CLASSIFICATION OF IMAGES BASED ON MEMORABLE SCORE USING VISUAL MEMORY SCHEMA AND COMPUTER VISION TECHNIQUES

PHASE I REPORT

Submitted by

HARSHA PRIYA G

in partial fulfillment of the requirements for the degree of

MASTER OF ENGINEERING IN

COMPUTER SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

ANNA UNIVERSITY, CHENNAI

NOVEMBER 2020

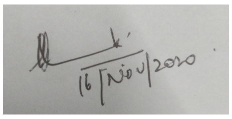
ANNA UNIVERSITY, CHENNAI

BONAFIDE CERTIFICATE

Certified that this Report titled **"CLASSIFICATION OF IMAGES BASED ON MEMORABLE SCORE USING VISUAL MEMORY SCHEMA AND COMPUTER VISION TECHNIQUES"** is the bonafide work of **HARSHA PRIYA G (2019207029)** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Dr. VALLI S
Professor and Head
Department of Computer Science and Engineering
Anna University
Chennai - 600 025



SIGNATURE

Dr. MARY ANITA RAJAM V
Professor
Department of Computer Science and Engineering
Anna University
Chennai - 600 025

ABSTRACT

Computer Vision techniques have become the new emerging field of technology which has the potential to solve the real time high demanding solutions in almost all the areas of interest. The main focus of the project is to predict how memorable an image will be among the humans using Computer vision techniques and Visual Memory Schema. This project has the capability to serve the purpose of enhancing the human understandability, diagnose memory problems, for effective retrieval of image search, Computer Graphics, summarization of Big data images and videos etc.

In this project we have shown that predicting an image memory is a task that can be targeted at current computer viewing techniques. We rated the memory uses a restricted test setting to obtain a reasonable quantity: define the image monument rate as the viewer's chances of receiving a duplicate of image within photo streams.

We have shown that there is a great balance between different viewers, and that some images are more memorable than others or no common items such as relatives or famous reminders. We will understand that memorability is a stable property of an image that is shared across different viewers.

ACKNOWLEDGEMENT

I wish to place on record my deep sense of gratitude to my supervisor **Dr. MARY ANITA RAJAM V**, Professor, Department of Computer Science and Engineering, Anna University, Chennai, for her highly valuable suggestions, encouragement and excellent guidance throughout the course of this Phase I project work.

I profoundly thank **Dr. VALLI S**, Professor and Head of the Department of Computer Science and Engineering, Anna University, Chennai, for her encouragement during this research work.

I wish to thank my panel committee members, **Dr. Chitrakala S**, Professor, Department of Computer Science and Engineering, Anna University, Chennai and **Dr. Manjula D**, Professor, Department of Computer Science and Engineering, Anna University, Chennai, for their valuable suggestions and guidance throughout the entire project work.

I wish to thank all the teaching and non-teaching members of the Department of Computer Science and Engineering, Anna University, Chennai for their cooperation and moral support. Last but not the least, I express my sincere thanks to my parents, for consistently and continuously providing tremendous support for completion of my project work.

HARSHA PRIYA G

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

CV	Computer Vision
HR	Hit Rate
FAR	False Alarm Rate
MI	Mutual Information
SVM	Support Vector Machine
SVR	Support Vector Regression
RBF	Radial Basis Function

CHAPTER 1

INTRODUCTION

1.1 COMPUTER VISION

Computer Vision which is abbreviated as CV, is a field of study that seeks to develop techniques to help computers to see and understand the content of digital footages and the world around. The matter of computer vision looks easy as a results of it is enjoyed by the people. It is an extension to the biological seeing of the world around us using computers. The driving factors behind the growth of computer vision is that amount of information that is there in today's world. We then use to train and make computer vision with a tremendous quantity of visual information which is more than 3 billion footage square measure shared on-line each day, the computing power required to analysis the data is presently accessible.

Computer vision has grown with new hardware and algorithms so that the accuracy rates for object identification, object classification is made more reliable. Today's systems have reached ninety percent accuracy from fifty percent making them further correct than humans at quickly reacting to visual inputs. Today the applications for computer vision have matured exponentially. By 2022, the computer vision and hardware market is foretold to attain 48.6 billion dollars.

1.2 EVOLUTION OF COMPUTER VISION

Early experiments in computer vision started in 1950s and it completely was the first place to use commercially to differentiate between oral and written communication by the 1970s. Before the deep learning bangs, the tasks that computer vision could perform were very limited. It required a lot of manual coding, effort by developers and human operators. For instance, to perform facial recognition, we have to perform the following steps:

- **Create a database:** One has to capture individual images of all the subjects one wanted to track in a specific format.
- **Annotate Images:** Then, for every individual image, one has to enter several key data points, such as distance between the eyes, the width of nose bridge, distance between upper-lip and nose, and dozens of other measurements that define the unique characteristics of each person.
- **Capture new images:**One has to capture new images, whether from photographs or video content. And then one has to go through the measurement process again, marking the key points on the image and also has to factor in the angle the image was taken.

After all this manual work, the application would finally be able to compare the measurements in the new image with the ones stored in its database and tell you whether it corresponded with any of the profiles it was tracking. In fact, there was very little automation involved and most of the work was being done manually. And the error margin was still large.

Machine learning provided a different approach to solving computer vision problems. With machine learning, developers no longer needed to manually code every single rule into their vision applications. Instead, they programmed "features," smaller applications that could detect specific patterns in images. They then used a statistical learning algorithm such as linear regression, logistic regression, decision trees or support vector machines (SVM) to detect patterns and classify images and detect objects in them.

Machine learning helped solve many problems that were historically challenging for classical software development tools and approaches. For instance, years ago, machine learning engineers were able to create a software that could predict breast cancer survival windows better than human experts. However building the features of the software required the efforts of dozens of engineers and breast cancer experts and took a lot of time develop.

Deep learning provided a fundamentally different approach to doing machine learning. Deep learning relies on neural networks, a general-purpose function that can solve any problem presentable through examples. When you provide a neural network with many labeled examples of a specific kind of data, it'll be able to extract common patterns between those examples and transform it into a mathematical equation that will help classify future pieces of information.

For instance, creating a facial recognition application with deep learning only requires you to develop or choose a reconstructed algorithm and train it with examples of the faces of the people it must detect. Given enough examples (lots of examples), the neural network will be able to detect faces without further instructions on features or measurements.

Deep learning is a very effective method to do computer vision. In most cases, creating a good deep learning algorithm comes down to gathering a large amount of labeled training data and tuning the parameters such as the type and number of layers of neural networks and training epochs. Compared to previous types of machine learning, deep learning is both easier and faster to develop and deploy. Most of current computer vision applications such as cancer detection, self-driving cars and facial recognition make use of deep learning. Deep learning and deep neural networks have moved from the conceptual realm into practical applications thanks to availability and advances in hardware and cloud computing resources.

1.3 APPLICATIONS OF COMPUTER VISION

CV In Self-Driving Cars: Computer vision enables self-driving cars to make sense of their surroundings.Cameras capture video from different angles around the car and feed it to computer vision software, which then processes the images in real-time to find the extremities of roads, read traffic signs, detect other cars, objects and pedestrians. The self-driving car can then steer its way on streets and highways, avoid hitting obstacles, and (hopefully) safely drive its passengers to their destination.

CV In Facial Recognition: Computer vision also plays an important role in facial recognition applications, the technology that enables computers to match images of people's faces to their identities. Computer vision algorithms detect facial features in images and compare them with databases of face profiles. Consumer devices use facial recognition to authenticate the identities of their owners. Social media apps use facial recognition to detect and tag users. Law enforcement agencies also rely on facial recognition technology to identify criminals in video feeds.

CV In Augmented Reality and Mixed Reality: Computer vision also plays an important role in augmented and mixed reality, the technology that enables computing devices such as smartphones, tablets and smart glasses to overlay and embed virtual objects on real world imagery. Using computer vision, AR gear detect objects in real world in order to determine the locations on a device's display to place a virtual object. For instance, computer vision algorithms can help AR applications detect planes such as tabletops, walls and floors, a very important part of establishing depth and dimensions and placing virtual objects in physical world.

CV In Healthcare: Computer vision has also been an important part of advances in health-tech. Computer vision algorithms can help automate tasks such as detecting cancerous moles in skin images or finding symptoms in x-ray and MRI scans.

1.4 CHALLENGES OF COMPUTER VISION

- Helping computers to see turns out to be very hard.
- Inventing a machine that sees like we do is a deceptively difficult task, not just because it's hard to make computers do it, but because we're not entirely sure how human vision works in the first place.
- Studying biological vision requires an understanding of the perception organs like the eyes, as well as the interpretation of the perception within the brain.
- Much progress has been made, both in charting the process and in terms of discovering the tricks and shortcuts used by the system, although like any study that involves the brain, there is a long way to go.

1.5 OBJECTIVE AND PROBLEM STATEMENT

The focus of the project is to predict how memorable an image will be using Computer vision techniques and Visual Memory Schema. In this project we do research on intrinsic image features that makes the image memorable. The main objective is to serve the purpose of enhancing the human understandability, diagnose memory problems, for effective retrieval of image search, Choosing a book cover, designing a logo, organizing images in a photo album, and selecting an image to decorate a website using Computer Graphics and summarization of Big data images and videos etc...

1.6 NEED FOR THE SYSTEM

- To serve the purpose of enhancing the human understandability and to diagnose memory problems.
- To gather the memory data for a large collection of images and determine which images left the trace in long term memory.
- For effective retrieval of images on searching using prediction algorithms.

1.7 ORGANIZATION OF THESIS

In this report, Chapter 2 gives the background on interactive discussion of related works. Chapter 3 describes the overall system design, algorithm and detailed description about each module and sub-modules. Chapter 4 describes the implementation of the proposed framework, results of the experiments conducted and performance evaluation.

CHAPTER 2

LITERATURE SURVEY

2.1 IMPORTANCE OF MEMORIES

E.Tulving *et al.* [6] in his work best describes how memories play an important role in learning. The work emphasises more on how the memories form the essential component to define ourselves.

