CELL PHONE IMAGES IN SOCIAL MEDIA:
AN ANALYSIS OF CELLPHONE IMAGE STRUCTURE BEFORE AND AFTER SOCIAL MEDIA COMPRESSION

by

NICHOLAS KIAN-SENG NG

B.S., Middle Tennessee State University, 2011

A thesis submitted to the
University of Colorado Denver
in partial fulfillment
of the requirements for the degree of
Master of Science
Recording Arts
2013
This thesis for the Master of Science degree by

Nicholas Kian-Seng Ng

has been approved for the

Recording Arts Program

by

Catalin Grigoras, Chair

Jeff M. Smith

Sam Brothers

June 17, 2013
Ng, Nicholas Kian-Seng (M.S., Recording Arts)

Cell Phone Images in Social Media: An analysis of cellphone image structure before and after Social Media processing

Thesis directed by Associate Professor Catalin Grigoras

ABSTRACT

Social media has made it easier than ever to publish pictures on the internet. Investigators now have to determine how to deal with pictures, posted to these social media websites, which become evidence in crimes. Traditional methods of authenticating the origin of pictures are not as effective because the processing applied by the social media websites eliminates or obscures much of this information. This thesis proposes several methods of image attribution that utilize popular image authentication techniques and presents test data to illustrate their use. Chapter 1 introduces to the topic of Image Attribution. Chapter 2 provides brief overview of the background analysis techniques that will be used. Chapter 3 describes how the test data that was taken and why. Chapter 4 proposes three different methods for answering three different forms of image attribution. Chapter 5 tests the reliability of the proposed methods with the cellphone image database uploaded to three separate social media websites. Chapter 6 provides a conclusion based on the results test results. Chapter 7 proposed future research in the field.

The form and content of this abstract are approved. I recommend its publication.

Approved: Catalin Grigoras
ACKNOWLEDGEMENT

I would like to give special thanks to Catalin Grigoras and Jeff Smith for guiding me in both my graduate education and throughout the creation of this thesis.

I would also like to thank my lovely girlfriend who has supported me through six years of out-of-state education providing me with encouragement and guidance throughout.

Last, but not least, I would like to thank my parents, who have been a constant source support for my academic goals. Without their support, my academic career and this thesis would not have been possible.
# TABLE OF CONTENTS

**CHAPTER**

I. INTRODUCTION ........................................................................................................... 1

II. IMAGE ATTRIBUTION TOOLS ..................................................................................... 3

   Dimensional Analysis .................................................................................................. 3

   Structural Analysis ....................................................................................................... 3

   Metadata Analysis ........................................................................................................ 4

   Quantization Table Analysis ......................................................................................... 4

   Compression Level Analysis ......................................................................................... 5

   Color Filter Array Analysis .......................................................................................... 6

   Photo Response Non-Uniformity Analysis ..................................................................... 7

III. TEST DATA ................................................................................................................... 8

   Defining the Size of the Image Database ...................................................................... 8

   Taking the Original Pictures ......................................................................................... 9

   Choosing Social Network Websites .............................................................................. 10

   Uploading to Social Media Websites ............................................................................. 11

   Downloading from Social Media Websites .................................................................... 12

IV. METHOD ...................................................................................................................... 14

   Social Media Website Image Attribution ...................................................................... 14

   Camera Model Image Attribution ............................................................................... 17

   Specific Camera Image Attribution .............................................................................. 19
CHAPTER I

INTRODUCTION

Social networking websites have become a group depository of personal photos for all to see. The public availability of these images provides evidence otherwise unavailable to law enforcement. However, this valuable investigative tool comes with its drawbacks.

There has not been much research done in images uploaded to social networking websites. Previous research has been done into the usefulness of artifacts left by social networking websites (Helenek) and a general survey of images that exist on Social Media websites. (Castiglione, Cattaneo and De Santis) No formal methods have been proposed to track images uploaded to social networking websites back to their origin.

