# Numerical Estimation of Visual Appeal of a Website

A PROJECT REPORT

Submitted by

**Mahima Bhargava (**15BCE1075**)**

in partial fulfillment for the award of the degree of

**Bachelor of Technology**

in

**Computer Science and Engineering**

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

**School of Computing Science and Engineering**

Vandalur - Kelambakkam Road, Chennai - 600 127

April - 2019

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

**School of Computing Science and Engineering**

# DECLARATION

I hereby declare that the project entitled “Numerical Estimation of Visual Appeal of a Website” submitted by me to the School of Computing Science and Engineering, VIT Chennai, 600127 in partial fulfilment of the requirements of the award of the degree of B.Tech CSE is a bona-fide record of the work carried out by me under the supervision of Dr. R. Jagadeesh Kannan. I further declare that the work reported in this project, has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma of this institute or of any other institute or University.

Place: Chennai                                                        Signature of Candidate (s)

Date:

(Mahima Bhargava )

School of Computing Science and Engineering

# CERTIFICATE

This is to certify that the report entitled "Numerical Estimation of Visual Appeal of a Website" is prepared and submitted by Mahima Bhargava (15BCE1075), to VIT Chennai, in partial fulfilment of the requirement for the award of the degree of B.Tech CSE programme is a bona-fide record carried out under my guidance. The project fulfils the requirements as per the regulations of this University and in my opinion meets the necessary standards for submission. The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma and the same is certified.

Signature of the Guide:

Name: Dr. R.Jagadeesh Kannan

Date:

Signature of the Internal Examiner
Name:
Date:

Signature of the External Examiner
Name:
Date:

Approved by the **B. Tech CSE Program Chair**

Name: Dr. B Rajesh Khanna
Date:

(Seal of SCSE)

# Acknowledgement

I wish to express my sincere thanks to a number of people without whom I could not have completed the thesis successfully.

I would like to place on record my deep sense of gratitude and thanks to my guide Dr. R.Jagadeesh Kannan, School of Computer Science and Engineering (SCSE), VIT, Chennai campus whose esteemed support and immense guidance encouraged me to complete the project successfully.

I would like to thank my Program Chair Dr.B Rajesh Kanna & Co-Chair, Dr. C Sweetlin Hemalatha, B. Tech. Computer Science and Engineering and Project Coordinator Dr. B V A N S S Prabhakar Rao, VIT Chennai campus, for their valuable support and encouragement to take up and complete the thesis.

Special mention to my Dean, Dr. Vaidehi Vijayakumar, & Associate Deans Dr. Vijayakumar V and Dr. Subbulakshmi T, School of Computing Science and Engineering (SCSE), VIT Chennai campus, for spending their valuable time and efforts in sharing their knowledge and for helping me in every minute aspect of software engineering.

I thank my management of VIT Chennai campus for permitting me to use the library resources. I also thank all the faculty members for giving me the courage and the strength that I needed to complete my goal. This acknowledgment would be incomplete without expressing the whole hearted thanks to my family and friends who motivated me during the course of my work.

I would also like to extend my appreciation to Dr Dipti Prasad Mukherjee (Professor & Head, Electronics and Communication Sciences Unit ,Indian Statistical Institute, Kolkata) for being the guiding force and providing me with domain knowledge, subject expertise, and validation of the results without which this project could not be completed. I am grateful for his patience, kindness and enthusiasm to work with a college student outside his working hours and his suggestions that have shaped my research.

Mahima Bhargava (15BCE1075)

## Table of Contents

**List of Tables**

**List of Figures**

**List of Abbreviations**

| Sr. No | Abbreviations | Full Form | Description |
|---|---|---|---|
| 1. | CCM | Color Complexity Measure | A feature which captures the color complexity of a region. |
| 2. | FLS | First Largest Segment | The largest segment/ region in a mean-shift segmented image |
| 3. | SLS | Second Largest Segment | The second largest segment/ region in a mean-shift segmented image |
| 4. | MSE | Mean Squared Error | Used to evaluate the accuracy of a model |
| 5. | HSI | Hue, Saturation, Intensity | A color space representation |
| 6. | MLR | Multiple Linear Regression | The statistical method for model generation. |

## EXECUTIVE SUMMARY

This study attempts at identifying image region features which influence visual complexity and colorfulness of a website and developing estimation models both these parameters individually. The attempt was initiated by exploring literature in this domain and selecting an approach which utilizes 22 features to capture the mechanism of human visual perception. These 22 features are categorized into three levels- global, local and salient regions. The global level features aim at capturing characteristics of the image which influence the first overall impression of the website for a user, without focusing on details. Local level features capture the details of various regions in an image. We have focussed on the first largest and second largest image segments of the website and extracted them by performing mean-shift segmentation on the original website screenshot. The local features are calculated based on these two regions. The salient region refers to the most visually vital region of a website image and salient features capture the characteristics of this region. The salient region is found by generating a saliency of the image, followed by employment of an automatic cropping algorithm which crops a rectangle of aspect ratio 16/16(a square) and retains 65% energy of the saliency map.

An experiment ,conducted via 4 surveys, has been designed which aims at collecting individual subjective ratings of visual complexity and colorfulness of all 40 websites, on a 9-point scale ranging from “Not at all Complex” to “Extremely Complex” and “Not at all Colorful” to “Extremely Colorful”, respectively. The average ratings for each website are calculated to be later employed for development of regression models.

Next, multiple linear regression is employed for development of models for complexity and colorfulness and features which influence the parameters individually have been identified. K-cross validation technique is employed for enhancing the robustness of the models.

## ABSTRACT

Aesthetics often influence the assessment of the usability and reliability of a product. Similarly, the appeal of a website has a great impact on the users as it influences the assessment of a website's trustworthiness and utility. A split second is sufficient for users' to form lasting judgments about a website's appeal without conscious noting of design details and content while seeing it for the first time . This first impression is influential enough to later affect their opinions of a site's usability and trustworthiness. Websites which seem beautiful are considered to be useful.

Visual complexity and colorfulness are two essential parameters that influence the visual appeal of a website. The aim is to estimate the objective measures of visual complexity and colorfulness of websites based on 22 global, local and salient features which represent the three factors- distribution of composition, colors and content of a website screenshot. In this paper, we have proposed models for estimation of the visual complexity and colorfulness of website screenshots objectively.

An experiment was conducted to collect the subjective ratings of visual complexity and colorfulness of a set of 40 website images and this data along with the 22 image features was employed to construct the models for the respective parameters. The results of the proposed models indicate that visual complexity values can be predicted with a mean squared error of 0.942 (5-fold cross-validation ) while colorfulness can be predicted with a mean squared error of 0.839 (5-fold cross validation).

# 1. INTRODUCTION

Advancement in technology has led to improved methods of catering to the needs of the target users. To ensure the growth of e-commerce, it is imperative to ensure ease of use and development of trust of the offered services and products. Although functionality and usability had been deemed essential in human-computer interaction, now, aesthetics too has been included as an important parameter due to its positive influence on people's behaviour and purchase intentions [1].

During present times, people are impatient and due to hectic schedules, prefer websites which are easy to discern and interact with. Often, presence of too many objects on the website confuses a new user, leading them to leave the website quickly. An organized and structured website, hence, would lure users, making them stay and explore it. Organized website is an indicator of an organized business, making users trust the site more.

Humans are always attracted to appealing things; be it a bright red sports car or a beautiful yellow flower. Consider two websites- in one, a flat 50% offer is written in normal black letters, in normal font ; in the other website, it is highlighted via bigger font and use of bright combination of colors. It is natural for the offer displayed in the latter to catch the user's, leading him/her to en-cash the offer. Thus, website aesthetics ends up playing a vital role in attracting users. Users extended stay often results in purchase and eventual satisfaction

Website aesthetics can be judged on various factors such as colorfulness, symmetry, balance, visual complexity, and others. In this study, visual complexity and colorfulness are the two factors which will be the focus points.

## 1.1 OBJECTIVES

The main aim is to objectively estimate the visual complexity and colorfulness of website screenshots.

The objectives are

- To extract the global region features for a set of 40 website screenshots
- To extract local features of a website screenshot with the help of EDISON System[2] and Image Region Analyzer App (Image Processing Toolbox- MATLAB)
- To extract the visually most important region in a website screenshot using its saliency map [3] and automatic cropping algorithm [4].

- To identify image region features which influence visual complexity and develop a model to estimate the same.
- To identify image region features which influence colorfulness of a website and develop a model to estimate the same.
- To ensure that the objective ratings of the respective parameters, obtained from respective models, present good estimates of the subjective assessment of the parameters by users. The subjective ratings concerned are collected via surveys design to suit this purpose.

## 1.2 BACKGROUND

With the rise of e-commerce, more and more online retailers are emerging. Website appeal has been acknowledged as an instrumental moderator of users' behaviour and perceptions. As such, online retailers need to ensure that their websites appeal to the users, leading their businesses to burgeon.

Website appeal refers to the quality of a website's perceptual features to generate user interaction, and attract users to a website[5]. Visual appeal influences the assessment of a website's trustworthiness and usability. Better visual appeal may be translated in the form of better commercial benefit for e-commerce system. A website appealing to users would attract repeated visits and stimulate positive purchase intentions, and establish higher standards of trust and usability [6].

First impression matters significantly as users are able to form immediate and consistent evaluation of the appeal of a webpage. The degree to which websites are deemed attractive by users after a short exposure remains the same even after a long exposure span [7]. Hence, ensuring positive impression on the user entails understanding the parameters that influence a user's first impression of a website which would help estimate user response for a website.

The need of the hour is to have a system which would enable website developers to design websites in a way that they appear to be highly appealing to probable customers and users visiting the site. A pre-requisite to development of such a system is the development of systems to estimate visual complexity and colorfulness as these are the two most essential factors which play a role in influencing visual appeal.

Visual complexity of a website can be considered as an inherent marker of cognitive load which greatly influence the user's interaction with the website[8]. Colorfuless while attracting users also influences the website's perceived trustworthiness [9],users' loyalty[9,10],and purchase intention[11]. Visual complexity and colorfulness are presented as measures which would enable designers to place the website on pre-defined complexity and colorfulness scales and gauge the improvement in their design.

### 1.3 MOTIVATION

The motivation behind this attempt was the lack of consideration of mechanism of human visual perception during development of predictive models of visual complexity and colorfulness of webites. Earlier attempts had mainly focused on distribution of spatial features without considering image features. In this paper, an attempt has been made to propose a model which takes image features on three levels- global, local and salient- to estimate the visual complexity as well as colorfulness of a website accurately.

Currently, the research for an optimum measure of appeal is still being conducted. The previous studies have significantly contributed to the research but have been unable to present a highly accurate result. It is thus, essential to improve the estimation model of the underlying parameters which influence appeal.

# 2. PROJECT DESCRIPTION AND GOALS

## 2.1 PROJECT DESCRIPTION

This project is an attempt to develop models which estimate visual complexity and colorfulness of a website screenshot objectively. The initial stage of the project aims as extraction of image region features on three levels- global, local and salient. These features had been designed in the study [12] to represent three factors- distribution of compositions, color and content which according to their survey, significantly affect the participants' assessment of complexity.

Global features refer to the attributes of the first impression people develop when they look at an image while local features reflect the regional information of an image. Salient region features represent the attributes of the most vital visual region of a given image. These extracted features are later employed as independent predictor variables for the dependent variables- visual complexity and colorfulness- in respective models. Multiple linear regression methodology has been employed along with k-fold cross validation technique for the identification of factors influencing visual complexity and colorfulness, respectively, and the coefficients of these factors for prediction purposes.

## 2.2 GOALS

The goals of the project include:

- Extraction of the image features for all 40 websites in the dataset
- Conducting of a survey to collect the subjective ratings of visual complexity and colorfulness , respectively, from participants for all websites
- Identification of features which influence visual complexity and colorfulness respectively
- Determination of best models for estimation of complexity and colorfulness, respectively and their validation using k-fold cross validation technique.
- Proposing justification for the influence of the features on complexity and colorfulness, respectively.
- Generating objective values of visual complexity and colorfulness such that it serves as an accurate estimate of subjective ratings for the same.

## 3. LITERATURE SURVEY

Estimation of website appeal has been the focus of many research attempts over the last few years. Most of them have employed distribution of spatial frequencies of the visual stimulus as factors for estimating complexity and colorfulness. A common such technique utilized was quad tree decomposition.

Previous research has demonstrated that a split second if sufficient to form a user's opinion without conscious noting of design details and content.[13] Websites which seem beautiful are considered to be useful. This is known as the "halo effect". For this purpose, low-level image statistics of static website screenshots[14] were employed to predict whether or not, the site will be liked by the user.

Zheng et al.'s [14] explored utilization of low-level image statistics(color, intensity and texture) for making judgements on a website's aesthetics. All images were analyzed in three steps beginning with calculation of low-level image statistics followed by quadtree decomposition. The concept of quadtree decomposition had been employed to iteratively decompose web-pages into quadrants of minimum entropy based on low-level image statistics. This decomposition allowed for description of these pages in terms of their respective organizational symmetry, balance and equilibrium. These parameters were assessed for their correlation with participants'subjective ratings of the same web pages on four aesthetic dimensions- Attractiveness, Pragmatic Quality, Hedonic Quality.
The study's evaluation has demonstrated that symmetry and balance have significant correlation with the user's ratings on aesthetics as compared to negligible relation to equilibrium.

Some studies have explored the concept of visual appeal in terms of visual appearance and aesthetic quality of a website. Reinecke et al., [15] claims that visual appeal of a website could be expressed in terms of visual complexity and colorfulness and demonstrates that aesthetics components of a website influence decisions of online users.

