Chapter 1
Problem Definition

1.1 Introduction to Saliency

The **salience** (also called **saliency**) of an item – be it an object, a person, a pixel etc. is the state or quality by which it stands out relative to its neighbors. Saliency detection is considered to be a key attentional mechanism that facilitates learning and survival by enabling organisms to focus their limited perceptual and cognitive resources on the most pertinent subset of the available sensory data.

Saliency typically arises from contrasts between items and their neighborhood, such as a red dot surrounded by white dots, a flickering message indicator of an answering machine, or a loud noise in an otherwise quiet environment. Saliency detection is often studied in the context of the visual system, but similar mechanisms operate in other sensory systems. What is salient can be influenced by training: for example, for human subjects particular letters can become salient by training.

Identifying visually salient regions is useful in applications such as object based image retrieval, adaptive content delivery [1, 2], adaptive region-of-interest based image compression, smart image resizing [3], image segmentation [4] and object recognition [5]. We identify salient regions as those regions of an image that are visually more conspicuous by virtue of their contrast with respect to surrounding regions. Similar definitions of saliency exist in literature where saliency in images is referred to as local contrast.

1.2 Salient Region detection techniques

The approaches for determining saliency can be based on biological models or purely computational ones or combination of both. Some approaches consider saliency over several scales while others operate on a single scale. In general, all methods use some means of determining local contrast of image regions with their surroundings using one or more of the features of color, intensity, and orientation. Usually, separate feature maps are created for

each of the features used and then combined to obtain the final saliency map. To date, a few methods have been proposed in literature to detect salient objects and regions from images. These are summarized as follow:

- Achanta's saliency detection method [6] determines salient regions in images using low-level features of luminance and color. This method uses a contrast determination filter which operates at various scales to generate saliency maps containing saliency values per pixel.
- The frequency tuned salient region algorithm [7] basically makes estimation regarding centre surround contrast using the features which are based on color and luminance value. This algorithm achieves pre-attentive, low level, bottom-up saliency. This method is inspired by biological center surround concept but is not based on any biological model [8].
- Maximum system surround method [9], assumption regarding the scale of object with respect to image border is made, thus bandwidth of the center surround filtering near the image boarder is varied using symmetric surrounds concept.
- Non parametric low level vision method [10] uses local feature to determine saliency by convolving the image with bank of filter using multi-resolution wavelet transform and scale-weighting function termed Extended Contrast Sensitivity Function (ECSF) has been optimized to better replicate psychophysical data on color appearance,
- Context aware saliency detection [11] unifies local and global saliency by measuring the similarity between each image patch and other image patches, both locally and globally and resemblance between patches used for saliency estimation.
- Spectral residual method [12] compute the log spectrum of input image and extract the spectral residual of an image in spectral domain, proposes a method to construct saliency map in spatial domain.

1.3 Problem Identified

All the techniques proposed in literature for salient region detection from images have some drawbacks and limitations which can be summarized as:

- There are some techniques which does not detect the object in presence of complex background.
- Only few techniques provide the saliency map of full resolution but object boundaries are not properly defined.
- Saliency map obtained from some technique are not uniformly highlighted.
- Some saliency detection techniques have very high computational complexity so cannot be used for real time applications.
- Technique which uses local features only fails to detect may object pixels.
- Technique which uses global features only detects many background pixels as object pixels.

1.4 Proposed Technique

In this thesis work, a new technique is proposed to improve the performance of saliency detection from color images keeping following aspects in mind to:

- Provide high perceptual quality saliency map.
- Uniformly highlight the salient region.
- Generate saliency map of full resolution.
- Obtain saliency map with low computational complexity.
- Incorporate both local and global features.
- Detect object boundaries properly.

1.5 Tool Used

MATLAB is used as simulator to implement the techniques. MATLAB provides high computing environment and advanced in-built functions for image processing.

Why matlab...?

Matlab is an integrated technical computing environment that combines numeric computation, advanced graphics and visualization. It is a high level programming

language that can communicate with its cousins, e.g. FORTRAN and C. MATLAB allows matrix manipulation, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs in other languages. The language, tools and built-in math functions enables to explore multiple approaches and reach a solution faster than other programming languages. Besides this Matlab is also used for

- Communications Systems
- Computational Biology
- Computational Finance
- Control Systems
- Digital Signal Processing
- Embedded Systems
- FPGA Design and Code design
- Image and Video Processing
- Technical Computing
- Test and Measurement

1.6 Performance Evaluation Matrices

Following parameters are taken to measure the performance of salient region detection techniques:

1.6.1 Precision

In pattern recognition and information retrieval with binary classification, precision is the fraction of retrieved instances that are relevant. High precision means that an algorithm returned substantially more relevant results than irrelevant.

1.6.2 Recall

In pattern recognition and information retrieval with binary classification, recall is the fraction of relevant instances that are retrieved. High recall means that an algorithm returned most of the relevant results.

For classification tasks, the terms true positives, true negatives, false positives, and false negatives compares the results of the classifier under test with trusted external judgments.

The terms positive and negative refer to the classifier's prediction (sometimes known as the expectation), and the terms true and false refer to whether that prediction corresponds to the external judgment (sometimes known as the observation).

Let us define an experiment from **P** positive instances and **N** negative instances for some condition. The four outcomes can be formulated in a 2×2 contingency table or confusion matrix, as follows:

		Condition positive	Condition negative
Test outcome	Test outcome positive	**True positive**	**False positive** (Type I error)
	Test outcome negative	**False negative** (Type II error)	**True negative**

Fig.1.1 – confusion matrix

Precision and recall are then defined as:

$$\textbf{Precision} = \frac{\textbf{True Positive}}{\textbf{True Positive} + \textbf{False Positive}}$$

$$\textbf{Recall} = \frac{\textbf{True Positive}}{\textbf{True Positive} + \textbf{False Negative}}$$

1.6.3 F-Measure

A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score:

$$\mathbf{F} = \mathbf{2} \cdot \frac{\mathbf{precision} \cdot \mathbf{recall}}{\mathbf{precision} + \mathbf{recall}}$$

This is also known as the F_1 measure, because recall and precision are evenly weighted.

A special case of the general F_β measure (for non-negative real values of β) is given as:

$$\mathbf{F_\beta = (1 + \beta^2) . \frac{precision . recall}{\beta^2 . precision + recall}}$$

Two other commonly used F measures are the F_2 measure, which weights recall higher than precision, and the $F_{0.5}$ measure, which puts more emphasis on precision than recall.

1.7 Organization of Thesis

The remaining part of the thesis is organized into five chapters:

- Chapter 2: The various types of digital images are described.
- Chapter 3: Literature survey explains various techniques and then problem is identified based on the literature survey.
- Chapter 4: proposed technique is described.
- Chapter 5: simulation setup parameters and then results.
- Chapter 6: include the concluding remarks of various results and the future scope of the thesis. In the end references of research papers and books are included.

Chapter 2

Introduction to Digital Images

A digital image can be considered as a large array of discrete dots, each of which has a brightness associated with it. These dots are called picture elements, or more simply pixels. A Grayscale digital image is shown in the figure Fig 2.1.

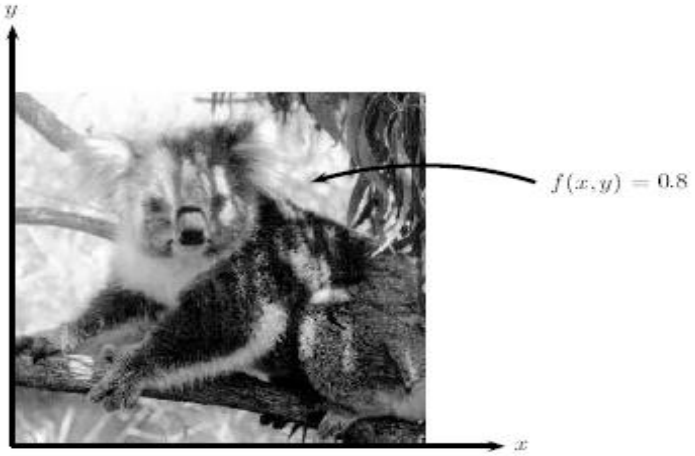


Fig 2.1 – Digital Image representation

2.1 Types of Digital Images

There are four basic types of images.

Binary – These images are black and white images. Only one bit is required to store a pixel. Bit zero represents black pixels while bit one represents white pixels.

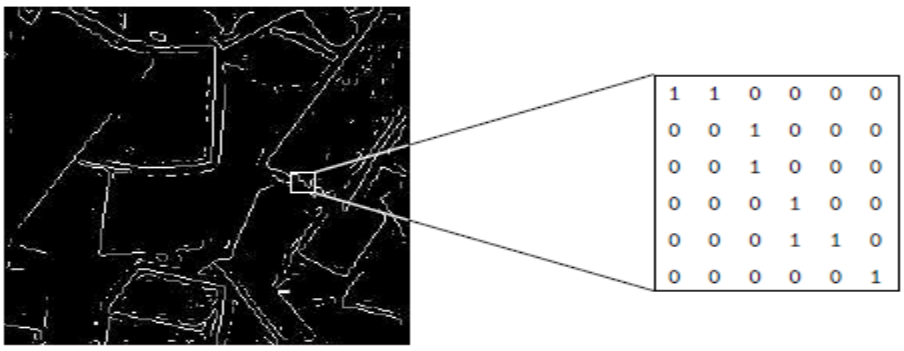


Fig.2.2 – Binary image

Grayscale – The image pixels are in the shade of gray. Pixel value ranges from 0(black) to 255(white). This means eight bits are required to store a pixel.

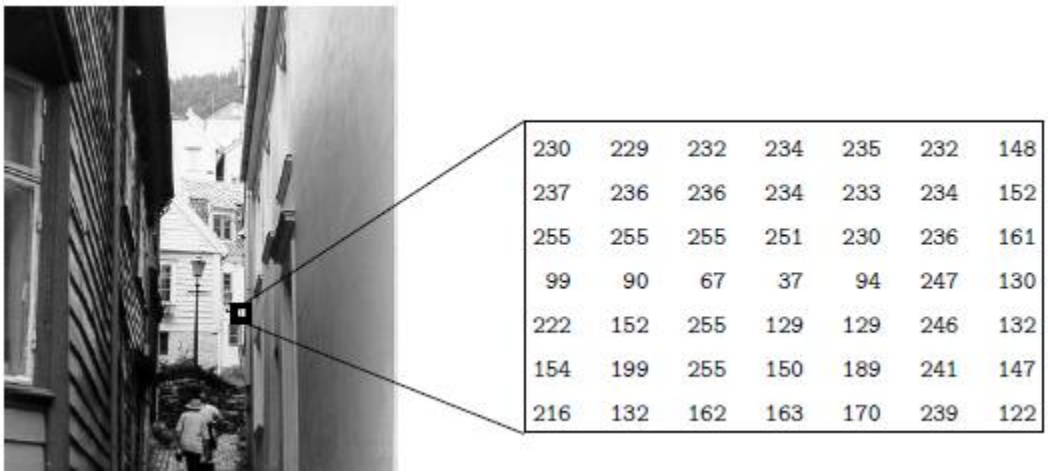


Fig.2.3- Grayscale image

True color - Here each pixel has a particular color and that color being described by the amount of red, green and blue in it. The total number of bits required for each pixel is 24, one byte for each color plane.

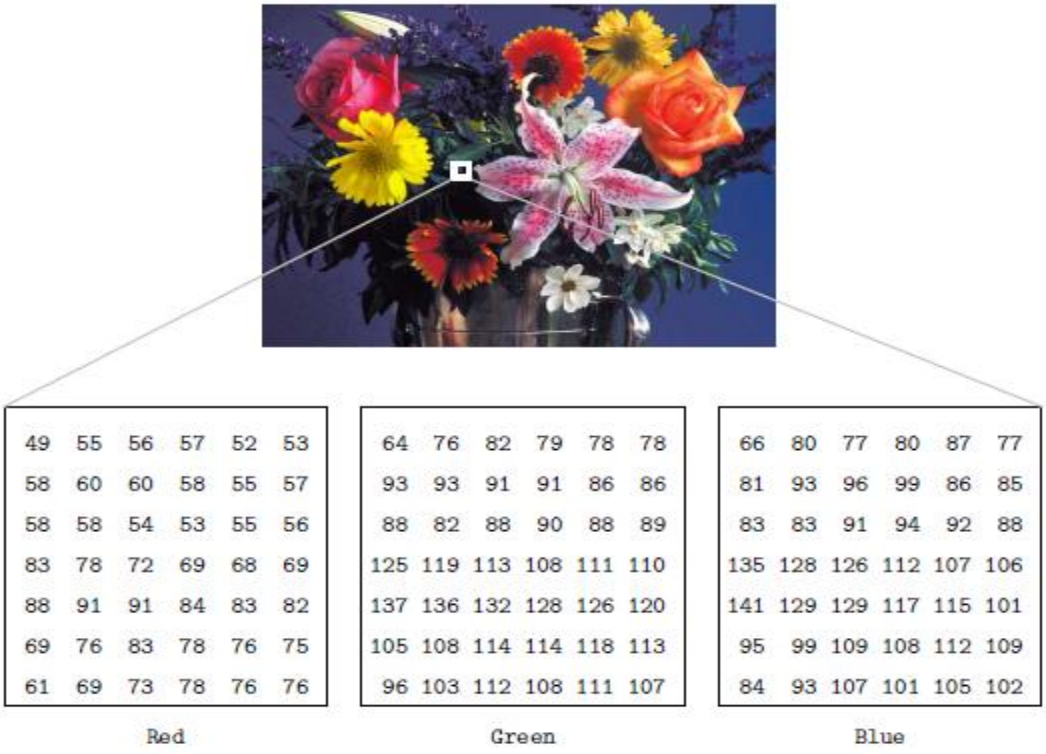


Fig 2.4 – True color image

Indexed - Each pixel has a value which does not give its color (as for True Color image), but an index to the color in the map. It is convenient if an image has 256 colors or less, for then the index values will only require one byte each to store.

