ESTIMATING THE EMOTIONAL CONTENT OF AN IMAGE FROM THE OBSERVER'S EYE SCAN PATERN

A Thesis

Presented to

The Faculty of the Department of Electrical and Computer Engineering

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree of

Master of Science in

Electrical Engineering

By

Shrivatsa Neerkaje

December, 2016

ESTIMATING THE EMOTIONAL CONTENT OF AN IMAGE FROM THE

OBSERVER'S EYE SCAN PATERN

An Abstract

of a

Thesis

Presented to

The Faculty of the Department of Electrical and Computer Engineering

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree of

Master of Science in

Electrical Engineering

By

Shrivatsa Neerkaje

December, 2016

Abstract:

The aim of the study is to predict the emotional gist of the image, namely the level of arousal (low or high) and kind of emotion (positive or negative) that the image elicits from the pattern of eye movements of a human observer. Images were selected based on their arousal and valence ratings. The observers (n=32) viewed the images in a random order and their pattern of eye movements was recorded with a head-mounted eye tracker. Features pertaining to saccades and fixation were extracted. Feature values obtained from the eye scan pattern data were fed into a random forest algorithm in MATLAB. Performing 10 fold cross-validation yielded a classification efficiency of 57% on low versus high arousal images, and 56% on positive versus negative valence images (*a priori* probability=50%). Several dynamic features were added to improve the efficiency though the effort proved to be unfruitful. Finally, the images were checked to see if they really show any difference by training them through a Convolutional Neural Network. The model showed a classification efficiency of 85% based on Valence and 75% based on Arousal.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 - INTRODUCTION

1.1 Objective

Every day we witness lots of events and experience a whole host of emotions. Some of the experiences are highly arousing (high arousal), and others are not (low arousal). Independent of arousal content, the basic kind of emotion elicited, i.e., its emotional valence, is a different entity altogether. An emotion can either be positive (positive or high valence), e.g., the sight of a spectacular waterfall or a pack of puppies playing in the park, or negative (negative or low valence), e.g., a violent or scary scene in a movie. Emotional response to a visual stimulus can be obtained by neurophysiological recordings in the amygdala and other key regions of the brain. However, these measurements are typically invasive and are not done in humans but rather on animals – if single cell or multicellular recordings are performed in said brain areas – or expensive and time-consuming – if a non-invasive technology like functional magnetic resonance imaging (fMRI) is used. A non-invasive, relatively easy and cheap measure that can be used to gauge the emotional response of an individual is thus required to circumvent these limitations.

Emotion can be identified from recording facial expression while viewing an image. This approach has been tried with notable success. However, the approach has some limitations. Snapshots of faces provide a static view and are limited ipso facto; on the other hand, videos of a face viewing an image do provide dynamic information but are computationally challenging to analyze and difficult to extract information pertaining solely to the emotional content of the image, mainly because of the overwhelming richness and diversity of information, e.g., facial features and expressions across the population of human viewers. Second, facial expression can be masked or overridden by conscious volition of the viewer, i.e., the viewer can decide to put up a “poker face”, which will reduce the capability of said approach. Alternatively, physiological monitoring of the individual, i.e. heart rate, blood pressure, heart rate variability, skin conductance response etc. can be used to measure emotional response, and these measures have also achieved a reasonable degree of success. These measures, with the exception of heart rate which is unidimensional and cannot code for the richness of emotion anyway, evolve over the period of several minutes and are slow, and furthermore, have low signal to noise ratios and are therefore highly susceptible to noise. The objective of the present study is to test if the pattern of eye movements a viewer makes while viewing an image with known emotional content along the dimensions of arousal and valence contains information embedded within of the emotional response elicited by the individual regardless of overt response or self-awareness during or following.

Human vision, while looking at any scene has a tendency of moving from a point to another to seek more information about its cognitive and behavioral goals. Eye movements follow a pattern. They do not have a fixed pattern and can be broadly divided into two: a rapid movement called saccades and a one point focus called fixations. The long term goal of this project is to see if eye movement pattern can help in successfully estimate the emotional content of any visual stimulus being viewed.

An image can be defined by several characteristics based on their contents. Two characteristics of an image namely valence and arousal are considered for the selection of images. Valence is a term used in the field of psychology to characterize and categorize the different kinds of emotions associated to images. Images that induce a positive emotion (joy) in the minds of subjects are considered as positively valenced images and the images which evokes\ negative emotions (anger, sadness) are considered as negatively valenced images. Arousal is a state of activity of mind which acts as a degree of stimulation on the thoughts of subjects. It varies across a spectrum from low to high. In other words, images that do not evoke a high degree of emotion are classified as low arousal images and those that evoke are classified as high arousal images. Both these characteristics are defined by ratings.

The aim of the project is to efficiently estimate the emotional content of an image by looking at the eye movements of the subject who has viewed. Several parameters viz. saccade amplitude, saccade duration, saccade peak velocity, saccade average velocity, fixation duration, number of saccades, pupil size, total duration, wiener entropy, entropy, number of fixations, saccade crossings and alpha values from detrended fluctuation analysis are used as features to aid in the classification process. A detailed view upon the classifiers and results will also be discussed in this report.

As an extension of the goal, we would try to investigate on the impact of exposure time and their own perceptions of each image. This is done by performing the following three experiments:

1) Image presentation with a Variable Time Duration

2) Image presentation with a Fixed Time Duration

3) User-input ratings of images

1.2 Background

Eye movements are generally perceived to be a precursor of the thought process of a person. Eyes act as a gateway to the perception process and are linked to human cognitive processes facilitating interaction between individuals. Humans are considered to be efficient in deciphering emotions of others (Schyns et al., 2009; Smith et al., 2005; Baron-Cohen et al., 2004). Though some can suffer from a disorder or deficit in distinguishing between the types of emotions (Sasson et al., 2007), six primary expressions like joy, sadness, fear, anger, disgust ad surprise (Ekman et al., 1975) can be deciphered by people in addition to self-conscious emotions i.e., shame, embarrassment (Hejmadi et al, 2000; Keltner et al, 1997) and pride (Tracy et al, 2004).

The eye has evolved 40 times in nature (Fernald, 1997), and a striking resemblance is that all living beings which have developed visual systems are able to control their gaze with their eye movements (Land 1995; Trueue, 2001). Eye movements work in synergy between perception and cognition. Eye movements don't perceive information in a random order but correspond according to the goals, inclinations of the brain. Eye movements follow a bottom-up approach for perceptual properties i.e., an eye tries to collect information from every spot it traverses, integrates the pieces of information to form a perception. This role makes saccadic patterns an important precursor to study perception and cognition.

Eye movements are fast, quickly corrective and can easily be triggered which makes study of cognition easier as certain low level changes might not be able to trigger other motor movements. Eye movements are very sensitive to any change in the visual representations. An example of this is an experiment where participants were exposed to two similar sounding words (Liebermann et al., 1957). Whenever a sound near the separating border of the two sounds were produced, the eye movements of the participants increased showing a higher activity of perception (McMurray et al., 2003).

Eye movement patterns show a striking resemblance between offline cognitive processes like reading, learning and online perceptual tasks (Barsalau, 1999; Damasio, 1989; Kosslyn et al., 1995, Martinez, 2001, Ryle, 1949). An experiment (Ballard at al., 1995) was conducted where four talking heads each reciting a quote. The talking heads were shown in any of the four quadrants on an empty screen. The talking head would appear during a sound clip and would disappear after the sound clip ended. When the sound clip of the quotes was played in the background and subjects were asked to verify if the quote was true. It was found that the participants were twice as likely to look into the quadrant at which the talking head was present. Another experiment (Spivey et al., 2000) was conducted with the participants facing an empty white screen. An audio story was played and the eye movements of the participants were recorded. The story went like this "Imagine that you are standing across the street 40m from a building. At the bottom there is a doorman in the blue. On the 10th floor is a woman hanging her laundry. On the 29th floor two kids are standing by the fire. On the very top floor there are two people screaming. While looking at the blank screen when the story was being played, the eyes of participants responded in a way similar to what they would have done if they were present in the actual scene by moving their eyes according to the story. The results suggest that eye movements convey information about internal representations and processes in the mind and information from eye scan patterns can be usefully harnessed to get a view into the inner workings of the human mind, perhaps even below the level of awareness.

Though eye movements are not necessary for vivid imagery (Hale et al., 1999), it appears that they accompany vividness (Antrobos 1969; Brandt et al., 1997; Demarais et al., 1998; Hebb, 1968; Laeng et al., 2002; Neissar, 1967). Frequency of eye movements increases during mental imagery (Clark, 1916; Goldthwait, 1933; Perky, 1910; Stoy, 1930; Totten, 1935) while an increase in the rapidness of eye movement during sleep corresponds to vividness in the dreams (Antrobus et al., 1969; Goodenough et al., 1959; Roffwarg et al., 1962). Real-time mental activity is provided by eye movements which cannot be provided by any other mechanism. It provides a semi continuous record of how partially active representations triggering a motor response thereby giving an insight into how cognition works.

Certain trends with respect to eye movements had emerged while discriminating emotions in some of the previous studies (Ekman et al., 1975). The pupil of the eyes widened for expressions like fear and constricted while coming across emotions like joy and happiness. The pattern of eye movements for fear suggest a perceptual enhancement to one's environment and the reverse pattern is observed for disgust (Susskind et al., 2008). An important finding is the approach used by eye movements in scanning through an image. Eye movements follow two types of image namely top down and bottom top approach while viewing an image. The moment an image is presented. The eye movements try to explore completely by looking around of the image with shorter fixations followed by a later period signified by shorter saccades due to increasing familiarity of the image (Buswell, 1935; Tatler et al., 2011; Yarbus, 1967).

Gaze positions tended to be concentrated on high levels of visual salience (high contrast) within images (Parkhurst et al., 2002; DeCarlo et al., 2002). In the past few years, lots of work has been done in saccades occurred during fixation. Saccades occur intermittently during fixation ranging upto an average of 2 to 3 saccades per fixation. The size of saccades were found to be upto 15 minutes of arc. The relatively larger sizes of saccades were attributed to use of illuminated rooms instead of dark rooms (Collweign et al., 2008). Interest in the topic of saccades occurring during fixation have been fueled by findings supporting that smaller saccades denote attention shifts. In accordance to the idea, a study (Hafed et al., 2002), showed that when a moving peripheral cue was shown to a subject, smaller saccades were produced along the direction of the cue. The size of the saccades were not 25% of the highest saccade. The performance of experiments also seen to be better with smaller than with larger saccades (Engbert et al., 2003). On the other hand saccades with a larger size than the standard size of microsaccades (15 min/arc) can improve the performance of visual tasks such as counting (Kowler et al., 1977), letter recognition (Kowler et al., 1987) or visual search (Schlingensiepen et al., 1986). This pattern suggests that micro saccades are more useful during active visual tasks whereas larger saccades are useful functions to representing larger visual arrays. But the links between attention and saccades show a different character when the performance cue marks the saccadic path during a sequence of saccades. (Gersch et al., 2008) found that when a color cue marks the saccadic path, the attention could be distributed even at locations that were previously fixated.

A study based on visibility modelling (Nejamic et al., 2005) constructed an ideal searcher model for finding the most probable representative saccade across a target. For this they studied the saccadic patterns of a visual search experiment where a small grating pattern was hidden in the environment of visual noise. They found that the search patterns as well as the spatial distribution of the landing locations can be predicted by a model in which each saccade was directed at a direction where the probability of finding the target was the maximum.

