Initial Plan
Image Forgery Detection

CM3203 Large One Term Individual Project
40 Credits

By Joshua Head
Supervised by Dr Yukun Lai

30th January, 2015

1 Project Description

We live in a world where images play a larger role than ever in our society. With the advent of digital systems and ever increasing hardware memory, photography is no longer an arduous task requiring careful planning or expensive equipment. Images can be previewed instantly after being captured, and can be sent to the other side of the world in a fraction of a second. A huge array of applications are available to modify these images, from simple smartphone applications that apply filters to huge photography suites that can do everything from remove objects to completely change the meaning of the image. Tools and algorithms within these applications have become so advanced that it is often very difficult to detect whether an image has been manipulated or not.

This causes a number of issues in understanding the trustworthiness of an image, as the absolute confidence that we once had in photographic proof has now been lost. Images cannot always be used as concrete evidence, such as in the case of CCTV footage of a suspect committing a crime; if the images cannot be proven to be genuine then they simply cannot be used in a trial. This produces a very real problem in that many criminal cases rely on such images in court as evidence, if they cannot be trusted then the credibility of the case is put in to dispute.

Wider social issues also arise from this problem. Magazines often use manipulated images of celebrities, achieving everything from wrinkle removal to increased muscle definition. These images often portray an unrealistic representation of an ideal body, which leads to decreased body satisfaction and can increase the risk of eating disorders such as anorexia and bulimia. Forged images are often seen in the media by the general public as truth, sometimes even being published by the media with the absolute belief that they are trustworthy when they are in fact not.

Image forgery detection techniques have emerged in order to verify the integrity of the contents of an image, and to prevent forgery using various processing techniques. This project aims to investigate these image forgery detection techniques, evaluating their effectiveness in handling various kinds of image forgeries and manipulation. Ensuring the trustworthiness of an image and regaining confidence in published or presented material is a key issue and of utmost importance in modern society.

2 Project Aims and Objectives

- Implementing a usable prototype that demonstrates the various available methods of image forgery detection. These include forgery detection based upon:
  - JPEG Compression Quantization and Compression Artefacts
    As a discrete cosine transform is applied on 8x8 macroblocks in JPEG Compression, this creates a noticeable fingerprint where lossy compression has been applied. By detecting anomalies between these macroblocks, we are able to potentially find where image tampering has occurred. It is also possible to detect where an image has been compressed twice, known as double compression. This does not always indicate tampering and could simply be a case of accidentally re-compressing the image, however it is a useful tool to compliment various other techniques.
  - Edge Detection using Standard Deviation
    Generally forged areas within an image will have harsher edges, being blended in less subtly with the surrounding area. Finding the Standard Deviation of the image in both the horizontal and vertical directions will find areas of high frequency change relative to their surrounding. Combining the two will allow us to detect harsh edge changes, and therefore some forms of forgery.
  - Clone Detection
    A popular way to remove an object from the scene is to clone the surrounding area in order to create a more realistic cover than would otherwise be possible via traditional means. This is often difficult to perceive to the naked eye, especially when the surrounding area is of a naturally repetitive nature, such as the sky, a beach or tree branches. Detecting areas that are identical to another is something that is unlikely to appear in genuine image, and is therefore very useful in detecting forgeries.
  - Resampling Detection
    Often forged images contain elements from other images that have been resized or rotated in order to achieve the correct scale. By scanning the image for areas that show signs of resampling, we are

able to potentially find objects that have been inserted after the original image was taken.

- Colour and Light Anomalies
  Whilst sometimes difficult to spot, altered sections of an image will never have identical lighting or colour conditions compared to the original source image. Detecting abnormalities in the lighting and chromatic variations in sections of the image allows us to detect some forms of forgery, especially those in which items have been added or removed from the image.

- Evaluating the runtime performance and detection effectiveness of these different algorithms on a variety of sample images. Using a large range of samples allows us to measure the performance of each implementation in a variety of situations, giving optimal and sub-optimal conditions for each algorithm, along with an average runtime and detection rate.