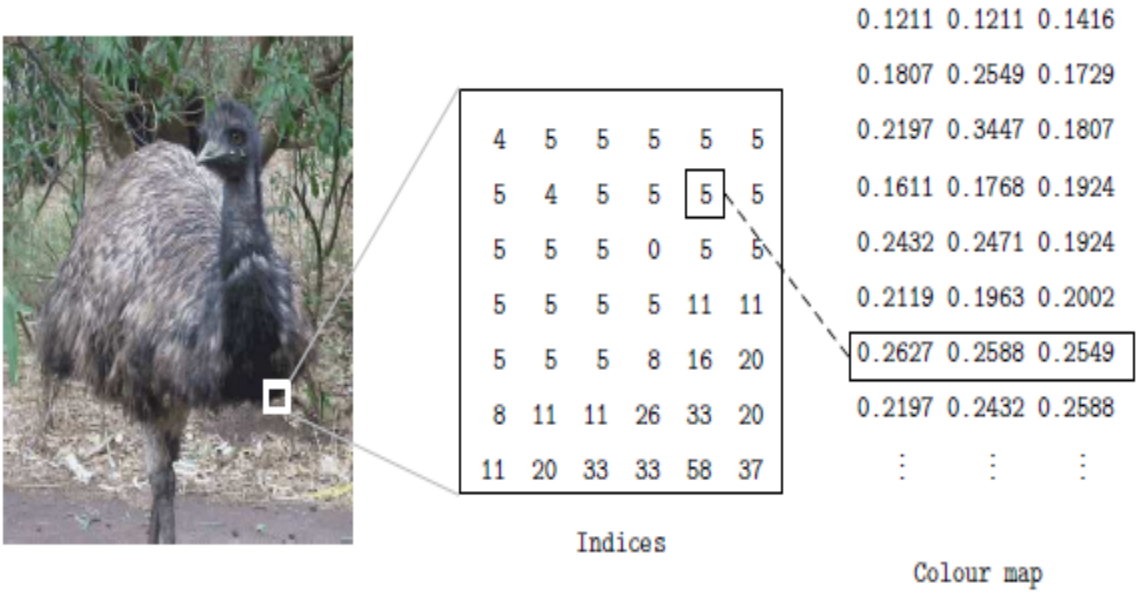


Fig. 2.5 – Indexed image

Chapter 3

Literature Survey & Problem Identification

Several techniques have been proposed in literature for saliency detection from true color images. In this section, first of all some of these techniques are explained which are implemented to compare the performance of proposed techniques. Then various problems of these techniques are explained which leads to problem identification of the proposed techniques.

3.1 Previous Techniques Implemented

3.1.1 Achanta Saliency Detection Method

This is a novel method to determine salient regions in images using low-level features of luminance and color. This method uses a contrast determination filter which operates at various scales to generate saliency maps containing saliency values per pixel. Finally, saliency map is obtained by combining the individual map obtained.

In this, saliency is determined as the local contrast of image region at various scales with respect to its neighborhood. The saliency map at a given scale can be obtained by using feature vector at a given scale. Feature vector can be calculated as distance between the average feature vector of the pixels with in image sub-region and average feature vector of pixels with in neighborhood.

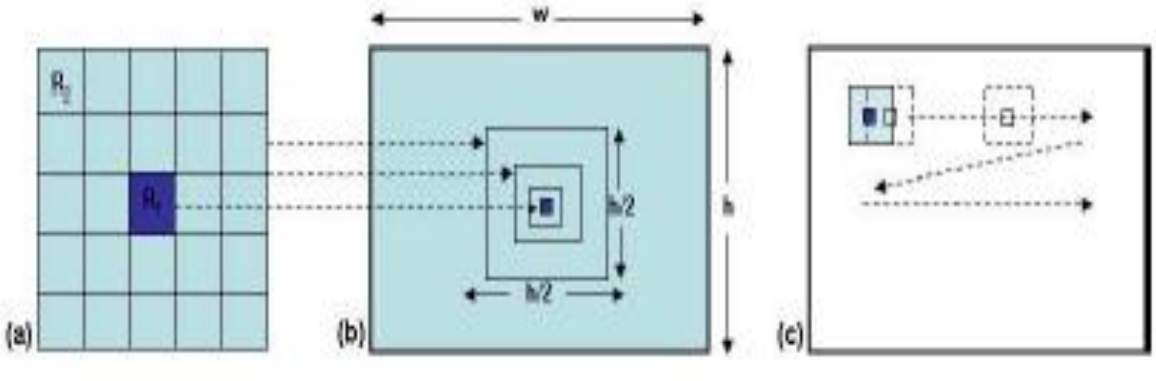


Fig 3.1- (a) Contrast determination filter with inner and outer region. (b) Variation of width of R_2 according to the scale. (c) Image filtering at one scale.

At a given scale, contrast based saliency value $d_{i,j}$ for the pixel position at (i, j) is given by the distance 'D' between the average feature vectors of pixels with in region R_1 to average feature vectors of pixels with in region R_2.

$$\mathbf{d_{i,j}} = \mathbf{D}\left\{\left(\frac{\mathbf{1}}{\mathbf{n_1}}\sum_{\mathbf{l=1}}^{\mathbf{n_1}}\mathbf{v_l}\right),\left(\frac{\mathbf{1}}{\mathbf{n_2}}\sum_{\mathbf{m=1}}^{\mathbf{n_2}}\mathbf{v_m}\right)\right\}$$

Where n_1 and n_2 denotes the number of pixels in region R_1 and R_2 respectively, v_l and v_m denotes the feature vector corresponding to the pixel. In this algorithm, CIE Lab color space [13] is used to obtained feature vectors of color and luminance. In CIE Lab color space, the contrast based saliency value $d_{i,j}$ is defined as:

$$\mathbf{d_{i,j}} = \|\mathbf{v_1} - \mathbf{v_2}\|$$

Where $v_1 = [L_1, a_1, b_1]$ and $v_2 = [L_2, a_2, b_2]$ are the average feature vector for region R_1 and R_2 respectively. To calculate the average feature vector for region R_1 and R_2, we will use integral image [14] approach for computational efficiency. Scaling is obtained by scaling the region R_2 instead of scaling the image. This results in the full resolution saliency map.
For an image of 'w' width, the width of region R_2 can be varied with in the range given by:

$$\frac{\mathbf{w}}{\mathbf{2}} \geq \mathbf{R_2} \geq \frac{\mathbf{w}}{\mathbf{8}}$$

For an image, filtering is performed at three different scales and final saliency map is obtained as the sum of saliency values across scales 'S':

$$\mathbf{S_{i,j}} = \sum_{\mathbf{S}} \mathbf{d_{i,j}}$$

Advantages

- Easy to implement.

- Generates high quality saliency maps of the same size and resolution as the input image. This is achieved by changing the filter size to achieve the change in scale rather than the original image.
- Method is effective on a wide range of images including those of paintings, video frames, and images containing noise.

Disadvantages

- Better highlight small salient region than larger one.
- Computational complexity is high.
- Precision, recall and f-measure are moderate.

3.1.2 Frequency Tuned Method

This algorithm achieves pre-attentive, low level, bottom-up saliency. This method is inspired by biological center surround concept but is not based on any biological model [5, 8]. This method is based on purely computational model [1, 6, 12].

The frequency tuned salient region algorithm [7] basically makes estimation regarding centre surround contrast using the features which are based on color and luminance value.

Algorithm:

Let w_l denotes the low frequency cut-off value and w_h denotes the high frequency cut-off value.

Criterion-

To highlight the salient region uniformly and to ensure the detection of larger salient object, the low frequency cut-off value should be low.

In order to achieve well defined boundaries and to retain the high frequency component from the original image, the high frequency cut-off value should be high.

To avoid coding artifacts, noise contamination and to remove texture pattern, high frequency component from the original image should be suppressed.

Thus, it can be said that the saliency map obtained from this method contains a wide range of frequencies, which is the combined output of several band pass filter with pass band range $[w_l , w_h]$ is appropriate.

Difference of Gaussian filter (DoG)

To achieve the saliency map in the desired pass band range, we choose the Difference of Gaussian filter for band pass filtering. Difference of Gaussian filter (DoG) closely approximate the Laplacian of Gaussian filter (LoG), therefore it is widely used in edge detection. This DoG filter is also used in saliency detection and interest point detection.
The Difference of Gaussian filter can be write as-

$$\mathbf{DoG(x,y) = \frac{1}{2\pi}\left(\frac{1}{\sigma_1^2} e^{-\frac{(x^2+y^2)}{2\sigma_1^2}} - \frac{1}{\sigma_2^2} e^{-\frac{(x^2+y^2)}{2\sigma_2^2}}\right)}$$

$$\mathbf{= G(x,y,\sigma_1) - G(x,y,\sigma_2)}$$

Where σ_1 and σ_2 are the standard deviation of Gaussian ($\sigma_1 > \sigma_2$).
A difference of Gaussian is a band pass filter whose pass band can be controlled by simply adjusting the ratio of standard deviation. If we combine several narrow band pass DoG filters with parameter as σ_1=ρσ and σ_2=σ then we define the summation over DoG in the ratio of standard deviation ρ as:

$$\sum_{n=0}^{N-1} \mathbf{G(x,y,\rho^{n+1}\sigma) - G(x,y,\rho^{n},\sigma)}$$

$$\mathbf{= G(x,y,\sigma\rho^{N}) - G(x,y,\sigma)}$$

Thus, for an integer N ≥ 0, which is simply the difference of two Gaussians (since all the terms except the first and last add up to zero) whose standard deviations can have any ratio K = ρ^N . That is, we can obtain the combined result of applying several band pass filters by choosing a DoG with a large K. If we assume that σ_1 and σ_2 are varied in such a way as to keep ρ constant at 1.6 (as needed for an ideal edge detector), then we essentially add up the output of several edge detectors (or selective band pass filters) at several image scales. This gives us a basic understanding of why the salient regions will be fully covered and not just highlighted on edges or in the center of the regions.

Selection of Parameter

Based on the argument, a strategic selection of σ_1 and σ_2 will provide an appropriate bandpass filter which is used to retain the desired spatial frequencies components from the original image while computing the saliency map. With $\sigma_1 > \sigma_2$, lower frequency cut-off value is determined by σ_1 and high frequency cut-off value is determined by σ_2. However, use of filters of a practical length, providing a correspondingly simple implementation, renders this approximation inaccurate.

To implement a large ratio in standard deviations, we drive σ_1 to infinity. This results in a notch in frequency at DC while retaining all other frequencies. To remove high frequency noise and textures, we use a small Gaussian Kernel keeping in mind the need for computational simplicity.

Estimation of Saliency Value:

Consider the image I which is having the width 'W' and height 'H', the saliency map 'S' for the Image is given by

$$\mathbf{S}(\mathbf{x}, \mathbf{y}) = |\mathbf{Im} - \mathbf{Iwhc}(\mathbf{x}, \mathbf{y})|$$

Where Im is the arithmetic mean pixel value of image and Iwhc is the Gaussian blurred version of the original image in order to remove noise, coding artifacts and fine texture details. The norm of difference is used as we are interested only in magnitude. This method is computationally quite efficient. Also we are operating over the full image without any down-sampling performed, so we obtain a saliency map of full resolution.
If we extend the above equation to use the feature of color and luminance, we can rewrite the equation as

$$\mathbf{S}(\mathbf{x}, \mathbf{y}) = \|\mathbf{Im} - \mathbf{Iwhc}(\mathbf{x}, \mathbf{y})\|$$

Where, Im is the mean image feature vector, Iwhc(x, y) is the corresponding image pixel vector value in the Gaussian blurred version (using a 5*5 separable binomial kernel) of the

original image, and ‖ ‖ is the L2 norm. Using the Lab color space, each pixel location is an [L, a, b] vector, and the L2 norm is the Euclidean distance.