Although aesthetic component of a website is an essential parameter for determining effectiveness of a website, it isn't the only parameter. Reinecke et al [15] among other researchers have considered demographic variables such as gender, educational level and age to be contributing factors. To construct the computational models for measuring the perceived visual complexity and colorfulness, an online study had been conducted to gather pertinent data; 548 volunteers were asked to rate visual appeal, colorfulness and visual complexity of a group of 450 websites. The models for complexity and colorfulness were then employed to develop a computational model for estimating visual appeal. Demographic factors such as user's gender, age and educational level were also considered in addition and the model accounts for 40% variation in aesthetic preferences. Since the sample was insufficient, an in-depth analysis could not be conducted in terms of demographics.

Moss and Gunn [16] established via experiments that male participants favoured websites developed by males over those developed by females, and vice versa for females. The tendency for females to favour female-developed sites was greater than the male tendency to prefer male-developed websites. Hsiu-Feng [17] investigated the aesthetic preferences of children for websites designed for them. The results of the experiment demonstrated that overall, children preferred website of moderate visual complexity as opposed to low or high complexities. When the results were viewed while considering gender, it was found that boys preferred high complexity websites while girls preferred low to moderate complexity. This result is partially in tune with Berlyne's theory of aesthetic preference. On the other hand, Bauerly and Liu's [18] study discovered that higher counts of elements, that is, higher complexity lowers ratings of appeal. Websites with moderate ranges of perceived complexity are preferred by consumers. Berlyne's theory is consistent with this as it implied that there is an inverted 'U-shape' relationship between complexity and appeal where moderate complexity is favoured more.

Later, in another study [19], it was demonstrated that aesthetic impressions differ significantly across individuals and also depend upon demographic variables. Several demographic factors that influence people's visual preferences have been discovered, as well as how they affect appeal by pointing out several between-group differences in aesthetic proclivity. A huge heterogeneous sample with regards to age, occupation, education, geographic location, and web experience has been employed for analysis in this paper. The authors constructed a computational model that combines a user's demographic data with computational image metrics for evaluating a website's colorfulness and complexity in order to predict a user's subjective assessment of visual appeal.

# 4. EXPLORATION AND IMPLEMENTATION OF AN EXISITING MODEL

The model for estimation of visual appeal in [15,16] has successfully combined two major features- complexity and colorfulness for estimation along with demographic variables such as age, geographic factors and gender. This paper has shown how easily low-level image statistics and spatial distribution of a website can be employed for estimating website aesthetics . The data which has been a part of this study has been employed for implementation. This dataset includes extracted features such as number of leaves and average hue, as well as complexity and colorfulness values.

In this approach, first perceptual models of visual complexity and colorfulness are generated using regression.

In multiple linear regression, a dependent variable(Y) is calculated as a linear combination of 'p' independent variables($X_i$) known as predictors.

$$\boldsymbol{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon} \tag{1}$$

Here, $\beta_i$ denotes the weights given to each of the predictor and '$\varepsilon$' denotes the residual random error.

In matrix form, it is represented as

$$\boldsymbol{Y = X\beta + \varepsilon} \tag{2}$$

Each column of X contains the values for a particular independent variable. The vector $\beta$ is a vector of unknown constants to be estimated from the data. Each element $\beta_i$ is a partial regression coefficient reflecting the change in the dependent

variable per unit change in the $i^{th}$ independent variable assuming all other independent variables are held constant. We aim to generate $\beta$ values such that the

$\sum\varepsilon_i^2$ (sum of residual squares) is minimized. This is known as least square estimation method.

$$\sum\varepsilon_i^2 = \sum(Y_i - Y^{\wedge}_i)^2 \tag{3}$$

The vector of fitted values Y^ in a linear regression model can be expressed as

$$Y^{\wedge} = X\beta^{\wedge} \tag{4}$$

Where, the least-squares estimator of β is (in vector form)

$$\beta^{\wedge} = (X'X)^{-1}X'Y \tag{5}$$

Description of various factors employed in the study [15]:-

- Number of leaves- The final number of leaves calculated by the space-based decomposition , which recursively divides an image into N evenly spaced content regions(leaves), until a region has no visible space divider or until a region is too small. [29]
- Text area & Non text area – The number of leaves that have been classified as text or non-text based on a set of heuristics.
- Number of text groups- Refers to the number of horizontal groups of text characters. Each group may represent a word, one-line text, multiple lines of text, or a paragraph.
- Number of image areas - Estimates the number of leaves that the algorithm identifies as separate images. Several adjacent images are counted as one image area.
- Colorfulness [30] - The sum of the average saturation value and its standard deviation where the saturation is computed as chroma divided by lightness in the CIELab color space,
- Colorfulness [31] - The weighted sum of the trigonometric length of the standard deviation in ab space and the distance of the center of gravity in ab space to the neutral axis,

- Percentage of each of the 16 colors listed by W3C- Black, Silver, Gray , White, Maroon, Red, Purple, Fuchsia, Green, Lime ,Olive, Yellow, Blue , Teal, Aqua, Navy.
- Hue – Average Hue values of pixels of the image in HSV color space
- Saturation- Average saturation of pixels in HSV space
- Value- Average Value of pixels in HSV space

Complexity Model [15]:-

The following independent variables had been employed to calculate the dependent variable, i.e,complexity –

- Text area & Non text area
- Number of leaves
- Number of text groups
- Number of image areas
- Colorfulness [30]
- Hue

Colorfulness model [15]-:

The following independent variables had been employed to calculate the dependent variable, i.e, colorfulness-

- Color percentages for- Gray ,White, Maroon, Green, Lime ,Blue ,Teal
- Saturation
- Colorfulness [31]
- Number of image areas,
- Number of leaves,
- Text area,
- Non text area

In the present study, an attempt has been made to modify these regression models to attain better results with reduced number of independent predictors.

Current study colorfulness model:

Colorfulness is predicted using the predictors =>

- Percentage of colors- gray ,white ,maroon, green, lime, blue : These colors seem to influence colorfulness best when compared to the remaining W3c colors.
- Saturation- It helps differentiate between the strength of two colors
- value – It represents the lightness of an image and
- number of image areas- represents the areas which will have more concentrated regions of colors.

Figure 4.1 demonstrates that the colorfulness values generated by the model of the present study provides good estimates to the value generated in [15] with reduced number of parameters.

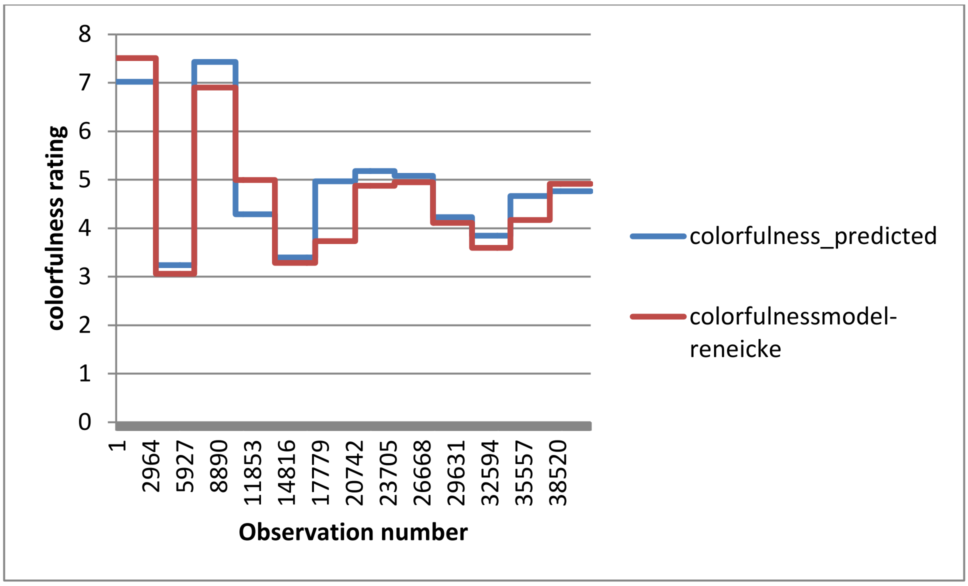



Figure 4.1: Demonstrates the difference in ratings between colorfulness values predicted in current study and the colorfulness ratings estimated in [15]

Current study complexity model-

Complexity is estimated using the predictors =>

- number of leaves by spatial decomposition- It is a marker of the number of objects present in an image,
- number of text groups- It is a measure of the number of homogenous, non-colorful area which require efforts for discrenment,
- colorfulness value estimated by the model in this study,

- non-Text Area,
- Text-area

Figure 4.2 demonstrates that the complexity values generated by the model of the present study provides good estimates to the value generated in [15] with reduced number of parameters.

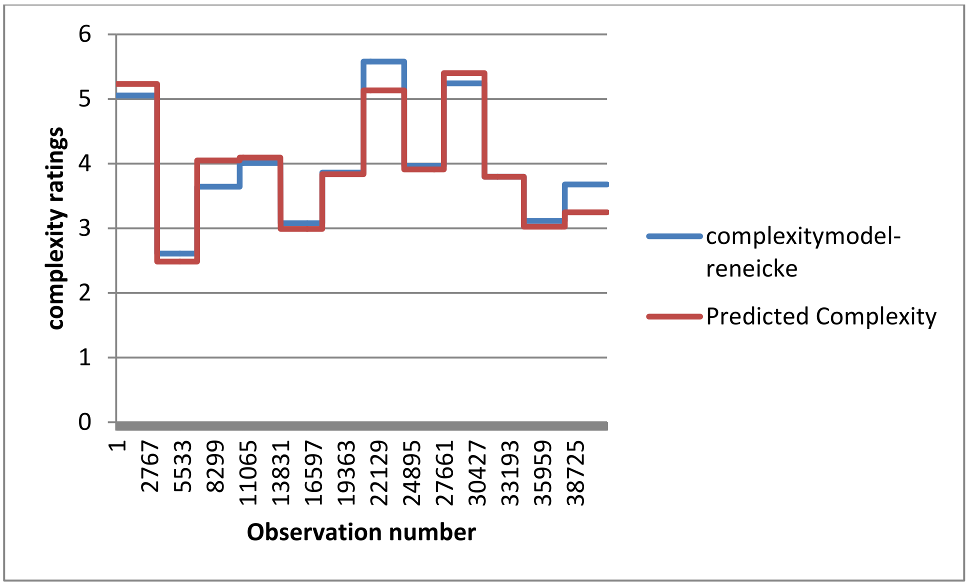



Figure 4.2: Demonstrates the difference in ratings between complexity values predicted in current study and the complexity ratings estimated in [15]

## 5. ENGINEERING DESIGN

The basic steps taken into account for the engineering design are as follows:

1. Problem Definition: The aim of the project scope was to find the relevant image features which influence visual complexity and colorfulness and development of predictive models for the same.

2. Background Research: Exploration of methods visual appeal estimation and computational models of colorfulness and complexity.

3. Gap Analysis : Mechanism of human visual perception had not been considered while development of models for visual complexity and colorfulness of website images.

4. Proposed Solution: Exploring image features as influencers of visual complexity and colorfulness, respectively.

5. Survey: Designing and launching surveys to collect the visual complexity and colorfulness for each of the 40 website images in the dataset.

6. Solution and Validation: Identifying the features which influence for visual complexity and colorfulness and construction of an appropriate regression model. Validation is done using k-fold cross validation process.

## 6. TECHNICAL SPECIFICATION

**Table 6.1: Software Specification**

| **Sr. No** | **Software** | **Version** | **Use case** |
|---|---|---|---|
| 1. | a.) MATLAB [25]<br>b.) MATLAB Image Region Analyzer APP [27] | 2018b | a.) For image processing and extraction of image features.<br>b.) For extracting Area and Perimeter of segments of the segmented image and generating binary masks to extract First largest segment and Second largest segment of the segmented website image respectively. |
| .2 | EDISON System[2] | - | For performing mean-shift segmentation on a website screenshot image and extracting the segmented image and boundary image. |
| 3. | R- Studio [24] | R version 3.4.3 | For development, implementation and validation of the predictive models for visual complexity and colorfulness of a website. |
| | | | |

# 7. DESIGN APPROACH AND DETAILS

## 7.1 EXPERIMENT: SUBJECTIVE ASSESSMENT OF THE PARAMETERS

To implement the first step of the proposed approach, an experiment was conducted to acquire the subjective assessment of the complexity and colorfulness of website images.

### 7.1.1 MATERIALS:

In order to ensure development of an accurate model, the subjective ratings of complexity and colorfulness are taken from participants. The stimuli consisted of a set of 40 website screenshots consisting of English language websites, foreign websites (using a different writing system), and websites that had been nominated for the Webby Awards in recent years. These websites have been taken from the study conducted in [4]. Websites selected had not received wide public exposure, represented a large variety of genres , and included a range of colorfulness and visual complexity levels.