2.2 CONSISTENCY ANALYSIS

T. F. Brady *et al.* [1] on their work illustrates the human memory capacity for visual information absorption and retrieval of images from the memory.

L. Standing *et al.* [5] on their work performed two experiments during which subjects learned either footage or descriptions of footage of images and then tested for recognition with either pictures or descriptions, altogether four mixtures. Recognition was best within the Picture-Picture condition, intermediate within the Picture-Word and Word-Word conditions which failed to take issue significantly, and worst within the Word-Picture condition. The additional variety of errors accessorial by dynamical to a metamorphosis condition (Picture-Word or Word-Picture) from the corresponding non-transformation condition (Picture-Picture or Word-Word respectively), was constant in either case. A model for recognition memory is projected that postulates that each pictorial stimuli and descriptive verbal stimuli are encoded in a very pictorial (or functionally equivalent) kind to that later transformations could also be applied by the experimental task. This model uses 2 parameters: a coffee background level related to pictorial secret writing and storage, and the next background level related to creating a verbal-pictorial transformation (or vice versa). The model is supported by a re-analysis of the information of Jenkins, Neale, and Deno (1967) and by the information of the 2 gift experiments. a further question, the chance of twin process of verbal and pictorial stimuli, was examined by using each footage and descriptions at the same time within the learning session and/or the take a look at the session. proof suggesting stereophonic operation was obtained in 3 out of 5 experimental conditions.

2.3 IMAGE MEMORABILITY

P. Isola *et al.* [3] on their work explained how the images exhibit the intrinsic property. Artists, advertisers, and photographers square measure habitually given with the task of making a picture can bear in mind ,whereas it's going to seem to be image memorability is only subjective, recent work shows that it's not associate degree paradoxical phenomenon: variation in memorability of pictures is consistent across subjects, suggesting that some pictures square measure in and of itself a lot of unforgettable than others, freelance of a subjects' contexts and biases.

During this paper work, they tend to use the publicly offered memorability dataset of P. Isola *et al.* [3] and increased the article and scene annotations with explainable spatial, content, and they tend to used a feature-selection theme with fascinating explaining-away properties to work out a compact set of attributes that characterizes the memorability of any person image. They discover that pictures of fogbound areas containing folks with visible faces square measure unforgettable, whereas pictures of vistas and peaceful scenes don't seem to be. Contrary to fashionable belief, uncommon or esthetical pleasing scenes don't tend to be extremely unforgettable. This work represents one in all the primary tries at understanding intrinsic image memorability, and opens a replacement domain of investigation at the interface between human knowledge and pc vision.

2.4 CLASSES OF IMAGE THAT PREDICT MEMORABILITY

Z. Bylinskii *et al.* [2] explains how the intrinsic and extrinsic effects on image memorability. Previous studies have known that pictures carry the attribute of memorability, a prophetical worth of whether or not a unique image are going to be later remembered or forgotten. Here the work has a tendency to investigate the interaction between intrinsic and foreign factors that have an effect on image memorability.

Firstly, they discovered that intrinsic variations in memorability exist at a finer-grained scale than antecedent documented. Secondly, they checked 2 foreign factors: image context and observer behaviour. Building on previous findings that pictures that area unit distinct with regard to their context area unit higher remembered, they proposed associate degree information-theoretic model of.

Their model mechanically predicted amendments in the context change of the memorability of natural pictures. Additionally to context, they studied a second foreign factor: wherever associate degree observer appearance whereas memorizing a picture. It seems that eye movements give extra info which will predict whether or not or not a picture are going to be remembered, on a trial-by-trial basis.

2.5 MEMORABILITY MAPPING

A. Oliva *et al.* [4] explains how to model the shape of the scene using a holistic representation of the spatial envelope. In this paper, their work tend to propose a process model of the popularity of universe scenes that bypasses the segmentation and also the process of individual objects or regions.

The procedure relies on a awfully low dimensional illustration of the scene, that their work tend to term the abstraction Envelope their work tend to propose a collection of sensory activity dimensions (naturalness, openness, roughness, expansion, ruggedness) that represent the dominant abstraction structure of a scene. Then, their work tend to show that these dimensions is also dependably calculable exploitation spectral and coarsely localized information.

The model generates a multidimensional area during which scenes sharing membership in linguistics classes (e.g., streets, highways, coasts) are projected closed along. The performance of the abstraction envelope model shows that specific info regarding object form or identity isn't a demand for scene categorization which modelling a holistic illustration of the scene informs regarding its probable linguistics class.

CHAPTER 3

SYSTEM DESIGN

3.1 INTRODUCTION

This chapter discusses the design of the system followed by implementation details. Initially the overall methodology is discussed followed by the implementation details for each of the modules in the system.

3.2 SYSTEM DESIGN

The system in Figure 3.1 aims at detailed explanation of the project of classifying the images based on memorable scores using visual memory schema.

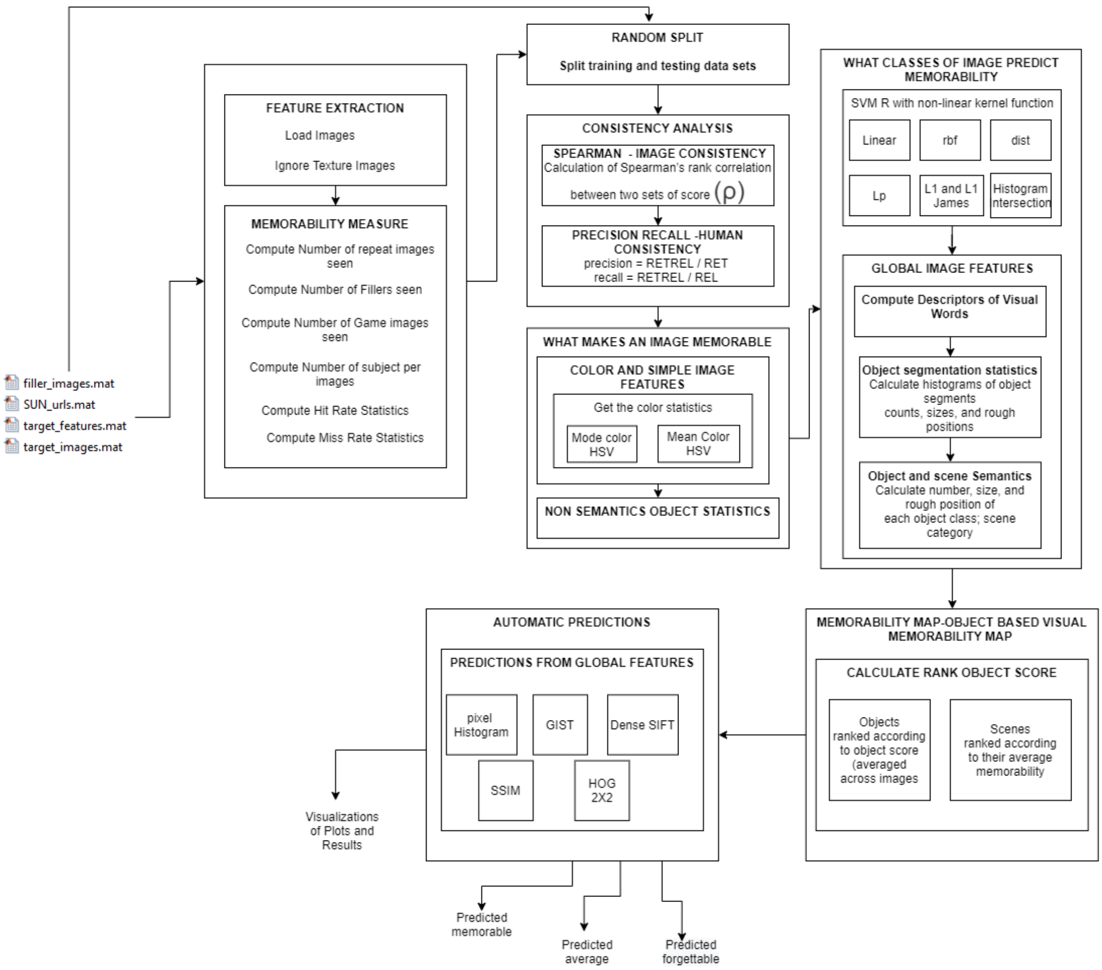


Figure 3.1: Detailed System Design

3.3 MODULE WISE DESCRIPTION

As shown in Figure 3.1, the system consists of following modules, namely The various phases of the project includes feature extraction, memorability measure, random split of the data sets, consistency analysis, image memorability, classes of the image that predict memorability, memorability map and automatic predictions. The mentioned modules of the proposed system are vividly explained in the following text.

3.3.1 Feature Extraction and Memorability Score

The aim of the feature extraction and memorability measure module is to compute the memorability measure and extract the necessary features, labels and annotations from the database. The input of the database is feed to the system in matrix form consisting of images of 2222 photographs from the SUN database ,2222 photos as target images and 8220 photos as filler images with pre-computed features and annotations.

We then, load the images and ignore the Texture Images for Feature extraction. Then, we compute the Memorability Measure to know the number of repeat images seen and Number of Filler images seen in Figure 3.2.

The aim of feature extraction is to improve the interpret-ability or perception of information in images for human viewers, or to provide 'better' input for other automated image processing techniques.

3.3.1.1 Pseudo code for feature extraction and measuring memorability

Step 1: Input the database in matrix form consisting of images of 2222 photographs from the SUN database, 2222 photos as target images and 8220 photos as filler images with pre computed features and annotations.

Step 2: Load the images and ignore the Texture Images for Feature extraction.

Step 3: Compute Memorability Measure:

a) Compute Number of repeat images seen.

b) Compute Number of Filler images seen.

c) Compute Number of Game images seen.

d) Compute Number of Subjects per image.

e) Compute Hit Rate Statistics.

INPUT: Database in matrix form consisting of images of 2222 photographs from the SUN database, 2222 photos as target images and 8220 photos as filler images with pre computed features and annotations.