Advantages:

- Uniformly highlight the salient region with well defined boundaries.
- The saliency maps generated by this method are of full resolution.
- The method is computationally efficient.

Disadvantages:

- In presence of complex background and large salient objects, it may fail to correctly identify the salient region.

3.1.3 Maximum Symmetric Surround Method

In this method [9], assumption regarding the scale of object with respect to image border is made, thus bandwidth of the center surround filtering near the image boarder is varied using symmetric surrounds concept.

Frequency tuned method treat the entire image as the common surround (abstracted as the average image CIELAB color vector) for any given pixel. The implicit premise is that in the absence of any knowledge of the scale of the salient object, it is best to pass all the low frequency content. We base our new saliency detection algorithm on the premise that we can make assumptions about the scale of the object of detection based on its position in the image.

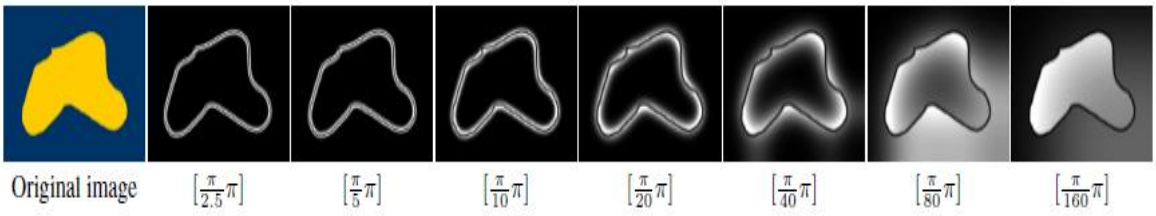


Fig.3.2 - Band-pass filtering output with increasing bandwidth

In Fig. 3.2 we note that the more central a pixel is within the salient object, the smaller has to be the low-frequency cut-off for detecting it. However, how central a pixel can be inside an object is limited by how far the pixel is from the boundary. That is, a pixel belonging to a salient object near the boundary will be less central inside the object. Therefore, assuming the salient object is fully within the image, and not cut-off by the image borders, we can afford to vary the bandwidth of the center-surround filter by increasing the low-frequency cut-off as we approach the image borders.

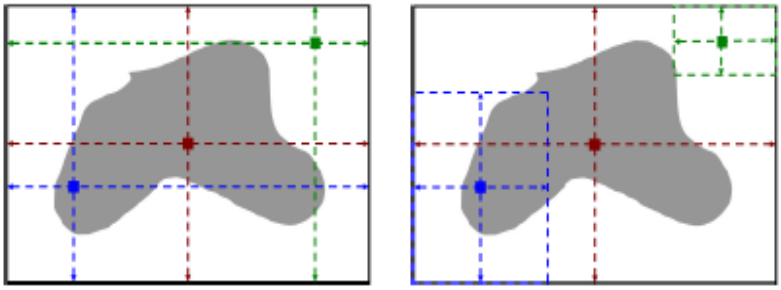


Fig. 3.3 (A) In Frequency tuned method for a pixel at the center (red) or elsewhere (blue), the surround regions used for computing saliency remains the same, namely the whole image area. (B) Maximum system surround uses surround regions (sub-images) that are symmetric w.r.t the pixel whose saliency needs to be computed.

In effect, as we approach the image borders we should use a more local surround region. We choose to do this by making the surround symmetric around the center with respect to the image borders as illustrated in Fig. 3.3 (B). This increases the low-frequency cut-off of the center-surround filter. By choosing a symmetric surround for each pixel (as the center), we implicitly treat each pixel to be at the center of its own sub-image (see Fig. 3.3 (B)). This is different from the method of Frequency tuned [7], where the entire image is used as the common global surround (abstracted as the average image CIELAB color vector) for any given pixel, resulting in an asymmetric surround for pixels that are not at the center of the image. This is explained graphically in Fig. 3.3 (A). Thus, for an input image of width 'w' and height 'h', the symmetric surround saliency value at the given pixel $S_{ss}(x, y)$ is obtained as:

$$\mathbf{S_{ss}(x, y) = \|I_m(x, y) - I_f(x, y)\|}$$

Where $I_m(x,y)$ is the average CIELAB vector of sub-image whose central pixel at position (x,y) is given by

$$\mathbf{I_m(x,y)} = \frac{\mathbf{1}}{\mathbf{A}} \sum_{\mathbf{i=x-x_o}}^{\mathbf{x+x_o}} \sum_{\mathbf{j=y-y_o}}^{\mathbf{y+y_o}} \mathbf{I(x,y)}$$

Where, offset x_0, y_0 and area A of sub-image computed as:

$$\mathbf{x_o = min(x, w\text{-}x)}$$
$$\mathbf{y_o = min(y, h\text{-}y)}$$
$$\mathbf{A = (2x_o + 1)\ (2y_o + 1)}$$

The sub-images obtained in the above equation are the maximum possible symmetric surround regions for a given pixel at the center. Consequently, the closer a pixel is to the edges, the narrower is its surround. To compute the CIELAB averages of these sub-images, we take the computationally efficient approach of using integral images as done by [6, 15].

Advantages:

- Narrowing the bandwidth near the image boarder causes the background to be less highlighted.
- Saliency map of full resolution.

Disadvantages:

- If the salient object is cut by image boarder i.e it is not completely inside the image, it is treated like the background and less likely be detected.

3.1.4 Compressed Domain Saliency Detection Method

This algorithm is based on the approach to detect saliency in compressed domain. Images are normally stored in compressed format like JPEG etc. This algorithm is design to extract feature for saliency detection from the coefficients of discrete cosine transform (DCT).
The discrete cosine transform (DCT) coefficients for a single block are shown in the figure Fig3.4.

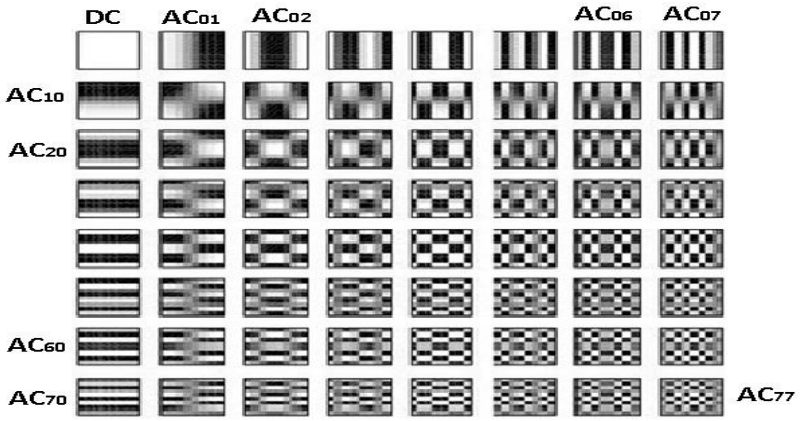


Fig3.4- DCT coefficients for one 8*8 block

DCT coefficients of the block contain both the AC coefficient and DC coefficients. DC coefficient is the measure of average energy of the block which is having 8*8 pixels while AC coefficients are used to measure the orientation information for the block. For a block of 8*8 pixels, it is having only one DC coefficient and 63 AC coefficients. To extract the color and intensity feature DC coefficients are used but to extract the orientation feature AC coefficients of the block are used.

When the JPEG image is in YCrCb color space, the DC coefficients give information related to the chrominance and luminance value. To extract information related to intensity and color feature, we have to transfer DC coefficients from YCrCb color space to RGB color space. Let R, G, B represent the red, green and blue color component from DC coefficients, then four broadly-tuned color channels can be generated as

r = R – (G + B) /2, **new red component;**

g = G – (R + B) /2, **new green component;**

b = B – (R + G) /2, **new blue component;**

y = (R + G) /2 – |R – B| /2 –B, **new yellow component;**

Intensity feature is given by

$$I = (R + G + B) /3;$$

Color features are denoted by C1 and C2, which are given as:

$$C1 = r - g;$$

$$C2 = b - y;$$

AC coefficients in YCrCb color space used to represent the orientation information for each block. In YCrCb color space, Cr and Cb component used to represent the color information and hence provide a little orientation information. Thus, we use the only the Y component to extract the orientation feature.

In the next step, we will obtain the feature map from the intensity, color and orientation feature calculated. Since the DC coefficients are used to calculate the color and intensity feature for every block, then the feature difference between the block i and block j can be calculated as:

$$d_{i,j}^{m} = f_i^m - f_j^m$$

Where m = 1, 2, 3 represent the intensity and color features respectively (one intensity feature and two color features).

Since AC coefficients are of the luminance component are used to calculate the orientation feature of image. Here a special distance known as Hausdorff distance [16] is used to calculate the difference between two AC Coefficient vectors from two different blocks. We are using the AC coefficients of first row and first column of luminance block which mathematically can be expressed as:

$$O = AC_{0k} \cup AC_{k0} \quad \text{(where } k = 1 \text{ to } 7\text{)}$$

The orientation feature difference for two blocks i and j is given by:

$$d_{i,j}^{4} = \max\left(H(O_i, O_j) , H(O_j, O_i) \right)$$

Where O_i and O_j represents the AC coefficients of block i and block j respectively and $H(O_i, O_j)$ can be calculated as

$$\mathbf{H(O_i, O_j) = \| o_i - o_j \|}$$

Where $o_i = \max(O_i)$, $o_j = \min(O_j)$ and $\|.\|$ is L2 norm.

The concept of visual sensitivity is used to determine the weight of the above calculated feature differences. The contrast sensitivity [17] as a function of eccentricity is given by:

$$\mathbf{CS(f,e) = 1 /(C_{so} \exp(a.f(e+e_1)/e_1)}$$

Where 'f' denotes the spatial frequency (cycles/degree), 'e' denotes the retinal eccentricity in degree between block i and j, 'C_{so}' denotes the minimum contrast threshold, 'a' denotes the spatial frequency decay constant, 'e_1' denotes the half resolution eccentricity.

Let $a_{i,j}$ = CS(f,e) is to represent weight for differences between blocks. From the above equation it can be concluded that the weighting factor is dependent on the retinal eccentricity between the block i and j. higher the value of retinal eccentricity, lower the value of weighting factor which cause the less contribution from $d_{i,j}$ in the final saliency value.
The saliency value for feature m for each block B_i^m is given by differences between block i and others in image and weighting factors for these differences by contrast sensitivities.

$$\mathbf{B_i^m = \sum_{j \neq i} a_{i,j}\ d_{i,j}^m}$$

Where B_i^m denotes the saliency value of feature 'm' for block i, $d_{i,j}^m$ is the difference of feature m between block i and j and $a_{i,j}$ is the weighting for block differences.

The final saliency map can be obtained by integrating the four feature maps B^m using coherent normalization based fusion method. The final saliency map is given as:

$$S = \sum \Upsilon\theta \; N(\theta) + \prod \rho \, \theta \, N(\theta)$$

Where N is the normalization operator, ϒ and ρ are parameter determining the weight for each component.

Advantages

- Saliency map can be used for image retargeting.
- Better precision, recall and f-measure value are obtained.

Disadvantages

- Object boundaries are not properly defined.
- Saliency map perceptual quality is very poor.
- Computational cost is high.
- Resolution map is not at full scale.

3.1.5 Non Parametric Low level Vision Method

This is a computational model which is a combination of simple, neurally-plausible mechanism which removes all the unnecessary arbitrary variables.

At first stage, the image is convolved with a bank of filters using multi-resolution wavelet transformation technique. This result in a spatial pyramid which is having wavelet planes oriented either vertically, horizontally or diagonally. The coefficients which are obtained from the wavelet transform and represent the spatial pyramid may be considered as the estimation of local oriented contrast.

Wavelet Transform for an image I can be expressed as:

$$WT\{Ic\} = \{w_{s,o}\} \quad s=1,2,\ldots,l \; ; \; o=h, v, d$$

Where Ic represent the one opponent channel of image I and WT represents the wavelet transform taken over an image opponent channel. $w_{s,o}$ denotes the wavelet planes at spatial scale ‘s’ and orientation ‘o’. ‘l’ denotes the number of scales used in the decomposition. As the transformation is using gabor-like basis function therefore value of ‘l’ is given by

$l=\log_2 M$, where M is largest dimension of the image. The wavelet coefficient which is centered at position (x,y) is denoted by $w_{x,y}$.

