Rochester Institute of Technology

RIT Scholar Works

Theses

2007

Identification and Ranking of Relevant Image Content

Mustafa I. A. Jaber

Follow this and additional works at: <https://scholarworks.rit.edu/theses>

Recommended Citation

Jaber, Mustafa I. A., "Identification and Ranking of Relevant Image Content" (2007). Thesis. Rochester Institute of Technology. Accessed from

This Thesis is brought to you for free and open access by RIT Scholar Works. It has been accepted for inclusion in Theses by an authorized administrator of RIT Scholar Works. For more information, please contact ritscholarworks@rit.edu.

IDENTIFICATION AND RANKING OF RELEVANT IMAGE CONTENT

by

Mustafa I. A. Jaber

A Thesis submitted in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

In

ELECTRICAL ENGINEERING

Approved by:

Prof.

Thesis Advisor – Dr. Eli Saber

Prof.

Thesis Committee Member – Dr. Sohail A. Dianat

Prof.

Electrical Engineering Department Head – Dr. Vincent Amuso

Department of Electrical Engineering,
The Kate Gleason College of Engineering,
Rochester Institute of Technology,
Rochester, New York

MAY 2007

THESIS AUTHOR PERMISSION STATEMENT

Title of Thesis : Identification and Ranking of Relevant Image Content
Name of Author : Mustafa I. A. Jaber
Degree : Master of Science
Program : Electrical Engineering
College : Kate Gleason College of Engineering

I understand that I must submit a print copy of my thesis or dissertation to the RIT Archives, per current RIT guidelines for the completion of my degree. I hereby grant to the Rochester Institute of Technology and its agents the non-exclusive license to archive and make accessible my thesis or dissertation in whole or in part in all forms of media in perpetuity. I retain all other ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Print Reproduction Permission Granted:

I, Mustafa I. A. Jaber, hereby **grant permission** to the Rochester Institute of Technology to reproduce my print thesis or dissertation in whole or in part. Any reproduction will not be for commercial use or profit.

Signature of Author: Mustafa I. A. Jaber Date: _____

Inclusion in the RIT Digital Media Library Electronic Thesis & Dissertation (ETD) Archive

I, Mustafa I. A. Jaber, additionally grant to the Rochester Institute of Technology Digital Media Library (RIT DML) the non-exclusive license to archive and provide electronic access to my thesis or dissertation in whole or in part in all forms of media in perpetuity.

I understand that my work, in addition to its bibliographic record and abstract, will be available to the world-wide community of scholars and researchers through the RIT DML. I retain all other ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation. I am aware that the Rochester Institute of Technology does not require registration of copyright for ETDs.

I hereby certify that, if appropriate, I have obtained and attached written permission statements from the owners of each third party copyrighted matter to be included in my thesis or dissertation. I certify that the version I submitted is the same as that approved by my committee.

Signature of Author: Mustafa I. A. Jaber Date: _____

DEDICATION

To my parents, family and friends.

To the Academy for Educational Development.

ACKNOWLEDGMENTS

I would like to express my gratitude to the wonderful people who made this achievement possible. It is a pleasure to present my sincere thanks to Dr. Eli Saber for advising my thesis and giving his support and guidance. Special thanks for Dr. Sohail Dianat for his comments and review throughout this work. I would also like to thank Dr. Vincent Amuso and Dr. Eric Peskin for their feedback.

I would like to extend my gratitude to Mr. Mark Shaw, Hewlett Packard – Color & Imaging, Boise, Idaho, for his continuous support and advice.

Many thanks to my colleagues Mr. Manoj Reddy, Mr. Luis Garcia, and Mr. Guru Balasubramanian for their wonderful ideas and comments.

Finally, I am grateful to the Electrical Engineering Department at the Rochester Institute of Technology and the Hewlett Packard Company for supporting and sponsoring this work.

Mustafa I. A. Jaber

ABSTRACT

The work in this thesis proposes an image understanding algorithm for automatically identifying and ranking different image regions into several levels of importance. Given a color image, specialized maps for classifying image content namely: weighted similarity, weighted homogeneity, image contrast and memory color maps are generated and combined to provide a perceptual importance map. Further analysis of this map yields a region ranking map which sorts the image content into different levels of significance.

The algorithm was tested on a large database that contains a variety of color images. Those images were acquired from the Berkeley segmentation dataset as well as internal images. Experimental results show that our technique matches human manual ranking with 90% efficiency.

Applications of the proposed algorithm include image rendering, classification, indexing and retrieval. Adaptive compression and camera auto-focus are other potential applications.

TABLE OF CONTENTS

Thesis Author Permission Statement	ii
Dedication	iii
Acknowledgments	iv
Abstract	v
Table of Contents	vi
List of Figures	viii
List of Table	ix
Chapter 1 Introduction.....	1
1.1 Motivation.....	1
1.2 Contributions.....	1
1.3 Thesis Organization	2
Chapter 2 Literature Review and Background.....	3
2.1 Literature Review.....	3
2.2 Background	5
2.2.1 Segmentation Algorithm.....	5
2.2.2 Quadtree Decomposition	7
Chapter 3 Human Preference Ranking Experiment	9
3.1 Image Set	9
3.2 Subjects	9
3.3 Experiment	10
3.4 Method	10
3.5 Findings.....	12
3.5.1 Importance Levels.....	12
3.5.2 Spatial Position Map	13
3.5.3 Image/Object Scale	14

