PERFORMANCE EVALUATION OF SALIENCY MAP METHODS ON REMOTELY SENSED RGB IMAGES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SELEN SÖNMEZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
GEODETIC AND GEOGRAPHIC INFORMATION TECHNOLOGIES

MAY 2016

Approval of the thesis:

PERFORMANCE EVALUATION OF SALIENCY MAP METHODS ON REMOTELY SENSED RGB IMAGES

submitted by **SELEN SÖNMEZ** in partial fulfillment of the requirements for the degree of **Master of Science in Geodetic and Geographic Information Technologies Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences**

Assoc. Prof. Dr. Uğur Murat Leloğlu
Head of Department, **Geodetic and Geographic Inf. Tech.**

Prof. Dr. Uğur Halıcı
Supervisor, **Electrical and Electronics Eng. Dept., METU**

Examining Committee Members:
Assoc. Prof. Dr. Uğur Murat Leloğlu
Geodetic and Geographic Inf. Tech. Dept., METU

Prof. Dr. Uğur Halıcı
Electrical and Electronics Eng. Dept., METU

Prof. Dr. İlhami Bayramin
Soil Science and Plant Nutrition Dept., Ankara University

Assoc. Prof. Dr. İlkay Ulusoy
Electrical and Electronics Eng. Dept., METU

Assoc. Prof. Dr. Mehmet Dikmen
Computer Engineering Dept., Başkent University

Date: 05.05.2016

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: SELEN SÖNMEZ

Signature :

ABSTRACT

PERFORMANCE EVALUATION OF SALIENCY MAP METHODS ON REMOTELY SENSED RGB IMAGES

Sönmez, Selen
M.S., Department of Geodetic and Geographic Information Technologies
Supervisor : Prof. Dr. Uğur Halıcı

May 2016, 88 pages

Predictive applications of human eye visualization so called saliency map computational models become more attractive in image processing studies. Saliency map highlights regions that are distinctive from their surrounding in the images in interest. In this study, various computational models for salient region detection are investigated on remotely sensed images. The computational methods considered are Itti-Koch, Graph-Based Visual Saliency, Saliency Detection by Combining Simple Priors, Frequency-tuned Salient Region Detection, Image Signature and Region Covariance based Saliency. For evaluation of the computational methods, a dataset containing 226 remotely sensed RGB images has been prepared. The dataset forestry and water surface images captured in three different levels. The saliency maps produced by the computational methods on the dataset are compared with the saliency maps extracted from data collected in experiment conducted on human subjects. In these experiments 20 subjects are participated and the data is collected by using Tobii T120 Eye Tracker device while the images in the dataset are presented to subjects on computer screen.

In the performance evaluation, the saliency maps obtained from human subjects are used as ground truth. The performances of the computational methods are determined by computing similarity of their results to ground truth. As similarity measure, Cosine correlation, Pearson correlation and Structural Similarity index are used. Our experimental evaluation demonstrated that Region Covariance based Saliency and Graph-Based Visual Saliency are the best saliency methods among those that we considered for saliency map generation of remotely sensed RGB images.

Keywords: Saliency Map, Eye Tracker, Image Processing, Correlation, Similarity Measurements

ÖZ

UZAKTAN ALGILANMIŞ RGB GÖRÜNTÜLERİNDE DİKKAT ÇEKERLİK HARİTASI METOTLARININ PERFORMANS DEĞERLENDİRMESİ

Sönmez, Selen
Yüksek Lisans, Jeodezi ve Coğrafi Bilgi Teknolojileri Bölümü
Tez Yöneticisi : Prof. Dr. Uğur Halıcı

Mayıs 2016, 88 sayfa

İnsan gözünün algılama sistemini tahmin edebilir çalışmalar, diğer bir deyimle dikkat çekerlik haritaları, imaj işleme alanında çok daha yaygın hale gelmiştir. Dikkat çekerlik haritaları, bir görüntü içerisinde yer alan farklı alanları veya objeleri ön plana çıkartmaktadır. Bu tezde dikkat çekerlik haritası oluşturan çeşitli uygulamalar uzaktan algılanmış görüntüler üzerinde incelenmiştir. Bu uygulamalar; Itti-Koch, Graph-Based Visual Saliency, Frequency-tuned Saliency Region Detection, ImageSignature, Saliency Detection by Combining Simple Priors ve Covariace based Saliency modelleridir. Dikkat çekerlik haritası uygulamalarının performans değerlendirmeleri için uzaktan algılama sistemleriyle elde edilen 226 tane RGB görüntüsü içeren veri seti hazırlanmıştır. Veri setinde yer alan imajlar orman ve su yüzeyi kategorilerine ayrılarak üç farklı seviyede elde edilmiştir. Veri seti kullanılarak, bu tezde incelenen uygulamalar ile oluşturulan dikkat çekerlik haritaları insan deneklerin katılımıyla gerçekleştirilen deney sonucunda oluşturulan dikkat çekerlik haritaları ile karşılaştırılmıştır.Deney, Tobii T120 göz izleme cihazı ve 20 deneğin katılımıyla tamamlanmıştır. Her deneğe veri setinde bulunan imajlar göz izleme cihazında yer alan ekranda gösterilmiştir.

Performans değerlendirme aşamasında, deneklerden elde edilen dikkat çekerlik haritaları referans doğrulaması için kullanılmıştır.Dikkat çekerlik haritası uygulamalarının performansları, elde edilen referans doğrulamaları baz alınarak, bu uygulamaların sonuçlarının benzerlik ölçümleri ile belirlenmiştir.Cosine korelasyonu, Pearson korelasyonu ve Structural Similarity indeks metotları benzerlik ölçümlerinde kullanılmıştır. Deney sonuçlarında Region Covariance basedSaliency ve Graph-Based Visual Saliency modellerinin uzaktan algılanmış RGB görüntüleri ile dikkat çekerlik haritası oluşturan ve incelenen diğer modellere göre en iyi metot oldukları saptanmıştır.

Anahtar Kelimeler: Dikkat Çekerlik Haritası, Göz İzleme Cihazı, Görüntü İşleme, Korelasyon, Benzerlik Ölçümü

To My Family

ACKNOWLEDGMENTS

I would like to thank my thesis advisor Prof. Dr. Uğur Halıcı of the Engineering Faculty at METU. She steered me in the right direction whenever she thought I needed it while she consistently allowed this thesis to be my own work.

I also like to thank my friends who were involved in the experiment part of this thesis study and contributed me to understand details of the processes in required experiment research: Cafer Ünsal, Başak Özkan, Ufuk Bölükbaşı, Gülin Demirparmak, Özgehan Özen, Nihan Ocak and Alaettin Zubaroğlu. Without their passionate participation and input, experimental results could not have been successfully conducted.

My sincere thanks also goes to my dear friends Arda Arıkoğlu, Görkem Can and Ayşe Yılmaz for supporting me spiritually throughout writing this thesis.

I would also like to acknowledge my director Tolga Yaşıt for supporting me to complete my research and writing this thesis whenever I need.

I must express my very profound gratitude to my father Bülent Sönmez and to my mother Sevim Sönmez for providing me with unfailing support and continuousencouragement throughout my studies and through the process of researching and writing this thesis.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

GIS	Geographical Information Systems
STD	Standard Deviation
PNG	Portable Network Graphics
GBVS	Graph-Based VisualSaliency
SDSP	Salient Detection by Combining Simple Priors
Achanta	Frequency-tuned Salient Region Detection
CovSal	Region Covariance based Saliency
WTA	Winner Take All Computation
DoG	Difference of Gaussians
GUI	Graphical User Interface
DCT	Discrete Cosine Transform
IDCT	Inverse Discrete Cosine Transform
SSIM	The Structural Similarity Index
RGB	Red, Green, Blue
Lab	International Commission on Illumination 1976 Color Space
CV	Coefficient of Variation
HS	Human Subject

CHAPTER 1

INTRODUCTION

1.1. Saliency Map and GIS Applications

Human eye is capable of recognizing the differences in an image of interest according to the distinctive features such as color, orientation, location and texture. In recent years, computational models which approach human eye behaviors have been developed. They identify the most significant areas of the given image for data extraction. Saliency map indicates how a certain location in an image is distinctive from its surround by using the features sensitive to differences (Koch & Ullman 1987). Methods in saliency map computation diversify. For example, warm colors such as red and yellow are more pronounced to human visual system than cold colors such as green and blue. The color channels representing green-red and blue-yellow information are performed to extract salient region. Band-pass filtering is another method to represent human visual system (Zhang, et al., 2013). Computation of saliency map such a system that simulates human eye behaviors with respect to eye movement in area of interest which can be called as important parts of the interested area. Considering the certain areas that include different patterns, unique parts which do not repeat and colors provides higher contrast from background and repetitive items, salient region detection become more efficient and saliency map outputs can vary as in Frequency-tuned Salient Region Detection (Achanta, et al., 2009) or GBVS and Itti-Koch (Biskupska, 2013).

Saliency map is generated based on low-level and high-level features. While low-level saliency features include visual features as color, edge and texture, high-level saliency features contain also priors as location, semantic and color. Therefore, detection probability of the salient regions becomes higher. Numerical results show that parts contain defined priors such as color, edge and location become vivid. It is also obvious that results of saliency map are expected to be observed with a deviation that is as small as possible in order to obtain accurate and precise information extraction. Experiments were executed and results of comparisons between different models with same image dataset confirm that high-level features provide more accurate and precise salient region detection (Shen, et al., 2012). On the other hand, it is a fact that human neuronal structure is implausibly fast and its ability to perceive is high. In case of generating saliency map, this is why computational models only approximate but cannot mirror the reality. In order to analysis performance of low-level or high-level computations, real-life experiments can be performed. Correlation between computational models and real-life measurements is a measure for determining the closeness of saliency map computations to human eye detection in reality.

Saliency maps can be used to supply a reliable pre-warning system solution for geographical information system based applications aiming to detect differences in specified areas that are periodically controlled with remotely sensed imaginary. Visual saliency detection proposes to reduce search space so that the areas of interest in images can be processed much more rapidly (Cui, et al., 2014). Additionally, saliency based segmentation is a better solution for whole image interpretation. If the most attractive objects or areas in remotely sensed images can be highlighted closer to human vision, the targets in the input image will have a higher probability to be identified or the image can be interpreted better even though no target is identified (Sharma & Ghosh, 2015). Vegetation health is one of the most common studies in agricultural work field. Pre-warning applications constitute importance in order to detect vegetation anomalies which response to environmental factors and human interventions (Asoka & Mishra, 2015). Color and texture changes in fields or forestry may emerge due to concrete construction, desertification, lack of nitrogen and potassium in soil. Another area related to saliency detection is coast security which contains marine facility. In case of illegal boat or ship accommodation in forbidden water zones, objects that are differed from background in remotely sensed image can be highlighted and tracked. Water level changes also provide desertification awareness according to size and distance differences between current and previous statuses. In order to recognize differences in such geographical images, saliency maps can be can be applied on set of images that are acquired in same location but different time.

1.2 Literature Survey

Saliency map generation is mainly conducted by top-down and bottom-up saliency. Feature maps are generated separately by decomposing input image according to features such as intensity, color, location and orientation of the input image. Bottom-up saliency uses low-level feature maps without any prediction like knowledge, expectations and goals. However, top-down saliency uses high-level feature maps which emphasis priors of the features in the image with prediction. Therefore, performance of saliency algorithms become vary according to the feature maps that are used (Shen, et al., 2012). Various techniques are available in producing multi resolution representation of images such as Gaussian and Laplacian pyramids (Burt & Adelson, 1983; Olkkonen & Pesola, 1996). A sequence of low-pass filtered images so-called Gaussian pyramids are obtained by convolution of weighting function iteratively within the input image. Therefore, image samples are reduced with a decreasing density in each iteration (Burt & Adelson, 1983).

In remote sensing, saliency methods have been employed some target recognition studies. For example; saliency features are used in Hu & Tao, 2007 for automatic road extraction, saliency map is used in order to extract candidate regions for ship detection from optical satellite images. Multiple saliency map models are also computed together for ship detection in Li, et al., 2015. Salient region detection also provides robust and effective computational result for precise airplane detection (Li, et al., 2011). Another area that uses saliency map is automatic cloud detection in satellite images (Hu, et al., 2015; Qi, et al., 2015).

A huge volume of high-resolution images can be acquired by satellite sensors, however; processing them can lead time consumption. In Zhang, et al., 2015, saliency detection is used in order to extract data patches for robust and more efficient feature learning in high-resolution satellite imaginary. Besides satellite imaginary, aerial photography images are also inputs for saliency detection algorithm for detection of outstanding objects (Rigas, et al., 2013). , Salient region detection is also used for small target detection in infrared images (Qi, et al., 2013).

In literature, a diversity of saliency map computations based on low-level or high-level features and combination of them has been proposed. Frequency-tuned Saliency Detection model proposed in Achanta, et al., 2009, is a frequency based method with low-level features providing well-defined boundaries of salient objects and regions. While Itti-Koch and GBVS models separate low-level features as maps, process them and combine to generate saliency map (Cui, et al., 2014), Salient Detection by Combining Simple Priors uses high-level feature maps as priors (Zhang, et al., 2013). Additional to complete bottom-up saliency process, GBVS uses graph based computations as normalizing the feature maps. Normalized feature maps express the salient region much more glaring way (Harel, et al., 2006).Considering their diversity, all these methods are examined and compared in this thesis study. Depending on bottom-up and top-down visual saliency models or combination of them, another model was included to our study. Region Covariance based Saliency computational model proposed in Erdem & Erdem, 2013, uses nonlinear integration of different features by using covariances to produce saliency map. It is indicated in Erdem & Erdem, 2013thatCovSal model is suitable for natural images. Image Signature is another computational saliency detection model proposed in Hou, et al., 2012. It is based on figure-ground separation in human visual system which has ability to detect separated features in images rapidly. Because of the fact that remotely sensed images are used in experimental studies of this thesis containing distinctive features, this model was included.

At first sight, unique features are observable variables by human eye. Therefore, decomposing the image in distinctive feature maps such pattern and color is a method to extract non-unique information or the regions close to background. Salient regions become highlighted by combination of these feature maps (Zhao, et al., 2015). Figure-ground is another factor to obtain distinctness for saliency region detection. As it is indicated in Koch & Ullman 1987, human visualization system perceives the rapid differences better and faster than smooth changes in images (Treisman & Gelade, 1980). Therefore, more separated figure-ground provides higher discreteness observation between the background and the objects or regions such as edges, corners, contrast color changes in image (Romantan, et al., 2002).Since the improvements in saliency computational models become fact, performance comparisons between them are also occurred in order to find such a model that approximates the reality best (Zhao, et al., 2015; Biskupska, 2013).

Heat map refers to color representation conducted by quantitative data. Colors in a heat map shows the distribution of the quantitative data that is stated in the image (Few, 2010; Krakov & Feitelson 2013). Eye tracker based saliency maps were exported according to distribution of eye movements collected on each input image in the given dataset in gray scale. To evaluate the similarity degree of computational and eye tracker based saliency maps, correlation between them can be performed. However, it is also needed that elimination of outliers which can be caused by reasons such as lighting and exposure should be computed so that each pixel value of the image is identified in the same range (Rao, et al., 2014).

Since different approaches such as frequency-based model or covariance-based model are used in salient region detection, saliency map computational models are compared with reality in consideration of performance to find such a model that approximates the reality best (Zhao, et al., 2015; Biskupska, 2013).

In many saliency map studies, quantitative comparisons were obtained by using different similarity metrics. For value-based metrics normalized scan path saliency metric is performed to quantify the saliency map values at the eye fixation locations. Another metric type is distribution-based metrics includes Pearson correlation to find similarity between two distributions. Area Under Curve (AUC) is a metric includes location-based category.AUC selects saliency map values from random points to form the negative set. A binary mask is then created to separate the positive samples from the negatives (Riche, et al., 2013) , (Radha, et al., 2014; Le Meur, et al., 2013; Borji, et al., 2013). Cosine similarity method is also used for comparison of similarity measurements of images (Zuva & Zuva, 2012; Liu, et al., 2008). It is indicated in Wang, et al., 2004 that natural images signals are highly structured. Structural Similarity (SSIM) index provides a more direct way to compare images containing distortions caused by acquisition and processing.

1.3 Overview of the Thesis

Remote sensing applications are beneficial in many fields such as military, agriculture, health sector, geodesy and advertisement. The working fields like these require measurements or observations in order to obtain data. This can lead time consumption and cost. In order to decrease or eliminate field work as much as possible, image processing methods have been developed. Saliency maps are used as pre-warning applications to predict land degradation, deforestation and vegetation anomalies as well as to identify targets such as ship, cloud, airplane in remotely sensed images.

The performance of computational saliency methods are mostly evaluated on images containing distinctive features such as edges, different orientation, textured backgrounds and the objects differ from image background.

As the saliency methods are used in many recent studies in analysis of remotely sensed images, it also became necessary to compare the performances of these methods on remotely sensed images. According to figure-background structure and location of objects or areas in images, different saliency map methods can provide a higher detection performance on salient regions of the image. This is the motivation behind this thesis study.

The aim of this thesis is to assess the performance of computational saliency models on remotely sensed images by comparing their results with the results obtained from human subjects by conducting eye-tracking experiments.

In this thesis, Itti-Koch, GBVS, Achanta, ImageSignature, SDSP and CovSal methods are selected in order to generate saliency maps on remotely sensed images. Itti-Koch and GBVS methods use intensity, color and orientation based features which are related to structure of remotely sensed images. Achanta method produces saliency map with highlighted details and low computational cost. ImageSignature method collects data in important parts of the images. By the motivation of human brain visual system, SDSP method generates three different saliency maps and combines them for the final saliency map. In CovSal method non-overlapping regions are extracted and dissimilarity between these regions is measured in order to produce final saliency map.

A dataset, containing remotely sensed images was prepared for comparison of the performances of the selected saliency methods. The images are captured from Google Map Maker service using GMapCatcher toolbox. Forestry and Water Surface are the categories that we considered and each of them includes three zoom levels as -2, -1 and 0. Since higher zoom levels more than 0 contains additional details which are noninformative for saliency. Forestry category contains 126 images as 42 images in each zoom level while Water surface category contains 120 images as 40 images in each zoom level.

According to high accuracy and precision in recognition of salient regions remotely sensed images were exported from the locations on earth surface providing high figure-background. Forestry and Water Surface categories are selected since the variation at the background is low and salient objects can be detected more precisely by human subjects without using semantic information (Zhao, et al., 2015). Therefore, visual separation is obtained with high-contrast between geographical objects and background of the image in interested location. Forestry and Water Surface categories are selected since the variation at the background is low and salient objects can be detected more precisely by human subjects without using semantic information.

The saliency maps produced by the computational methods on the dataset are compared with the saliency maps extracted from data collected in experiment conducted on human subjects. In these experiments 20 subjects participated and the data is collected by using Tobii T120 Eye Tracker device while the images in the dataset are presented to subjects on computer screen.

In the performance evaluation, the saliency maps obtained from human subjects are used as ground truth. The performances of the computational methods are determined by computing similarity of their results to ground truth. As similarity measure, Pearson correlation, Cosine correlation and SSIM index methods are used for comparisons. Besides Cosine correlation and Pearson correlation methods require low computational cost, they are also mostly used methods for saliency map comparisons. Remotely sensed images contain distortions and irregular features.

SSIM index is a suitable method to compare remotely sensed imagery based saliency maps. Saliency maps produced by the investigated computational models were compared with saliency maps extracted from Tobii T120 Eye Tracker device by each selected similarity measure. Resulting coefficient values were interpreted by mean, standard deviation and coefficient of variation calculations.

Additionally, accuracy measurements were performed to compare computational applications with human based experiment with the dataset containing 226 remotely sensed RGB images and appropriate experimental design.

1.4 Outline of the Thesis

The structure of the thesis is explained below.

Chapter 2 explains Itti-Koch, GBVS, Achanta, ImageSignature, SDSP and CovSal methods. This chapter contains detailed information of these methods and results are shown.

Chapter 3 includes experimental part of the thesis. Experimental design, data collection, outcomes and image processing methods are included in this chapter.

Chapter 4 comprises the result of experiment with the comparison of saliency map implementations in terms of performance measurements by mean, standard deviation and coefficient of variation calculations.

Chapter 5 concludes this study with an interpretation of the findings of the study and theoretical support for those findings.

CHAPTER 2

SALIENCY MAP COMPUTATIONS

Computational visual attention models become more important to understand human vision system by the inspiration of psychological and neurobiological studies. These models are also significant solutions in order to recognize objects, patterns, colors, localizations that are different from background information. Therefore, salient region recognition becomes more interested in computational study fields.

According to needs and requirements many saliency models have been developed. The main goal of the saliency computational models is to extract feature maps by several features and then combine them into a single map to generate a map which is mostly called as saliency map. The extracted features are indicated according to the basics of human visual system as orientation, color, intensity (Carrasco, 2011). Main steps of computational models of salient region detection are feature extraction, generating saliency map by combining feature maps and determining the focus of the attention in saliency map. Moreover, saliency map models based on predictions of human visual system are also defined so that results have better approximate solutions which are able to estimate human visualization patterns.

2.1 Itti-Koch Saliency Map Model

The Itti-Koch model is a biologically-inspired model based on human visual search strategies.It uses feature integration theory explains human visual search strategies (Treisman & Gelade, 1980). The general architecture of Itti-Koch Model is given in Figure 2.1.

