Image manipulation and complexity recognised by humans

Xiang Guo

u5541646

A thesis submitted in partial fulfilment of the degree of
Bachelor of Advanced Computing (Honours) at
The Research School of Computing Science
Australian National University

October 2016
I declare that, to the best of my knowledge, this thesis is my own original work and does not contain any material previously published or written by another person except where otherwise indicated

Xiang Guo
28/10/2016
Acknowledgements

I would like to give my best appreciation to all people who provided help to me to finish my academic research and this report.

First, I would like give my thanks to my supervisor, Professor Tom Gedeon, who gave a lot of valuable advice during the whole process of my project. Especially the suggestions of the design of my experiment help me to improve it greatly.

I would also like to give my gratitude to my supervisor, Doctor Sabrina Caldwell, who gave plenty of suggestions on both of my academic research and report writing. Sabrina is always there ready to support me, especially at the last period of the semester when I got really tight in my schedule and I really appreciate it.

Thanks to Haolei Ye, who provide the framework for my experiment and lots of useful information helping me to improve my performance and modify the framework. Also thanks to Zikai Zhao, who gave great ideas on my dataset design. I would like to thanks Chris Chow to provide information about eyetrieb integration as well.

Also thanks to the participants that took part in my experiment providing valuable results to my research. Their participations is the foundation of the success of my project.

Finally thanks to the course convenor, Professor Weifa Liang, who gave a lot of advice about time management, presentation and report writing.
Abstract

This paper describes the experiment conducted to research how people recognise image manipulation and complexity, including an literature survey, developing the experiment’s aim, modification of the tools and prototype, analysis and reconstruction of the database, the design of the experiment methods and procedure, and analysis of final results. In addition, the paper compares the different measurements of image complexity, which are rated by human visualisation and JPEG compression.

The issues caused by manipulated images in a wide range of areas raise the question of how well people can identify image manipulations. According to the related studies, the characteristics of images could influence people’s judgement of image manipulation. Hence, according to the literature survey, the image’s complexity is thought to be a factor which is correlated with image manipulation, which is part of the experiment. The results indicate people is not good at identifying image manipulations, and the accuracy to find out the manipulated image being manipulated is less than 50%. Also the result shows that higher image complexity could decrease people’s abilities to recognise image manipulations.
Contents

Acknowledgements ........................................................................................................................................3

Abstract .........................................................................................................................................................4

1. Introduction ...............................................................................................................................................7

2. Literature Survey

   2.1 Image Manipulation ..............................................................................................................................8
   2.1.1 Issues Caused by Image Manipulation ............................................................................................8
   2.1.2 Previous Work ..................................................................................................................................8
   2.1.3 Manipulation Methods .....................................................................................................................9

   2.2 Image Complexity ................................................................................................................................10

   2.3 Eye Gaze Tracking ...............................................................................................................................11

3. Aim .........................................................................................................................................................13

4. Tools

   4.1 Eye Tribe ..............................................................................................................................................14

   4.2 Mondrian Question Prototype ...........................................................................................................14

5. Database

   5.1 CASIA 2.0 ..........................................................................................................................................15

   5.2 IMCRD ...............................................................................................................................................16

   5.3 Construction of IMCRD .......................................................................................................................16
   5.3.1 Scope and Structure Identification .................................................................................................17

   5.3.2 Analysis of CASIA 2.0 ...................................................................................................................17

   5.3.3 Reconstruction ................................................................................................................................18

6. Method

   6.1 Principles .............................................................................................................................................20

   6.2 Image Pool .........................................................................................................................................21

5 / 38
6.3 Question Pool ................................................................. 22
6.4 Display Algorithm ......................................................... 22
6.5 Experiment Procedure .................................................... 24

7. Results and Discussion

7.1 General Statistics of Results .............................................. 26
7.2 Specific Examples .......................................................... 27
7.3 JPEG Compression VS Human Decisions ............................. 28

8. Conclusion and Future Work

8.1 Conclusion ................................................................. 29
8.2 Future Work ............................................................... 29

References ................................................................. 31

Appendix ................................................................. 32
1. Introduction

Nowadays, digital images have become an important method for people to receive information, like news, advertisements, study and social network website and smartphone applications. In addition, normal people have easy access to image processing tools like Photoshop, gimp and image editing smartphone apps. Therefore, most digital images can easily be manipulated by anyone using one of these tools. These manipulated images could cause issues in many areas of society. This raises the interest in determining how well people recognise image manipulation in their daily life. There are several automatic systems which have been developed to detect the manipulation of images [10]. This also raises interest in comparing the performance of these systems with humans. Hence, it was decided to conduct an experimental study on the human recognition of image manipulation.

According to the related work [1], the image’s characteristics could influence people’s ability to identify manipulation. Hence, image complexity is treated as a variable in this experiment, in order to determine which image characteristics effects manipulation recognition. More details are presented in the Literature Survey section.

In this experiment, each participant will be shown a series of images and asked ‘Do you believe this image is manipulated or not?’ or ‘How would you rate the complexity of this image?’ In addition, the participants’ gaze will be tracked by Eye Tribe and stored in a database for future analysis. Finally, the results of the experiment are analysed and presented in the Results section.

This paper presents a full description of the experiment conducted in this research on image manipulation and complexity associated with gaze tracking technology. The description includes the literature survey, which is the study of related work, the aim of the experiment, the tools that were modified and used to establish and process the experiment, the databases which were analysed and reconstructed, the design of the experiment methods and procedure, and, finally, the analysis of the experiment results.
2. Literature Survey

2.1 Image Manipulation

2.1.1 Issues Caused by Image Manipulation

Our daily lives are filled with digital images. Facebook users upload more than 350 million images each day [3]. The widely available digital image processing software, like Photoshop and smartphone applications, makes it easy for people to manipulate digital images. All 12 of the participants in this experiment stated that they have experience in manipulating images. In such a situation, the digital images around us could potentially present wrong or misleading information which has been proved to cause issues in range of areas including scientific research [2], news [11] and legal evidence.

