CARDIFF SCHOOL OF COMPUTER SCIENCE AND INFORMATICS

Saliency in Image and Vision Computing

Kenneth Cynric M. Dasalla

supervised by
Dr. Hantao Liu

moderated by
Dr. Yukun Lai

CM3203 One Semester Individual Project 40 Credits
May 5, 2017

ABSTRACT

Most saliency benchmark datasets lack distorted images for saliency models to predict fixation locations. In this project, a new dataset is created consisting of original images along with its distorted versions categorised into different image content. Eye tracking experiments are conducted to create ground truth fixation locations. The aim is to find differences between distorted versions and non-distorted versions. Different evaluation metrics are used to generate a range of findings. Our comparisons showed that there are some minor differences in visual perception between the distorted and non-distorted versions.

Acknowledgements

I would like to thank my supervisor, Dr. Hantao Liu, for the guidance he has given me throughout my project. It has been amusing to work with him.

I also would like to thank my friends who continues to see me as role model. All we can do is push one another to keep going. I would like to thank my family for their continuing support allowing me to be where I am now.

Most of all, I thank God for giving me so many blessings in life. Proud to be INC.

Table of Contents

Table of Figures

1. Introduction

Understanding human perception has been quite a challenge throughout Computer Vision. Models are constantly being evaluated and improved on their ability to predict where people look in images. These models have been considered to be biased on small scale datasets. More and more datasets are being created to boost saliency research to prevent any biased models.

From recent datasets [1][2], images are either categorised or distorted. I wanted to investigate how different distortion types can affect human perception of saliency on different categories.

The aim of the project is to find differences or similarities between how humans look at images that are distorted against images that are untouched (not distorted) on a completely new dataset.

The scope of the project will consider decisions made when selecting images fit for dataset creation. I shall also look at the importance of experiment design when collecting human fixation data of the images.

My approach will involve three different aspects to achieve the aim. These aspects are dataset creation, data collection and data analysis. It is hoped that the dataset created during this project can be used to even more boost saliency research and look at aspects of distortions when improving the models.

2. Background

2.1 MIT Saliency Benchmark

This website was my introduction to saliency research during my CUROP project last summer. There are a lot of research papers that tackle saliency research. These papers look at different aspects of saliency such as understanding the evaluation metrics [3] to benchmarking computational models [4]. Models are submitted and evaluated here against numerous image datasets. There are also sections on different evaluation metrics and code for the saliency maps generated by the models. One of the most popular datasets used are MIT300 and CAT2000. In CAT2000 [5], we are presented with 4000 images categorised into 20 categories.

2.2 CUROP – (Cardiff University Research Opportunity Programme)

From my CUROP Project, I have been able to collect 200 images in total over the internet [6]. Images from Unsplash are licensed under Creative Commons Zero [7] which means that they are royalty-free and will not violate any copyright issues. All of the images are high quality and are suitable for degradation when adding distortions. These images have been resized into 1920 x 1080 pixel resolution.

2.3 Categories of Images

Inspired by CAT2000, I have been able to allot the images into 10 categories. Each category consists of 20 similar images.

The categories used are:

Action, Black and White, CGI, Indoor, Object, Outdoor Man-made, Outdoor Natural, Pattern, Portrait, Social

To clarify how the images are allotted, I have added a small description of each category.

Action - Images that contain some form of movement

Black and White - Captured images of objects or people that are in grayscale

CGI - These images are not taken by cameras but instead are generated by computers which involve mostly video game characters

Indoor - Images captured inside different buildings from museums, to trains, libraries and houses.

Object – Objects with different shapes and sizes and content

Outdoor Man-Made - Images captured outside mostly involving buildings of different forms

Outdoor Natural - Images that capture sceneries with no human architecture

Pattern - Images that show repetition of shapes

Portrait – Close-up shots of different people of different age and culture

Social - Images with interactions between different people

2.4 Distortion Types

There are at least 20 different distortions [1] that can be used to degrade the images. In my CUROP project, I have created a graphical user interface that utilises seven different distortion types to be applied on a given image. Upon selecting an image and distortion type, the user is able to view three levels of degradation. These levels can be manipulated using the sliders and the result is instantly shown in the corresponding display section. When choosing the distortions, I shall refer to the seven distortions that I have already implemented as I have been able to understand the process within each individual distortion type. The seven distortion types mentioned are as follows: Additive Gaussian Noise, Gaussian Blur, Contrast Change, Motion Blur, Colour Saturation, JPEG Compression and Lossy Compression of Noisy Image.

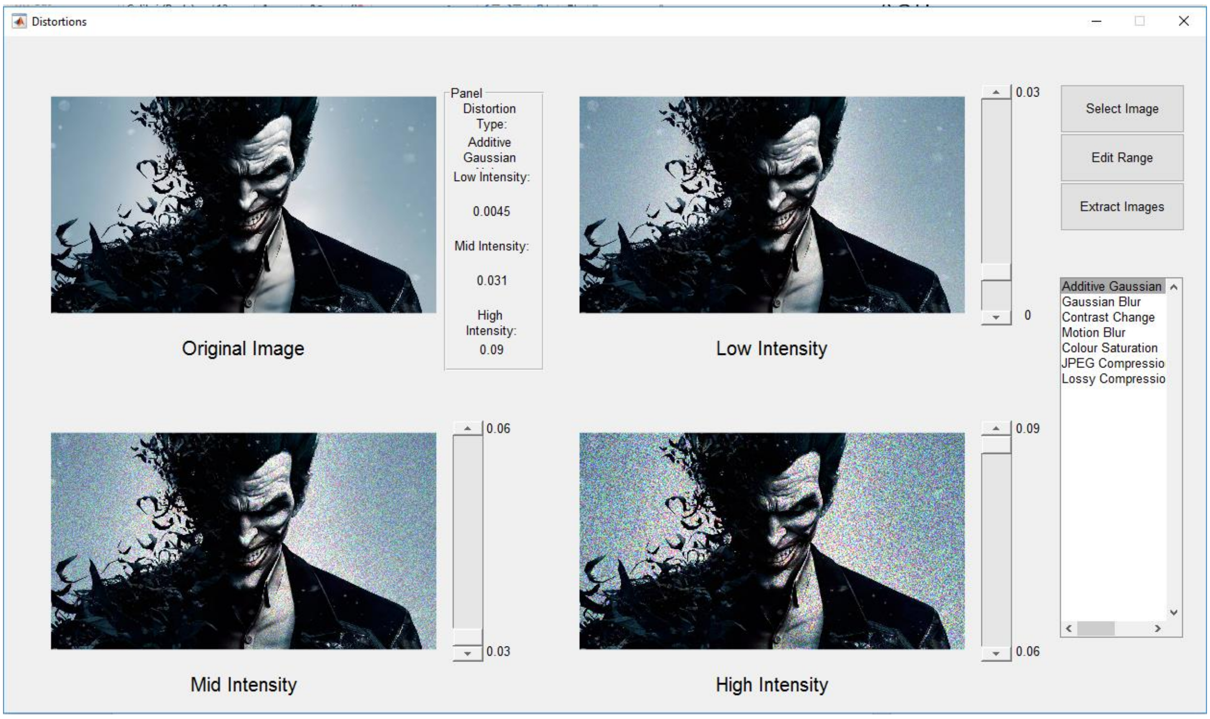


Figure 1: Distortion GUI created in CUROP

2.5 Directory Setup

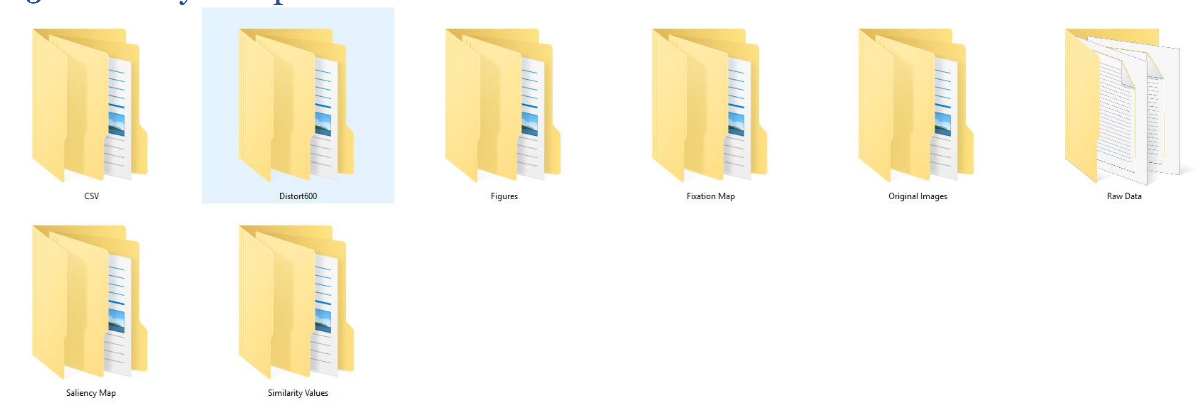


Figure 2: Directory

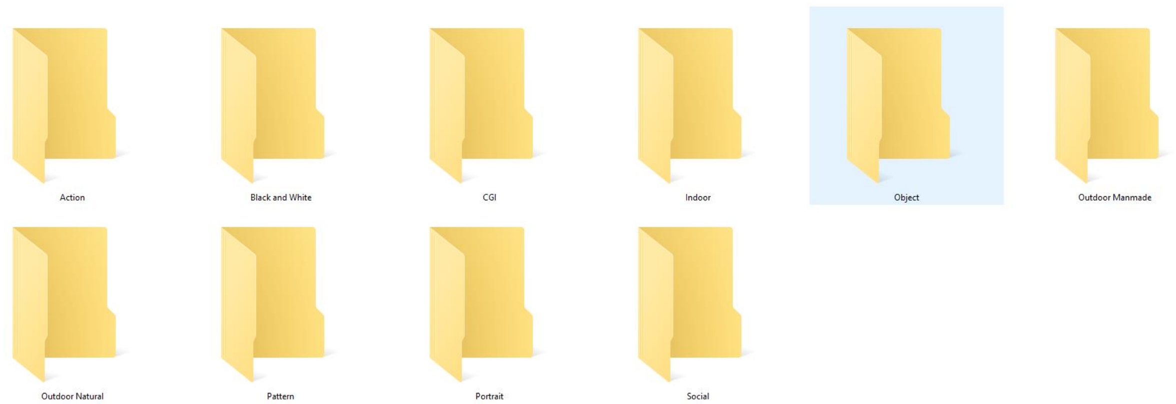


Figure 3: Sub Directory

Throughout the report, I shall be referring back to the two figures that shows the directory and sub directory. The figures will be used as a guide all throughout the different sections of the project aim. All throughout the next sections I shall be discussing individually the different folders that are shown within Figure 2. Within each directory contains the folders for each individual category for the image categories.

3. Design

3.1 Specifications

In this part of the report, I shall be discussing the requirements and considerations needed during the dataset creation, data collection and data analysis. For dataset creation, I shall further examine how to generate a suitable dataset and what are the constraints that were used. For data collection, I shall look at different aspects of experimental design and what problems may arise and how to prevent them. For data analysis, I shall consider different aspects of representing the data collected.

3.1.1. Dataset Creation

Selecting suitable images is a crucial step when it comes to data creation as these images will be used all throughout the project. These images shall be used all throughout the eye tracking experiment and for analysis. Selecting distortion type and distortion levels are also a key task as it can affect the experimental design as it controls the scale of the dataset.

3.1.1.1. Image Selection

There are a couple of factors that needs to be considered when selecting an image. Some are more necessary than others. One of the factors that needs to be considered is photographer bias [8] as they cause the fixation of users at the centre of the screen. This factor needs to be prevented to get a biased dataset and therefore reduction of centre biased images needs to be applied. Other factors needed to be considered are minimising noise and blur as these shows that the images are already distorted. Factors that can be considered as well are finding images with multiple objects to stimulate the gaze of the user. For the CGI category, it is preferable to select images that are less realistic as observer's may treat the category differently.

Overview of the constraints when selecting images are as follows:

- Not centre-biased
- Minimum noise and blur
- Contain Multiple Objects when possible
- Least realistic for CGI category

3.1.1.2. Distortion Type

When selecting the distortion types, we shall need to consider the ease of implementing the distortions. As mentioned in 3.4, I have been able to achieve seven different distortion types which can be re-used to create the dataset. When selecting the distortion type, I want to be able to differentiate between the distortion types and therefore similar distortion types can be discarded for the purpose of the project. I want to be able to control the levels of distortions as well and therefore the I need to have fully understood the distortions first.