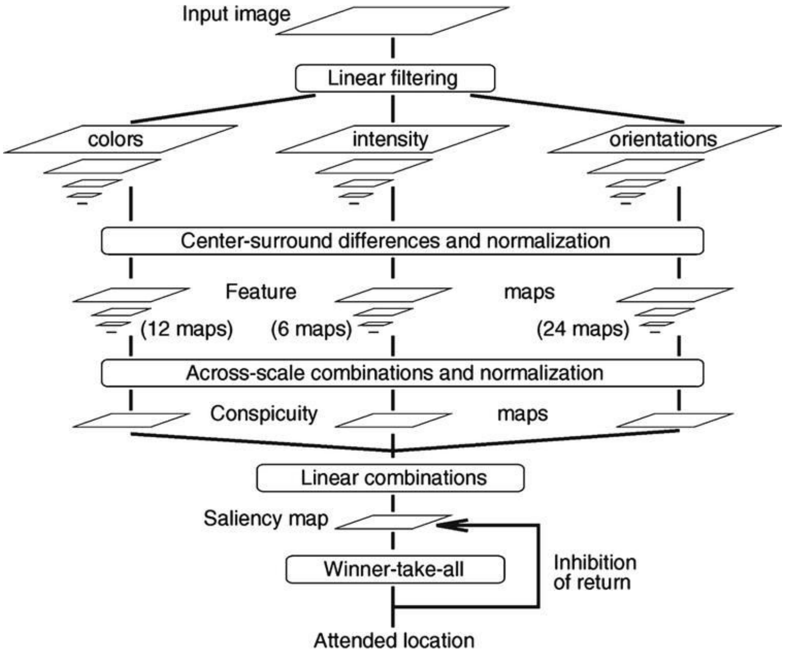


Figure 2.1:General architecture of Itti-Koch model

First step in this model is to decompose the input image into feature maps which are generated based on color, intensity and orientation with different scales. According to neurobiological studies, it is purposed that these are the basic features perceived by human visualization system (Wolfe, 1994; Treisman & Gormican, 1988). For each feature map, Gaussian pyramids are obtained by applying Gaussian filters iteratively in different scales (between zero and eight scale in eight octaves) and the subsamples of the feature maps are generated (Itti, et al., 1998; Itti& Koch, 2001).

According to neurobiological studies, only locations that differ from their surround achieve attention persistently in comparison with neighboring locations.This mechanism is also called as center and surrounds (Cavanaugh, et al., 2002). Each feature map is computed by a center-surround operation to determine contrast.

In order to eliminate amplitude differences in feature maps and to emphasize a small number of regions with strong peaks, normalization is applied so that values of the maps are fixed to a range. In normalization process, maximum activity is compared with overall image. This step measures the difference of the most active locationfrom the average. As the second step of this model, saliency map is computed by combining of all feature maps (Itti, et al., 1998).

2.1.1 Feature Maps

Center-surround implementation is the difference between fine (center) scale and coarse (surround) scale for a given feature.The center is the pixel in scales $c \in \{2,3,4\}$ and the surround is the correspondingpixel at scale$s = c + \delta \,, \delta \in \{3,4\}$. The across-scale difference, denoted as "⊖" below, is obtained by the interpolation to the finer scale and point-by-point subtraction between two feature maps. Itti-Koch model generates nine different spatial scales of feature maps which are Gaussian pyramids in scales$\sigma \epsilon [0 \dots 8]$ denoted as $I\ (\sigma)$.

In order to construct Gaussian pyramids, firstly an intensity image (I) is created by averaging the R,G,B color channels of the input image as given in Equation 2.1.

$$\mathrm{I = (r + g + b)/3} \tag{2.1}$$

Mammals' visual system is much more sensitive to the changes in contrast between dark and bright. For example dark center is surrounded by bright or bright center is surrounded by dark. This is the reason for the creation of intensity feature maps.

Additionally, four color channels Red, Green, Blue, Yellow (Equation 2.2) and their corresponding Gaussian pyramids R(σ), G(σ), B(σ) and Y(σ) are constructed.

$$\mathrm{R(\sigma) = r - (g + b)/2\ (red)}$$

$\mathrm{G}(\sigma) = \mathrm{g} - (\mathrm{r} + \mathrm{b})/2$ (green)

$\mathrm{B}(\sigma) = \mathrm{b} - (\mathrm{r} + \mathrm{g})/2$ (blue)

$\mathrm{Y}(\sigma) = (\mathrm{r} + \mathrm{g})/2 - |\mathrm{r} - \mathrm{g}|/2 - \mathrm{b}$ (yellow) (2.2)

The intensity feature map is constructed by using Equation 2.3for center scales as $c \in \{2,3,4\}$ and surround scales$s = c + \delta, \delta \in \{3,4\}$, where $\ominus$ is the operator calculating center-surround differences.

$\mathrm{I}(\mathrm{c},\mathrm{s}) = |\mathrm{I}(\mathrm{c}) \ominus \mathrm{I}(\mathrm{s})|$ (2.3)

Second set of feature maps are concerned withcolor sensitivity, which, in mammals, neurons in cortex are excited by one color and inhibited another. This is so-called double-color opponent system. For example, if red color is excited then green is inhibited and on the contrary if green is excited then red is inhibited. Such spatial and chromatic opponency exists for the red/green, green/red, blue/yellow, and yellow/blue color pairs in human primary visual cortex (Itti, et al., 1998). Therefore color based feature maps are created as in Equation 2.4 and Equation 2.5.

$\mathrm{RG}(\mathrm{c},\mathrm{s}) = |(\mathrm{R}(\mathrm{c}) - \mathrm{G}(\mathrm{c})) \ominus (\mathrm{G}(\mathrm{s}) - \mathrm{R}(\mathrm{s}))|$ (2.4)

$\mathrm{BY}(\mathrm{c},\mathrm{s}) = |(\mathrm{B}(\mathrm{c}) - \mathrm{Y}(\mathrm{c})) \ominus (\mathrm{Y}(\mathrm{s}) - \mathrm{B}(\mathrm{s}))|$ (2.5)

The third set of feature maps are related to orientation. Since Gabor filters are the approximately model the receptive field sensitivity of mammals to orientation, Gabor pyramids, $O(\sigma, \theta)$ in scale of $\sigma \in [0 \ldots 8]$ and the orientation of

$\theta \in \{0{,}45\ {,}90\ {,}135\}$ are used in order to obtain local orientation information (Itti, et al., 1998). As it is shown in Equation 2.6, orientation based feature maps are created.

$$\mathrm{O(c, s, \theta)\ =\ |O(c, \theta)\ \ominus\ O(s, \theta)|} \tag{2.6}$$

As a result of these steps, 6 feature maps for intensity, 12 for color and 24 for orientation, totally 42 feature mapsare generated.

2.1.2 Combination of Feature Maps

The combination of feature maps constitutes the saliency map. Before the summation of feature maps, they are performed with a normalization operator to compare responses associated with meaningfulactivation spots in the map and to ignore homogenous areas. Difference between global maximum activity and average activation defines the promotion level of the location. Higher difference indicates the corresponding location stands out and the map is strongly promoted.

Three “conspicuity maps” are created with the scale $\sigma = 4$. They are obtained by across-scale addition, denoted by "$\oplus$", as shown in Equations 2.7, 2.8 and 2.9, where $\mathcal{N}$ is the normalization operator consist of the following;

1. Normalizing the values in the map to a fixed range $[0..M]$, in order to eliminate modality-dependent amplitude differences;
2. finding global maximum M of each map and computing the average $\overline{m}$of all its other local maxima; and
3. globally multiplying the map by $(\ M - \overline{m})^2$.

$$\bar{I} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c=4} \mathcal{N}\left(I(c,s)\right) \tag{2.7}$$

$$\bar{C} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} \left[\mathcal{N}\left(RG(c,s)\right) + \mathcal{N}\left(BY(c,s)\right)\right] \tag{2.8}$$

$$\bar{O} = \sum\nolimits_{\theta \in \{0^\circ,45^\circ,90^\circ,135^\circ\}} \mathcal{N}\left(\bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} \mathcal{N}\left(O(c,s,\theta)\right)\right) \tag{2.9}$$

Motivation of separated conspicuity maps is that similar features represent stronger saliency while discrete or unrelated features contribute independently to the saliency map. After normalization stepof conspicuity maps, Saliency Map is generated by the summation of them into single map as given in Equation 2.10.

$$S = \frac{1}{3}\left(\mathcal{N}(\bar{I}) + \mathcal{N}(\bar{C}) + \mathcal{N}(\bar{O})\right) \tag{2.10}$$

2.1.3 Winner-Take-All Computation (WTA)

Even if the computed Saliency Map already highlights the most attended region in input image, Itti-Koch model finds locations of salient regions by a winner-take-all(WTA) network. Motivation of WTA is biological and related to visualization mechanism in human brain. Neurons compete with each other for activation. While a neuron is with highest activation, another neuron’s activation reduces but does not vanish. Therefore, as it is introduced in Lee, et al., 1999, WTA is applied to saliency map of the input image in order to find the location of focus of attention.

2.1.4 MATLAB Code for Itti-Koch Saliency Map Model

The MATLAB code for Itti-Koch model is implemented by Caltech and available at http://www.vision.caltech.edu/~harel/share/GBVS.php.This implementation regarding the Itti-Koch saliency map model outputs saliency map and heat map as a layer on the input image. The sample output is given in Figure 2.2. In order to speed up computation, Saliency Map Toolbox creates totally 14 feature maps, 2 of them being for intensity, 4 for color and 8 for orientation while standard Itti-Koch model execution. Additionally, final blur is applied to master saliency map to improve eye movement predictions.

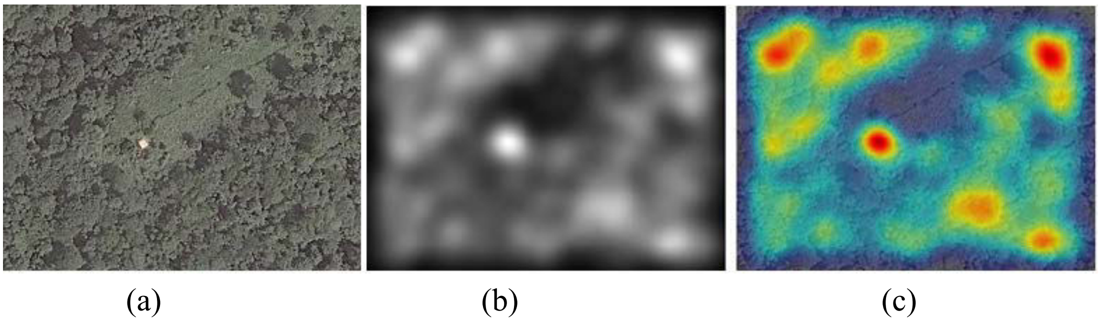

(a) (b) (c)

Figure 2.2: (a) Original remotely sensed image(b) Itti-Koch saliency map (c) Itti-Koch saliency map output as a heat map layer on the original remotely sensed image

2.2 Graph-Based Visual Saliency (GBVS)

GBVS presents a method based on topological structure to find saliency map. The process is divided into three stages; extraction of features to feature vectors, forming activation maps from feature vectors and normalization and combination of activation maps into a single saliency map. Markov chain approach is interpreted in both forming activation maps and normalization of them (Biskupska, 2013).

For a given feature map, each pixel is considered as a node. By connecting every node in the given feature map, fully-connected directed graph is created. The directed edges between the nodes are assigned to a weight. Dissimilarity is proportional with the weight of the edge between the nodes. The weights are then normalized to 1. The result is the activation map.

In normalization phase, the activation map is performed with Markov chain algorithm. Directed edges between the nodes in activation map are assigned to weights. Again, the weights are normalized to 1. The goal of normalizing the activation map is concentrating mass on activation maps. Normalized activation maps are combined into saliency map (Harel, et al., 2006).

2.2.1 Activation Maps

Dissimilarity measurement between pixel $M(i,j)$ and pixel $M(p,q)$ as defined in Equation 2.11.

$$d\big((i,j) \parallel (p,q)\big) \triangleq \left|\log \frac{M\,(i,j)}{M\,(p,q)}\right| \tag{2.11}$$

where $\triangleq$ refers to difference equality (delta equal to) between the pixels denoted as $M(i,j)$ and $M(p,q)$ while (i,j) and (p,q) are the locations of the pixels in the given feature map M respectively.

A fully-connected graph G_a is obtained by connecting every pixel of the given feature map Mwhich is same in Itti-Koch model. In G_a, the directed edge from pixel $M(i,j)$to pixel $M(p,q)$is assigned to a weight defined in Equation 2.12 below.

$$w\,\big((i,j),(p,q)\big) \triangleq d\big((i,j) \parallel (p,q)\big)\,.\,F(i-p,j-q) \tag{2.12}$$

where $F\ (a,b)\ \triangleq \exp(-\ \frac{a^2+b^2}{2\sigma^2}\ )$ and σ is a free parameter used in GBVS algorithm.

By normalizing the weights to 1 with and used in Markov Chain.Equilibrium distribution of Markov chain forms the activation map A. The resulting activation map A shows that the mass flows to the nodeshave high dissimilarity with their surrounding nodes.

2.2.2 Normalization of Activation Map

After activation map A is generated,anotherfully-connected graph, $G_n$is constructed for normalization of activation map. The weights of the directed edges between pixel $A(i,j)$and pixel $A(p,q)$in $G_n$are calculated by the Equation 2.13.

$$w\left((i,j),(p,q)\right)\ \triangleq\ \ A(p,q)\ .F(i-p,j-q) \tag{2.13}$$

where $A(p,q)$ is the pixel value at position (p,q) in activation map A

Again, weights are normalized to 1 and resulting graph is a Markov Chain that allows computing equilibrium distribution as a basis to create output saliency map (Harel, et al., 2006; Biskupska, 2013).

2.2.3 GBVS Saliency Map

The processes explained in section 2.2.1 are applied to each given feature map. After activation maps are generated, they are performed with the processes given in section 2.2.2 seperately. Each normalized activation map is combined into saliency map.

The GBVS saliency map contains values between zero and one. The value closer to one, indicates more importance in GBVS saliency map while the closer pixel value to zero is less important.

2.2.4 MATLAB Code for GBVS

GBVS model implementation for MATLAB is also available atwww.vision.caltech.edu/~harel/share/GBVS.php.Sample input image, output GBVS saliency map and layered GBVS saliency map on input image as heat map can be seen in Figure 2.3below. GBVS saliency map is in gray scale includes pixel values between zero and one. Gaussian blur kernel applied to final saliency map so that accuracy of GBVS saliency algorithm is improved. Related process is given with details in Greco & La Cascia, 2013.

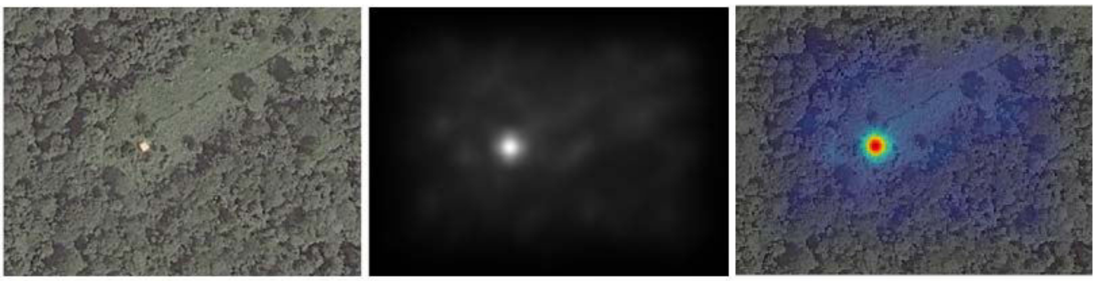

(a) (b) (c)

Figure 2.3: (a) Sample remotely sensed input image (b) GBVS saliency map (c) GBVS saliency map as a heat map layer on the original remotely sensed image

2.3 Frequency-tuned Salient Region Detection (Achanta)

In Achanta model, a method is used to find salient regions with high resolution, well-defined edges and efficient computation. This algorithm estimates center-surround contrast by using color and luminance featuresbased on a frequency-tuned approach. Spatial frequencies are investigated by using an approach motivated by different saliency detection models and Difference of Gaussian (DoG) on color and luminance features to generate saliency map. Table 2.1 indicates the requirements for a saliency detector.

Table 2.1: Requirements for saliency map generation in Achanta

Target	Requirement
Emphasize the largest salient objects.	Depending on very low frequencies from input image
Uniformly highlight whole salient regions	Depending on very low frequencies from input image
Establish well-defined boundaries of salient objects	Retaining high-frequencies from input image
Disregard high frequencies arising from texture, noise and blocking artifacts.	Difference between arithmetic mean pixel value of input image and Gaussian blurred version of input image.
Efficiently output full resolution saliency maps	Operating without down sampling

In order to obtain first four requirements in Table 2.1, a wide range of frequenciesare investigated in this saliency model. Low frequencies (denoted by ω_{lc}) and high frequencies (denoted by ω_{hc}) are used. Hence pass bands are determined as$[\omega_{lc},\omega_{hc}]$. Additionally, any down-sampling is not operated with the input image. Hence, the last requirement in Table 2.1 is also obtained (Achanta, et al., 2009).

2.3.1 Binomial Filter

Instead of applying a set of continuous DoG filters, Bionomial filters are applied to the given image (Achanta, et al., 2009; Hatipoglu, et al., 2014).

Binomial filter simulates low pass filter (Li & Gao, 2014). It can be used to eliminate noise and texture caused by high frequencies and to obtain computational simplicity the 5 X 5 Separable Binomial filter ($\omega_{hc} = \pi/2.75$) given in Equation 2.14 is used in

Achanta, et al., 2009; Li & Gao, 2014; Preim & Botha, 2013.

$$\frac{1}{16}[1,4,6,4,1] \rightarrow \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix} \quad (2.14)$$

2.3.2 Computing Saliency Map

To compute saliency map of Achanta model, firstly, the input image is blurred with a [3 x 3] Gaussian filter. Then, it is converted to Lab color space and decomposed into the bands which are L (lightness), a (green-red information) and b (blue-yellow information). Each band is filtered by the binomial filter given in Equation 2.14 in section 2.3.1.

Difference between the binomial filtered bands and the mean of the bands ofthe Gaussian blurred input image is the basis to compute saliency map. For the input image I, resulting saliency map $S(I)$ is computed by using Equation 2.15.

$$S(I)= \left(L_{\omega_{hc}} - L_{\mu}\right)^2 + \left(a_{\omega_{hc}} - a_{\mu}\right)^2 + \left(b_{\omega_{hc}} - b_{\mu}\right)^2 \quad (2.15)$$

where $L_{\omega_{hc}}$, $a_{\omega_{hc}}$ and $b_{\omega_{hc}}$ are the binomial filtered bands and L_{μ}, a_{μ} and $b_{\mu}$are the means of the bands in Lab color space.

2.3.3 MATLAB Code for Achanta Saliency Map Method

The MATLAB code of Achanta saliency map method is available at http://ivrlwww.epfl.ch/supplementary_material/RK_CVPR09/index.html. It is implemented by IVRG - Images and Visual Representation Group. In the MATLAB implementation, binomial filter is not applied to bands of the Gaussian blurred input image in Lab color space. The bands are only converted to gray scale. Then, mean of each band is calculated. Sample outputs obtained by the Achanta method are provided in Figure 2.4.

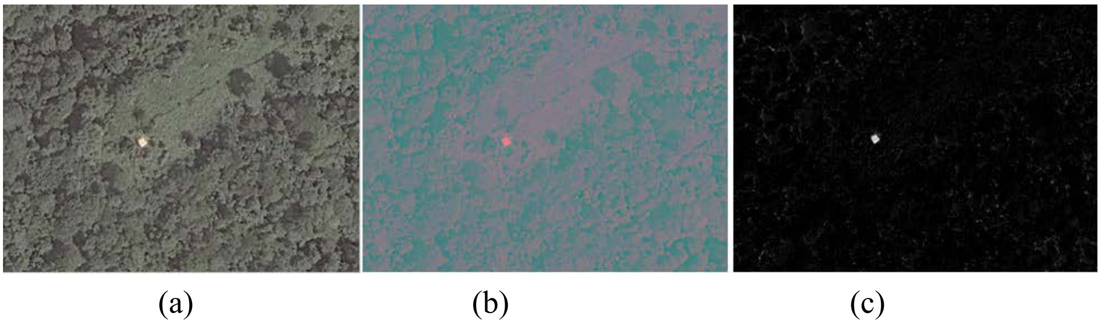


Figure 2.4: (a) Original input image (b) Gaussian blurred input image in Lab color space (c) Final saliency map of Achanta model

2.4 Saliency Map by Image Signature

Sign function of the Discrete Cosine Transform (DCT) of an image is defined as a simple image descriptor which contains information about the foreground of an image. In Image Signature model, this approach is used for the solution of Figure-ground separation problem. By using Inverse Discrete Cosine Transform (IDCT) of the image signature, it is proven that the image energy at the locations of a spatially sparse foreground, relative to a spectrally sparse background is concentrated (Hou, et al., 2012).

2.4.1 Discrete Cosine Transform (DCT) Approach

In this model firstly, input image is decomposed into the bands in RGB and Lab color spaces. Then, each band is transformed by DCT. By taking account the sign of the input image bands in transformed domain, positive DCT coefficient points are selected (Hou, et al., 2012; Hatipoglu, et al., 2014).

Signum function, denoted by $sign(x)$, is 1 if the corresponding element of x is greater than zero, 0 if the corresponding element of x equals zero, -1 if the corresponding element of x is less than zero.