2.1.2 Previous Work

One of the previous studies that is similar to this project is the research conducted by Caldwell et al. [1]. In her experiment, 80 participants viewed 14 manipulated or unmanipulated images. She focused on a verbal analysis of the participants’ response to questions asked when they were viewing images. Based on the analysis of those verbal responses, she stated that logic could influence the accuracy of the participants’ ability to recognise image manipulation [1]. According to Caldwell, logic is often applied by the participants to help them determine the validity of the images [1]. The logic could include the part of image being pixelated, a different expression to other people in the image and something appearing in an environment where it could not be [1]. However, even with the help of logic, the participants’ ability to accurately identify image manipulation was still low. Caldwell stated that the accuracy rate was only 50.1% [1].

The participants in Caldwell’s experiment were allowed to observe the images closely and think for a relatively long time which enabled them to apply logic to assist in their decision. Thus, logic plays an important role in influencing people’s recognition of image manipulation. In real life, people would not spend a few minute staring at every image they saw and consciously use logic to determine the validity of the images. Hence, in order to test people’s ability to recognise image manipulation in their daily lives, in this experiment the participants will only be able to view each image for 15 seconds. This will ensure that the participants do not have enough time to build up robust logic to support their decisions and will more closely replicate how they view images in their daily lives.

In addition, Caldwell stated that the salience of the image could influence how people perceive it [1]. She states the image’s characteristics can determine one’s eye path on the image to some extent. Harding and Bloj’s work also proved that luminance could attract people’s attention through their investigation of eye movement data [4]. All of
this evidence shows that many characteristics could influence people’s ability to identify manipulation. In addition, it is difficult to obtain acute information on eye movement data as many elements can distract people’s gaze. Caldwell stated in her conclusion that eye tracking data is helpful in association with identifying image manipulation, however, it is not always true [1]. This raise the question, what elements could influence the accuracy of image manipulation recognition and how do they interact with each other? With the conclusion of the literature survey on image complexity (which will be introduced in the next section), it was determined that image complexity would be the second variable in the experiment, apart from the first variable, manipulation. As a ‘start-up’ experiment, image complexity is general enough to cover some of the important characteristics of images which could provide a direction for more specific topics for future work.

2.1.3 Manipulation Methods

There are many manipulation methods from simple to professional. As this project focusses manipulations that could lead to misunderstandings or wrong information which could cause issues, rather than the manipulation and adjusting the saturation to make the images more beautiful and attractive, four simple manipulation methods are used on the images in this experiment.

![Figure 2.1.3.1 Copy](image1)

![Figure 2.1.3.2 Insertion](image2)
The first method (Figure 2.1.3.1) is copy, which is copying some of the elements within the image. The second method (Figure 2.1.3.2) is insertion which is cutting certain elements from other images and inserting them into the image which is being manipulated. The next method is remove, which is removing some of the elements from the image and fitting the removed parts to the environment in the image in order to make it appear that there were no such elements in the image. For example, in Figure 2.1.3.3 the vent on the wall is removed, and its location is manipulated to fit the wall so that it appears as if there was never anything there. The fourth method of manipulation is replacement, which replaces parts of the image with elements of another image as shown in Figure 2.1.3.4.

2.2 Image Complexity

The first question to address when discussing image complexity is the definition of complexity. According to Webster’s dictionary (1986), mentioned by Rigau, Feixas and Sbert, the definition of a complex object is ‘an arrangement of parts, so intricate as to be hard to understand or deal with’ [8]. That is, complexity describes the difficulty of doing something. A different definition of image complexity could be derived from the characteristics of the images this study focuses on. The previous work on image complexity indicates that there are two rough definitions of image complexity, which are identified from human and computer perspectives respectively. From the human perspective, Rigau et al. believes that image complexity is related to the entropy of images, which measures ‘uncertainty, ignorance, surprise, or information’ [8] and these
elements are identified by human’s visual reflections. From the computer’s perspective, Rigau mentioned the book by Hopcroft et al. which states that image complexity can be defined by the amount of computational resources which are needed to process the image to achieve certain goals [8][5].

Different methods are applied to measure image complexity based on the kind of image definition. For the human perspective, Silva et al. proposed measuring the time that people spent observing an image in order to describe it to certain level [9]. Silva also mentioned some other methods in the visual image complexity field, including fractal theory, fuzzy theory and information theory [9]. From the computer perspective, Mulhern and Sawey provide an interesting idea: measure the image complexity using JPEG Compression [7]. Mulhern et al. states that jpeg compression describes ‘subjective measures of image complexity’ in ‘highly detailed and colorized environments’ [7]. For computers, the further a jpeg compression can compress one image, the less complex the image is for computers as it requires less space and computational resources to process.

These studies present several methods for measuring image complexity. However, in this study, due to the limitations on the time that the participants can spend on each image, image complexity will be measured in a very simple way. The measurement method is asking the participants to rate the complexity from 0 (simple) to 100 (complex). In addition to saving time and being simple, this method has another important advantage; it also collects the participants’ opinions of image complexity. Combined with the participants’ judgements of image manipulation, it is possible to determine the correlation between image complexity and manipulation. In addition, it is meaningful to compare the results of human’s decisions with jpeg compression, in order to determine whether jpeg compression could reflect a subjective measure of image complexity. If yes, then there would be an automatic image complexity measurement method.

