CHAPTER 2

LITERATURE SURVEY

2.1  INTRODUCTION

This chapter elaborately deals with the state-of-the-art techniques for attention modeling and their application to image segmentation and image compression. The chapter consists of four major parts. First part deals with the various computational models used in saliency detection. Special emphasis is given to spatial and frequency based visual saliency method. The other two parts illustrate the application of methods for visual saliency detection in image segmentation and image compression. Literatures on visual saliency based image segmentation and image compression are also discussed.

2.2  COMPUTATIONAL MODELS FOR SALIENCY DETECTION

In any natural setting, a scene is viewed by humans without any effort in mind and the eyes automatically turn to the objects which appear prominently distinct from their background. This wonderful ability of human is often referred to as “visual attention”. Researchers have expended a lot of effort in developing a model which will accomplish or mimic the innate ability of human attention. Researchers from the neuroscience, cognitive theory and pattern recognition have tried to model this wonderful behavior of human beings.

Concepts of feature integration theory and guided search model were used subsequently, in the major findings of computational models. The
computational model of visual attention takes an image as input and yields a topographical map which is called “saliency map” as output. The research area of visual saliency detection is more vibrant in the field of computer vision. The computational models for saliency detection find application in many of the computer vision tasks such as image retargeting, image resizing, image retrieval, image compression etc. Several methods such as pixel-wise differences, spatiotemporal cues, center-surround differences, graph based visual saliency and psychological attention models have been successfully applied to tackle detection of visual saliency.

Generally, the models for visual saliency detection aim to identify salient regions or may predict human fixations. Both the types of models produce saliency map as output. Fixation prediction models generally identify popped-out portions in the salient regions, but on the contrary, salient region detection models yield the smoothly connected regions. The strength of the saliency detection models is that they are immensely useful in applications like video summarization, image quality assessment, image editing and human robot interaction. In this present study, visual saliency detection models which are of the latter type, i.e. salient region detection have been studied extensively and applied successfully in image segmentation and image compression.

The visual saliency detection models are capable of predicting the important regions of an input image/scene. By using the saliency map the behavioral predictions are also studied. The existing computational models for saliency detection use a variety of input features. The models can be differentiated based on the input features, such as contrast, entropy, pixel similarity, symmetry, spatial and frequency analysis. In addition, center bias and other border effects are also used to construct the saliency map.

The pioneer model for saliency detection was developed by Itti & Koch (1998). In this method, biological plausible set of features are taken to
construct the visual saliency model. These features mimic the behavior of visual processing in humans. Individual feature map is created from the pre-attentive features. Winner Take All (WTA) mechanism is adopted for combining the feature maps. This work was extended to improve the performance of this approach. Similar to Itti Koch’s model, Harel et al. (2006) proposed graph based visual saliency model with performance enhancement of Itti’s approach. It uses the concept of dissimilarity in the saliency detection method. Gao et al. (2008) used center surround hypothesis in the visual saliency detection. The saliency map was obtained at a coarser scale. So these models suffered from the local information loss. In order to circumvent this, two types of normalization were performed. One is iterative filtering and the other is non-linear normalization. These normalization procedures were inefficient in fusing the feature maps.

Thereafter, many researchers have developed various models for visual saliency detection in the spatial domain and as well as in the frequency domain. Each of the approaches has its own pros and cons. Bruce & Tsotsos (2005) have proposed a visual saliency model based on the concept of information maximization. Independent Component Analysis (ICA) was used to obtain the image feature set. Ma & Zhang (2003) proposed saliency detection making use of fuzzy growing algorithm. Local contrast information is used to find the saliency map. Extended Saliency or E-saliency was proposed by Avraham & Lindenbaum (2010). It was a graphical method to find the salient parts of an image. By using stochastic approach, the probability of interest in a scene/image was identified. Jianming Zhang & Stan Sclaroff (2013) proposed saliency detection with the help of Boolean maps. Saliency map was obtained by leveraging the topological arrangements of Boolean maps.

There are literatures which describe various models for saliency detection based on statistics, stochastic concepts and probability. These
concepts served as basis for other models. WTA and center surround mechanism were generally used in the development of visual saliency models in larger context, because it gave vibrant interpretation and computation on saliency models. These mechanisms were implemented using the local features such as color, contrast, orientation, texture, shape and KL divergence and weighted histogram.