In Image Signature model, an image is represented as the combination of foreground information, denoted by f, and background information, denoted by b. It is assumed that $f$sparsely supported in the standard spatial basis and $b$sparsely supported in the basisof the Discrete Cosine Transform. Let I_r, I_g and I_b be the red, green and blue channels of the input image where $c = I_r, I_g, I_r$. By taking into account separating $f$and b, Image Signature of an input image $I_c$is formed by given Equation 2.16.

$$ImageSignature\ (I_c) = sign(DCT(I_c)) \tag{2.16}$$

where,DCT(I_c) is Discrete Cosine Transform of the RGB color channels of the input image I_c while $sign(DCT(I_c))$ corresponds to Signum function of Discrete Cosine Transform applied input image I_c.

The foreground information is then calculated by Inverse Discrete Cosine Transform is performed with $ImageSignature(I_c)$ (Equation 2.16) to calculate foreground of the input image I_c by Equation 2.17.

$$\overline{I_c} = IDCT\left(sign\big(DCT(I_c)\big)\right) \tag{2.17}$$

where, $\overline{I_c}$ is the reconstructed image containing foreground of the input image I_c in spatial domain, $IDCT$denotes Inverse Discrete Transform.

2.4.2 Saliency Map

For the final saliency map, the processes explained in section 2.4.1 are applied to each band. The reconstructed images of the separated bands are squared and combined. This process is denoted by $(\overline{RGB}\, \text{O}\, \overline{RGB})$for RGB color space and $\left(\overline{Lab}\, \text{O}\, \overline{Lab}\right)$ for Lab color space. The Gaussian blur is applied to the combined image to produce final saliency map by using Equations 2.18 and 2.19.

$$S_{RGB}(I) = g * (\overline{RGB}\text{O}\, \overline{RGB}) \tag{2.18}$$

$$S_{Lab}(I) = g * \left(\overline{Lab}\text{O}\, \overline{Lab}\right) \tag{2.19}$$

where $\overline{RGB}$ and $\overline{Lab}$ represent the combination of reconstructed band images in spatial domain of RGB and Lab color spaces respectively. $S_{RGB}(I)$ is saliency map for RGB color space while$S_{Lab}(I)$is saliency map for Lab color space. g indicates Gaussian filter, $*$ is the convolution operator and O is the Hadamard (entry wise) product operator.

2.4.3 MATLAB Code for Image Signature Saliency Map Model

The MATLAB implementation of Image Signature saliency method is developed by MIT Saliency Benchmark and the MATLAB source code of this model is available athttp://saliency.mit.edu/results_old.html. It produces two final saliency maps as in Figure 2.5 below.

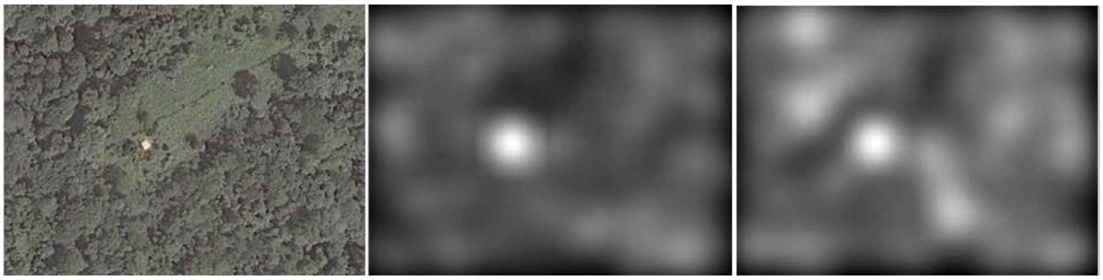


Figure 2.5: (a) Sample remotely sensed input image (b) Saliency map output for RGB color space (b) Saliency map output for Lab color space

2.5 Saliency Detection by Combining Simple Priors (SDSP)

SDSP algorithm is consist of three simple priors which are frequency, color and location. By combination of these priors' separate saliency maps, the final SDSP saliency map is generated (Zhang, et al., 2013).

2.5.1 Frequency Prior

As it is explained in Achanta, et al., 2009, to highlight salient regions and emphasize salient objects a band-pass filtering method is used similar to detection of salient region in human visual system. Therefore, log-Gabor filter, denoted as $g(\mathbf{x})$, is used. This filter isexpressed approximately in frequency domain as $G(\mathbf{u})$ as given in Equation 2.20, where $\mathbf{x} = (x, y)$ refers to the location on the input image and $\mathbf{u} = (u, v)$ refers to the location in frequency domain.

$$G(\boldsymbol{u}) = \exp\left(-\frac{\left(\log\frac{\|\mathbf{u}\|_2}{\omega_0}\right)^2}{2\sigma_F^2}\right) \tag{2.20}$$

where ω_0 is the center frequency of the filter and σ_F is the filter's bandwidth.

The input image I is given in RGB. Firstly, RGB image is converted to Lab space and three channels are defined $f_L(I), f_a(I)$and $f_b(I)$. The frequency prior saliency map, denoted by$S_F(I)$, is calculated by using Equation 2.21.

$$S_F(I) = ((f_L * g)^2 + (f_a * g)^2 + (f_b * g)^2)^{1/2} \tag{2.21}$$

where $*$ is convolution operator and g is the log-Gabor filter.

2.5.2 Color Prior

It is proposed in Zhang, et al., 2013 that people pay more attention in warm colors rather than cold colors. As it is explained in Section 2.3.1, the given input image is converted to Lab space. Where a-channelrefers to green-red information while b-channel represents blue-yellow information. If a pixel has a higher value in a or b, it would provide warmer color information. For example, higher value in a-channel, it would seem greenish or higher value in b-channel, it would seem bluish. L-channel is for lightness.

Color saliency map is calculated pixelwise by using Equation 2.22.

$$f_{an}(I) = \frac{f_a(I) - mina}{maxa - mina}\ ,\ f_{bn}(I) = \frac{f_b(I) - minb}{maxb - minb} \tag{2.22}$$

where I is the given input image, $f_a(I)$ and $f_b(I)$ is in $[0,1]$ range. $maxa$ and $maxb$ refer to maximum values of input image while $mina$ and $minb$ refer to minimum values.

Color saliency map of a point for a given image is defined as given in Equation 2.23where σ_c is a parameter for bias correction (Li, et al., 2011). Since the warm colors have a bias towards attention (Zhang, et al., 2013).

$$S_c(I) = 1 - exp\left(-\frac{f_{an}^2\,(I) + f_{bn}^2\,(I)}{\sigma_c^2}\right) \tag{2.23}$$

2.5.3 Location Prior

Another prior for SDSP algorithm is location which implies that people pay more attention on the center of an image. Location saliency is expressed as a Gaussian map as given in Equation 2.24. Since the central areas will have a bias, for the accuray bias is corrected by σ_D parameter (Li, et al., 2011; Zhang, et al., 2013).

$$S_D(I) = exp\,\left(-\frac{\|I - c\|_2^2}{\sigma_D^2}\right) \tag{2.24}$$

where $\sigma_D$is a parameter for bias correction

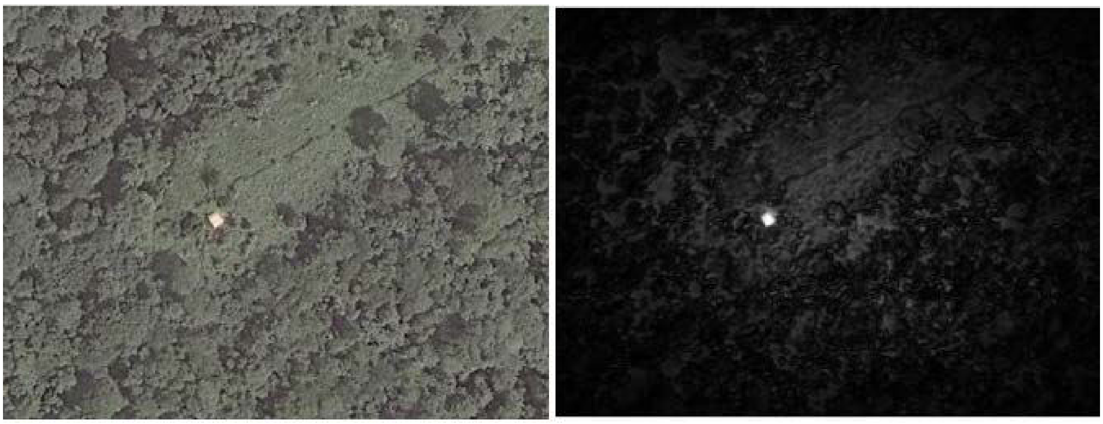
2.5.4 SDSP Saliency Map

SDSP saliency map is generated by combining prior saliency maps that are $S_F(I)$, $S_D(I)$ and $S_C(I)$ using Equation 2.25 below.

$$SDSP\,(I) = S_F(I).S_D(I)\,.S_C(I) \tag{2.25}$$

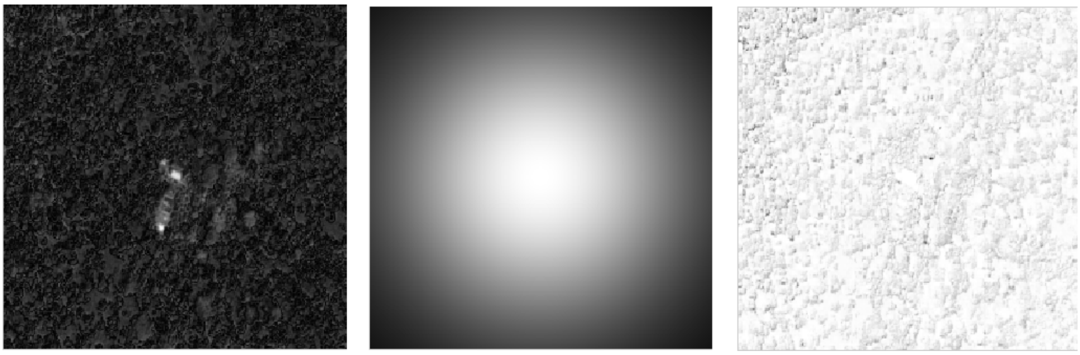
2.5.5 MATLAB Code for SDSP Saliency Map

The MATLAB code for SDSP model is implemented by Tongji University available athttp://sse.tongji.edu.cn/linzhang/va/SDSP/SDSP.htm. It provides each prior based saliency maps and the combination of them as the final saliency map. Sample outputs of SDSP model are given in Figures 2.6 and 2.7.



(a) (b)

Figure 2.6: (a) Sample original remotely sensed input image (b) Sample output of final SDSP saliency map



(a) (b) (c)

Figure 2.7: (a) Sample frequency prior saliency map (b) Sample location prior saliency map (c) Sample color prior saliency map

2.6 The CovSal Saliency Model

This model firstly decomposes given input image based on color, orientation and spatial feature points. Then, non-overlapping regions are extracted from the feature points. The covariances of these non-overlapping regions are calculated by examining the surrounding regions. The non-overlapping regions with similar characteristics have similar covariances while the non-overlapping regions with dissimilar characteristics have dissimilar descriptors. Therefore, dissimilar regions to their neighboring regions represent salient areas or object in the given input image (Erdem & Erdem, 2013).

2.6.1 Region Covariances

CovSal computational saliency model uses seven-dimensional feature vector calculated considering color, orientation and location. Formal definition for the feature vector $F\ (x, y)$is given in Equation 2.26.

$$F\ (x,y) = \left[L(x,y)\ \ a(x,y)\ \ b(x,y) \left|\frac{\partial\ I(x,y)}{\partial x}\right| \left|\frac{\partial\ I(x,y)}{\partial y}\right|\ x\ y\right]^T \tag{2.26}$$

where a and b are the values of the pixel (x, y)in Lab color space. $\left|\frac{\partial\ I}{\partial x}\right|, \left|\frac{\partial\ I}{\partial y}\right|$are the edge orientation information of the given input image I while (x, y) denotes the pixel location.

A region R in the feature image extracted from the input image (F) can be represented by a$d\ x\ d$ covariance matrixas given in Equation 2.27.

$$C_R = \frac{1}{n-1} \sum_{i=1}^{n} (f_i - \mu)\ (f_i - \mu)^T \tag{2.27}$$

where $\{f_i\}_{i=1\ldots n}$ denotes d-dimensional feature points in region R and μ is mean of these points. In CovSal $n = 7$ due to feature vector size.

CovSal model uses Equation 2.28 in order to calculate distance between two covariance matrix (Erdem & Erdem, 2013).

$$\rho\ (C_1\ , C_2) = \sqrt{\sum_{i=1}^{n} ln^2 \lambda_i (C_1, C_2)} \tag{2.28}$$

where $\{\lambda_i(C_1, C_2)\}_{i=1\ldots n}$ represents generalized eigenvalues and $\{x_i\}_{i=1\ldots n}$ are eigenvectors of the covariance matrixes C_1 and C_2. $n = 7$ due to the CovSal model uses 7-dimensional feature vector.

It is indicated in Erdem & Erdem, 2013, the first-order statics corresponding to the mean vector of the features is included to distinguishing between two different distributions of features.

Let C be a dxd covariance matrix, related Sigma Points $S = \{s_i\}$ which can be computed considering Cholesky decomposition by using the Equation 2.29.

$$s_i = \begin{cases} \alpha\sqrt{dL_i} & if\ 1 \leq i \leq d \\ -\alpha\sqrt{dL_i} & if\ d + 1 \leq i \leq 2d \end{cases} \tag{2.29}$$

whereL_i is the ith column of the lower triangular matrix L obtained with Cholesky decomposition ($C = LL^T$). $d = 7$due to the feature vector size used in the CovSal model and α is a parameter. The CovSal model takes $\alpha = \sqrt{2}$.

By the motivation of Sigma Points (Julier & Uhlmann, 1996), mean vector of the features are concatenated to the previous 7-dimensional feature vector. This enriched feature vector is denoted as $\psi(C)$ given in Equation 2.30.

$$\psi(C) = (\mu, s_1 , \dots \ s_d \ , , s_{d+1}, \dots , s_{2d} \ )^T \tag{2.30}$$

where s_i represents sigma points and μ is the mean vector of the features.

2.6.2 Model 1: Saliency Using Covariance Features (CovSal-C)

For the first saliency map output of the CovSal, the feature covariances of the non-overlapping regions are used only. In Model 1, firstly, the given input image I is reshaped to square form. It is then decomposed to square non-overlapping regions which are size of $k \ x \ k$ pixels.The saliency of a square non-overlapping region is calculated by comparing it with its neighboring the non-overlapping regions. If it displays distinctive characteristics locally, it is defined as salient.

Let R_i be the region under consideration and $\{R_j\}$ be the regions in the given radius of r. The dissimilarity measurement $d(R_i, R_j)$between the region R_i and the regions $\{R_j\}$ is defined by given Equation 2.31.

$$d(R_i, R_j) = \frac{\rho \ (C_1 \ ,C_2)}{1+ \|x_i - x_j\|} \tag{2.31}$$

where C_1 and C_2 are the covariance matrices, x_i and x_j denoting the image coordinates of the center of the regions R_i and R_j, respectively.

More formally, the saliency of the region R_i is given by Equation 2.32.

$$S\ (R_i) = \frac{1}{m}\sum_{j=1}^{m} d\ (R_i, R_j) \quad (2.32)$$

where m refers to most similar regions around of R_i is found considering to the dissimilarity measure $d(R_i, R_j)$.

It is noted that in Erdem & Erdem, 2013the region size k determines the resolution of the saliency map. Therefore, the computed saliency maps of the non-overlapping regions are resized to obtain the final saliency map at the resolution of the original given input image I.

2.6.3 Model 2: Saliency Using Covariance and Mean Features (CovSal-CM)

For the second saliency map output of the CovSal model, mean information of the features are incorporated into covariance-based model explained in section 2.6.2. The dissimilarity measurement $d^{'}(R_i, R_j)$ between the region R_i and the neighboring regions $\{R_j\}$ is defined by given Equation 2.33.

$$d'(R_i, R_j) = \frac{\|\psi(C_1)-\psi(C_2)\|}{1+\|x_i - x_j\|} \quad (2.33)$$

where$\psi\ (C_1)$ and $\psi\ (C_2)$ are the enriched feature vectors of the covariance matrixes C_1 and $C_2$with the incorporated the mean vector of the features while x_i and x_j denoting the image coordinates of the center of the regions R_i and R_j, respectively.

Saliency map of the region R_i based on covariance and mean calculations is given by Equation 2.34.

$$S\ (R_i) = \frac{1}{m}\sum_{j=1}^{m} d'\ (R_i, R_j) \quad (2.30)$$

where m denotes the most similar regions around of R_i is found considering to the dissimilarity measure $d'(R_i, R_j)$.

Again, at scale k, the computed saliency maps of the non-overlapping regions are resized.

2.6.4 MATLAB Code for CovSal Saliency Map

The MATLAB code for the CovSal model is implemented by Hacettepe University Department of Computer Engineering and is available at http://web.cs.hacettepe.edu.tr/~erkut/projects/CovSal/, last visited on May 2016. It produces two saliency map. The first saliency map is the output for the first model based on region covariances only while the second saliency map is the output of the second model performed according to region covariances and means. The sample outputs of the CovSal model are given in Figure 2.8.

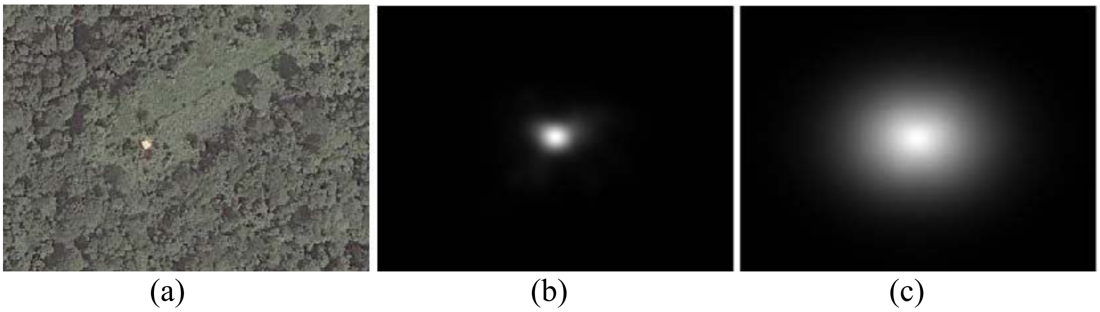


Figure 2.8: (a) Sample original input image (a) Sample saliency map of Model 1: Saliency using covariance features only (b) Sample saliency map of Model 2: Saliency using covariance and mean features

2.7 Similarity Measurements

In the following sub-sections, details of image comparison methods used in this thesis are explained. These methods are used to calculate similarities between the saliency map of Tobii T120 Eye Tracker and the saliency maps of computational models executed in MATLAB.

2.7.1 Normalization and Vector Form of Saliency Maps

For similarity measurements, the images should be coverted to vector forms. The saliency maps are grey level images repsented in two dimensional matrixes while vector forms of saliency maps are one dimensional arrays. A pixel (x, y) in an image of size $n_x \times n_y$ is repsented by index i in the vector form as given in Equation 2.31.

$$i = (y - 1)n_y + x \tag{2.31}$$

where $i = 1 \dots n$ and $n = n_x n_y$

In order to reduce noise caused by outliers and to improve interpretability of information in saliency maps, a normalization method is applied to each saliency map as described in Equation 2.32.

$$N_i = I_i \,/\, \sum_{i=1 \dots n} I_i \tag{2.32}$$

whereI_i is the pixel value at location i and$N(x, y)$denotes the normalized input image.

The input images are saliency maps, which are grey level images, each pixel having value between 0 and 1. By computing each saliency map with Equation 2.32, sum of all pixel values of Tobii T120, GBVS, Itti-Koch, Achanta, ImageSignature, SDSP and CovSal saliency maps are equal to 1 individually.

2.7.2 The Structural SIMilarity (SSIM) Index

Digital images are subject to a wide range of distortions during acquisition, transmission, reproduction and processing. In consideration with error sensitivity of image signals, Structural Similarity Index method is developed based on luminance, contrast and structural terms (Wang, et al., 2004).

The comparisons between input images X and Y for luminance (l), contrast (c) and structural terms (s) are defined as in Equations 2.33, 2.34 and 2.35 below.

$$l\,(X,Y) = \frac{2\,\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1} \tag{2.33}$$

Luminance comparison function, with constant C_1

$$c\,(X,Y) = \frac{2\,\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2} \tag{2.34}$$

Contrast comparison function, with constant C_2

$$s\,(X,Y) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3} \tag{2.35}$$

Structural comparison function, with constant C_3

where μ_X and μ_Y are means, σ_X and σ_Y refer to standard deviations of the input images X and Yrespectively while σ_{XY} denotes cross-covariance for input images X and Y.

By the combination of these three comparisons, structural similarity index between images $X$and Y forms as in Equation 2.36;

$$SSIM\ (X,Y) = [l\ (X,Y)]^{\alpha}[c\ (X,Y)]^{\beta}[s\ (X,Y)]^{\gamma} \tag{2.36}$$

Parameters > 0 , $\beta > 0$ and $\gamma > 0$ are used to adjust relative importance for three components.