2.3 Eye Gaze Tracking

Eye gaze tracking is the process of tracking where someone is looking by using sensors to detect their eye gaze. Eye gaze tracking can show useful information about how people view images and what attracts their attention. The eye gaze tracking technology has proved to be useful in image visualisation. For example, in Caldwell’s experiment, the eye gaze data assisted in recognising manipulations [1]. In addition, Itti and Baldi use eye gaze tracking in their surprise model research and they state that over 70% of people tend to lock their gaze onto the areas which is estimated to be more surprising than other areas [6]. According to these studies, eye gaze tracking could be useful in this experiment. As the time the participants spent on each image is limited, they will only focus on the areas of the image that is the most attractive. With eye gaze tracking, we can analysis what elements are more likely to attract the participants and identify the relationship between these elements and the image complexity rating given by the participants. Caldwell’s study also indicates that the eye gaze data could support the
study of image manipulation [1].
3. Aim

The first aim of this experiment is to determine to what extent people can recognise image manipulation when they view images normally. The ‘normally’ in this experiment means within the time limit for the participants to view each image. The second aim is to determine how people measure the image complexity by viewing the images for a certain amount of time and identifying the differences among the other methods of measuring image complexity. The third aim is to analyse whether a correlation exits between image manipulation and complexity. The fourth aim is to record the eye gaze tracking when people view images for future research.
4. Tools

4.1 Eye Tribe

Eye Tribe is the device which is used in this experiment to track and record the participants’ eye gaze while they are viewing images. Eye Tribe is used in this experiment for two reasons. The first one is mobility. The components of Eye Tribe are one small device and one cable which connects the device to the computer. Therefore, the equipment does not need to be operated in a room which is specially prepared for the experiment and it does not require a specific environment to facilitate peak performance. The second reason is that the device relies on infrared illumination [12], which enables the device to function correctly when the participant is wearing glasses.

4.2 Mondrian Question Prototype

The Mondrian Question Prototype is a display framework for the Mondrian Draw Project which was developed by Haolei Ye in Python and Django. This framework was developed to display Mondrian images. In addition, it provide rating functions below the image which allows the participants to rate how they like the image and store the results on the database. The framework provides different rating methods for the participants, like sliding sliders and clicking dots, to research how people interact with these selection methods. In the future, more display methods will added to this framework in order to research the interaction between humans and different kinds of images display and selection interfaces.

As this experiment is about displaying images to people and recording the data collected on each participant, the Mondrian Question Prototype is the right display solution for it. The requirements of this experiment provide some advice for the development of the framework. For example, in this experiment, the participants should answer different questions about manipulation and image complexity, therefore, the framework added a text display function which allowed it to display text on the screen in order to ask different questions of the participants. This function is also useful for Ye’s project as it could display text instructions and ask different kinds of questions in order to observe the interaction between human and the different kinds and forms of questions displayed.

As the framework is just a prototype, it requires two modification prior to its implementation in this experiment. First, Eye Tribe must be integrated with the framework. With the help of the HCC Workshop, which has done this before, the main challenge for this project is to set up the ajax and transfer Eye Tribe every time the participant clicks the ‘next’ button. As the framework does not have the control, a display algorithm should be developed to perform the role of control in order to decide which image will be displayed, what question will be asked and which type of selection method will be shown. The Design section will present more detailed information on the display algorithm.
5. Database

5.1 CASIA 2.0

CASIA 2.0 is a collection of natural colour images and images manipulated with the four methods which were introduced in Literature Survey. This dataset was created by Dong, J. et al. for research on the detection of image manipulation [13]. The dataset contains two documents; Au and Tp. Au contains 7491 authentic images while Tp contains images from Au altered using the four manipulation methods. The format of the files are JPEG, BMP and TIFF. The files in BMF and TIFF provide images in an uncompressed format which provides samples for this experiment to compare the human and JPEG measurements of image complexity.

The names of the images in Au and Tp are formatted to contain the image’s information, enabling the analysis in the following paper. For images in Au, the name format is (Content_Category)_(File Index) [13]. The content_category indicates which category (will be explained in next part) the image belongs to and the Index labels the image by giving the image a unique number. In Tp, the name format is (Operation Type)_(Image source)_(Tampered Region Size)_(Post-processing)_(Source Image Index1)_(Source Image Index2)_(File Index) [13]. The most useful parts are Source Image Index1 and Source Image Index2 indicating the two authentic images that were used to generate this tampered image.

All of the images are divided into nine categories according to the image content, which are animal, architecture, article, character, indoor, nature, plant, scene and texture. In order to separate these images into each category and perform a statistical analysis, some code work needs to be done. A simple program was written in Python, for more information on this code work, similar works are introduced in the IMCRD section. As one tampered image could use two or more authentic images, there is some overlap between the categories. Therefore, if one tampered image used parts from two images which belonged to categories A and B respectively, then the tampered image would appear in both categories. Consequently, the sum of the tampered images in each category would larger than actual number of tampered images. According to Table 5.1, the tex (texture) category contains the least images, which is only about 2.68% in total. Another category with only a small percentage is ind (indoor). The remaining categories account for almost equal percentages which range between 11% and 15%.

<table>
<thead>
<tr>
<th>Category</th>
<th>Au</th>
<th>Percentage</th>
<th>Tp</th>
<th>Percentage</th>
<th>Total Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>ani(animal)</td>
<td>1094</td>
<td>14.60%</td>
<td>810</td>
<td>13.29%</td>
<td>14.02%</td>
</tr>
<tr>
<td>arc(architecture)</td>
<td>1075</td>
<td>14.35%</td>
<td>652</td>
<td>10.70%</td>
<td>12.71%</td>
</tr>
<tr>
<td>art(article)</td>
<td>913</td>
<td>12.19%</td>
<td>751</td>
<td>12.33%</td>
<td>12.25%</td>
</tr>
<tr>
<td>cha(character)</td>
<td>980</td>
<td>13.08%</td>
<td>917</td>
<td>15.05%</td>
<td>13.96%</td>
</tr>
<tr>
<td>Category</td>
<td>Valuable Images</td>
<td>IMCRD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>----------------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Au</td>
<td>Percentage</td>
<td>Original Images</td>
<td>Percentages</td>
<td></td>
</tr>
<tr>
<td>ani (animal)</td>
<td>280</td>
<td>14.48%</td>
<td>46</td>
<td>15.33%</td>
<td></td>
</tr>
<tr>
<td>arc (architecture)</td>
<td>224</td>
<td>11.58%</td>
<td>37</td>
<td>12.33%</td>
<td></td>
</tr>
<tr>
<td>art (article)</td>
<td>224</td>
<td>11.58%</td>
<td>37</td>
<td>12.33%</td>
<td></td>
</tr>
<tr>
<td>cha (character)</td>
<td>250</td>
<td>12.93%</td>
<td>41</td>
<td>13.67%</td>
<td></td>
</tr>
<tr>
<td>ind (indoor)</td>
<td>171</td>
<td>8.84%</td>
<td>25</td>
<td>8.33%</td>
<td></td>
</tr>
<tr>
<td>nat (nature)</td>
<td>267</td>
<td>13.81%</td>
<td>39</td>
<td>13.00%</td>
<td></td>
</tr>
<tr>
<td>pla (plant)</td>
<td>215</td>
<td>11.12%</td>
<td>31</td>
<td>10.33%</td>
<td></td>
</tr>
<tr>
<td>sec (scene)</td>
<td>213</td>
<td>11.01%</td>
<td>31</td>
<td>10.33%</td>
<td></td>
</tr>
<tr>
<td>txt (texture)</td>
<td>90</td>
<td>4.65%</td>
<td>13</td>
<td>4.33%</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1934</td>
<td>100.00%</td>
<td>300</td>
<td>100.00%</td>
<td></td>
</tr>
</tbody>
</table>