This presentation will provide several methods to track the origin of images posted to social media websites. While Image Authentication techniques are traditionally used to determine the validity of images, some can be useful for determine the source of an image. According to the SWGIT Best Practices for Image Authentication, “Forensic Image Authentication is the application of image science and domain expertise to discern if a questioned image or video is an accurate representation of the original data by some defined criteria.” This paper will differentiate methods of source identification as Image Attribution.

Traditional eye witnesses of crimes are being replaced by video and image recordings. Today, almost everyone owns a cellphone and crimes are being increasingly captured on cellphone cameras. Unlike the traditional camera, people always carry their cellphones and they are becoming increasingly user friendly. Due to proliferation and ease of use, cellphones are becoming increasingly important to law enforcement investigations.
As society is technologically evolving, evidence of crimes is increasingly posted to social networking websites. Investigators must determine where photos that find their way onto social networking websites originated. Investigators need to be able to trace exactly what social networking website an image came from and the camera that captured the image, however, the upload process makes it difficult for investigators to determine where uploaded images originate.

Cellphone cameras were chosen for this paper because each cell phone creates a photo in a unique way which is useful for image attribution. Several social networking websites were chosen because each alters uploaded images to normalize and reduce file size differently, providing a distinctive signature. Identifiers from the unique method that cellphone cameras use to capture photos and the distinctive signature produced by social networking websites are the basis of this study.

Three ways of attributing images to sources were proposed based on image authentication analysis techniques and distinctive social media processing features.

A database of cellphone camera images was prepared to provide a baseline for comparison and test images to determine the successfulness of the proposed Image Attribution methods. The database of images uploaded to social networking websites will be compared to determine distinctive signs of each one of the social networking websites uploading process.

The final results of this research will show that images uploaded to the internet can be identified given a sufficient database of images to compare and known identifiers of cellphone cameras and social networking website upload systems.
CHAPTER II
IMAGE ATTRIBUTION TOOLS

Image Manipulation has become easier since the advent of the computer and digital image editing software. Image Authentication provides a method of detecting image manipulation. Many of these techniques that are used to authenticate an image to an alleged source can also be used for Image Attribution. These techniques will be briefly covered in this chapter.

**Dimensional Analysis**

Dimensional analysis is the simplest of the Image Attribution methods. An image source creates a digital image with a fixed height and width. Some sources may allow for multiple dimensional settings. Comparing an image database, with all possible dimension settings for the image source established, against an unknown image’s dimensions can establish a set of possible sources.

**Structure Analysis**

Structure Analysis relies on the way a source creates an image. Images analyzed in this project, JPEGs, adhere to a general structure making them universally readable. However, many aspects of the structure that can vary between sources or even be excluded all together. These structural variations can be combined to form signatures, which can be attributed to a capture device or image processing. However, capture device information is changed dramatically when sent though social media processing. This method is usually successful in attributing an image to the social media that it was processed by but not always the original source.
**Metadata Analysis**

Metadata analysis explores information imbedded in the image file data structure. Metadata can be read by simply looking for ASCII information in the hexadecimal code (Figure 9) or through an EXIF reader such as the one built into Windows (Figure 1). A majority of the information that is embedding in JPEG images is indiscernible in ASCII and must be read with an EXIF reader. A typical capture device will provide a series of information that contains standard information such as the Camera Make and Model, as well as manufacturer specific information. A social media website may remove or alter this information to decrease file size and maintain user privacy. However, this processing can add its own metadata information, making it possible to attribute the image to the social media website.

![Figure 1 Iphone 4 Metadata](image)

**Quantization Table Analysis**

Quantization Table Analysis relies on the way JPEG files determine compression level. Quantization Tables values control the amount of compression that is applied to an image. Two 8x8 quantization tables exist for each JPEG separately compressing the luminance and chrominance channels. The values of the 8x8 matrices can vary from 1 to
255 with higher values producing more compression. Figure 2 shows the two quantization tables (chrominance and luminance) contained in an original iPhone 4 image. Because the Quantization Tables have such a strong effect on subsequent image quality and file size, they vary considerably between sources. Since social media websites rely on small file sizes to decrease internet loading time, images from these sources are characterized by high quantization table values.