3.1.1.3. Distortion Level
Selecting the number of distortion level can impact the size of the dataset. If there are m numbers of distortion types and I select n different levels. The size of version per image would be 1 + (m * n) the amount. Knowing how to generate the levels is important as increasing or reducing the dataset maybe needed later on.

3.1.2. Data Collection
Within the section, I shall consider the different aspects of experimental design when collecting human data from an eye tracker. I propose that the dataset is split into different groups with equal image content per category. These groups shall be used as different experiments where the observers are allotted equally. This experimental design was inspired by SIQ288 [9] and initially I shall follow the same technique used as guide. There shall some alterations on the experimental design as the images used may not be the same amount.

3.1.2.1. Experiment Setup
Similarly, in CAT2000 [5], the observers are given five seconds to view each image. In between the viewing of the images, the observer is shown a two second grey screen as visual masking [10].

3.1.2.1.1. Task
For every eye tracking experiment, the observers are given a task. Some observers are given tasks such as memory tasks [11] or searching tasks [12]. For this project, the users are given no tasks and can freely view the images. With no task at hand, I want to see if the way observers see an image can differ with untouched original images against distorted versions.

3.1.2.1.2. Display
When conducting an eye tracking experiment, I must ensure that correct or adequate display monitor is used. The resolution of the display must be able to accommodate the full resolution of the images. Other factors that needs to be considered includes not using displays with dead pixels or has the potential to flicker during the experiment.

3.1.2.1.3. Eye Tracker
The Eye Tracking System that will be used is the SensoMotoric Instrument (SMI) Red-m with a sampling rate of 250 Hz. The eye tracker is powerful enough to detect the observer's eye even with spectacles or eyewear. The Eye Tracker includes a software package called the Experiment Suite Scientific Software containing SMI Experiment Center and SMI BeGaze. Factors that needs to be considered when using the eye tracker includes:

- Calibration
- Distance from eyes to the eye tracker
- Detection of the eyes

3.1.2.1.4. Environment

Some control variables needed to be considering when conducting the experiment include the lighting condition of the room. I need to ensure that the experimental room to be used has adequate light source and. Any excessive noise must be considered as well as it can distract the concentration or perception of the observer during the experiment.

3.1.2.2. Observers

There should be an adequate number of observers per image. As mentioned in 4.1.2, splitting the images into groups means that a lot of observers are needed. Factors needed to be considered when selecting observers include the gender, age and their availability. I shall also need to further research how many observers per image is required to get a reliable data collection.

3.1.2.3 Groupings

Since the dataset will be large scale, I shall need to group the images for smaller manageable experiments. Within each experiment, a selected number of observers will view participate. The group will consist of equal distribution of category type, distortion type and distortion levels. I also need to consider how many times the same image version are viewed by each observer.

3.1.3 Data Analysis

In this section, I shall look at exporting the results of the experiment as raw data and how to manipulate the data to generate a more meaningful visualisation. When analysing the raw data, the fixation points are important when understanding where the observers looked at the most. From these fixations points, saliency maps can be generated as a more meaningful representation. As an analysis, I want to be able to compare the original images with the distorted versions and therefore more research in this aspect is required. Within the MIT Saliency Benchmark, there are several evaluation metrics that can be used to analyse saliency maps. Further research on these metrics are needed.

3.1.3.1 Data Extraction

Understanding the format of the raw data is important, especially for data mining. Within BeGaze, there are several ways to format and export raw data. I shall consider different ways of formatting the raw data and finding which format is the most suitable for data mining. Since the data are raw, I may need to parse through the data and make it more readable and more usable for image mapping.

3.1.3.2 Image Mapping

The use of saliency maps for analysis is recommended to get results as there are several documentations available [13] which aids to find similarities between the distortions and the original images. I shall need to research and understand how to construct saliency maps beforehand using the raw data exported from the experiments.

3.1.3.3. Evaluation Metrics
Using the code generated for saliency research, I should be able to get similarity values of the distorted versions against the original images. When getting the values, I need to store these values for visualisation. At first, I shall need to get the evaluation metrics that are suitable for my saliency maps.

3.1.3.3 Visualisation
Displaying the evaluation metric results in figures will help as visual aid for finding any differences or similarities between distortions and original. Creating a meaningful and good representation of the results are essential and therefore I shall need to look at different ways to interpret and visualise the results.

3.2 Flow Chart
The flowchart presented on the following shows the visual representation of my upcoming implementations. It can be broken down into 3 subsections: Dataset Creation, Data Collection and Data Visualisation as represented in different colours.

The flow chart shows the structure and procedure used on the implementation from start to finish.

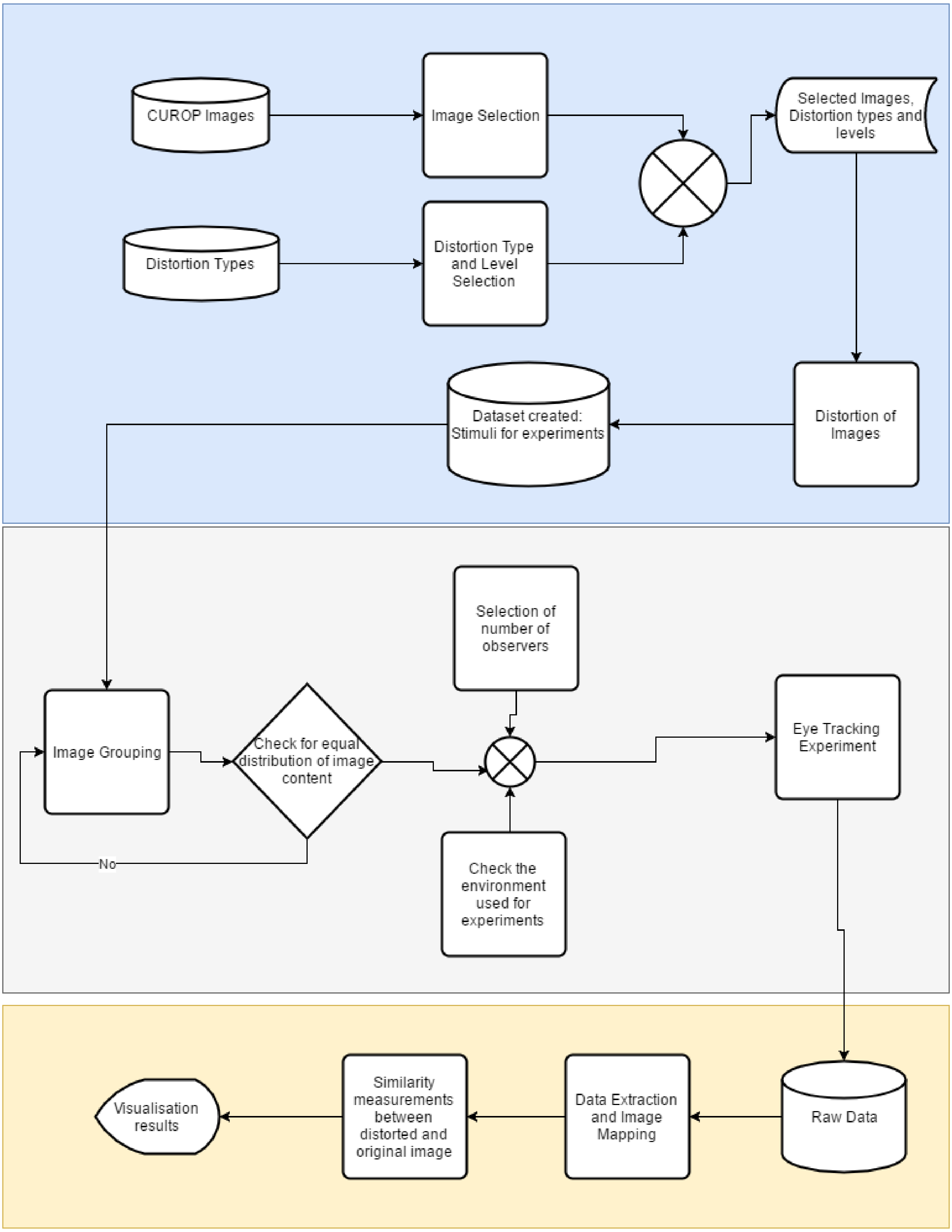


Figure 4: Flow Chart

4. Implementation

This section talks about the implementations made from the specifications that was previously mentioned. I shall discuss different aspects of how the specification was met and other changes or alterations made in order to proceed to the next stages. Within the implementations, I shall talk about different versions made and the changes or problems detected to create the next versions. The implementations that were altered include the selection of the image, the grouping of the images and the format of the raw data.

I have used two programming which are MATLAB and Python due to the availability of the code and also the usefulness of the libraries. For MATLAB, I have focussed on tasks that deals with image processing as there are some useful functions and documentations that are readily available. For Python, I have focussed on data processing and visualisation using the library pandas as it is powerful in manipulating dataframes.

4.1 Dataset Creation

For this section, I shall talk about the dataset Distort600. With 3 distortion levels and 3 distortion types, 9 alternative versions of the original image are created. In total, we would have 10 versions of the same image. With 60 original images, it would accumulate to 600 images. I shall explain in more detail why we have used 3 distortion types and 3 distortion levels. The contents of Distort600 is put within the Distort600 directory in figure 2.

4.1.1 Image Selection

Using the constraints given in the design specification, there were a lot of images that were ruled out and not included in our dataset. Some images were considered as photography biased as the content of the image is exactly at its centre. Some images had some intentional distortions as a photography technique to make the images more aesthetically pleasing. There should be an equal number of selected images per category while also considering the number of distortion type and level. During the selection, with the guidance of my supervisor, I have decided to use 7 images per category. Choosing 7 images per category means that we have 70 Original Images in total. As 70 is not advisable for grouping, I have decided to reduce the images to 6 per category. The 7^{th} image will be used as a reserved image if alterations are needed during the experiment. Upon selecting the images here are the results.

Category	Selected Images from CUROP project	Reserved Images
Action	1,6,7,11,14,19	16
Black And White	1,5,9,10,13,14	12
CGI	5,14,15,16,19,20	1
Indoor	2,5,6,7,8,13	17
Object	1,2,5,6,7,9	15
Outdoor Manmade	3,4,6,8,14,15	10
Outdoor Natural	3,7,9,13,14,16	19
Pattern	2,3,12,13,19,20	5
Portrait	1,2,4,12,16,17	5
Social	8,10,13,14,16,20	1

Figure 5: Initial Image Selection

These are the initial selection of the image along with the reserved images. As a reference, I shall include along with this report the 200 images collected in CUROP. There were some changes in the image selection and are shown in red during the pilot tests in the Data Collection section. The images marked red are replaced with the reserved images. A new image is selected for categories that require more than one change.

Category	Selected Images from CUROP project	Renumbered as
Action	1,6,16,11,2,19	1,2,3,4,5,6
Black And White	1,5,9,10,13,12	1,2,3,4,5,6
CGI	20,5,14,15,16,19	1,2,3,4,5,6
Indoor	2,5,6,7,8,13	1,2,3,4,5,6
Object	1,2,5,6,7,9	1,2,3,4,5,6
Outdoor Manmade	10,4,6,8,15,13	1,2,3,4,5,6
Outdoor Natural	3,7,9,13,14,16	1,2,3,4,5,6
Pattern	2,3,12,13,19,20	1,2,3,4,5,6
Portrait	1,2,4,12,16,17	1,2,3,4,5,6
Social	20,8,10,13,14,16	1,2,3,4,5,6

Figure 6: Final Image Selection

Figure 6 shows the final images selected as our Original Images. The numbers that have the red font colour shows the changes made. Notice as well that the order of the images is changed. In Social category, image 20 is at the start of the selected image. This is because image 20 is renumber to image 1 in my Original Images directory. Going back to figure 1, the selected 60 images is saved within the 'Original Images' directory in their corresponding subdirectories in figure 2.

4.1.2 Distortion Type

I have decided to use 3 different distortion types which are Contrast Change, Motion Blur and JPEG Compression. My supervisor and I have agreed that 3 different distortions are sufficient as a sample data for my final project. Adding more distortions will create an even larger dataset which may be hard to collect data due to the restriction of time. With the distortion types selected, they are simple to implement as there are built-in MATLAB Functions available. These distortions can also be distinguished from one another which is one of considerations as stated in my specification.

Contrast Change - achieved using one of MATLAB's documentation on Contrast Enhancement [14]. The images in RGB colour space is converted to L*a*b* colour space where the luminosity of the image can be manipulated. There are three techniques to contrast change in MATLAB which are imadjust, histeq and adapthisteq, however, we have used the latter as our distortion. This technique ensures that given any luminosity parameter it performs histogram equalization on small data regions that matches the specified histogram and prevent amplifying noise. The luminosity parameter is used to achieve different levels of distortions.