5.2 IMCRD

The Image Manipulation and Complexity Recognition Dataset (IMCRD), which was formed from CASIA, is used as the dataset for this experiment. IMCRD consists of 300 different pairs of images which were selected from CASIA. Each pair of images in IMCRD contains one original image which has not been manipulated and one corresponding manipulated image, which is the manipulated version of the original image in the same pair. The original images in IMCRD cover all of the categories in CASIA, and the number of images in each category and their percentages is shown in Table 5.2. It is evident that each category contains an almost equal number of images, except the texture and indoor categories which are much smaller. In addition, the methods of manipulation used on the manipulated images in IMCRD cover all of the methods mentioned in CASIA. Therefore, there is no significant emphasis on the differences among all categories.

5.3 Construction of IMCRD

The construction of IMCRD included three main steps; are scope and structure
identification, analysis of CASIA and reconstruction. Scope and structure identification includes determining the number of images and the structure of the dataset. The analysis of CASIA is to check the validity of the images in the dataset and conduct a statistical analysis. Finally, reconstruction includes the selection of 300 pairs of images from CASIA, and renaming and reforming the files.

5.3.1 Scope and Structure Identification

To form the IMCRD, the number of images and the structure of the dataset should be decided. The previous work on this topic, which was conducted by Sabrina Caldwell, provides some references for these decisions. Caldwell stated that her experiment used 14 manipulated and unmanipulated images [1]. There are two pairs of corresponding original and manipulated images and the participants’ reactions to these pairs of images contributed to interesting conclusions on eye gaze tracking [1]. To identify how well people can recognise manipulation, it is necessary to compare their eye gaze on the original and the corresponding manipulated images. The reason is that the comparison could eliminate the effect from the image itself on image manipulation and complexity recognition and potentially reveal some important facts like Caldwell’s work.

However, this experiment differs from Caldwell’s in terms of the amount of images used. In Caldwell’s work, the grounded theory she used allowed her to conduct a detailed analysis of each reflection of the participants on every image [1]. This indicates that her strategy is to dig deeply and be specific, therefore, 14 was a suitable number of images for her experiment. However, this experiment include two elements of images, which are manipulation and complexity. Compared with 14 images, a large amount of samples could potentially contain data which could indicate the relationship between these two elements and have the ability to provide general and solid evidence to support any conclusions which are drawn from this experiment. Consequently, with the combination of participant estimation and method design, which will be discussed in Method section, the number of images in IMCRD is set to 300 pairs. Initially, the number of images was set to 600 pairs. However, Zikai Zhao, who is conducting the Mondrian Draw project, which also involves human visualisation of images, suggested halving the number of images. His experience suggested that people would become impatient if presented with over 200 images. Zhao stated that only one participant had the patience to view over 200 images and the time he spent viewing each image had been reduced by approximately 50% by the end of the experiment. Compared to the Mondrian images, which only include straight lines and four colours filling the blocks, viewed by the participants in Zhao’s experiment, the real photos in CASIA involve more details and variations which could attract people’s interest longer. Hence, 300 pairs (each participant will view 300 images) is a reasonable number for IMCRD.

5.3.2 Analysis of CASIA 2.0

The next step to forming the IMCRD is to analyse the source database CASIA 2.0. The structure of the file name which was mentioned in the last section makes it possible to
check how many corresponding manipulated images in the Tp document each original image in the Au document has, which will be called the Matching Check. The Matching Check was done in Python as it provides a convenient method, ‘os.walk’, for walking the document, and reading and storing file names. However, prior to conducting the Matching Check, all of the files’ names should be checked to determine whether all of them actually fit the structure in order to increase the accuracy of the Matching Check, this process is called the Name Validation Check. Three file names were found to differ from the name structure that was defined by the database. The core of the Matching Check is splitting and recombining the file name. To identify whether an image in the Tp document is the corresponding manipulated image of an original image in Au, the category names and corresponding index numbers should be matched. To do this, Python provides another convenient tool, the string split and combination method, to split file names by ‘_’ and form the category name and the corresponding index standard form by recombination.

![Figure 5.3.2 Statistics of Corresponding Manipulated and Original Images](image)

According to **Figure 5.3.2** and **Table 5.1**, 5539 out of 7491 original images in CASIA do not have corresponding manipulated images, therefore, they were not involved in manipulation. For the other 1952 images, which are called valuable images in this paper, parts from each of them appear in at least one manipulated image. In addition, **Table 5.2** shows the category statistics of these 1952 (valuable) images, which indicates that there is no preference for selecting images from certain categories to generate the manipulated images.