In order to simplify expression of structural similarity index of two input images, following values are used as $\alpha = \beta = \gamma = 1$ and $C_3 = \frac{C_2}{2}$. Simplified structural similarity index, denoted as $SSIM\ (X,Y)$, between two input images X and Y is given by Equation 2.37.

$$SSIM\ (X,Y) = \frac{(2\ \mu_X\mu_Y + C_1)\ (\ 2\sigma_{XY} + C_2)}{(\ \mu_X^2 + \mu_Y^2 + C_1)\ (\sigma_X^2 + \sigma_Y^2 + C_2)} \tag{2.37}$$

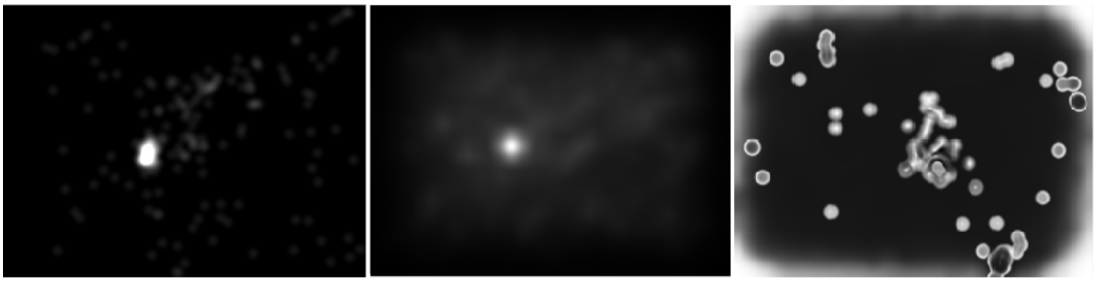
where$\mu_x$and μ_y are means, σ_x and σ_y refer to standard deviations and σ_{xy} denotes cross-covariance for input images$X$ and Y. C_1andC_2 are constant values.

SSIM map, denoted by S_ssim, obtained by given Equation 2.38.

$$S\_ssim = \frac{(2\ g(X)g(Y)+C_1)(2\ (g(XY)-g(X)g(Y))+C_2)}{(g(X)^2+\ g(Y)^2+\ C_1)((g(X^2)-g(X)^2)+(g(Y^2)-\ g(Y)^2)+C_2)} \tag{2.38}$$

where g is the Gaussian filter with size 11 × 11 and sigma=1.5. $g(X)$ and $g(Y)$ are the filtered input images X and Y respectively. $C_1$and C_2 are constant values.

Each raw saliency map of the computational saliency models as GBVS, Itti-Koch, Achanta, ImageSignature, SDSP, CovSal is compared with the raw saliency map of Tobii T120 by computing Equations 2.37 and 2.38. In MATLAB, the corresponding SSIM function has two outputs as SSIM index value and SSIM map. SSIM value is in the range between 0 and 1. 0 value means that compared images are completely different from each other, while value 1means they are exactly same. Sample SSIM map showing the SSIM coefficients is given by Figure 2.9.



(a) (b) (c)

Figure 2.9: (a) Raw saliency map of Tobii T120 (b) Raw saliency map of GBVS (c) SSIM map displaying visual difference between (a) and (b)

2.7.3 Cosine Similarity

Cosine similarity method is used to quantify the similarity between two input images which are normalized saliency maps obtained from the computational saliency models and Tobii T120 Eye Tracker device.

Cosine similarity coefficients between input images X and Y, denoted as $C\ (X,Y)$, were obtained according to the vector form of input images X and Y by given Equation 2.39 (Manning, et al., 2009);

$$C\ (X,Y) = \frac{\sum_{i=1}^{n} X_i Y_i}{\sqrt{\sum_{i=1}^{n} X_i^2}\sqrt{\sum_{i=1}^{n} Y_i^2}} \qquad (2.39)$$

where X_i and Y_i are the pixel values of input images X and Y.

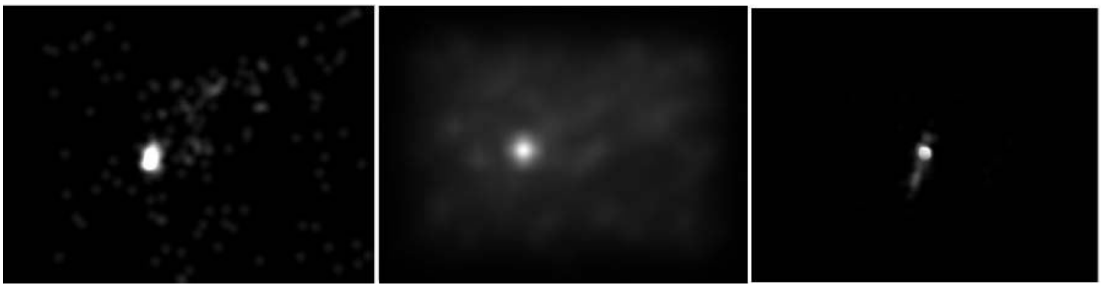
The closer value to 1 for cosine of the angle between compared saliency maps means higher similarity while value 0 indicates complete dissimilarity (Manning, et al., 2009).

The pixel value at location i of Cosine similarity map, denoted as S_cos_i, is calculated by given Equation 2.40.

$$S\_cos_i = \frac{X_i Y_i}{\sqrt{\sum_{i=1}^{n} X_i^2}\sqrt{\sum_{i=1}^{n} Y_i^2}} \qquad (2.40)$$

where X_i and Y_i are the pixel values at location i of input images X and Y.

As for SSIM measurement described in previous section, corresponding Cosine similarity function produces both Cosine similarity coefficient and Cosine similarity map of the compared saliency maps in MATLAB. Sample visual output of Cosine similarity measurement is given by Figure 2.10.



(a) (b) (c)

Figure 2.10: (a) Normalized Tobii T120 saliency map (b) Normalized GBVS saliency map (c) Cosine similarity map displaying visual difference between (a) and (b)

2.7.4 Pearson Correlation Similarity

Pearson correlation similarity is another method to measure the similarity between raw saliency maps of computational models and Tobii T120. Pearson correlation coefficient, denoted by $P(X,Y)$, is obtained by given Equation 2.41;

$$P\ (X,Y) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(X_i - \mu_X)}{\sigma_X}\right)\left(\frac{(Y_i - \mu_y)}{\sigma_Y}\right) \tag{2.41}$$

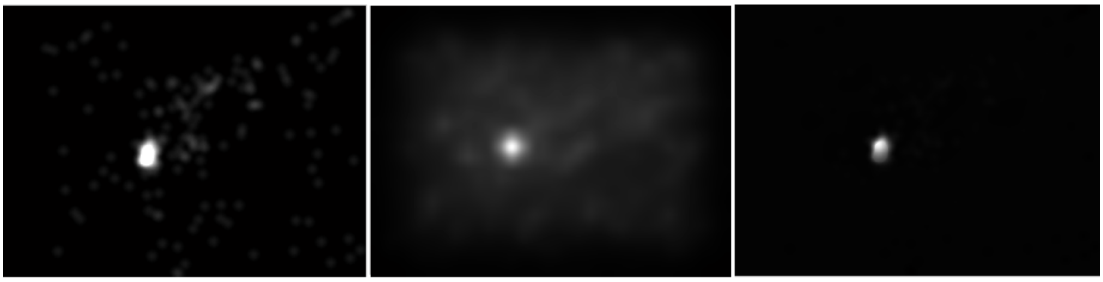
where X and Y are the input images while X_i and Y_i are the pixel values at location i and n denotes the total number of the pixels. μ_X, μ_y, σ_X and σ_Y are the mean and standard deviations of the input images X and Y respectively.

Pearson similarity map, denoted as S_p_i, is calculated by given Equation 2.42.

$$S\_p_i = \left(\frac{(X_i - \mu_X)}{\sigma_X}\right)\left(\frac{(Y_i - \mu_y)}{\sigma_Y}\right) \tag{2.42}$$

where $\mu_X$and μ_y are the means values and σ_X and σ_Y are the standard deviations of the input images X and Y respectively while X_i and Y_i are the pixel values at location i.

By computing Equations 2.41 and 2.42 in MATLAB, Pearson correlation coefficients and Pearson similarity maps for the compared saliency maps are obtained. Sample visual output of Pearson similarity measurement is given by Figure 2.11.



(a) (b) (c)

Figure 2.11: (a) Normalized Tobii T120 saliency map (b) Normalized GBVS saliency map (c) Pearson similarity map displaying visual difference between (a) and (b)

CHAPTER 3

EYE TRACKING EXPERIMENTS

In this study, to obtain ground truth information for comparing the computational saliency map models, an eye tracking experiment is conducted. In the following sub-sections, detailed information about the dataset and the experiment processes are explained.

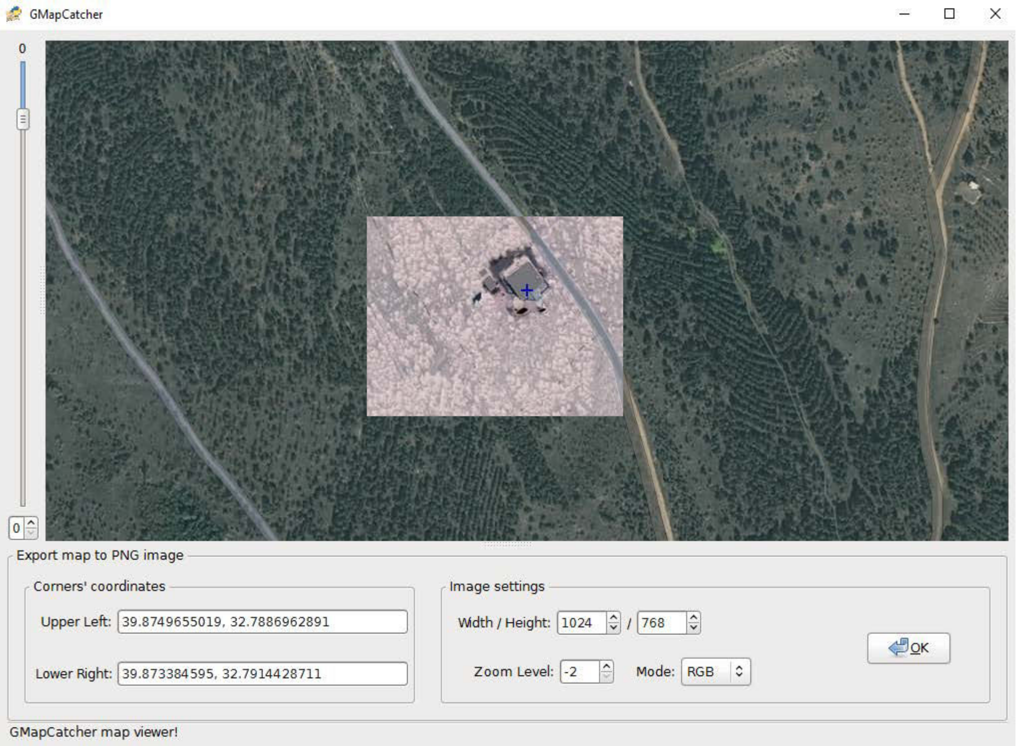
3.1 The Dataset

The regions on earth surface provide higher figure-background separation are selected for the data set. Two remotely sensed image categories, which are forestry and water surface, are considered. Total dataset contains 226 remotely sensed RGB images with size of 1024 x 768 pixels.

Input images are collected by GMapCatcher toolbox which is a map viewer program available at code.google.com/p/gmapcatcher/. GMapCather has a feature to export remotely sensed RGB images which are provided by Google Map Maker service. This feature is provided by a custom GUI called as Export Map. Screenshot of Export Map feature is given in Figure 3.1. Export Map also allows user to adjust the size and the zoom level of the remotely sensed RGB image.

Table 3.1: Number of images

Level	Forestry	Water Surface
0	42	40
-1	42	40
-2	42	40
Total	126	120



Export map to PNG image

Corners' coordinates

Upper Left: 39.8749655019, 32.7886962891

Lower Right: 39.873384595, 32.7914428711

Image settings

Width / Height: 1024 / 768

Zoom Level: -2 Mode: RGB

GMapCatcher map viewer!

Figure 3.1: GUI of GMapCatcher Export Map feature displaying map tiles in specified coordinates

Three different zoom levels as 0, -1 and -2, where level -2 provides the most closer view, are used in exporting the remotely sensed images. In order to obtain better visual separation between the background and the objects or the areas in the images, minimum zoom level is chosen as -2 and maximum zoom level is chosen as 0. Further zoom levels cause more objects to be included in the input image while lower zoom levels captures more details both may be lead to wrong interpretation of comparison results between the saliency maps of the computational models and Tobii T120 eye tracker device.

For the same coordinates, remotely sensed images in dataset are exported at 0,-1 and -2 zoom levels. 0.18 km, 0.09 km and 0.045 km are the height information of zoom levels 0, -1 and -2 respectively. Three different zoom levels that are 0, -1 and -2 have spatial resolution as 1.2 m x 1.2 m, 0.6 m x 0.6 m and 0.3 m x 0.3 m respectively. Spatial resolution $m \; x \; n$ is calculated by given Equation 3.1;

$$m = \frac{D_h \; (x,y)}{1204}, n = \frac{D_v \; (x,y)}{768} \tag{3.1}$$

where m is horizontal spatial resolution and n is vertical spatial resolution. D_h and D_v are real horizontal and vertical distances respectively.

By using the coordinates of upper-left and lower-right corners of the exported image provided by exporting feature of GMapCatcher,D_h and D_v calculations are completed by an offline application which is called as FizzyCalc. It is implemented by FizzyMagic and available at www.fizzymagic.net/Geocaching/FizzyCalc/. Real distances and spatial resolutions of the zoom levels are given in Table 3.2.

Table 3.2:Input image attributes according to the zoom levels

Zoom Level	0	-1	-2
Height	0.18 km	0.09 km	0.045 km
Horizontal Distance	1.222674 km	0.611336 km	0.305668 km
Vertical Distance	0.910868 km	0.455434 km	0.227717 km
Spatial Resolution	~1.2 m x 1.2 m	~0.6 m x 0.6 m	~0.3 m x 0.3 m

Exported image extension is .png and it contains three color channels as Red, Green and Blue. Sample exported images in zoom levels 0, -1 and -2 are given in Figures 3.2 and 3.3. The input images in the dataset for forestry and water surface categories are shown in Appendix A and Appendix B respectively.

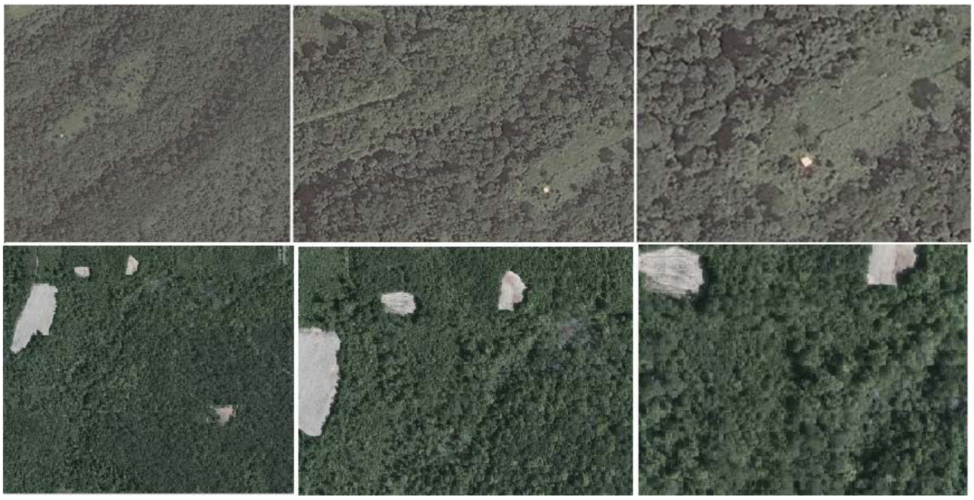

(a) (b) (c)

Figure 3.2: Sample exported images for forestry category (a) Level 0 (b) Level -1 (c) Level -2

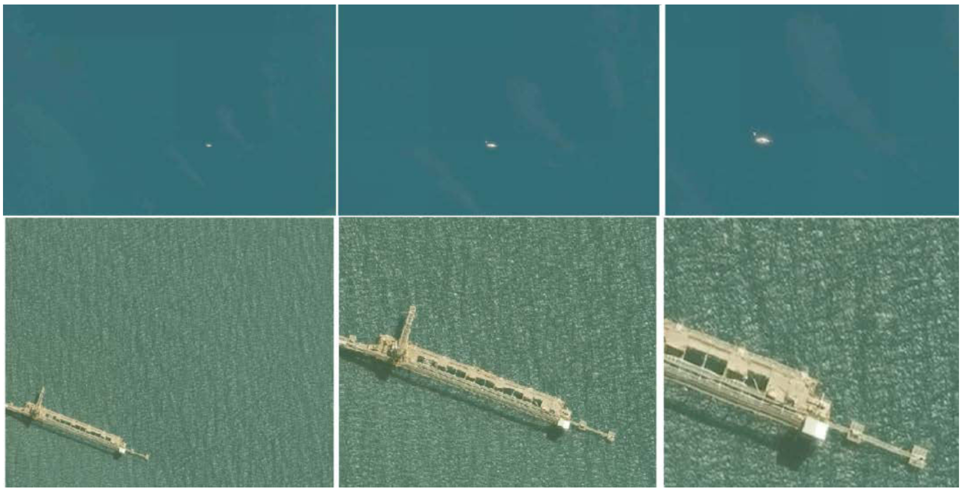

(a) (b) (c)

Figure 3.3: Sample exported images for water surface category (a) Level 0 (b) Level -1 (c) Level -2

3.2 Eye Tracker Hardware

Tobii T120 Eye Tracker device is located in Human Computer Interaction (HCI) Research and Application Laboratory at METU Computer Center. Tobi T120 shown in Figure 3.4consists of the eye tracker hardware and test computer. It collects the user's eye movement data such as how long and how many times he/she looks at a certain point on the screen. The data export tool includes heat maps and gaze plots as both layer on input image and a single image itself. Figure 3.5 shows sample heat map outputs of Tobii T120 Eye Tracker device while Figure 3.6 shows sample gaze plot outputs.

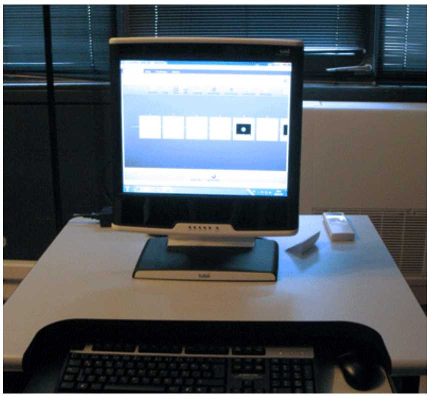
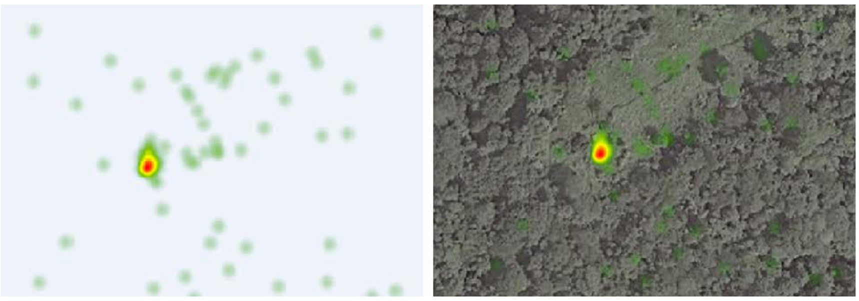
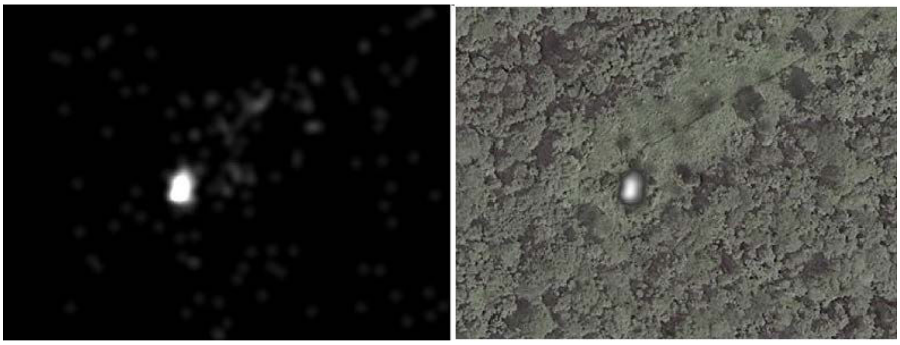


Figure 3.4:Tobii T120 Eye Tracker on test computer in HCI lab



(a) (b)

Figure 3.5: (a) Tobii T120 heat map output as a single image (b) Tobii T120 heat map output as a layer on input image



(a) (b)

Figure 3.6: (a) Tobii T120 gaze plot output as a single image (b) Tobii T120 gaze plot output as a layer on input image

3.3 Calibration

Before the experiment, human participants are asked to follow specific points on the screen. Firstly, sitting position is set for calibration procedure. Eyes are displayed as two white dots which must be located between green ranges as it is shown in Figure 3.7. In vertical axis, distance between the subject's eyes and the screen is expected to be 60 cm ± 3cm. In horizontal axis, the location of the subject’s eyes is expected to be centered. Otherwise, the eye tracker hardware is unable to detect eyes or it detects the subject's eye inaccurately.