Chapter 4 Proposed Ranking Algorithm.....	15
4.1 Pre-processing Module	15
4.2 Importance Map Generation Module.....	18
4.2.1 Weighted Similarity Map.....	18
4.2.2 Weighted Homogeneity Map.....	21
4.2.3 Image Contrast Map.....	22
4.2.4 Memory Color Map	24
4.3 Region Ranking Module	26
Chapter 5 Results and Discussions.....	29
Chapter 6 Conclusions and Future Work	33
References	34

LIST OF FIGURES

Figure 1: Overview of proposed approach	2
Figure 2: Overview of the segmentation algorithm in use	6
Figure 3: Quadtree decomposition	7
Figure 4: Result of region ranking experiment	10
Figure 5: Image used in the experiment, Sydney opera house	11
Figure 6: Spatial position map	13
Figure 7: Block diagram of the proposed algorithm	17
Figure 8: Weighted similarity map module	20
Figure 9: Numerical example shows the weighted similarity map generation	21
Figure 10: Weighted homogeneity map module	23
Figure 11: Memory color map module	24
Figure 12: Region ranking module	27
Figure 13: Example of the histogram segmentation procedure	28
Figure 14: Results of the proposed algorithm for portrait-like images	30
Figure 15: Results of the proposed algorithm for landscape images	32

LIST OF TABLES

Table 1: Experiment result of the Sydney opera house image in Figure 4	12
Table 2: Estimated parameters	25
Table 3: Match to human manual ranking	31

CHAPTER 1: INTRODUCTION

1.1 MOTIVATION

The performance of many systems in image processing can be enhanced by adding more intelligence to the process. Only certain image region(s) are of visual interest to an average observer; also referred to as *Region of Interest* (ROI) which is automatically computed using the psychometric aspects of input image segments. Importance Map (IM), a criterion to classify image regions in relation to their perceptual importance [1], is another representation of regions of visual interest.

1.2 CONTRIBUTIONS

In this thesis, a segmentation map of the input image and the IM are used to automatically identify image regions and rank them according to their perceptual importance into different levels; referred to as *Region Ranking Map* (RRM). This saves the intelligence and ensures simplicity of the allocation. The algorithm is based on a human preference experiment that allocates regions of different priority to the average observer. Experiment findings such as the importance level of objects and their spatial positions will be recorded and studied. These findings, in addition to low level features from literature, are used to develop and evaluate the proposed algorithm. The algorithm consists of three stages. Firstly, the input image is segmented and converted to a suitable color space at a preprocessing stage. Secondly, an IM is computed. It is a combination of four maps: *weighted similarity*, *weighted homogeneity*, *image contrast*, and *memory color*. Finally, a region ranking module that uses the segmented image to rank and sort

the IM regions into different levels according to their importance is utilized. The flowchart in Figure 1 provides an overview of the proposed approach.

Figure 1: Overview of proposed approach.

The proposed algorithm assigns different levels of significance to objects and regions of the image in hand while ROI algorithms in literature usually give the result as a mask which presents the main ROI and rejects all other regions (considered as background). The segmentation map is not required to generate the IM in the proposed algorithm. It also differs from the IM algorithms proposed in [1] and developed in [4], [5] in terms of sorting image regions adaptively into certain levels of significance. Using experiments to develop and evaluate the proposed algorithm is another major contribution in this work.

1.3 THESIS ORGANIZATION

The rest of this thesis is organized as follows. Chapter 2 provides the literature review and background. Chapter 3 presents the experiment. The proposed algorithm is described in Chapter 4. Chapter 5 demonstrates results and discussion. Conclusions are drawn in Chapter 6.

CHAPTER 2: LITERATURE REVIEW AND BACKGROUND

2.1 LITERATURE REVIEW

Osberger and Maeder [1] used image segmentation to define the IM. Regions' contrast, size, shape, and location, in addition to foreground / background classification are used to assign a priority value for the segmented regions. The IM is generated by finding the squared sum of these factors. This approach showed promising results but it was limited by the success of the segmentation in use and was applied to gray-scale images only. Similar algorithm is used in [2] to assess image quality. A Neural-network based approach is used in [3] to segment image regions and to characterize the perceptual importance of particular regions. Data training and user feedback are required for this approach. Nyguen *et al.* [4] tried to overcome the segmentation drawback by using four low level features to generate the IM. However, these features (contrast, relative brightness, variance and edge density) are application specific and cannot be extended to image understanding in general. Pardo [5] enhanced the metric proposed in [1] and used it to extract semantic objects in previously segmented images.

Many computational, context-free approaches tried to automatically identify the region of interest. A significant work on defining visual ROI using computational algorithms was done by Privitera and Stark [6]. They proposed a set of features based on a series of experiments. They concluded that four tools: wavelets, symmetry, contrast, orientation

and edges per unit area are important to identify ROI. However, they did not combine these features to check the overall performance relative to the human visual system. ROI was determined in [7] to be use in image retrieval. It combines color histograms with the spatial information as criteria of region importance. Marques *et al.* [8] identified ROI based on extracting the most salient points within an image. Wavelet transform is used in identifying ROI [9], [10]. Xiangyangt *et al.* [9] proposed an image retrieval algorithm based on the ROI using the Discrete Wavelet Transform (DWT). They used color, texture, and position as features for determining ROI. Moreover, Zhang *et al.* [10] used the *wavelet modulus maxima* edge detection and mean shift color region segmentation technique to identify ROI. However, this approach is limited to low depth of field and landscape images only.