Numerous studies have shown strategic deployment of attention on certain type of emotions. Upon exposure to fearful images it was noted that rhesus monkeys look at the nose whereas for expressions like yawn their focus was on mouth (Nahm et al., 1997). Human beings on the other hand have displayed a varied fixation patterns across images and have differed individually as well (Walker-Smith et al., 1977). Such a type of deployment of attention can also be a result of the traits of a person. Spider phobics usually have a slower eye movement when exposed to fearful images (Pflugshauft et al., 1936). Pessimists try to look more onto the negative aspects of an image when compared to optimists (Segerstrom, 2001). Subjects under the condition of neuroticism tend to look more into the eyes of an image (Perlman et al., 2009). All these examples reveal that there is a wide difference in the eye movement patterns with each of them driven by certain biases.

Different types of fixation patterns can be seen with respect to emotions. A study titled “Eye Movements during Emotion Recognition in faces” (Schurgin et al., 2014) was done with an aim of finding the patterns on different emotions given by faces. A distinct fixation pattern was found in the emotive images like a cluster of fixations around the lips for faces that represent joy and a cluster near the eyes for faces that represent sadness. The pattern was more evident for images that had a high emotional content and low but evident for images with a lower emotional content. An emotional or neutral stimulus plays an important role in the fixation patterns of eye movement (Carniglia et al., 2012). Both gaze duration and fixation duration were found to be longer for emotive images than neutral images thus showing that emotional and animated images tend to produce an adaptive behavioral tendency.

Deep learning is a powerful technique which is useful in building a computational learning model to perform complex tasks by using multi-level processing layers. Following the design of a neural network model called Neocognitron (Fakushima K, 1988), which was created for visual pattern recognition, the use of deep learning algorithms have increased. Deep Learning has the ability to build high level features by working on raw unstructured data. There are several algorithms that come under deep learning viz. Deep Learning neural networks, Convolutional Deep Learning Neural Networks, Recurrent Neural Network etc. The signal input is transformed across the different layers of hidden neurons. The transformation chain of a signal from the input to output is called as Credit Assessment Path (CAP). Convolutional Neural Networks (CNN) have been commonly used in image classification applications. CNNs perform favorably when compared to MLP based algorithms in the case of face recognition (Lawrence S et al., 1997). A CNN with a deep hidden layer is considered to improve the accuracy of classification (Simonyan K et al., 2014).

CHAPTER 2 - HYPOTHESIS

All the studies listed above show that there is a distinct division in the eye movement patterns while looking at images with various states of emotions. Our experiments would try to extend the study of eye movement patterns by trying to estimate the emotional content of an image by looking at the eye movements of a person who sees that image. The data acquisition and analyses would be done based on the hypothesis that eye movements contain sufficient information with which an emotional gist of an image can be predicted with a high accuracy.

CHAPTER 3 - METHODS

3.1 Participants

Thirty two undergraduates from the Psychology department of the University of Houston (15 females, age: 23.5 years ± 2.9 years) having normal or corrected to normal vision participated in the study. They were asked to read the terms and conditions with regard to the details of experiment after which their signatures were taken as a consent of approval for participation.

3.2 General Apparatus

The visual stimuli were presented on a 21″ View Sonic Graphic Series 225f monitor, which was set inside a dark room. The monitor had an image aspect ratio of 16:10 and was connected to a Windows PC. This display PC was connected to the Host PC used for modulating and generating the data obtained from a subject's eyes. The images that were displayed had a resolution of 1024x768 pixels. A chin rest was fixed at a distance of 1 meter from the screen. A head mounted eye movement tracker (Eyelink II) manufactured by SR Research was used to track the eye movements of the subject. The eye tracker had three miniature cameras of which two were used to monitor eye movements and one for head movement traction. The eye cameras recorded the eye movements at the rate of 500 Hz. The eye movement tracker was connected to a Host PC which was used for calibration, validation and for controlling parameters like pupil threshold, corneal reflection etc.

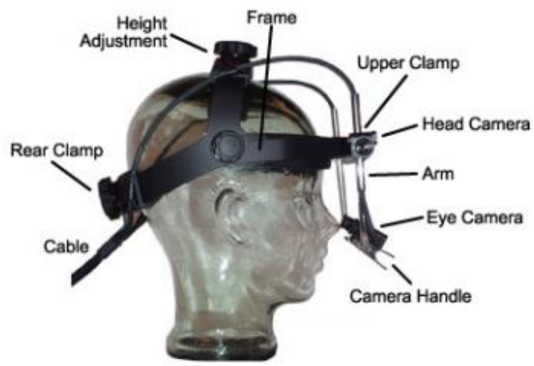

Figure 1. Parts of the Eyelink Headband
(NOTE- Source: EyelinkII User Manual)

3.3 General Procedure

The visual stimulus used were in the form of images. The images shown to the subjects were selected on the basis of arousal and valence ratings varying from low to high. So in general the images that were to be presented can be broadly classified into four categories viz. High Valence Low Arousal, High valence Low Arousal, Low Valence High Arousal and Low Valence Low Arousal. A collection of 200 images taken from IAPS collections belonging to the mentioned categories were taken in such a way that there was a clear distinction between them. The participant was asked to keep his head's position fixed by resting his/her chin on the chin rest. The Eyelink II eye tracker equipment was mounted to the subject's head. The two miniature eye cameras were adjusted in such a way that a clear image of the eye could be captured in the host PC. The next step was to fit the cameras' foci on the pupils. This was done by adjusting the image threshold color (blue). The color was adjusted in such a way that only the pupil area was blue in color. When this happens, the cameras can correctly detect the pupil which is denoted by the appearance of a green rectangular block. Once the initial settings were done, the eye movements were matched to certain data points on the computer screen. This was done by a process called calibration. During this process, a black spot appears in the middle of the screen. The subject was asked to focus on the middle of the black spot and press the 'Enter' key in the keyboard. Once the focus is proper the spot moves to different points in the screen and the subject is asked to follow the spot. Once the process of calibration is over, it has to be validated. During validation, the spots are again presented at the same points at which calibration was done and the subject is asked to follow it. This step is performed to make sure that the calibration was done properly.

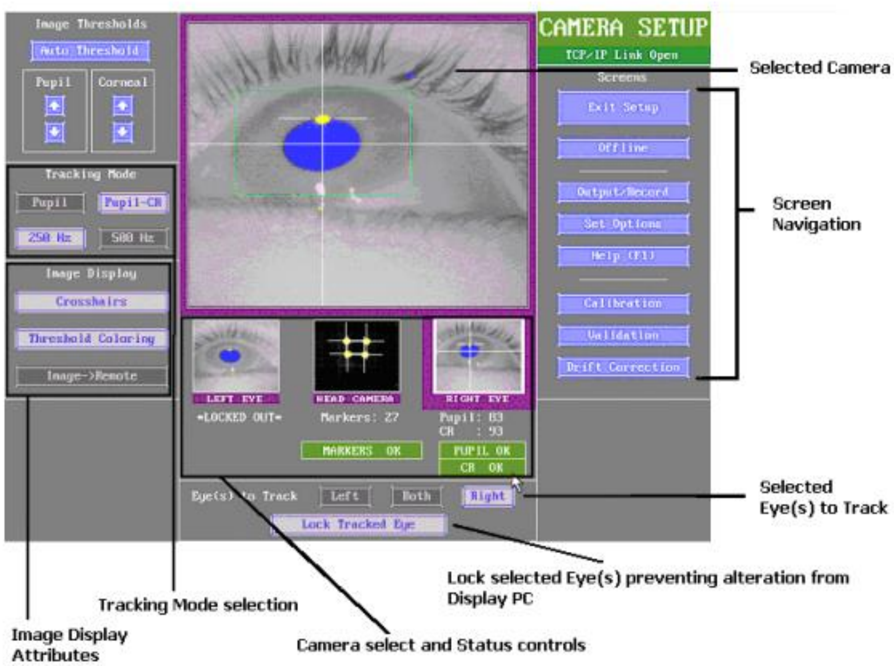

Figure 2. Eyelink II Setup screen
(NOTE- Source: Setup Screen of EyelinkII)

The experiment dataset consisted of 200 images shuffled in a random order. Before each image was presented to the subject, a drift correction was performed. The drift correction process eliminates the minor changes in the head position of the subject. After this process the image was presented to the subject. The task of the observer was to look at whatever interests him in the images.

3.4 Dataset selection

The dataset consists of 200 images selected from the International Affective Picture System (IAPS). 50 images of each class namely i.e., High Valence Low Arousal (HVLA), High Valence High Arousal (HVHA), Low Valence Low Arousal (LVLA), Low Valence High Arousal (LVHA). The distribution of the dataset based on their valence and arousal ratings is given by Figure 3. Note that the differential grouping of the data denotes to the representation of four different classes.

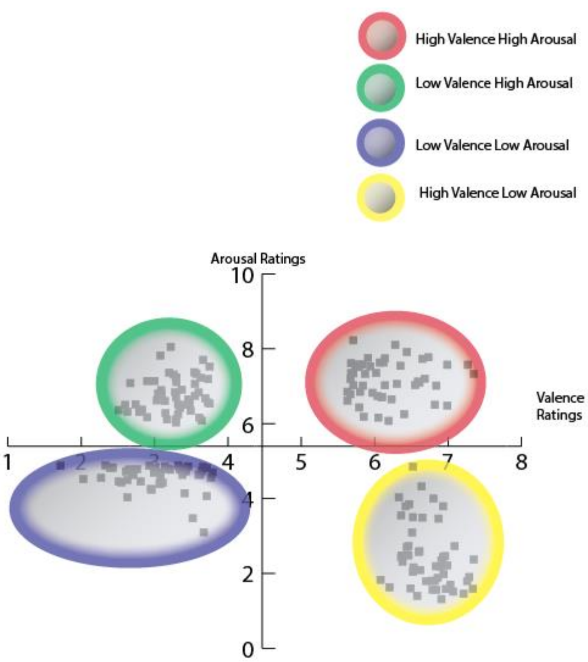

Figure 3. Graph representing the distribution of images in terms of their Arousal and Valence ratings, and divisions into different classes depending on their combination

CHAPTER 4 – DATA COLLECTION

4.1 Experiment 1: Image Presentation without a Time Limit

The goal of the experiment is to find the relation between eye movement patterns and the visual content of the images by giving the choice of duration of image presentation to the subject. The experiment is set up in such a way that the image can be changed only by the subject by clicking the space bar button. The motive of giving the control to the subject is to check if there is an inherent bias to a particular category of image.