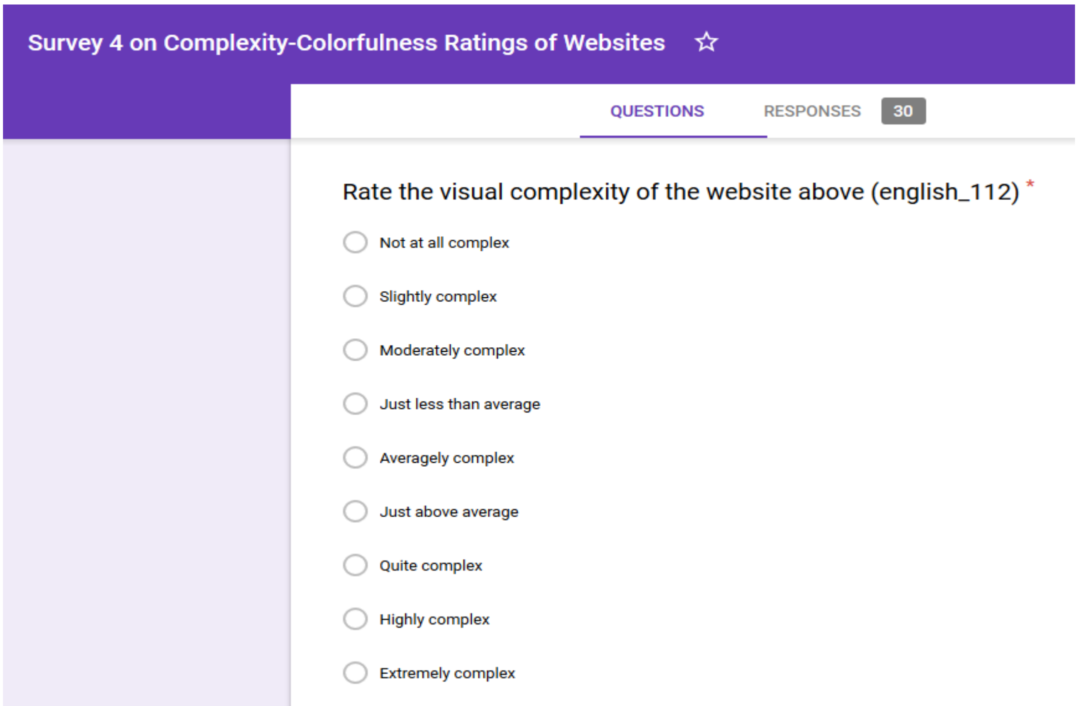
### 7.1.2 PROCEDURE:

The experiment was designed in the form of 4 surveys and conducted with the help of Google platform. Each survey had 10 websites from the dataset of 40 websites. Each survey began with two questions regarding the parameters which influence participants' opinion of complexity, followed by questions on their age and gender. The participants were also asked to select the factors which according to them, contributes to visual complexity. The five factors presented were:

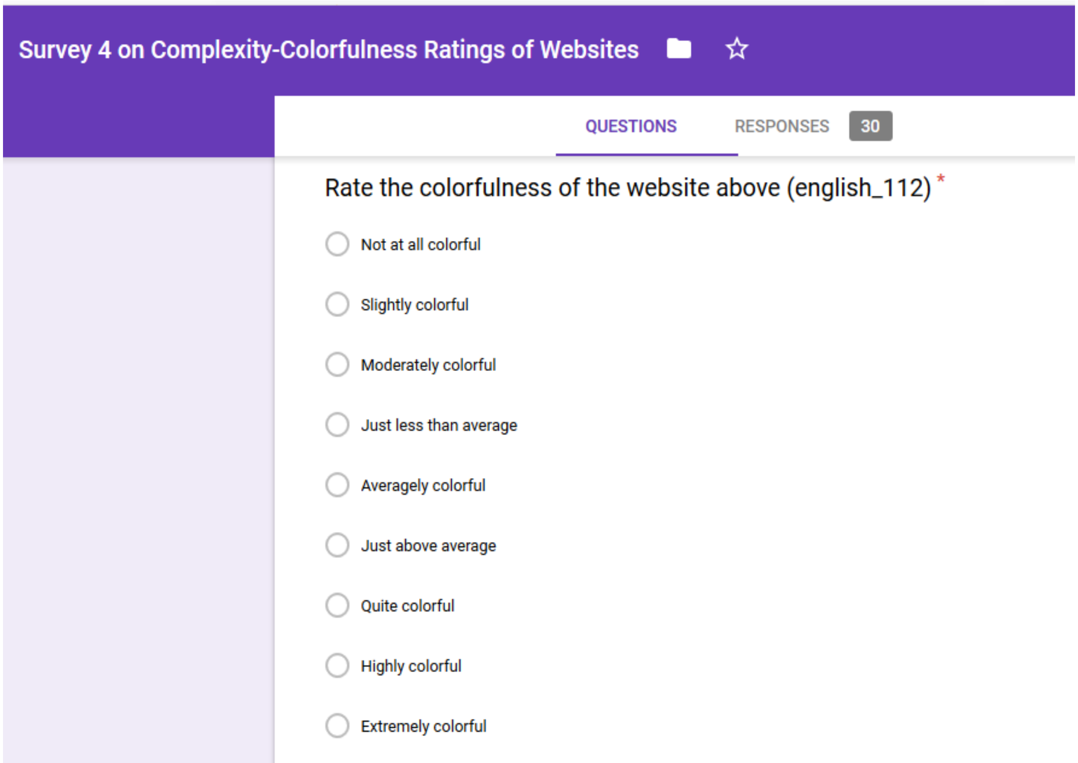
- Color
- Distribution of composition
- Symmetry
- Content and
- Contrast

Next, participants are requested to assess complexity and colorfulness of the presented screenshots using 9 point scales which ranged from 'Not at all complex' to 'Extremely complex' and 'Not at all colorful' to 'Extremely colorful' respectively. Figure 7.1 (a) and (b) displays these scales.

All stimuli were of their original size and presented on a white background.




(a)




(b)

Figure 7.1 : (a) The scale employed for collecting subjective visual complexity values, (b) The scale employed for collecting subjective colorfulness values

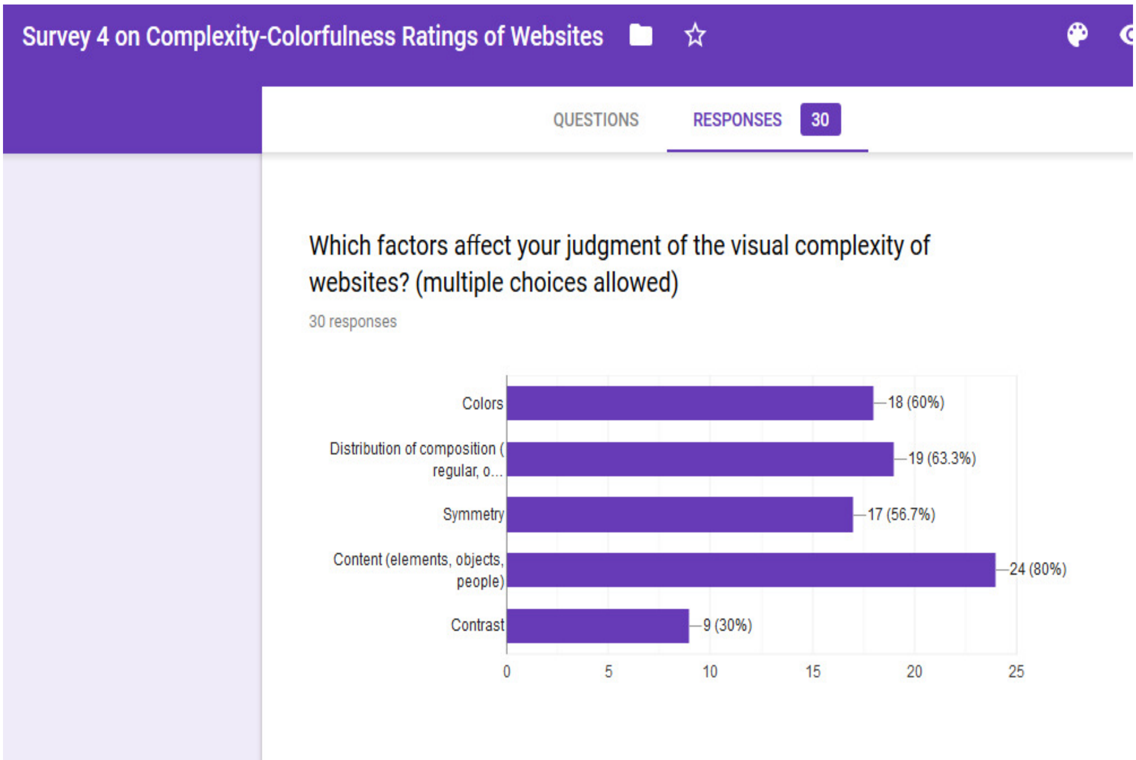
### 7.1.3 RESULTS:

The entries where saved in a Google sheet linked to the Google document.

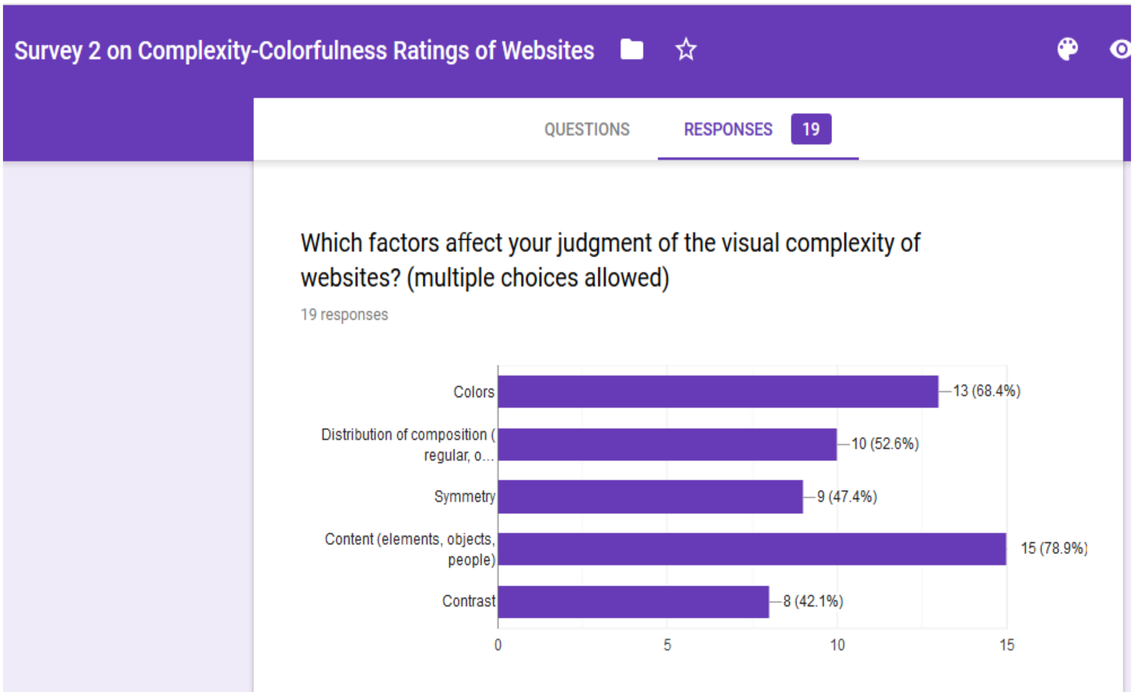
It was found that three factors- Distribution of composition, content and colors- received the highest votes as influencers of visual complexity. This result is consistent with the results in [21]. Figure 7.2 (a) and (b) shows these ratings for survey 4 and survey 2 respectively.

To represent these three factors, 22 features have been designed to estimate the visual complexity [21] and colorfulness of websites.

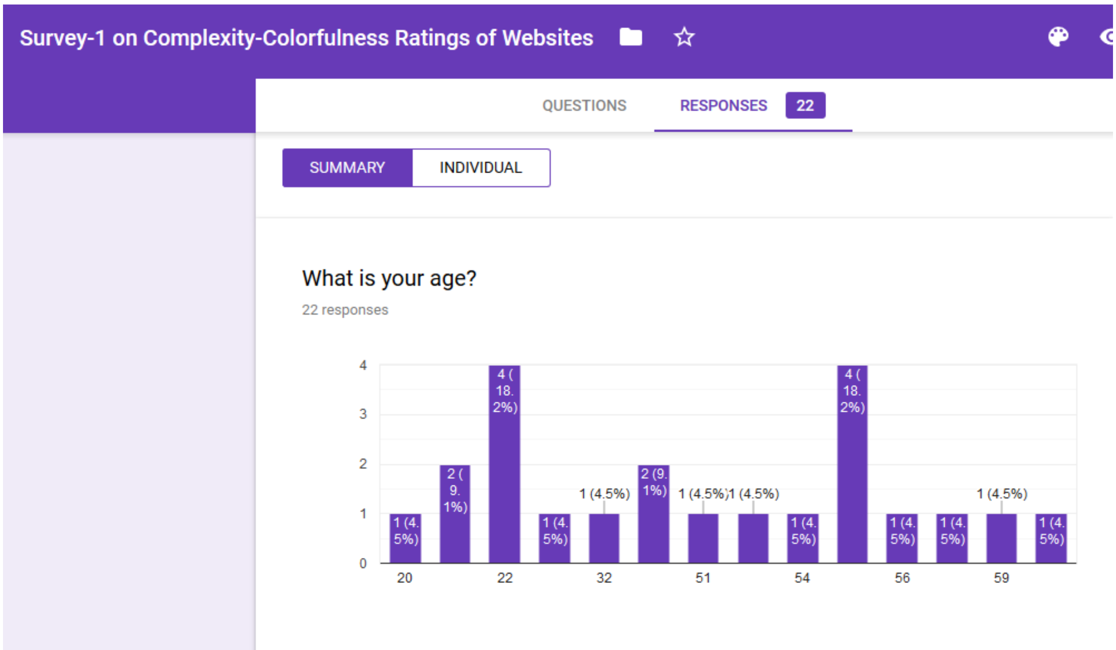
The visual complexity and colorfulness ratings were averaged for each website respectively and saved to be used later, during the development of the proposed models. The average complexity and colorfulness of a website is mentioned in Figure 7.3 and while Figure7.4 .shows the summary of ratings for complexity and colorfulness of a website.