### 5.3.3 Reconstruction

The 300 original images were selected from the valuable images mentioned above. The Matching Check also helps to select these images automatically by copying the original
images in another document when their corresponding manipulated images are detected. To compress 1952 images into 300, every sixth image was selected to form the draft of the original images of IMCRD. Three principles were set to check the validity of these 300 images.

- These 300 images are different from each other.
- The content of images in same category should be different when the manipulation methods are same.
- The manipulation methods should be different when the content of the images is similar.

Over 20 pairs of original images were found to be identical but with different file names. The second principle aims to identify the effect of one manipulation method on images with different content. The third principle could provide comparable results of different manipulation methods on similar image content.

The 300 manipulated images were selected manually. There were two principles for selecting the corresponding manipulated images.

- The original image should be the main body of the corresponding manipulated image.
- When one original image has multiple corresponding manipulated images, select one randomly.

The final step to form the ICMRD is to rename and reformatted the files to achieve the consistency of images for this experiment. The file name form is “Index_Type_OrignalName”. The Index is an integer number from 001 to 300 labelling the 300 pairs of images. Type could be O for original images and M for manipulated images. OrignalName is the original file name which could retain the original information from CASIA. The underscores ‘_’ in the original name were converted into dash ‘-‘, so that OrignalName could be separated from the Index and Type by underscores as a whole part. All images were reformatted into JPEG to compress the size of the image due to the limitations of the display system for experimental usage and to achieve consistency in the format of the images in the dataset. The renaming and reformattion was done with the aid of irfanview, which is provide powerful batch function for image operations. All images are attached in the Appendix.
6. Method

6.1 Principles

In this experiment, the participants will view images in IMCRD displayed by the framework. There will be two questions for the participants to answer about the images. The first one is about manipulation, which is ‘Do you believe this image is manipulated or not?’ The second one is about complexity, which is ‘How would you rate the complexity of this image?’ To ensure the results of the experiment can achieve the aims and provide solid evidence to support the conclusion, a series of principles should be developed to facilitate a scientific and rigorous methodology. These six principles are:

1. Each participant should finish the experiment individually without references from outside.

2. Each participant needs to answer exactly one question for each image which they view.

3. Each participant views exactly one image from each pair from IMCRD.

4. Each participant views an equal number of original and manipulated images.

5. Each participant answers an equal number of questions about manipulation and complexity on both original images and manipulated images.

6. Both images in each pair in IMCRD are viewed equal times at the end of the experiment.

7. The manipulation and complexity questions are answered equal times for each image.

As the aim is to determine to what extent humans can identify manipulation of an image, no participants should view both images from one pair from the IMCRD. The comparison of the original and corresponding manipulated image could increase the possibility of the participant identifying the manipulation.

As each participant needs to view 300 images, the participants may give up or be less focused on the images due to the long time period of the experiment. Hence, each image will be matched with one question to shorten the exam time, and enable the participant to focus on the only question for each image.

Because each participant in this experiment is a unique individual, the experiment should collect their ideas on manipulation and complexity on both original and manipulated images equally to ensure each participant makes the same contribution to each aspect of the results.

As the comparison of the participants’ ideas on the original and manipulated images is
an important part of this experiment, the total views on both images in each pair should be equal. In addition, the comparison of the results of image manipulation and complexity is another important part of the experiment, therefore, the total number of questions matched with each image should be divided equally into manipulation and complexity.

6.2 Image Pool

According to the third and sixth principles in 6.1, the number of participants should be divisible by two. Furthermore, when principle 7 is taken into consideration, the number of participants should be divisible by four. Due to the tight schedule for this experiment, it was determined that 12 participants would be used.

Figure 6.2 shows the design of the images pools and question arrangements for each participant. Let P1, P2, … P12 indicate 12 participants, and (O001, M001), (O002, M002), … (O300, M300) indicate 300 pairs of images, while O indicates original image, and M indicates manipulated image. To satisfy the principles:

P1 will view: O001 – O150, M151 – M300

P2 will view: M001 – M025, O026 – O175, M176 – M300

P3 will view: M001 – M050, O051 – O200, M201 – M300

…

P12 will view: O001 – O125, M126 – M275, O276 – O300

This is presented in Figure 6.2. Locate all original images of IMCRD in one row and the corresponding manipulated image in the second row in the same sequence of first
row. These two rows will form one of the twelve big squares in Figure 6.2 which indicates a copy of IMCRD. The smaller blocks represent 25 images in a row. Each participant will view the images in the red or blue blocks but not those in the white blocks. The images in the red blocks will be asked questions about manipulation and the blue blocks will be asked questions about complexity.

Figure 6.2 shows that each participant will view 50 images which are different from the neighbouring participant and no participant has the exact same image pool. This will allow each participant view some images which were viewed by other which could provide comparative results from different views on the same image. In addition, each set of 25 images will be viewed by a different group of participants to decrease the effect of individual differentiation.

In addition, according to Figure 6.2, each set of 25 images will not be attached to the same type of questions as the nearby blocks, which is due to sequence of the images in the dataset. In IMCRD, the images in same category are located together. Attaching different types of questions to neighbouring block could prevent all of the images in one category being presented with the same questions.

6.3 Questions Pool

There are two question pools for this experiment, which are the Basic Question Pool (BQP) and the MC (manipulation and complexity) Question Pool (MCQP). The BQP aims to collect basic information on the participant which may influence their perception of the images and for convenience of data analysis. For instance, the names of the participants could be used to label the results, including their answers to the questions in addition to the images and Eye Tribe data. The other questions in the BQP are regarding their gender, age, department, and image manipulation skills. The differences in this background information could probably influence people’s ability to identify manipulation and their opinions on image complexity, especially their department and manipulation skills. A participant’s educational background may influence how they view images. For instance, art students may recognise manipulation due to luminance differences, while computer students may focus on the logical issues in images. In addition, the level of manipulation skills could definitely play an important role in manipulation recognition.