At second stage, it is required to calculate the contrast energy around wavelet coefficients. For the estimation of contrast energy it is required to convolve the local region around the wavelet coefficient with some binary filter 't'. The filter should be chosen in such a way that its shape should vary in accordance with the orientation of wavelet plane. Contrast energy around a wavelet coefficient $w_{x,y}$ is denoted by $e_{x,y}$. Thus, for horizontal wavelet plane contrast energy can be computed as:

$$\mathbf{e_{x,y} = \sum_j \ w_{x-j,y} \ 2t_j}$$

Where t_j denotes the j-th coefficients of one dimensional filter t. The contrast energy for coefficients is computed at each spatial scale and each spatial location. The center surround effect produced due to the interaction of center region and surround regions. To model center-surround effect, the energies of center region $e^c_{x,y}$ and surround region $e^s_{x,y}$ are compared using

$$\mathbf{c_{x,y} = (e^c_{x,y})^2 / (e^s_{x,y})^2}$$

The energies of surrounding regions are calculated in the same manner as that of central region however the only difference lies in the definition of the binary filter used. To generate the final center surround energy measure, a non linear scaling of $c_{x,y}$ is performed which is given by

$$\mathbf{d_{x,y} = c^2_{x,y} / (1+c^2_{x,y}) ;} \qquad \mathbf{d_{x,y} \in [0, 1]}$$

When $d_{x,y} \to 0$, represents central activity is much lower than surrounding activity. Similarly when $d_{x,y} \to 1$, represents central activity is much more than surrounding activity. The size of center and surround region are used to define the size of corresponding filter.

It is well known that color appearance in an image is dependent on the spatial frequency. Mullen [19] described the human sensitivity to local contrast in color opponent channel with a Contrast Sensitivity Function(CSF). The above idea is extended to extended contrast

sensitivity function which is parameterized by spatial scale and normalized center-surround contrast energy.

The ECSF is given by

$$\mathbf{ECSF(d,s) = d \,.\, f(s) + q(s)}$$

Where f(s) is given by

$$\mathbf{f(s)} = \begin{cases} \mathbf{a e}^{-\frac{(s-k)^2}{2\sigma_1^2}} & \mathbf{s \le k} \\ \mathbf{a e}^{-\frac{(s-k)^2}{2\sigma_2^2}} & \text{otherwise} \end{cases}$$

in the above equation, s denotes the spatial scale of wavelet plane, a is the scaling constant. σ_1 and σ_2 represents spread of spatial sensitivity of f(s). The parameter k defines the peak spatial scale sensitivity of f(s). This shows that normalized contrast center surround activity 'd' of wavelet coefficients is modulated by f(s).A function q(s) is also added to ensure the non-zero lower bound on ECSF which is given by:

$$\mathbf{q(s)} = \begin{cases} \boldsymbol{e}^{-\frac{(s-k1)^2}{2\sigma_3^2}} & \boldsymbol{s \le k1} \\ \mathbf{1} & \boldsymbol{otherwise} \end{cases}$$

σ_3 represents the spread of spatial sensitivity of q(s).

Thus it can be concluded that the extended contrast sensitivity function (ECSF) is the weighted function of normalized contrast centre-surround energy $d_{x,y}$ at a location, which produced the final response as:

$$\mathbf{p_{x,y} = ECSF(d_{x,y}, s_{x,y})}$$

The weight $p_{x,y}$ is used to modulate the wavelet coefficients $\mathbf{w_{x,y}}$ **.** The output image channel that contains the color appearance illusions are obtained by performing the inverse wavelet transform on the wavelet coefficients $\mathbf{w_{x,y}}$ at each location, scale and orientation multiplied by the weighted response $\mathbf{p_{x,y}}$ is given by

$$\textbf{Ico(x,y) = Inverse WT } \{p_{x,y,s,o} * w_{x,y,s,o}\}$$

Where Ico(x,y) represent the output image channel which is modified from original image by 'p' weight. The weight assigned to each wavelet coefficient cause the output image to go under blurring effect or enhancing effect. The color of modified locations have either been averaged to surrounding color or sharpen to be less similar to surrounding. Thus we can say, those image locations that undergo enhancement are salient while those locations that undergo blurring are non salient.

By neglecting the image wavelet coefficients and performing the inverse wavelet transform only on the weight computed at each spatial and orientation location, provide a direct method to estimate the saliency for each opponent channel.

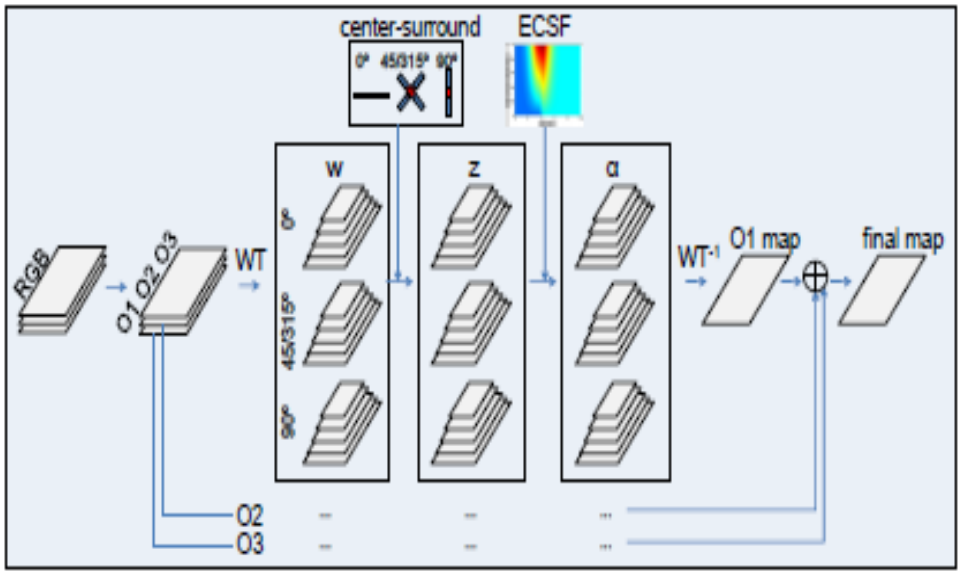


Fig3.5- Block diagram of non parametric low-level vision method

Saliency map for an image channel Ic at location (x,y) is estimated as

$$\textbf{SALMAP(x,y) = Inverse WT } \{ p_{x,y,s,o} \}$$

In order to calculate the final saliency map SM, saliency map for each opponent channel is calculated and Euclidean norm is used to calculate the final saliency map, which is given by

$$\mathbf{SM} = \sqrt{\boldsymbol{S1^2 + S2^2 + S3^2}}$$

Where S1, S2, S3 denote saliency map of each opponent channel and SM is the final saliency map.

Advantages:

- Integration of multi-scale information by the use of wavelet Transform.
- Use of Extended Contrast Sensitivity Function with biologically justifiable parameter.
- Less sensitive to the low frequency edges in image.

Disadvantages:

- Strength of centre-surround contrast energy is highly dependable on size of binary filter.
- The model performs best only when mid range frequencies are enhanced and low or high freq are inhibited.
- This algorithm used to detect many background pixels as it does not consider any global feature.

3.1.6 Context Aware Saliency Detection Method

This algorithm is designed to detect the salient objects together with part of the unique background which surrounds them. This algorithm basically focuses on both global and local feature. It follows the principles of

1. Local low level consideration, which include the features of contrast and color.
2. Global consideration ,to pop out the frequently occurring feature, and maintain those which deviate from norm
3. Visual consideration, which determine center of gravity about which structured is to be formed.
4. High level consideration, include feature regarding salient object position.

Firstly, we will define the single-scale local-global saliency based on the principle 1 to 3. Then we extend the approach to multiple scales and at last principle 4 is used as post processing step.

Single-Scale Local-Global Saliency

Consider a single patch of scale s at each pixel. Thus a pixel i is considered to be salient, if the appearance of patch p_i centered at pixel i is distinctive with respect to all other image patches.

Let $d_{col}(p_i, p_j)$ be the euclidean distance between patch p_i and p_j.The image is in the format of CIE L*a*b color space is converted from RGB color space. The Euclidean distance is normalized in the range [0, 1]. For pixel i to be salient, the normalized Euclidean color difference should be high for all values of j.

To examine the patch uniqueness, instead of calculating its dissimilarity with all other image patch, we would like to consider K most similar patches by $d_{col}(p_i, p_j)$. If the most similar patches are highly different from the reference patch p_i , then automatically all the other patches will also be highly different and hence the pixel i is considered as salient pixel. In our experiment the value of most similar patches K is chosen to be 64.

According to principle 3, positional distance between the patches is an important factor to determine the patch saliency. All the background similar patches are spread near and far away in the image while salient patches tend to be group together.

Let $d_{pos}(p_i, p_j)$ be the Euclidean distance between the position of patch p_i and p_j. The dissimilarity measure index is given by

$$\mathbf{d(p_i, p_j)} = \frac{\mathbf{d_{col}(p_i, p_j)}}{\mathbf{1 + c.d_{pos}(p_i, p_j)}}$$

where c is a constant ,Dissimilarity measure index is directly proportional to color difference and inversely proportional to positional distance. For a pixel i to be salient, $d(p_i, p_j)$ should be high for all j∈ [1,K].

The single scale saliency value at scale s for pixel i is given by

$$\mathbf{S_i^s = 1 - exp\{-\frac{1}{K}\sum_{j=1}^{K} d(p_i^s, p_j^s)\}}$$

Multi-scale Saliency Enhancement

Multi-scale are used to decrease the saliency of the background pixel and increase the saliency of foreground pixel. A pixel is considered as salient if it is consistently different from all other patches at multiple scales.

The single scale saliency value at scale s for pixel i is given by

$$\mathbf{S_i^s = 1 - exp\{-\frac{1}{K}\sum_{j=1}^{K} d(p_i^s, p_j^s)\}}$$

The Saliency map at each scale is normalized in the range [0, 1] and interpolated back to normal image size.

Let R = $\{s_{1,}s_{2,\ldots,}s_M\}$ denotes the patch size to be consider for pixel i. The saliency value at pixel i is the mean of saliency values taken at different scale.

$$\mathbf{\bar{S}_i = \frac{1}{M}\sum_{s\in R} S_i^S}$$

Larger the value of $\bar{S}_i$, pixel is more salient.

Including the Immediate Context

This suggests that areas that are closed to foci of attention should be explored more significantly then other far away region when region surrounding the object convey the background context, it is assumed to be salient.

Let $d_{foci}^{s}(i)$ be the Euclidean positional distance between pixel i and closest focus of attention pixel at scale s. The redefine pixel saliency is given by

$$\mathbf{\hat{S}_i = \frac{1}{M}\sum_{s\in R} S_i^S\,(1 - d_{foci}^{s}(i))}$$

This shows that saliency of interested background will be increased by the equation.

Including the center prior feature

In the daily life it is experienced that pixels near the centre of image captures the more human visual attention than the pixels which are near the boarder of the image. Whenever a picture is taken, it is normal consideration to put salient object to the centre of image. This shows that location which are near to the centre of image are more likely to be salient than those region which are far from centre. This feature can be modeled using the Gaussian map. Let $G(\sigma_x,\sigma_y)$ be a 2D Gaussian positioned at the center of the image. Our final saliency of a pixel is defined as:

$$\mathbf{S_i = \widehat{S_I} G_i}$$

Where G_i is the value of pixel i in the map G.

Advantages

- This method incorporates both local and global features.
- This algorithm also detect context of image.

Disadvantage

- Time complexity is very high.
- It does not uniformly highlight the salient Region.

3.1.7 Spectral Residual Approach Method

Saliency detection using spectral residual approach is basically a frequency domain based algorithm which used to detect the salient region more efficiently and very fast as compared to other algorithm. A basic principle to detect the visual saliency is suppressed the frequently occurring feature response while maintaining those feature which deviate from the norm. This algorithm is used to calculate global feature i.e frequency feature instead of calculating the local features like color, contrast, orientation etc. The aim of the algorithm is to minimize the most redundant information and to preserve the non redundant information. The steps of the algorithm are as follow.

For a given image, down-sampled the input image to a size related to the visual scale. The selection of the visual scale will affect the visual saliency. Changing the visual scale will change the resulting saliency region.

Transform the down-sampled image from spatial domain to the frequency domain and log spectrum is computed for the down-sampled image. Let I(x) represent the down-sampled image. Transform of the image is represented by FF

$$\mathbf{FF = fft2\ (I(x))}$$

Log spectrum of image is given by

$$\mathbf{LS(f) = log\ (abs\ (FF))}$$

Phase spectrum of the image can be given by

$$\mathbf{PS(f) = angle\ (FF)}$$

We know that, statistical singularities in spectrum are responsible for the salient region in the images. In order to calculate these spectral singularities, we have to define the spectral residue.

Spectral residue is defined as:

$$\mathbf{SR(f) = LS(f) - AS(f)}$$

Where AS(f) denotes the general shape of the log spectrum or average spectrum. AS(f) can be obtained by convolving the input image with filter h(f)

$$\mathbf{AS(f) = h(f) * LS(f)}$$

Where h(f) is a n * n matrices of ones.