OUTPUT: Computed Memorability Score and Visualizations.

INFERENCE: Thus, memorability scores are a good measure of correct memories.

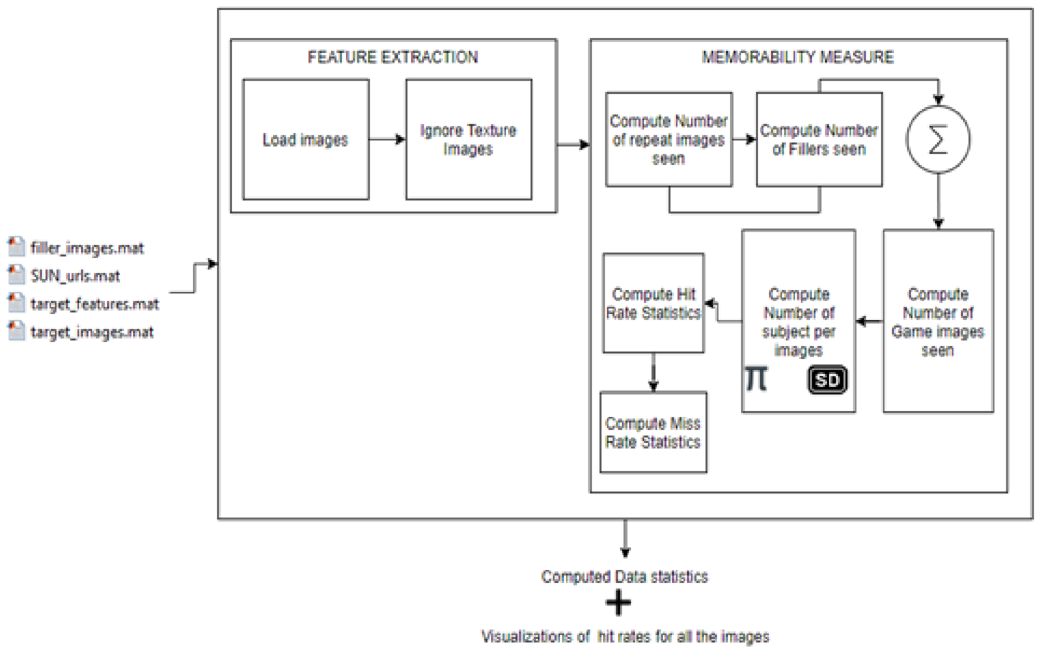


Figure 3.2: Feature Extraction and Measuring Memorability

3.3.2 Random Splits of Training and Testing Data

The dataset is split into testing and training data using random splits of 25 sets.

3.3.2.1 Pseudo code for Random Splits

Step 1: Input the database in matrix form consisting of images of 2222 photographs from the SUN database, 2222 photos as target images and 8220 photos as filler images with pre-computed features and annotations.

Step 2: Generate many random splits.

Step 3: Get the target detection vector and Ignore the texture images.

Step 4: Split Half subject Hit Rate Ratio.

Step 5: Split the target detection Vectors into training and testing Hit Rate Ratios.

Step 6: Output consists of Splited training and testing data in matrix format.

INPUT: Database in matrix form consisting of images of 2222 photographs from the SUN database, 2222 photos as target images and 8220 photos as filler images with pre computed features and annotations.

OUTPUT: Splited training and testing data in matrix format seen in Figure 3.3

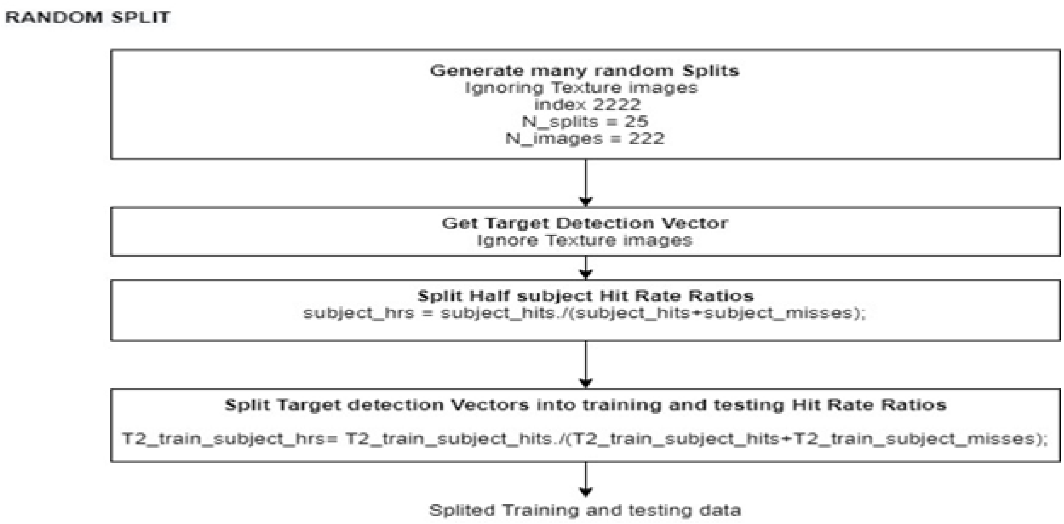


Figure 3.3: Random Splits of Training and Testing Data

3.3.3 Consistency Analysis

This module Consistency Analysis shows how high human-to-human memorability consistency can be and how image is consistent.This module shows that our data has enough consistency and that would be possible to predict image memorability.

3.3.3.1 Pseudo code for consistency analysis

Step 1: Input the Splited training and testing images for analysis of consistency model.

Step 2: Compute the Spearman's Rank Correlation for image Consistency analysis

a) Compute Mean Scores per Images

b) Compute Consistency of Image

Step 3: Compute Precision Recall for Human Consistency Analysis

a) Sort the images by their score given by the first half of the participants.

b) Calculate the cumulative average memorability according to the second half of the participants

c) Use the Object detector function to generate the structure with the detection annotation.

d) Get the object Confidence and compute the precision Recall.

INPUT: Splited training and testing images for analysis of consistency model.

OUTPUT: The percent of times that these images were remembered by an independent set of participants is elaborated in Figure 3.4

INFERENCE: Analysis shows how high human-to-human memorability consistency can be. Thus, our data has enough consistency that it should be possible to predict image memorability.

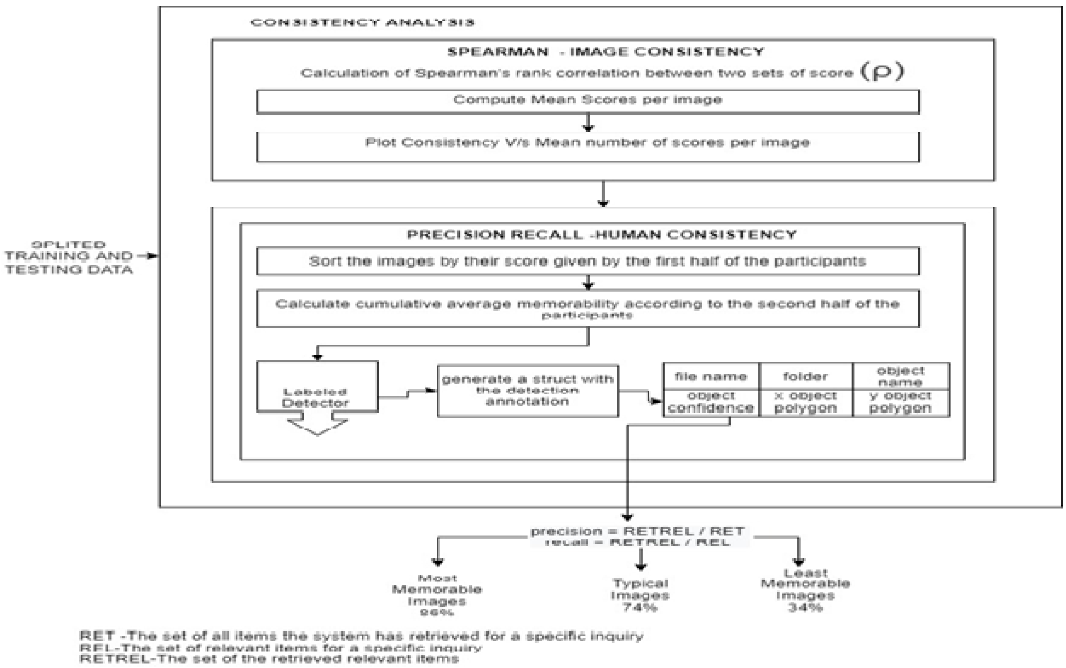


Figure 3.4: Consistency Analysis

3.3.4 Image Memorability

This module helps us to understand that the simple image features, as well as non-semantic object statistics, do not correlate strongly with memorability score and that statistics do not predict memorability.The pseudo code of image memorability is shown below:

3.3.4.1 Pseudo code for image memorability

Step 1: Input the system with Memorable Score with Visualization of Wide range of memorability and high inter subject consistency.

Step 2: Color and Simple image feature analysis using memorability as a function of a color.

1) Get the sorted target data Hit Rates.

2) Get the color Statistics.

a) Compute Mode Color Histogram Vector.

b) Compute Mean Color Histogram Vector.

Step 3: Compute the Non-Semantics Object Statistics

a) Compute Object Counts V/s Hit Rate.

b) Compute Object mean Area V/s Hit Rate.

Step 4: Visualizations of Color Statistics and Non Semantics Object Statistics.

INPUT: Memorable Score with Visualization of Wide range of memorabilities and high inter subject consistency.

OUTPUT: Visualizations of Color Statistics and Non Semantics Object Statistics as explained in Figure 3.5

INFERENCE: Simple image features, as well as non-semantic object statistics, do not correlate strongly with memorability score and that statistics do not predict memorability.