- From this data we are then able to distinguish which algorithms are best suited for certain image subsets, and how easy or difficult each type of forgery is able to be detected by each method.

3 Ethics

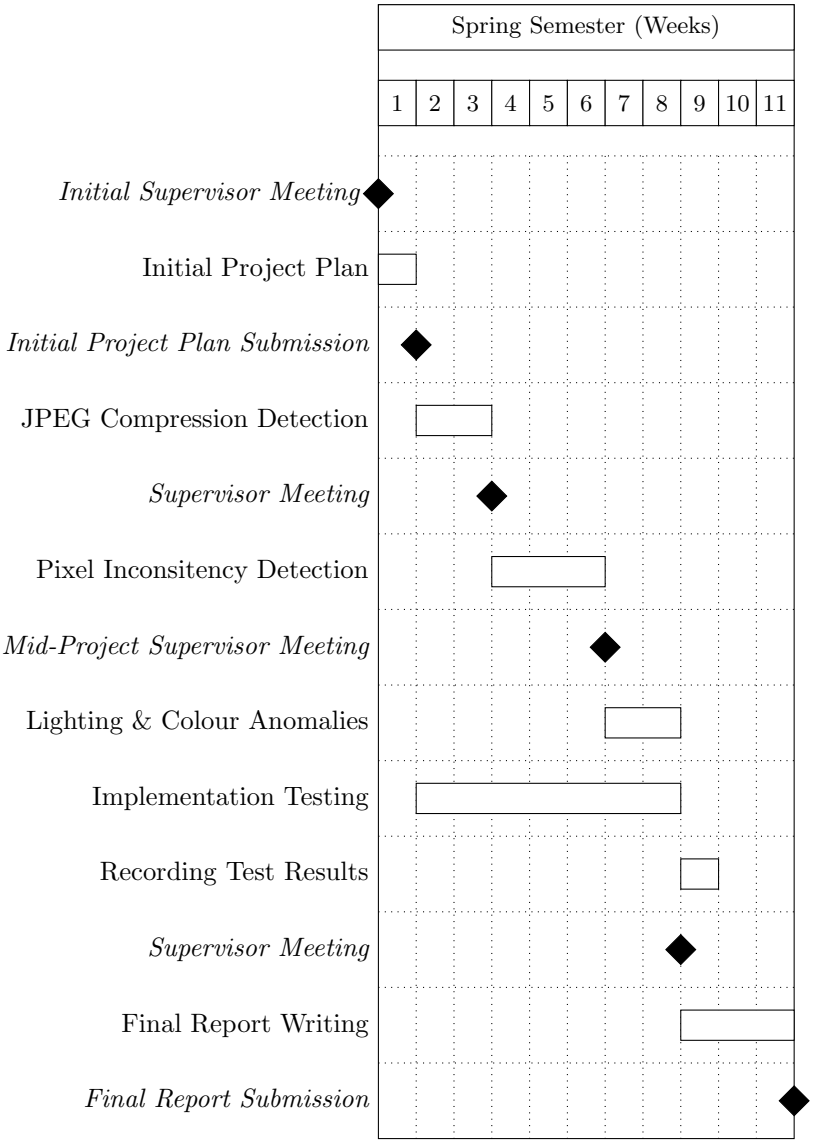
There are no ethical issues with this project, and therefore no ethical approval is required in regards to any aspect of research. Whilst image forgery detection has a wide range of applications, including its use regarding potentially sensitive images, we will not be discussing these nor using the project for anything but a wide range of standard library images.

4 Work Plan

The work plan is specified in weeks of the spring semester as opposed to dates in order to improve the readability of the plan.

Whilst major supervisor meetings have been highlighted as milestones, I am making regular contact with my supervisor in the form of weekly meetings in order to discuss any issues or questions in person.