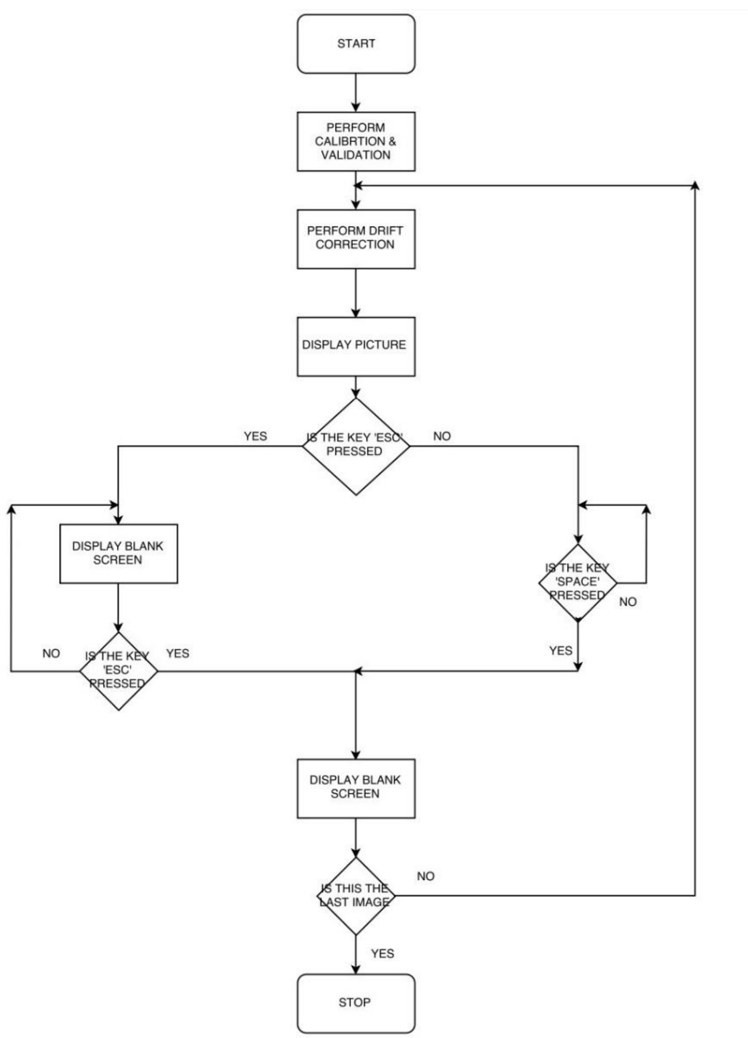

Figure 4. Sequential Representation of Experiment 1

CHAPTER 5 - FEATURE EXTRACTION

The following features were extracted from the saccades and fixations recorded on each for each subject.

1) Saccade amplitude

It denotes the angular distances traversed by the eyes while looking at the picture. The amplitudes of saccades that are created between the onset of the visual stimulus i.e., the moment at which the images appears on the screen and the withdrawal of the same. The second highest value of all the saccade amplitude values is considered to represent an image. For e.g., the representative saccade amplitude in the Figure 7 is 0.30 degree.

2) Average saccade velocity

A saccade throughout its course has a variable velocity. A single saccade moves through its path with different velocities. The term Average Saccade Velocity denotes the average of all the velocities with which a saccade traverses. The second highest value of all the average velocities is taken to represent the image. For e.g., the representative saccade amplitude in the Figure 7 is 43 deg/s.

3) Peak saccade velocity

It denotes the peak value of the velocity with which a saccade moves across a image. Generally when a stimulus is initiated, the velocity gradually rises to a peak post which it falls down. The velocity profiles have a canonical Gaussian or a bell shaped form. For e.g., the representative saccade amplitude in the Figure 7 is 122 deg/s.

4) Saccade duration

This parameter denotes the duration for which an individual saccade is existent. Several such saccades can be seen during the course of the stimulus. The second highest value of all the saccade durations is considered in order to avoid any outliers. For e.g., the representative saccade amplitude in the Figure 7 is 843 ms.

5) Fixation duration

A fixation occurs between two consecutive saccades when a person stops moving his eyes for a while and gazes at a specific spot. Fixations are generally longer in the earlier stages and decrease as time progresses. The second highest value of the all the fixation durations is considered.

6) Total duration

Total duration is the sum of a saccade duration and its successive fixation duration. It is a combination of both saccade and fixation events.

7) Total number of saccades

This feature denotes the total number of saccades created while the image is on the screen from its onset to recession. For e.g., the number of saccades in the sample figure Figure 7 is 16.

8) DFA

Detrended Fluctuation Analysis (DFA) is a technique that is used to assess the self-affinity of a signal. This technique is widely used to perform non stationary time series analysis. In our case, the input signal is the path traversed by the eye which is obtained from the gaze coordinate recordings. DFA applies a least square fit algorithm and tries to generalize a set of points to a straight line. The scaling component that is obtained from this is used as a feature. The scaling factor is calculated as the slope of the straight line fit to the log scale of the signal.

The analysis is performed by dividing the signal into a number of time scale windows Y_j (L) of length L samples. A least square straight line fit is applied to each of the single windows. The slope and y-intercept of the straight line fits are represented by 'a' and 'b' respectively.

` The minimum error across every single point inside a point is calculated by subtracting the straight line fit from the actual signal. The fluctuation over every single window is calculated by

$$F(L) = \sqrt{[\tfrac{1}{L} * \textstyle\sum_{j=1}^{L} (Y_j - ja - b)^2]}\,, \tag{1}$$

where L is the Length of the window, Y is the actual line, a is the slope of the best fit, b is the y-intercept of the best fit and F denotes the fluctuation of the actual w.r.t the best fit.

After the calculation of fluctuation across all the windows, a logarithmic graph of F(L) vs L is constructed. It is found that the relation representing the self-affinity of a signal is given by

$$F(L) \propto L^{\alpha}, \tag{2}$$

where the scaling component, α is found to be slope of straight line fit to the log-log graph of F(L) vs L. If the value of α is equal 0.5, the inference is that the pattern of eye movement is uncorrelated. If it is greater than 0.5 then a correlation exists in the eye movement value series.

9) Number of Saccade crossings

The saccades imprints that are created on the images are in random directions and there are a lot of intersections between them. The number of such intersections for a single stimulus is taken as a feature. For e.g., the number of saccades that is present in the sample figure 10 is 5.

10) Pupil size

The pupil's size keeps changing when a person keeps looking at a stimulus for a long time. The general notion is that pupil relaxes while looking at a scary stimulus and constricts while exposed to happiness inducing stimulus. The second maximum value of the pupil size is taken as a parameter.

11) Number of fixations

This feature denotes number of fixations that are created from the onset to the withdrawal of the image. There are three subdivisions under this parameter:

a) Fix 1000: Fixation duration > 1000ms

The number of fixations that were present for longer than 1000ms were considered.

b) Fix 500: Fixation duration > 500ms

The number of fixations that were present for longer than 500ms were considered.

c) Fix0-200: Fixation duration between 0 and 200ms

The number of fixations that lasted between 0 and 200ms were considered.

12) Wiener entropy

Wiener entropy is a measure of the flatness of the spectrum. The value ranges from 0 to 1. The value '1' relates to absolute flatness indicating that the power of spectrum is equal in all the spectral bands whereas '0' denotes the low degree of flatness. This measure is applied to the fixation map of each image (Fixation map shows the distribution of fixation over an image) by dividing the image into five different zones or bands. The zone separation is done by dividing the whole fixation map into 16 equal blocks as shown in Figure. 5. The 4 blocks in the middle are taken as four separate zones. The remaining blocks are combined into one single zone. The reason for considering Wiener entropy across the five zones is to see if there is an inclination to one of the four zones in the middle for various images differing valent and arousal values. The reason for combining the blocks in the outer boundaries into one is that many of the images had no fixations in at least one of the blocks in the outer side. So the plan was to combine all the fixations in the outer blocks into one so that a lack of fixation in one of the middle blocks can be easily detected. The Wiener entropy is measured on the number of fixations that land in each of these five zones. Wiener entropy for the whole spectrum is calculated by

$$W = \frac{\sqrt[N]{\prod_{n=0}^{n=N-1} x(n)}}{\frac{\sum_{n=0}^{n=N-1} x(n)}{N}} \quad , \tag{3}$$

where W is the Wiener coefficient, x(n) is the number of fixations in each zone n (1-5) and N=5.

The value of W approaching 0, represents a presence of most of the durations in selected bins across a spectrum. Similarly if the value of W is anywhere near to 1, the deduction is that the durations are spread across all the bins of the spectrum.

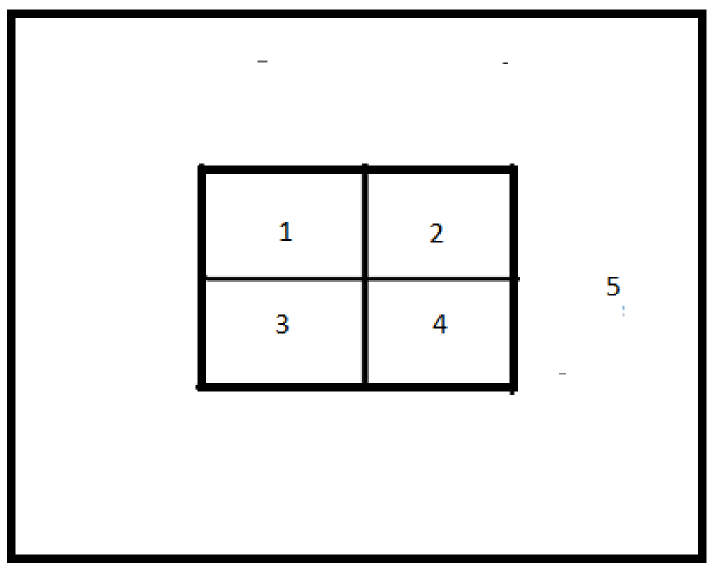

Figure 5. Zonal division of an image for calculating the wiener entropy of fixations

13) Entropy

Entropy is statistical measure of randomness in the system. The value ranges between 0 and $\log_2 n$, where n is the number of bins in the histogram. In the present case, there are n=16 bins and therefore, the maximum value is 4. As the value of entropy increases from 0 to 4, the randomness increases. The fixation map is divided into sixteen equal blocks as shown in Figure. 6. The measure is applied on the number of fixations in all the bins. The reason for applying entropy on all the fixations is to check the trend of eye movement across all parts of the image which is not found by the previously calculated Wiener entropy. Entropy is calculated by

$$E = -\sum_{n=0}^{n=N-1} x(n) * \log_2 x(n) \quad , \tag{4}$$

where E is Entropy coefficient, x (n) is the number of fixations in the zones n (1-16) and N=16.

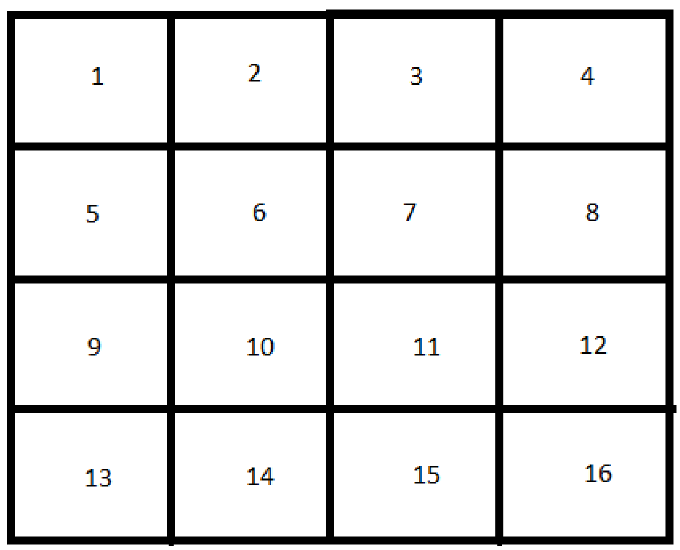

Figure 6. Zonal division of an image for calculating the Entropy of number of fixations

The value of E when equal to 4 represents an equally distributed spectrum whereas 0 represents a concentration of fixation in a single bin.

CHAPTER 6: OUTPUT

The Figure. 7 shows a sample output of an example image seen by a subject. The picture shows the pattern of eye movements of the subject as a series of saccades (white lines) interspersed by stable fixations (shown as circles of varying diameter as a function of fixation duration). The 16 block divisions that can be seen are inserted to calculate the certain interest area based features i.e., Wiener Entropy, Entropy etc.