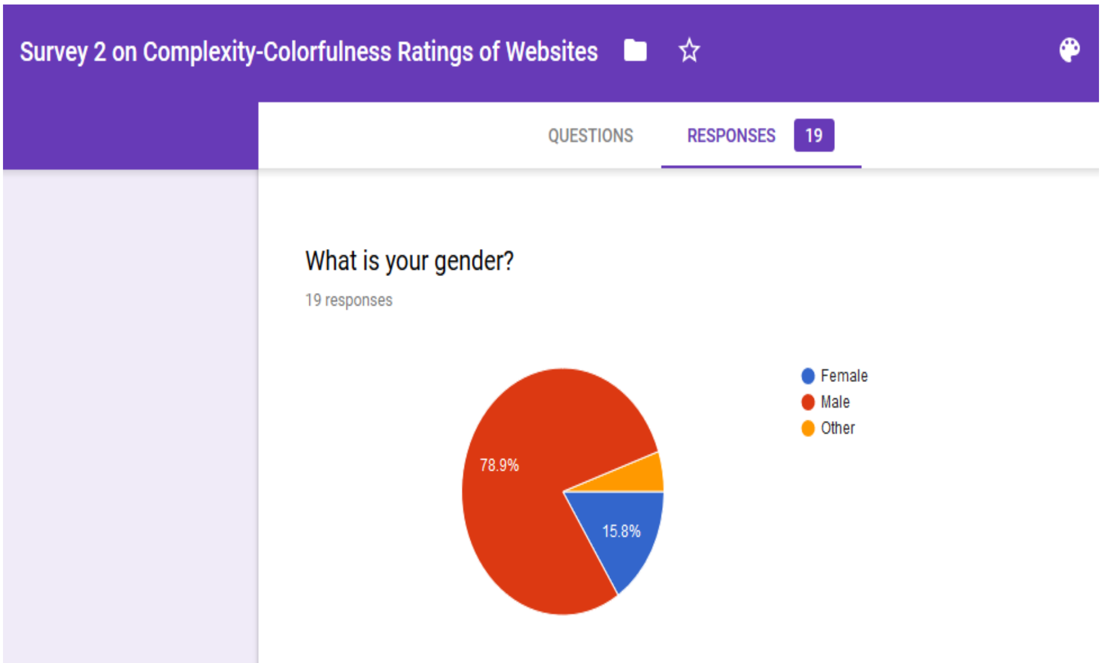

(a)


Survey 2 on Complexity-Colorfulness Ratings of Websites
QUESTIONS
RESPONSES 19
Which factors affect your judgment of the visual complexity of websites? (multiple choices allowed)
19 responses
Colors
13 (68.4%)
Distribution of composition ( regular, o...
10 (52.6%)
Symmetry
9 (47.4%)
Content (elements, objects, people)
15 (78.9%)
Contrast
8 (42.1%)
0
5
10
15

(b)


Survey-1 on Complexity-Colorfulness Ratings of Websites
QUESTIONS
RESPONSES 22
SUMMARY
INDIVIDUAL
What is your age?
22 responses
0
1
2
3
4
20
22
32
51
54
56
59

(c)




(d)

Figure 7.2: (a) Survey 4 –summary of participant ratings on factors influencing visual complexity; (b) Survey 2 – summary of participant ratings on factors influencing visual complexity; (c) Age statistics of participants in survey 1; (d) Gender statistics of participants in survey 2

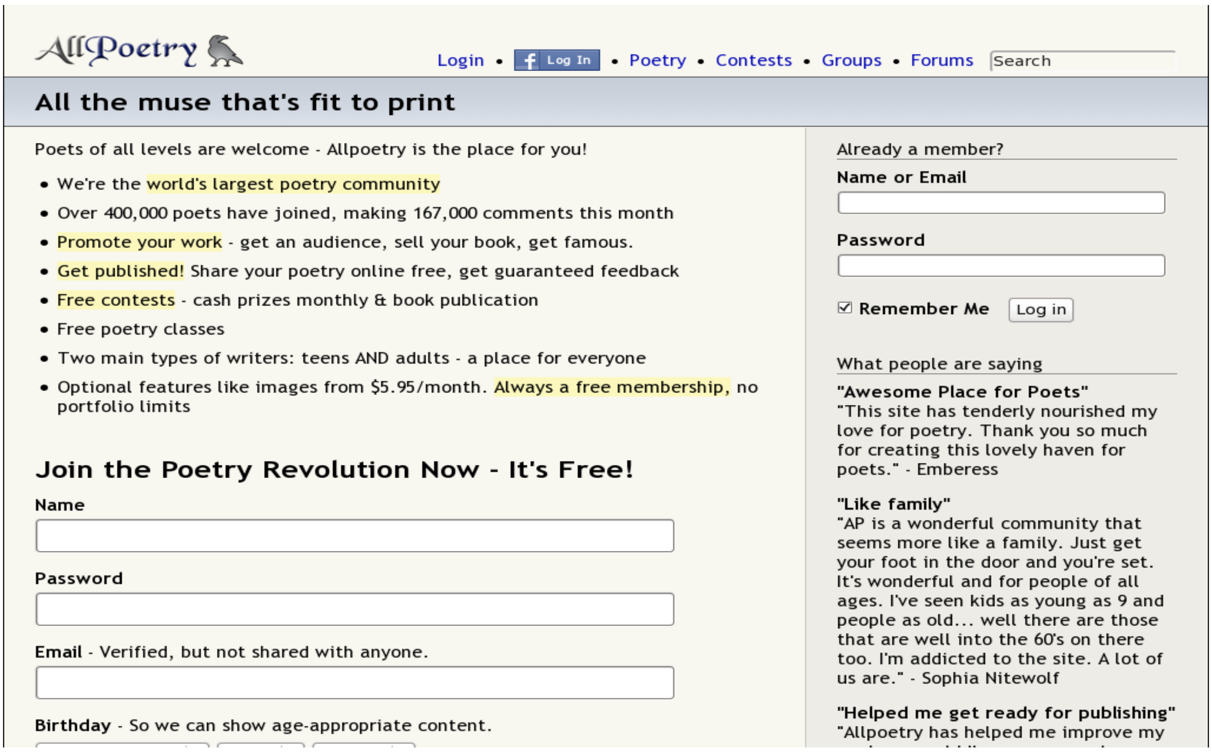
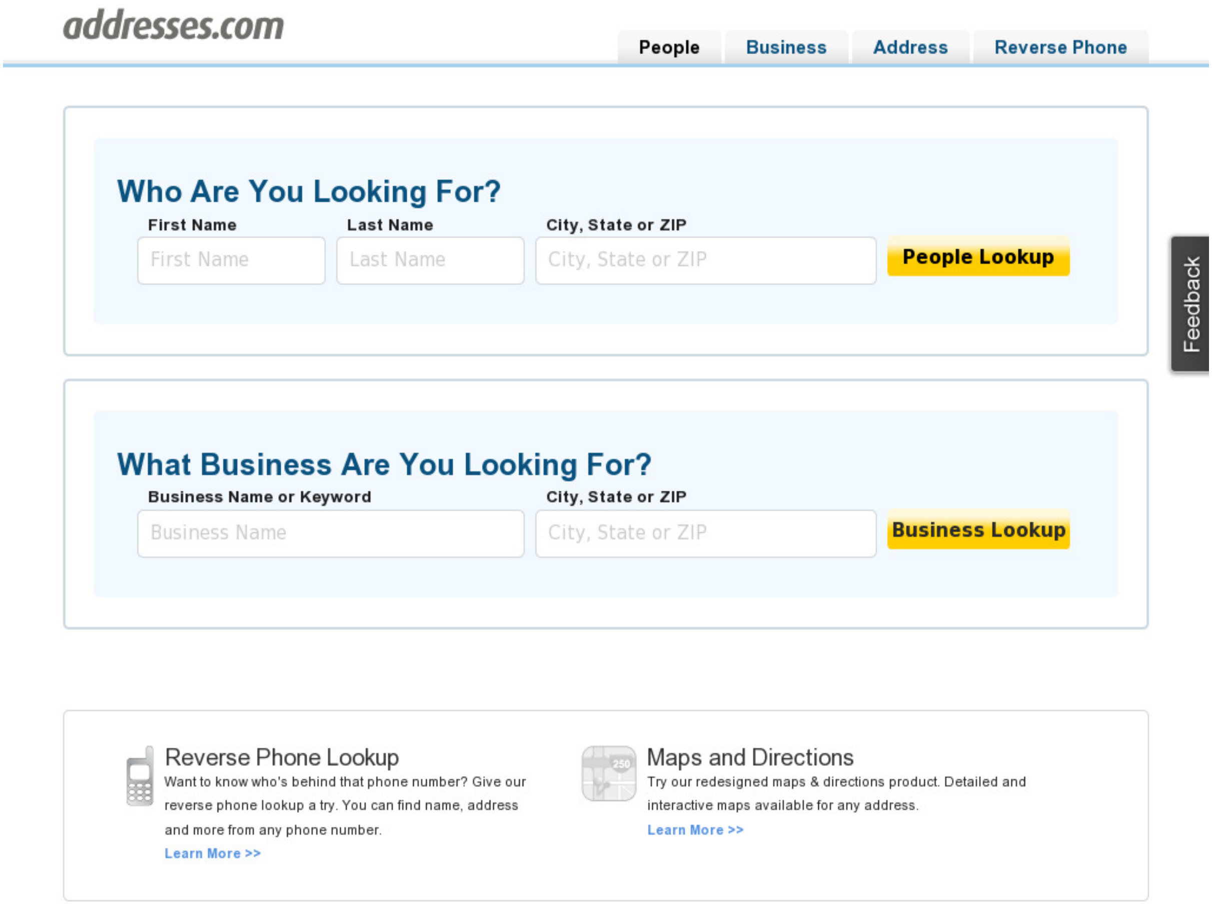
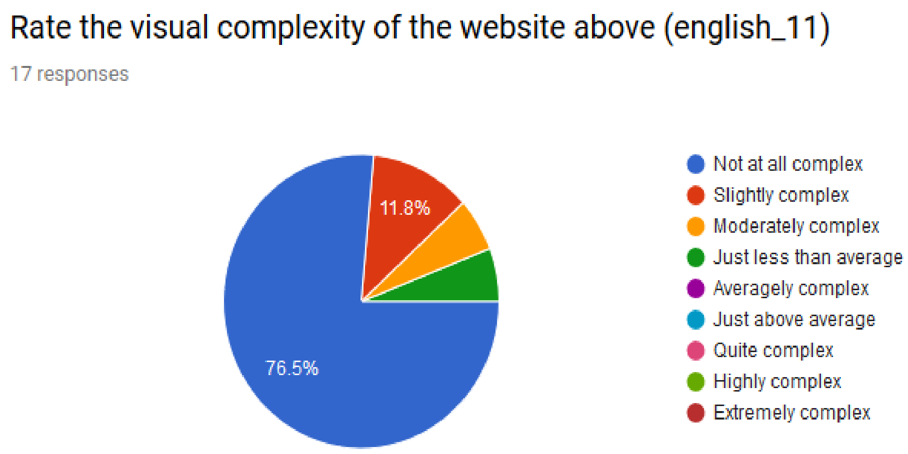


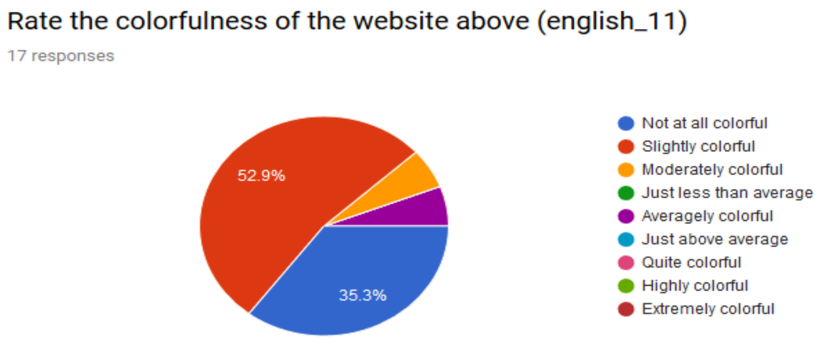

Figure 7.3: A website with mean visual complexity rating of 2.066667 and mean colorfulness rating of 2.333333


addresses.com
People
Business
Address
Reverse Phone
Who Are You Looking For?
First Name
Last Name
City, State or ZIP
People Lookup
What Business Are You Looking For?
Business Name or Keyword
City, State or ZIP
Business Lookup
Feedback
Reverse Phone Lookup
Want to know who's behind that phone number? Give our reverse phone lookup a try. You can find name, address and more from any phone number.
Learn More >>
Maps and Directions
Try our redesigned maps & directions product. Detailed and interactive maps available for any address.
Learn More >>


(a)


Rate the visual complexity of the website above (english_11)
17 responses
11.8%
76.5%
Not at all complex
Slightly complex
Moderately complex
Just less than average
Averagely complex
Just above average
Quite complex
Highly complex
Extremely complex


(b)




(c)

Figure 7.4 : (a) Original Website with ID English_11; (b) Summary of visual complexity ratings for this site; (c) Summary of colorfulness ratings for this site

## 7.2 FEATURE EXTRACTION

The image features considered have been categorized into three levels- global level, local level and saliency levels. Global image features refer to characteristics which influence the first impressions of website on users. Local image features, on the other hand, focuses upon the regional details of a website. Salient features capture the information extracted from the most important visual part of the image. The 22 features calculated for this study are summarized in Figure 7.5.