Based on the literatures, the following steps are investigated for the modeling of saliency detection.

Step 1 Obtain low level features

Step 2 Bind/Fuse the feature maps for the corresponding features

Step 3 Incorporate features in a top down fashion for final saliency detection

In step 1, the visual attention modeling process involves many low level features. Color is one of the major features. So the selection of color channels plays a significant role in this step. RGB, Lab, YCbCr, RGBYI are the major color channels which were used in the saliency detection literatures. Usually RGB, Lab color channels were used in the saliency detection. Recently, a few literatures are quantifying RGBYI color channels in saliency detection. Riche et al. (2013) used YCbCr color channel for the visual saliency detection. YCbCr color channels more or less match human vision, compare to RGB color space. Xiaolong Ma et al. (2015) proposed saliency detection based on Singular Value Decomposition. In this method, YCbCr color space was employed for singular value calculation of image. Borji (2012) used RGBYI color channels in the estimation of saliency. Imamoglu et al. (2013) preferred Lab color channels in the saliency detection process.
Step 2 involves the compilation of feature maps. This is done by linear combination, hierarchical inference, pixel-to-pixel wise, manipulation of feature maps or feature vectors etc.

Step 3 incorporates specific cues/tasks of an image/scene. Higher order statistics were added to tackle tasks which involved faces, buildings, people, car, vehicle, etc.

Ma & Zhang (2003) used only color features to compute the saliency model. Color contrast of each pixel was considered to generate this model. Actual saliency results were not obtained by this model, because only the color features were considered. Zhang Peng & Wang Run-sheng (2004) also considered texture and intensity features in addition to color to construct the saliency detection model. Contrast is considered as an important cue in the visual attention process. So Jie Liu & Shengjin Wang (2015), Tong et al. (2015) used local and global contrast to detect the salient region of an image. Global contrast was computed through low level features. Anton Garcia Diaz et al. (2012) proposed saliency detection using adaptive whitening of color features. Keeping the statistical structures of an image in mind, hierarchical adaptation was employed. Qi et al. (2015) also used contrast information to help saliency detection. In this method, color contrast was utilized by measuring color distribution with respect to background prior.

Many techniques for saliency detection available in the literature come under the umbrella of spatial domain analysis. However, spatial techniques, in fact, have downsides like higher computational effort and less speed (Ru-Je Lin & Wei-Song Lin 2014). So researchers shifted their attention to transform domain approaches. The phase spectrum of Fourier transform is used by Guo et al. (2008) to model visual saliency detection. Li J et al. (2013) proposed saliency detection by using the phase spectrum of Fourier transform and Gaussian kernel. Fourier transform provides satisfactory output for
stationary signals only, whereas Wavelet transform could represent the singularities in a much better manner than what Fourier transforms could. So, in the recent studies, Wavelet transform is generally used to model visual saliency. Immagolou et al. (2013) used multi-scale Wavelet transform to model visual saliency. Inverse Wavelet transform was applied to get the feature maps and thereby producing the saliency map. Xiaolong Ma et al. (2015) proposed saliency model based on Wavelet transform and entropy. Saliency maps of different color channels were combined using the entropy and Wavelet coefficients. Even though the Wavelet transform is good in representing the point singularities, it fails to address the directional features. So, highly directional Wavelet transforms were introduced in the visual saliency research. The Curvelet transform was introduced in the visual saliency detection by Sheng-hua Zhong et al. (2011). The directional information of Curvelet transform and spatial localization of Gabor transform were harnessed to obtain the saliency map. To further sharpen the directional singularities, Ripplet transform is expected to give good results in visual saliency detection.

2.3 VISUAL SALIENCY BASED IMAGE SEGMENTATION

The application of visual attention modeling has become widespread in the areas of segmentation, compression, recognition, resizing, tracking, etc. Visual saliency is one of the intellectual abilities that bring out any pixel from its neighborhood and thereby grabs the attention of the beholder. In this study, saliency based image segmentation and saliency based image compression are rigorously discussed.