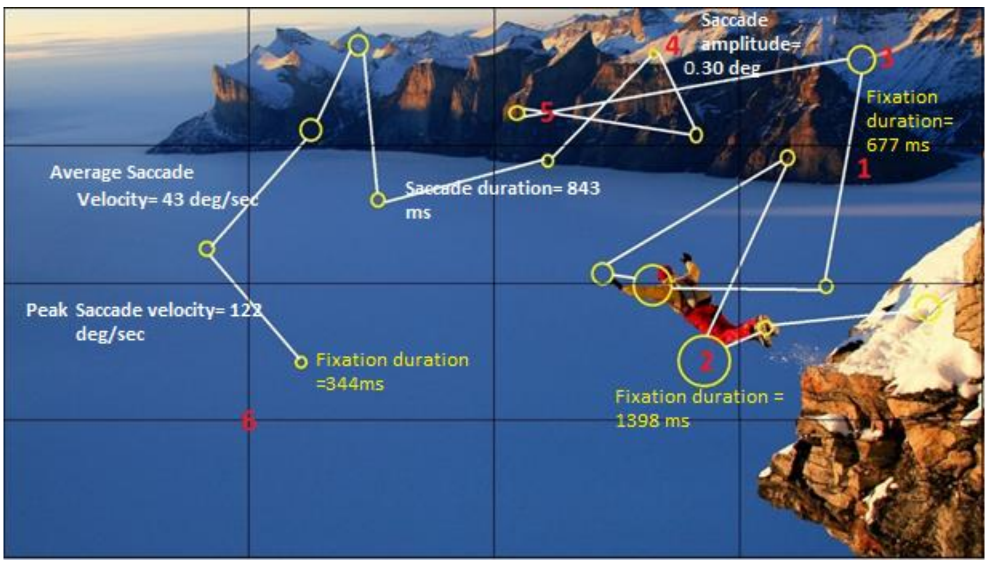

Figure 7. Sample picture showing the eye movement pattern of a subject

Table 1. The relation between the markings in Fig. 7 and the features they represent

Number on the image	Structure represented	Features that can be extracted
1	Saccade	Saccade Amplitude, Peak Saccade Velocity, Average Saccade Velocity, Saccade duration, Number of Saccades, Total duration, DFA
2	Fixation	Number of Fixations with a duration >1000 ms, Fixation duration, Total duration
3	Fixation	Number of Fixations with a duration >500 ms, Fixation duration, Total duration
4	Fixation	Number of Fixations with a duration between 0 and 200 ms, Fixation duration, Total duration
5	Saccade intersection	Number of Saccade intersections
6	Interest area represented by 16 blocks	Wiener entropy, Entropy
-	Not represented in the image	Pupil size

CHAPTER 7 - ANALYSIS

7.1 Classifier

The pattern recognition toolbox in MATLAB was used to program and design the classifier. The complete dataset contains a set of 83200 data points (200 images X 32 participants X 13 features). In order to classify this complex dataset a Random Forest Algorithm (Treebagger) was used. The optimal number of trees to be used was calculated by finding the out of bag classification error for different number of tree.

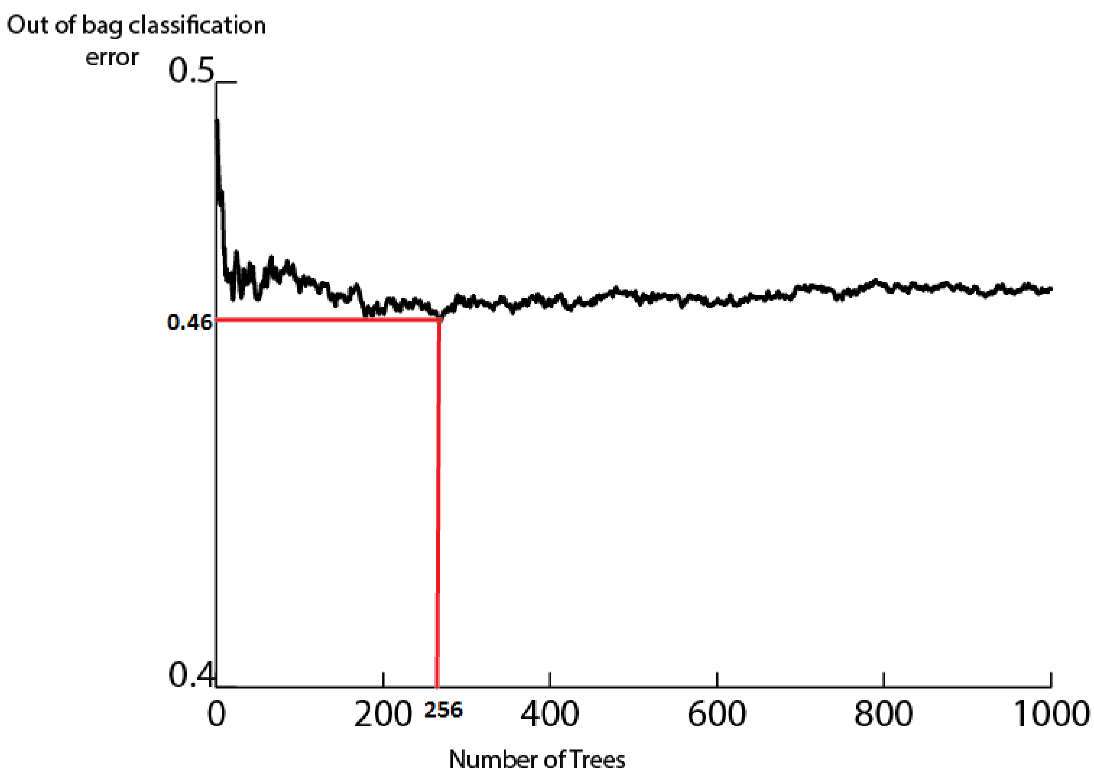

Figure 8. Graph representing the out of bag classification error for different number of trees

As it is evident from Figure. 8, the error reaches a minimum value (0.46) at N=443 where N represents the Number of trees. The number of predictors to be picked at random is equal to the square root of the total number of variables i.e., $\sqrt{13}$ = 3.61. Hence the possible number of trees for considering all the combinations is equal to $^{13}C_4 = 715$. To be on a safer side the number of trees is arbitrarily considered as 1000. A ten fold cross validation is performed on the dataset and an average of the ten classification efficiencies is calculated.

CHAPTER 8: PRELIMINARY RESULTS

The results of experiment 1 obtained by applying the random forest algorithm with the motive of classifying images based on their valence and arousal can be seen in the below table.

Table 2. Efficiency of various parameters considered individually

Feature number	Features	Arousal	Valence
1-7	7 saccade & fixation parameters	54	51
8	DFA	55.6	57
9	Saccade crossing	53.5	51
10	Pupil Size	52	51
11a	Fixation duration > 1000	50.3	50.5
11b	Fixation duration > 500	51	51.4
11c	Fixation between (0-200) ms	52	50.6
11a + 11b +11c	Combination of 3 fixation duration thresholds	50.5	51.3
12	Wiener Entropy	52	50.4
13	Entropy	53.6	52.7
12+13	Wiener Entropy+ Entropy	56	52.2
	Total efficiency	57	56

the total efficiency obtained by taking all the 13 parameters were found to be 57% for low arousal vs high arousal image classification and 56% for low valence vs high valence image classification. The efficiencies obtained by considering individual and certain combinations can be seen from figure 9 and figure 10.

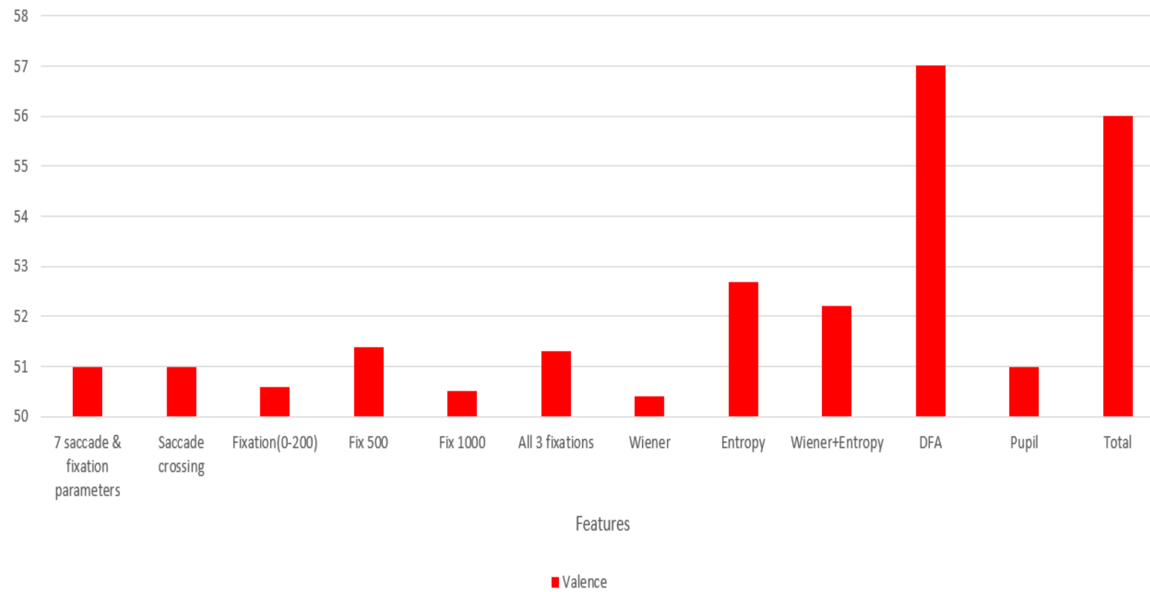

Figure 9. Classification efficiencies of various combination of parameters based on valence

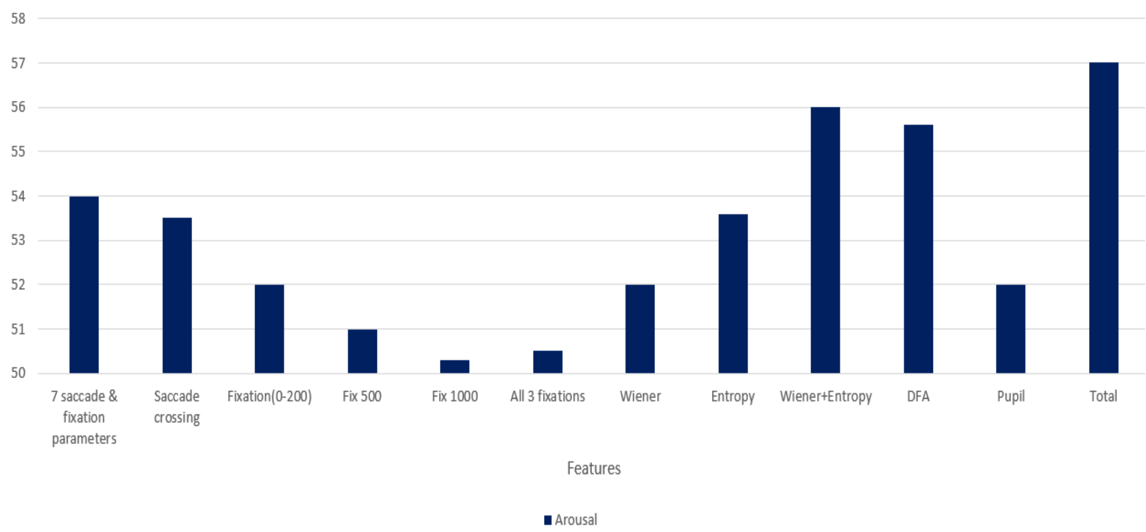

Figure 10. Classification efficiencies of various combination of parameters based on arousal

8.1 Discussion

The two critical observations obtained from the results were

1) The classification efficiencies obtained from experiment 1 is just slightly better than chance

2) Classification performance of some of the features i.e., DFA, Wiener entropy, entropy are comparable with the overall result.