Computational approaches were also proposed to model the human *Visual Attention* (VA) system. VA systems are presented in [11], [12], [13] as computational models of visual salience, a map of the most salient points in an image. Color, intensity, and orientation were used to form the saliency map in these systems. Ko *et al.* [14] created an *attention window* based on the distribution of salient points in the image where color, texture, normalized area, location, and shape of the segmented regions are used to extract salient regions and determine their importance scores. The dissimilarity between neighborhoods in an image is used in [15], [16] to present a visual attention algorithm. Similar approach was proposed to identify quantification of DNA damage in cells as ROI [17]. Bradley and Stentiford [18] developed the VA system in [15], [16] and introduced it in the JPEG 2000 coding algorithm. It is based on suppressing areas of the image with patterns that are

repeated elsewhere. Torralba [19] used a *global scene configuration* to model attention guidance and showed that the low-level features can be used to predict the location, scale, and appearance of objects in the scene. Han *et al.* [20] formulate the attention objects as a Markov Random Field by integrating computational visual attention mechanisms with attention object growing techniques. A new technique to extract objects of visual importance is proposed in [21]. Salient objects are extracted by applying a segmentation algorithm on a combination of image edge and color maps. The algorithm showed promising results but a quantitative evaluation is needed to measure the performance.

2.2 BACKGROUND

As previously mentioned, a segmentation algorithm is required at the preprocessing step. In addition, the *weighted homogeneity map* used in finding the IM is based on the Quadtree decomposition. Summaries of the segmentation algorithm in use and the Quadtree decomposition are introduced in the following Sections.

2.2.1 Segmentation Algorithm

The *Dynamic Color Gradient Thresholding* (DCGT) segmentation algorithm [22] is used in this work. This approach employs vector-based color gradient method [23] and Otsu's automatic threshold [24] to perform a dynamic threshold-based segmentation. It segments the image by placing emphasis on the use of color-homogenous regions and color transitions without generating edges. The use of color gradient to aid in the region growing process rather than for generating edges avoids issues of thresholding and disconnected edges.

A weighted vector-based color gradient map is used to provide the groundwork upon which seeds are generated and region growing is automated. Seeds here refer to 4-neighborhood connected pixels where gradient is below a specified threshold. A dynamic threshold operator is applied to this gradient map to govern the growth process. To ensure consistency of the segmentation with the image regions, region growing is followed by a similarity measure-based region-merging step. This produces an optimally segmented image. Figure 2 shows the block diagram of the segmentation algorithm used in this work as given in [22].

Figure 2: Overview of the segmentation algorithm in use.

2.2.2 Quadtree Decomposition

Quadtree (Q_t) decomposition is defined in [25] as a simple technique to represent an image at different levels of resolution. A natural gray-level image usually can be divided into different size non-overlapped square blocks with variable amount of details and information. The union of these blocks is the entire image [26]. Q_t decomposition is a powerful technique which divides the image into 2-D homogeneous (in the property of interest) regions. Block variance and a certain threshold are some possible test criteria that can be used to determine the homogeneity of each block.

Figure 3: Quadtree decomposition, (a) data objects set, (b) the corresponding Quadtree data structure.

An example showing the procedures of finding the Q_t decomposition of an image can be summarized as follows. Suppose a set S of n data objects in a given plane. Also, let R denote a square region that contains all the data objects of S . The Q_t data structure is a partition tree T such that the root r of T is associated with the region R . Higher levels in T are obtained by subdividing R into four equal-sized squares R_1, R_2, R_3 , and R_4 , and each square R_i is associated with a potential child of the root r . Specifically, a child v_i of r is created, if the square R_i contains a point in S . If a square R_i contains no points in S , then we do not create R_i . This process of refining R into the squares R_1, R_2, R_3 , and R_4 is called a *split*. Figure 3 illustrates an example of data objects and the associated Quadtree. Details of Q_t decomposition can be found in [27].

CHAPTER 3: HUMAN PREFERENCE RANKING

EXPERIMENT

The experiment aims to identify and rank image regions into different levels of importance. In other words, to answer the question: “which image regions constitute the most important segments?” Findings of this experiment helped in building the proposed algorithm.

3.1 IMAGE SET

A set of fifty images with different sizes were used. The set consisted of a variety of portraits and landscape images. Portrait images refer to face only images, mug shots, and profiles for males and females of different ages. Landscape images consist of indoor settings for people and objects, as well as outdoor settings, people, city scenes, and buildings. Paintings, digital, and 3D art were also used in the experiment. Images vary from simple to complex. Simple images have an obvious important object with a simple background, while complex images may show multiple objects of interest with or without a complex background.

3.2 SUBJECTS

Eight human subjects, six males and two females, participated in this experiment. They are between 22 and 40 years of age. Three of them have considerable professional experience in the image processing field. Observers were not familiar with the images.

3.3 EXPERIMENT

In the experiment, four colored markers (red, green, blue, and black) were used to manually outline the perceived segments or regions. The condition was to use all colored markers, wherever possible. There was no time restriction to complete the experiment. Five levels of importance were identified in the following order: red, green, blue, black, and 'without'. The observers were asked to keep a suitable reading distance (around 12 inches when printed on 8.5 X 11 inches) while doing the experiment.