4.1 Gantt Chart



Implementation testing and modification will occur continuously throughout weeks 2 - 8, as the different types of algorithms are developed and tested on sample data. In addition, the Project Viva is not shown above as the exact date of this is currently unknown, however it will be roughly scheduled in weeks 4-5 of the exam period. Whilst this will undeniably be a busy time, as I am only taking one exam it is more than reasonable to be able to prepare for the viva after the final report is completed and submitted.

4.2 Week 1

4.2.1 Initial Supervisor Meeting

This has consisted of meeting Dr Yukun Lai and discussing the basis of the project, future aims and how best to proceed with the initial report. Regular weekly meetings have been set up in order to ensure that any issues can be solved in person and that the project continues to be on track for completion.

4.2.2 Initial Project Plan

The initial plan has been the main focus of week 1, along with some background research on the methods involved as part of the project.

4.3 Weeks 2 - 4

Work will begin on in-depth research and implementation of various forgery detection methods. In order to better structure and plan the time required, I have split the algorithms into three categories based upon their operative nature, JPEG Compression Techniques, Pixel Inconsistency Methods and Lighting and Colour Anomalies. This is so that the work can be reduced into manageable sections and can be given more realistic timeframes.

4.3.1 JPEG Compression Detection

This implementation mainly involves exploring differences between the macroblocks of JPEG quantization, and using this information to detect anomalies that may have been caused by image forgery. I am also interested in looking at JPEG double-compression artefacts to further improve detection rates.

4.3.2 Quarterly Supervisor Meeting

This more extensive meeting provides an opportunity to discuss progress so far and also to outline the effectiveness of the implemented algorithms, including the results of the test data and a demonstration of the prototype in its current state.

4.4 Weeks 4 - 6

4.4.1 Pixel Inconsistency Detection

Once JPEG compression algorithms are run and tested, work will continue on using methods more focused on differences between adjacent pixels or clusters of pixels. I will be specialising in edge detection by calculating the standard deviation of images, detecting cloning by finding clusters of similar pixels within images and detecting resampling artefacts within an image.

Whilst there are a few different algorithms to be implemented in this stage, they all operate in a similar fashion (by comparing clusters of pixels), and therefore it is reasonable to allow two weeks to complete this task.

4.4.2 Mid-Project Supervisor Meeting

As a major milestone marking the half way point of the project, this will allow extensive discussion of the progress that has been made, in addition to discussing future plans and outlining, in detail, the final six weeks of the project.

4.5 Weeks 6 - 8

4.5.1 Lighting & Colour Anomalies

The third type of algorithms that will be researched and developed are ones based on colour and lighting differences within images. This includes techniques such as detecting anomalies using Chromatic Aberration, where we exploit the failure of an optical system to perfectly focus lights of varying wavelengths. This can be used to detect areas of colour that do not correctly match the rest of the image, despite looking perfectly reasonable to the naked eye.

4.6 Weeks 8 - 11

4.6.1 Recording Test Results

Once all implementations have been sufficiently trialled on test images, a set of 20-30 sample images will be used in order to fully test the efficiency of each algorithm. These sample images be grouped based upon specific types of forgeries, for example added objects, removed objects, altered features etc. The ability of each algorithm to detect the forgery, plus its running time will be recorded as test results. It is important to note that the sample images will remain consistent between experiments, in order to produce fair results. Run time will be calculated based on an average of five runs, using the inbuilt MATLAB timer function on the same machine.

4.6.2 Supervisor Meeting

This will be the final extensive supervisor meeting before the deadline, and will provide the chance to discuss any potential questions involved with writing the report and any queries based on the test data collected.

4.6.3 Final Report Writing

Whilst the report will be continuously worked upon and updated throughout the eleven weeks, the majority of the work will happen within the final three weeks, as this is when the results of the algorithms will become apparent. Work will continue on the report until the final deadline towards the end of week eleven.