CHAPTER 9: IMPROVISATIONS

9.1 Selective Structures

9.1.1 Approach

When an observer is given a chance to look at an image without any restriction, the first few seconds would be the time during which he/she would try to look at anything which interests him/her. Hence there is a very high chance that the structures i.e., saccades and fixations formed during the initial seconds of the viewing time might hold a lot of information than the rest. In order to effectively utilize that information, only the first five saccades and fixations where considered and parameters where extracted from them. The features thus extracted where used for training a Random Forest Classifier which was later validated using a 10-fold cross validation technique.

9.1.2 Result and Inference

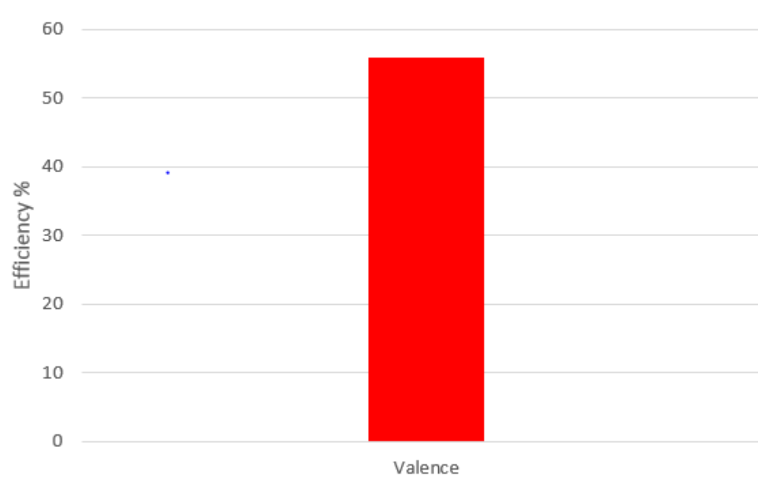

Figure 11. Classification efficiencies using Selective Saccades based on Valence

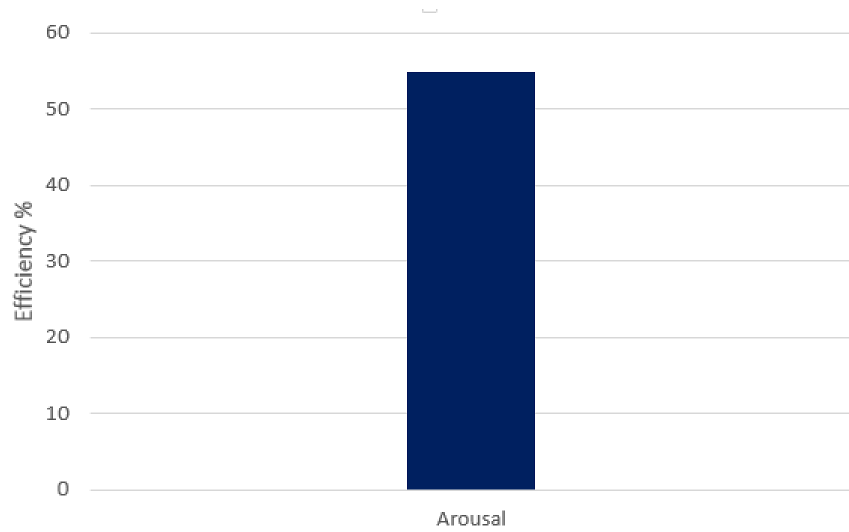

Figure 12. Classification efficiencies using Selective Saccades based on Valence

Figures 11 and Figure 12 show the result of using the testing set on the Classifier. There was no improvement seen as the classification efficiencies were still in the lower 50s. An assumption to the below average classification efficiency could be the fact that parameters used were all static.

9.2 Dynamic features

9.2.1 Features

All the features that were discussed previously are static. Static variable at times provide limited information whereas dynamic features tend to provide a deeper insight as they provide a continuous representation over time domain. Figure. 13 is an example of a dynamic feature representing the variation of Saccade amplitude over successive saccades.

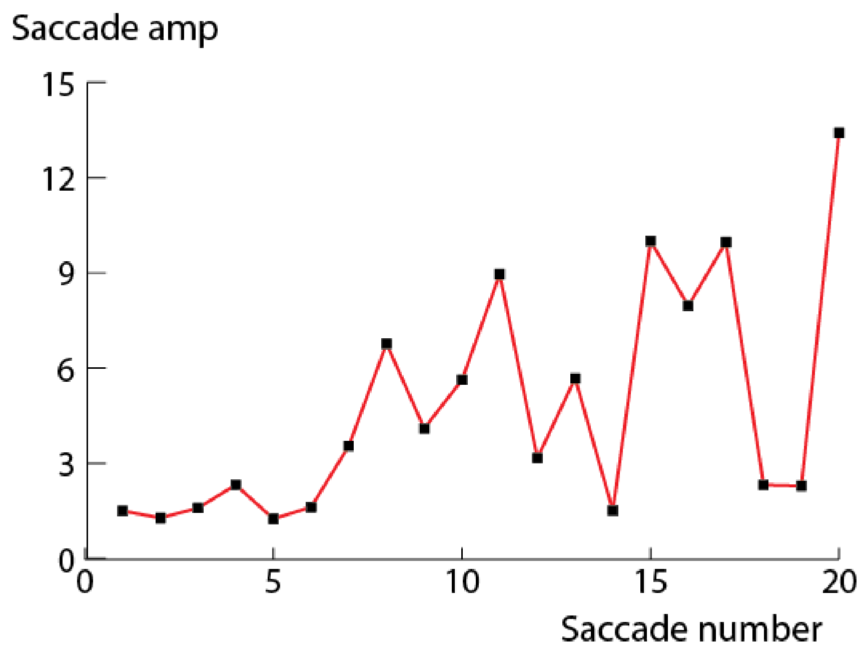

Figure 13. Distribution of Saccade amplitude over successive saccades

Two kinds of dynamic features were considered for the design.

a) Saccade amplitude vs Saccade number

b) Fixation duration vs Fixation number

After the construction of the dynamic features, the following measures were calculated and used as predictors for classification.

1) Number of peaks

Upon designing the distribution of the fixation durations and saccade amplitudes over time, an initial analysis over a set a graphs showed slight variations in the variability of the graph. Hence the number of peaks for every single image viewed by each subject was calculated and used one of the predictors.

2) FFT

The signal formed in the time domain is transformed into the frequency domain using a Fast Fourier Transform (FFT) algorithm. The number of samples taken is extended to 256 for all output data from the images taken across all the subjects. The first set of 25 samples from the output are taken in the form of a predictor vector as the rest of the samples do not possess much meaningful data. The output obtained can be calculated from

$$X(k) = \sum_{n=0}^{N-1} x\,(n) * e^{\frac{-2*\pi*i*k*n}{N}} \quad , \tag{5}$$

where X(k) is the output signal in the frequency domain, x(n) is the input signal, N represents the number of samples.

3) Normalized power

Another important characteristic pertaining to a signal is the power of the signal. The quantity is given by

$$P = \frac{1}{N} * \sum_{t=1}^{N} x^2(t) \ , \tag{6}$$

where P is normalized power, N is the number of samples, x is the input signal. The normalized Power of the signals representing each image is taken as the third predictor.

9.2.2. Result and Inference

The dynamic features that were calculated upon the saccades' and fixations' variations over the time domain. The classification efficiency was slightly above chance as shown in Figure 14 and Figure 15. Thus selecting dynamic features did not bring a considerable improvement.

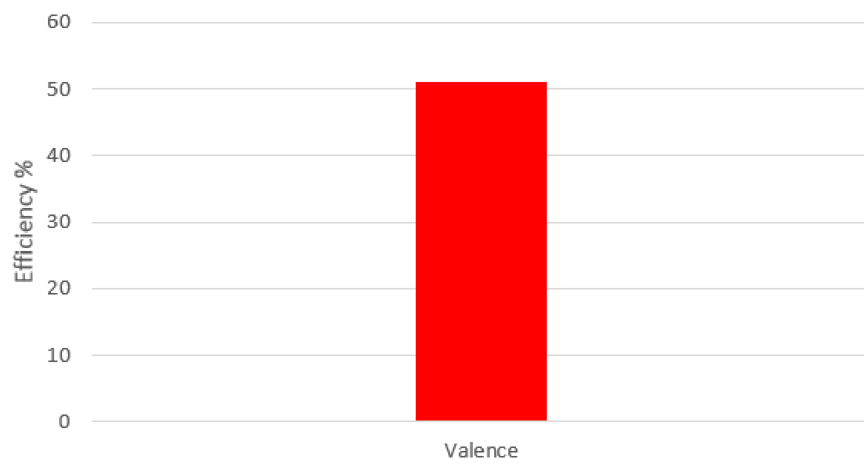

Figure 14. Classification Efficiency using Dynamic Features based on Valence

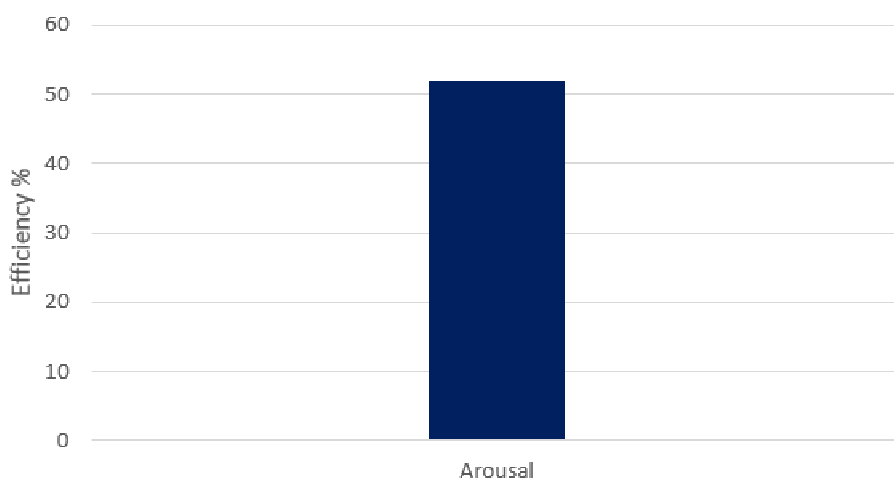

Figure 15. Classification Efficiency using Dynamic Features based on Arousal

9.3 Multiple image set

9.3.1 Design

In order to check if the ratings of the images have a connection with the classification efficiencies of the classifier, the images were further divided based on their rating levels. Three individual divisions were done. In the first division, 10 images having the top ten highest ratings (was done separately for valence and arousal ratings) and 10 images having ten lowest ratings were considered and were used for classification. The image set was later increased to 40 with images belonging to the top 20 extreme ratings on both the ends of the scale and analyzed. Another division with 50 extremum images was later considered. The classification efficiency for each of the division is shown in Figure 14 and Figure 15.

9.3.2 Result and Inference

The gradual decrease in the peaks of the graphs seen in Figure 16 and Figure 17 as the number of images of images grows from the extreme ends of the scale to middle range indicates that the variation of classification efficiency could be linked to the inherent content of the images. In other words, the classifier performed better when there was a distinct difference in the rating levels of the images.