The MCQP includes two questions, which are ‘Do you believe this image is manipulated or not?’ and ‘How would you rate the complexity of this image?’ In the framework there are six options for participants to select in response to the first question which include Yes, Not Sure and No. There is an even number of options to prevent the participants sitting on the fence regarding the decisions. For the second question, there is a slider for participants which allows them to rate the complexity from 0 (very simple) to 100 (very complex).

6.4 Display Algorithm
To display the images in the image pools to the corresponding participants, the algorithm generating image list applied by Python was designed and modified for the framework to enable this function. This algorithm selects the images out of the IMCRD according to the design of the image pool and the index of participant for each object. In addition, the algorithm attaches the corresponding question, selection type (if the participant answers the question by choosing a number or sliding the slider) and the labels of the selection type. Figure 6.2 shows that for the first half of the participants, the index of the original images for each participant will increase by 25 compared to the participant immediately above. For the second half of the participants, the index of the manipulated images will add 25 for the next participant.

Inputs: 
L: a list, contains the paths of all images in IMCRD
Par: an integer which is the index of the participant

Output: 
L_Im, L_Qu, L_Se, L_La: lists, which respectively contain images (that will be viewed by this participant), questions (that match to each image in L_Im), selection types (that indicate which selection type will be provided to each questions in L_Qu), and labels (that provide labels for each selection type in L_Se).

Variables: 
In: integer, the index part of the image name, indicates which pair the image belongs to.
Ty: character, the ‘O/M’ part of the image name, indicates the type of image

DisplayAlgorithm (L, Par):
1. For image in L:
2. 
3. In = image.Index
4. Ty = image.Type
5. if par <= 6:
6. 
7. if (par – 1) * 25 < In <= ((par – 1) * 25 + 150):
8. 
9. else:
10. 
11. if Ty == ‘O’:
12. 
13. L_Im.append(image)
14. 
15. else:
16. 
17. if Ty == ‘M’:
18. 
19. L_Im.append(image)
20. 
21. else:
22. 
23. if (par – 7) * 25 < In <= ((par – 7) * 25 + 150):
24. 
25. else:
26. 
27. if Ty == ‘M’:
28. 
29. L_Im.append(image)
30. 
31. else:
32. 
33. if Ty == 'O':
34. 
35. L_Im.append(image)
36. 
37. Random.shuffle(L_Im)
19. for x in range(0, 300):
20.     In = L_Im [x].Index
21.     Ty = L_Im [x].Type
22.     if par%2 == 1:
23.         if 0 < n <= 25 or 50 < n <= 75 or 100 < n <= 125 or 150 < n <= 175 or 200 < n <= 225
24.             or 250 < n <= 275:
25.             L_Se.append (2)
26.             L_Qu.append ("Do you believe this image is manipulated or not?")
27.             L_La.append (["Yes", 'Not Sure', 'No'])
28.     else:
29.             L_Se.append (4)
30.             L_Qu.append ("How would you rate the complexity of this image?")
31.             L_La.append (["Very Simple", 'Medium', 'Very Complex'])
32. else:
33.         if 0 < n <= 25 or 50 < n <= 75 or 100 < n <= 125 or 150 < n <= 175 or 200 < n <= 225
34.             or 250 < n <= 275:
35.             L_Se.append (4)
36.             L_Qu.append ("How would you rate the complexity of this image?")
37.             L_La.append (["Very Simple", 'Medium', 'Very Complex'])
38.     else:
39.             L_Se.append (2)
40.             L_Qu.append ("Do you believe this image is manipulated or not?")
41.             L_La.append (["Yes", 'Not Sure', 'No'])
42. return (L_Im, L_Qu, L_Se, L_La)

Due to the regular design of the image pool, the display algorithm is simple and its time complexity is $O(n)$.

6.5. Experiment Procedure

At the start of the experiment, the EyeTribe will be set up and the Calibration function will be executed in order to optimise the eye gaze tracking performance. Second, the participant will read the instructions on the first part of the experiment, which is the collection of basic information. The framework will display all of the basic questions in the BQP one by one when the participants click next button. If the participants are unwilling to provide the corresponding information then they may skip the question by selecting the next button without answering.

When the first part is finished, the answers provided by the participants will be stored to the database and the system will jump to the second part of the experiment. The participants need to read the instructions and click the next button to begin when they are ready. In the second part, the system will display the images generated by the display algorithm one by one. For each image, there will be a question from the MCQP which will be matched to the current image by the display algorithm shown under the current image for the participants to answer. This question must be answered before proceeding.
to the next image. The participants will need to slide the slider to rate the complexity or choose ‘Yes,’ ‘No,’ or ‘Not sure’ to indicate whether the image has been manipulated. The ‘next’ button and the function for answering the question will only be enabled after the image has been displayed for three seconds. This aims to let the participants view each image for at least three seconds before they answer the question and proceed to the next image. However, the participants are not expected to spend too much time viewing one image, as the experiment is investigating how people view images in their daily lives. If the participants spend more than 15 seconds on one image, they will be reminded to answer the question and proceed to the next image. After they view all of the images prepared for them, the data will be stored to the database and the experiment is concluded.
7. Results and Discussion

7.1 General Statistics of Results

There were 12 participants in this experiment. All of these participants are male, and their average age is 21.67. They have the same educational background, which is a bachelor of advanced computing. In terms of their manipulation skills, two of them rate their skills as ‘not at all,’ which means they do not have manipulation experience. Four selected ‘A little,’ which means they have little experience of image manipulation, while five selected a higher manipulation skill level, which is ‘general.’ The remaining participant rated their skills as ‘Professional’ which is the highest level people can choose in this experiment.

<table>
<thead>
<tr>
<th>Table 7.1 Overview of Manipulation Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image Type</td>
</tr>
<tr>
<td>Accuracy</td>
</tr>
<tr>
<td>Number of images</td>
</tr>
<tr>
<td>Average of image complexity</td>
</tr>
<tr>
<td>Mean Accuracy</td>
</tr>
</tbody>
</table>

In Table 7.1, Accuracy indicates the percentage of participants to correctly answered the questions ‘Do you believe this image had been manipulated or not?’ That is, the percentage of participants that correctly identified whether the image was the original or had been manipulated. The next row indicates the number of images which were accurately identified. Take the first column for example, the numbers indicate that there were 69 images which nobody could accurately identify as manipulated or not.