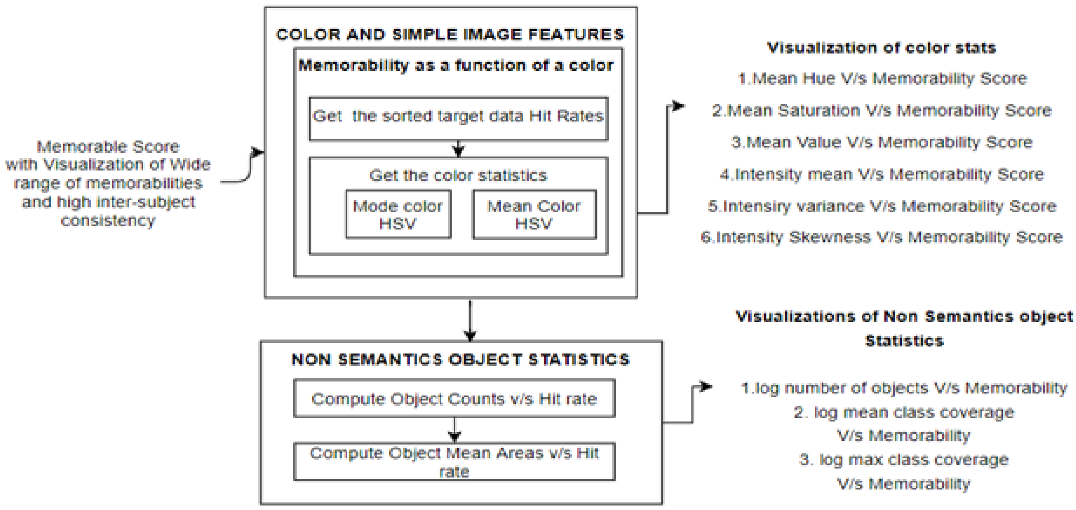


Figure 3.5: Image Memorability

3.3.5 Finding Classes of Image that Predict Memorability

This module elaborates the analysis of object semantic regressions and demonstrates that if a system knows which objects an image contains.It is able to predict memorability with a performance not too far from human consistency. It explains that semantic images contribute for image memorability.

3.3.5.1 Pseudo code for finding classes of image predict memorability

Step 1: Input the system with Computed results of Non Semantics Object Statistics.

Step 2: Use Prediction algorithm for effective correlation with memorability score

1) Calculate Support Vector Machine Regression with Non-linear

Kernel Functions.

2) Use the computation of kernel functions with global image features.
   a) Compute Descriptors of Visual words.

Step 3: Calculate histograms of object segments, count, sizes and rough positions for object segmentation statistics.

Step 4: Calculate number size and rough position of each object class scene category.

INPUT: Computed results of Non Semantics Object Statistics.

OUTPUT: Average memorability for top N ranked images(%) V/s Predicted Rank N as explained in Figure 3.6

INFERENCE: Thus, analysis of object semantic regressions demonstrate that if a system knows which objects an image contains, it is able to predict memorability with a performance not too far from human consistency. Semantic images contribute for image memorability.

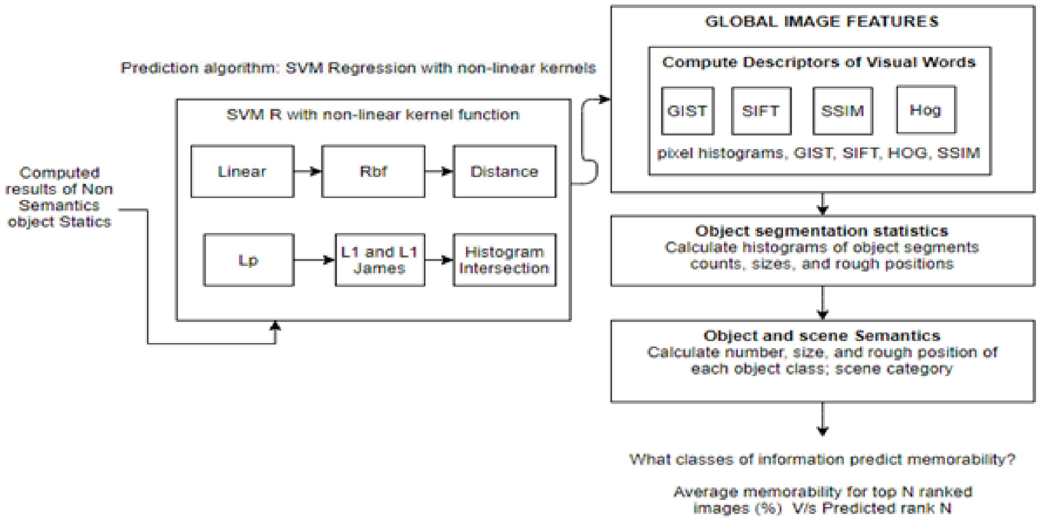


Figure 3.6: Finding Classes of Image that Predict memorability

3.3.6 Memorability Map

This module explains that the memorability maps shade each object according to how much the object adds to, or subtracts from the image's predicted memorability.

To quantify the contribution of an object i to an image, we take a prediction function f, that maps object features to memorability scores and calculate how its prediction m changes when we zero features associated with object i from the current image's feature vector $(a_1, ..., a_n)$.

This gives the score s_i for each object in an given images:

$$m_1 = f\ (a_1,..........., a_i, ...a_n) \qquad (1)$$

$$m_2 = f\ (a_1,..........., 0\ , ...a_n) \qquad (2)$$

$$s_i = m1 - m2 \tag{3}$$

For prediction function we use our SVR on Labeled Multi scale Object Areas, trained above and we plot memory maps on test set images. In red we show objects that contribute to higher predicted memorability and in blue are the objects that contribute to lower predicted memorability. Brightness is proportional to the magnitude of the contribution the results of average measured memorability of each sample set.

3.3.6.1 Pseudo code for memorability map

Step 1: Input the system with the Average memorability for top ranked images (%) V/s Predicted rank N.

Step 2: Calculate the object exclusion Factor.

a) Remove the feature at index.
b) Make the new prediction without features.
c) Get the Exclusive Factor Calculation.

Step 3: Create Memorability map objects.

a) Visualize objects based on memorability map.

Step 4: Calculate Object Scores.

Step 5: Calculate Rank Object Scores.

INPUT: Average memorability for top ranked images (%) V/s Predicted rank N

OUTPUT: Visualizations of Objects ranked according to object score (averaged across image),Scenes ranked according to their average memorability.

INFERENCE: It helps to visualize what content makes an image memorable as explained in Figure 3.7

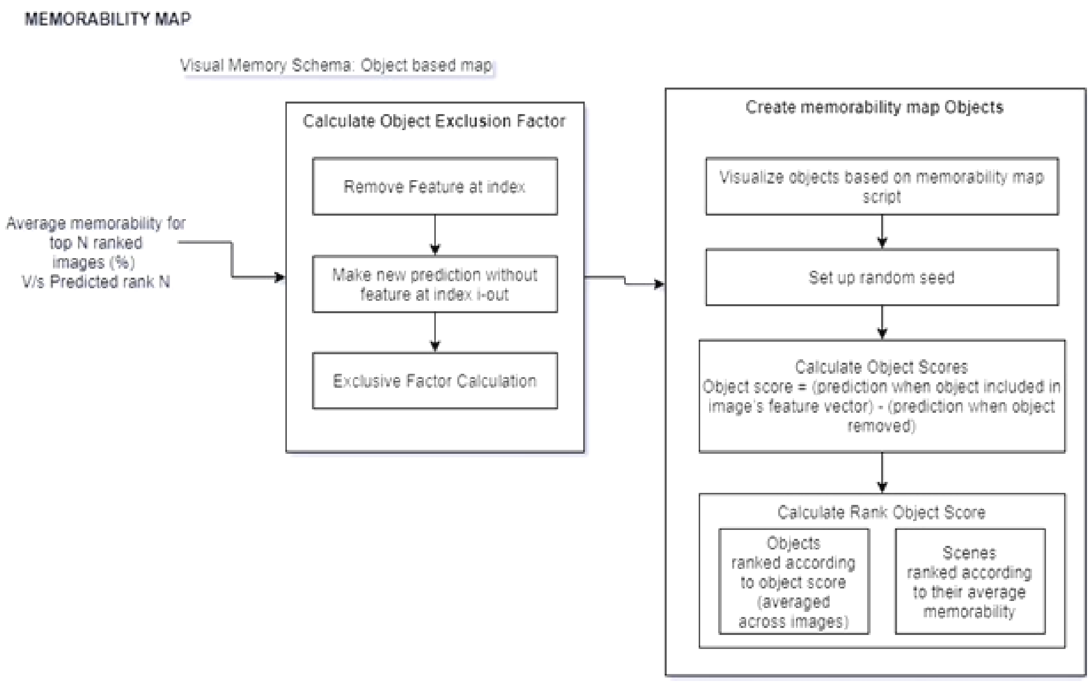


Figure 3.7: Memorability Map

3.3.7 Automatic Predictions

In this module the predicted memorability rank is found on the basis of global image features with the empirical memorability rank and the mean rank error between predicted and measured ranks across each set of images.

3.3.7.1 Pseudo code for automatic predictions

Step 1: Input the system with the object and scenes ranked according to the object score.

Step 2: Load the Prediction Data using Prediction function.

Step 3: Make Predictions from the Global Features using the following:

a) Pixel Histograms

b) GIST

c) Dense SIFT

d) SSIM

e) HOG 2x2

INPUT:

Objects and scenes ranked according to the object score.