Figure 4: Result of region ranking experiment, (a) original image, (b) manually outlined image.

An example illustrated in Figure 4 shows the original image in addition to the experimental result where the red region (the man's face) is determined to be the most important region whereas the frame (left untouched) is least important for one of the observers.

3.4 METHOD

The procedure used to collect and analyze experiment results is introduced here. For each image in the data set, a table of all image objects and segments is created. Ranking result

(level) of that object is recorded for all observers that participated in the experiment. The average value of these readings is utilized as the significance level of the object. The entire process is illustrated by the example shown in Figure 5 and Table 1. The figure shows the *Sydney Opera House*. S_1 to S_8 are the subjects' results where 4 – 0 in descending order means marker colors: red, green, blue, black and 'without', respectively which represent levels of importance as red, green, blue, gray, and black.

Table 1 shows that an average observer recognizes the Opera House as the most important object and the bridge as the second level of significance. The background (water and sky) in Figure 5 is assigned to the fourth level of importance.

Figure 5: Image used in the experiment, Sydney opera house (nationalgeographic.com).

Table 1: Experiment result of the Sydney opera house image in Figure 4.

Object	S₁	S₂	S₃	S₄	S₅	S₆	S₇	S₈	S_{Avg}	Rank
Opera House	4	3	4	4	4	4	4	4	3.875	4
Bridge	3	2	3	3	2	4	0	3	2.500	3
Water with light	2	1	0	0	3	3	0	2	1.375	1
Water	0	1	0	2	3	3	0	2	1.375	1
Lights	0	0	4	0	1	2	0	0	0.875	1
City in background	0	0	0	0	0	1	0	0	0.125	0
Sky	0	1	0	0	3	3	0	1	1.000	1
Red lights	0	4	4	0	1	2	0		1.571	2
Title (Text)	0	0	0	0	0	0	0	1	0.125	0

3.5 FINDINGS

Three features were chosen and examined to drive findings from the conducted experiment viz. importance levels, spatial position and image/object scale. The first feature shows how objects from different categories have different levels of importance. For example, at which level of importance would a human face probably be? The second feature ascertains where the position of objects at a certain level of importance would be. The third feature studies the effect of image/object scale on the observer's behavior.

3.5.1 Importance Levels

Objects in an image are classified according to their visual importance to different levels. However, boundaries between these levels are not rigid. For example, a region may be classified as the first level of importance by one observer and second by another. First or second importance level include features that identify the human face and body, such as skin, hair, and clothes. Objects with a large size at the image center are classified to the first or second level as well. The second or third level consists of natural components

with unusual color such as a red or golden sky, green sea, etc. the third or fourth level contains images with people relatively small in size, people in the background and natural components such as sky, clouds, land, water and grass. Backgrounds (simple and complex), frames, and text (titles, logos, and dates) are usually labeled as the fourth or fifth level of importance.

Figure 6: Spatial position map, (a) SPM in 3D modeled by Gaussian distributions, (b) SPM where white means the highest priority and black means the least priority, (c) SPM where 0 = least important and 4= most important.

3.5.2 Spatial Position Map

Spatial Position Map (SPM) is found by analyzing the manually outlined image set. Image region is sub-divided into nine non-overlapping blocks as shown in Figure 6c. The major levels in each block are averaged and saved as SPM for that image. The average

across all SPMs of images used in the experiment generates the general SPM as given in Figure 6c. It represents different image regions by numbers where the image center has the first level of importance, the middle outer thirds along the larger dimension have the third level of importance and the middle thirds along the smaller dimension have the fourth value of priority. Furthermore, the experiment showed that corners are the least important regions in the image. See Figure 6b. the mathematical model of this map is detailed in Section 4.1.

3.5.3 Image/Object Scale

Analyzing experiment outputs shows that every observer reacts differently to the same object at different scales. For a fixed image size, if the main object is given in a large scale (relatively covers the whole image); the observer gives attention to all objects in details and assigns them into different levels of importance. On the other hand, if the same object is given in a smaller scale in the same image, the observer assigns all segments of the main object into one level of importance. In a similar manner, image size has a similar effect. If an image is printed with different resolutions, observers ranked objects in the image with higher resolution (greater size) into more detailed regions than the images with lower resolution (smaller size).

CHAPTER 4: PROPOSED RANKING ALGORITHM

The block diagram of the proposed algorithm is shown in Figure 7. The algorithm consists of three stages. Input image is segmented and converted to a suitable color space at the preprocessing module. The SPM and the size of the processing block are also found at the first stage. The second module finds the IM which is created by combining four feature maps (*weighted similarity*, *weighted homogeneity*, *image contrast*, and *memory color*). Finally, the region ranking map is generated by combining the IM with the segmented image.

4.1 PRE-PROCESSING MODULE

Finding the IM requires transforming the input image to the *YES* color domain [28]. Y represents the luminance channel and E , S denote the chrominance components. This space has been chosen as it reduces variations in chrominance due to changes in luminance. It is defined as a linear transformation from *RGB* color space [28].

$$\begin{bmatrix} Y \\ E \\ S \end{bmatrix} = \begin{bmatrix} 0.253 & 0.684 & 0.063 \\ 0.500 & -0.500 & 0.000 \\ 0.250 & 0.250 & -0.500 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

The DCGT segmentation algorithm [22] is used to find the segmentation map at this stage. Other segmentation algorithms can be found in [29], [30], [31].