Motion Blur - MATLAB fspecial [15] and imfilter [16] commands are used to achieve this distortion. fspecial can has three parameters which are the type, length and angle. To achieve motion blur, the type parameter is set to 'motion' and also, I have set the angle parameter to 45 degrees. The level of distortion is changed by the length parameter. Given the three parameters, fspecial returns square sized kernel. Varying the length changes the size and content of the kernel to be used for imfilter. Distortion is achieved through convolution of the image and the kernel.

JPEG compression – is achieved using MATLAB's imwrite built-in function. Within the function, we can specify the type of image to be saved and also the quality of the output file.

4.1.3 Distortion Level

Making full use of my GUI created in CUROP, my supervisor and I have agreed to use 3 levels of distortions. The sliding tiles helped in finding a suitable parameter for the different levels of distortions for the different distortion type. These parameters are applied to all of the images from all categories. The parameters per distortion used are as follows:

	Low Distortion	Mid Distortion	High Distortion
Contrast Change	0.5	1.5	2.5
Motion Blur	5	15	25
JPEG Compression	1.5	10	18

Figure 7: Parameters to achieve distortion levels

4.1.4 Distortion Code: distort.m

Using the code, I have used for my GUI, I have created a MATLAB script that will loop through each image within the 'Original Images' directory and create a distorted version of them with the 3 chosen distortion types along with the 3 levels of distortions. Each generated stimulus is then saved to the corresponding category within the Distort600 directory. As the Distort600 has been created, we shall refer to the images within the Distort600 as stimuli.

4.2 Data Collection

This section will consider different aspects of data collection. I shall be discussing the preparations made when asking for observers to participate, creating the experiment protocols, briefing the observers and setting up the experiments.

4.2.1 Participation Information Sheet

Before going straight to the experiments, I have had to give some information to the potential observers. I created a participation information sheet that gives an introduction on my final year project. There are several sections given within the information sheet such as the purpose of the project, actions to be taken if they want to participate, the duration of the experiment, advantages and disadvantages in taking part and confidentiality of the experiment.

4.2.2. Experiment Setup

Following the specifications listed in 4.1.2.1, I have used also used the SMI Experiment Centre within SMI's Experiment Suite Scientific Software. The Experiment Centre allows me to create experiments for eye tracking and add different stimulus within an experiment.

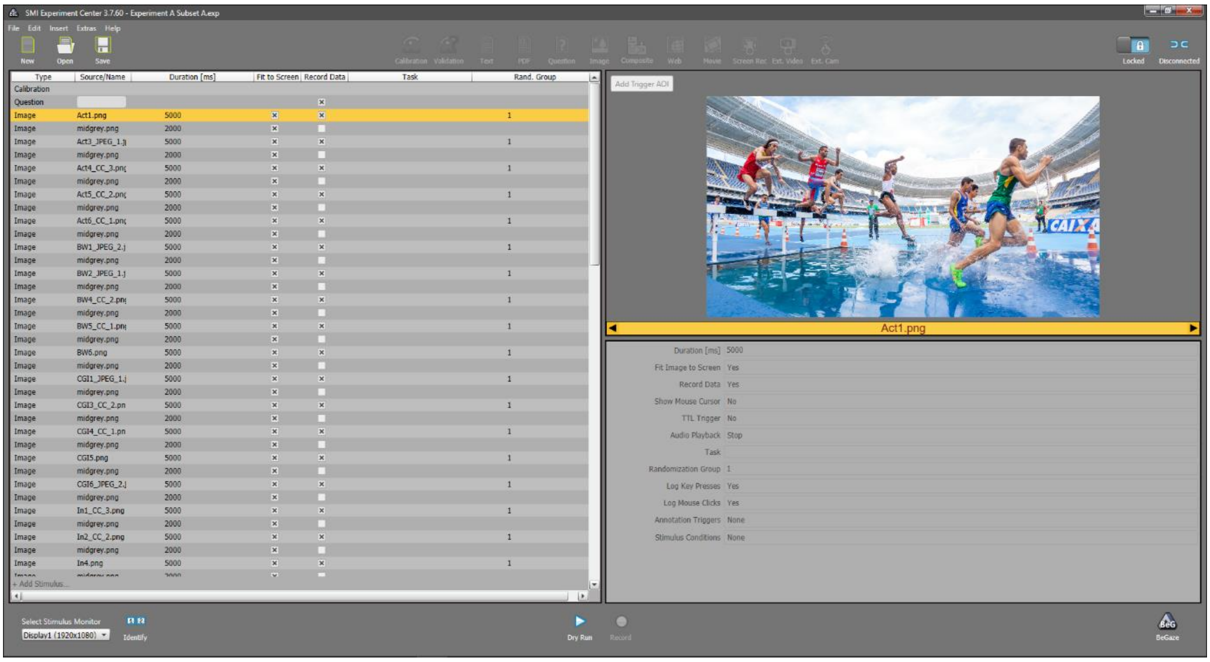


Figure 8: SMI Experiment Centre

The experiment has been setup to randomise the viewing order of the stimuli. Re-running the experiment will present a different order of the stimuli each time. At the very beginning of each experiment, I have configured the experiment to start with a 9-point calibration.

4.2.2.1. Display

Following the given specifications, I have used an LCD Screen 19-inch laptop which has the capabilities to achieve a resolution of 1920x1080 pixels. The laptop is part of the package that came along with the eye-tracker and therefore it is suitable to be used as a display for eye tracking experiments. There are also no dead pixels on the display and for every experiment, I made sure there are no dust on the display.

4.2.2.2. Eye Tracker

As for the distance of the observer, it has been maintained between 60 to 65 cm from the observer's eyes and the eye tracker. The angle of the laptop has been kept the same all throughout the entire data collection. Before the start of the experiments, the observers are briefed about the experiment setup and told to refrain from too much head movements to prevent or reduce any anomalies. For the calibration of the eye tracker, I have used 9-point calibrations to at the start of each experiment.

4.2.2.3. Environment

The environment that I have used is a recording studio within the university. I have been given access to use the studio as an experimental room. The room is secluded from the main halls where noise may be generated and the room is filled with acoustic foam. The room is silent and the lighting condition is adequate for the experiment. During the experiment, I leave the room so that the observer has full attention on the stimuli.

4.2.2.4. Problem – Fatigue

Viewing images for a long period of time can cause eye strain, discomfort and fatigue. It is therefore essential that the experimentation is structured to reduce this. Minimising experimentation time is required while getting adequate results.

4.2.3 Observers

According to [17], it is advisable that we have at least 16 observers per image to have a reliable collection of eye tracking data. 16 observers per image will be difficult to achieve as the dataset will need to be split into groups first. 16 observers cannot view 600 images at once as it will take a lot of time and availability for the observers. If the images are split into 2 groups with 300 images per group, I will need 2 x the number of observers which is 32. It shall be too short for the span of the project and therefore the number of observers per image is reduced. In [17], we can look at the inter-observer agreement (IOA) value depending on the number of observers. "The IOA refers to the degree of agreement in saliency among observers viewing the same stimulus". Using the IOA, it is recommended that I get 10 observers per image to be able to get an acceptable eye tracking data.

4.2.3.1 Problem - Availability

The target audience for the observers are pupils and staff within the school university. However, availability of the observers is only within the academic schedules. Some students are unavailable due to the need of attendance of lectures and clubs. Some staff are also unavailable. I have setup a doodle poll [18] that I have used that allows the user to select his available time from the time range that I have set. With the availability of observers being a problem, I have restricted the number of observers to 7 per image.

4.2.4. Groupings

This section shall look at the approaches that I made when grouping the images. I have made 3 different versions and changes were made based on the previous versions. When the groupings are made, a pilot test of experiment was executed to make further improvements.

4.2.4.1 GROUPING VERSION I

Following the experiment design in [9], I shall split Distort600 into 3 different groups and furthermore split to subgroups as shown in the diagram. Each observer is allotted into either groups A, B or C. They are then shown two sub experiments A and B within their experiment group. The observers are given breaks in between to reduce fatigue and strain in the eyes. Each sub group would contain 100 images each.

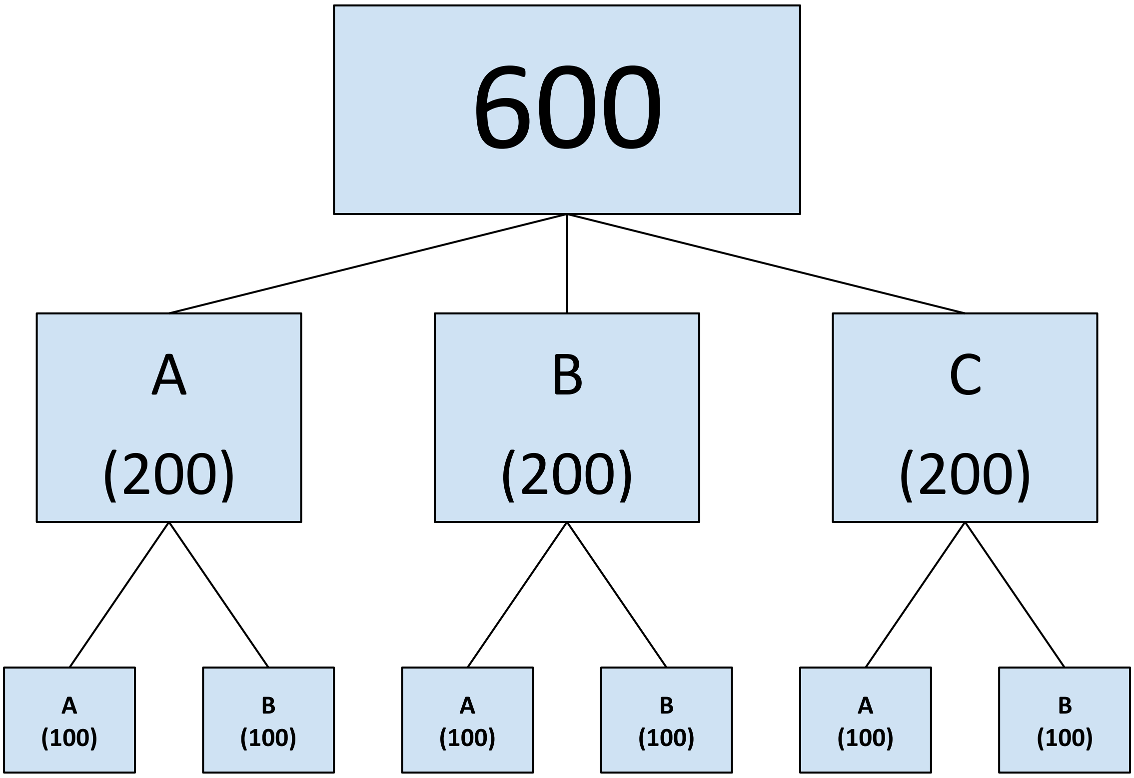


Figure 9: Grouping Version I

The stimuli are split into 6 subgroups which are:

Experiment A Subset A
Experiment A Subset B
Experiment B Subset A
Experiment B Subset B
Experiment C Subset A
Experiment C Subset B

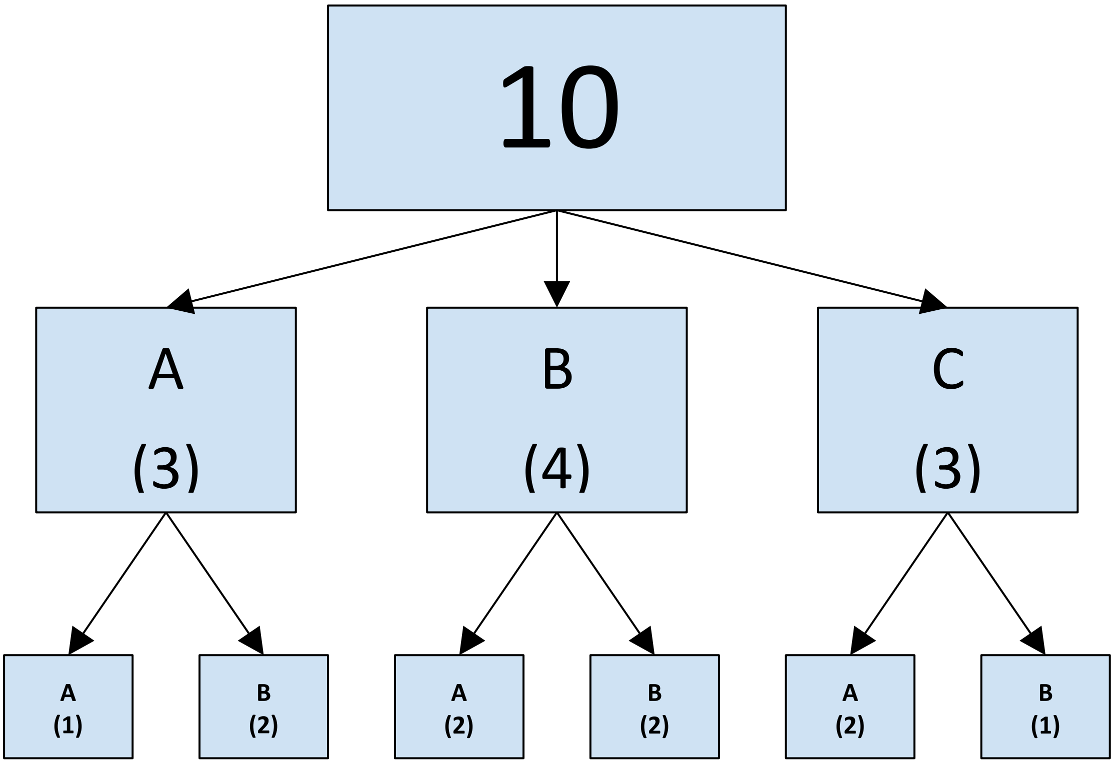


Figure 10: Splitting the image versions into 3 groups

Figure 10 shows how an image with 10 different versions have been split as equally as possible. Each subset containing no more than two repeated images. The maximum number of times an observer may see the same image content is restricted to 4 as shown in the example of Experiment B. This is to reduce visual perception and visual memory of the repeated images to prevent or minimise the effect of looking at the image differently. The number of image content is randomised and A=3, B=4, C=3 is just an example. Other alternatives include, A=4, B=3, C=3 and A=3, B=3, C=4.