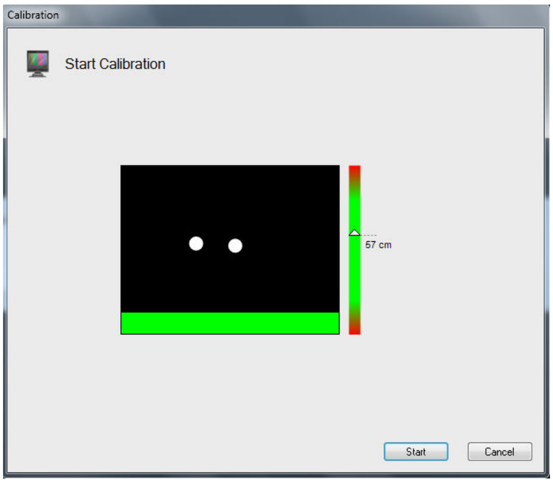


Figure 3.7: Calibration tool of Tobii T120 Eye Tracker

After eye position is fixed, the subject is informed about standing firm and not to blink during whole calibration process for preventing incorrect data collection of his/her eye movements on the screen. By clicking start button in calibration tool given by Figure 3.7, the red points start to appear and to move on the screen continuously as shown in Figure 3.8.

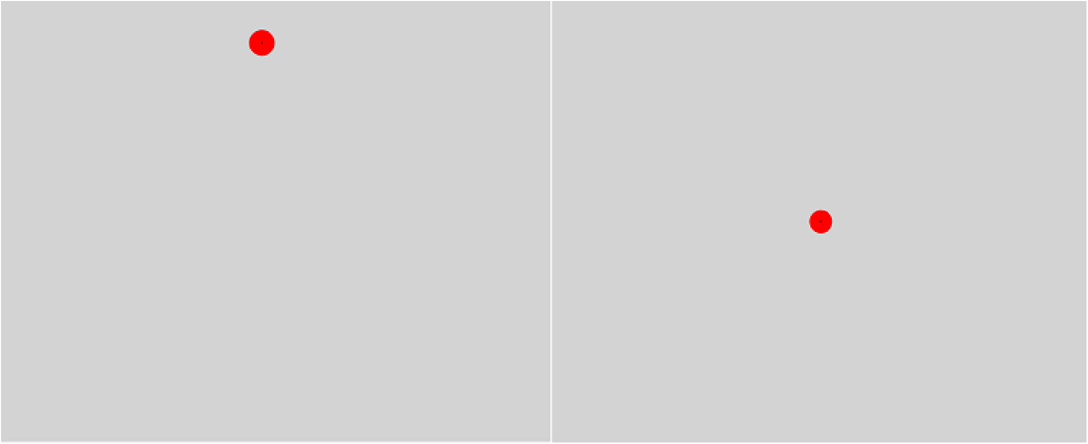
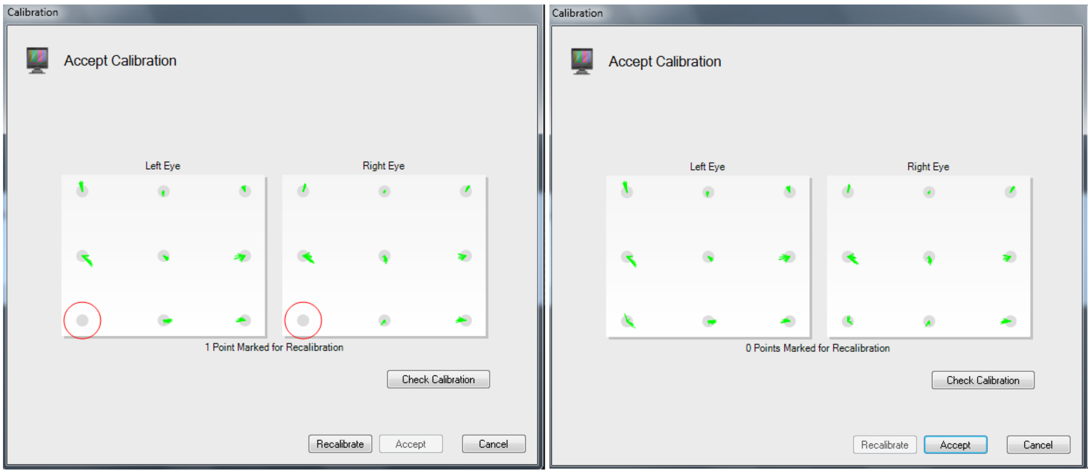


Figure 3.8: Specified points while calibrating eyes before the experiment

The red points contain black dots in the middle of them. It is also important to indicate to the participant that following black dots while red points move not only stop. If the black dots are not tracked in whole duration of eye calibration, Tobii T120 can not capture the correct location of eye movement even though the participant looks at correct location on the screen.

After calibration procedure is completed, two separate panels for each eye are displayed on the screen to check the result of the calibration as given in Figure 3.9. The green parts in Figure 3.9 show how well calibration is completed. If any eye movement is not captured during calibration process, there is not green part inside of the corresponding circle. In this case, it is expected that calibration should be repeated for the corresponding circle. Furthermore, if green parts are scattered too much to outwards of the circle, calibration for the corresponding circle should be repeated.



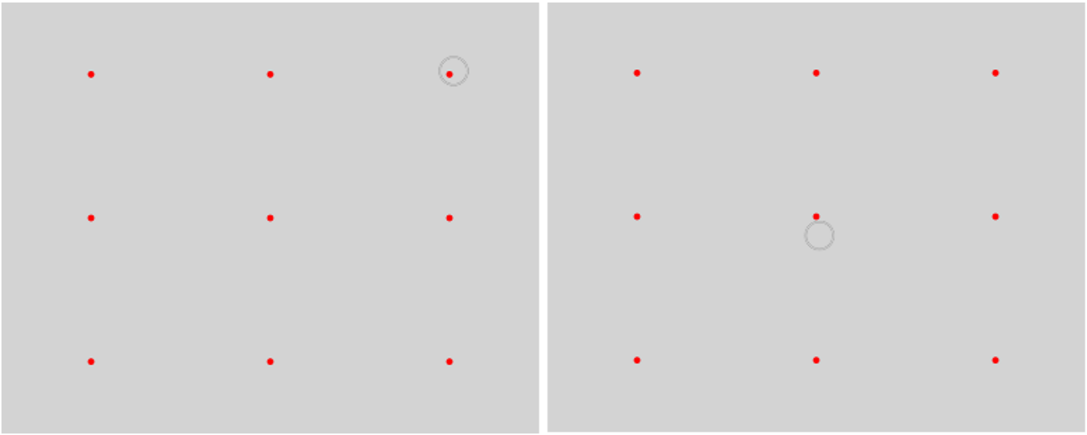
1 Point Marked for Recalibration

0 Points Marked for Recalibration

(a) (b)

Figure 3.9: (a) Eye calibration could not capture eye movement (b) Acceptable eye calibration to start experiment

Tobii T120 also contains another calibration check tool as given in Figure 3.10. The subject is expected to look at the center of the red points on screen. If the red point thatis located inside of the circle moves with the eye, calibration process is completed successfully. Otherwise, calibration procedure should be repeated for the corresponding circle. In Figure 3.10, difference between successful calibration and inadequate calibration can also be seen.



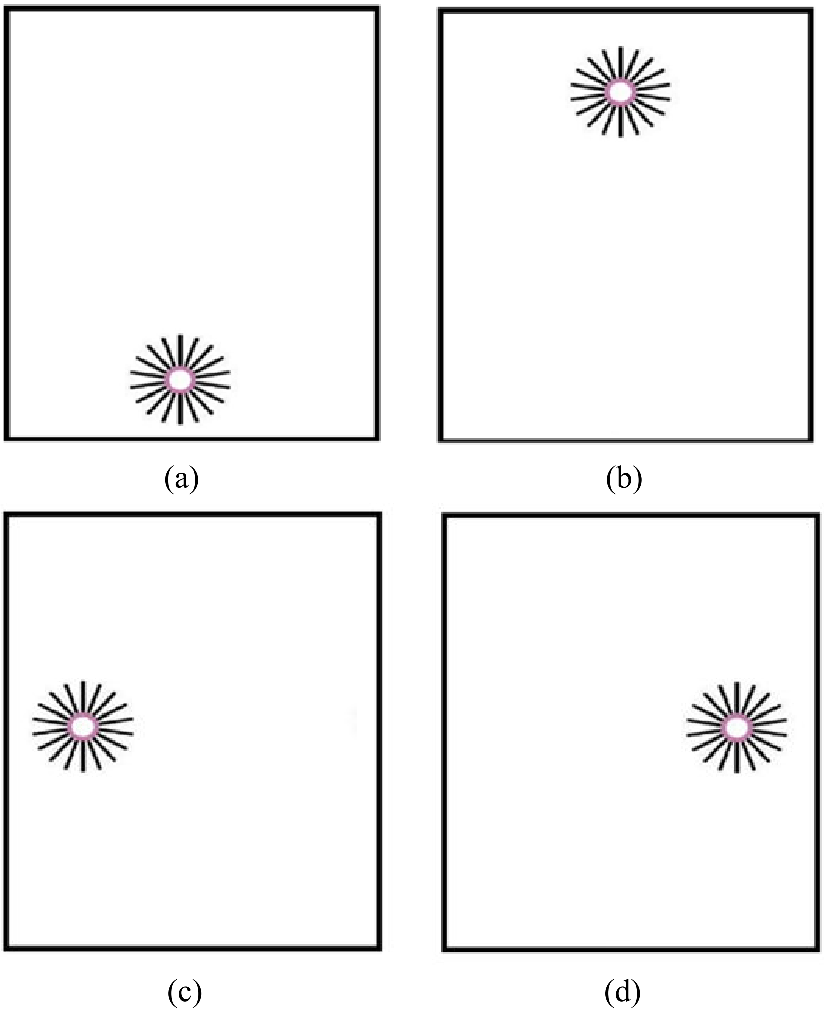
(a) (b)

Figure 3.10: (a) Correct data capture of eye movement (b) Incorrect data capture of eye movement due to insufficient calibration

3.4 Eye Tracking Experiments

In experiments, separate test projects are created with the dataset containing 226 remotely sensed RGB images. Two test projects are executed separately for forestry and water surface categories.20 people participated in both test projects to obtain ground truth information.They are selected based on their eye structure. They have big eyes and do not use contact lenses and glasses. Additionally, they do not have any visual impairment. For forestry category, images are presented to 7 females and 13 males while in water surface category 8 females and 12 males. In order to fix subjects' eye movements to Tobii T120 screen, their eyes are calibrated using the software provided by Tobii T120 system. Calibration should be performed according to the targets given by Tobii T120 software on the screen. Otherwise, it collects data in wrong locations even though the subject looks at the exact location. Therefore, it is required that the subjects to stay firm during the calibration process and whole experiment duration.

Since subject's attention and concentration are very important in order to collect data with high accuracy and precision, an instructor image given by Figure 3.11 is also located between two input images in test projects. Fix cross sign is located according to next input image. For example, if expected salient region or object are located in right-hand side of the input image, fix cross sign is located in left-hand side. Therefore, additional and missing looking are prevented. Hence, visualization outcomes of Tobii T120 and interpretation of them are expected to be more reliable.



(a) (b)

(c) (d)

Figure 3.11:(a) Next info is located in top side of the screen, fix cross is located in bottom side (b) Next info is located in bottom side of the screen, fix cross is located in top side (c) Next info is located in right side of the screen, fix cross is located in left side (d) Next info is located in left side of the screen, fix cross is located in right side

Before the eye tracking experiments are performed, two separate test projects are created for forestry and water surface categories. Then, the input images are aligned randomly and the fix-cross signs are located between the input images. After the image sequences of both test projectsare completed, calibration process is performed. Firstly, 126 input images of forestry category and the fix-cross signs located between the input images are represented to the subjects. The input images in forestry category are displayed for 6 seconds while the fix-cross signs are displayed for 2 seconds separately. Total time duration for forestry category is 30 minutes. After forestry category is completed, the subject has a break until he/she is ready for the second part of the experiment which is water surface category. For the second part of the experiment, the calibration process is performed again. 120 input images of water surface category are displayed for 6 seconds while the fix-cross signs, which are located between the input images are displayed for 2 seconds separately. Total time duration for water surface category is 28 minutes. Sample sequences of both categories are given in Figures 3.12 and 3.13.

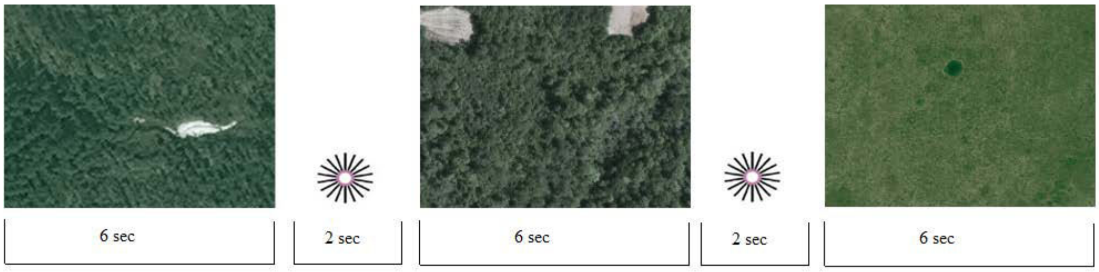


Figure 3.12: Sample image sequence for forestry category

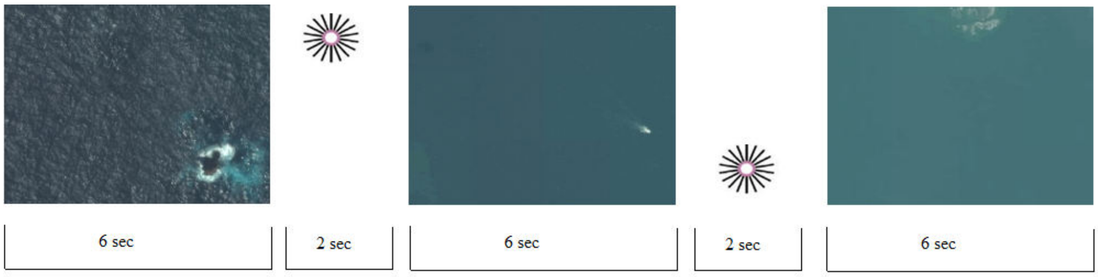


Figure 3.13: Sample image sequence for water category

3.5 Eye Tracking Outputs

In this experiment, grey scale heat maps are obtained by using gaze plot export feature of Tobii T120 as given by Figure 3.14. This feature is capable of an adjustment such as zero saliency regions are assigned to 0 value and the regions with higher salient regions are assigned to the values that are closer to 1.

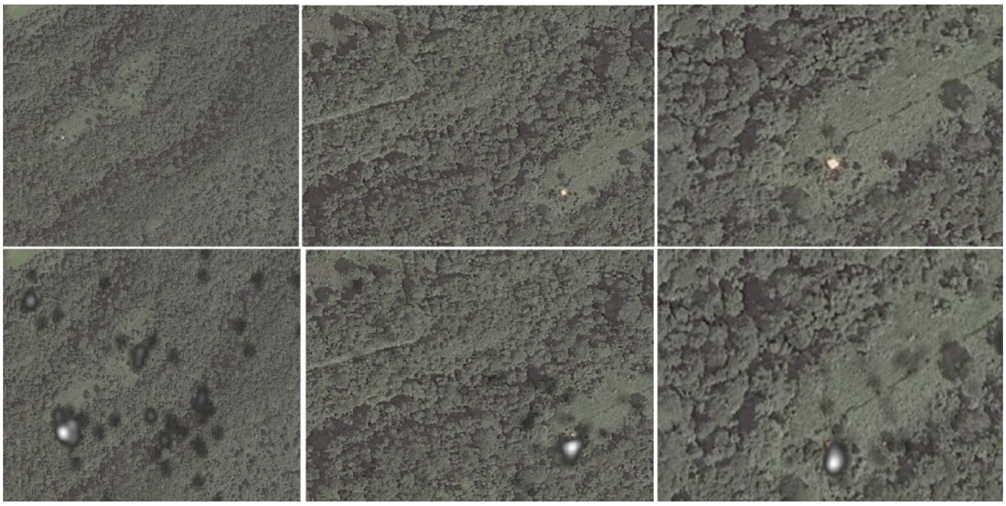


Figure 3.14: Gaze plots of sample input images as gray scale heat map layer

Moreover, Tobii T120 includes weighted gaze samples feature, which is the percentage calculated by dividing the number of eye tracking samples that were correctly identified, by the number of eye movements, indicates how useful the recording of both eyes for analysis by measuring the percentage of captured eye movements during a recording. For forestry category, weighted gaze sample contains %92 usable gaze plots while %93 for water surface category. Furthermore, individual gaze samples are also obtained. They provides the percentage of captured eye movements of each subject individually; minimum usable percentage for individual gaze plots is %74 while the maximum percentage is %99 in the experiment.

Sample outcomes of weighted and individual gaze plots as image layers are given in Figure 3.15.

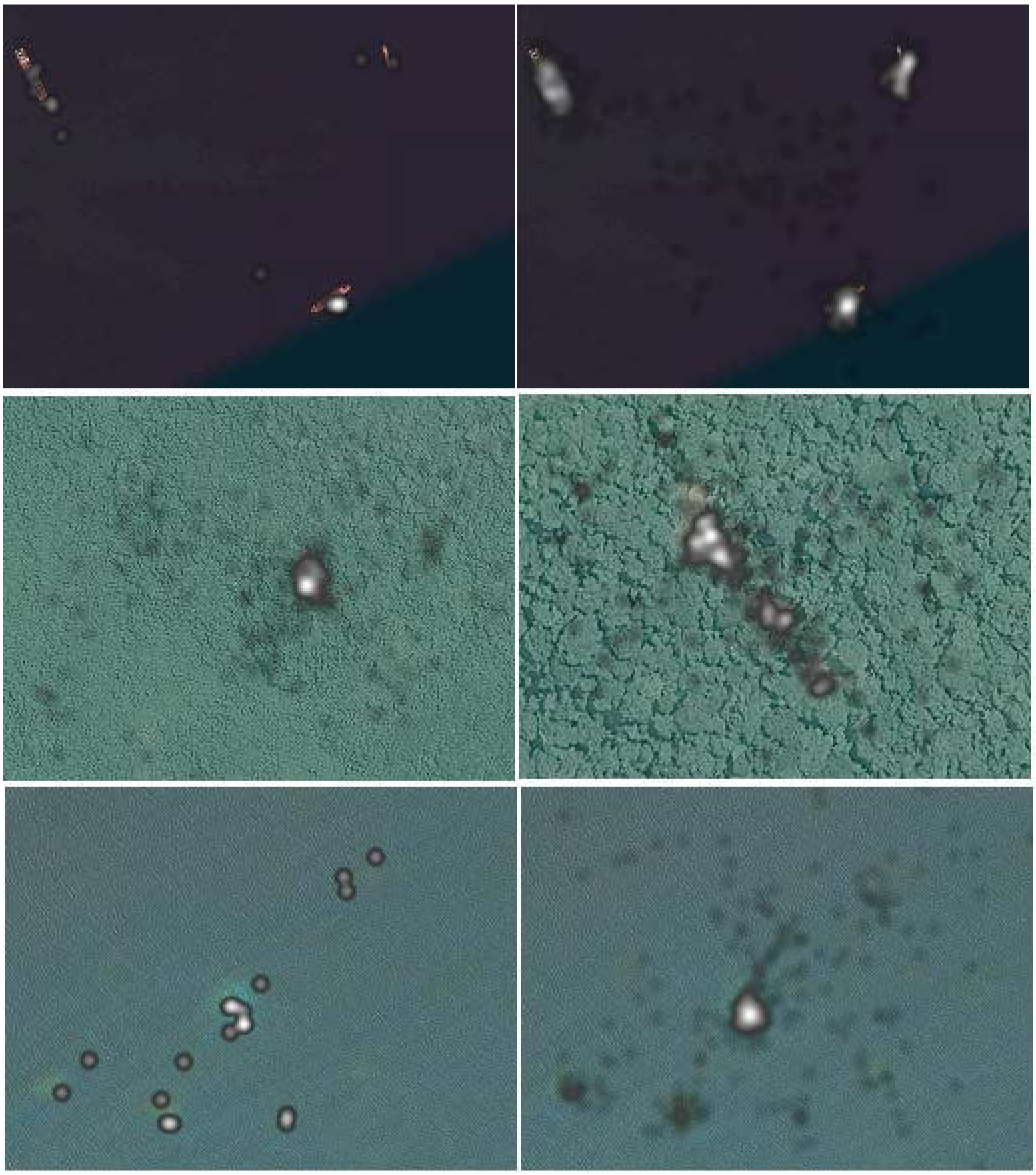


Figure 3.15 : (a) Individual gaze plot layers (b) Weighted gaze plot layers considering all subjects together

Additionally, each weighted gaze plot is also exported within specified time intervals as given in Table 3.3. For accuracy measurement, each time interval is compared with others.

Table 3.3: Time intervals for forestry and water surface categories

Start Time	Finish Time	Category	Total Exported Image Count
0.0	2.0	Forestry	126
2.0	4.0	Forestry	126
4.0	6.0	Forestry	126
1.0	6.0	Forestry	126
1.0	3.5	Forestry	126
3.5	6.0	Forestry	126
0.0	2.0	Water Surface	120
2.0	4.0	Water Surface	120
4.0	6.0	Water Surface	120
1.0	6.0	Water Surface	120
1.0	3.5	Water Surface	120
3.5	6.0	Water Surface	120

Since subjects lose concentration on single image for time intervals longer than 6seconds, input images are presented to them for 6 seconds. Besides locating instructor images between the input images, weighted gaze plots are exported in between 1.0 sec and 6.0 sec. Therefore, resulting gaze plots contain more accurate and precise eye movement data.