Han *et al.* [20] modeled the saliency map by a normalized Gaussian function with the center located at the image center. However, the experiment showed that this modeling is

good but not fully accurate. SPM in this work is modeled using six Gaussian distributions with mean of 127 and variance 60 for an 8 bit image. Four Gaussians (set a) are located at the one-third corners and the other two (set b) are located along the center of middle outer thirds of the image (along the larger dimension, see Figure 6b). Weights of set a are ones while weights assigned to set b are one-half. This weighting scheme awards highest priority to the center region (consistent with the experiment findings). Tests show that better results are obtained by multiplying these weights by the entropy of the area that the corresponding Gaussian covers. The proposed SPM does not depend on the color domain of the input image, thus it can be computed before or after the conversion to the *YES* color space.

We assume that the quality of the final output will be judged by a human. Thus the reading distance comes to the picture and so the size of the *processing block* (B) which is proportional to the size of the image in hand.

For an input image with size $M_1 \times M_2$ and block size $L \times L$, number of blocks in M_1 and M_2 direction are defined as K_1 and K_2 , respectively.

$$K_1 = \frac{M_1}{L}, \quad K_2 = \frac{M_2}{L} \quad (2)$$

Figure 7: Block diagram of the proposed algorithm. Lighter regions represent higher importance; priority is represented in colors in descending order: red, green, blue, gray, and black.

4.2 IMPORTANCE MAP (IM) GENERATION MODULE

The importance map is a criterion that represents the image regions in relation to their perceptual importance. It is utilized as a weighted sum of different features of an image of size $M_1 \times M_2$ as given in (3)

$$IM(i, j) = \frac{\sum_{k=1}^N w_k(i, j) F_k(i, j)}{\sum_{k=1}^N w_k(i, j)} \quad (3)$$

where F_k is a feature, w_k its weighting factor, N is the number of features. $i = 0, 1, 2, \dots, M_1 - 1, j = 0, 1, 2, \dots, M_2 - 1$. Due to the lack of prior knowledge about the significant feature, they will be assigned equal weights. The features proposed in this algorithm are *weighted similarity map*, *weighted homogeneity map*, *contrast map*, and *memory color map*. Detailed discussion of above features follows:

4.2.1 Weighted Similarity Map (WSM)

The experiment (see Section 2) gives a clear identification for ROI in terms of spatial position. It shows that if an object is located at image center and has a unique color and texture, it gains the highest priority. Therefore, the WSM assigns different weights to object in the input image according to their spatial position. These weights are represented by the SPM, see Figure 6.

The measure of similarity of a region across the entire image is defined by this similarity map. Each channel is divided into square windows of size $L \times L$. For input image with size $M_1 \times M_2$, the numbers of blocks are K_1 and K_2 in M_1 and M_2 directions, respectively. Norm of B is given by [32]

$$\|B(k_1, k_2)\| = \sqrt{\lambda_{\max}(B(k_1, k_2)^T B(k_1, k_2))} \quad (4)$$

where $\lambda_{\max}$ is the largest eigenvalue, $k_1 = 0, 1, \dots, K_1-1$, $k_2 = 0, 1, \dots, K_2-1$. This results in three maps for Y , E and S channels. For consistency, the maps are scaled to the range 0 – 255.

The similarity map represents similar regions with same norm values. However, to test the uniqueness of regions in image center, SPM and a similarity distance d are included in modeling the *weighted similarity map* (WSM). For each channel, it is found as follows: for each value of $\|B(k_1, k_2)\|$ in the similarity map, all positions within the intensity range $(\|B(k_1, k_2)\| - d, \|B(k_1, k_2)\| + d)$ are saved in W_c and used to locate the corresponding pixel values in the SPM.

The Weighted Similarity Map is defined as the average of the corresponding values of W in the SPM and is given by

$$WSM(k_1, k_2) = \frac{1}{C} \sum_{c=1}^C SPM(W_c) \quad (5)$$

where C is the number of pixels in W_c , $k_1 = 0, 1, \dots, K_1-1$, $k_2 = 0, 1, \dots, K_2-1$. This procedure gives three WSM for the three different channels. The average of channels' WSM is scaled to the range 0 – 255 and used as the image WSM.

This procedure gives three WSM for the different channels. The average of channels' WSM is scaled to the range 0 – 255 and used as the image WSM. Figure 8 shows the flowchart of modeling the WSM for a color image.

Figure 8: Weighted similarity map module.

Example: Figure 9 shows numerical values for the norm and SPM. The most important region (image center) is shown as 10 while regions with the least priority are presented as ones in the SPM. For $d = 1$ and $n_1 = 25$, the positions in the SPM that corresponds to norm values in the range 24 – 26 are averaged and stored as the WSM for $n_1=25$. Using the same procedure, the result is shown for $n_2 = 4$ and $d = 1$.

Figure 9: Numerical example shows the weighted similarity map generation, (a) norm matrix, (b) spatial position map, (C) weighted similarity map.

4.2.2 Weighted Homogeneity Map (WHM)

A weighted Quadtree decomposition of the input image is used to find the *weighted homogeneity map* (WHM). The criterion used to determine the homogeneity of each block in the decomposition is given in (6) where α is a threshold value.

$$\max(Block) - \min(Block) \leq \alpha \quad (6)$$

The Quadtree decomposition of a rectangular image results in blocks of different sizes. A region with larger blocks indicates high homogeneity and hence less information. This defines a criterion such that the smaller the size of the block, the higher its (importance) weight. Accordingly, the region near strong edges will have the highest priority.