OUTPUT:

1) Visualizations of Average memorability for top N ranked Images V/s Predicted Rank for all global features as explained in Figure 3.8

2) Visualization of Predicted

a) Memorable

b) Average

c) Forgettable

3) Visualization of Errors Over Predicted and Errors Under Predicted

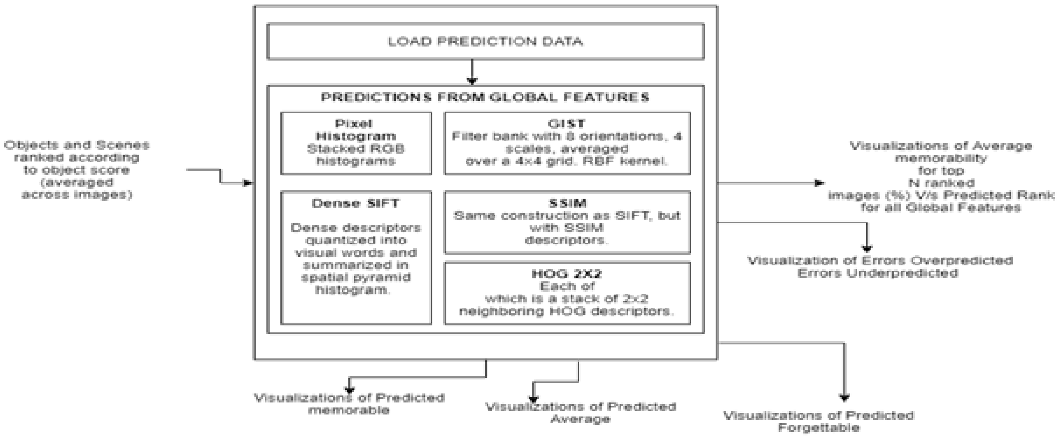


Figure 3.8: Automatic Predictions

CHAPTER 4

RESULTS AND DISCUSSION

The implementation of the proposed system ”CLASSIFICATION OF IMAGES BASED ON MEMORABLE SCORE USING VISUAL MEMORY SCHEMA AND COMPUTER VISION TECHNIQUES” is detailed in this chapter. This chapter includes the experimental setup,test cases executions,experimental results and evaluation metrics of the system.

4.1 EXPERIMENTAL SETUP

Dataset Name : SUN DATABASE

Source : http://labelme.csail.mit.edu

Accessibility Type : Free

The experimental set up consists of the database which is feed in matrix form consisting of photographs from the SUN database, 2222 photos as target images and 8220 photos as filler images with pre computed features and annotations and memorability measurements from ”Memory Game”.

In the memory game participants received feedback whenever they pressed a key i.e a green symbol shown at the center of the screen for correct detection. and a gray X for an error. Image sequences were broken up into levels that consisted of 120 images each. Each level took 4.8 minutes to perform. Participants could complete at most 30 levels and were able to exit the game at any time.

The target and filler images represents a random sampling of the scene categories from the SUN database. In the memory game participants were presented with a sequence of images, each of which is displayed with a 1.4 second gap in between image presentations as shown in Figure 4.1. The role of fillers were twofold:

1. They provide a spacing between the first and second repeat of a target.

2. The responses of the repeated fillers constituted a 'vigilance task' that allow us to continuously check that participants were attentive to the task.

Repeat of images occurred on the fillers with a spacing of 1-7 images and on the targets with a spacing of 91-109 images.

C, C++, Scripts, .mat for implementation. Mat Lab Software is used for execution of code and visualizing the results and plotting in Windows OS platform.

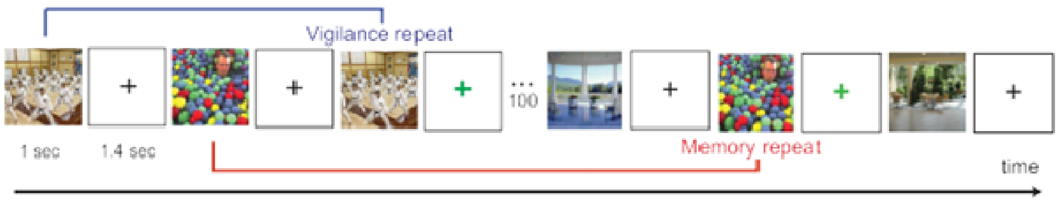


Figure 4.1: Number of Repeats Seen on a Long Stream of Images in a Memory Game

4.2 TEST CASES

The proposed research work is tested with various test cases such as Natural Scene Images, Outdoor Images, Indoor Images and Focused Images and tested for its accurate predictability as shown in Table 4.1

Table 4.1: Consistency Table

Spearman rank correlation ρ	Percentage
Predictions and measurements	75 %
Global features predictions	46 %

4.3 EXPERIMENTAL RESULTS

In Figure 4.2 the results of Feature extraction and measuring memorability shows the score of each image which is defined as the percentage of correct detection by participants. On average, each image each image was scored by 78 participants. The average memorability score was 67.5% .Average false alarm rate was 10.7%. As, the false alarm rate was low in comparison with correct detection.We infer that memorability scores are good measure of correct memories as shown in Figure 4.3

```
Command Window
C:\Users\Harsha\Desktop\CSE\Prediction

>> CalculateBasicStatsForCVPR
how many subjects cover >= 90% of target repeats:
   347

how many subjects saw any target repeats:
   643

how many subjects entered the game (made it past the demo):
     0

avg number of levels played by subjects who entered the game (made it past demo
mean num scores per image:
   77.8447

std num scores per image:
    7.8100

mean hit rate:
    0.6750

std hit rate:
    0.1356

mean false alarm rate:
    0.1072
```

Figure 4.2: Feature Extraction and Measuring Memorability

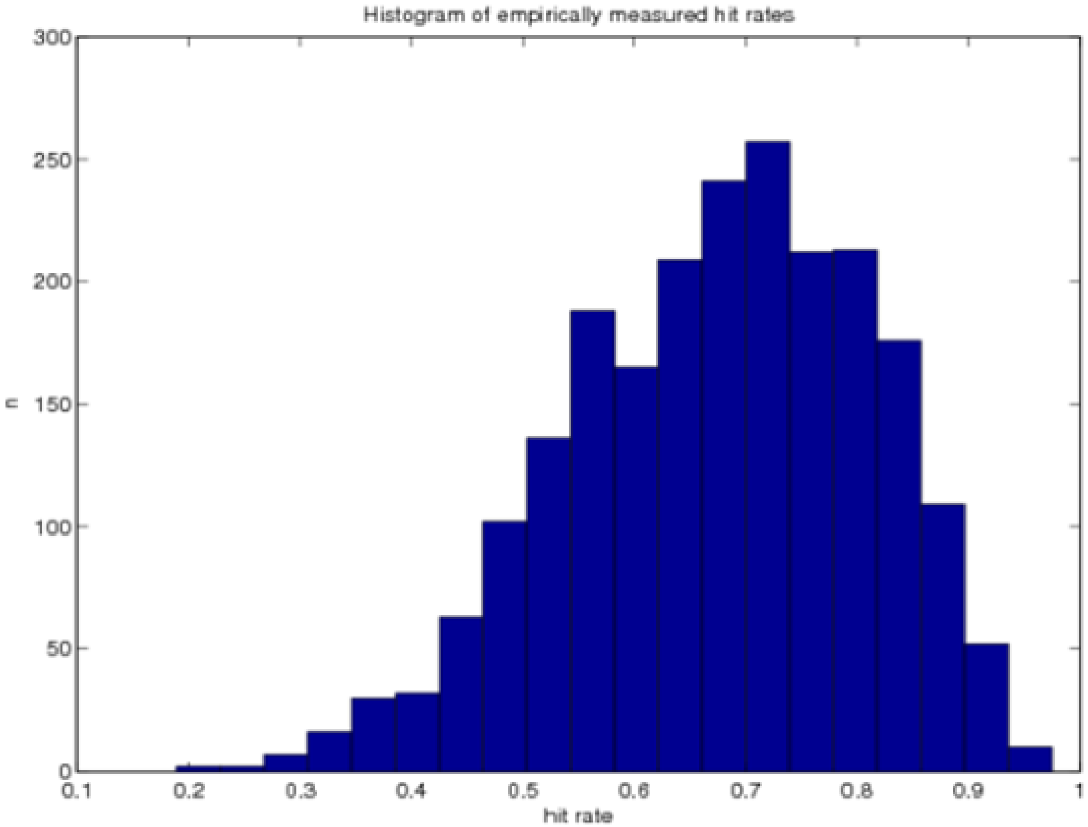


Figure 4.3: Histogram for Feature Extraction and Measuring Memorability

25 sets of Splited training and testing data in matrix format seen in Figure 4.4 and the sample data set after spilt are shown in Figure 4.5. The hit rate of Group 1 set of people is in subject hrs1 file.The hit rate of Group 2 set of people is in subject hrs2 file.