Pilot Test Result: Although the images are split equally in terms of image content, the image distortion types and levels were not equal. Some groups had too much contrast change distortion and some were too highly distorted. Also, some of the images selected do not have the exact dimensions as 1920x1080 and some are slightly smaller in height. These images were changed as shown in figure 6.

```
G:\Final Year Project\Experiment\Versions\Experiment[V1]\Experiment A\Experiment A>python checkcontent.py
CC1: 17
CC2: 9
CC3: 7
JPEG1: 10
JPEG2: 6
JPEG3: 12
MB1: 14
MB2: 5
MB3: 10
orig: 10
Total Images: 100
```

Figure 11: Content Check 1

```
G:\Final Year Project\Experiment\Versions\Experiment[V1]\Experiment C\Experiment A>python checkcontent.py
CC1: 10
CC2: 8
CC3: 12
JPEG1: 11
JPEG2: 10
JPEG3: 11
MB1: 11
MB2: 8
MB3: 9
orig: 10
Total Images: 100
```

Figure 12: Content Check 2

I have created a python script that allows me to see the image contents within a directory. The script is located within each individual subset directory and it will count the number of distortion type and level within the directory. From the result of the python script there are only 6 mid-level JPEG Compression in Experiment A Subset A while in Experiment C subset A, there are 10. An alternative grouping must be used and ensure that all content are equally distributed.

4.2.4.2. GROUPING VERSION II

To make the image content as equal as possible per subset, allotting the images in each group then subset must be consistent. I have proposed an allotting procedure that allows equal distribution of image content within the subsets of each individual group.

	Orig	CC1	CC2	CC3	JPEG1	JPEG2	JPEG3	MB1	MB2	MB3
Act1	AA	BA	CA	AB	BB	CB	AA	BA	CA	AB
Act2	BA	CA	AB	BB	CB	AA	BA	CA	AB	BB
Act3	CA	AB	BB	CB	AA	BA	CA	AB	BB	CB
Act4	AB	BB	CB	AA	BA	CA	AB	BB	CB	AA
Act5	BB	CB	AA	BA	CA	AB	BB	CB	AA	BA
Act6	CB	AA	BA	CA	AB	BB	CB	AA	BA	CA
BW1	BA	CA	AB	BB	CB	AA	BA	CA	AB	AA
BW2	CA	AB	BB	CB	AA	BA	CA	AB	BB	BA
BW3	AB	BB	CB	AA	BA	CA	AB	BB	CB	CA
BW4	BB	CB	AA	BA	CA	AB	BB	CB	AA	AB
BW5	CB	AA	BA	CA	AB	BB	CB	AA	BA	BB
BW6	AA	BA	CA	AB	BB	CB	AA	BA	CA	CB
CGI1	CA	AB	BB	CB	AA	BA	CA	AB	AA	BA
CGI2	AB	BB	CB	AA	BA	CA	AB	BB	BA	CA
CGI3	BB	CB	AA	BA	CA	AB	BB	CB	CA	AB
CGI4	CB	AA	BA	CA	AB	BB	CB	AA	AB	BB
CGI5	AA	BA	CA	AB	BB	CB	AA	BA	BB	CB
CGI6	BA	CA	AB	BB	CB	AA	BA	CA	CB	AA
In1	AB	BB	CB	AA	BA	CA	AB	AA	BA	CA
In2	BB	CB	AA	BA	CA	AB	BB	BA	CA	AB
In3	CB	AA	BA	CA	AB	BB	CB	CA	AB	BB
In4	AA	BA	CA	AB	BB	CB	AA	AB	BB	CB
In5	BA	CA	AB	BB	CB	AA	BA	BB	CB	AA
In6	CA	AB	BB	CB	AA	BA	CA	CB	AA	BA
Ob1	BB	CB	AA	BA	CA	AB	AA	BA	CA	AB
Ob2	CB	AA	BA	CA	AB	BB	BA	CA	AB	BB
Ob3	AA	BA	CA	AB	BB	CB	CA	AB	BB	CB
Ob4	BA	CA	AB	BB	CB	AA	AB	BB	CB	AA
Ob5	CA	AB	BB	CB	AA	BA	BB	CB	AA	BA
Ob6	AB	BB	CB	AA	BA	CA	CB	AA	BA	CA

	Orig	CC1	CC2	CC3	JPEG1	JPEG2	JPEG3	MB1	MB2	MB3
OM1	CB	AA	BA	CA	AB	AA	BA	CA	AB	BB
OM2	AA	BA	CA	AB	BB	BA	CA	AB	BB	CB
OM3	BA	CA	AB	BB	CB	CA	AB	BB	CB	AA
OM4	CA	AB	BB	CB	AA	AB	BB	CB	AA	BA
OM5	AB	BB	CB	AA	BA	BB	CB	AA	BA	CA
OM6	BB	CB	AA	BA	CA	CB	AA	BA	CA	AB
ON1	AA	BA	CA	AB	AA	BA	CA	AB	BB	CB
ON2	BA	CA	AB	BB	BA	CA	AB	BB	CB	AA
ON3	CA	AB	BB	CB	CA	AB	BB	CB	AA	BA
ON4	AB	BB	CB	AA	AB	BB	CB	AA	BA	CA
ON5	BB	CB	AA	BA	BB	CB	AA	BA	CA	AB
ON6	CB	AA	BA	CA	CB	AA	BA	CA	AB	BB
Pat1	BA	CA	AB	AA	BA	CA	AB	BB	CB	AA
Pat2	CA	AB	BB	BA	CA	AB	BB	CB	AA	BA
Pat3	AB	BB	CB	CA	AB	BB	CB	AA	BA	CA
Pat4	BB	CB	AA	AB	BB	CB	AA	BA	CA	AB
Pat5	CB	AA	BA	BB	CB	AA	BA	CA	AB	BB
Pat6	AA	BA	CA	CB	AA	BA	CA	AB	BB	CB
Po1	CA	AB	AA	BA	CA	AB	BB	CB	AA	BA
Po2	AB	BB	BA	CA	AB	BB	CB	AA	BA	CA
Po3	BB	CB	CA	AB	BB	CB	AA	BA	CA	AB
Po4	CB	AA	AB	BB	CB	AA	BA	CA	AB	BB
Po5	AA	BA	BB	CB	AA	BA	CA	AB	BB	CB
Po6	BA	CA	CB	AA	BA	CA	AB	BB	CB	AA
Soc1	AB	AA	BA	CA	AB	BB	CB	AA	BA	CA
Soc2	BB	BA	CA	AB	BB	CB	AA	BA	CA	AB
Soc3	CB	CA	AB	BB	CB	AA	BA	CA	AB	BB
Soc4	AA	AB	BB	CB	AA	BA	CA	AB	BB	CB
Soc5	BA	BB	CB	AA	BA	CA	AB	BB	CB	AA
Soc6	CA	CB	AA	BA	CA	AB	BB	CB	AA	BA

Figure 13: Grouping Version II

Figure 13 shows the allotment that I have used. There are two patterns that within this setup. First is the pattern category type. The initial pattern is AA, BA, CA, AB, BB, CB. Going down the Orig column, each Act is allotted with the initial pattern. With Black and White the pattern is changed such that the first allotment AA is put at the back creating a new pattern which is BA, CA, AB, BB, CB, AA. Going down each category the pattern is altered in the same way. On the next pattern, we shall look at the distortion type and how they are allotted. When the Orig has been filled in, we have groups of patterns per category. When we move to the next distortion type which is the low-level contrast change distortion, the group of patterns are shifted up. The pattern in Action category is shifted to the Social category. The pink highlight shown is a guide to show how that pattern has shifted along the different distortions.

```
G:\Final Year Project\Experiment\Versions\Experiment[V2]\Experiment A Subset A>python checkcontent.py
CC1: 10
CC2: 10
CC3: 10
JPEG1: 10
JPEG2: 10
JPEG3: 10
MB1: 10
MB2: 10
MB3: 10
orig: 10
Total Images: 100
```

Figure 14: Content Check 3

Pilot Testing: During the pilot testing of an experiment, we have concluded that the experiment is too long and can cause fatigue and eye strain. During an experiment a sub-experiment with 100 images would last approximately 15 minutes long. The total experiment duration would therefore last approximately 30 minutes. The experiment length must be reduced.

4.2.4.3. GROUPING VERSION III

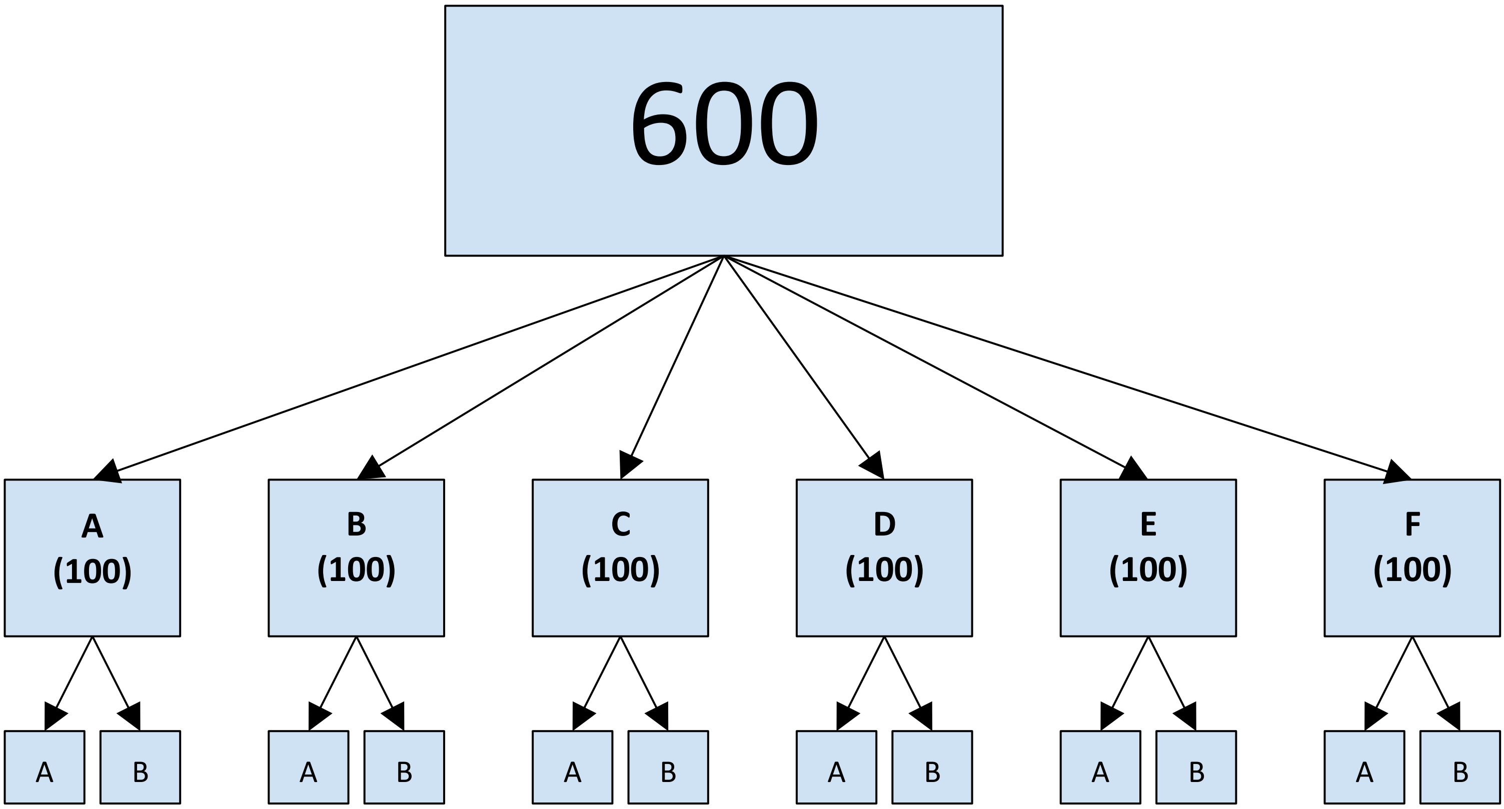


Figure 15: Grouping Version III

I have extended my groups to six and the initial groups that I have used are the subgroups from version II. The content of the groups is therefore equal and creating subgroups from these groups will not create any unequal distribution. Reducing each subset into 50 images will reduce the length of each sub-experiment by half therefore reducing the total duration of an experiment per observer. Each sub-experiment would last approximately 7 minutes long which I think is suitable for an eye tracking experiment. The observers are then given break in between sub-experiments to rest the eyes. From the problem identified earlier in 5.2.2.4, we have reduced eye strain and fatigue.