However, the module fails when an edge occurs near image borders such as the frame of a portrait image. Therefore, scaling the weighted Quadtree image with the SPM overcomes this drawback and gives the weighted homogeneity map. This procedure gives three WHM for the different channels. The average of channels' WHM is scaled to the range 0 – 255 and used as the image WHM. The approach to compute the WHM for image is illustrated in Figure 10.

4.2.3 Image Contrast Map

Region contrast is a strong low-level visual attractor [1], [4], [6]. Regions with high contrast attract more attention than their neighbors and therefore gain higher visual importance. The luminance channel in *YES* color domain is used to estimate the relevant contrast of an image region to the overall image brightness. *IC* is computed as follows:

$$IC(k_1, k_2) = \left| (Y_m(k_1, k_2) - Y_M) / Y_M \right| \quad (7)$$

where Y_m is the mean value in the square window B of size $L \times L$, Y_M is the overall mean luminance of the image, $k_1 = 0, 1, \dots, K_1 - 1$, and $k_2 = 0, 1, \dots, K_2 - 1$. This map is also scaled to the range 0 – 255.

Figure 10: Weighted homogeneity map module.

4.2.4 Memory Color Map

The memory color map (MCM) is based on the algorithm proposed by Saber *et al.* [28] where a statistical model classifies pixels into four different classes viz. skin, sky, grass (memory colors) and the ‘other’ class. Gaussian distributions are used to model the memory color classes in the *YES* color space. The contour of the Gaussian of each class i defines an ellipse (8). It maps the *ES* domain into a scalar statistic λ at each pixel g .

$$[x_g - m_i]^T K_i^{-1} [x_g - m_i] = \lambda_g^i \quad (8)$$

Where,

$$x_g = \begin{bmatrix} E_g \\ S_g \end{bmatrix}, \quad m_i = \begin{bmatrix} m_E \\ m_S \end{bmatrix}, \quad \text{and} \quad K_i = \begin{bmatrix} \sigma_E^2 & \sigma_{ES} \\ \sigma_{ES} & \sigma_S^2 \end{bmatrix} \quad (9)$$

Figure 11: Memory color map module.

Table 2: Estimated parameters.

Class	Skin	Sky	Grass
$m_i = \begin{bmatrix} m_E \\ m_S \end{bmatrix}$	$m_1 = \begin{bmatrix} 26.01 \\ 25.34 \end{bmatrix}$	$m_2 = \begin{bmatrix} -20.58 \\ -35.46 \end{bmatrix}$	$m_3 = \begin{bmatrix} -8.97 \\ 20.15 \end{bmatrix}$
$K_i = \begin{bmatrix} \sigma_E^2 & \sigma_{ES} \\ \sigma_{ES} & \sigma_S^2 \end{bmatrix}$	$K_1 = \begin{bmatrix} 12.32 & 6.51 \\ 6.51 & 8.82 \end{bmatrix}$	$K_2 = \begin{bmatrix} 3.13 & 3.09 \\ 3.09 & 8.88 \end{bmatrix}$	$K_3 = \begin{bmatrix} 6.18 & 0.12 \\ 0.12 & 10.43 \end{bmatrix}$
Universal Threshold	$t_u^1 = 18.42$	$t_u^2 = 27.63$	$t_u^3 = 9.21$

Seventy-five colored images are used as a training set to find the mean vector and covariance matrix of E & S channels for skin tones, sky, and grass classes. The estimated parameters of the training data are shown in Table 2.

The value of λ represent the probability that pixel g belongs to a given class. A small value of λ indicates that the color of pixel g is near to the center of the ellipse, and thereby likely to belong to that class and vice versa. λ_g^i is calculated for each pixel in the image and compared with the universal threshold t_u^i . If λ_g^i comes out to be less than t_u^i it is labeled as a pixel from class i else as the ‘other’ class. (The apriori probabilities of classes are assumed to be equal.)

Experiment findings (Section 2) form the memory color map. They assign higher priority to skin pixels (human face and hands are at the first level of importance) and lower level of importance to sky and grass (usually come as background in landscapes). Therefore, the MCM represents skin pixels as 255, sky and grass pixels as 0, while 128 is assigned to pixels from the ‘other’ class. The final MCM is scaled to the range 0 – 255. Figure 11 shows a flowchart of the MCM generation process.

4.3 REGION RANKING MODULE

Inputs to the ranking module are the segmented image (pre-computed) and IM (see Figure 12). An intermediate map is generated by averaging the importance values that correspond to each segment. This map has levels of importance equal to (or less than) the number of segments in the original segmented image. A histogram segmentation algorithm based on Otsu's automatic threshold [24] is used to classify segments of the intermediate map up to five levels as detailed below.

The histogram of the intermediate image is first split into 2 levels. These 2 levels are further divided in a similar fashion to yield 4 levels. This procedure is continued until 16 different levels are obtained. A simple iterative merging process is applied on these threshold values. It finds the distance D_i between classes and merges the classes separated by the minimum distance. Values of thresholds between classes T_i are updated and the merging process is iterated until five classes (four values of T) remain.

$$D_i = T_{i+2} - T_i \quad (10)$$

where $i = 1, 2, \dots, 16$. Figure 13 illustrates an example of the histogram segmentation procedure in use where Figure 13a shows the histogram of an intermediate map (mean values of importance for image segments). Figure 13b shows the iterative splitting technique applied on the histogram in Figure 13a. It clarifies that T_8 is found in the first iteration, T_4 & T_{12} result from the second round and so on to find all the threshold values. T_5 and T_{11} are zeros which imply flat segment. The merging procedure is shown in Figure 13c where T_8 is the first threshold to be deleted. T_2 is deleted in the second iteration and so on to give five classes with thresholds T_4 , T_7 , T_{10} and T_{13} .