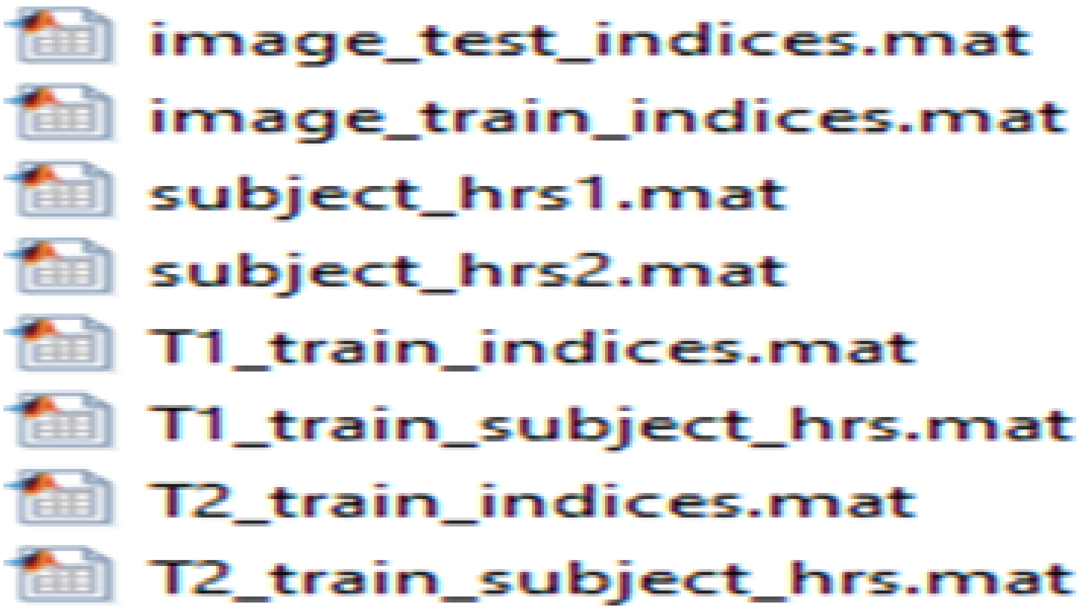


Figure 4.4: Random Split

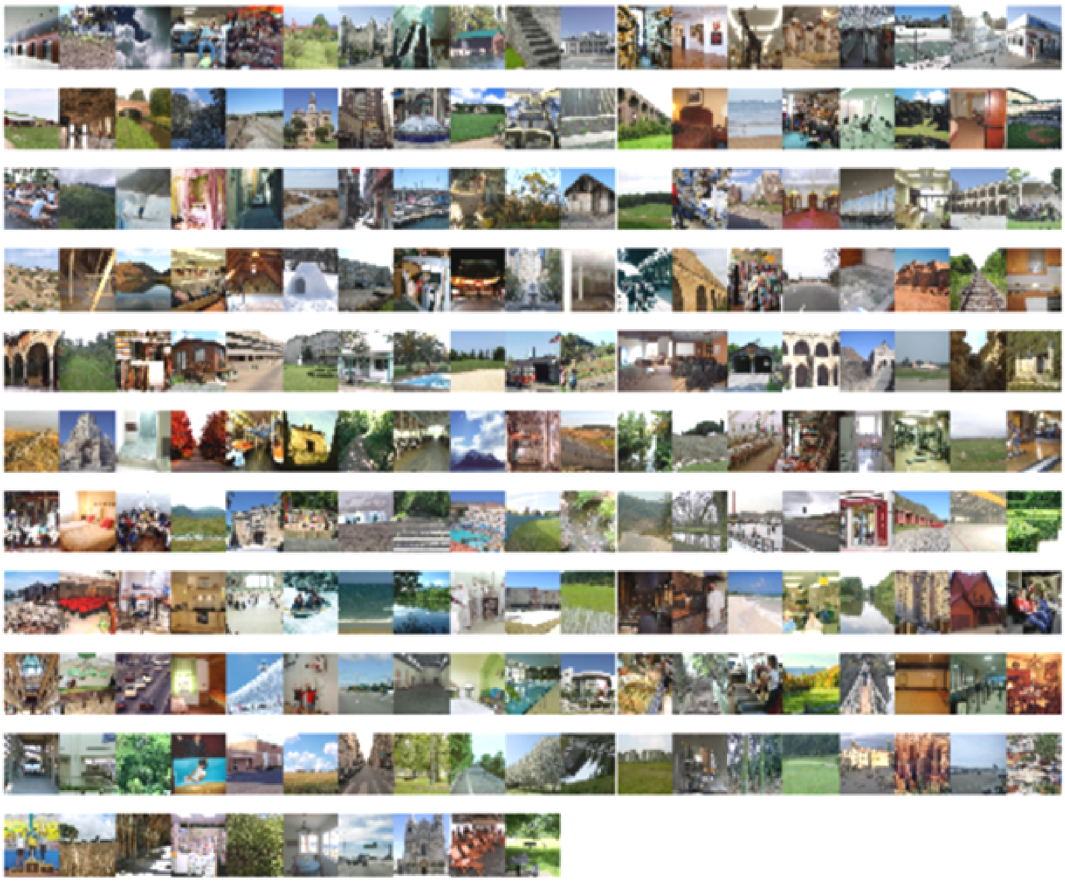


Figure 4.5: Database sample set after split

In Figure 4.6 shows the results of human consistency. Participants were split into two Group 1 and Group 2. Images were ranked by memorability scores from participants in group 1 and plotted against the cumulative average memorability scores given by the participants in the two groups.

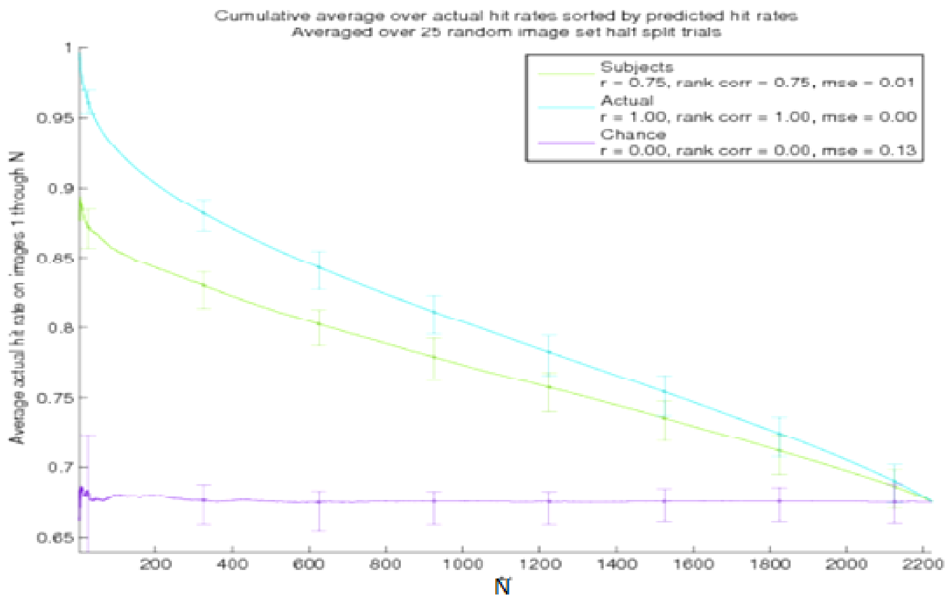


Figure 4.6: Human Consistency

Figure 4.7 shows the Spearman's rank correlation between subject groups 1 and 2 as a function of the mean number of scores per image and it accounts to 75% thus proving that image is consistent for making automatic predictions.

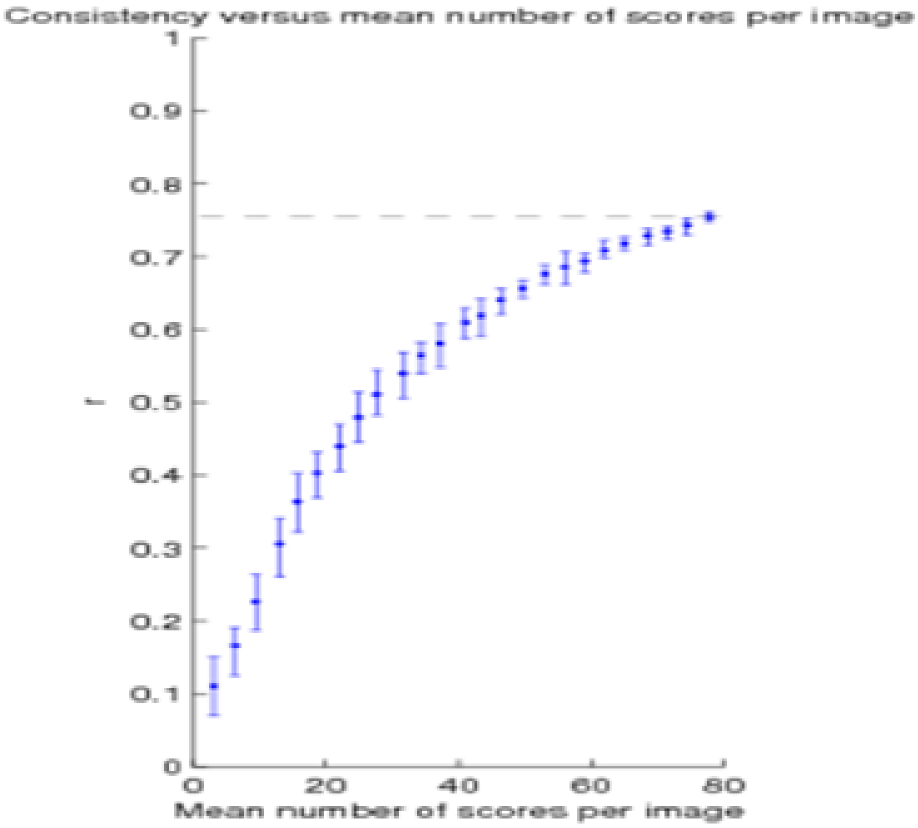


Figure 4.7: Spearman's Correlation for image consistency

Figure 4.8 the visualizations of color and simple image features of images are plotted with the correlation between memorability and basic pixel statistics to determine whether or not an image will be memorable. In the results we can conclude that mean hue is weakly predictive of memory. The mean hue transition from red to green to blue to purple, memorability tends to go down to -0.16. This weak correlation is due to blue and green outdoor landscapes being remembered less frequently than more warmly colored human faces and indoor scenes. Mean saturation and mean value also exhibit weaker correlation with memorability.