Image segmentation is one of the key areas in the field of pattern recognition and has attained significant level of maturity in research and in industry over the last few decades. Even though there were prominent advances in this field, none of the techniques could be taken as a general method for all applications. In many of the computer vision problems, the
image segmentation is not well defined. The image segmentation is defined as a process of splitting the set of pixels in an image into several subsets. It also delineates background and foreground of an image.

Segmentation approaches broadly fall into six categories. They are classified based on region, edge, feature, graph, cluster and energy. Many hybrid approaches are also implemented for various applications. The typical algorithms for region based image segmentation can be listed as region growing, splitting, and merging. The region growing technique identifies the seeds initially. The adjacent pixels are grouped into these seeds depending on the similarity in terms of intensity, color and texture. Hence, the selection of the seed points plays a vital role in image growing techniques. The region splitting and merging techniques divide the input image into a series of small regions. The region merging is done by considering the dissimilarity measurements in color, area and binary partition tree.

In edge based image segmentation, different pixels are identified through Laplace and gradient methods which connect the image foreground and background. Common edge detectors like those of Prewitt, Sobel, Robert under gradient concepts are used to detect contour. However, these may be very sensitive to noise. Watershed segmentation algorithm is one of the popular methods in the edge based image segmentation. In order to separate the two different regions, the Watershed lines are found out. This is the main objective of the Watershed image segmentation.

Feature based image segmentation rely on image features which are extracted from an input image. Input features are grouped together according to characteristics of color, texture, edge etc. The edge information of the image and the spatial structure are not conserved generally.
The next category of image segmentation is based on graph theoretic concepts. The graph based image segmentation is one of the hybrid approaches utilizing features and spatial information. Similarity and proximity are the metrics used in the visual grouping. In these methods, usually a graph is generated in which each vertex corresponds to a pixel or a region. In the weighted graph, the weighted are attached to the color texture features. Graph cut, min cut, normalized min cut are some of the standard methods which are used in the color image segmentation. Normalized cut algorithms require high computational complexity. So it is not applicable for the real time image processing applications. Shi & Malik (2000) proposed a method for color image segmentation. Normalized cut algorithm is implemented by utilizing the solution of Eigen systems. Tao et al. (2007) used mean shift along with the normalized cut in their proposed method for image segmentation method.

In clustering based image segmentation, the input image is viewed as multidimensional data. The input image is clustered based on the homogeneity criterion. It is reported that his method also produces better segmentation results, but it suffers from over-segmentation. Feature extraction is an important step in the clustering based segmentation.

Image segmentation based on optimizing an evaluation function over features and spatial information also exists in the literature. Different types of evolutionary algorithms have been devised. Genetic algorithm, particle swarm optimization and honey bee optimization are the majority of evolutionary type algorithms. By using these types of algorithms the search space can be reduced and there is also a possibility of getting stuck in local minima.

In recent years, visual saliency detection is being highly researched by researchers from multiple fields including individuals from the fields of psychology, computer vision and neurobiology in order to unravel the secret
of how human beings perceive visual inputs. It is also applied in other methods for image segmentation. The methods for visual saliency detection aim at detecting which part of an image mostly captures the human perceptual attention.

In general, segmentation of salient region involves two major tasks - saliency detection and salient region segmentation. Salient region/object segmentation is one of the key techniques in applications like object recognition, object tracking, image editing, image retrieval, and 2D/3D conversion. Saliency map is used to identify the salient pixels. The salient regions and non-salient regions are segmented using any one of the machine learning techniques. It is quite different from the general segmentation methods which aim at partitioning an image into intelligible regions. Many methods for object segmentation have inbuilt saliency detection models. Outwardly, saliency detection and region/object segmentation are independent but a knotty connection exists in between them. Conversely, it is also true that object/region segmentation is helpful in detecting saliency in images. In fact, the exact region for segmentation is the main objective of object level saliency detection.

Many segmentation methods rely on important information from the saliency map. Ko & Nam (2006) selected important regions from the saliency map using machine learning techniques. A trained SVM was used to perform segmentation. In Han et al. (2006), the main segmentation seeds were retrieved from a saliency map. The standard Markov Random Field is used for the object segmentation which utilizes the basic image features such as color, edge orientation and luminance. In Achanta et al. (2009) method, the salient objects were distinguished from their background using threshold method. The average saliency values were considered while setting the threshold.
In Fu et al. (2008) method, spectral residual approach was applied to create the saliency map which was used to generate seeds for graph-cut implementation. Li et al. (2011) used saliency map for the extraction of salient regions which were used in the graph cut segmentation. Similarly Lee et al. (2012) used the feature maps of color, intensity and orientation to generate the saliency map which was used in region merging. A simple particle swarm optimization method was employed.