Graphically SR(f), LS(f) and AS(f) can be shown as

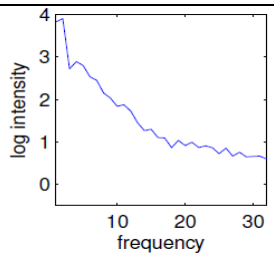
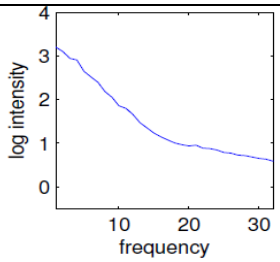
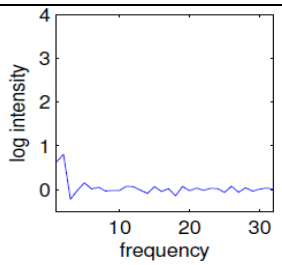
Input image	LS(f)	AS(f)	SR(f)
			

Fig 3.6 - Log spectrum, average spectrum and spectral residual for an input image

The effect of filter size h(f) to the resulting saliency map can be shown in the figure

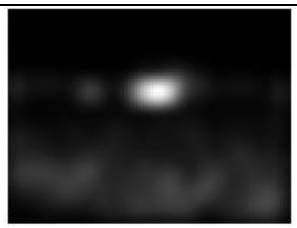
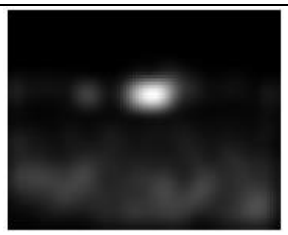
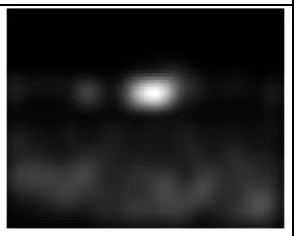
Input image	3 * 3 filter	5 * 5 filter	7 * 7 filter
			

Fig 3.7 - Spectral residual saliency map for different filter size h(f)

In this algorithm, the spectral residual obtained represents the saliency part of the image. Spectral residual are also interpreted as the unexpected portion of image. To construct the saliency map in spatial domain, inverse Fourier transform of the spectral residual is taken.

Saliency Map = abs [Ifft (exp(SR(f) + PS(f)))^2]

In order to obtain the saliency map with better visual effect, the saliency map is smooth by a Gaussian filter.

Advantages:

- It is based on purely computational model, no prior knowledge is required for saliency detection.
- This is very fast algorithm, it can be used for real time application.

- Very simple implementation.

Disadvantage:

- Saliency map obtained are not of full resolution.
- Do not detect object boundaries very accurately.
- Do not include any local feature so many object pixels are counted as background pixels.

Chapter 4

Proposed Technique

This algorithm is designed to generate saliency map with higher precision and with very low computational cost using hybrid features. Features of color, position, luminance and frequency are calculated to generate the individual saliency map. Finally, these individual saliency maps are combined to generate the final saliency map. The basic block diagram of proposed technique is shown in the figure Fig.4.1. In the following section, first of all we summarize the concept of RGB to CIE Lab color space conversion then we discuss the proposed technique.

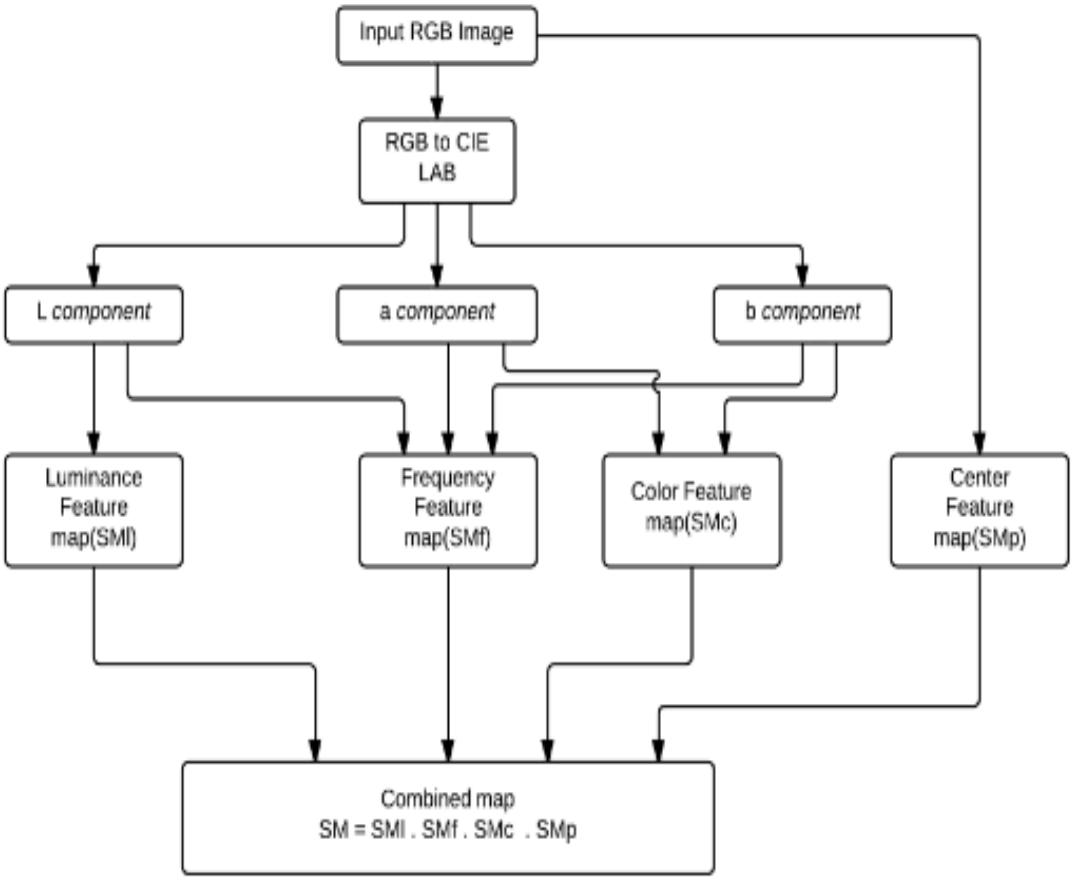


Fig. 4.1 – Flow Chart of proposed algorithm

4.1 RGB to CIE Lab color space conversion

The major problem that exists with the RGB color space is the colorimetric distances between the individual colors do not correspond to perceived color differences. This is due to the fact that human eye is not equally sensitive to these three colors. Therefore it is of particular interest for a perceptually uniform color space where a small perturbation in a component value is approximately equally perceptible across the range of that value [19]. The CIE solved the problem with the development of the three-dimensional Lab color space. The axis 'a' extends from green (-a) to red (+a) while the axis 'b' from blue (-b) to yellow (+b). The brightness (L) increases from the bottom to the top of the three-dimensional model as shown in figure Fig. 4.2.

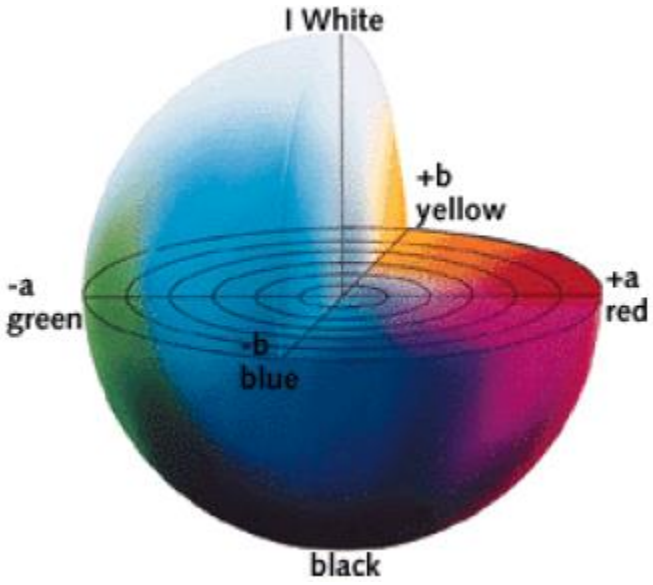


Fig 4.2 – CIE Lab color space

RGB model output of physical device while Lab matches human perception of color. Lab color space separates the luminance of image from the color component of the image. It is perceptually uniform i.e a change of same amount of color value will produce the same amount of visual importance.

To obtain the LAB image from the RGB image, we first convert RGB image to XYZ color space and then XYZ color space is converted to LAB color space.

The linear RGB values in the range [0, 1] can be converted to the corresponding CIE XYZ values in the range [0, 1] using the following matrix transformation:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4125 & 0.3576 & 0.1804 \\ 0.2127 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9502 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Then Lab color space is calculated from XYZ as below:

$$L^* = 116 \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16$$

$$a^* = 500 \left[\left(\frac{X}{X_n}\right)^{\frac{1}{3}} - \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} \right]$$

$$b^* = 500 \left[\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - \left(\frac{Z}{Z_n}\right)^{\frac{1}{3}} \right]$$

With the constraint that $\frac{X}{X_n}, \frac{Y}{Y_n}, \frac{Z}{Z_n} > 0.01$. This constraint is satisfied for most practical practices hence complete formula as given in [20], can be ignored. X_n, Y_n, Z_n are the XYZ parameters of white reference point.

4.2 Frequency Feature Calculation

Salient object detection mechanism which is based on the behavior that human visual system detects the visual scene may be approximated by integrating the band pass filter response obtained for each opponent channel such as CIE Lab color channel. This is inspired by the frequency tuned method of Achanta et.al [7] where he uses Difference of Gaussian filter for band pass filtering. But in our method, Log- Gabor filter is used instead of the Difference of Gaussian filter for frequency feature calculation. There are few advantages of Log-Gabor filter.

- Log-Gabor filter can be constructed with any arbitrary bandwidth.
- This filter does not have any DC component.
- The transfer function is having extended tail at high frequencies which makes it more suitable to encode images better than other ordinary band pass filters [21, 22].

4.2.1 Log-Gabor filter

Gabor filters are used to obtain localized frequency information. These filters offer the best simultaneous localization of spatial and frequency information. But this filter has two main drawbacks. The maximum bandwidth of a Gabor filter is limited to approximately one octave and Gabor filters are not optimal if one is seeking broad spectral information with maximal spatial localization.

An alternative to the Gabor function is the Log-Gabor function proposed by Field [1987]. Log-Gabor filters can be constructed with any arbitrary bandwidth and the bandwidth can be optimized to produce a filter with minimal spatial extent. Field suggests that natural images are better coded by filters that have Gaussian transfer functions when viewed on the *logarithmic* frequency scale. (Gabor functions have Gaussian transfer functions when viewed on the *linear* frequency scale). On the linear frequency scale the log-Gabor function has a transfer function of the form

$$LG(w) = \exp\{-\log(\frac{w}{w_0})^2 / 2\log(\frac{c}{w_0})^2\}$$

Where, w_o is the filter's centre frequency. To obtain constant shape ratio filters the term c/w_o must also be held constant for varying w_o. For example, a c/w_o value of 0.74 will produce a filter bandwidth of approximately one octave, 0.55 will produce in two octaves, and 0.41 will produce three octaves.

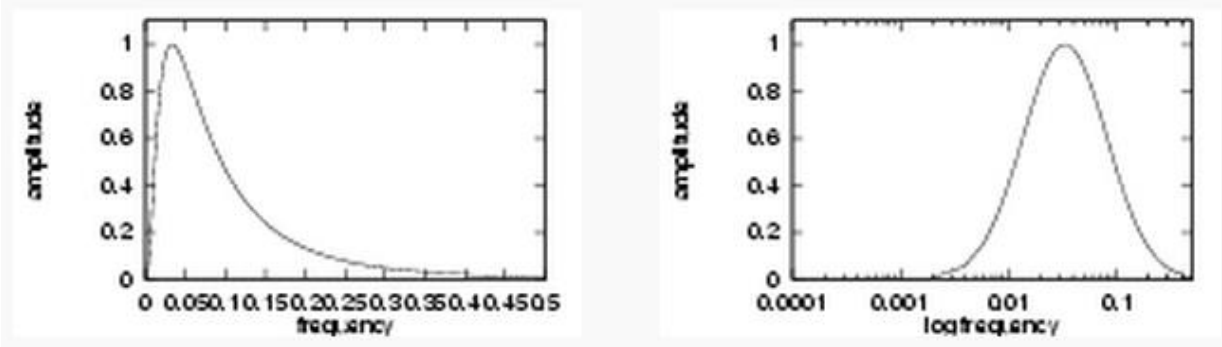


Fig. 4.3 Frequency response of log Gabor filter

Field's studies of the statistics of natural images indicate that natural images have amplitude spectra that fall off at approximately 1/w. To encode images having such spectral characteristics one should use filters having spectra that are similar. Field suggests that log Gabor functions, having extended tails, should be able to encode natural images more efficiently than, ordinary Gabor functions, which would over-represent the low frequency components and under-represent the high frequency components in any encoding. Another point in support of the log Gabor function is that it is consistent with measurements on mammalian visual systems which indicate we have cell responses that are symmetric on the log frequency scale.