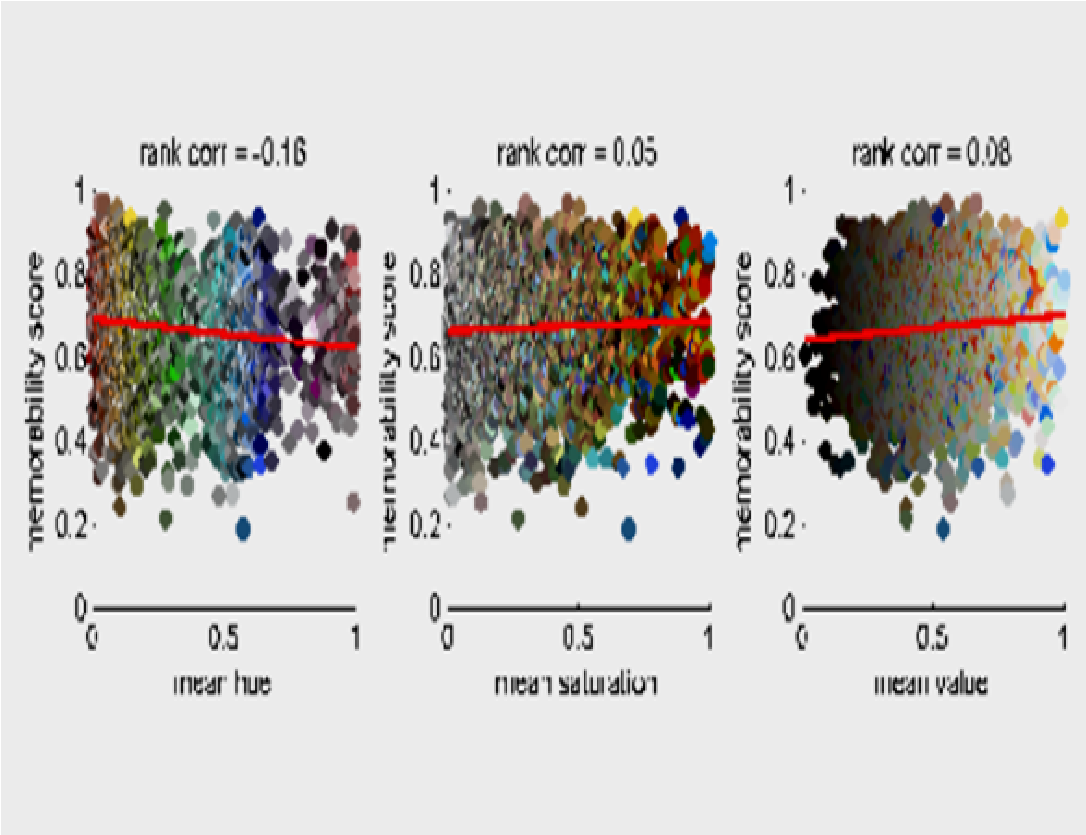
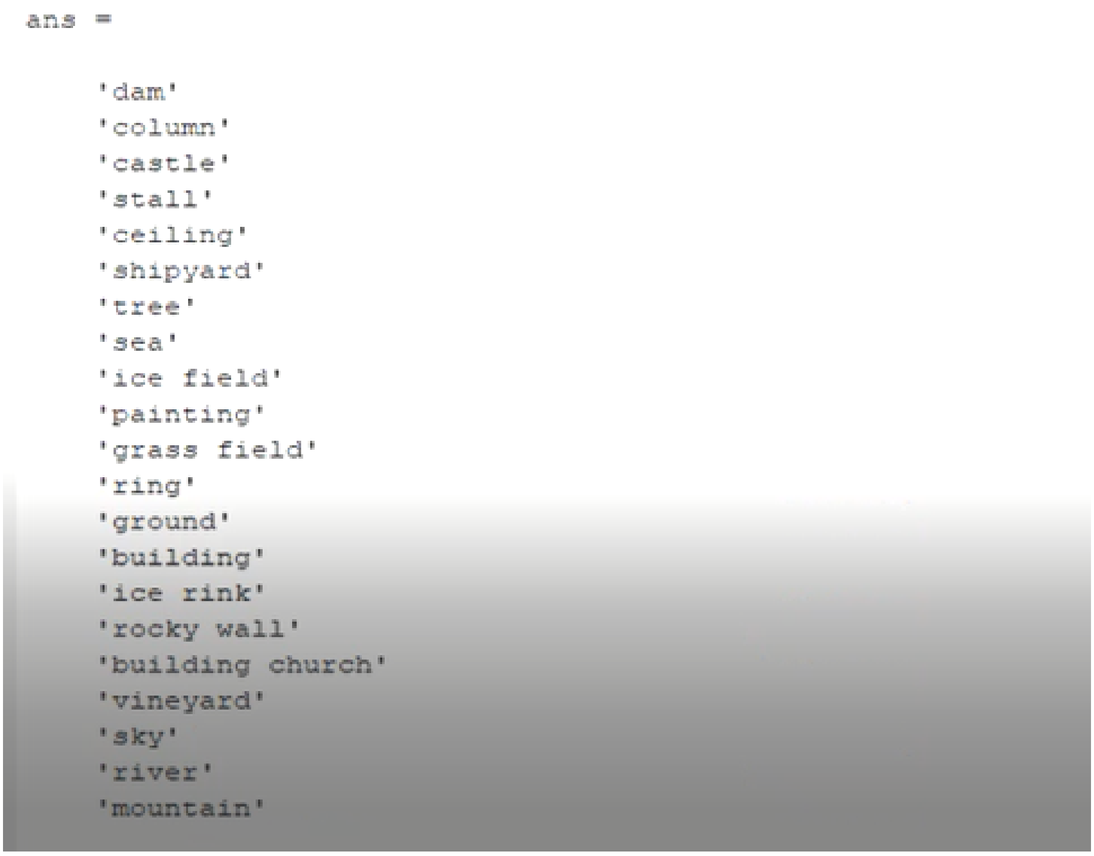


Figure 4.8: Visualization of Color Stats

Figure 4.9 shows that each image in the target set was segmented into object regions and each of these segments was given an object class label using Label Me function.



```
ans =

    'dam'
    'column'
    'castle'
    'stall'
    'ceiling'
    'shipyard'
    'tree'
    'sea'
    'ice field'
    'painting'
    'grass field'
    'ring'
    'ground'
    'building'
    'ice rink'
    'rocky wall'
    'building church'
    'vineyard'
    'sky'
    'river'
    'mountain'
```

Figure 4.9: Segmentation of Target Images

Figure 4.10 shows the quantified degree to which our data can be explained by non-semantics object statistics. We found that simple image features as well as non -semantic object statistics do not correlate strongly with memorability score.

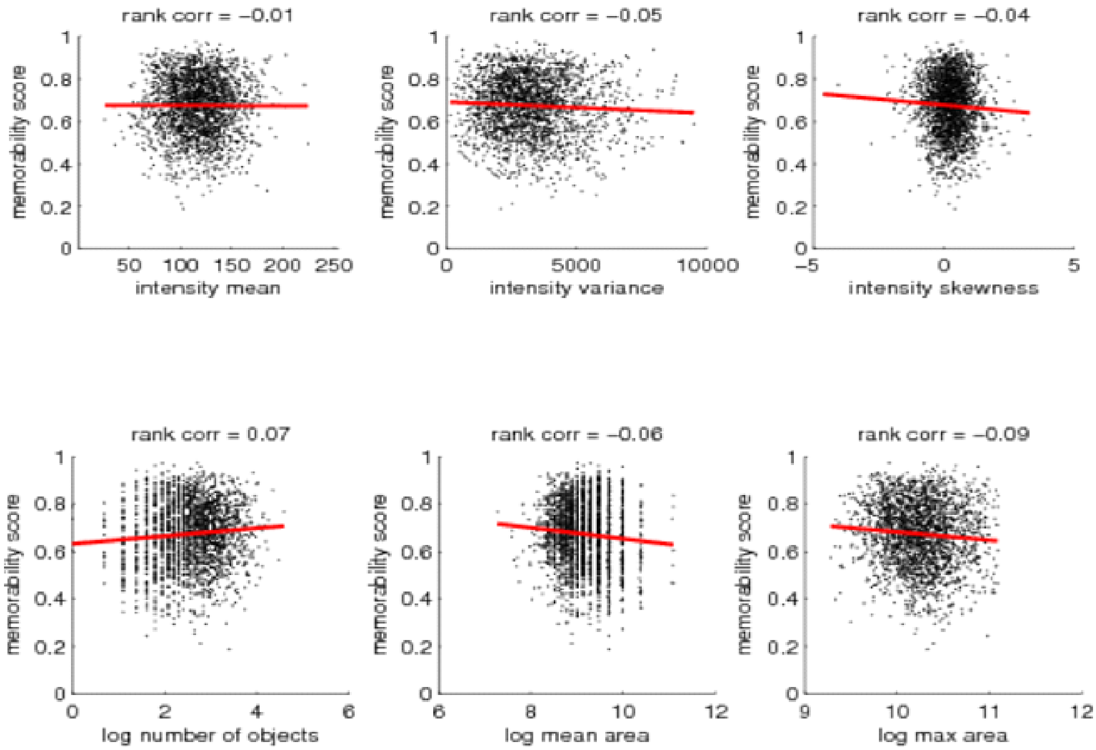


Figure 4.10: Visualization of Non-Semantics Object Stats

Figure 4.11 shows the set of images with actual mean, min and max for top memorability, middle memorability and bottom memorability set of images.