In order to avoid having insignificant region(s) in the RRM, an adaptive level adjustment module is applied. If the size of any region for a particular level is less than 5% of the total image size, pixels representing this level are merged into the lower level of importance, and the lower levels are promoted. Another step is to fill holes in the RRM where an area of a lower level is surrounded by pixels of higher level of importance. Hole sizes less than or equal to 3% of the image are filled.

Figure 12: Region ranking module.

Figure 13: Example of the histogram segmentation procedure, (a) histogram of an intermediate map, (b) iterative splitting technique, (c) iterative merging.

CHAPTER 5: RESULTS AND DISCUSSIONS

The proposed algorithm has been implemented using Matlab 7.2® and successfully illustrated on a database of 50 color images. RRM for seventeen simple and complex images from Berkeley segmentation dataset [33] have been shown in Figure 14 and Figure 15 using a similarity distance $d = 3$ (Section 4.2.1) and a threshold value $\alpha = 69$ (Section 4.2.2). Colored regions in the RRM represent priority in the following descending order: red, green, blue, gray and black. Results may not always have five levels of importance (simple images). The simulation results generally match the findings of the experiment (Section 3.5).

RRMs for simple portrait-like images are shown in Figure 14a & 14b where red color (highest priority) depicts the main object. Portrait images (Figure 14c & 14d) show that face, hair, body, and background are classified in descending priority order, which is the same ranking as found in the experiment. Figure 14e shows that the main object in the image is ranked to the highest level of importance while the background constitutes all other levels. Figure 14f and 14g show that the proposed algorithm assigns humans first and second levels of importance (red and green). Objects with large size at image center are classified to the first level (Figure 14g) while background comes at a lower levels of priority. A more complex scene is shown in Figure 14h where several disconnected similar objects in the image are assigned the same level of priority.

The proposed algorithm also shows commendable results for simple landscape images (Figure 15a -15c) where it ranks the object of interest at the highest importance level.

Figure 14: Results of the proposed algorithm for portrait-like images.

Although the bears in Figure 15b touch the lower image edge, they still have more priority than the background (touches the other three edges). Figure 15d – 15f show results for more complex scenes where the main object consists of segments of different color and texture. Figure 15c assigns the man (face, hands and uniform) more importance than the background. Similar performance is found in Figure 15e & 15f. Performance of the proposed algorithm on landscapes with multiple objects of interest is shown in Figure 15g – 15i. It manages to rank the main objects in the scene into the first and second level of priority (red and green colors).

A subjective evaluation for the proposed algorithm is shown in Table 3. Eight observers evaluated the proposed algorithm across 50 different images. They were given the input image in addition to the RRM of the proposed algorithm and were asked to match them to what they anticipated. Their feedback was a score out of 100 for each image. Category 1 means results did not correlate well with human expectation; categories 2, 3 and 4 illustrate that the RRM matched some, several and all important objects in the scene, respectively. Table 3 shows that the proposed algorithm is effective on 90% of images in the database.

Table 3: Match to human manual ranking

Category	< 50%	50%-74%	75%-99%	100%
	1	2	3	4
Percentage	10%	19%	29%	42%

Figure 15: Results of the proposed algorithm for landscape images.

CHAPTER 6: CONCLUSIONS AND FUTURE WORK

We have proposed a new algorithm for identifying and ranking image regions according to their visual importance. Four feature maps (weighted similarity, weighted homogeneity, image contrast, and memory color maps) are used to generate the region ranking map. It presents image content into different levels of perceptual importance. The algorithm was tested on a database of 50 images with competitive performance. The average computational time is 15 seconds on an image of size 512 x 512 using MATLAB 7.2® running on a 3.2 GHz dual core processor machine.

The algorithm can be used to perform intelligent region classification, object identification and scene analysis. Future works include enhancing the generation of the importance map by utilizing information regarding image content and viewing perspective. Another potential improvement would be to have different weights for the four features depending on the applications. Automatic weight selection technique will be studied for applications such as image rendering or image data compression.

REFERENCES

- [1] W. Osberger and A. Maeder, "Automatic Identification of Perceptually Important Regions in an Image," *14th International Conference on Pattern Recognition*, 1998.
- [2] W. Osberger, N. Bergmann, and A. Maeder, "An Automatic Image Quality Assessment Technique Incorporating High Level Perceptual Factors," *IEEE International Conference on Image Processing*, 1998.
- [3] H. Wong and L. Guan, "Characterization of Perceptual Importance for Object-Based Image Segmentation," *IEEE International Conference on Image Processing*, Vol. 3, pp. 54-57, 2000.
- [4] A. Nyguen, V. Chandran, S. Sridharan, and R. Prandolini, "Importance Assignment to Regions in Surveillance Imagery to Aid Visual Examination and Interpretation of Compressed Images," *International Symposium on Intelligent Multimedia, Video and Speech Processing*, 2001.
- [5] A. Pardo, "Extraction of Semantic Objects from Still Images," *IEEE International Conference on Image Processing*, 2002.
- [6] C. Privitera and L. Stark, "Algorithms for Defining Visual Regions-of-Interest: Comparison with Eye Fixations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, issue 9, pp. 970 – 982, 2000.
- [7] X. Hongli, X. De, and G. Yong, "Region-Based Image Retrieval Using Color Coherence Region Vectors," *7th International Conference on Signal Processing*, Vol. 1, pp. 761- 764, 2004.
- [8] O. Marques, L.M. Mayron, G.B. Borba, and H.R. Gamba, "Using Visual Attention to Extract Regions of Interest in The Context of Image Retrieval," *ACM SE'06*, 2006.