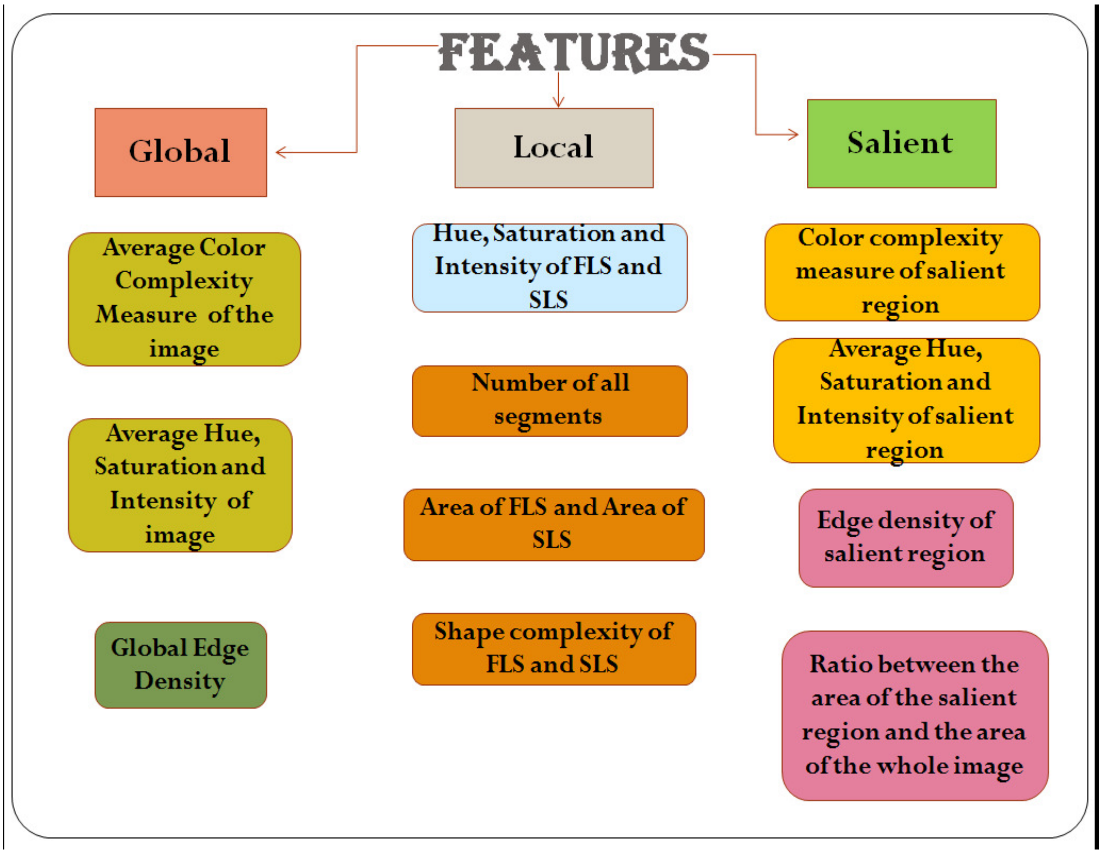



Figure 7.5: The features extracted from the website images

### 7.2.1 GLOBAL FEATURES

When viewing a product for the first time, a person first perceives the overall impression of the item before focusing upon the details and contents. The global features are described below-:

1) **Color Complexity Measure [20] :**

This feature is a measure of the color complexity of a given image. It measures the color complexity about the interesting pixel by using the absolute color deviation with Gaussian weighting about the neighbourhood of that pixel. The global CCM is the average of the CCM values of all pixels in an image. Figure 7.6. shows the basic steps of calculation.

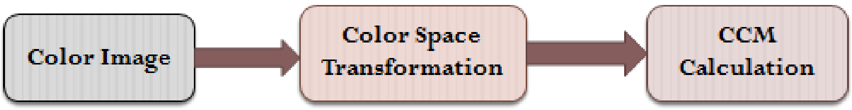



Figure 7.6 : The steps in calculation of CCM of a given pixel

The CCM of an interesting pixel is defined as:-

$$\phi(i,j) = \sum_{x,y\in\Omega} G_{\alpha}\left(1-exp\left(-\frac{E(c(x,y),c(i,j))}{\gamma}\right)\right) \qquad (6)$$

Where,

$G_{\alpha}$ denotes the Gaussian weighting function , $\gamma$ is a normalizing factor and usually taken as 14, $\Omega$ represents the 3 x 3 neighborhood, c(i,j) is the average value within a local mask $\Omega$(neighborhood) centered at (i, j) and is calculated as-

$$\bar{c}(i,j) = \frac{1}{N}\sum_{x,y\in\Omega_{(i,j)}} c(x,y) \qquad (7)$$

and E(c(x,y),c(i,j)) is the Euclidean distance between two points in a color space and is measured as:

$$E(c(i,j), c(x,y)) = \sqrt{(L_{ij} - L_{xy})^2 + (a_{ij} - a_{xy})^2 + (b_{ij} - b_{xy})^2} \qquad (8)$$

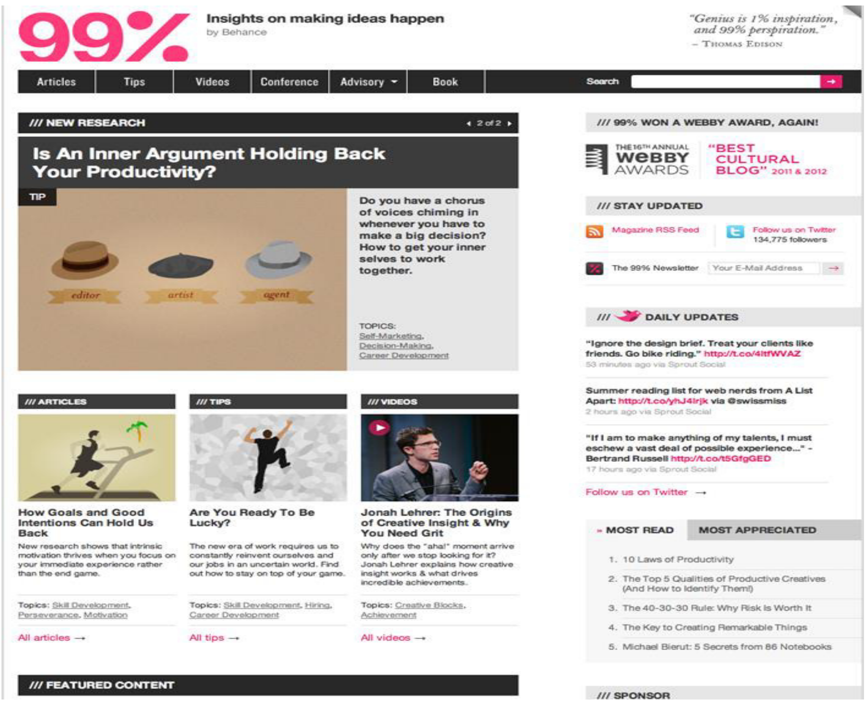
Where, L,a,b represent the L*a*b axes in the CIE Lab space.

Color difference measure is considered as it a marker of the human visual perception of colors. CIELab space has been considered as it represents all colors that are visible and discernible by human eyes.

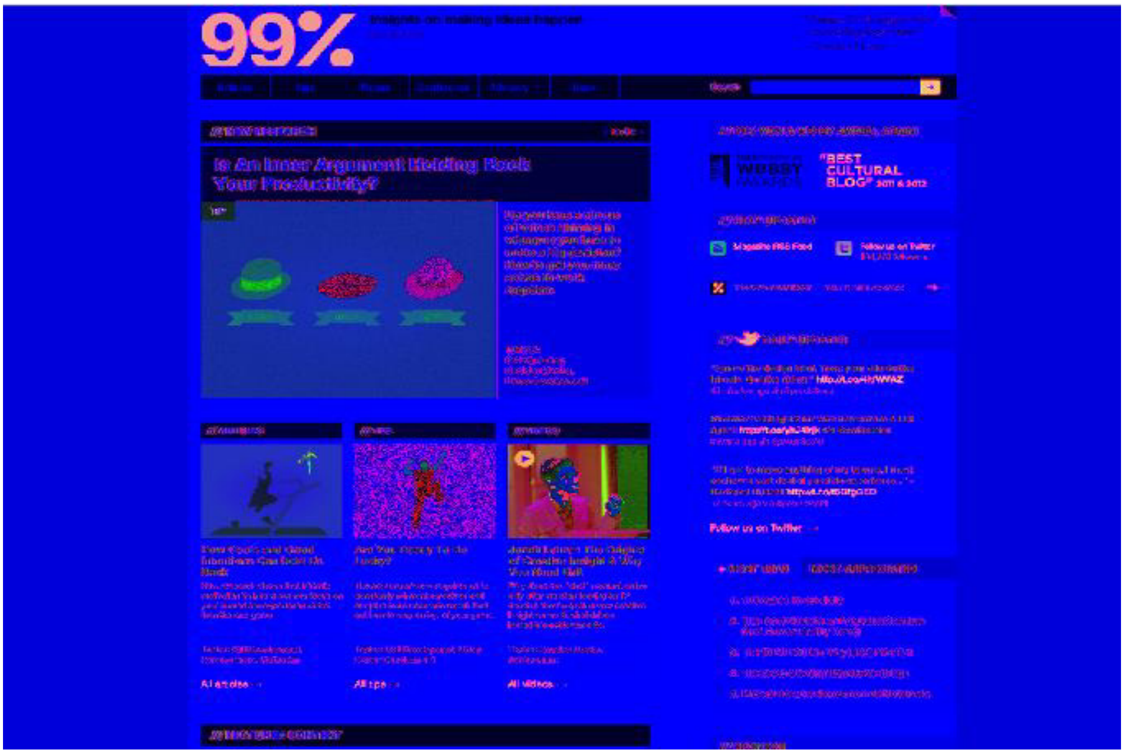
Large CCM value implies the presence of a high spatial color pattern variation in the local neighbor and small CCM value is an indicator that the particular pixel is located in a homogenous region.

### 2.) Basic Color Features:

There are numerous ways of representing color. The three dimensional RGB color space though commonly used, does not serve well in the context of color perception by human eye. On the other hand, HSI color space allows for effective manipulation of colors and resembles the mechanism of human eye perception[26]. HSI also has a greater dynamic range of saturation. Therefore, basic color features are calculated by converting the website image into HSI color space and measuring the mean hue, saturation and intensity values of the image , respectively. Figure 7.7 (a) and (b) display the website screenshots in RGB and HSI color spaces respectively.




(a)




(b)

Figure 7.7: (a) The original image in RGB color space, (b) The image in HSI color space

a.) Hue : the average of hue reflects upon the colourful aspect of an image in relation to a viewer's perception[21]. It is measure of color purity.

$$F_H = \sum_{m} \sum_{n} \frac{I_H(m,n)}{MN} \tag{9}$$

Where, M and N are the number of rows and columns of the image, respectively, and $I_H(m, n)$ is the hue value at the pixel (m, n).

b.) Saturation : this parameter is used to take into account the 'faded' effect of colors, or in a way, it helps distinguish between a strong color from a weak one.

$$F_S = \sum_{m} \sum_{n} \frac{I_S(m,n)}{MN} \tag{10}$$

Where, M and N are the number of rows and columns of the image, and $I_S(m, n)$ is the saturation value at the pixel (m, n).

c.) Intensity : This is a measure of the lightness of an image, indicating how light or dark a color is in the entire image. This quantity reflects the tone of an image.

$$F_I = \frac{\sum_m \sum_n I_I(m,n)}{MN} \tag{11}$$

Where, M and N are the number of rows and columns of the image, and $I_I(m, n)$ is the intensity value at the pixel (m, n).

**3.) Edge Density** :

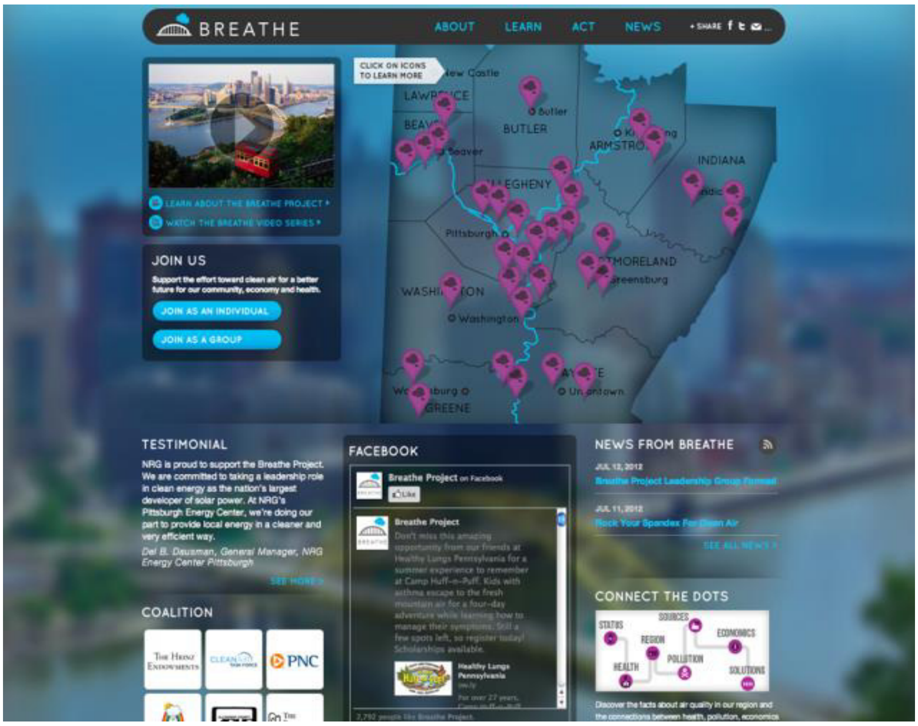
The edge density of the website image can be determined by the ratio between the pixel number of the extracted edges and that of the entire image. The number of edges in an image is an inherent marker of the number of objects in an image. A high edge density implies presence of many objects, giving rise to high perceived complexity [22].

To calculate the edge density, first, the 'Canny algorithm' is employed for edge extraction. In the edge detected image, which is the output of the mentioned algorithm, the edge pixels are marked in white while the other pixels are black. Figure 7.8 (a) and (b) depict the original website screenshot and the edge detected image, respectively.