<table>
<thead>
<tr>
<th>Luminance</th>
<th>Chrominance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1 1 1 2 4 5 6</td>
<td>1 1 2 4 9 9 9 9</td>
</tr>
<tr>
<td>1 1 1 2 5 6 5</td>
<td>1 2 2 6 9 9 9 9</td>
</tr>
<tr>
<td>1 1 2 4 5 6 5</td>
<td>2 2 5 9 9 9 9 9</td>
</tr>
<tr>
<td>1 1 2 5 8 8 6</td>
<td>4 6 9 9 9 9 9 9</td>
</tr>
<tr>
<td>1 2 3 5 6 10 10 7</td>
<td>9 9 9 9 9 9 9 9</td>
</tr>
<tr>
<td>2 3 5 6 8 10 11 9</td>
<td>9 9 9 9 9 9 9 9</td>
</tr>
<tr>
<td>4 6 7 8 10 12 12 10</td>
<td>9 9 9 9 9 9 9 9</td>
</tr>
<tr>
<td>7 9 9 9 11 10 10 9</td>
<td>9 9 9 9 9 9 9 9</td>
</tr>
</tbody>
</table>

**Figure 2 iPhone 4 Quantization Tables**

**Compression Level Analysis**

Quantization values determined by the Quantization Tables, is applied to the DCT coefficients of the image. This lossy process generates periodicity in the DCT coefficients. Compression Level Analysis focuses on this periodicity by analyzing the second derivative of the DCT coefficients. (Popescu and Farid, Statistical Tools for Digital Forensics) The traditional method of analysis is to visually analyze second derivative graphs as shown in Figure 3. A first generation JPEG image will produce a distinctive center spike with two minor spikes to the left and right. Subsequent compression will lead to many small spikes indicating increased periodicity due to the recompression. Each camera source will produce a distinctive frequency histogram and the location and shape of the spikes can be used to match a camera source image with a processed image. A database of known images from the camera is valuable as a point of comparison.
Color Filter Array Analysis

Since traditional camera pixels only capture one of the three primary colors (red, green or blue), a Color Filter Array is used to capture an image with a combination of the three colors. The camera using a Color Filter Array, a mosaic of red, green and blue filters, captures a single color per pixel and then interpolates all red, green, and blue values based on neighboring pixels. (Popescu and Farid, Exposing digital forgeries in color filter array interpolated images) This interpolation produces periodicity unique to the color filter array used in the camera. Different color filter arrays use different combinations of red, green, and blue filter producing different frequencies in the histogram. The red, green, and blue histograms of the an uncompressed image are different from one another due to the placement and relative numbers of red, green and blue filters contained in the color filter array. Subsequent compression, which is done with the signal broken up into chrominance and luminance, will produce similarity between the red, green and blue histograms. Subsequent use of Color Filter Array did not provide any usable results for Image Attribution and its usage was scrapped from this project.
Photo Response Non-Uniformity Analysis

Photo Response Non-Uniformity Analysis relies on the noise generated by the camera optics. Each camera model will contain different components leading to noise in the image captured. Even cameras within the same make and model will have varying noise signatures because of minor variation that occurs when producing camera imaging sensors. The Photo Response Non-Uniformity is best isolated by capturing several images from the same source and averaging them together. A proper set of images for PRNU requires good lighting and random scene content. PRNU is specific to the actual source camera used to take the picture due to the unique manufacturing error produced in every camera. (Lukas, Fridrich and Goljan) Figure 5 is an example of a PRNU sample generated from 1000 images that will be compared against test images.

Figure 5 Iphone 4 PRNU sample generated from 1000 images
CHAPTER III
TEST DATA

A database of cellphone camera images was compiled to test the proposed Image Attribution methods. These images were then uploaded to three social media websites, Facebook, Myspace, and Google Plus, to determine which unique signatures remained and what new signatures were formed for identification. The alterations done by the social media website image upload process were then recorded and compared against other social networking websites.

Defining the Size of the Image Database

The large corpus was captured for two reasons: to look for possible sources of error in the workflow and to provide enough images for high quality PRNU samples.