The Tobii T120 outcomes are computed in MATLAB to obtain saliency map. Tobii T120 saliency map is a single channel image in grey scale. Hence, the comparison with the computational saliency models can be performed with the same data type in MATLAB. Sample for Tobii T120 saliency map is given by Figure 3.16.

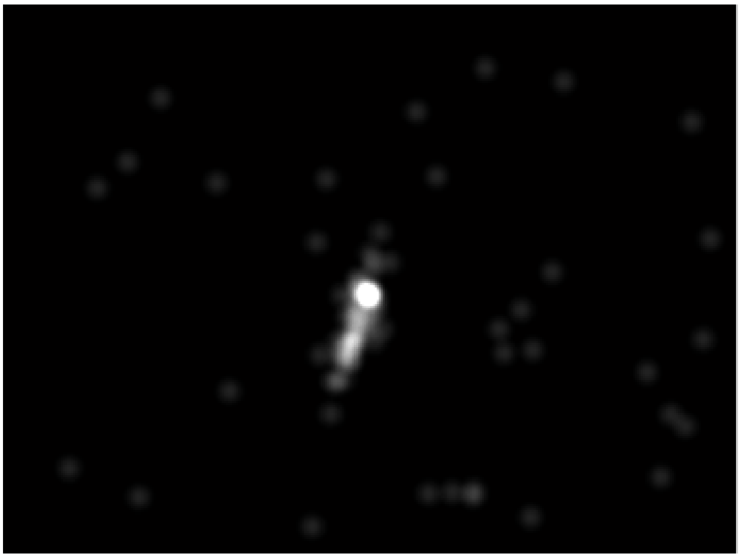


Figure 3.16:Tobii T120 saliency map in MATLAB

CHAPTER 4

RESULTS

In this thesis, we proposed a study of how well saliency map computational models approximate the eye tracking experiment performed with the eye tracker device Tobii T120. Results were obtained by measuring the similarities of saliency maps obtained by eye tracking experiments and saliency map outputs of computational methods which are GBVS, Itti-Koch, Achanta, ImageSignature, SDSP and CovSal.

4.1 Analysis of Similarity Between Tobii T120 Saliency Maps Obtained for Different Time Intervals

For each image, weighted and individual gaze plots are collected for 6 seconds using Tobii T120 eye tracker device.

In order to analyze the similarity of Tobii T120 saliency maps extracted for different time intervals, weighted gaze plots of each input image were exported in [0.0- 2.0] sec, [2.0- 4.0] sec and [4.0- 6.0] sec intervals. Also, Tobii T120 saliency maps for [1.0 - 3.5] sec and [3.5 - 6.0] sec intervals are collected for further analysis. SSIM requires raw input images. Therefore, image adjustment such as normalization is not performed. Hence, data loss is prevented. Similarities between the time durations were measured by computing SSIM according to the categories. Mean value, standard deviation (STD) value and coefficient of variation (CV) percentage corresponding to the SSIM results of forestry and water surface categories are given in Table 4.1 and Table 4.2 respectively. CV percentage is calculated by given Equation 4.1.

$$CV = \%100 \times \sigma \ / \ \mu \tag{4.1}$$

whereσ is the standard deviation and μ is the mean value.

Table 4.1:SSIM comparisons of forestry images in specified time intervals

SSIM Mean / STD / CV			
Time Intervals	[2.0 - 4.0] sec	[4.0 - 6.0] sec	[3.5 - 6.0] sec
[0.0 - 2.0] sec	0.665 / 0.056 / %8	0.664 / 0.056 / %8	
[2.0 - 4.0] sec		0.999 / 0.014 / %1	
[1.0 - 3.5] sec			0.999 / 0.016 / %2

Table 4.2:SSIM comparisons of water surface images in specified time intervals

SSIM Mean / STD / CV			
Time Intervals	[2.0 - 4.0] sec	[4.0 - 6.0] sec	[3.5 - 6.0] sec
[0.0 - 2.0] sec	1 / 0 / %0	0.999 / 0.007 / %1	
[2.0 - 4.0] sec		0.999 / 0.007 / %1	
[1.0 - 3.5] sec			0.998 / 0.006 / %1

According to Table 4.1,the subjects focus more in salient regions between the time range in [0.2 - 6.0] sec in forestry category. For water surface category, images contain less textured background than forestry images. He/she does not focus on additional details. Therefore, the deviations are more stable in comparison with forestry category as indicated in Table 4.2. According to Tables 4.1 and 4.2 it is decided to use [1.0 - 6.0] sec time intervals for Tobii T120 saliency map extraction because it eliminates inconsistency in the [0.0 - 1.0] sec interval and also provides more sampling points compared to [2.0-6.0] sec time interval.

4.2 Analysis of Similarity Between Tobii T120 Saliency Maps Obtained for Different Participants

In order to measure similarity between the subjects, 12 individual gaze plots for time interval [1.0 - 6.0] sec were exported randomly from whole dataset. Each individual gaze plot was executed with the others. As for time duration comparisons, the individual gaze plots are compared within SSIM calculation without image adjustment. Table 4.3 contains mean value, standard deviation (STD) value and coefficient of variation (CV) percentage of each participant.

Table 4.3:Sample individual gaze plot comparison results

Subject No	**Mean / STD / CV**	**Subject No**	**Mean / STD / CV**
1	0.9456 / 0.0239 / %3	11	0.9516 / 0.0241 / %3
2	0.9546 / 0.0247 / %3	12	0.9512 / 0.0212 / %2
3	0.9449 / 0.0233 / %2	13	0.9463 / 0.0204 / %2
4	0.9372 / 0.0185 / %2	14	0.9455 / 0.0189 / %2
5	0.9491 / 0.0216 / %2	15	0.9397 / 0.0241 / %3
6	0.9545 / 0.0228 / %2	16	0.9411 / 0.0206 / %2
7	0.9535 / 0.0255 / %3	17	0.9519 / 0.0234 / %2
8	0.9375 / 0.0227 / %2	18	0.9408 / 0.0219 / %2
9	0.9556 / 0.0226 / %2	19	0.9375 / 0.0022 / %2
10	0.9516 / 0.0232 / %2	20	0.9540 / 0.0114 / %2

According to Table 4.3,each individual Tobii T120 saliency map provides a high similarity with other individual maps since the mean values are close to 1 and coefficient of variation percentage is close to %0.

4.3 Analysis of Similarity Between Tobii T120 Saliency Maps and Computational Saliency Maps

Figures 4.1 - 4.3 show histogram graphs of the computational saliency map models individually in forestry category while Figures 4.4 - 4.6 show histogram graphs in water surface category. The histogram graphs indicate the similarity values of each saliency map between minimum and maximum similarity in the range [0..1]. Additionally, on the Figures 4.1 - 4.6 also the mean values (m), standard deviations (STD) and coefficient of variation (CV) percentage are given for each model.

A similarity value with mean=1, STD=0 and %CV=0 corresponds to exact similarity between the saliency maps of computational models and Tobii T120.

In Figure 4.1 and Figure 4.3, it is shown that CovSal-C has more similarity values closer to 1 while in Figure 4.2, GBVS has more correlation values collected closer to 1 for forestry category.

For water surface category, Figure 4.4 indicates CovSal-C contains more similarity values closer to 1. In Figures 4.5 and 4.6, GBVS has more similarity values closer to 1.

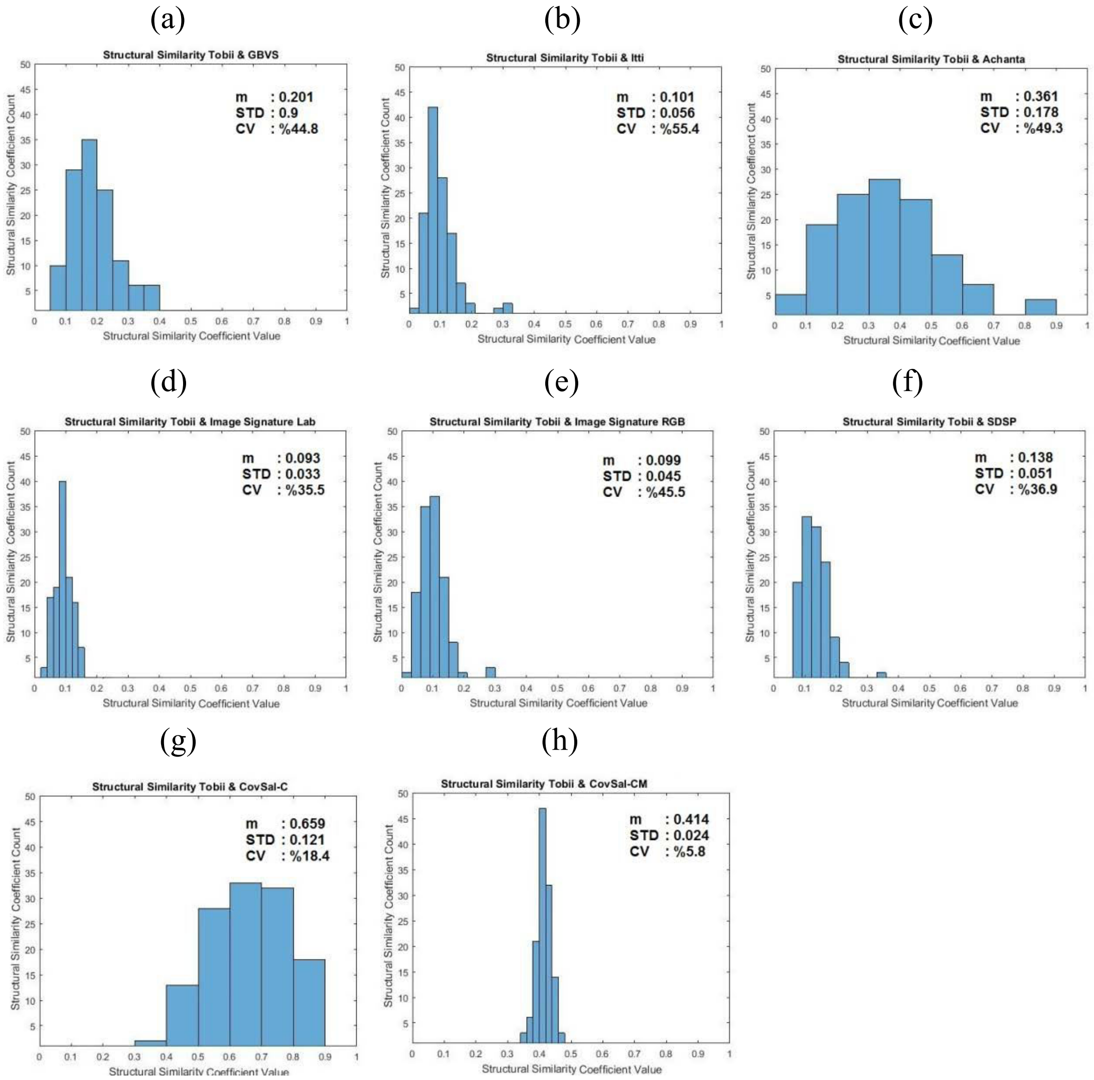


Figure 4.1:SSIM coefficient histogram of forest category saliency maps (a) Tobii and GBVS (b) Tobii and Itti-Koch (c) Tobii and Achanta, (d) Tobii and ImageSignature - Lab (e) Tobii and ImageSignature - RGB (f) Tobii and SDSP (g) Tobii and CovSal-C (h) Tobii and CovSal-CM

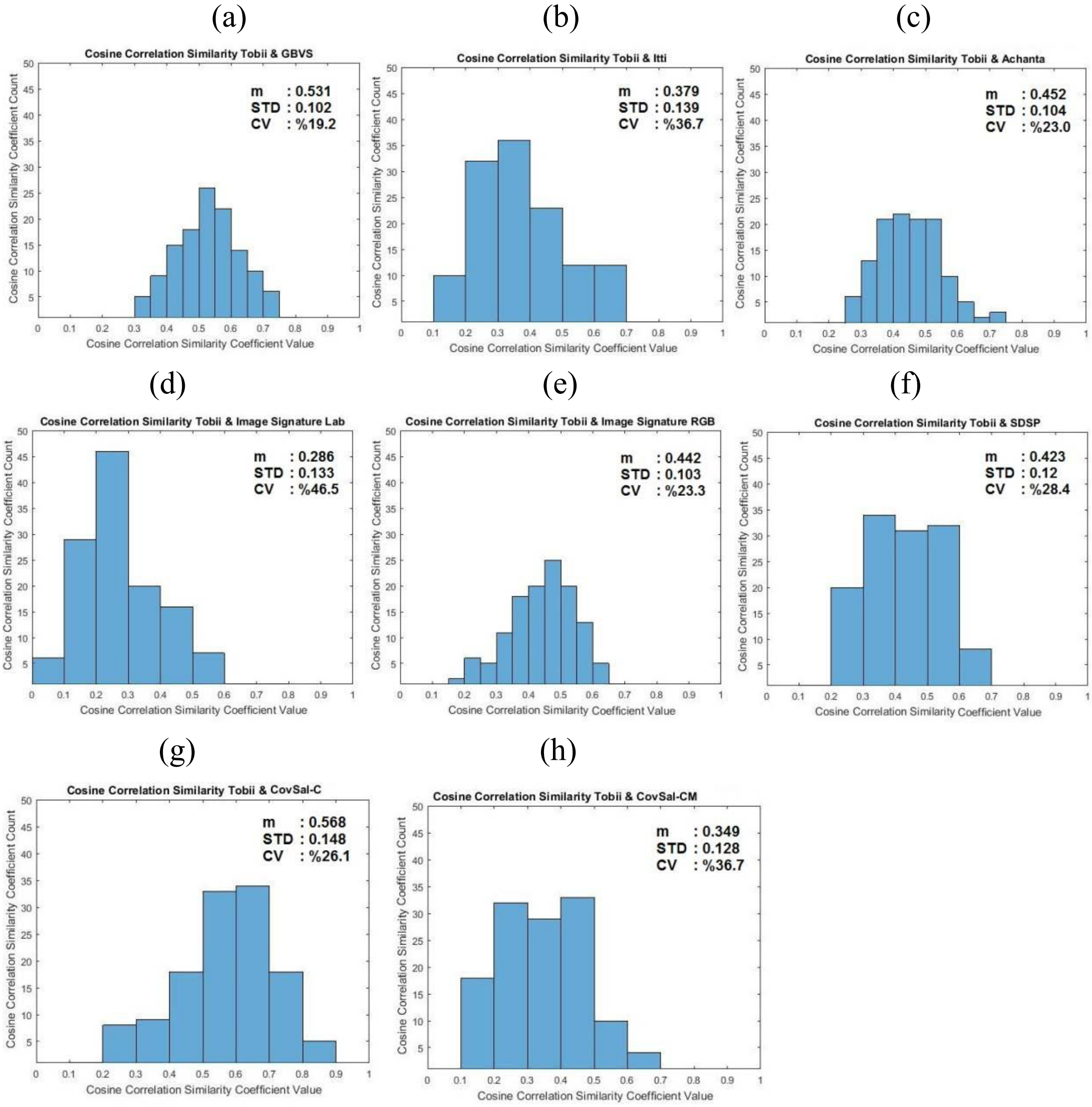


Figure 4.2: Cosine correlation coefficient histogram of forestry category saliency maps (a) Tobii and GBVS (b) Tobii and Itti-Koch (c) Tobii and Achanta (d) Tobii and ImageSignature - Lab (e) Tobii and ImageSignature - RGB (f) Tobii and SDSP (g) Tobii and CovSal-C (h) Tobii and CovSal-CM

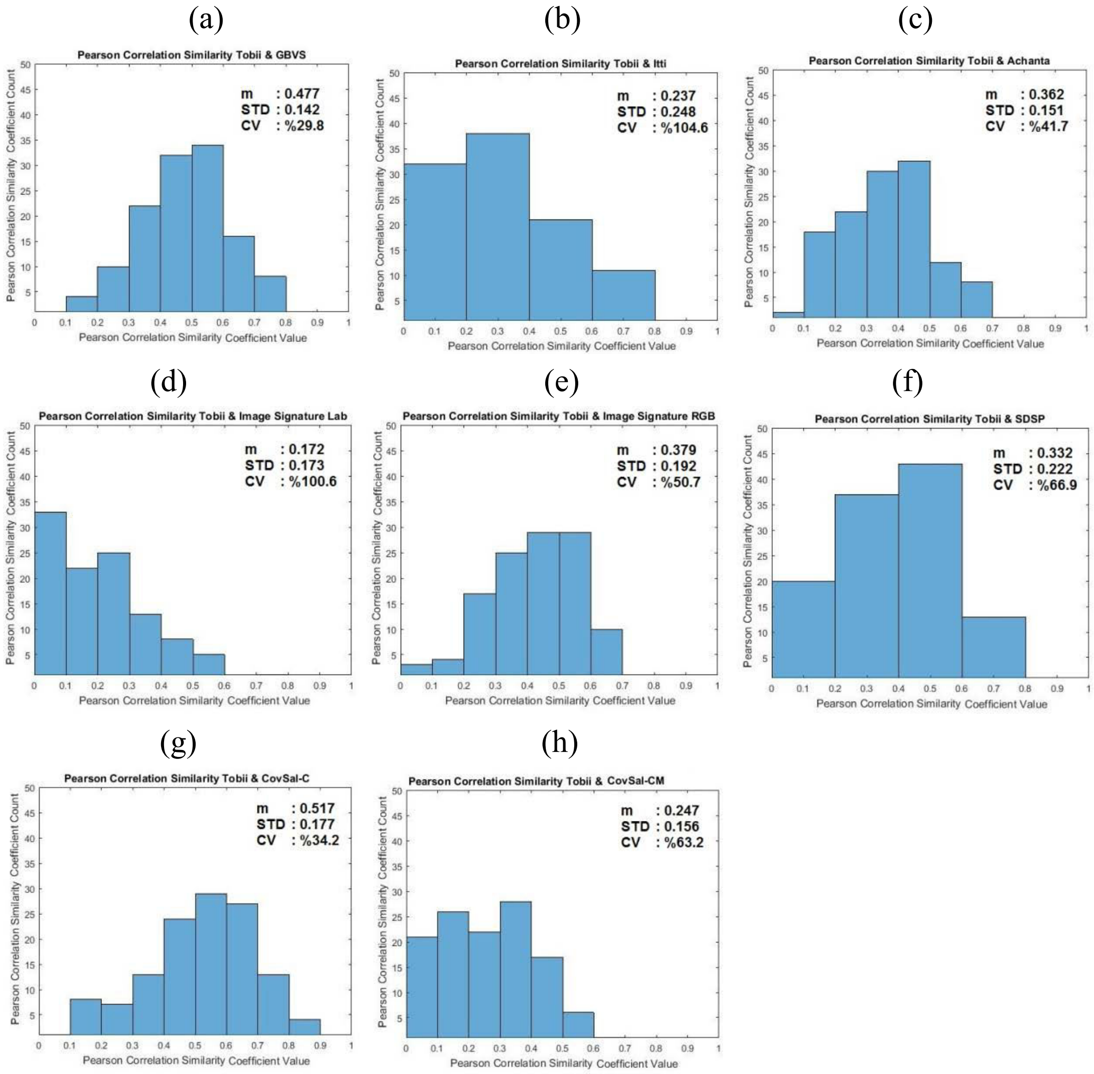


Figure 4.3: Pearson correlation coefficient histogram of forestry category saliency maps (a) Tobii and GBVS (b) Tobii and Itti-Koch (c) Tobii and Achanta (d) Tobii and ImageSignature - Lab (e) Tobii and ImageSignature - RGB (f) Tobii and SDSP(g) Tobii and CovSal-C (h) Tobii and CovSal-CM