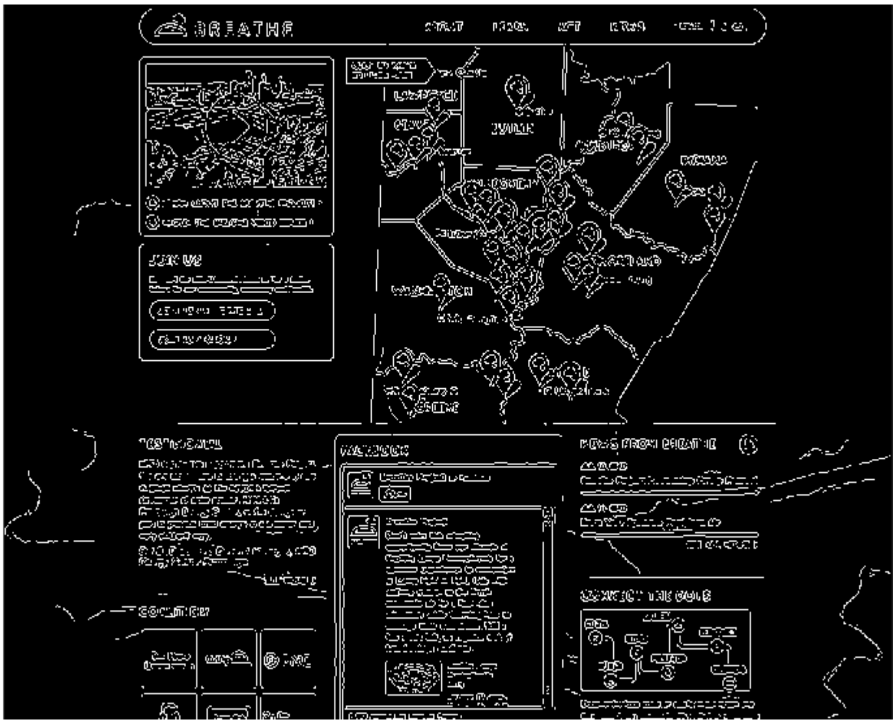
The 'nnz( )' function of MATLAB returns the number of non-zero pixels in an image. When applied to the edge detected image, it can return the number of border pixels. The total number of image pixels is the product of the number of rows and columns in the image.

Global edge density is calculated as-:

$$\textit{Edge density} = \frac{\textit{Edge Pixels}}{\textit{Total number of Image Pixels}} \tag{12}$$




(a)



(b)

Figure 7.8: (a) The original website screenshot, (b) The edges detected in the image

## 7.2.2 LOCAL FEATURES

For the extraction of the next set of features, image segmentation is done using mean-shift algorithm with the help of EDISON System [2]. The website before and after segmentation are shown in Figure 7.9 (a) and (b), respectively.

The first largest segment(FLS) and the second largest segment(SLS) are the basis for computation of remaining features.


(a)


(b)

Figure 7.9 : (a) The original website screenshot, (b) The mean-shift segmented image

**1.) Hue, saturation & lightness of the two largest segments**

The average Hue, Saturation and lightness of FLS and SLS both, are computed. Thus, these are the 6 color features are calculated based on HSI color space in this part. The calculation is similar to equations (9), (10) and (11) respectively, but instead of total number of image pixels, the number of pixels in the FLS or SLS is considered.

**2.) Number of all segments**

In the image segmentation procedure, the website image is clustered and segmented according to the bandwidths in the spatial and feature space domain. The segmented pixels form small regions which is the total number of segments.

**3.) Areas of top two largest segments**

A large segments marks a highly homogenous region which implies a gentle visual perception. Areas of FLS and SLS are extracted using the Image Region Analyzer App of MATLAB. The SLS of a website is highlighted as shown in Figure 7.10 and Figure 7.11 shows how the App can be utilized for generating a binary mark which can help extract a required segment.

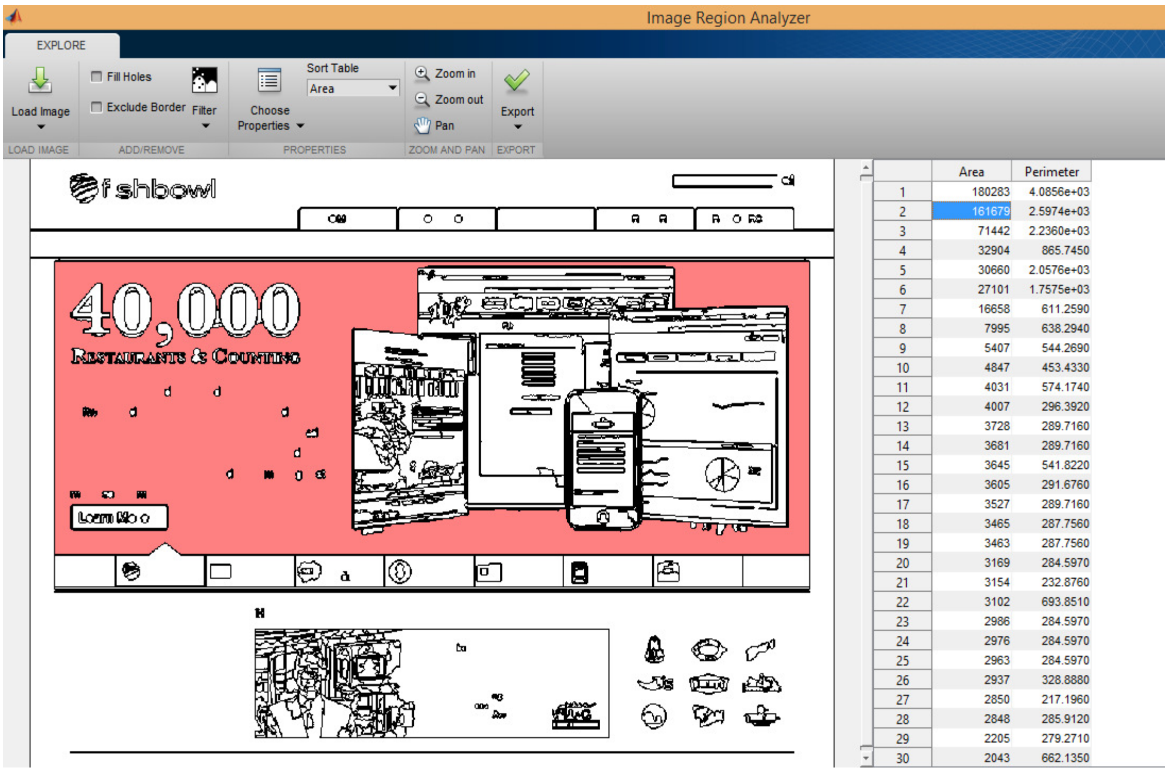



Figure 7.10: Image Region Analyzer; SLS has been highlighted


Figure 7.11 : Binary mask to obtain the SLS of a website

**4.) Shape complexity of the first two largest segments**

The contours of the regions with different shapes elicit a different visual perception [21]. Presence of coherent, structured objects reduces stress on cognitive mechanisms of humans [28], rendering perception less taxing. If a segment is of definite shape such as a rectangle or circle, it would allow easy grouping and perception of points in that area. On the other hand, irregular shape of the segment would not catch the user's eye as it will be more difficult to make out its form and the features of this area would not be able create a good visual impact. The shape complexity of each segment[21] is calculated as follows:

$$f_{sc} = \frac{P^3}{4\pi R} \tag{13}$$

where, P is the perimeter of a segment (FLS or SLS), and A is the area of that segment.

### 7.2.3 SALIENT FEATURES

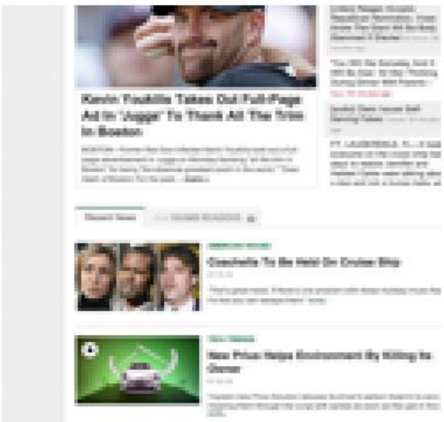
Salient region refers to the most visually important region in an image. When the observers see an image, they are easily attracted to this region at the first glance. This sections focuses upon the extraction of features from this region.

An open-source software[3] , coded in MATLAB employs Context-Aware Saliency Method[4] for saliency map generation.

The visually most important region is extracted from the saliency map of the image by employing the automatic cropping algorithm described in [23]. The aspect ratio employed in this study is r= 16/16 (a square). Aspect ratio is the ratio of width to height of the rectangle to be used for cropping. The area is cropped such that it preserves atleast 65% of the total energy of the saliency map. Figure 7.12 (a), (b) and (c) shows the original website, its saliency map and the most salient area of the website.

(a)

(b)



(c)

Figure 7.12: (a) The original website screenshot, (b) The saliency map of the website, (c) The salient region of the image

**1.) Color Features**

The average CCM and the basic colors of the salient area are extracted.

a) Color complexity measure- of the salient region is calculated.
   The equation employed is similar to the equation (6), but CCM values of salient region pixels alone are averaged.

b) Basic Colors – The HSI space is employed to represent numerous colors perceived by human eye. The Hue, Saturation and Intensity values of the salient region is computed. The calculations are similar to the equations eq(9), eq(10) and eq(11), but instead of total pixels of an image, the values are averaged for salient region pixels only.

**2.) Content Features** – To the capture the content features of the salient area, two features have been considered-

a) the edge density of the salient area-
   It is computed by running the 'Canny algorithm' on the salient region, followed by utilization on nnz( ) function on the edge detected image and calculating the edge density in a manner similar to equation (12) .

b) the area ratio of the salient region to the total image area is calculated –
   It is found by multiplying the height and width of the cropping rectangle which is found by employing the algorithm in [23].

$$\boldsymbol{Area\ Ratio = \frac{Area\ of\ Salient\ Region}{Total\ Area\ of\ the\ Image}} \qquad (14)$$

## 7.3 Development of Regression Models

Multiple Linear Regression has been employed for development of the models for complexity and colorfulness. The data is collected A backward elimination approach has been taken for the same. For each model, all features are taken for the regression model K-fold cross validation is employed for model evaluation and validation due to the availability of limited test data. Mean Squared Error of predictions, presented as a result of K-fold validation is a marker of model accuracy. A model with lesser value of MSE is considered more accurate. The features which improve model accuracy upon removal from the model are considered poor predictors and are excluded. On the other hand, if model accuracy drops significantly on removal of a feature, it implies that the feature influences the particular parameter (complexity/colorfulness) and is retained.

It is essential to rigorously test the model's performance to ensure accuracy. One way is to ensure that the model equation employed will perform well is by 'building' it on a different subset of training data, and predicting on the remaining data. This is done by splitting the data into 'k' mutually exclusive random sample portions. In each iteration, one of the portions is employed as test data, while the remaining (K-1 portion) data is used to build the model and the mean squared error of the predictions is calculated. This is done for each k random samples.MSE tends to punish outliers severely ,that is, it increases significantly even for a single outlier.

In this study, K- fold cross validation has been performed with three different K values- K=5,10,40.

### 7.3.1 Visual Complexity Model

The final set of predictor variables is as follows:-

- Global Average CCM
- Global Average Hue
- Global Average Intensity
- Area of FLS
- Shape Complexity of FLS
- Number of segments in the image
- Average Hue of FLS
- Average Intensity of FLS
- Average Intensity of SLS
- Average Intensity of Salient Region
- Area ratio (Area of salient region by Total area)
- Edge Density of salient region
- Salient region average CCM

Table 2 demonstrates the summary of the complexity model generated for objective estimation of complexity. The 'Estimate' column represents the coefficient values for each of the predictors of complexity while 'Residual Standard Error' gives the difference in the values obtained from the regression model and the actual values of complexity.

```
Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          -8.871e+01  7.210e+01  -1.230   0.2296
Mean_CCM             -2.070e+01  4.461e+01  -0.464   0.6464
Avg_hue_global        7.045e-01  1.825e+00   0.386   0.7026
Avg_intensity_global  3.951e+00  3.757e+00   1.052   0.3027
Area_FLS              8.422e-07  1.241e-06   0.679   0.5034
Shape_Complexity_FLS -6.882e-06  4.734e-06  -1.454   0.1580
numRegions            4.201e-04  3.190e-04   1.317   0.1993
F_mean_hue           -1.288e+00  1.095e+00  -1.176   0.2502
F_mean_I             -1.458e+00  1.821e+00  -0.801   0.4307
S_mean_I             -3.129e-01  6.916e-01  -0.452   0.6547
Salient_Intensity    -3.220e+00  2.330e+00  -1.382   0.1789
Area_ratio            3.869e+00  1.514e+00   2.555   0.0168 *
Salient_edgeDensity  -1.171e+00  6.519e+00  -0.180   0.8588
Salient_CCM           5.142e+01  6.198e+01   0.830   0.4143
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8075 on 26 degrees of freedom
Multiple R-squared:  0.4735,    Adjusted R-squared:  0.2102
F-statistic: 1.798 on 13 and 26 DF,  p-value: 0.09835
```

Table 7.1: The final regression model for visual complexity

Figure 15 displays the original vs predicted complexity values for each of the folds in a 5-fold cross validation model.