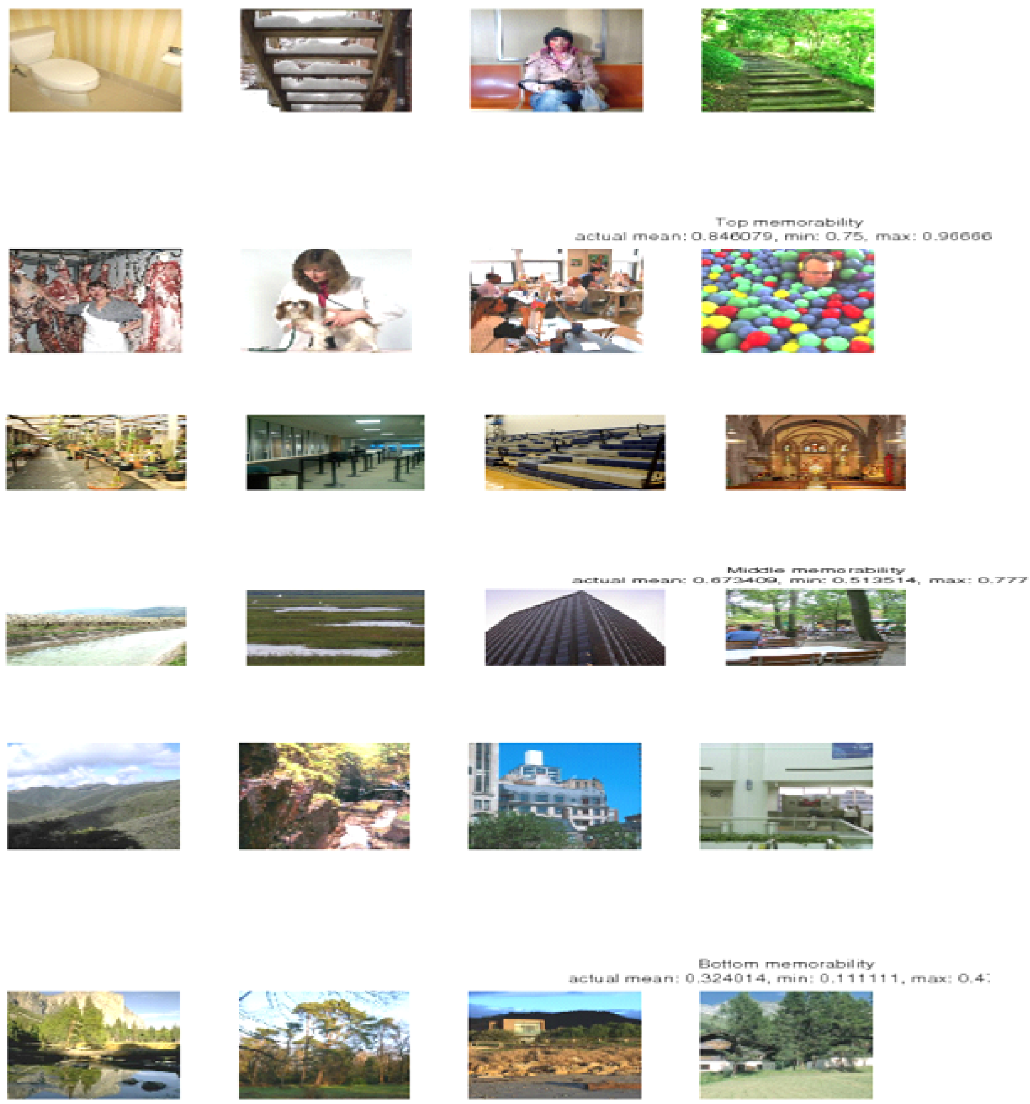


Figure 4.11: The Percentage of times that these Images were remembered by an Independent Set of Images

Figure 4.12 shows memorability map. These maps shade each object according to how much the object adds to, or subtracts from, the image's predicted memorability. For prediction function we use our SVR on Labeled Multi scale Object Areas, trained above and we plot memory maps on test set images.

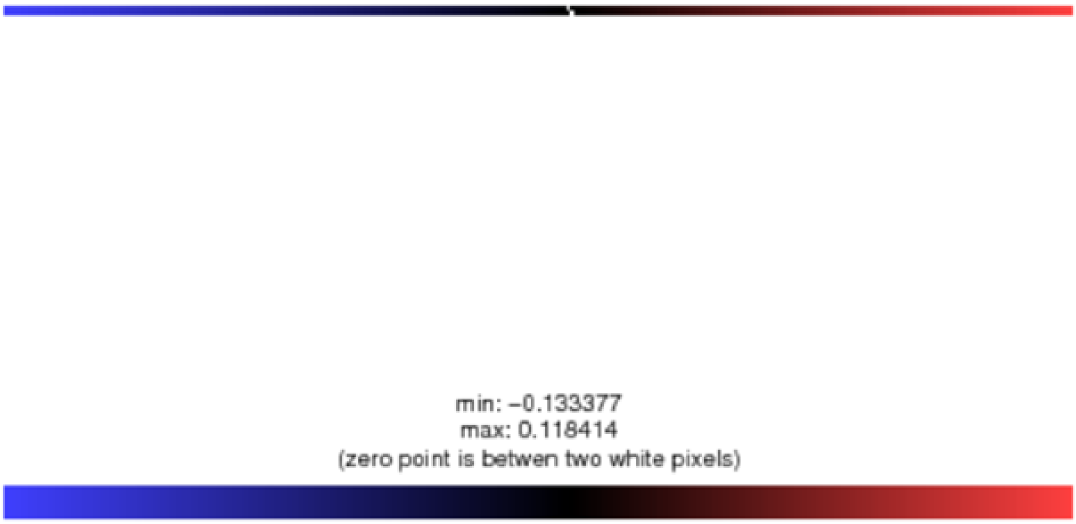


Figure 4.12: Memorability Map Color Scale

In red we show objects that contribute to higher predicted memorability and in blue are the objects that contribute to lower predicted memorability. Brightness is proportional to the magnitude of the contribution the results of average measured memorability of each sample set. Figure 4.13 shows predicted highly memorable of 91% . Figure 4.14 shows the results of predicted as typical memorability 68%. Figure 4.15 shows Predicted as unmemorable 55%

Figure 4.13: Predicted as Highly Memorable 91%

Figure 4.14: Predicted as Typical Memorable 68%

Figure 4.15: Predicted as unmemorable 55%

Figure 4.16, Figure 4.17, Figure 4.18 shows the sample images on which the global features regression performed poorly. These predictions produce clear visual distinctions, but may fail to notice more subtle cues that make certain images more memorable than others.

Figure 4.16: Predicted As Most Memorable 87%

Figure 4.17: Predicted As Typical Memorability 68%

Figure 4.18: Predicted As Least Memorable 52%

Figure 4.19 shows the results of predicted memorability rank on the basis of global image features with the empirical memorability rank and the mean rank error between predicted and measured ranks across each set of images.

Figure 4.19: Automated Predictions

4.4 PERFORMANCE METRICS

This section explains the various performance metrics that are considered to evaluate the system.

Memorability

Memorability is the measure of probability of a correctly detecting a repeat after a single view of an image in a long stream of images. Mathematically it is defined as Equation 4.1

$$\text{Memorability} = \text{Probability of correctly detecting a repeat after a single view} \tag{4.1}$$

Object Score

Object Score is the difference between the prediction when object included in image's feature vector and prediction when object removed. Mathematically it is defined as Equation 4.2

$$\text{Object Score} = P_O - P'_O \tag{4.2}$$

Where, P_o is Prediction when object included in image's feature vector and P'_O is Prediction when object removed.

Hit Rate (HR)

Hit Rate is the ratio of Number of Images chosen correctly as remembered by observers and Total number of their occurrences as a repeat image. Mathematically it is defined as Equation 4.3

$$\text{Hit Rate} = \frac{\text{Number of Images chosen correctly as remembered by observers}}{\text{Total number of their occurrences as a repeat image}} \tag{4.3}$$

False Alarm Rate (FAR)

False Alarm Rate is the ratio of hits of an Image to the total number of its occurrences as a second-stage-filler that is the non-repeat image. Mathematically it is defined as Equation 4.4

$$\text{False Alarm Rate (FAR)} = \frac{\text{False Hits of an Image}}{\text{Total number of occurrences as a second stage filler image}} \tag{4.4}$$

Spearman linear correlation coefficient ρ

The Spearman correlation between two variables is equal to the Pearson correlation between the rank values of those two variables; while Pearson's correlation assesses linear relationships, Spearman's correlation assesses monotonic relationships whether linear or not. If there are no repeated data values, a perfect Spearman correlation of +1 or -1 occurs when each of the variables is a perfect monotone function of the other. Mathematically it is defined as Equation 4.5.

$$\rho_{A,B}^{2D} = \frac{1}{n} \cdot \sum_{i,j} \frac{(A(i,j) - \mu_A) \cdot (B(i,j) - \mu_B)}{\sigma_A \cdot \sigma_B} \tag{4.5}$$

Mutual information(MI)

Mutual information is the confidence scale that allows us to produce ROC curves that provided us with a sensitivity measure of the experiment and mathematically defined as Equation 4.6.

$$I_{A,B} = \sum_{a \in A} \sum_{be\ B} p(a,b) \cdot \log \left(\frac{p(a,b)}{p(a)p(b)} \right) da \cdot db \tag{4.6}$$

Precision

Precision is the ratio of the set of retrieved relevant items to the set of all items the system has retrieved for a specific inquiry.

$$\text{Precision} = \frac{\text{RETREL}}{\text{RET}} \tag{4.7}$$

where RETREL-The set of the retrieved relevant items. RET -The set of all items the system has retrieved for a specific inquiry.

Recall

Recall is the ratio of the set of retrieved relevant items to the The set of relevant items for a specific inquiry.

$$\text{RECALL} = \frac{\text{RETREL}}{\text{REL}} \tag{4.8}$$

where RETREL-The set of the retrieved relevant items.REL-The set of relevant items for a specific inquiry

CHAPTER 5

CONCLUSION

5.1 CONCLUSION AND FUTURE WORK

Making memorable photos is a challenging task in viewing and taking pictures, and is often presented as an obscure concept that is difficult to measure. Surprisingly, there has never been a previous attempt to measure this asset systematically in photo collections, and using computer viewing techniques to extract memory automatically. Scaling auto-image formats is an active domain for research many applications. For example, image memory can be used to extract, from a collection of images, those that can be most memorable viewers. This can be used to select image images, covers, user-friendly connectors, etc.

In this project we have shown that predicting image memory is a task that can be targeted at current computer viewing techniques. We rated the memory uses a restricted test setting to obtain a reasonable quantity: define the image monument rate as the viewer's chances of receiving a duplicate of image within photo streams. We have shown that there is a great balance between different viewers, and that some images are more memorable than others or no common items (such as relatives or famous reminders).

This work is the first attempt to measure the useful quality of each image. In the work to come it will be interesting to investigate the relationship between image memory and other steps such as object value, creativity, and image quality.

REFERENCES

1. T. F. Brady, T. Konkle, and G. A. Alvarez. A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of vision*, 11(5):4–4, 2011.

2. Z. Bylinskii, P. Isola, C. Bainbridge, A. Torralba, and A. Oliva. Intrinsic and extrinsic effects on image memorability. *Vision research*, 116:165–178, 2015.

3. P. Isola, D. Parikh, A. Torralba, and A. Oliva. Understanding the intrinsic memorability of images. In *Advances in neural information processing systems*, pages 2429–2437, 2011.

4. A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 42(3):145–175, 2001.

5. L. Standing and P. Smith. Verbal-pictorial transformations in recognition memory. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 29(4):316, 1975.

6. E. Tulving. " episodic and semantic memory." pp. 381-403 in organization of memory, edited by e. tulving and w. donaldson. new york: Academic press. 1972.