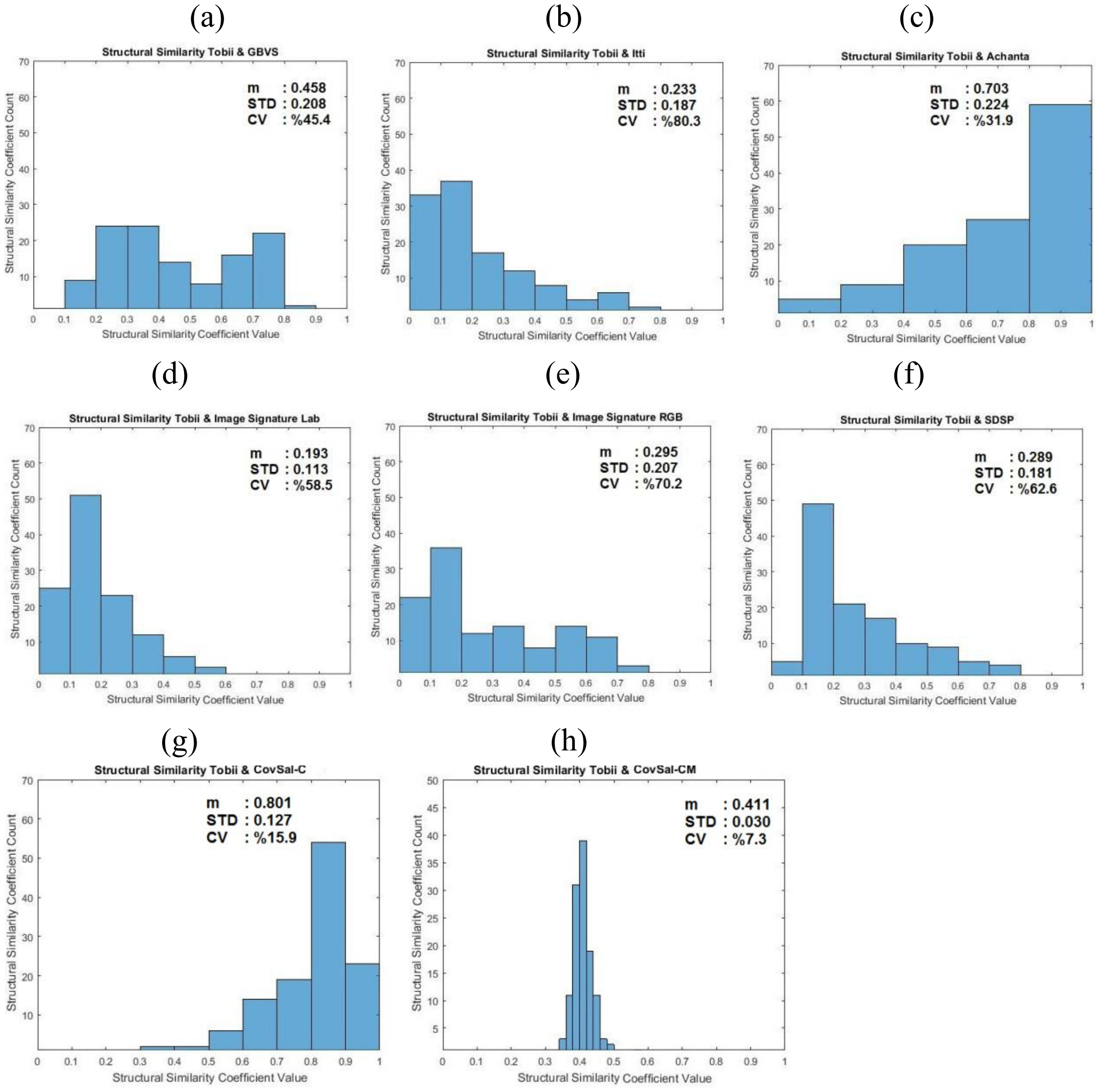


Figure 4.4: SSIM coefficient histogram of water surface category saliency maps (a) Tobii and GBVS (b) Tobii and Itti-Koch (c) Tobii and Achanta (d) Tobii and ImageSignature - Lab (e) Tobii and ImageSignature - RGB (f) Tobii and SDSP (g) Tobii and CovSal-C(h) Tobii and CovSal-CM

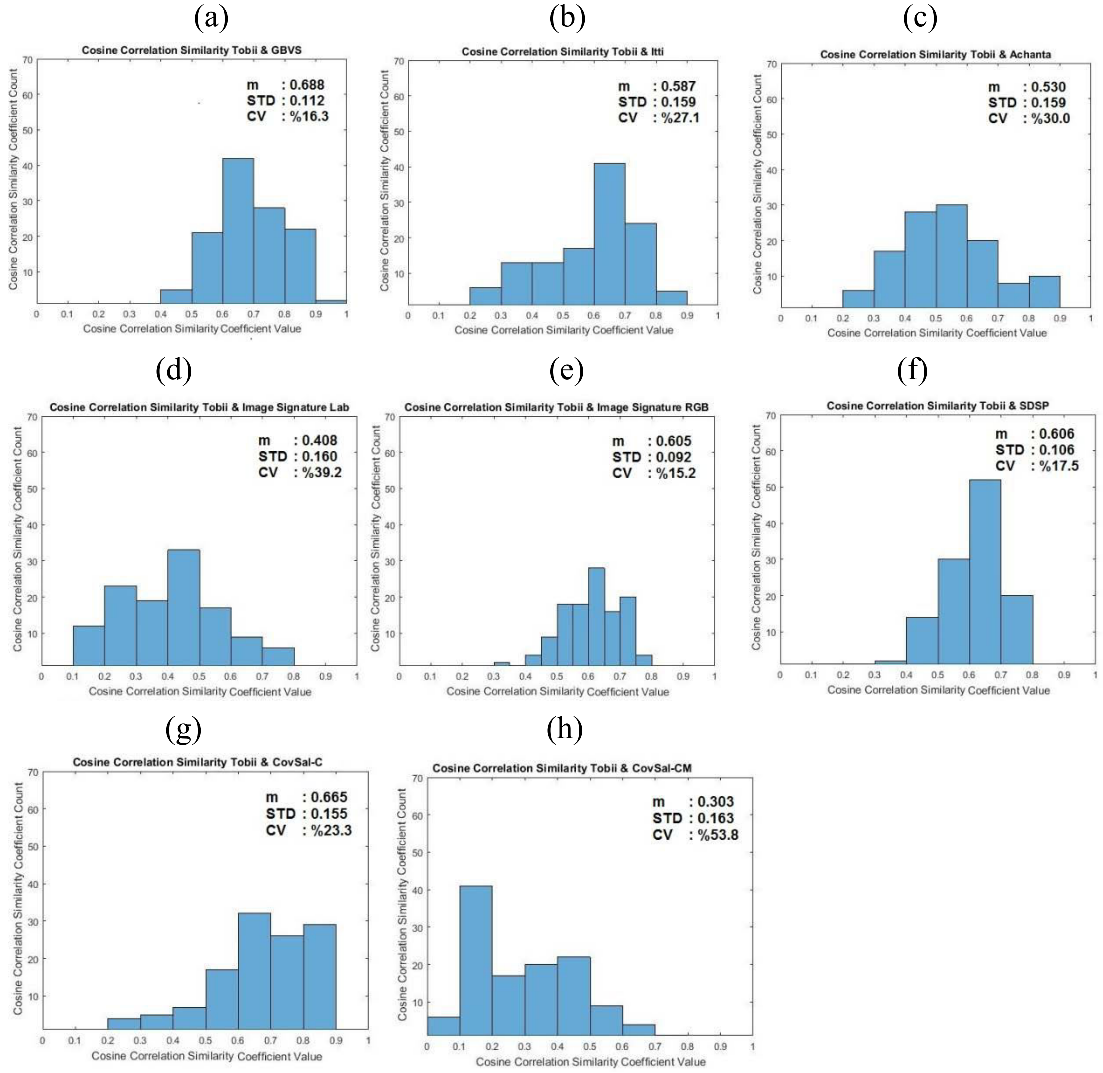


Figure 4.5: Cosine correlation coefficient histogram of water surface category saliency maps (a) Tobii and GBVS (b) Tobii and Itti-Koch (c) Tobii and Achanta (d) Tobii and ImageSignature - Lab (e) Tobii and ImageSignature - RGB (f) Tobii and SDSP (g) Tobii and CovSal-C (h) Tobii and CovSal-CM

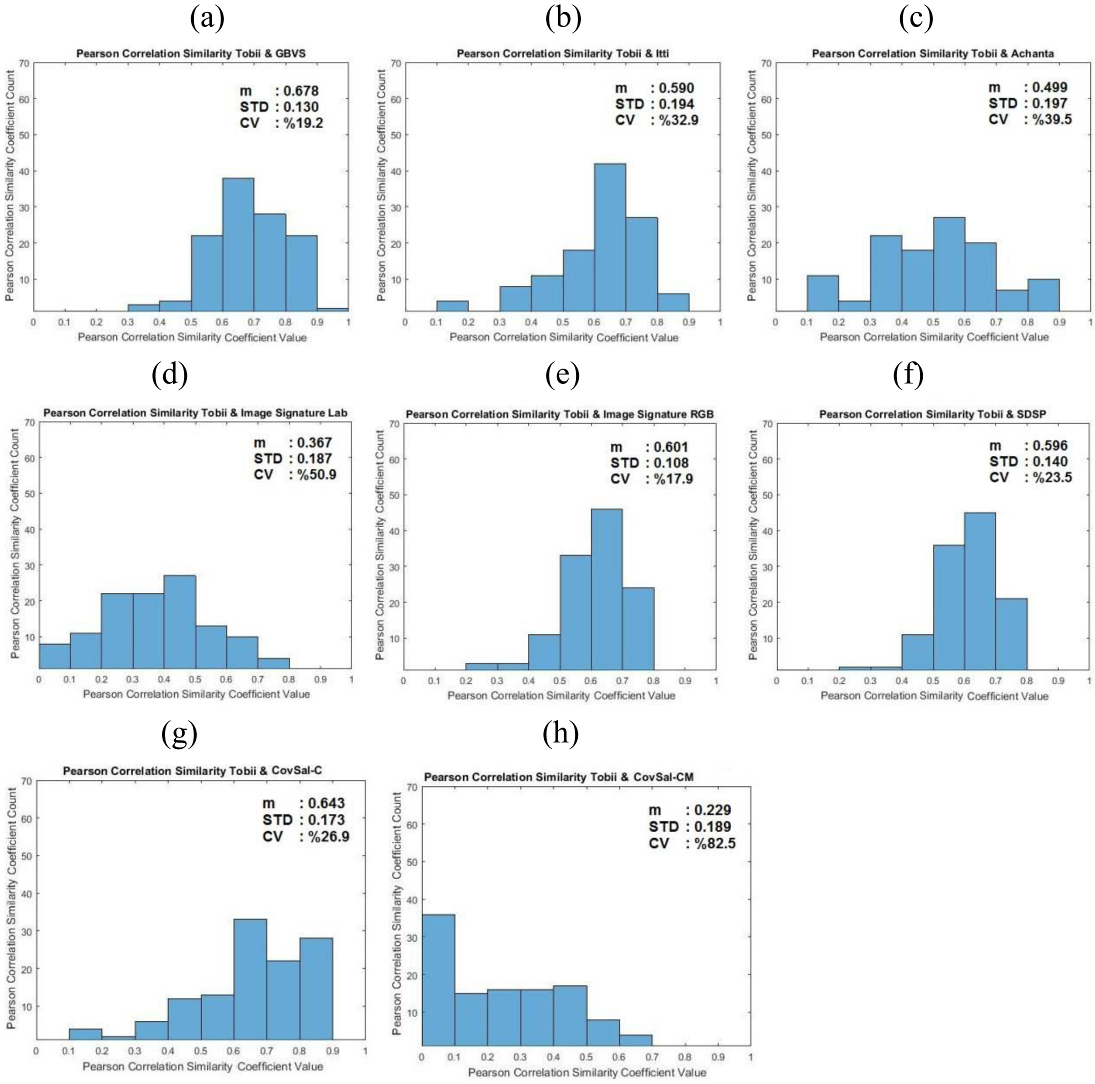


Figure 4.6: Pearson correlation coefficient histogram of water surface category saliency maps (a) Tobii and GBVS (b) Tobii and Itti-Koch (c) Tobii and Achanta(d) Tobii and ImageSignature - Lab (e) Tobii and ImageSignature - RGB (f) Tobii and SDSP (g) Tobii and CovSal-C (h) Tobii and CovSal-CM

Table 4.4 and Table 4.5summarize mean, standard deviation (STD) values and coefficient of variation (CV) percentage for the SSIM, Cosine and Pearson correlation similarities of the saliency maps produced by computational models and Tobii T120 saliency maps in both categories (126 images for forestry, 120 images for water surface).Computational saliency maps of both categories were also compared with the saliency map of Tobii T120 according to the levels 0, -1 and -2. Tables4.6-4.8 contain forestry category results while Table 4.9- 4.11containwater surface category results.

In order to obtain highest similarity to identify the best fit saliency map model, the correlation coefficients of the model are expected to be close to 1. For SSIM index and Pearson correlation methods CovSal-C model provides more coefficient close to 1 in forestry category among other saliency map models. It is also important that standard deviation value of correlation values is expected to be close to 0 and coefficient of variation percentage is expected to be close to %0 to obtain highest precision. GBVS provides smaller standard deviation and mean value higher than 0.5 according to Cosine correlation.

For water surface category, GBVS model is the best fit saliency map model among others according to Cosine correlation and Pearson correlation methods. It provides highest mean value in comparison with the others. According to SSIM index correlation method, CovSal-C method has highest mean value. However, GBVS model has a smaller standard deviation value and a smaller coefficient of variation percentage according to Cosine correlation and Pearson correlation results in comparison with CovSal-C model.

Additionally, non-salient regions which are textured backgrounds are eliminated by CovSal model more effectively than GBVS model. Forestry images have more low and intermediate informative regions caused by textured background while water surface images provides more unique background.

Table 4.4: Mean, STD and CV values of similarity measurement coefficients in Forestry category for 126 images

Model Name	SSIM Index Mean / STD / CV	Cosine Similarity Mean / STD / CV	Pearson Correlation Similarity Mean / STD / CV
GBVS	0.201 / 0.09 / %44.8	0.531 / 0.102 / %19.2	0.477 / 0.142 / %29.8
Itti-Koch	0.101 / 0.056 / %55.4	0.379 / 0.139 / %36.7	0.237 / 0.248 / %104.6
Achanta	0.361 / 0.178 / %49.3	0.452 / 0.104 / %23.0	0.362 / 0.151 / %41.7
ImageSignature -Lab	0.093 / 0.033 / %35.5	0.286 / 0.133 / %46.5	0.172 / 0.173 / %100.6
ImageSignature -RGB	0.099 / 0.045 / %45.5	0.442 / 0.103 / %23.3	0.379 / 0.192 / %50.7
SDSP	0.138 / 0.051 / %36.9	0.423 / 0.12 / %28.4	0.332 / 0.222 / %66.9
CovSal-C	0.659 / 0.121 / %18.4	0.568 / 0.148 / %26.1	0.517 / 0.177 / %34.2
CovSal-CM	0.414 / 0.024 / %5.8	0.349 / 0.128 / %36.7	0.247 / 0.156 / %63.2

Table 4.5: Mean, STD and CV values of similarity measurement coefficients in Water Surface category for 120 images

Model Name	SSIM Index Mean / STD / CV	Cosine Similarity Mean / STD / CV	Pearson Correlation Similarity Mean / STD / CV
GBVS	0.458 / 0.208 / %45.4	0.688 / 0.112 / %16.3	0.678 / 0.13 / %19.2
Itti-Koch	0.233 / 0.187 / %80.3	0.587 / 0.159 / %27.1	0.59 / 0.194 / %32.9
Achanta	0.703 / 0.224 / %31.9	0.530 / 0.159 / %30.0	0.499 / 0.197 / %39.5
ImageSignature -Lab	0.193 / 0.113 / %58.5	0.408 / 0.16 / %39.2	0.367 / 0.187 / %50.9
ImageSignature -RGB	0.295 / 0.207 / %70.2	0.605 / 0.092 / %15.2	0.601 / 0.108 / %17.9
SDSP	0.289 / 0.181 / %62.6	0.606 / 0.106 / %17.5	0.596 / 0.14 / %23.5
CovSal-C	0.801 / 0.127 / %15.9	0.665 / 0.155 / %23.3	0.643 / 0.173 / %26.9
CovSal-CM	0.411 / 0.03 / %7.3	0.303 / 0.163 / %53.8	0.229 / 0.189 / %82.5

Input images of forestry category provide better seperation between the actual salient region and backgroundat level 0. For level -2, the actual salient regions become more remarkable since the details such as color, edges and orientation become more apparent. However, at level -1 the details are not apparent as much as level -2 and figure-ground seperation is lesser than level 0. Therefore, subjects focusmore on details until they find the actual salient region at level -1. Hence,for level -1, standard deviation values of CovSal-C and GBVS model is higher than level 0 and level -2. GBVS model provides a smaller coefficient of variation percentage according to Cosine similarity rather than CovSal-C model. However, mean value ofCovSal-C model is higher than GBVS model according to each comparison method as given in

Tables 4.6-4.8. Therefore, CovSal-C model is the best fit model for 0, -1 and -2 levels of forestry category.

Table 4.6: Forestry category. Level: 0, Total input image: 42

Model Name	SSIM Index Mean / STD / CV	Cosine Similarity Mean / STD / CV	Pearson Correlation Similarity Mean / STD / CV
GBVS	0.198 / 0.078 / %39.4	0.523 / 0.089 / %17.0	0.467 / 0.123 / %26.3
Itti-Koch	0.112 / 0.059 / %52.7	0.415 / 0.12 / %28.9	0.3 / 0.196 / %65.3
Achanta	0.358 / 0.192 / %53.6	0.421 / 0.081 / %19.2	0.316 / 0.122 / %38.6
ImageSignature -Lab	0.104 / 0.037 / %35.6	0.248 / 0.087 / %35.1	0.124 / 0.123 / %99.2
ImageSignature -RGB	0.111 / 0.044 / %39.6	0.458 / 0.087 / %18.9	0.395 / 0.131 / %33.2
SDSP	0.136 / 0.054 / %39.7	0.441 / 0.091 / %20.6	0.365 / 0.149 / %40.8
CovSal-C	0.665 / 0.123 / %18.5	0.57 / 0.138 / %24.2	0.519 / 0.171 / %32.9
CovSal-CM	0.411 / 0.026 / %6.3	0.339 / 0.126 / %37.2	0.232 / 0.155 / %66.8

Table 4.7:Forestry category. Level: -1, Total input image: 42

Model Name	SSIM Index Mean / STD / CV	Cosine Similarity Mean / STD / CV	Pearson Correlation Similarity Mean / STD / CV
GBVS	0.205 / 0.098 / %47.8	0.539 / 0.11 / %20.4	0.498 / 0.146 / %29.3
Itti-Koch	0.091 / 0.051 / %56.0	0.374 / 0.15 / %40.1	0.243 / 0.271 / %111.5
Achanta	0.349 / 0.186 / %53.3	0.461 / 0.107 / %23.2	0.388 / 0.152 / %39.2
ImageSignature -Lab	0.088 / 0.032 / %36.4	0.287 / 0.132 / %45.9	0.179 / 0.172 / %96.1
ImageSignature -RGB	0.098 / 0.05 / %51.0	0.452 / 0.114 / %25.2	0.415 / 0.211 / %50.8
SDSP	0.142 / 0.056 / %39.4	0.435 / 0.129 / %29.7	0.367 / 0.227 / %61.9
CovSal-C	0.68 / 0.127 / %18.7	0.592 / 0.161 / %27.2	0.548 / 0.189 / %34.5
CovSal-CM	0.411 / 0.021 / %5.1	0.345 / 0.114 / %33.0	0.249 / 0.144 / %57.8

Table 4.8: Forestry category. Level: -2, Total input image: 42

Model Name	SSIM Index Mean / STD / CV	Cosine Similarity Mean / STD / CV	Pearson Correlation Similarity Mean / STD / CV
GBVS	0.198 / 0.093 / %46.9	0.542 / 0.098 / %18.1	0.484 / 0.142 / %29.3
Itti-Koch	0.101 / 0.056 / %55.4	0.36 / 0.144 / %40.0	0.187 / 0.264 / %141.2
Achanta	0.378 / 0.161 / %42.6	0.484 / 0.105 / %21.7	0.398 / 0.146 / %36.7
ImageSignature -Lab	0.088 / 0.026 / %29.5	0.334 / 0.154 / %46.1	0.222 / 0.201 / %91.3
ImageSignature -RGB	0.09 / 0.037 / %41.1	0.433 / 0.097 / %22.4	0.349 / 0.204 / %58.5
SDSP	0.136 / 0.045 / %33.1	0.406 / 0.127 / %31.3	0.282 / 0.26 / %92.2
CovSal-C	0.636 / 0.113 / %17.8	0.556 / 0.129 / %23.2	0.501 / 0.157 / %31.3
CovSal-CM	0.422 / 0.022 / %5.2	0.365 / 0.146 / %40.0	0.262 / 0.172 / %65.6

Since, the subjects deal with less details in water surface images, their eye movement data is collected much more on actual informative salient regions rather than forestry images. Therefore, smaller filter is sufficient to eliminate additional detected pixels in order to provide more approximative solution to reality. CovSal uses larger filter than GBVS, it also reduce the pixels in actual salient regions.

Tables 4.9 - 4.11 also show that GBVS is the best fit model for water surface images at zoom levels 0,-1 and -2 individually. Even though CovSal-C model has a mean value higher than 0.5 for each comparision method, it is also clear that GBVS model provides a higher mean value with a smaller coefficient of variation percentage than CovSal-C model.