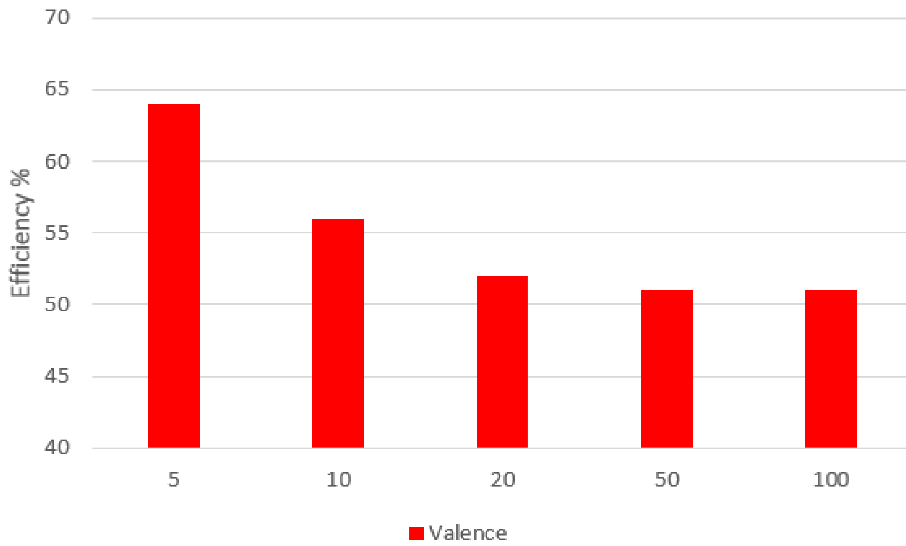

Figure 16. Classification Efficiencies over a different set of images based on Valence

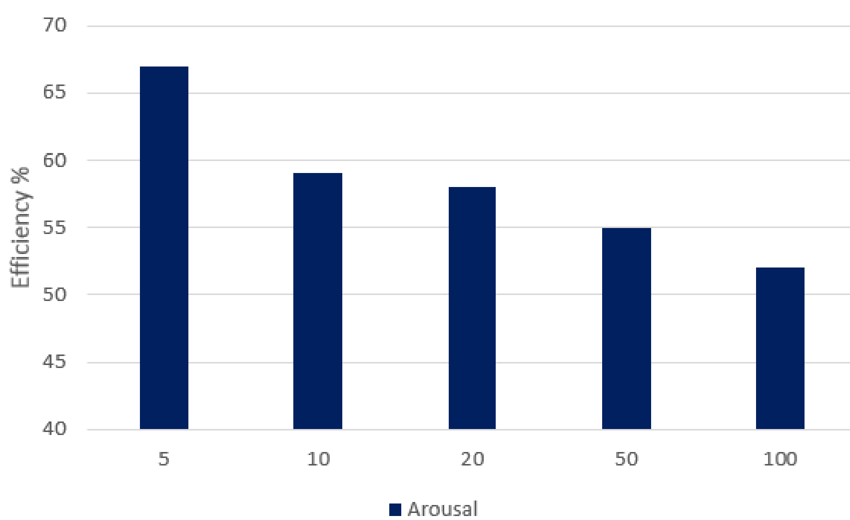

Figure 17. Classification Efficiencies over a different set of images based on Arousal

9.4 Deep Learning

All the previous methods used for classifying images based on their emotional content proved to be unfruitful. But one takeaway was that there was significant increase in efficiency when images with a high difference in ratings were considered. This shows that there could be some information hidden in the pictures itself. In order to find the information, the raw images without the saccade and fixation markings had to be analyzed. Though Random forest is a powerful classifier, an algorithm that can perform a high level data abstraction on the feature data has a possibility to come up with a better classification model to estimate the emotional content of the visual images on all the experiments. For this purpose a deep learning algorithm would be used.

9.4.1 Convolutional Neural Network

Convolutional Neural Network has shown very high accuracies in several previous cases of image classification when it comes to consist of multiple processing layers .In order to perform these tasks, the Theano library of the software language PYTHON would be used. The library is used in Python for performing numerical computations. The use of a GPU would be considered later depending on the complexity of the task which would later be analyzed during implementation.

9.4.1.1 Design

For the design of CNN shown in Figure. 12, the Keras library in Python with a Theano background was used. The model was implemented using an Intel Core i5 processor running at a frequency of 1.70GHz. A GPU was not required for the implementation as the number of input images were small.

All the images in the dataset were considered for training the model and a 10 fold cross validation was performed to obtain the classification efficiency of the model. The model that was designed consisted of 7 stages as shown in Figure 18. The images were imported into the system and were converted into 3-dimensional arrays consisting of RGB values. A 3x3 convolution was first performed and the output was matched to a feature map. The output was then inputted to a pooling layer. The purpose of having a pooling layer in between was to reduce the spatial size and computational complexity. This procedure was followed two more times post which the data was inserted into a model to be trained.

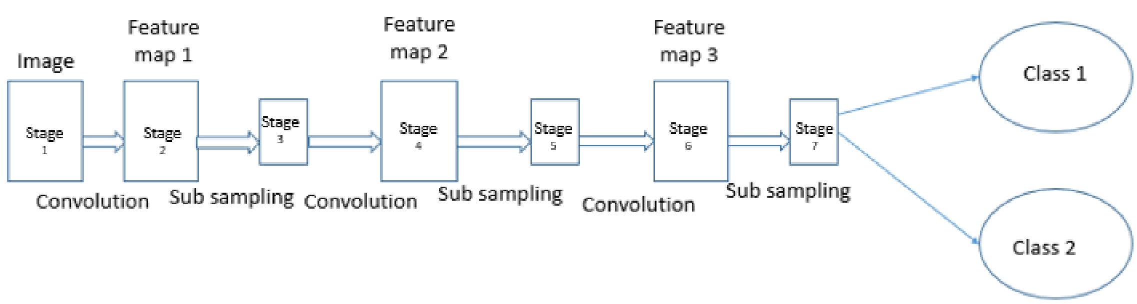

Figure 18. Block diagram of Convolutional Neural Network

A 10 cross validation was performed so that all the datapoints were used both for training and testing. The results obtained from the CNN, presented in Figure 19 and Figure 20 show that the images were classified at an efficiency if 85% and 75% based on valence and arousal ratings on the training set and at an efficiency of 75% and 66% on the testing set. The higher efficiency percentage on the raw images without the structures produced by eye movements suggest that could be inherent information present in the images defined by definitive ratings.

Figure 19. Classification Efficiencies using Convolutional Neural Network based on Valence

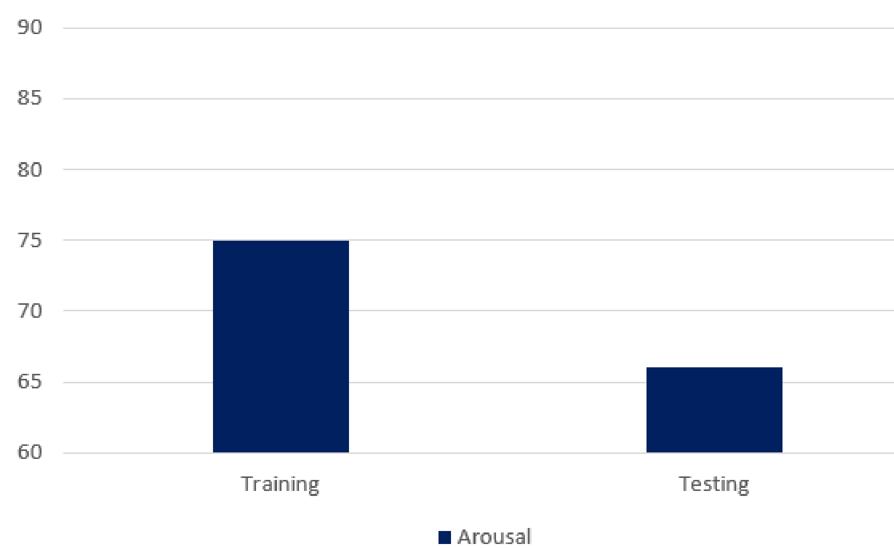

Figure 20. Classification Efficiencies using Convolutional Neural Network based on Arousal

CHAPTER 10: CONCLUSION

Our study found that the eye scan patterns of our sample performed slightly above chance in the task of predicting the emotional gist of an image, which was indicative of poor classifier performance. This leads to two possibilities: either we have disproved that there is not sufficient information in eye scan patterns to predict the emotional gist of an image or that we have not provided enough information to warrant any conclusion. One possibility is that the images were chosen from the IAPS database with ratings of arousal and valence provided. However, these ratings could vary from individual to individual and obtaining ratings from our sample as opposed to relying on externally provided ratings may prove beneficial. Our experimental sessions are an hour to an hour and a half long and fatigue could have been a factor leading to improper or poor viewing of the images. One possible change to reduce the effect of fatigue would be to fix the time duration of the image. Combined, both changes could help obtain more reliable data and consequently, better classifier performance.

Another confound in our experiments could be the fact that valence and arousal may not be independent of each other. For example, an image with low arousal might not be appealing to a person so his/her approach would not differ if it is an image with low or high valence. Hence, segregation of images prior to presentation is another aspect to be taken care of. Choosing images on the extremes of the rating scale could prove fruitful in reference to the trend shown in the study where the classification efficiency started decreasing after more number of images were added in the decreasing distance from the middle of the rating scale. This would also make the objective modest wherein eye movements on images with a very high difference in ratings can be worked upon instead of going through images with slight nuances. Moreover, it would be a good idea to confine all our studies on images with high arousal ratings.

Finally, it may be that each image is sufficiently different from all others that it would be hard to pick out information about the emotional content of an image. For example, two images may be scary and have identical valence ratings but for different reasons, i.e. different items (e.g. gun or a bloody scene of a bomb attack) or even the same item but at different positions in the different images; either of those changes would change the pattern of eye movements and fixations. Moreover, these differences would be further enhanced when taking into account differences in eye scan patterns across individuals or even the same individual looking at the same image multiple times. These changes in the pattern could obscure any similarities in the emotional content of the image. In summary, it may be the case that despite the improvements proposed, there may not be enough information contained in an image to be able to reliably pull out information about its emotional content. Future experiments will help us resolve this issue.

CHAPTER 11: FUTURE WORK

Our future plan of action is to focus on improving the test performances of our classifiers. There are several ways of approaching this.

1) Modifying experiment 1 by fixing the exposure duration of an image. The reason for this modification is to restrict the inter subject variability in viewing a particular image.

2) The other reason for getting an average classification efficiency can be due to the fact that some of the observers can have an idiosyncratic reaction to a given image and thus his arousal and valence rating of the image might differ from that of the ones specified in the IAPS database. So instead of arbitrarily defining arousal and valence values for an image, we plan to get the inputs for these values from the subjects. This can help us in improving the classification efficiency.

11.1 Experiment 2: Image Presentation with Time Limit

It is possible that the viewer becomes disinterested in Experiment 1. From Figure and as the experiment progressed, (s)he spent progressively less time on each image and did not process each image as thoroughly.

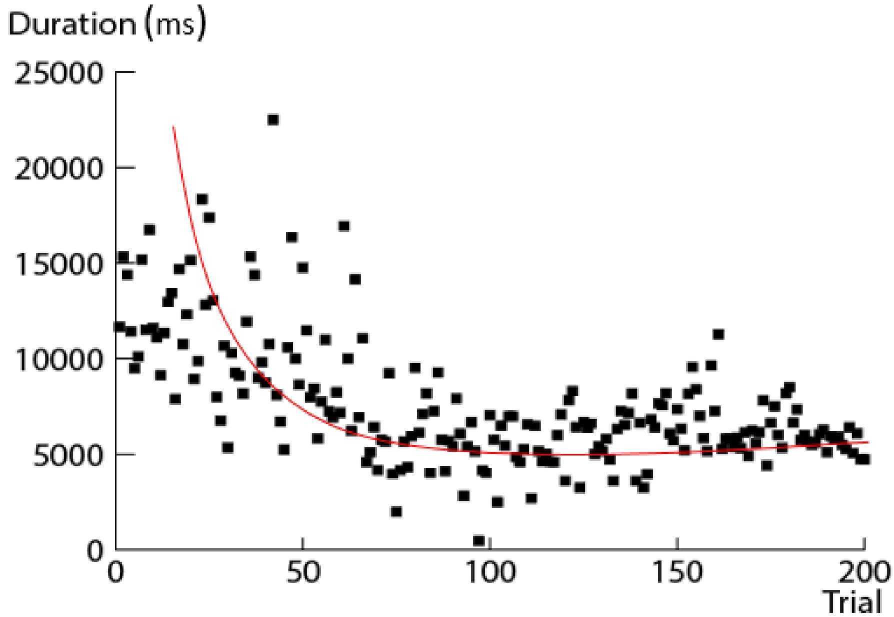

Figure 21. A subject's view time duration of each trial

Our results (Figure 21) show a decrease in viewing duration on the experiment pyramid. For this reason we decided to pre determine the amount of time that the image remains on the screen. The image duration for the new experiment is determined by taking the median of the means of viewing time of each subject who participated in experiment 1. The flow of experiment 2 is given by Figure 22.