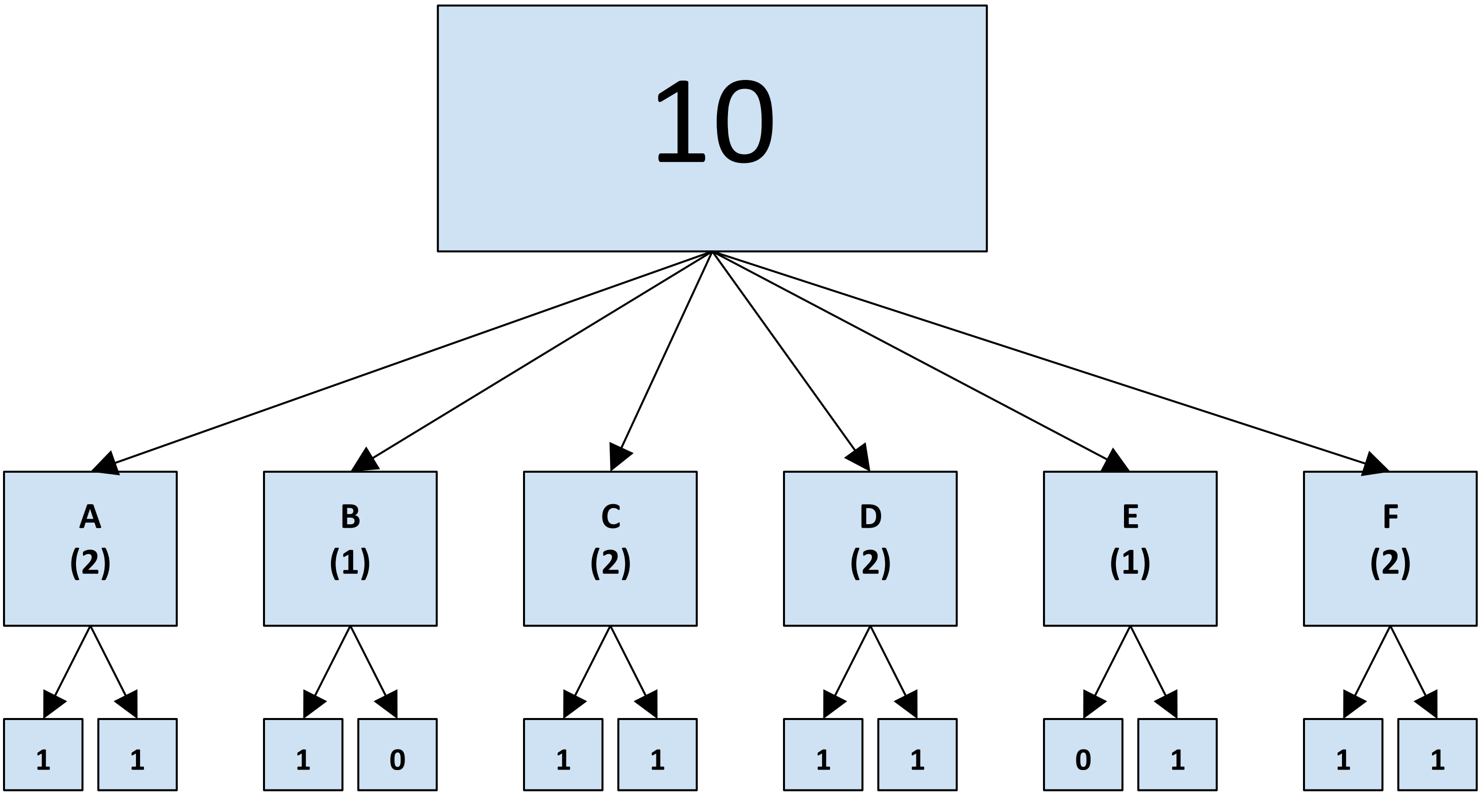


Figure 16: Splitting the image versions into 6 groups

In the first 2 versions of the groupings, the observers will see at least 3 repetitions of the same image. However, in this grouping, the same images can only be seen twice at maximum. This would further help with any deviation and influence with visual perception and visual memory. In fact, out of the 100 images the observer will see 40 images that are repeated twice during both sub experiments. This will not create an impact towards the results as the duplicates are split into each sub groups.

4.2.4.4. Problem: Experiment Length

From the groupings that has been considered, we have identified the experiment length as a problem. We have dealt with the experiment length by reducing the amount of eye tracking session and prolonged strain of the eyes. Increasing the number of groups has helped in reduction of experiment length as well. We have resulted in 6 experiment groups but we cannot increase this any further. As mentioned in [17], we hope to get 7 observers per image, which means we need total of 42 observers with 7 observers per group.

4.3 Data Analysis

When extracting the data and visualising the results, I shall be using two programming language which are MATLAB and Python. There are a lot of documentaries in MIT Saliency Benchmark that I can use that are written in MATLAB. When it comes to manipulating data values, I am comfortable with using Python's pandas data analysis library.

4.3.1. Data Extraction

Within SMI's Experiment Suite Scientific Software, I can extract the experiment results using BeGaze. There are several ways of formatting that can export the Raw Data and within this section, I have included 2 different ways. Both ways are saved in text files.

4.3.1.1. RAW DATA FORMATTING I

```
                  Experiment A Subset A_P01_002_Trial002 Events
[BeGaze]
Converted from: D:\Users\Kenric\Documents\SMI Experiment Centre\Results
Path\Experiment A Subset
A\P01-[6afa400a-01bf-454c-95d6-1d05f3ec1b47]\P01-eye_data.idf
Date:    10.04.2017 23:25:19
Version:        BeGaze 3.7.42
Sample Rate:    250
Subject:        P01
Description:    Run2

Table Header for Fixations:
Event Type      Trial   Number  Start   End     Duration        Location X
Location Y      Dispersion X    Dispersion Y    Plane   Avg. Pupil Size X
Avg. Pupil Size Y

Table Header for User Events:
Event Type      Trial   Number  Start   Description

UserEvent       1       1       2619337233      # Message: In5_JPEG_2.jpg
Fixation L      1       1       2619339288      2619550937      211649  496.89
350.11  24      47      -1      17.87   17.87
Fixation R      1       1       2619339288      2619554946      215658  492.03
312.48  52      71      -1      20.56   20.56
Fixation R      1       2       2619611072      2619807359      196287  1051.20
552.28  60      76      -1      20.17   20.17
Fixation L      1       2       2619619090      2620147332      528242  994.79
642.77  44      41      -1      17.73   17.73
Fixation R      1       3       2619827389      2620155324      327935  1040.56
562.12  49      83      -1      19.74   19.74
Fixation R      1       4       2620195283      2620491198      295915  1143.28
633.90  32      56      -1      18.79   18.79
Fixation L      1       3       2620207271      2620491198      283927  1124.79
710.32  28      30      -1      16.90   16.90
Fixation L      1       4       2620555224      2620819296      264072  1784.38
976.97  89      34      -1      16.20   16.20
```

Figure 17: Raw Data Formatting I

The first formatting of the raw data involves a file per stimuli per observer. A text file would contain general information such as subject (the observer ID) and stimuli (image) shown. The main information needed are the fixation points which are the x and y coordinates where the observer looks at the most. The first few lines of the raw data file are not wanted and data mining it means that I need to understand the structure of the first few lines and skip it. Data mining this type of file will take long to process. Each of the 6 groups would have 350 raw data text files stating the observer ID and stimuli used. Analysing all 350x6 files are not recommended as parsing through each file will take longer.