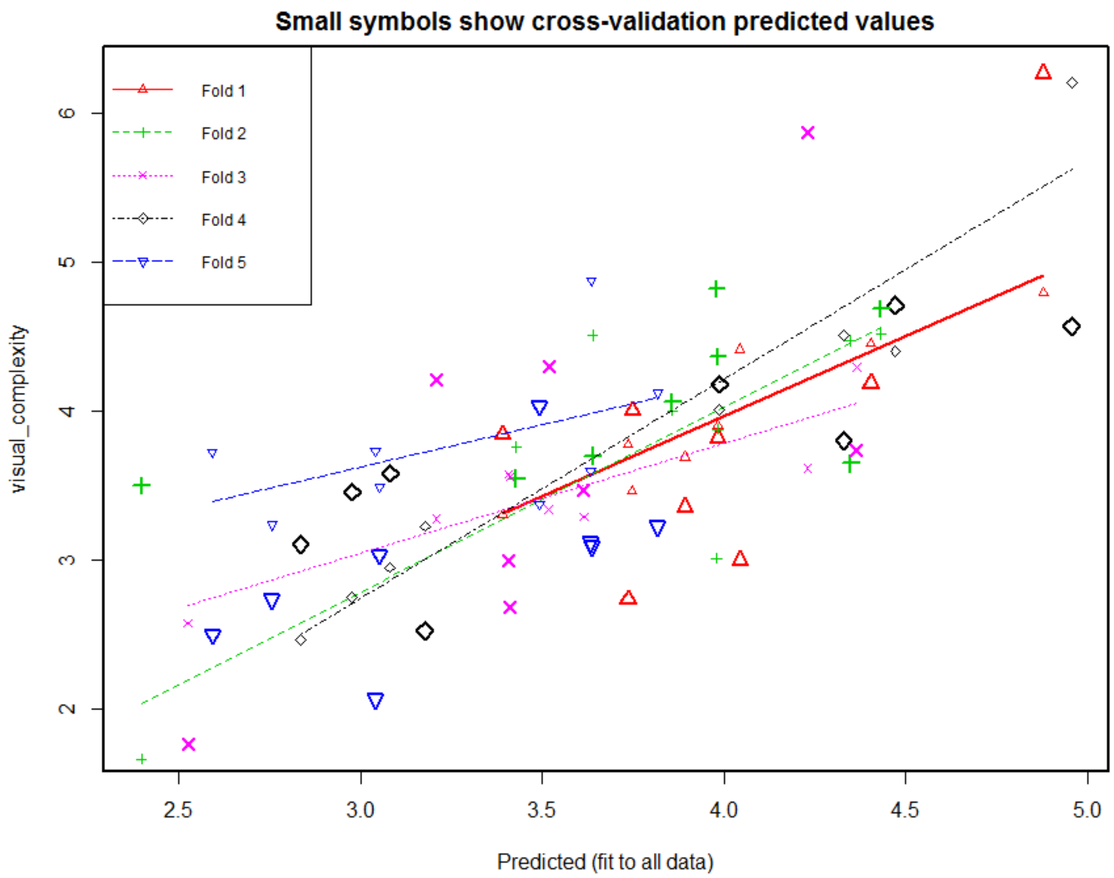



Figure 7.13: 5-fold cross validation for visual complexity model

Table 3 displays the accuracy of the visual complexity model when validated with k-fold cross validation models with k values of 5,10 and 40.

| k-value | Corresponding overall MSE value |
|---|---|
| 5-fold | 0.942 |
| 10-fold | 1.17 |
| 40-fold | 1.09 |

Table 7.2 : Summary of MSE values generated from various k-fold validations for visual complexity model

### 7.3.2 Colorfulness Model

MLR is employed for model generation.

The final set of predictor variables is as follows:-

- Global Average Hue
- Global Average Intensity
- Number of segments in the image
- Average Hue of FLS
- Average Saturation of FLS
- Average Intensity of FLS
- Average Hue of Salient Region
- Average Saturation of Salient Region
- Average Intensity of Salient Region

Table 4 demonstrates the summary of the model generated for objective estimation of color. The ‘Estimate’ column represents the coefficient values for each of the predictors of colorfulness while ‘Residual Standard Error’ gives the difference in the values obtained from the regression model and the actual values of colorfulness.

```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)          3.277233   0.554601    5.91  1.8e-06 ***
Avg_hue_global       2.587295   1.705344    1.52   0.1397    
Avg_intensity_global -0.670972  2.388216   -0.28   0.7807    
numRegions           0.000521   0.000224    2.32   0.0271 *  
F_mean_hue           3.400715   0.946676    3.59   0.0012 ** 
F_mean_S            -4.898546   2.004701   -2.44   0.0206 *  
F_mean_I             1.202844   0.979589    1.23   0.2290    
Salient_Hue         -4.788773   1.710602   -2.80   0.0089 ** 
Salient_Saturation   7.523421   2.411432    3.12   0.0040 ** 
Salient_Intensity   -0.875107   2.039979   -0.43   0.6710    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.573 on 30 degrees of freedom
Multiple R-squared:  0.775,     Adjusted R-squared:  0.707 
F-statistic: 11.5 on 9 and 30 DF,  p-value: 1.53e-07
```

Table 7.3: The final regression model for colorfulness

Figure 16 displays the original vs predicted values for each of the folds in a 5-fold cross validation model. Figure 17 displays a sample output presented during k-fold validation.

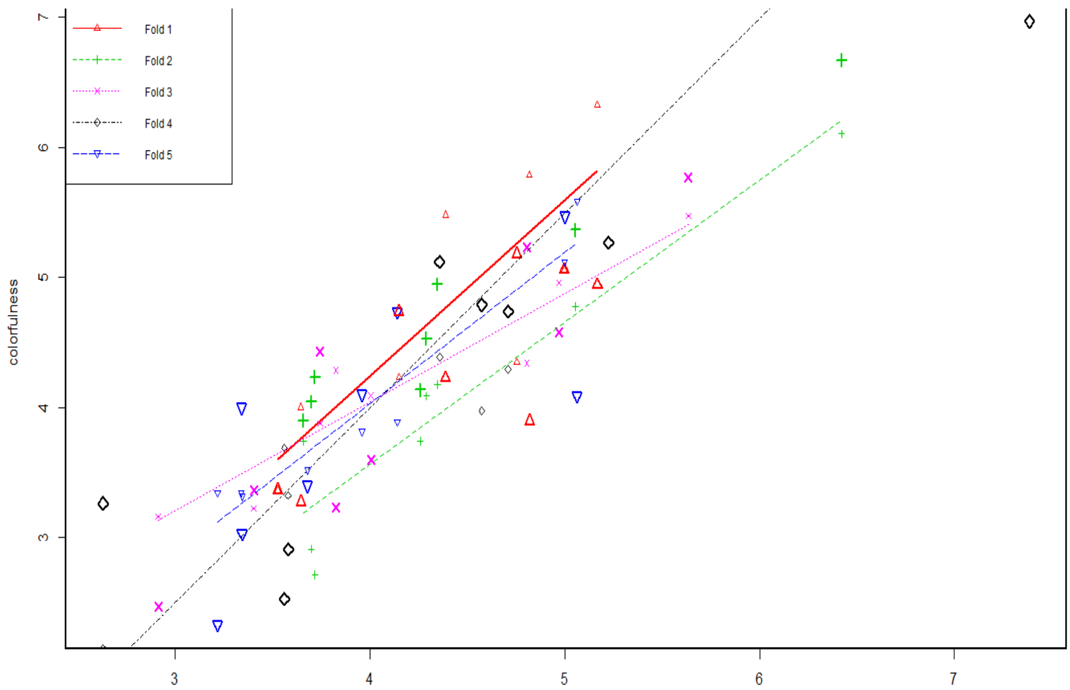



Figure 7.15 : 5-fold cross validation for colorfulness model

```
fold 1
Observations in test set: 8
                  2      3     9        11     15     19     34      36
Predicted     3.52533  4.82 4.149 4.99708  5.17 4.753  4.39   3.643
cvpred        3.36547  5.78 4.230 5.05314  6.33 4.350  5.48   4.002
colorfulness  3.36842  3.89 4.737 5.05882  4.94 5.176  4.23   3.273
CV residual   0.00296 -1.89 0.507 0.00569 -1.38 0.827 -1.25  -0.729

Sum of squares = 8.52    Mean square = 1.07    n = 8

fold 2
Observations in test set: 8
                 16    17    20     26     32     33    38     39
Predicted     4.290 3.72 3.657 6.420 4.344 4.257 3.70 5.054
cvpred        4.088 2.71 3.741 6.103 4.175 3.740 2.91 4.775
colorfulness  4.529 4.24 3.900 6.667 4.947 4.136 4.05 5.364
CV residual   0.441 1.52 0.159 0.564 0.772 0.397 1.13 0.588

Sum of squares = 5.24    Mean square = 0.65    n = 8

fold 3
Observations in test set: 8
                   5     12    23      24     27     30      31     40
Predicted      2.912  3.82 3.744  4.004 5.635 3.403  4.970 4.804
cvpred         3.161  4.29 3.879  4.091 5.471 3.224  4.960 4.340
colorfulness   2.474  3.24 4.433  3.600 5.767 3.368  4.579 5.227
CV residual   -0.688 -1.05 0.554 -0.491 0.296 0.145 -0.381 0.887

Sum of squares = 3.17    Mean square = 0.4    n = 8

fold 4
Observations in test set: 8
                    4     6    10     14    18     22     29      35
Predicted     4.574 2.63 4.709  3.56 4.36  5.22  7.39  3.580
cvpred        3.976 2.15 4.294  3.69 4.39  7.11  9.05  3.326
colorfulness  4.789 3.26 4.737  2.53 5.12  5.27  6.97  2.909
CV residual   0.814 1.11 0.443 -1.16 0.73 -1.85 -2.09 -0.417

Sum of squares = 11.9    Mean square = 1.49    n = 8
```

Figure 7.15: Sample folds of 5-cross validation for colorfulness

Table 5 displays the accuracy of the colorfulness model when validated with k-fold cross validation models with k values of 5,10 and 40.

| k-value | Corresponding overall MSE value |
|---|---|
| 5-fold | 0.839 |
| 10-fold | 0.515 |
| 40-fold | 0.474 |

Table 7.4 : Summary of MSE values generated from various k-fold validations for colorfulness model

# 8. CHALLENGES AND CONSTRAINTS

- Perceived visual complexity and colorfulness may differ according to age, gender and other demographic factors. This study does not take into account these factors and requires more work for a specific user segment.

- Only 40 websites have been used for this study. It does not guarantee that it will represent a random website drawn from the internet as accurately.

- Subjective ratings of the two parameters, collected from a very small group of participants (around 86 participants) do not necessarily cover ratings of a heterogeneous sample.

- Factors influencing visual complexity and colorfulness are numerous and this study does not cover all possible aspects of such factors.

# 9. GANTT CHART

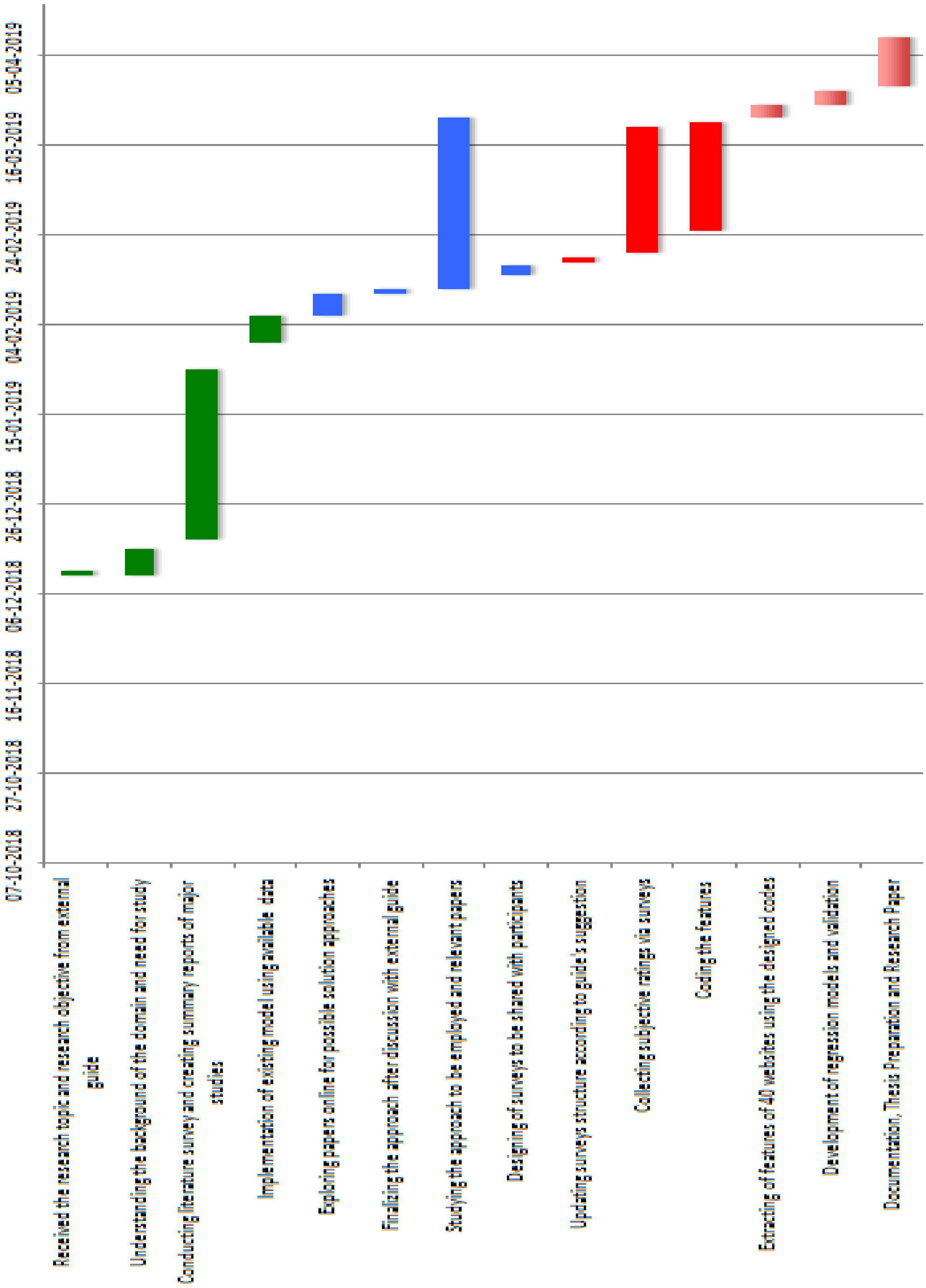



Figure 9.1: Gantt Chart

# 10. CONCLUSION

## 10.1 Summary

Visual complexity is an essential marker of human beings' ability to comprehend and perceive visual stimuli. Colorfulness attracts attention and at the same time, develops feelings of trustworthiness as well as enhances purchase intentions. Hence, it becomes essential to estimate these parameters to help e-commerce burgeon. In this paper, we first conduct surveys to collect the subjective ratings of visual complexity and colorfulness of a set of 40 websites. The age and gender of the participants are also recorded along with their opinion on factors which influence visual complexity of websites. It was found that three factors- Distribution of composition, colors and content influence the complexity most. Based on these three factors, 22 features were extracted for each of the 40 websites.