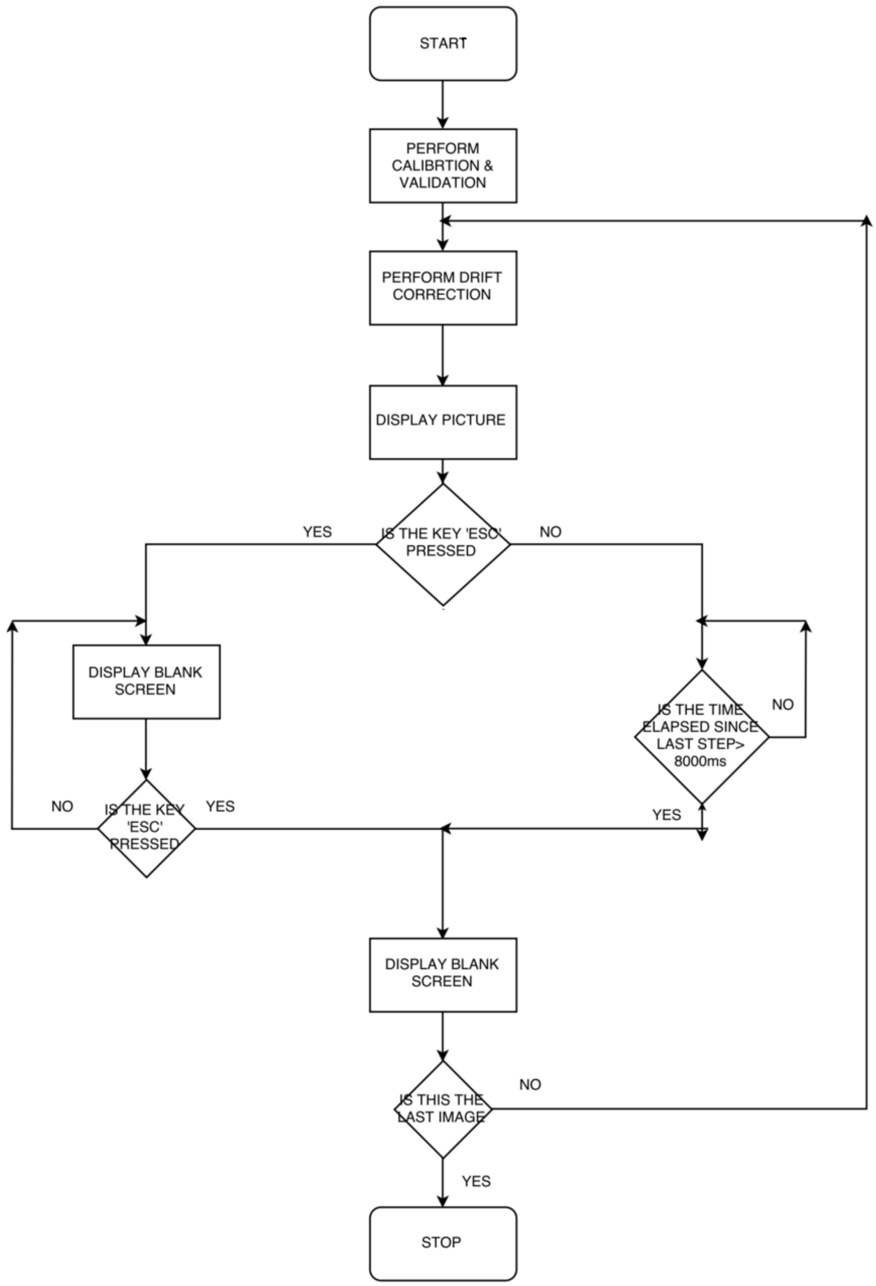

Figure 22. Sequential representation of Experiment 2

11.2 Experiment 3: User-Input Ratings from Subjects:

In experiment 1, the arousal and valence ratings were provided for each image in the IAPS database. A number of undergraduates were run and their ratings were averaged to obtain the representative ratings for each image. Our observations is that some images could have a different effect on different individuals. In experiment 3, we will be obtaining Arousal/Valence ratings for each image and use their ratings for the classification. Image presentation time will be either constant (as in Experiment 2) or be viewer determined (as in Experiment 1) depending on which yields a better overall classification efficiency.

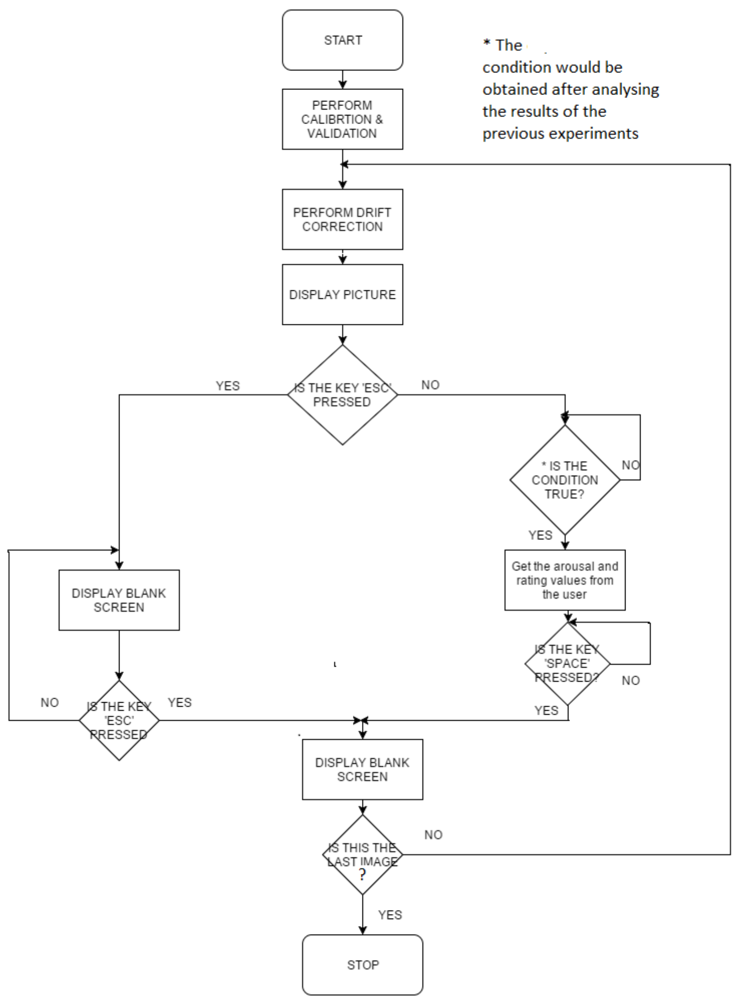

Figure 23. Sequential Representation of Experiment 3

REFERENCES

Antrobus, J., Antrobus, S. (1969). “Rapid Eye Movements and Rapid Eye Movement Periods”. *Psychophysiology*, 6(1), pp.45-48

Ballard, P., Hayhoe, M.M, Pelz, J.B (1995). “Memory Representations in Natural Tasks”. *Cognitive Neuroscience*, 7(1):66–80

Barsalou, L.W (1999). “Perceptual Symbol Systems”. *Behavioural Brain Science*, 22(4):577-609

Brandt, A., Stark, W. (1997). “Spontaneous Eye Movements during Visual Imagery Reflect the Content of the Visual Scene”. *J Cogn Neurosci*, 9(1):27-38

Buswell, G. (1935). “After Science: What?”. *The Phi Delta Kappan*,17(4), pp.114-115

Collewijn, K. (2008). “The Significance of Microsaccades for Vision and Oculomotor Control”. *Journal of Vision*, 8,20.

Clark, J.J (1999). “Spatial Attention and Latencies of Saccadic Eye Movements”. *Vision Res*. 39(3):585-602.

Dacher, K., Brenda, N., Buswell, V. (1997). “Embarrassment: Its Distinct Form and Appeasement Functions”. *Psychological Bulletin*, W. 122. No, 3, 250-270

Decarlo, D., Santella, A. (2002). “Stylization and Abstraction of Photographs”. *ACM Transactions on Graphics*, Vol.21(3)

Demarais, M., Cohen, H. (1998). “Evidence for Image-Scanning Eye Movements during Transitive Inference”. *Biological Psychology*, 49(3), pp.229-247

Engbert, R., Kliegl, R. (2003). “Microsaccades Uncover the Orientation of Covert Attention”. *Vision Research*, 43 (9), pp. 1035–1045

Fernald, R.D (1997). “The Evolution of Eyes”. *Brain Behav Evol*, 50(4):253-9.

Fukushima, K.(1980). “Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position.” *Biol. Cybernetics* 36, 193 202

Gersch, T.M, Kowler, E., Schnitzer, B.S., Dosher, B.A. (2008). “Visual Memory during Pauses between Successive Saccades.” *J Vis*, 8(16):15.1-18.

Goldwait, C. (1933). “Relation of Eye-Movements to Visual Imagery.” *The American Journal of Psychology*, 45(1), pp.106-110

Goodenough, D., Shapiro, A., Holden, M., Steinschriber, L., Brewster, L. (1959). “A Comparison of "Dreamers" and "Nondreamers": Eye Movements, Electroencephalograms, and the Recall of Dreams.” *The Journal of Abnormal and Social Psychology*, 59(3), pp.295-302

Hafed, M., Clark, J.J. (2002). “Microsaccades as an Overt Measure of Covert Attention Shifts.” *Vision Research*, 42 (22), pp. 2533–2545

Hale, S.M., Simpson, H.M. (1971). “Effects of Eye Movements on the rate of Discovery and the Vividness of Visual Images.” *Perception & Psychophysics*, Issue 2, pp 242-246

Hebb, D.O. (1968). “Concerning imagery.” *Psychological Review*, 75(6), pp.466-477

Hejmadi, A., Davidson, R., & Rozin, P. (2000). “Exploring Hindu Indian Emotion Expressions: Evidence for Accurate Recognition by Americans and Indians.” *Psychological Science*, 11, 183-187.

Kosslyn, S.M., Behrmann, M., Jeannerod, M. (1995). “The Cognitive Neuroscience of Mental Imagery.” *Neuropsychologia*, 33(11):1335-44

Kowler, E., Anton, V. (1987). “Reading Twisted Text: Implications for the Role of Saccades.” *Vision Research*, 27(1), pp.45-60

Kowler, E., Steinman, R. (1977). “The Role of Small Saccades in Counting.” *Vision Research*, 17 (1) pp. 141–146

Laeng, B., Teodorescu, D. (2002). “Eye Scanpaths during Visual Imagery Reenact those of Perception of the same Visual Scene.” *Cognitive Science*, 26(2), pp.207-231

Land, M.F. (2006). “Eye Movements and the Control of Actions in Everyday Life.” *Progress in Retinal and Eye Research*, 25: 296–324

Liberman, A.M., Harris, K.S., Hoffman, H.S., Griffith, B.C. (1957). “The Discrimination of Speech Sounds within and across Phoneme Boundaries.” *J Exp Psychol*, 54(5):358-68.