The transfer function of Log-Gabor filter in frequency domain can also be represented as

$$\mathbf{LG(w) = exp\ \{\ --log(\frac{w}{w_0})^2\ /\ 2\sigma_1^2\}}$$

Where w = {w1,w2} represents the co-ordinate in frequency domain, w_0 represents the centre frequency of filter and σ_1^2 represent the bandwidth controlling parameter.

Spatial domain representation lg(x) of the filter cannot be obtained analytically because singularity exists at the origin due to logarithmic function. However, approximation of lg(x) can be obtained by taking the inverse Fourier transform of LG(w).

Let the image I(x) is a RGB image. The image is first converted to the CIE Lab space. Let $I_l(x), I_a(x)$ and $I_b(x)$ represent the three opponent channels, then frequency saliency map is given by

$$\mathbf{SMf(x)\ = \{(I_l * lg)^2\ +\ (I_a * lg)^2\ +\ (I_b * lg)^2\}^{\frac{1}{2}}\ (x)}$$

Where * denotes the convolution operation.

4.3 Color Feature Calculation

In daily life, it is noted that warm colors such as red and yellow capture more visual attention than cold colors like blue and green. This method is used to calculate color feature saliency based on the model of warmness and coldness of color

In CIE Lab color space, 'L' denotes the luminance component, 'a' channel denotes the green to red color information while 'b' channel denotes the blue to yellow color information. Higher (Lower) pixel value in a channel denotes that the color is more likely to be red (green). Similarly, higher (lower) pixel value in 'b' channel denotes that color is more likely to be yellow (blue).Thus, our saliency map will generate the higher saliency value for warm color than cold color.

To obtained the color saliency, linear mapping is performed such as

$$\boldsymbol{I_a(x) :\rightarrow I_{an}(x) \in [0,1] \quad \text{and} \quad I_b(x) :\rightarrow I_{bn}(x) \in [0,1]}$$

$$\boldsymbol{I_{an}(x) = \frac{I_a(x) - mina}{maxa - mina};}$$

$$\boldsymbol{I_{bn}(x) = \frac{I_b(x) - minb}{maxb - minb};}$$

Where mina(maxa) denotes the minimum(maximum) value of $I_a(x)$. Where minb(maxb) denotes the minimum(maximum) value of $I_b(x)$. In a color plane, the point $\{I_{an}(x)=0, I_{an}(x)=0\}$ denotes the coldest point and hence least significant. The color saliency at point x is define as

$$\boldsymbol{SMc(x) = 1 - \exp\left(-\frac{I_{an}^2(x) + I_{bn}^2(x)}{\sigma_2^2}\right)}$$

4.4 Centre feature Calculation

In the daily life it is experienced that pixels near the centre of image captures the more human visual attention than the pixels which are near the boarder of the image. Whenever a picture is taken, it is normal consideration to put salient object to the centre of image. This

shows that location which are near to the centre of image are more likely to be salient than those region which are far from centre. This feature can be modeled using the Gaussian map. The location saliency at x can be expressed using Gaussian map

$$\mathbf{SMp(x) = exp(-\frac{\|x-c\|^2}{\sigma_3^2})}$$

Where 'c' represent the center of image and σ_3 is a parameter.

4.5 Luminance feature Calculation

This feature calculation is based on the fact that salient object in an image occupy area less than the 50% of the image area. To obtained the luminance saliency, linear mapping is performed such as

$$\mathbf{I_l(x) :\rightarrow I_{ln}(x) \in [0, 1]}$$

$$\mathbf{I_{ln}(x) = \frac{I_l(x) - meanl}{maxl - minl};}$$

Where mina(maxa) denotes the minimum(maximum) value of I_l(x) . meanl denotes the average pixel value over image. The Luminance saliency at point x is define as

$$\mathbf{SMl(x) = 1 - exp(-\frac{I_{ln}^2(x)}{\sigma_4^2})}$$

Where, σ_4 is a parameter.

4.6 Combining Features

The final saliency map can be computed by combining the individual feature saliency map which is given by;

SM(x) = SMf(x) . SMc(x) . SMp(x) . SMl(x)

Where, SM(x) denotes the final saliency map.

Chapter 5

Simulation and Results

5.1 Setup parameters

Algorithms are developed in MATLAB to detect salient region from images. MATLAB is used because of large number of advanced inbuilt functions and image processing toolbox. The setup parameters are as shown in table 5.1.

Table 5.1 – Setup Parameters

Image Database	SED 1 (100 images)
Image type	.jpg
Image Format	RGB
Processor	Core i5 @ 2.5 Ghz , 4GB RAM
Simulation Tool	Matlab R2008a

5.2 Qualitative Evolution

To compare the result of our algorithm with other saliency detection method, These algorithms are run over SED 1 [24] database having 100 color images of single object and where ground truth is obtained by asking people to select the salient regions where the object is present. For qualitative evolution few mages are selected from database to show the saliency maps of different algorithms.