- [9] W. Xiangyangt, Y. Hongying, and H. Fengli, "A New Regions-of-Interest Based Image Retrieval Using DWT," *IEEE International Symposium on Communications and Information Technology*, Vol. 1, pp. 132- 135, 2005
- [10] K. Zhang, H. Lu, M. Duan, and Q. Zhao, "Automatic Salient Regions of Interest Extraction Based on Edge and Region Integration," *IEEE International Symposium on Industrial Electronics*, 2006.
- [11] L. Itti, C. Koch, and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, pp. 1254-1259, 1998.
- [12] L. Itti and C. Koch, "A Saliency-Based Search Mechanism for Overt and Covert Shifts of Visual Attention," *Vision Research*, Vol. 40, pp. 1489-1506, 2000.
- [13] L. Itti and C. Koch, "Feature Combination Strategies for Saliency-Based Visual Attention Systems," *Journal of Electronic Imaging*, Vol. 10 (1), pp. 161–169, 2001.
- [14] B. Ko, S. Kwak, and H. Byun, "SVM-Based Salient Region(s) Extraction Method for Image Retrieval," *17th International Conference on Pattern Recognition*, 2004.
- [15] F. Stentiford, "An Estimator for Visual Attention Through Competitive Novelty with Application to Image Compression," *Picture Coding Symposium*, pp. 101–104, 2001.
- [16] F. Stentiford, "An Evolutionary Programming Approach to Simulation of Visual Attention," *Congress on Evolutionary Computation*, pp. 851–858, 2001.
- [17] F. Stentiford, N. Morley, and A. Curnow, "Automatic Identification of Regions of Interest with Application to the Quantification of DNA Damage in Cells," *SPIE*, Vol. 4662, pp. 244-253, 2002.

- [18] A. Bradley and F. Stentiford, "Visual Attention for Region of Interest Coding in JPEG 2000," *Journal of Visual Communication and Image Representation*, Vol. 14, pp. 232–250, 2003.
- [19] A. Torralba, "Modeling Global Scene Factors in Attention," *Journal of the Optical Society of America*, Vol. 20, no. 7, pp. 1407- 1418, 2003.
- [20] J. Han, K. Ngan, M. Li, and H. Zhang, "Unsupervised Extraction of Visual Attention Objects in Color Images," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 16, no. 1, 2006.
- [21] K. Park and Y. Moon, "Automatic Extraction of Salient Objects Using Feature Maps," *The 32nd International Conference on Acoustics, Speech, and Signal Processing*, pp. 617-620, 2007.
- [22] G. Balasubramanian, *Unsupervised Color Image Segmentation Using a Dynamic Color Gradient Thresholding Algorithm*, Thesis (M.S.)-Rochester Institute of Technology, 2006.
- [23] H. C. Lee and D. R. Cok, "Detecting Boundaries in a Vector Field," *IEEE Transactions on Signal Processing*, Vol. 39, no.5, pp. 1181–1194, 1991.
- [24] N. Otsu, "A Threshold Selection Method from Gray-Level Histogram," *IEEE Transactions on SMC*, pp. 62-66, 1978.
- [25] E. Shusterman and M. Feder, "Image Compression Via Improved Quadtree Decomposition Algorithms," *IEEE Transactions on Image Processing*, Vol. 3, issue 2, pp. 207-215, 1994.
- [26] W. Mahmoud and D. Jackson, "Improved Quadtree Decomposition / Recomposition Algorithm for Fractal Image Compression," *IEEE Southeastcon*, 1999.

- [27] M. T. Goodrich and R. Tamassia, *Algorithm Design: Foundations, Analysis, and Internet Examples*, New York City, John Wiley & Sons, 2002.
- [28] E. Saber, A. M. Tekalp, R. Eschbach, and K. Knox, "Automatic Image Annotation Using Adaptive Color Classification," *Graphical Models and Image Processing*, Vol. 58, no. 2, 1996.
- [29] E. Saber, A. M. Tekalp, and G. Bozdagi, "Fusion of Color and Edge Information for Improved Segmentation and Edge Linking," *Image and Vision Computing Journal*, Vol. 15, no. 10, 1997.
- [30] Y. Deng and B. S. Manjunath, "Unsupervised Segmentation of Color-Texture Regions in Images and Video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, pp. 800-810, 2001.
- [31] B. Sumengen, *Variational Image Segmentation and Curve Evolution on Natural Images*, PhD thesis, University of California, Santa Barbara, 2004.
- [32] G. Strang, *Linear Algebra and Its Application*, 4th Edition, Thomson Learning, 2006.
- [33] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A Database of Human Segmented Natural Images and Its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics," *IEEE Conference on Computer Vision*, 2001.