Martinez, M. (2001). “Mass Customization: Designing for Successful Learning.” *International Journal of Educational Technology*, 2 (2).

McMurray, B., Tanenhaus, M.K., Aslin, R.N., Spivey, M.J. (2003). “Probabilistic Constraint Satisfaction at the Lexical/Phonetic Interface: Evidence for Gradient

Effects of within-category VOT on Lexical Access." *J Psycholinguist Res* .Jan; 32(1):77-97.

Najemnik, J., Geisler, W.S. (2005). "Optimal Eye Movement Strategies in Visual Search." *Nature*, 434, 387-391.

Nahm, F.K., Perret, A., Amaral, D.G., Albright, T.D. (1997). "How do Monkeys look at Faces? *J Cogn Neurosci,* 9(5):611-23.

Parkhurst, D., Law, K., Niebur, E. (2002). "Modeling the role of Salience in the Allocation of Overt Visual Attention." *Vision Res*. 42(1):107-23

Perky, C. (1910). "An Experimental Study of Imagination." *The American Journal of Psychology*, 21(3), pp.422-452

Perlman, S.B., Morris, J., Brent, C., Vander, B., Green, S., Doyle, J.L., Pelphrey K.A. (2009). "Individual Differences in Personality Predict how People look at Faces." *PLoS One* 4: e5952

Pflugshaupt, T., Mosimann, P., Schmitt, W.J., Wartburg, R.V., Wurtz, P., Luthi, M., Nyffeler, T., Hess, C.W., Muri, R.M. (2007). "To Look or not to look at Threat? Scanpath Differences within a group of Spider Phobics." *J Anxiety Disord*;21(3):353-66..

Roffwarg, H., Dement, W., Muzio, J., Fisher, C. (1962). "Dream Imagery: Relationship to Rapid Eye Movements in Sleep." *Archives of General Psychiatry*, 7, pp. 235–258

Ryle, G. (1949). The Will. Chapter 3 of 'The concept of mind'. ISBN 0-226-73295-9

Schlingensiepen, K.H., Campbell, F.W., Legge, G.E. and Walker, T.D. (1986). "The Importance of Eye Movements in the Analysis of Simple Patterns." *Vision Research*, 26 (7), pp. 1111–1117

Schurgin, M.W., Nelson, J., Iida, S., Ohira, H., Chiao, J.Y., Franconeri, S.L. (2014). "Eye Movements during Emotion Recognition in Faces." *J Vis.Nov* 18;14(13):14.

Segerstrom, S.C. (2001). "Optimism, Goal Conflict, and Stressor-related Immune Change." *J Behav Med*, 24(5):441-67

Stoy, E. (1930). "A Preliminary Study of Ocular Attitudes in Thinking of Spatial Relations." *The Journal of General Psychology*, 4(1-4), p.379-385

Susskind, M. (2008). "Expressing Fear Enhances Sensory Acquisition." *Nature Neuroscience* 11, 843 - 850

Tatler, W., Hayhoe, M., Land, F., Ballard, H. (2011). "Eye Guidance in Natural Vision: Reinterpreting Salience." *Journal of vision*,11(5), pp.5

Totten, E. (1935). "Eye Movement during Visual Imagery." *Comp Psychol Monogr*;11:46

Tracy, J.L., Robins T., Richard, W. (2004). "Show Your Pride: Evidence for a Discrete Emotion Expression." *Psychological Science*, 15(3), pp.194-197

Treue, S. (2001). "Neural correlates of Attention in Primate Visual Cortex." *Trends Neurosci*, 24(5):295-300.

Walker, S., Gail, D., Alaistar, C., Findlay, S. (1977). "Eye Movement Strategies Involved in Face Perception. Perception." 42(11), pp.1120-1133

Yarbuz, A.L. (1965). “Eye Movements And Vision.” Institute of Information Transmission of the Academy of Sciences of the USSR by Nauka Press. Moscow.

APPENDIX

Table 3. Valence and Arousal Ratings of the images from IAPS database

Image	Valence	Arousal
1050.jpg	3.46	6.87
1052.jpg	3.5	6.52
1114.jpg	4.03	6.33
1120.jpg	3.79	6.93
1201.jpg	3.55	6.36
1321.jpg	4.32	6.64
1333.jpg	6.11	3.17
1419.jpg	6.54	3.48
1450.jpg	6.37	2.83
1525.jpg	3.09	6.51
1602.jpg	6.5	3.43
1603.jpg	6.9	3.37
1604.jpg	7.11	3.3
1610.jpg	7.82	3.08
1620.jpg	7.37	3.54
1650.jpg	6.65	6.23
1670.jpg	6.81	3.05
1812.jpg	6.83	3.6
1900.jpg	6.65	3.46
1910.jpg	6.71	3.29
1930.jpg	3.79	6.42
1931.jpg	4	6.8
1932.jpg	3.85	6.47
2000.jpg	6.51	3.32
2010.jpg	6.25	3.32
2037.jpg	6.42	3.35
2104.jpg	4.42	3.11
2190.jpg	4.83	2.41
2200.jpg	4.79	3.18
2206.jpg	4.06	3.71
2208.jpg	7.35	5.68

Image	Valence	Arousal
2270.jpg	6.28	3.15
2304.jpg	7.22	3.63
2320.jpg	6.17	2.9
2358.jpg	6.56	3.73
2360.jpg	7.7	3.66
2370.jpg	7.14	2.9
2383.jpg	4.72	3.41
2389.jpg	6.61	5.63
2393.jpg	4.87	2.93
2440.jpg	4.49	2.63
2441.jpg	4.64	3.62
2446.jpg	4.7	3.79
2480.jpg	4.77	2.66
2491.jpg	4.14	3.41
2493.jpg	4.82	3.34
2500.jpg	6.16	3.61
2501.jpg	6.89	3.09
2512.jpg	4.86	3.46
2560.jpg	6.34	3.49
2570.jpg	4.78	2.76
2598.jpg	7.19	3.73
2722.jpg	3.47	3.52
2811.jpg	2.17	6.9
2830.jpg	4.73	3.64
3000.jpg	1.59	7.34
3005.1.j	1.63	6.2
3010.jpg	1.79	7.26
3053.jpg	1.7	7.03
3068.jpg	1.31	6.91
3069.jpg	1.8	6.77
3071.jpg	1.88	6.86

Image	Valence	Arousal
3266.jp	1.56	6.79
3400.jp	2.35	6.91
3500.jp	2.21	6.99
3530.jp	1.8	6.82
5000.jp	7.08	2.67
5010.jp	7.14	3
5020.jp	6.32	2.63
5030.jp	6.51	2.74
5120.jp	4.39	3.07
5130.jp	4.45	2.51
5200.jp	7.36	3.2
5250.jp	6.08	3.64
5260.jp	7.34	5.71
5410.jp	6.11	3.29
5450.jp	7.01	5.84
5460.jp	7.33	5.87
5470.jp	7.35	6.02
5534.jp	4.84	3.14
5551.jp	7.31	3.26
5621.jp	7.57	6.99
5623.jp	7.19	5.67
5626.jp	6.71	6.1
5629.jp	7.03	6.55
5700.jp	7.61	5.68
5711.jp	6.62	3.03
5720.jp	6.31	2.79
5750.jp	6.6	3.14
5760.jp	8.05	3.22
5764.jp	6.74	3.55
5779.jp	7.33	3.57
5780.jp	7.52	3.75

Image	Valence	Arousal
5971.jpg	3.49	6.65
6230.jpg	2.37	7.35
6260.jpg	2.44	6.93
6300.jpg	2.59	6.61
6312.jpg	2.48	6.37
6313.jpg	1.98	6.94
6315.jpg	2.31	6.38
6350.jpg	1.9	7.29
6370.jpg	2.7	6.44
6510.jpg	2.46	6.96
6540.jpg	2.19	6.83
6550.jpg	2.73	7.09
6560.jpg	2.16	6.53
7006.jpg	4.88	2.33
7025.jpg	4.63	2.71
7030.jpg	4.69	2.99
7031.jpg	4.52	2.03
7036.jpg	4.88	3.32
7037.jpg	4.81	3.71
7038.jpg	4.82	3.01
7040.jpg	4.69	2.69
7060.jpg	4.43	2.55
7110.jpg	4.55	2.27
7130.jpg	4.77	3.35
7150.jpg	4.72	2.61
7175.jpg	4.87	1.72
7180.jpg	4.73	3.43
7184.jpg	4.84	3.66
7186.jpg	4.63	3.6
7217.jpg	4.82	2.43
7224.jpg	4.45	2.81

Image	Valence	Arousal
7502.jpg	7.75	5.91
7545.jpg	6.84	3.28
7595.jpg	4.55	3.77
7700.jpg	4.25	2.95
7705.jpg	4.77	2.65
7900.jpg	6.5	2.6
8021.jpg	6.79	5.67
8030.jpg	7.33	7.35
8034.jpg	7.06	6.3
8080.jpg	7.73	6.65
8090.jpg	7.02	5.71
8116.jpg	6.82	5.97
8161.jpg	6.71	6.09
8170.jpg	7.63	6.12
8178.jpg	6.5	6.82
8179.jpg	6.48	6.99
8180.jpg	7.12	6.59
8185.jpg	7.57	7.27
8186.jpg	7.01	6.84
8190.jpg	8.1	6.28
8191.jpg	6.07	6.19
8193.jpg	6.73	6.04
8200.jpg	7.54	6.35
8210.jpg	7.53	5.94
8251.jpg	6.16	6.05
8260.jpg	6.18	5.85
8300.jpg	7.02	6.14
8340.jpg	6.85	5.8
8341.jpg	6.25	6.4
8370.jpg	7.77	6.73
8380.jpg	7.56	5.74

2210.1.j	4.7	3.08
2210.jpg	4.7	3.08
2215.jpg	4.63	3.38
2216.jpg	7.57	5.83
2221.jpg	4.39	3.07
2240.jpg	6.53	3.75

3100.jpg	1.6	6.49
3102.jpg	1.4	6.58
3110.jpg	1.56	6.84
3120.jpg	1.58	6.97
3130.jpg	1.79	6.7
3170.jpg	1.46	7.21

5800.jp	6.36	2.51
5811.jp	7.23	3.3
5833.jp	8.22	5.71
5870.jp	6.78	3.1
5875.jp	6.03	3.29
5891.jp	7.22	3.29

7234.jpg	4.23	2.96
7270.jpg	7.53	5.76
7325.jpg	7.06	3.55
7340.jpg	6.68	3.69
7491.jpg	4.82	2.39
7501.jpg	6.85	5.63

8400.jpg	7.09	6.61
8470.jpg	7.74	6.14
8475.jpg	4.85	6.52
8485.jpg	2.73	6.46
8490.jpg	7.2	6.68
8496.jpg	7.58	5.79

Image	Valence	Arousal
8499.jpg	7.63	6.07
8501.jpg	7.91	6.44
8502.jpg	7.51	5.78
9001.jpg	3.1	3.67
9050.jpg	2.43	6.36
9156.jpg	6.43	5.79
9210.jpg	4.53	3.08
9250.jpg	2.57	6.6
9360.jpg	1.83	6.08
9405.jpg	4.03	2.63
9410.jpg	1.51	7.07
9600.jpg	2.48	6.46
9700.jpg	4.77	3.21
9810.jpg	2.09	6.62
9921.jpg	2.04	6.52