Fig. 5.1- Input Images

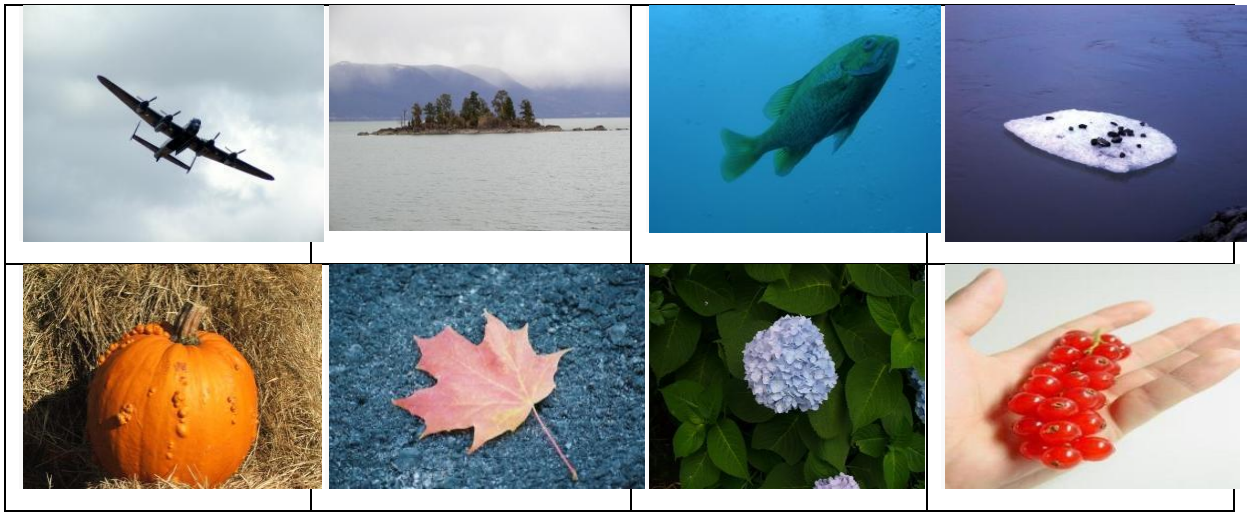


Fig. 5.2 - Ground truth

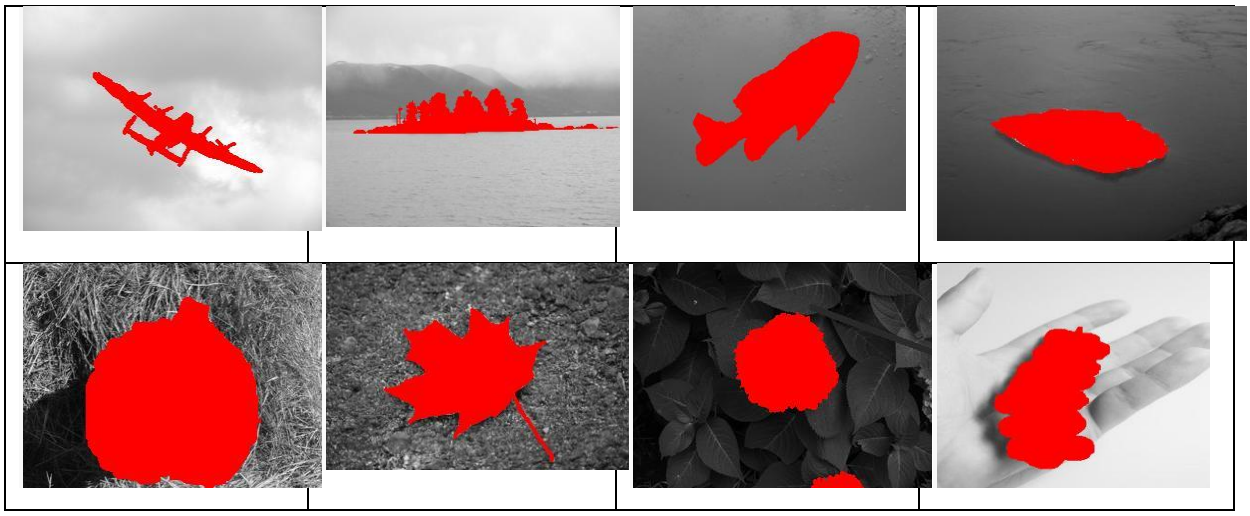


Fig. 5.3- Achanta method saliency map

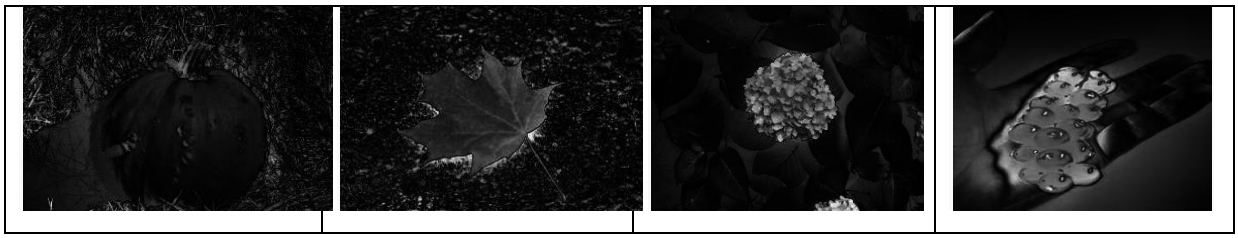


Fig. 5.4- Frequency tuned method saliency map

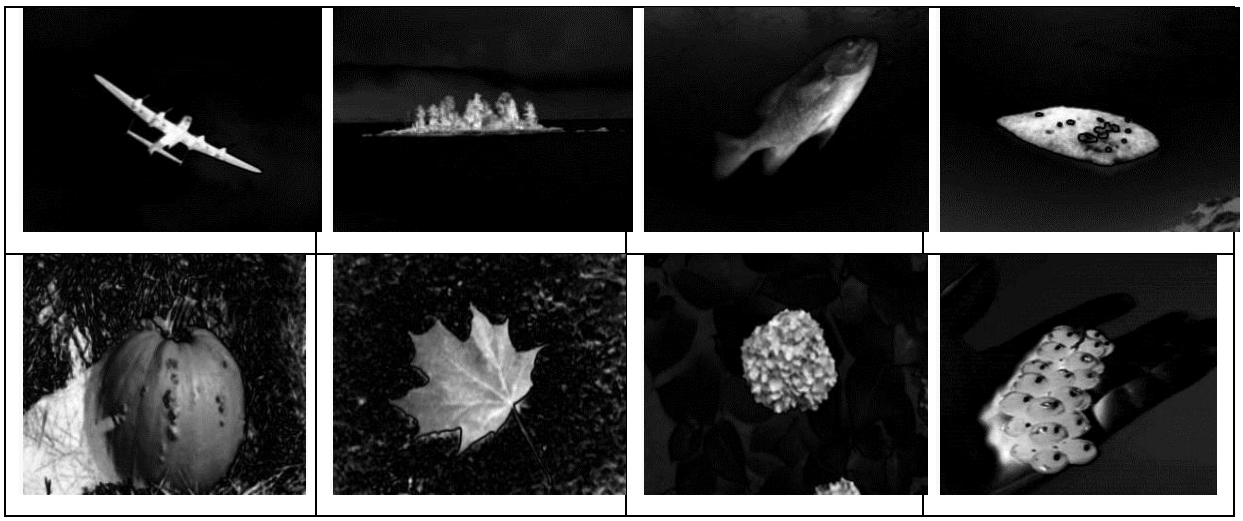


Fig. 5.5 - Maximum system surround method saliency map

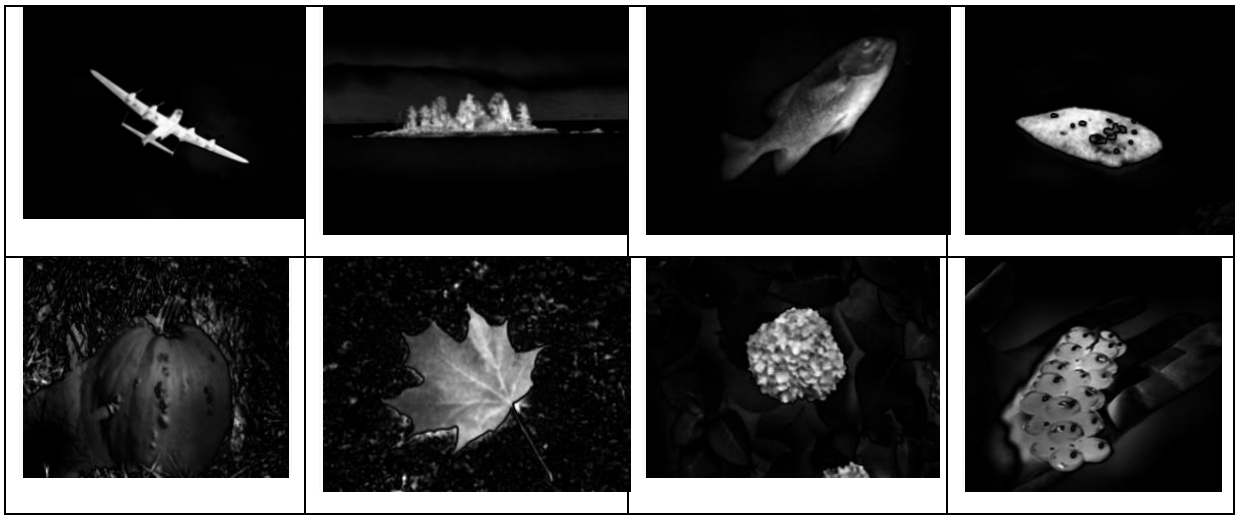


Fig. 5.6 -Compressed domain method saliency map

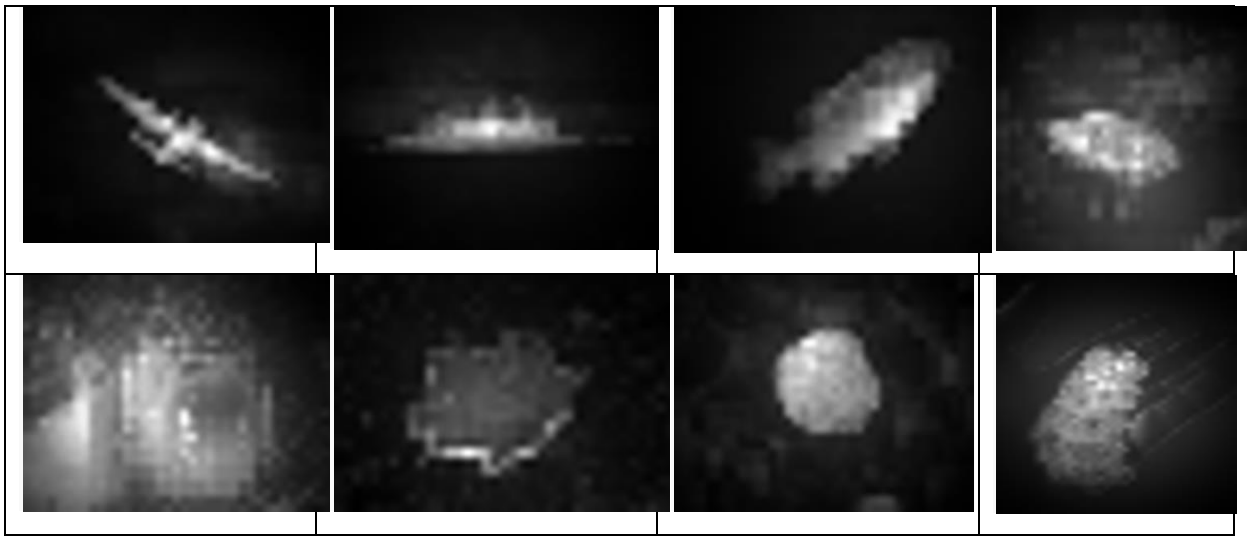


Fig. 5.7-Non Parametric low-level vision method saliency map

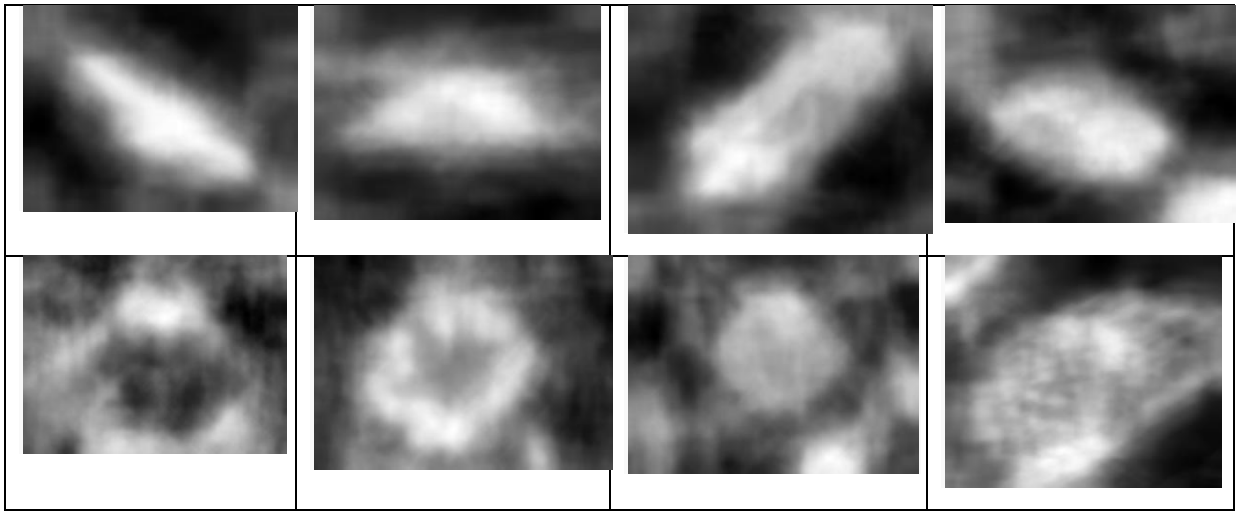
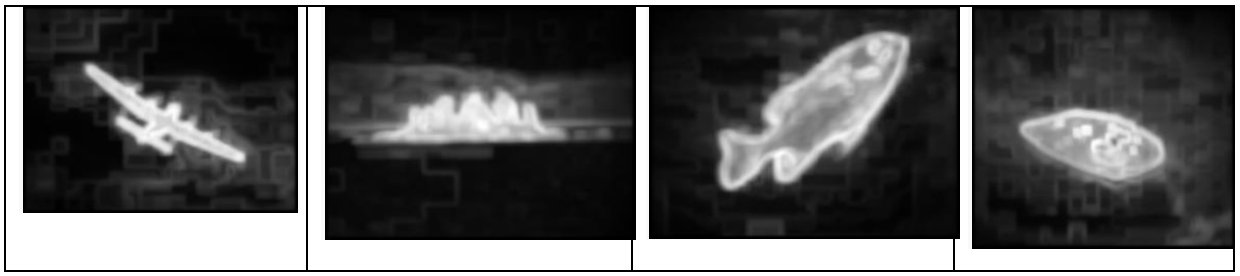


Fig.5.8-Context aware method saliency map



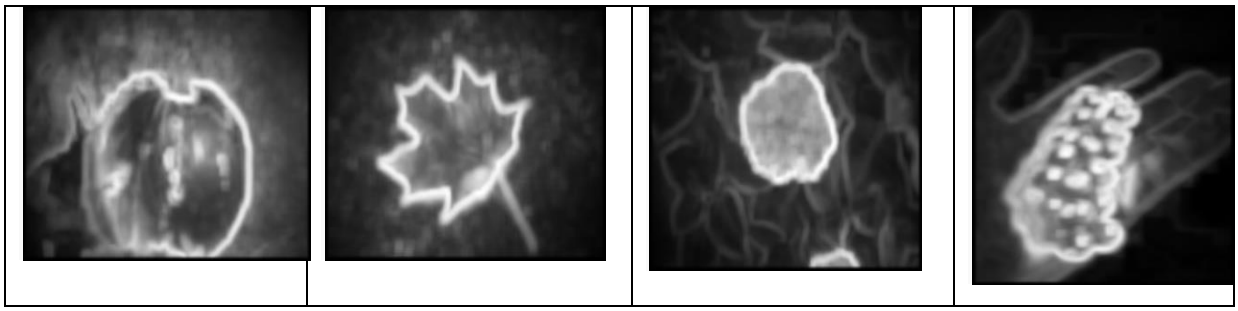


Fig.5.9- Spectral residual method saliency map

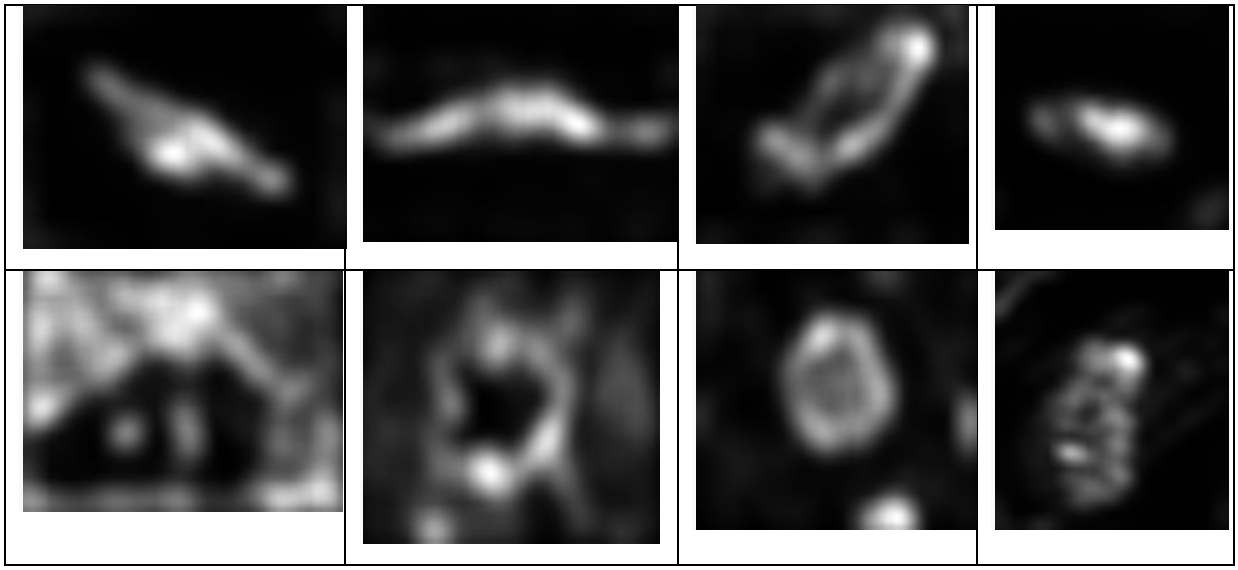
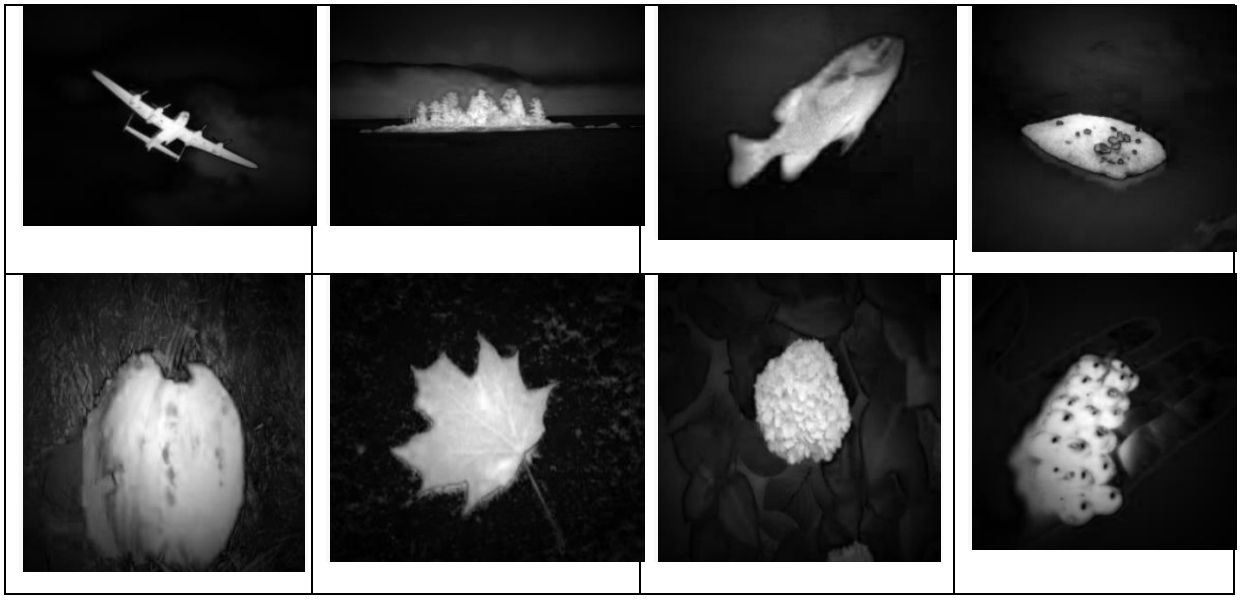


Fig.5.10- Our method saliency map



5.3 Quantitative Evolution

To obtain a quantitative evolution of salient region detection algorithm, Results are compared on the basis of Precision, Recall and F-measure value (As explained in chapter 1). We compare our method with seven other states of art methods on the common available data set SED1 which contains 100 images with ground truth. The following two ways are adopted for quantitative evolution.