The table shows that the situation is similar to Caldwell’s results [1], people are better at identifying the original images as being unmanipulated (62.59%) than at identifying manipulated images that have being manipulated (48.78%). In addition, with a larger sample and database, the accuracy (48.78%) with which people recognise manipulation is lower than in Caldwell’s research (50.1%).

In addition, it is evident that when the accuracy decreases, the image complexity increases, especially for manipulated images. This may indicate that, the higher image complexity increases the difficulty of recognising manipulations.
Figure 7.1 Distribution of Image Complexity

Figure 7.1 shows the distribution of image complexity. From Figure 7.1, it is clear that most of the images have medium complexity. Another interesting factor which this figure shows is that manipulation decreases the image complexity rather than increasing it as was assumed before. In Figure 7.1, the number of manipulated images is almost larger than the number of original images when the image complexity is smaller than 50, and smaller than the original images when image complexity is larger than 50, which indicates that manipulation decreases image complexity.

7.2 Specific Examples

The two images in Figure 7.2 are a typical example for this experiment. Both images used the copy method to copy one part of the image itself. As the image on the left is relatively more complex, it is more difficult for the participant to determine that the same section of flowers appears in three places. In the right image, the copied part is evident as there is little to distract attention and thus it becomes more obvious. The accuracy for both images support the theory that higher image complexity increases the
difficulty of recognising manipulation.

7.3 JPEG Compression VS Human Decisions

In order to compare the JPEG Compression measurement and human decision measurement of image complexity, 201 non-jpeg format images were compressed by JPEG Compression and the compression ratio were calculated. The simpler image for JPEG Compression will have a bigger compression ratio, while the more complex image will be rated by a big number by human. To conveniently compare two methods, the reciprocal of the compression ratio for each image is calculated and scaled up by 50 times to match the value of the image complexity rated by humans. The results are shown in Figure 7.3

According to Figure 7.3, human decisions experience a much larger variance than JPEG compression. While the curve of the JPEG compression matches the trend of human decisions in some small waves, the three highest and most significant peaks do not match and two of them are contrary to the trend of the human decisions’ curve. This indicates that the JPEG compression does not always work as humans do. Therefore, this method cannot represent the human’s opinions of visual complexity, however, it can still accurately describe the image complexity in terms of computational resources.
8. Conclusion and Future Work

8.1 Conclusion

Compared with Caldwell’s work [1], this project utilised a much larger sample which enabled it to generate additional data and consider the characteristics of the image (image complexity). The results show that image manipulation could decrease the image complexity, and a high complexity level could reduce people’s ability to recognise image manipulation.

For image complexity itself, participants tended to rate images with a medium complexity value. This may be due to the participants’ differing opinions on the complexity of the same image (one rated it low complexity, while another rated it high complexity) which is supported by the experiment data.

The comparison between the JPEG compression measurement and human decision measurement of image complexity shows that while the two methods have some commonalities, they also had some notable differences.

During the project’s development, it provided a lot of useful feedback for the Mondrian Question Prototype. Furthermore, the work done on CASIA 2.0 and the information extracted could be a useful addition to the dataset, which is a well-known reference. Finally, the results of the experiment are useful to help the studies of image manipulation detecting software assess their software’s performance with humans.

8.2 Future Work

The most important region of future development is analysis of the eye gaze data. Due to the time limitations of this project, there was insufficient time to analyse the eye gaze data collected from each participant. However, it can be estimated that with the assistance of eye gaze data, people would have a greater chance of identifying image manipulation and the study of eye gaze data could give a clearer and more precise conclusion on the correlation between manipulation and image complexity.

Another important future task is improving the design of the experiment, which was also caused by the limitation of time. For example, the function of Eye Tribe should be improved to enable framework to store the eye gaze data for each image individually, rather than storing the data for each participant in one file, to simplify the analysis process. In addition, other details, like the sequence of image presentation and image scaling, should be considered further to improve the user experience. Finally, participants from a larger area with different educational backgrounds, age and manipulation skills should be involved in this experiment to increase the diversity of the sample.

With the identification of the correlation between image manipulation and image
complexity and the study of how humans interact with image complexity through eye
gaze analysis, the image complexity could be abstracted into more specific definitions
of image characteristics and the elements of images which influence people’s ability to
recognise image manipulation could be identified.

In addition, the results of this experiment could act as a reference for the studies of
image manipulation detecting software development, which could be used as standard
to determine how effective the manipulation detection software is compared with
humans.
References


[12] https://theeyetribe.com/

1. Appendix

PROJECT DESCRIPTION:
- Literature survey
- Construct a software tool to collect responses to manipulated images and image complexity
- Extend tool for display and recording of eye gaze of subjects while viewing selected images
  - each image with identified areas of interest from specifications file
  - interface with Eye Tribe eye gaze device for eye gaze location and pupil dilation signals
- Analysis of image manipulation database and reconstruction.
- Design, conduct and evaluate experiments to detect subject responses to manipulated images and image complexity
- Optional task: extend experiments to different kinds of manipulated images
  - differentiate between additions / omissions / modifications / colour changes and so on
- Integrate all code into HCC Workshop tool
- Write report

ASSESSMENT (as per course’s project rules web page, with the differences noted below):

<table>
<thead>
<tr>
<th>Assessed project components:</th>
<th>% of mark</th>
<th>Due date</th>
<th>Evaluated by:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report: name style: ___________ (e.g. research report, software description...)</td>
<td>45</td>
<td></td>
<td>Leana Copeland</td>
</tr>
<tr>
<td>Artefact: name kind: ___________ (e.g. software, user interface, robot...)</td>
<td>45</td>
<td></td>
<td>Tom Gedeon</td>
</tr>
<tr>
<td>Presentation:</td>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