4.3.1.2. RAW DATA FORMATTING II

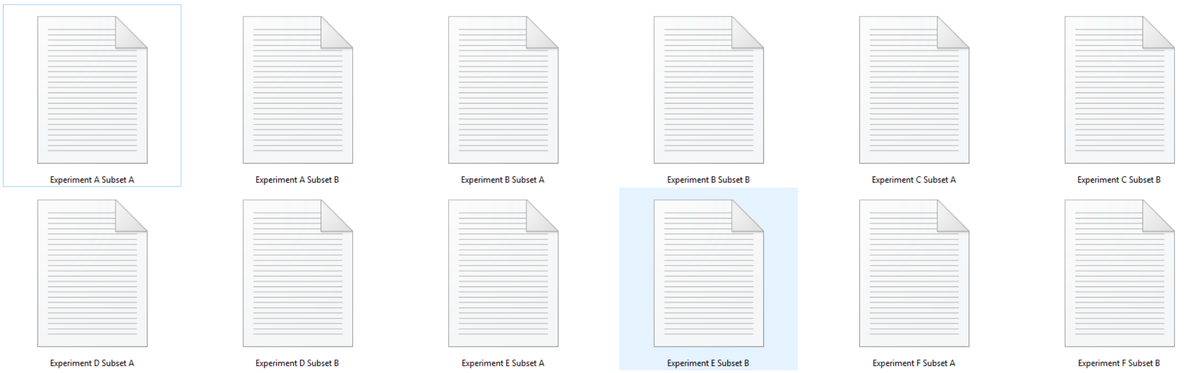


Figure 18: Raw Data text files per subset

An alternate formatting that I have used is to have all contents of fixations per subset across all participants and stimuli. All text files has a table with columns Stimulus, Participant, Eye L/R, Index, Event Start Trial Time [ms], Event End Trial Time [ms], Event Duration [ms], Fixation Position X [px], Fixation Position Y [px]. With only 12 text files and manageable content, it should be easier to extract and manipulate the data to get the desired sections that I need within the data file.

```
                         Experiment A Subset A
Stimulus        Participant     Eye L/R Index    Event Start Trial Time [ms]
Event End Trial Time [ms]       Event Duration [ms]     Fixation Position X [px]
Fixation Position Y [px]
In5_JPEG_2.jpg  P01     Right   1       2.1     217.7   215.7   492.0   312.5
In5_JPEG_2.jpg  P01     Left    1       2.1     213.7   211.6   496.9   350.1
In5_JPEG_2.jpg  P01     Right   2       273.8   470.1   196.3   1051.2  552.3
In5_JPEG_2.jpg  P01     Left    2       281.9   810.1   528.2   994.8   642.8
In5_JPEG_2.jpg  P01     Right   3       490.2   818.1   327.9   1040.6  562.1
In5_JPEG_2.jpg  P01     Right   4       858.1   1154.0  295.9   1143.3  633.9
In5_JPEG_2.jpg  P01     Left    3       870.0   1154.0  283.9   1124.8  710.3
In5_JPEG_2.jpg  P01     Left    4       1218.0  1482.1  264.1   1784.4  977.0
In5_JPEG_2.jpg  P01     Right   5       1234.0  1358.1  124.1   1812.9  913.2
In5_JPEG_2.jpg  P01     Right   6       1386.1  1474.1  88.0    1816.7  859.3
In5_JPEG_2.jpg  P01     Left    5       1542.1  1790.0  248.0   1822.6  379.7
In5_JPEG_2.jpg  P01     Right   7       1574.0  1790.0  216.0   1913.3  366.9
In5_JPEG_2.jpg  P01     Left    6       1838.0  2038.0  200.0   1668.4  407.9
In5_JPEG_2.jpg  P01     Right   8       1842.0  2042.0  200.0   1735.5  344.3
In5_JPEG_2.jpg  P01     Left    7       2150.0  2454.0  304.0   171.4   906.5
In5_JPEG_2.jpg  P01     Right   9       2206.0  2446.0  240.0   131.1   896.9
In5_JPEG_2.jpg  P01     Right   10      2498.0  2826.1  328.1   305.6   974.9
In5_JPEG_2.jpg  P01     Left    8       2498.0  3237.9  739.9   353.6   1000.9
In5_JPEG_2.jpg  P01     Right   11      2994.1  3229.9  235.8   347.2   973.3
In5_JPEG_2.jpg  P01     Right   12      3289.9  3401.9  112.0   147.6   604.1
In5_JPEG_2.jpg  P01     Left    9       3305.9  3745.9  440.0   168.8   741.1
In5_JPEG_2.jpg  P01     Right   13      3417.9  3749.9  332.0   164.6   639.5
In5_JPEG_2.jpg  P01     Right   14      3817.9  4034.0  216.1   684.8   521.8
In5_JPEG_2.jpg  P01     Left    10      3817.9  4046.0  228.1   663.0   590.9
In5_JPEG_2.jpg  P01     Right   15      4054.0  4233.9  179.9   630.7   569.5
In5_JPEG_2.jpg  P01     Left    11      4074.0  4237.9  163.9   613.6   655.2
In5_JPEG_2.jpg  P01     Right   16      4285.9  4681.9  396.0   1030.5  665.7
In5_JPEG_2.jpg  P01     Left    12      4289.9  4681.9  392.0   1036.6  772.0
In5_JPEG_2.jpg  P01     Right   17      4725.9  5013.9  288.0   1177.7  628.6
In5_JPEG_2.jpg  P01     Left    13      4737.9  5013.9  276.0   1160.8  722.2
ON4_CC_3.png    P01     Right   1       0.6     140.5   139.9   514.8   3.9
ON4_CC_3.png    P01     Left    1       0.6     136.5   135.9   518.9   93.9
ON4_CC_3.png    P01     Right   2       212.5   444.4   232.0   686.7   507.7
ON4_CC_3.png    P01     Left    2       216.5   444.4   228.0   694.3   563.8
ON4_CC_3.png    P01     Right   3       492.4   888.4   396.0   1055.8  524.4
ON4_CC_3.png    P01     Left    3       496.4   892.4   396.0   1052.9  581.9
ON4_CC_3.png    P01     Left    4       940.4   1120.4  180.0   1232.9  308.3
ON4_CC_3.png    P01     Right   4       944.4   1120.4  176.0   1265.3  305.0
```

Figure 19: Raw Data File Content

The figure above shows a screenshot of the first few lines of the text file. The Stimulus seen is In5_JPEG_2.jpg which is an Indoor type image with a mid-level JPEG Compression distortion. This formatting is much easier to data mine as we can separate the data into stimulus groups. Line 2 and 3 shows values of index 1 for In5_JPEG_2.jpg for Participant 1, the only difference is the one is for the left eye and the other is for the right eye. We can also see that start time, end time, duration, fixation x and fixation y are similar. To get a more meaningful data, we need to be able to select the same participant, stimulus and index to get the average of the left and right eye.

4.3.1.3 Raw Data Extraction Code

Using pandas library in python, I can read each individual raw data text files as dataframes. Pandas is very powerful when analysing dataframes as there are a lot of ways to analyse them. When data mining the raw data files, I shall create comma-separated value files per stimulus and save them in the 'CSV' file directory as shown in figure 2.

```
for i in Stimulus_list:
    df2 = df[df.Stimulus == i] # selects the data with
    corresponding stimuli e.g. 'Po2_MB_1'
    df2 = df2.groupby(['Participant', 'Index']).mean() #selects
    mean value of fixations (mean L and R eyes)
```

Figure 20: Raw Data Extraction Code: readtxt.py

The main feature of the python script is reading the raw data as a large dataframe with the same columns as in the raw data text file. The large dataframe is then sliced to smaller dataframes for each stimulus. We can further compress each dataframes by grouping them by their Participant and Index. We group them in this way to select both left and right eye of the corresponding Participant and Index. We can then find the mean of the left and right eye's fixation values, start time, end time and duration. The fixation values are the coordinates of the pixels. The compressed dataframe is then saved per stimuli.

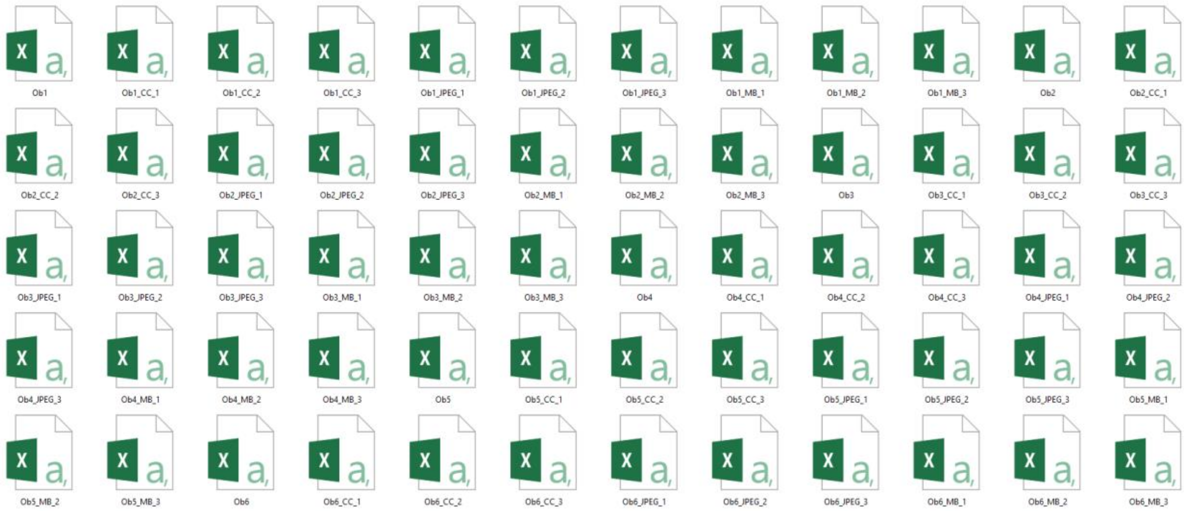


Figure 21: Example CSV file in Object Category

Participant	Index	Start_Time	End_Time	Duration	X	Y
P01	1	3.2	281.3	278	864	438.15
P01	2	327.2	555.2	228	933.95	382.7
P01	3	611.2	937.2	326	589.35	336.4
P01	4	1015.2	1373.2	358	342.5	292.95
P01	5	1455.2	1707	251.8	1597.6	299.9
P01	6	1761	1995.05	234.1	1302.65	463.05
P01	7	2037.1	2225.1	188.05	1231.3	623.5
P01	8	2303.1	2447.1	144	1454.6	692.4
P01	9	2513.1	2727.2	214.05	1737.55	608.75
P01	10	2775.2	3117.1	341.9	1652.75	519.25
P01	11	3177.1	3433.1	256	862.9	449.15
P01	12	3479.1	3668.9	189.8	805.4	512.85
P01	13	3708.9	4039	330.1	942.9	394.4
P01	14	4087	4421	334	627.9	402.5
P01	15	4463	4701	238	554.85	290.05
P01	16	4751	4987	236	283.15	256.85

Figure 22: Example CSV content of a stimulus

In figure 22, the contents of the CSV file include all of the seven observers who observed this particular stimulus. Each observer would have different lengths of indexes as they are the fixation points recorded during the eye tracking experiment.

4.3.2. IMAGE MAPPING

In this section, we shall discuss the two different mappings required to use the metric scores given in MIT Saliency Benchmark. These are Fixation Maps and Saliency Maps. Fixation Maps are the pixel to pixel representation of the fixations collected in the raw data. Saliency maps are created from fixation points.

4.3.2.1 FIXATION MAPS

Fixations are like visual representation of the raw data. I have created a zero matrix with dimension 1920x1080 which are the same as the stimuli and for every fixation point of a stimulus I change it to a 1 to create a binary matrix. These are then saved to create the fixation maps of the stimuli. The corresponding fixation maps are saved with the corresponding category in the ‘Fixation Map’ directory in figure 2.

4.3.2.2. SALIENCY MAPS

Saliency maps are the maps that are used in the metric score in [19]. To create the saliency maps, I need to convolve the fixation map with a Gaussian Blob [20]. The saliency maps generated are saved within the 'Saliency Map' directory in figure 2.

```
xy_coords = matrix(:, 4:5);
fixation_map = zeros(1080, 1920);
for j = 1:length(xy_coords)
    line = xy_coords(j, :);
    if (line(2) <=1080 && line(1) <=1920)
        fixation_map(line(2), line(1)) = 1;
    end
end

saliency_map = conv2(fixation_map, gauss, 'same');
saliency_map = im2uint8(mat2gray(saliency_map));
```

Figure 23: Saliency and Fixation map code

The xy_coords in figure 23 are the fixation coordinates used from the generated CSV files. The for loop will loop through each (x,y) coordinate and change the corresponding matrix location to 1. Within the for loop, an if statement was created as there are some anomaly in the raw data. These anomalies are (x, y) coordinates that are bigger than the dimensions of the image. Without the if statement, MATLAB would resize the fixation map matrix to a bigger dimension than the stimuli.

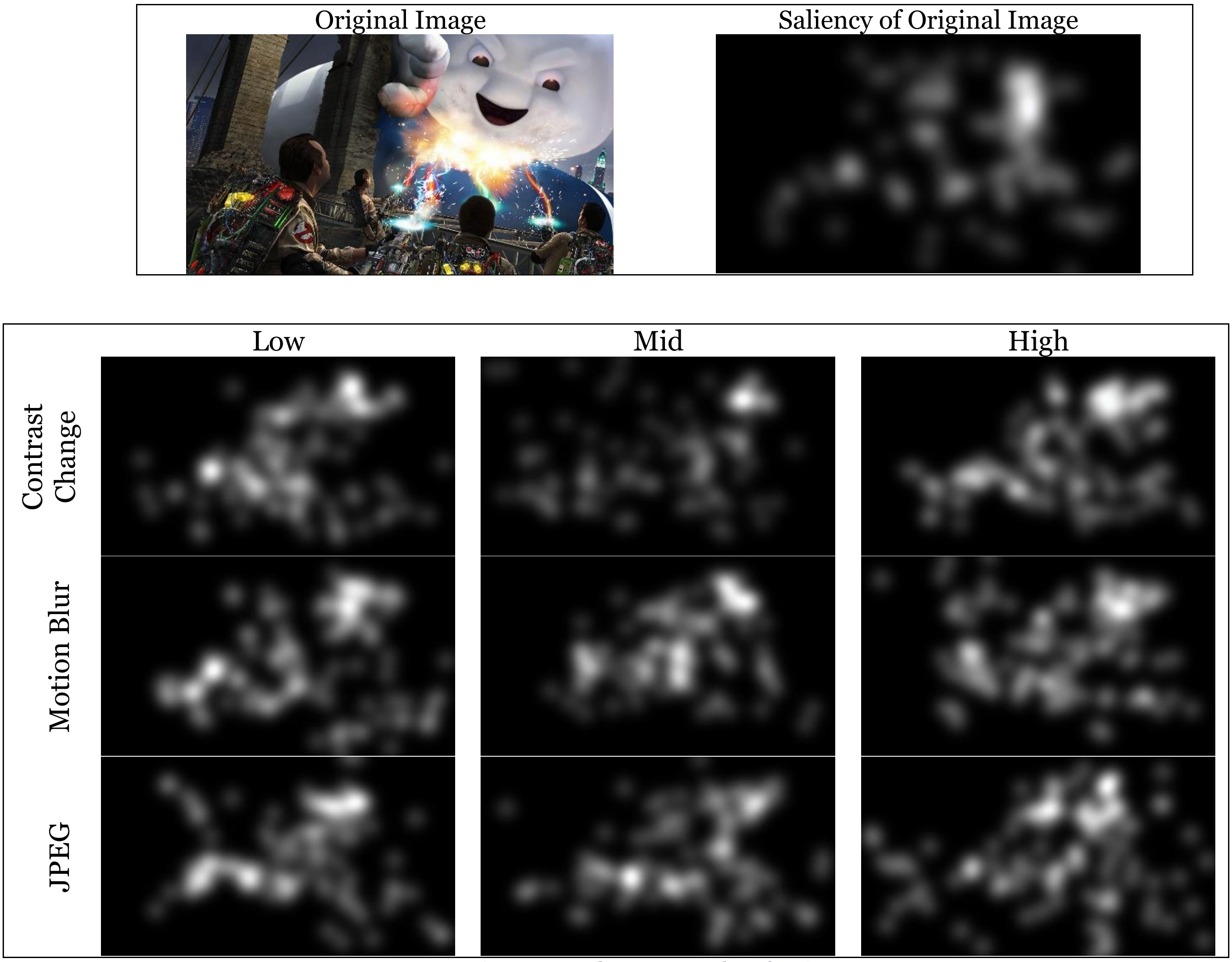


Figure 24: Saliency examples of an image

Figure 24 shows all the different saliency maps of the original image and the 9 distorted versions. These saliency maps are creating from observers. By visual evaluation, it is clearly seen that there are slight differences in how the observers see the images. On the saliency of the original image, the main focus is on Stay Puft Marshmallow Man's face, but looking at the different distortions, there is more wide spread in fixation.