5.3.1 Fixed Threshold Segmentation

The evolution of the performance of saliency detection algorithm is to be done with reference to salient object segmentation. For a given saliency map which have values in the range [0 255], the salient object is threshold at T_o, where T_o, varies from 0 to 255. With this, various precision-recall pairs are obtained and a precision recall curve for an image can be drawn. Precision-recall curve for an image database can be calculated by averaging the precision-recall curve for all the images. The comparison of our algorithm with other states of art algorithm is shown in the figure.

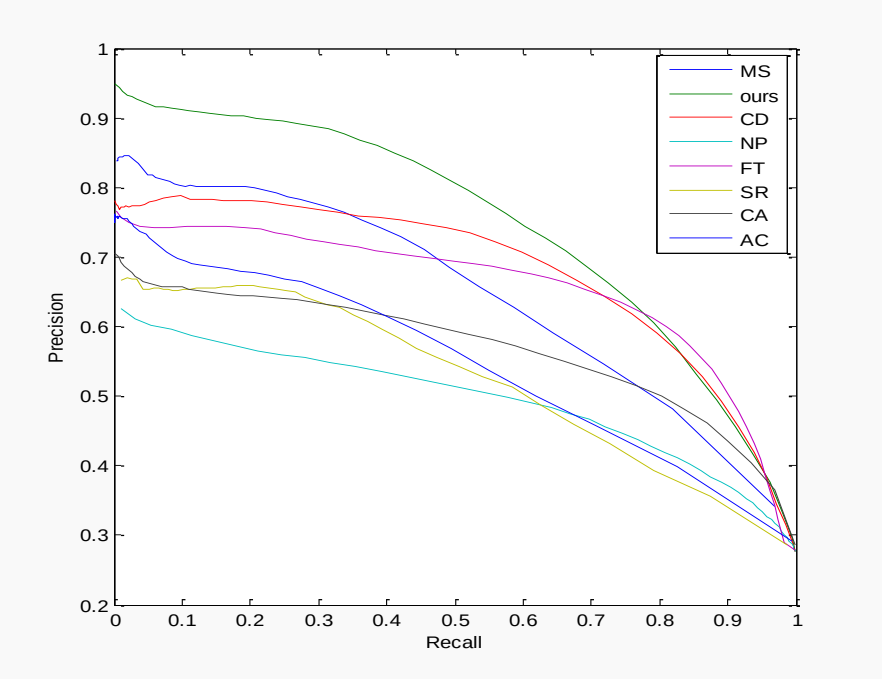


Fig. 5.1- Fixed threshold based comparison of precision, recall & F-measure

5.3.2 Adaptive Threshold Segmentation

In this, an image dependent adaptive threshold is used to segment the object for the image itself . For an image saliency map, values lie in the range 0 to 255 and the segmentation is done with the adaptive threshold that can be obtained from the matlab command graythresh. Using, this threshold, we can binarized the saliency map to extract the salient object from the image. Precision, Recall, F-measure values for an image are calculated by adaptive thresholding. Finally, Precision, Recall and F-measure value for image database can be calculated by averaging these values for all the images. For each image F-measure can be calculated as

$$\textbf{F-measure} = \frac{(1+ b^2).\textbf{Precision} .\textbf{Recall}}{b^2.\textbf{Precision}+\textbf{Recall}}$$

For our experiment b^2 is set to 0.5. The average F-measure predicts the overall saliency detection accuracy of an algorithm (as explained in chapter 1). The Precision, Recall and F-measure value for different algorithm are shown below in the table

Table 5.2 – Precision, Recall, F-measure

Method	Precision	Recall	F-measure
Achanta (AC)	0.6594	0.3076	0.4774
Frequency Tuned(FT)	0.6016	0.3473	0.4874
Maximum system Surround (MS)	0.7572	0.3526	0.5477
Compressed Domain(CD)	0.7029	0.6451	0.6825
Non parametric (NP)	0.4215	0.8148	0.5024
Context aware (CA)	0.6234	0.7230	0.653
Spectral residual (SR)	0.5843	0.553	0.5735
Ours	0.7662	0.5766	0.6945

These Precision, Recall and F-measure can be shown graphically for different algorithm as shown in the figure

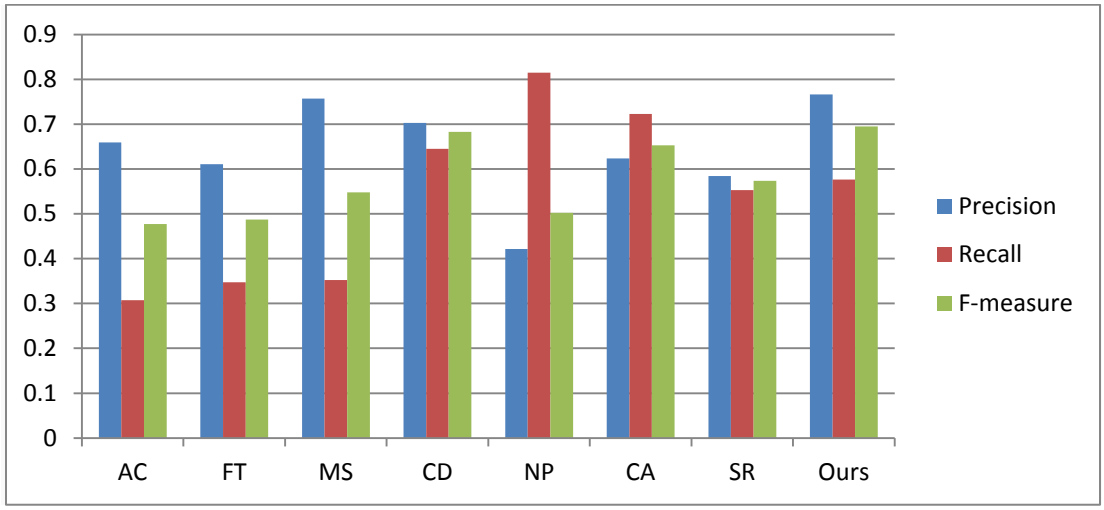


Fig. 5.12- Adaptive threshold based comparison of precision, recall & f-measure

5.4 Computational Cost

Computational cost can be defined as the time taken to run the algorithm over the image dataset. In our work, all the algorithms are run over SED1 dataset and computational time is calculated which is as shown in table 5.3

Table 5.3- Computational Cost

Method	**Run Time**
AC	5794.6 sec
FT	15.2 sec
MS	32.7 sec
CD	485.4 sec
NP	151.3 sec
CA	4 h 32 min
SR	4.54 sec
Ours	14.72 sec

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

In this thesis, we studied several techniques for detecting the salient region from the images. Some of the techniques are implemented in the work and analyzed in terms of various performance matrices. By understanding advantages, drawbacks and limitation of these techniques, we proposed an optimum technique for salient region detection from color images. We proposed a hybrid feature based technique for this purpose. The proposed modifications give several advantages as follows:

- Proposed technique generates best perceptual quality of saliency maps.
- The technique is very simple and effective.
- Computational cost of the algorithm is very low.
- The salient region in the images is uniformly highlighted with this algorithm.
- The proposed algorithm detects the proper object boundaries from the image.
- This method produced the saliency map of full resolution.
- This technique incorporate both global and local features which results in high precision, recall and f-measure value as compared to other state of art methods.

6.2 Future Scope

In the future work, the saliency map could be further enhanced using high-level factors, such as recognized objects or face detection. For example, one could incorporate the face detection algorithm of [23], which generates 1 for face pixels and 0 otherwise. The saliency map can then be modified by taking the maximum value of the saliency map and the face map. This modification will decrease the computational cost and will increase the detection accuracy of the algorithm. We view this step as an a posteriori refinement of the saliency map and hence exclude it from our experiments.

References

1. Y.-F. Ma and H.-J. Zhang. Contrast-based image attention analysis by using fuzzy growing. In *Proceedings of the Eleventh ACM International Conference on Multimedia*, pages 374{381, November 2003.
2. V. Setlur, S. Takagi, R. Raskar, M. Gleicher, and B. Gooch. Automatic image retargeting. In *Proceedings of the 4th International Conference on Mobile and Ubiquitous Multimedia (MUM'05)*, pages 59-68, October 2005.
3. S. Avidan and A. Shamir. Seam carving for content-aware image resizing. *ACM Transactions on Graphics*, 26(3):10, July 2007.
4. J. Han, K. Ngan, M. Li, and H. Zhang. Unsupervised extraction of visual attention objects in color images. IEEE Transactions on Circuits and Systems for Video Technology, 16(1):141–145, 2006.
5. U. Rutishauser, D. Walther, C. Koch, and P. Perona. Is bottom-up attention useful for object recognition? IEEE Conference on Computer Vision and Pattern Recognition, 2, 2004.
6. R. Achanta, F. Estrada, P. Wils, and S. S¨usstrunk, "Salient region detection and segmentation," International Conference on Computer Vision Systems, vol. 5008, pp. 66–75, 2008.
7. R. Achanta, S. Hemami, F. Estrada, and S. S¨usstrunk, "Frequency-tuned salient region detection," IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1597–1604, June 2009.
8. L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(11):1254–
   1259, 1998.
9. R. Achanta and S. Susstrunk, "Saliency detection using maximum symmetric surround," *ICIP'10*, pp. 2653-2656,2010.
10. Naila Murray, Maria Vanrell, Xavier Otazu, and C. Alejandro Parraga, "Saliency Estimation using a non parametric low level vision model" 2010.

11. S. Goferman, Z. M. Lihi, and A. Tal, "Context-aware saliency detection," IEEE transaction of pattern recognition and machine intelligence, vol. 34, no.. 10, october 2012.
12. X. Hou and L. Zang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Computer. Vis. Pattern Recognition.*, Jun. 2007, pp. 1–8.
13. R. W. G. Hunt. *Measuring Color*. Fountain Press, 1998.
14. Yuming Fang, Zhenzhong Chen, Weisi Lin, Chia-Wen Lin, "Saliency-based Image Retargeting in the Compressed Domain" ACM 978-1-4503-0616-4/11/11.
15. S. Frintrop, M. Klodt, and E. Rome, "A real-time visual attention system using integral images," in International Conference on Computer Vision Systems, March 2007.
16. R. T. Rockafellar, and R. J.-B. Wets. Variational Analysis. Springer-Verlag, 2005.
17. W. S. Geisler, and J. S. Perry. A real-time foveated multi-solution system for low-bandwidth video communication. *SPIE*, 3299, 1998.
18. K. Mullen. The contrast sensitivity of human color-vision to red green and blue yellow chromatic gratings. Journal of Physiology, pages 381–400, 1985.
19. G.Wyszecki and W. S. Styles. Color Science: Concepts and Methods, Quantitative Data and Formulae. Wiley, New York, NY, 1982.
20. K. N. Plataniotis, A. N. Venetsanopoulos, Color Image Processing and Applications. Springer, Berlin, 2000.
21. P. Kovesi, "Image features from phase congruency," *Videre: J. Comp.Vis. Res.*, vol. 1, pp. 1-26, 1999.
22. D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am.A*, vol. 4, pp. 2379-2394, 1987.
23. P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," *Proc. IEEE Computer Vision and Pattern Recognition*, 2001.
24. S. Alpert, M. Galun, R. Basri, and A. Brandt, "Image segmentation by probabilistic bottom-up aggregation and cue integration," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2007, pp. 1–8.
25. Rafael Gonzalez and Richard E. Woods. Digital Image Processing. Addison-Wesley, second edition, 2002