Based on these 22 global, local and salient level features, we have developed regression models for complexity and colorfulness of websites by identifying the particular features which influence each of the factors independently. These models have been validated using k-fold cross validation technique. The results of the proposed models indicate that visual complexity values can be predicted with a mean squared error of 0.942 (5-fold cross-validation ) while colorfulness can be predicted with a mean squared error of 0.839 (5-fold cross validation).

## 10.2 How this project supports the present solution

Presently the models for visual complexity and colorfulness for website appeal have not taken into account the mechanism of human visual perception. This project focuses on utilizing image features which can objectively estimate the two parameters as per human perception system. This study has enabled identification of features which influence visual complexity and colorfulness of websites, respectively, aiding the development regression models for the same. These models have been validated using k-fold cross validation system.

## 10.3 Future implementations

As human perception is vital in both the fields of psychology as well as computer science, there is a constant effort in studying the factors that influence it. The factors influencing visual complexity are not limited to the ones considered in this study. Presence of too much text makes the website uninteresting while too many images might make content unclear. It is essential to find an appropriate balance between these factors while finding a way to structure them in an appealing way. The font and style of text too may play a role in moderating the complexity of a website.

The contrasts of color employed may also impact the overall colorfulnness of an image. A black and white contract may not seem colorfuless while a combination of red and orange may enhance visual appeal.

Too colourful or very simple website might reduce appeal instead of enhancing it. It is important to find the estimated range of colorfulness and complexity at which a user finds a website appealing. What seems appealing to a teenager might not seem appealing to a person in late 60s. Preference of people of different gender might differ and their might be an impact of geographical and cultural factors on the preference of complexity, colorfulness and overall appeal of a website on a person.

In the future, I will focus on enlarging my image database as well as gather responses from a more heterogeneous group of participants to increase the robustness of the proposed method. I also aim at exploring other factors influencing complexity and colorfulness while also considering demographics in my attempt. I also hope to extend my efforts in the development of a model for the estimation of visual appeal of websites for target groups.

## REFERENCES


[1] Tractinsky, N., Cokhavi, A., Kirschenbaum, M., and Sharfi, T. Evaluating the Consistency of Immediate Aesthetic Perceptions of Web Pages. Int. Journal of Human-Computer Studies 64 (2006).

[2] http://rci.rutgers.edu/~meer/RIUL/research/code/EDISON/index.html

[3]https://cgm.technion.ac.il/Computer-GraphicsMultimedia/Software/Saliency/Saliency.html#Software

[4] S. Goferman, L. Zelnikmanor, A. Tal, Context-aware saliency detection, IEEE Trans.Pattern Anal. Mach. Intell. 34 (10) (2012) 1915–1926.

[5] N. Bonnardel, A. Piolat, and L. Le Bigot, The impact of color on website appeal and users' cognitive processes, Displays, vol. 32, no. 2, pp. 69-80, 2011.

[6] Lindgaard, G., Fernandes, G., Dudek, C., and Brown, J. Attention Web Designers: You Have 50 Milliseconds to Make a Good First Impression! Behaviour & Information Technology 25, 2 (2006), 115–126.

[7] Kim, J., Lee, J., Choi, D., 2003. Designing emotionally evocative homepages: an empirical study of the quantitative relations between design factors and emotional dimensions. International Journal of Human–Computer Studies 59 (6), 899–940.

[8] Simon Harper, Eleni Michailidou, and Robert Stevens. 2009. Toward a definition of visual complexity as an implicit measure of cognitive load. ACM Trans. Appl. Percept.6,2, Article 10 (March 2009), 18 pages. DOI=http://dx.doi.org/10.1145/1498700.1498704

[9] Cyr, D., Head, M., and Larios, H. Colour Appeal in Website Design Within and Across Cultures: A Multi-method Evaluation. Int. Journal of Human-Computer Studies 68, 1-2 (2010), 1–21.

[10] Cyr, D. Modeling Website Design across Cultures: Relationships to Trust, Satisfaction and E-loyalty. Journal of Management Information Systems 24, 4 (2008), 47–72.

[11] Hall, R. H., and Hanna, P. The Impact of Web Page Text-background Colour Combinations on Readability, Retention, Aesthetics and Behavioural Intention. Behaviour & Information Technology 23, 3 (2004), 183–195.

[12] Guo, Xiaoying & Qian, Yuhua & Li, Liang & Asano, Akira. (2018). Assessment model for perceived visual complexity of painting images. Knowledge-Based Systems. 10.1016/j.knosys.2018.06.006.

[13] Lindgaard, G., Dudek, C., Sen, D., Sumegi, L., and Noonan, P. An Exploration of Relations Between Visual Appeal, Trustworthiness and Perceived Usability of Homepages. ACM Transactions on Computer-Human Interaction 18, 1 (2011).

[14] Zheng, X., Chakraborty, I., Lin, J., and Rauschenberger, R. Correlating Low-level Image Statistics with Users' Rapid Aesthetic and Affective Judgments of Web Pages. In Proc. CHI (2009).

[15] K. Reinecke, T. Yeh, L. Miratrix, R. Mardiko, Y. Zhao, J. Liu, and K.Z. Gajos, Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, 2013, pp. 2049-2058.

[16] Moss, G., and Gunn, R. Gender Differences in Website Production and Preference Aesthetics: Preliminary Implications for ICT Education and Beyond. Behaviour & Information Technology 28, 5 (2000), 447–460

[17]Hsiu-Feng, W. Picture Perfect: Girls' and Boys' Preferences Towards Visual Complexity in Children's Websites. Computers in Human Behavior(2013).

[18]Bauerly, M., and Liu, Y. Effects of Symmetry and Number of Compositional Elements on Interface and Design Aesthetics. Int. Journal of Human-Computer Interaction 3 (2008), 275–287.

[19]Katharina Reinecke and Krzysztof Z. Gajos. Quantifying visual preferences around the world. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14, pages 11-20, New York, NY, USA, 2014. ACM.

[20] K.J. Yoon, I.S. Kweon, Color image segmentation considering of human sensitivity for color pattern variations, Proc. SPIE, (2001), pp. 269–278.

[21] C. Li, T. Chen, Aesthetic visual quality assessment of paintings, IEEE J. Sel. Top Signal Process. 3 (2) (2009) 236–253.

[22] M. Cardaci, V.D. Gesù, M. Petrou, M.E. Tabacchi, A fuzzy approach to the evaluation of image compexity, Fuzzy Sets Syst. (2009) 1474–1484.

[23] J. Chen, G. Bai, S. Liang, Z. Li, Automatic image cropping : a computational complexity study, 2016 Comput. Vision Pattern Recognit. (2016) 507–515.

[24] https://cran.r-project.org/bin/windows/base/old/3.4.3/

[25] https://in.mathworks.com/downloads/

[26] http://ijarcet.org/wp-content/uploads/IJARCET-VOL-3-ISSUE-3-618-622.pdf

[27] https://in.mathworks.com/help/images/ref/imageregionanalyzer-app.html

[28] https://www.pnas.org/content/99/23/15164

[29] Ha, J., Haralick, R.M., & Phillips, I.T. (1995). Recursive X-Y cut using bounding boxes of connected components. ICDAR.

[30] Yendrikhovskij, S. N., Blommaert, F. J. J., and De Ridder, H. Optimizing Color Reproduction of Natural Images. In Proc. Color Imaging Conference: Color Science, Systems, and Applications (1998), 140–145.

[31] Hasler, D., and Suesstrunk, S. Measuring Colourfulness in Natural Images. In Proc. SPIE/IS&T Human Vision and Electronic Imaging, vol. 5007 (2003), 87–95.

# APPENDEX 1:

# CODE SNIPPETS

```
close all;
clc;

image = imread('D:\ISI\supplemental\My_Stimuli\webby\16.png');
rgb= mat2gray(image);
HSI= rgb2hsi(rgb); % convert rgb values to [0,1]
figure,imshow(HSI);
img     = imread('D:\ISI\supplemental\My_Stimuli\webby\segmented16_segm.png');
imgbdr     = imread('D:\ISI\supplemental\My_Stimuli\webby\segmented16_bndy.png');
display(img);
figure, imshow(img);
Gray= rgb2gray(imgbdr);%boundary of segmented image
figure, imshow(Gray);
BW = imbinarize(Gray); %convert image to binary

%Get area and perimeter values from the propstruct
FLS_area = propsStruct(1).Area;
FLS_peri = propsStruct(1).Perimeter;
SLS_area = propsStruct(2).Area;
SLS_peri = propsStruct(2).Perimeter;

%Calculate the shape complexity of the two segments
Shape_complexity_FLS= (FLS_peri * FLS_peri * FLS_peri) / ( 4 * pi * FLS_area);
Shape_complexity_SLS= (SLS_peri * SLS_peri * SLS_peri) / ( 4 * pi * SLS_area);

display(FLS_area);
display(FLS_peri);
display(Shape_complexity_FLS);
display(SLS_area);
display(SLS_peri);
```

```
function [mean_hue, mean_saturation, mean_intensity]= calculationMeanValues(mask,hsi_image)
sum_hue=0;
sum_saturation=0;
sum_intensity=0;
count=0;

for i=1:size(hsi_image,1)
    for j=1:size(hsi_image,2)
        if (mask(i,j)==0)
            count=count+1;  %number of pixels in the given region
            sum_hue= sum_hue + hsi_image(i,j,1);
            sum_saturation=sum_saturation + hsi_image(i,j,2);
            sum_intensity=sum_intensity + hsi_image(i,j,3);
        else
            continue;
        end
    end
end
%display(sum_hue);
%display(count);
 mean_hue= sum_hue / count ;
 mean_saturation= sum_saturation / count;
 mean_intensity= sum_intensity/ count;
end

function [RGB_Image] = convertBinImage2RGB(BinImage)
  RGB_Image = uint8( BinImage(:,:,[1 1 1]) * 255 );
end

```

```
% Read the desired image file.
Image=imread('D:\ISI\supplemental\My_Stimuli\webby\16.png');

% Display the original image.
figure,imshow(Image);
title(' Original Image: ');

%convert from RGB to Grayscale
Grayimg = rgb2gray(Image);
figure, imshow(Grayimg)
title(' Grayscale Image: ');
%Canny edge detection algorithm is employed
edgeImg = edge(Grayimg,'Canny');
figure,imshow(edgeImg);
title('Edges Detected in the image');

%calculating the number of edge pixels
NumEdgePixels= nnz(edgeImg);
display(NumEdgePixels);

x= size(edgeImg,1);
y=size(edgeImg,2);
display(x);
display(y);
imgSize= x*y;
%display(imgSize);

%calculating edge density
EdgeDensity= NumEdgePixels / imgSize;
display(EdgeDensity);
```

```
function [i,j,w,h] = FixedAspRatioRectangle(G,T,r,G_plus,G_plus_c)
i = 0;
j = 0;
m= size(G,1);
n= size(G,2);
w = Inf;
h = Inf;
i1 =1;
i2=1;
T1 = T* G_plus(m,n);
Smin = -1 ;
while( i2 <= m && i1 < m)
h0 = i2 - i1 + 1;
w0 = ceil(h0 * r) ;
if (w0 > n)
    i1 = i1 + 1 ;
else
    if (i1==1)
        a = G_plus_c(i2,:) - 0 ;
    else
           a = G_plus_c(i2,:) - G_plus_c(i1 - 1,:) ;
    end
    %display(a);
    [j1,S0] = maxSubarrayFL(a,w0,T1) ;
    if (j1 > 0)
        if ((w0 * h0) < (w * h)|| (((w0 * h0) == (w * h)) && S0 > Smin))
            i = i1;
            j = j1;
            w = w0;
            h = h0 ;
```

```
            Smin = S0;
        end
        i1 = i1 + 1;
    else
        i2 = i2 + 1 ;
    end
end
end
end

function [j1,Smax] = maxSubarrayFL(a,w,T1)
 j1 = 0;
 Smax = -1;
 n= length(a);
 a_plus= zeros(size(a));
 for p= 1:n
     for q= 1: p
         a_plus(p)=a_plus(p) + a(q);
     end
 end
 if (T1 > 0 && w > 0)
     for st = 1 : (n - w + 1)
         if st==1
             S0 = a_plus(st + w - 1) - 0 ;
         else
           S0 = a_plus(st + w - 1) - a_plus(st - 1) ;
         end
         if (S0 >= T1 && S0 > Smax)
             j1 = st;
             Smax = S0;
```