4.3.2.3 Saliency Metrics

During my research, I have found different saliency metrics that can be used for my saliency maps. Using [13], I have been able to select the correct metrics that I can use for my saliency maps. The metrics that I can use include AUC, KLD, NSS and CC. Due to the shortage in time, I have decided to use CC as my metric to analyse the similarity of the original image and the distorted versions. Given two saliency maps, CC returns the linear correlation coefficient which is ranged between -1 to 1. Either 1 or -1 indicates that there is a correlation between images and a 0 indicates that there is no correlation at all. Since both saliency maps will be the same image content, we can assume that the correlation will be of range 0 to 1.

```
result = zeros(3,3);
result(1, 1) = CC(im_orig, im1);
result(1, 2) = CC(im_orig, im2);
result(1, 3) = CC(im_orig, im3);
result(2, 1) = CC(im_orig, im4);
result(2, 2) = CC(im_orig, im5);
result(2, 3) = CC(im_orig, im6);
result(3, 1) = CC(im_orig, im7);
result(3, 2) = CC(im_orig, im8);
result(3, 3) = CC(im_orig, im9);
```

Figure 25: Creation of Similarity Matrix

I have created a MATLAB script similaritymatrixcode.m that reads all 10 different versions of the same image. The linear correlation coefficient is calculated for each distortion against the original version. Each calculation is then stored in a 9x9 matrix then converted into a txt file. The matrix format is as follows:

Row,Low_Level, Mid_Level,High_Level

CC,Value,Value,Value

MB,Value,Value,Value

JPEG,Value,Value,Value

The text files are saved in the Similarity Values directory in the corresponding image category in figure 2. Within each category, 6 text files are generated for each original image which contain the similarity matrix for the original images against the distorted version.

Figure 26: Example content of similarity matrix files in a category

Ob1.txt

```
Row,Low_Level,Mid_Level,High_Level
CC,0.639006589357963,0.69691463209605,0.732449518353056
MB,0.678501559988183,0.673923291332104,0.659969574418278
JPEG,0.556088171938907,0.692143053342749,0.711693912480328

```

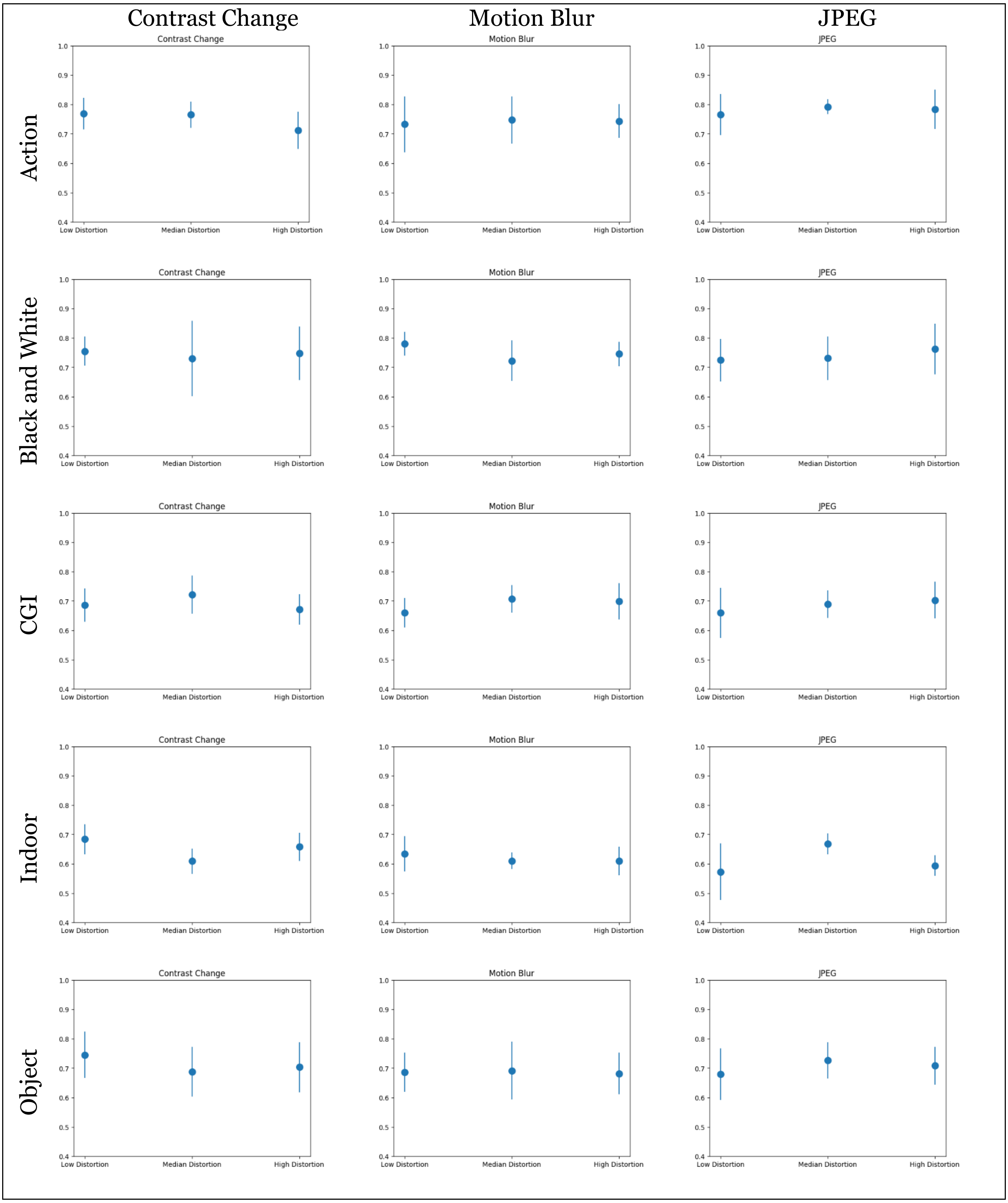
Figure 27: Example content of a similarity matrix file

4.3.3. VISUALISATION

In this section, we shall discuss about creating figures that allows us to understand how similar the distorted versions are to the original images. We shall look at plotting graphs including error bars and an overview of the result through all category. For this section, I shall create the figures using python's matplotlib.pyplot and pandas library.

4.3.3.1. Average Figures per category

Using python's pandas library and matplotlib.pyplot, I have generated different visual figures. I have created plots that average the similarity of the distortions per category. These plots contain error bars with 95% confidence intervals. In figure 28, it shows that Portrait type shots has the smallest error bar


Contrast Change
Motion Blur
JPEG
Action
Black and White
CGI
Indoor
Object
Contrast Change
Motion Blur
JPEG
1.0
0.9
0.8
0.7
0.6
0.5
0.4
Low Distortion
Median Distortion
High Distortion

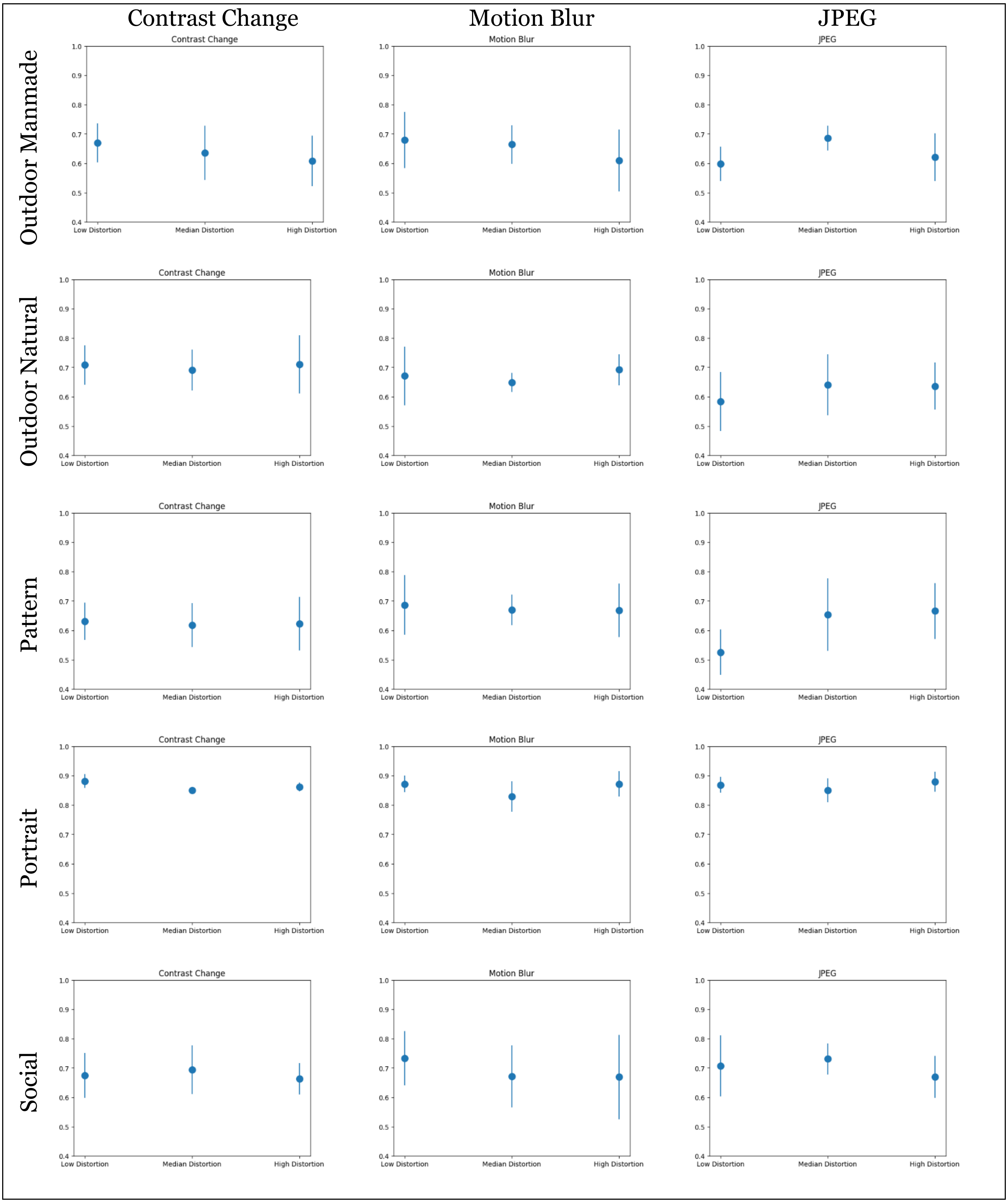


Figure 28: Figures Per Category

4.3.3.2. Average similarity through all images

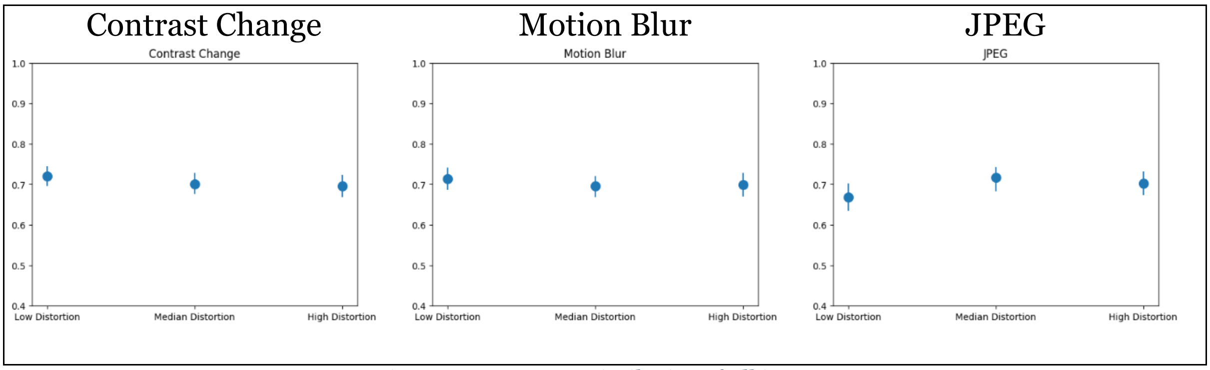


Figure 29: Average similarity of all images

This figure shows the overview of the similarities of the three distortions with their original image across all the categories. Looking at the visual representation, the linear coefficients are ranged between 0.6 and 0.8. The lowest is JPEG compression at low distortion as it is the most pixelated image.