32 / 38
INDEPENDENT STUDY CONTRACT

Note: Enrolment is subject to approval by the projects co-ordinator

SECTION A (Students and Supervisors)

Unit ID: u5541646
Surname: Guo
First Names: Xiang
Project Supervisor: Tom Gedeon, Sabrina Caldwell
Course Supervisor (A.R.Y academic):
Course Code, Title and Unit: COMP4560

Semester: S2 Year: 2016  S1 Year: 2017
Project Title: Image manipulation and complexity recognition by humans

Learning Objectives:
Experience with construction of advanced experiment software interface
Experience with design of human centred computing experiments
Experience with data analysis

Project Description:
• Literature survey
• Construct a software tool to collect responses to manipulated images and image complexity
• Extend tool for display and recording of eye gaze of subjects while viewing selected images
  — each image with identified areas of interest from specifications file
  — Interface with Eye Tribe eye gaze device for eye gaze location and pupil dilation signals
• Analysis of image manipulation database and reconstruction.
• Design, conduct and evaluate experiments to detect subject responses to manipulated images and image complexity
• Optional task: extend experiments to different kinds of manipulated images
  — Differentiate between additions, omissions, modifications, and colour changes and so on
• Integrate all code into HCC Workshop tool
• Write report

Research School of Computer Science  Form updated Jan-12
ASSESSMENT (as per course’s project rules web page, with the differences noted below):

<table>
<thead>
<tr>
<th>Assessed project components:</th>
<th>% of mark</th>
<th>Due date</th>
<th>Evaluated by:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report: name style:</td>
<td>45</td>
<td></td>
<td>Leona Copeland</td>
</tr>
<tr>
<td>(e.g. research report, software description...)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Artefact: name kind:</td>
<td>45</td>
<td></td>
<td>Tom Gedeon</td>
</tr>
<tr>
<td>(e.g. software, user interface, robot...)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presentation:</td>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MEETING DATES (IF KNOWN):

Weekly

STUDENT DECLARATION: I agree to fulfil the above defined contract:

………………………………………………………………………………... 13 July 2016 ……..

Signature
Date

SECTION B (Supervisor):

I am willing to supervise and support this project. I have checked the student’s academic record and believe this student can complete the project.

………………………………………………………………………………... 13 July 2016 ……..

Signature
Date

REQUIRED DEPARTMENT RESOURCES:

SECTION C (Course coordinator approval)

…………………………………………………………………………………………………

Signature
Date

SECTION D (Projects coordinator approval)

…………………………………………………………………………………………………

Signature
Date

Research School of Computer Science

Form updated Jun-12
3. Appendix

The framework from Haolei Ye initially is held on pythonanywhere (https://www.pythonanywhere.com/) a platform that has already installed Python. And the version of Python that the framework use is Python 3.5. And finally, it is integrated into HCC Workshop. The work I did in this framework mainly locate in view.py

Other code works I did are also in Python 3 and I only run them on PC with windows 10.

For the detailed information of each code work, please look at the next appendix, which contains the information of which parts are external resource, which parts are work I did myself. It also include the description for each code work.
4. Appendix

All codes in files contained by Other Codes, clamed here are my own work

This python code is firstly used to count how many images for each kind of image form. And then it is used to separate images in CASIA into different category according to the information in their names and copy them in corresponding categories.

This python code is used to check whether all files' names in the documents of CASIA are fit to the formula of the name standard. The input of the function getFilename() is the path of the document that contains images, and print file name which is not fitted to the standard formula.

This python code has two functions to check whether some parts from one file name is in the other name. For example:

name A: a1_n1; name B: a2n2_a3n3.

origCheck() is to check whether a1n1 is equal to a2n2
origFullCheck() is to check whether a1n1 is equal to a2n2 or a3n3

Both inputs for these two functions are two file names and return True or False.
The reason why the function works like this is because of the name structure of authentic and tampered images in CASIA.
For example one authentic image's name is '001_O_Au-ani00003', the tampered version of this image's name is '001_M_Tp-S-NNN-S-ani00003-ani00003-00834'. If a1n1 equals to a2n2 means that these two images could in a pair, one is authentic version, another one is tampered version of this authentic images. If a1n1 equals to a3n3, means one small part of this authentic image is used to generate the tampered image, which indicate even the authentic is used to generate the tampered image, they are not a pair of image.

This python code uses functions in namecheck and OriginalCheck.py to separate the authentic images in Au
document by
whether they have the corresponding tampered version (or whether they are used to generate tampered images) and
them into corresponding documents.

nameform.py
This is the display algorithm I introduced in the paper. This algorithm is used to pick the images from dataset and
attach corresponding questions for each participants according to the design of image pool.

result.py
This python code is used to process the results gather from participants. As the framework is used for other purpose
and
I don't have enough to change the format that the framework use to store data in database, the data of the results are
stored in weird structure. I have to read the data from text file and process them with a series of split, recombining...
to form the structure I wanted. Then I integrate 12 results from participant into two list. Two list contains results of
manipulation for each image respectively.

Also, this code contain the data analysis functions which generated the data for table 7.1

analysis.py
This python code is used to analyse the results of complexity and form the data that generate the Figure 7.1

readsize.py
This python code is used to read the file sizes from the original image and the size of corresponding compressed
version
which is compressed JPEG compression by iifanview. Meanwhile, the function calculate the reciprocal of
compression rate
times 50, and pair with the corresponding results from human decision and finally form the data that generate
Figure 7.3

For the code in Framework document
This framework is done by Haolei Ye, u5870415

The ajaxsetup code in COMP4540\selection\templates\akatsuki\index.html and
COMP4540/selection/templates/shimakaze
and eyetrie.js COMP4540/selection/static/javascript are from HCC Workshop in Gitlab
https://gitlab.anu.edu.au/HCC/workshop

My own code work mainly locates in COMP4540/selection/views.py, the display algorithm part.