Saliency directed color image segmentation was presented by Li & Ngan (2008) in which the saliency map was used in performing region merging for the applications of face segmentation. Valenti (2009) presented the real time salient region detection. Curvedness, iso-centricity, rarity of color edges are used to calculate pixel wise saliency. Graph based segmentation is adopted to segment the salient regions.

The recent studies have shown that the saliency maps generated using visual computational models yield useful information for the image segmentation process. Xuefei Bai & Wenjian Wang (2014) proposed SVM-saliency for color image segmentation wherein the salient regions and the background are extracted by using visual saliency models. Saliency map based positive and negative training datasets were identified. Using the trained SVM, salient regions and their background were segmented.

Esa Rahtu et al. (2010) proposed the CRF based segmentation method which identifies salient and non-salient regions of an image or video by optimizing a energy function. The Conditional Random Field (CRF) is used in this approach since the proposed method looks for pixel level segmentation. Optical flow motion cues are used in the saliency detection which yields stabilized output and reliable recall rates. Dwarikanath Mahapatra & Joachim M Buhmann (2015) proposed visual saliency based MRI segmentation. The informative samples for active learning are identified by the
visual saliency detection. Hence visual saliency detection is widely applicable in medical images also.

Shuzhen Li et al. (2015) proposed saliency cuts methods which are considered as special type of object segmentation. Saliency map was used to generate the seed and to segment the salient region. The graph cut algorithm was integrated to improve the segmentation results. Huynh Trung Manh & Gueesang Lee (2013) proposed visual saliency based small object detection. Gaussian Mixture Model (GMM) was used to locate the object of interest and thereby utilizing visual saliency analysis for object segmentation. The advantage lies in narrowing down the search region. The results are not the same (much better) as that of the traditional GMM segmentation.

Shui Linlin (2015) proposed adaptive target segmentation based on visual saliency. The salient targets were identified using the visual saliency model. The bottom-up feature channels of color, intensity, direction and energy were used to obtain the saliency map. Pre-segmented images were computed through saliency map. A maximum entropy criterion was adopted for target entropy calculation. Guokang Zhu et al. (2014) introduced tag saliency especially for top down attention. Hierarchical segmentation and auto tagging were made use of in this method.

Yanulevskaya et al. (2013) investigated the visual saliency detection based on spotlight attention theory. The complex images were segmented into proto-objects and then they are assessed for individual proto-object saliency. Two types of object-wise saliency were introduced. The first one deals with different objects which are different from their surroundings and the second one deals with an object which contains rare information. These two types with complementary characteristics are explained in this method. All these methods are investigated through natural images or video sequence. Apart
from these applications, saliency detection is also applied in real time medial images. Based on the saliency detection, the segmentation is carried out.

2.4 VISUAL SALIENCY BASED IMAGE COMPRESSION

Detection of salient regions is employed in many of the computer vision and pattern recognition tasks such as image compression, object recognition, content-based image retrieval, image collection browsing, image editing, visual tracking and human-robot interaction. Owing to the development of multimedia communication technology, it is now mandatory to perform image compression, when transferring an image from one end to another. Generally, prominent regions of an image highly degraded at low bitrates. The standards of compression like JPEG/JPEG-2000, MPEG 4 do not handle the salient regions well.

Nabil Ouerhani et al. (2001) proposed the adaptive color image compression based on biologically inspired visual attention. The initial stage perceptual salient regions of interest were identified automatically. Adaptive coding scheme was used to allocate higher number of bits for the salient regions. The results were compatible with the JPEG standards.

Stella et al. (2009) used visual saliency to achieve image compression. At individual scales saliency is measured through the Laplacian pyramid. Their proposed compression algorithm decreased the entropy of the image with respect to the saliency map in each scale. Nitin Dhavale and Laurent Itti (2003) proposed visual computational model based image compression. It successfully located the regions of interest of the human and thereby it is applied for image compression.