4.3.3.3. Similarity of images per category

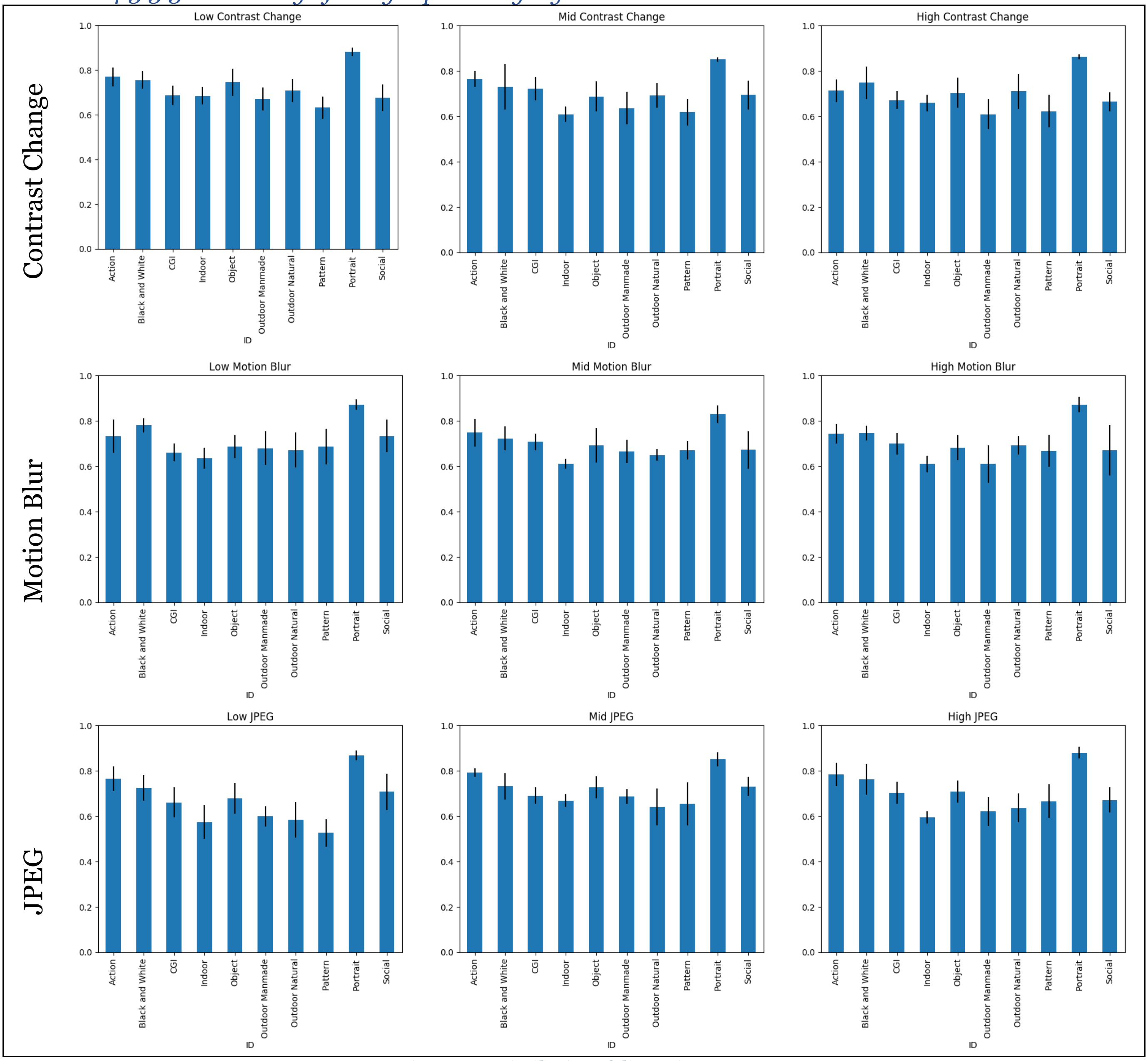


Figure 30: Similarity of distortion per category

This visualisation shows the similarity of the distortions per category and split into different distortion types. Low distortion JPEG compression has the least correlation as it has the lowest bar graphs for most of the categories. I can also conclude that Portrait type shots have the least impact even when distortions are added. This is because humans have been taught to look at certain features of the face.

5. FUTURE IMPROVEMENTS

In this section, I shall discuss about the different improvements that I could have made throughout my implementation. Also, I discuss about further work that can be continued or inspired from this project.

5.1 Consideration of Centre Biased Images

When selecting my images, I discarded centre biased images by eye and used no scientific evidence. With more research and more understanding of saliency, I could have looked at justifying evidence as to how an image is centre biased with an algorithm. This is something that can be considered when selecting appropriate original images for the dataset creation.

5.2 Variety of Distortion Types and Levels

As I have restricted this project to three types and three levels of distortions, we can improve the work by taking looking at more distortion types and levels. With the GUI that I have created in my CUROP project, I am able to use the distortions mentioned in [24] to generate more varieties of distortions. The distortions may or may not have different impacts when viewing the images.

5.3 Saliency Models

Apart from the human data, saliency models can be generated against the dataset that I have constructed. As mentioned in my initial plan, I wanted to look at any differences between how a person would view an original or distorted image against the models. Due to the shortage in time, it was not advisable to meet this plan as a well-planned and strong dataset must be created first.

5.4 Saliency Maps of Distort600

The saliency maps created are from 7 observers per image. As researched and mentioned in 5.2.3.1, I aimed to get 7 observers due to their availability and short length of the project. To create a stronger saliency, more experiments can be done to increase the number of observers per image. A strong eye tracking data collection would have at least 16 observers per image [17].

5.5 Similarity Metrics

As mentioned in 5.3.2.3, there are a variety of similarity metrics that can be used to analyse the saliency maps. And since saliency research itself is evolving, more metrics will be created that may be more informative than the metric used in this project. Using different metrics, I could have more variety of visualisations.

6. Conclusions

As a conclusion of my overall project, I have been able to develop a new dataset consisting of several distortions on different categories. This dataset was used to collect human fixation data which was then analysed for different findings. My main aim is to find any differences on human perception on images that have distorted versions and non-distorted versions. Furthermore, I wanted to see if different categories affect the perception. The process and design has been well thought as I wanted to get the data as reliable as possible. With my findings, I can conclude that there is a slight difference in perception between distorted and non-distorted images. Adding levels of distortions does not affect the perception even more on Colour Contrast and Motion Blur distortions. However, there is a noticeable difference on very pixelated JPEG Compression (low distortion) against high distorted JPEG compression. I can also conclude that Portrait type shots has the least impact on finding a difference in perception across all different distortions in comparison to the original image. Altering the categories of images e.g. Action, Black and White, Indoor does not have much impact on the difference of perception between distorted and non-distorted images.

7. Reflection

This project has had a lot of learning curve as I have been looking at this field ever since my summer internship. During my summer internship, I was still unaware of saliency research and the capabilities of the field. However, due time and more research I was able to appreciate this field of work.

I am rather curious on how saliency models would predict saliencies in my new data set and if given the opportunity I would like to further study and learn more.

One of the essential skills that stood out for me is having patience and careful planning. The structure of my final report has changed so much in comparison to the initial plan. It is true that everything takes longer than we think and that things cannot be rushed. I was eager to compare models against data without thinking much about the data collection. As shown in my initial planning, I was overly ambitious in quickly creating the dataset. However, I now appreciate the careful planning of creating the dataset and I have been seen the importance of the dataset. Having a good foundation is better than a rushed outcome.

I have also learnt the importance of reading and researching more information on algorithms or methods used from previous papers in order to prevent having to 'invent the wheel' and create everything from scratch. What I would to improve on, is to be able to understand carefully other people's work and how I could implement those into my work.

8. References

[1] Ponomarenko.info. (2017). *Nikolay Ponomarenko homepage - TID2013*. [online] Available at: http://www.ponomarenko.info/tid2013.htm

[2] E.C. Larson and D.M. Chandler, "Most Apparent Distortion: Full-Reference Image Quality Assessment and the Role of Strategy," *Journal of Electronic Imaging*, 19 (1), March 2010

[3] Bylynskii, Z., Judd, T., Olivia, A., Torralba, A. and Durand, F. (2016). "What do different evaluation metrics tell us about saliency models?" arXiv preprint arXiv:1604.03605

[4] Tilke, J., Durand, F. and Torralba, A. (2012). "A Benchmark of Computational Models of Saliency to Predict Human Fixations", *MIT Technical Report*

[5] Borji, A. and Itti, L. (2015). CAT2000: "A Large Scale Fixation Dataset for Boosting Saliency Research, *CVPR 2015 workshop on "Future of Datasets""*, arXiv preprint arXiv:1505.03581

[6] Unsplash.com. (2017). *Unsplash | Free High-Resolution Photos*. [online] Available at: https://unsplash.com/

[7] Creativecommons.org. (2017). *Creative Commons — CC0 1.0 Universal*. [online] Available at: https://creativecommons.org/publicdomain/zero/1.0/

[8] Tseng, P., Carmi, R., Cameron, I. G. M., Munoz, D.P. and Itti, L. (2009) "Quantifying center bias of observers in free viewing of dynamic natural scenes", *journal of vision*, 9 (7)

[9] Zhang, W. and Liu, H. (2017) "SIQ288, A Saliency Dataset for Image Quality Research", *Multimedia Signal Processing (MMSP), 2016 IEEE 18th International Workshop*

[10] Ansorge, U., Francis, G., Herzog, M. H. and Ogmen, H. (2008). "Visual masking and the dynamics of human perception, cognition, and consciousness A century of progress, a contemporary synthesis, and future directions.", *Advances in Cognitive Psychology*, 3(1-2): 1-8

[11] Bylinsjki, Z., Isola, P., Bainbridge, C., Torralba, A. and Oliva, A. (2015). "Intrinsic and extrinsic effects on image memorability", *Vision Research*

[12] Ehinger, K.A., Hidalgo-Sotelo, B., Torralba, A. and Oliva, A. (2009). "Modelling search for 900 scenes: A combined source model of eye guidance", *Psychology Press, an imprint of the Taylor & Francis Group, an Informa business*, 17(6/7), 945-978

[13] N. Riche, M. Duvinage, M. Mancas, B. Gosselin, T. Dutoit, "Saliency and human fixations: state-of-the-art and study of comparison metrics", *Proc. of the IEEE Int. Conf. on Comput. Vis.*, pp. 1153-1160, 2013.

[14] Uk.mathworks.com. (2017). *Contrast Enhancement Techniques - MATLAB & Simulink Example - MathWorks United Kingdom*. [online] Available at: https://uk.mathworks.com/help/images/examples/contrast-enhancement-techniques.html

[15] Uk.mathworks.com. (2017). *Create predefined 2-D filter - MATLAB fspecial - MathWorks United Kingdom*. [online] Available at: https://uk.mathworks.com/help/images/ref/fspecial.html

[16] Uk.mathworks.com. (2017). *N-D filtering of multidimensional images - MATLAB imfilter - MathWorks United Kingdom*. [online] Available at: https://uk.mathworks.com/help/images/ref/imfilter.html

[17] Zhang, W. and Liu, H. (2016) "Towards a reliable collection of eye-tracking data for image quality research: challenges, solutions and applications", *IEEE trasnsactions on image processing*

[18] Beta.doodle.com. (2017). *Doodle: Eye Tracking Session*. [online] Available at: https://beta.doodle.com/poll/4wchinwfzkyxm8pw#table

[19] GitHub. (2017). *cvzoya/saliency*. [online] Available at: https://github.com/cvzoya/saliency/tree/master/code_forMetrics

[20] Sites.google.com. (2017). *Evaluation measures - SaliencyEvaluation*. [online] Available at: https://sites.google.com/site/saliencyevaluation/evaluation-measures