Using a large test corpus allowed me to find sources of error that may not have been present in a smaller dataset. Most of the processing chain in this project contains unknown ‘black box’ processes and only a large amount of images could say whether they worked as expected or not. The large test corpus showed many possible pitfalls a user might fall into when trying to perform image attribution in real casework. These errors are covered in more detail later on.

Working with such a large test corpus for this experiment was done also to determine how high of a PRNU match could be obtained with social media compressed images. Photo Response Non-Uniformity (PRNU) relies on the proper extraction of camera noise. Several images from the same camera are averaged together and filtered so that only the camera noise remains. Too small of a data set will result in a PRNU sample with scene details as well as the noise signature. This becomes less of an issue with larger data sets.
**Taking the Original Pictures**

Creating the test corpus for this project can be broken down into several steps: capturing the images, choosing the social networking websites, uploading the image, and downloading the images.

Capturing images was done with strict criteria to ensure similar conditions for comparison. These criteria included GPS settings, orientation, and image content.

The camera settings were chosen based primarily on the default settings. In the cameras tested, this meant the highest quality and resolution settings. The only setting that was changed from default was the GPS location services. All phones with GPS location services had the feature turned on.

The orientation of the camera was controlled as well. All pictures were taken in landscape mode. The actual position of the camera varied between the side addressed and front addressed cameras. In front addressed cameras, such as the Blackberry 8650, pictures appear in landscape mode when the camera is positioned with the camera lens positioned at the top. In side addressed cameras, such as the iPhone 4, pictures appear in landscape mode when the camera lens is positioned to the side. This provides one realistic orientation for the front addressed cameras but two realistic orientations for the side addressed cameras. Typically a user will not take a picture with a front addressed camera with the camera positioned at the bottom. However, side addressed camera users may choose to take pictures with the camera lens to the left or the right. Most modern phones actually rotate the display based on the user’s chosen orientation. This variation can cause issues with noise profile comparisons with evidence images. All pictures taken with side addressed cameras in the test corpus were taken with the camera lens on the left. If a picture was taken from the right, a 180° rotation can be performed for a proper noise profile comparison. Further orientation issues are discussed when social media processing is applied.
The image content of the pictures is important for PRNU analysis. All images were taken during the day, outdoors, with sufficient sunlight. This was done to limit the amount of spurious noise generated and provide a more varied intensity and color information. Because PRNU comparison is light dependent, similar lighting was used for database and mock evidence.

Choosing Social Networking Websites

Three social networking websites were picked after early testing according to several criteria: a substantial user base, bulk upload capabilities, and bulk download capabilities.

The first criterion, a substantial user base, was established so that findings of this study would be useful for real life investigations. There are an endless number of social media websites available and it was not practical to include small user base websites in this study. Should data be required from other Social Networking Websites in the future, the principles and methodology of this experiment can be easily applied. The three social media websites chosen garner a large percentage of social media usage.

The second criterion, bulk upload capabilities, was established due to the size of the test corpus. With an input test corpus size of 9,900 pictures, uploading each image independently would be inefficient. The three social media websites chosen allow for bulk image upload. Tumblr was originally considered but was eliminated due to a lack of bulk uploading capabilities and a daily upload limit.

The third criterion, bulk download capabilities, was also established due to the size of the test corpus but was later relaxed. Two of the social networking websites, Google Plus and Facebook, allow for bulk downloads. The Google Plus’ bulk download feature, Download album, is shown in Figure 6. The third website, Myspace, restricted the bulk download feature to law enforcement. A third-party bulk download software was used to
acquire images from Myspace. Further explanation of this process will be provided later on.