Li et al. (2011) performed video compression based on computational models of visual attention. The salient regions were encoded
with higher quality compared with non-salient regions. The salient regions were given higher priority compared to other regions. But it led to visible artifacts in the non-salient part where the quality of image is poor. The artifacts also sometimes may draw the end-user’s attention. In severe cases, the artifacts become prominent and captured the viewer’s attention – pseudo saliency.

Lijuan Duan & Chunxia Ke (2012) proposed a image compression technique based on saliency detection and Independent Component Analysis (ICA). First, the input image was transformed using ICA. The transformed coefficients were numbered with set zero coefficient percentage. The sparse nature of independent component analysis was used in their method. It was compared with DCT based compression method.

Hadi Hadizadeh (2013) dealt with visual saliency methods for video compression and transmission in his dissertation. Saliency based video coding was investigated. The main concept was that the highly salient regions have higher appeal to percept than the lesser salient regions. The quality of an image or a video was manipulated towards the user most attended regions. His method effectively performed video coding expect in two major cases. If any region is rich in saliency, then its saliency can be increased after the compression, assuring the quality of the image/video to remain high. The reason is that the users notice highly salient regions in the scene. If the region is less salient, then its saliency will be decreased after the compression task, resulting in lower saliency regions with lesser quality.

Tien Ho Phuoc et al. (2012) proposed visual saliency based data compression for image sensors. The adaptive image compression is presented in each block. First the saliency value is obtained, then Haar Wavelet transform is applied for the compression. This framework gives lesser memory
and compact operators. The data stored in the image sensors are very much reduced and image quality is also not affected.

Fabio Zund et al. (2013) proposed the content based image compression using visual saliency methods. Initially the saliency map is obtained from the video, automatically or inputted by the user. The salient regions are identified with nonlinear image scaling. Salient image regions are given higher pixel count and non-salient regions are given lesser pixel count. Extant compression techniques were utilized to compress the non-linearly down-scaled images and in the receiver’s end it was up-scaled. This method supports anti-aliasing effect which reduced aliasing in highly scaled regions.

In order to reduce the redundant information in the dynamic scenes, the visual saliency in video was proposed by Qin Tu et al. (2015). Based on the video visual saliency map, the redundant information was removed. In this paper, video visual saliency is the catalyst of video compression technique. DCT is utilized to perform the video compression and this technique is adopted in MPEG-1, MPEG-4 and H.265/HEVC standards.

Chenlei Guo et al. (2010) and Hadi (2014) investigated saliency based compression techniques. To adhere to the saliency values, the transmitted coefficients were modified. These methods could not handle salient regions well and also suffered from complicated computations. Souptik Barua et al. (2015) developed Wavelet based image compression technique for images and videos. The algorithm was designed to obtain the saliency values in the Wavelet domain and then the corresponding image/video coefficients are transmitted. It preserved the important regions of an image/video.

In the review of literatures, various techniques are proposed for saliency based image compression. Most of the methods had less compression ratio. It is due to the higher computations in each step. The number of
computations in Walsh Hadamard Transform (WHT) is lesser compared to other transforms. In the proposed method, Walsh Hadamard Transform (WHT) based visual saliency is used in the image compression task.

2.5 SUMMARY

The inferences drawn from the literature survey are summarized below:

i. The spatial domain based visual saliency detection techniques suffer from higher computational cost and slower in speed.

ii. Transform domain approaches produce better results compared to the spatial domain approaches

iii. Orientation singularizes of an image are not captured by the Wavelet transform and higher order Wavelet transforms.

iv. Identification of salient regions reduces the search space for segmentation.

v. Till date, perfect segmentation algorithms of salient regions are poorly studied.

vi. Salient region segmentation approaches are very slow and are often used in preprocessing operation for most of the complex tasks.

vii. Finding the salient information in image can effectively reduce the computation cost and simpler hardware in the compression process.
Based on the exhaustive literatures, the main focus of the current research is:

The orientation singularities of an image should be addressed in the visual saliency detection models.

i. To address the orientation singularities of an image, the novel visual saliency detection model based on Ripplet transform is proposed.

ii. To reduce the search space in image segmentation, visual saliency detection is incorporated in leukocytes segmentation and salient region segmentation

iii. To reduce the number of computations in compression, Walsh Hadamard transform based visual saliency is applied.