
Improved Spectral Residue	- 0.0186	0.5838	0.5746
Phase Spectrum	0.1974	0.5684	0.3670
Contrast Based	0.4070	0.3655	0.5350
GBVS	0.1764	0.6374	0.6865
Block DWT	0.1409	0.1590	0.1307
WCVS	- 0.0975	0.5414	0.4310
MSCA	0.2048	- 0.3054	0.3014

Table II shows the correlation between human labeled images and the threshold output images of Type 1 of various methods. Here the threshold level is assumed as the mean of the saliency map. The anomalies in figure A in some methods can be attributed to the three stage nature of the image (i.e the most salient grass, then the human labeled rabbit and finally the background).

VII. SUMMARY

Table III summarizes all the methodologies discussed in this paper. A study over the problems addressed by each method and the types of images each method works for is made.

VIII. CONCLUSIONS AND FUTURE WORK

The problem of ROI extraction remains a challenging research area. Specifying where a region in an image is potentially an ROI is itself a task that requires subjective evaluation. Another challenge lies in defining an objective method or a baseline for evaluating the performance of various ROI extraction algorithms. Another challenge is to properly address the problems of channel selection and saliency reversal. In this paper color is considered as the feature of the image. However, it is very likely that there are some other features such as edge and symmetry which also should be considered. What feature and how many features should be extracted according to the target are questions which require further research.

TABLE III SUMMARY

APPROACH	METHODOLOGY	CHANNEL SELECTION PROBLEM	SALIENCY REVERSAL PROBLEM	MULTISCALE ANALYSIS	WORKS FOR IMAGE TYPES	ROBUSTNESS TO NOISE
Saliency Based						
	1.Itti-Koch	Addressed by taking all channel features	Not addressed	Performed	Natural (T1) images	Robust
	2.Spectral residue	Not addressed (only Intensity fm considered)	Not addressed	Not performed (only one scale used)	Natural(T1) images	Robust
	3.Improved spectral residue	Addressed by using k-means clustering	Addressed by calculating the spatial variances	Not performed	Natural(T1&T2), Partial results for medical and camouflaged images	Robust
	4.Phase spectrum	Addressed by considering salient regions are sparse in fm's	Not addressed	Performed	Natural(T1),camouflaged, Partial results for medical images	Not Robust
	5.Contrast based	Addressed by taking all channel features	Conditionally addressed(extracts regions of high contrast)	Not performed	Medical, Natural(T1 and T2) images	Robust
	6.GBVS	Addressed (by combining activation maps of each feature map)	Not addressed	Not performed	Natural(T1), Partial results for camouflaged images	Robust
Wavelet based						
	1.Block DWT	Not addressed (only intensity fm considered)	Addressed	Not performed (only 1 level DWT applied)	Natural(T2), Partial results for Natural(T1), Medical, Camouflaged images	Robust
	2.WCVS	Not addressed (only intensity fm considered)	Not addressed	Not performed	Natural(T1), Camouflaged images	Not Robust
Clustering Based						
	1.MSCA	Not addressed as clustering is based on pixel intensities	Addressed	Not performed	Natural(T2), Partial results for Natural(T1), Medical images	Robust

REFERENCES

[1] L Itti, C Koch and E Niebur, " A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", Proc. IEEE Transactions on Pattern Analysis and machine Intelligence, vol. 20,no. 11,November 1998.

[2] Xiaodi Hou, Liquing Zhang, "Saliency Detection: A Spectral Residual Approach", Proc. IEEE Conference on Computer Vision and Pattern recognition, June 2007.

[3] Zheshen Wang, Baoxin Li, "A Two Stage Approach to Saliency Detection in Images", Proc. IEEE Conference on Acoustics, Speech and Signal Processing, March 2008.

[4] Bin Zhang, Yafeng Zheng, Qiaorong Zhang, "Extracting Regions of Interest Based on Phase Spectrum and Morphological Approach", Proc. ISECS International Colloquium on Computing, Communication, Control and Management, May 2009.

[5] Qiaorong Zhang, Huimin Xiao, "Extracting Regions of Interest in Biomedical Images", Proc. International seminar on Future BioMedical Information Engineering, December 2008.

[6] J.Harel, C.Koch, P.Perona, "Graph Based Visual Saliency", Proc. NIPS, December 2006

[7] W.Xiangyang, Y.Hongying, H.Fengli,"A New Regions of Interest Based Image Retrieval Using DWT", Proc. ISCIT, October 2005.

[8] Q.Zhou, L.Ma, M.Celenk and D.M. Chelberg, "Content Based Image Retrieval Based on ROI Detection and Relevance Feedback", Multimedia Tools Appl, 27(2), 2005.

[9] J.Goldberger, S.Gordon and H.Greenspan, "Unsupervised Image Set Clustering Using an Information Theoretic Framework", IEEE Trans on Systems, Man and Cybernetics, vol 37, no:5, October 2007

[10] Qiaorong Zhang, Yafeng Zheng, Yafeng Zheng, "Automatically Extracting Salient Regions in Natural Images", Proc. ISECS International Colloquium on Computing, Communication, Control and Management, May 2009.

[11] R.C Gonzales, R.E Woods, Digital Image Processing, 2nd Edition, Prentice Hall, ISBN 0-201-18075-8.