Table 4.9: Water Surface category. Level: 0, Total input image: 40

Model Name	SSIM Index Mean / STD / CV	Cosine Similarity Mean / STD / CV	Pearson Correlation Similarity Mean / STD / CV
GBVS	0.437 / 0.211 / %48.3	0.661 / 0.123 / %18.6	0.638 / 0.153 / %23.9
Itti-Koch	0.243 / 0.18 / %74.1	0.594 / 0.139 / %23.4	0.592 / 0.161 / %27.2
Achanta	0.698 / 0.231 / %33.1	0.446 / 0.147 / %32.9	0.382 / 0.199 / %52.1
ImageSignature (Lab)	0.207 / 0.101 / %48.8	0.333 / 0.154 / %46.2	0.278 / 0.193 / %69.4
ImageSignature (RGB)	0.297 / 0.203 / %68.0	0.591 / 0.104 / %18.0	0.568 / 0.127 / %22.0
SDSP	0.267 / 0.154 / %57.7	0.601 / 0.097 / %16.1	0.583 / 0.116 / %19.9
CovSal-C	0.77 / 0.145 / %18.8	0.637 / 0.148 / %23.2	0.612 / 0.172 / %28.1
CovSal-CM	0.406 / 0.037 / %9.1	0.261 / 0.136 / %52.1	0.17 / 0.159 / %93.5

Table 4.10: Water Surface category. Level: -1, Total input image: 40

Model Name	SSIM Index Mean / STD / CV	Cosine Similarity Mean / STD / CV	Pearson Correlation Similarity Mean / STD / CV
GBVS	0.478 / 0.207 / %43.3	0.696 / 0.102 / %14.7	0.692 / 0.107 / %15.5
Itti-Koch	0.228 / 0.202 / %88.6	0.591 / 0.158 / %26.7	0.597 / 0.193 / %32.3
Achanta	0.712 / 0.219 / %30.8	0.542 / 0.152 / %28.0	0.524 / 0.174 / %33.2
ImageSignature -Lab	0.205 / 0.127 / %61.9	0.403 / 0.144 / %35.7	0.369 / 0.16 / %43.4
ImageSignature -RGB	0.317 / 0.217 / %68.4	0.609 / 0.082 / %13.5	0.611 / 0.093 / %15.2
SDSP	0.33 / 0.211 / %63.9	0.606 / 0.098 / %16.8	0.597 / 0.118 / %19.8
CovSal-C	0.828 / 0.096 / %11.6	0.686 / 0.145 / %21.1	0.67 / 0.156 / %23.3
CovSal-CM	0.415 / 0.03 / %7.2	0.334 / 0.161 / %48.2	0.271 / 0.185 / %68.3

Table 4.11: Water Surface category. Level: -2, Total input image: 40

Model Name	SSIM Index Mean / STD / CV	Cosine Similarity Mean / STD / CV	Pearson Correlation Similarity Mean / STD / CV
GBVS	0.458 / 0.209 / %45.6	0.708 / 0.106 / %14.9	0.705 / 0.118 / %16.7
Itti-Koch	0.228 / 0.183 / %80.3	0.579 / 0.181 / %31.3	0.581 / 0.227 / %39.1
Achanta	0.698 / 0.226 / %32.3	0.602 / 0.141 / %23.4	0.593 / 0.156 / %26.3
ImageSignature -Lab	0.168 / 0.107 / %63.7	0.488 / 0.146 / %29.9	0.455 / 0.167 / %36.7
ImageSignature -RGB	0.271 / 0.204 / %75.3	0.618 / 0.087 / %14.1	0.627 / 0.095 / %15.2
SDSP	0.271 / 0.173 / %63.8	0.614 / 0.122 / %19.9	0.609 / 0.179 / %29.4
CovSal-C	0.805 / 0.128 / %15.9	0.672 / 0.168 / %25.0	0.649 / 0.186 / %28.7
CovSal-CM	0.411 / 0.024 / %5.8	0.315 / 0.184 / %58.4	0.246 / 0.209 / %84.9

CHAPTER 5

CONCLUSION & FUTURE WORK

In this thesis, we evaluated the saliency map performances by calculating between the computational saliency models and Tobii T120 eye tracker device.

We firstly constituted a dataset containing remotely sensed RGB images in forestry and water surface categories. The experiment then was conducted for forestry and water surface categories separately. 20 human subjects participatedto the experiment. Each image in the dataset was presented to the subjectsparticipated in the experiments employing Tobii T120 eye tracker device. Individual andweighted gaze plots form the output saliency map of Tobii T120 for each image in the dataset.

Secondly, each saliency map methods that we investigated in this thesis were executed in MATLAB and their saliency maps were gathered for the images in the dataset.

After completing the experiment and obtaining all saliency maps, we have measured the similarities between each saliency maps produced by computational methodsand the Tobii T120 saliency map for both forestry and water surface categories by three different similaritymeasurementswhich are Cosine correlation, Pearson correlation and Structural SIMilarity index.

Finally, results were analyzed and best fit computational models are determined for forestry and water surface categories.

In forestry category, textured background contains more details than water surface category. Therefore, the subjects focus on non-informative regions until they find the actual salient region. Since CovSal-C model measures the similarity between the regions extracted from the input image, it eliminates non-salient regions better than Graph-based Visual Saliency method. However, for water surface category, the subjects do not focus on additional details as much as the images in the forestry category. Hence, gaze plot data is collected on the actual salient region much more than forestry category. Graph-based Visual Saliency method applies a smaller Gaussian Kernel filter to the saliency maps than CovSal-C method. In order to eliminate additional pixels which are detected as salient for the saliency maps produced by the computational models in water surface category, smaller Gaussian filter is sufficient. If the filter size become larger, main gaze plots collected on the actual salient regions are also eliminated. By considering the image category structres and the algorithms of the computational saliency methods, CovSal-C is the best model for forestry category while Graph-based Visual Saliency method is the best fit model for water surface category.

Salient region detection is used in GIS to detect objects such as airplane, ship, cloud, road and to predict vegatation anomalies, deforestration and land degradation. According to the results in this thesis, Graph-based Visual Saliency method can approximate the reality in object detection based GIS applications. For the GIS applications of vegetation, CovSal-C model can provide a better result corresponding to the reality. The dataset used in this thesis can be extented by adding desert and highland based remotely sensed RGB images. Desert based images can provide less detailed figure-background compared to highland based images. Therefore, object recognition based GIS applications such as airplane, man made structures or road detection can be developped by using Graph-based Visual Saliency method for desert images while using CovSal-C model for highland images.

REFERENCES

Asoka, A., & Mishra, V. (2015). Prediction of vegetation anomalies to improve food security and water management in India. *Geophysical Research Letters*,*42*(13), 5290-5298.

Biskupska, M. (2013). Bottom-up saliency maps: a review. *Elektronika: konstrukcje, technologie, zastosowania*, *54*(7), 53-57.

Borji, A., Sihite, D. N., & Itti, L. (2013). Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study. *Image Processing, IEEE Transactions on*, *22*(1), 55-69.

Burt, P. J., & Adelson, E. H. (1983). A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics (TOG)*, *2*(4), 217-236.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision research*,*51*(13), 1484-1525.

Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons.*Journal of neurophysiology*, *88*(5), 2530-2546.

Cui, X., Tian, Y., & Ma, L. (2014). Top-down visual saliency detection in optical satellite images based on local adaptive regression kernel. *Journal of Multimedia*, *9*(1), 173-180.

Erdem, E., & Erdem, A. (2013). Visual saliency estimation by nonlinearly integrating features using region covariances. *Journal of vision*, *13*(4), 11-11.

Few, S. (2006). Multivariate Analysis Using Heatmaps. *Perceptual Edge,* 1-8.

Goshtasby, A. Ardeshir. "Similarity and dissimilarity measures." *Image registration*. Springer London, 2012. 7-66.

Greco, L., & La Cascia, M. (2013). Saliency Based Aesthetic Cut of Digital Images. In *Image Analysis and Processing–ICIAP 2013* (pp. 151-160). Springer Berlin Heidelberg.

Harel, J., Koch, C., & Perona, P. (2006). Graph-based visual saliency. In*Advances in neural information processing systems* (pp. 545-552).

Hatipoglu, P., Aytekin, Ö, Ulusoy, İ, & Halici, U. (2014). Saliency Analysis For High Resolution Satellite Images With Challenging Contents. *ICT Innovations 2014 Web Proceedings,* 97-106.

Hou, X., Harel, J., & Koch, C. (2012). Image signature: Highlighting sparse salient regions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *34*(1), 194-201.

Hu, X., & Tao, V. (2007). Automatic extraction of main road centerlines from high resolution satellite imagery using hierarchical grouping. *Photogrammetric Engineering and Remote Sensing*, *73*(9), 1049.

Hu, X., Wang, Y., & Shan, J. (2015). Automatic Recognition of Cloud Images by Using Visual Saliency Features. *Geoscience and Remote Sensing Letters, IEEE*, *12*(8), 1760-1764.

Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature reviews neuroscience*, *2*(3), 194-203.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (11), 1254-1259.

Julier, S. J., & Uhlmann, J. K. (1996). *A general method for approximating nonlinear transformations of probability distributions*. Technical report, Robotics Research Group, Department of Engineering Science, University of Oxford.

Koch, C., & Ullman, S. (1987). Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of intelligence* (pp. 115-141). Springer Netherlands.

Krakov, D., & Feitelson, D. G. (2013, May). Comparing performance heatmaps. In *Job Scheduling Strategies for Parallel Processing* (pp. 42-61). Springer Berlin Heidelberg.

Lee, D. K., Itti, L., Koch, C., & Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nature neuroscience*, *2*(4), 375-381.

Le Meur, O., & Baccino, T. (2013). Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior research methods*, *45*(1), 251-266.

Li, C., Huang, R., Ding, Z., Gatenby, J. C., Metaxas, D. N., & Gore, J. C. (2011). A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI. Image Processing, IEEE Transactions on, 20(7), 2007-2016.

Li, J., & Gao, W. (2014). Object-Based Visual Saliency Computation. In *Visual Saliency Computation* (pp. 73-100). Springer International Publishing.

Li, W., Xiang, S., Wang, H., & Pan, C. (2011, September). Robust airplane detection in satellite images. In *Image Processing (ICIP), 2011 18th IEEE International Conference on* (pp. 2821-2824). IEEE.

Li, Z., Yang, D., & Chen, Z. (2015, August). Multi-layer Sparse Coding Based Ship Detection for Remote Sensing Images. In *Information Reuse and Integration (IRI), 2015 IEEE International Conference on* (pp. 122-125). IEEE.

Liu, H., Song, D., Rüger, S., Hu, R., & Uren, V. (2008). Comparing dissimilarity measures for content-based image retrieval. In Information Retrieval Technology (pp. 44-50). Springer Berlin Heidelberg.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval* (Vol. 1, No. 1, p. 496). Cambridge: Cambridge university press, 120-121.

Olkkonen, H., & Pesola, P. (1996). Gaussian pyramid wavelet transform for multiresolution analysis of images. *Graphical Models and Image Processing*,*58*(4), 394-398.

Ouerhani, N., Von Wartburg, R., Hügli, H., & Müri, R. (2004). Empirical Validation of the Saliency-based Model of Visual Attention. *Electronic Letters on Computer Vision and Image Analysis, 3*(1), 13-24.

Preim, B., & Botha, C. P. (2013). *Visual Computing for Medicine: Theory, Algorithms, and Applications*. Newnes, 117-118.

Qi, S., Ma, J., Lin, J., Li, Y., & Tian, J. (2015). Unsupervised Ship Detection Based on Saliency and S-HOG Descriptor From Optical Satellite Images.*Geoscience and Remote Sensing Letters, IEEE*, *12*(7), 1451-1455.

Qi, S., Ma, J., Tao, C., Yang, C., & Tian, J. (2013). A robust directional saliency-based method for infrared small-target detection under various complex backgrounds. *Geoscience and Remote Sensing Letters, IEEE*, *10*(3), 495-499.

Radha, D., Amudha, J., & Jyotsna, C. (2014). Study of Measuring Dissimilarity between Nodes to Optimize the Saliency Map. *Int.J.Computer Technology & Applications, 5*(3), 993-1000.

Rao, Y. R., Prathapani, N., & Nagabhooshanam, E. (2014). Application of normalized cross correlation to image registration. *International Journal of Research in Engineering and Technology*, *3*(05), 12-16.

Rajashekar, U., Van Der Linde, I., Bovik, A. C., & Cormack, L. K. (2008). GAFFE: A gaze-attentive fixation finding engine. *Image Processing, IEEE Transactions on*, *17*(4), 564-573.

Riche, N., Duvinage, M., Mancas, M., Gosselin, B., & Dutoit, T. (2013). Saliency and human fixations: state-of-the-art and study of comparison metrics. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1153-1160).

Rigas, I., Economou, G., & Fotopoulos, S. (2013). Low-level visual saliency with application on aerial imagery. *Geoscience and Remote Sensing Letters, IEEE*, *10*(6), 1389-1393.

Romantan, M., Vigouroux, B., Orza, B., & Vlaicu, A. (2002). Image indexing using the general theory of moments. In *Proceedings 3rd COST 276 Workshop on Information and Knowledge Management for Integrated Media Communications* (pp. 108-113).

Sharma, A., & Ghosh, J. K. (2015). Saliency Based Segmentation of Satellite Images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *2*(3), 207.

Shen, X., & Wu, Y. (2012, June). A unified approach to salient object detection via low rank matrix recovery. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 853-860). IEEE.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological review*, *95*(1), 15.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention.*Cognitive psychology*, *12*(1), 97-136.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, *13*(4), 600-612.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search.*Psychonomic bulletin & review*, *1*(2), 202-238.

Zhang, F., Du, B., & Zhang, L. (2015). Saliency-guided unsupervised feature learning for scene classification. *Geoscience and Remote Sensing, IEEE Transactions on*, *53*(4), 2175-2184.

Zhang, L., Gu, Z., & Li, H. (2013, September). SDSP: A novel saliency detection method by combining simple priors. In *Image Processing (ICIP), 2013 20th IEEE International Conference on* (pp. 171-175). IEEE.

Zhao, R., Ouyang, W., Li, H., & Wang, X. (2015). Saliency detection by multi-context deep learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1265-1274).

Zuva, K., & Zuva, T. (2012). Effectiveness of Image (dis)similarity Algorithms on Content-Based Image Retrieval. *International Journal of Engineering and Science*, *1*(1), 31-35.

ONLINE REFERENCES

Computational Vision at CALTECH,
http://www.vision.caltech.edu/~harel/share/GBVS.php, last visited on May 2016

FizzyMagic, http://www.fizzymagic.net/Geocaching/FizzyCalc/, last visited on May 2016

GMapCatcher an offline map viewer, https://code.google.com/p/gmapcatcher/, last visited on May 2016

Hacettepe University Department of Computer Engineering,
http://web.cs.hacettepe.edu.tr/~erkut/projects/CovSal/, last visited on May 2016

IVRG - Images And Visiual Representation Group,
http://ivrlwww.epfl.ch/supplementary_material/RK_CVPR09/index.html, last visited on May 2016

MIT Saliency Benchmark, saliency.mit.edu/results_old.html, last visited on May 2016

Tongji University, http://sse.tongji.edu.cn/linzhang/va/SDSP/SDSP.htm, last visited on May 2016

APPENDIX A

FORESTRY CATEGORY IMAGES

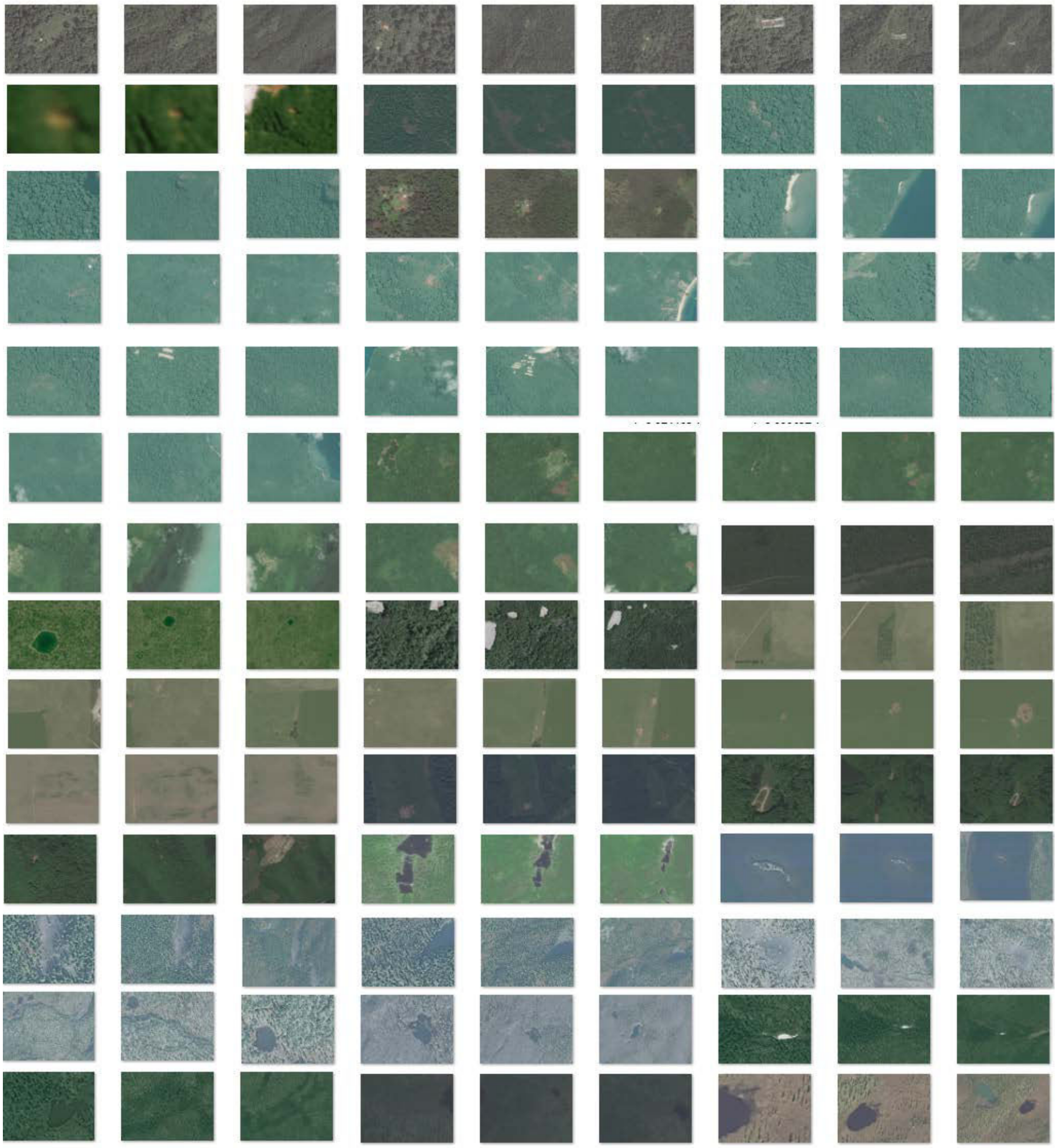


Figure A.1: The input images of forestry category at zoom levels 0, -1 and -2

APPENDIX B

WATER SURFACE CATEGORY IMAGES

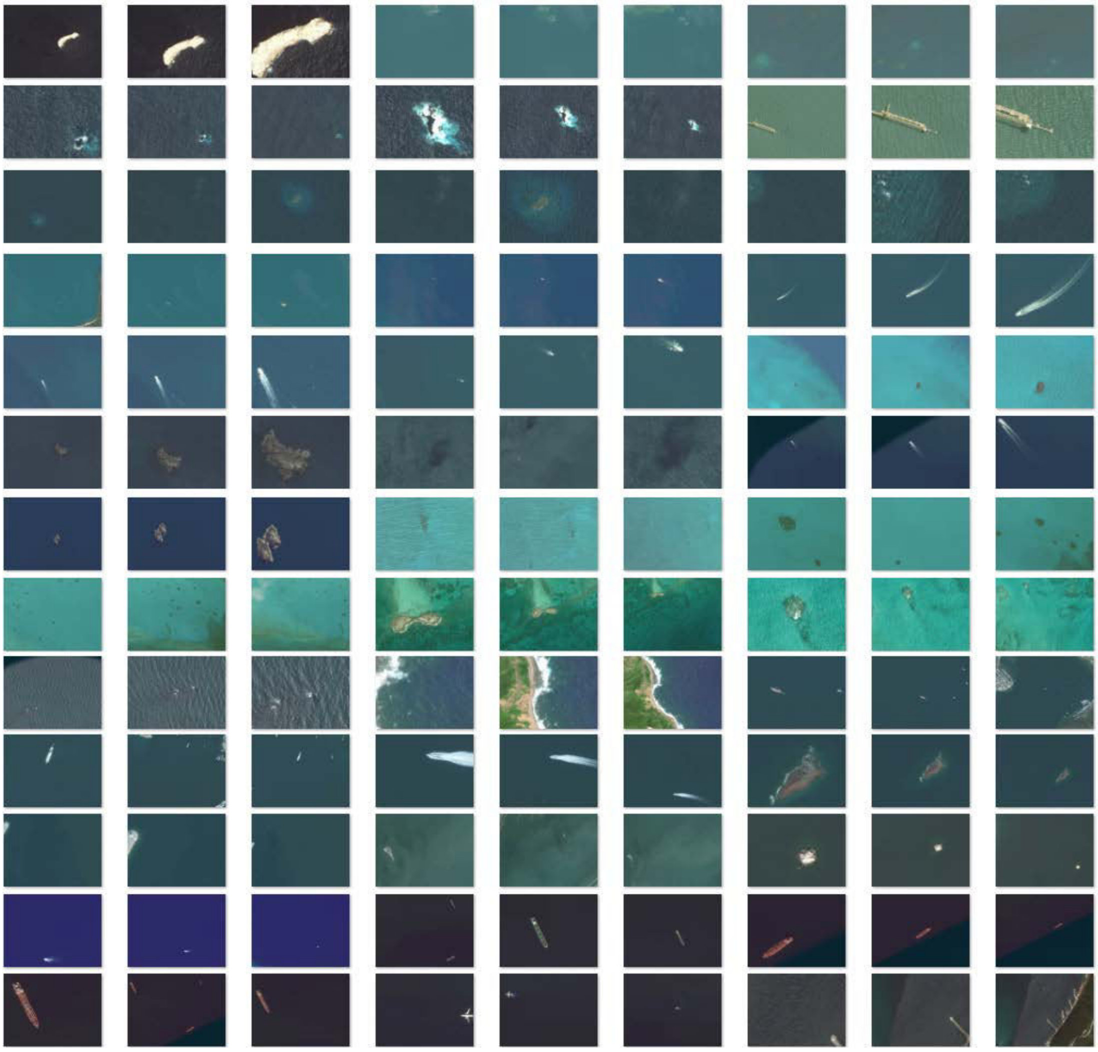


Figure B.1: The input images of forestry category at zoom levels 0, -1 and -2