![Google Plus Download Album Feature](image)

**Figure 6 Google Plus Download Album Feature**

**Uploading to Social Media Websites**

Uploading the images was done with each social media website’s standard PC upload interface. This was done to mimic the traditional method that a user would use to upload images and to ensure no further processing was done by 3rd party software. 1,100 images were uploaded to Facebook, Myspace and Google Plus from each of the 9 cameras. Images were uploaded in 100 image sets to avoid size restrictions and upload errors. All three social networking websites produced upload errors in which an image was uploaded twice or an image was not uploaded at all. These errors were rare and were fixed manually by deleting duplicates and uploading missing images. Google Plus and Myspace provided no options for customizing the upload quality but Facebook allowed for ‘High Quality' as seen in Figure 7. ‘High Quality' was chosen for all images uploaded to Facebook. Disabling ‘High Quality’ was not pursued in this project.
Downloading from Social Media Websites

Downloading the images was done differently for each social media website because each website had different download options. Google Plus allows for album and archive download, Facebook allows for archive download, and Myspace provided no bulk image download options available to the general public.

Google Plus provided the easiest bulk image download with both a single album and archive download. A photo archive was available through Google’s Takeout service. However, when downloaded, the photo archive contained an incomplete collection of the images uploaded to Google Plus. Subsequently, the single album download feature in Google Plus was used and all albums were downloaded without incident.

Facebook provided an archive download which among other data, was supposed to include all photos taken on the account. This archive included a multitude of account information but similar to the Google Takeout had an incomplete archive of photos. Without a single album download function, images were downloaded one at a time. This however was found not to be optimal because some images are displayed in a lower
resolution than what was is available through the archive download. Once this was determined, photos that were included in the archive were deleted from the Facebook account so that the missing images would be provided in a subsequent archive download. This was repeated several times, and several albums were downloaded. After several repeated archive download, the archive download stopped downloading correctly. Another Facebook account was created, and the rest of the albums were downloaded from this secondary account. Technical support inquiries provided no useful explanation for the archive download issues.

Myspace provided no archive download of any kind to the general public. Myspace only provides archive downloads with a legal request. Myspace required that images be downloaded independently. A third party bulk image downloader was used to perform this function shown in Figure 8. The program's output was verified to be identical to images manually downloaded.

![Bulk Image Downloader interface](image-url)
CHAPTER IV

METHOD

In an image attribution investigation, there are three possible questions asked: What social media was an image processed with, what camera model was the image captured with, and can the specific camera be matched? This chapter will propose methods for answering each of these questions.

Social Media Website Image Attribution

The first question, “what social media was an image processed with?”, involves looking for similarities between all images uploaded to a single social network website. This type of analysis is most important for investigative use. Determining the source of an unknown image can lead to further evidence, a higher resolution version of the image, and more information about the image itself. The information used for Social Media Website Image Attribution are metadata information, image dimensions, file structure and quantization tables. Overall this analysis is the least complicated and time-consuming especially if you already have a database of known identifiers.

Image metadata is information written into images that provides details about an image, such as the Make and Model of the capture device or the GPS location where the image was taken. Cellphone images were chosen as the test corpus partially because they contain a large amount of metadata information. A sample of the information contained in a cellphone image is provided in Figure 9 from a hex viewer. Each row shows information from the byte offset designated on the far left. The first column next to the byte offset show the hexadecimal representation at the particular byte offset, the second column shows the ASCII representation and the third column provides a
description of the contents of each offset.

Figure 9 Metadata displayed by ExifToolGUI

Determining the source of an image with original metadata information intact is relatively easy with the Make and Model included. Information can be interpreted in JPEG images with an Exif reader as shown in Figure 10. This is especially useful for metadata that is not in written in ASCII. However, social media websites modify image metadata, eliminating most of the information. This makes determining the camera source difficult but provides a tool for determining the social media website that the image was processed with. Each social media website studied so far removes and adds different information allowing for easy identification.
Modern cameras have evolved over time to produce larger resolution images in order to increase the possible overall quality of the image. Large resolution images become important if you want to crop or print an image. However, internet images are produced as small as possible to limit the file size and because most people only want to view images on the internet in a fraction of their original resolution. Social Media Websites compensate for large resolution uploaded images by resizing them. Some social media websites even adjust the relative width and height of images called the

![Figure 10](image.png)

Further information can be decoded from the imbedded Exif information.
aspect ratio. Since each social media website resizes images differently, knowledge of each websites image dimensions can be used to determine which social media an image was processed with.

Most cellphone images are JPEG compressed files. JPEG files use quantization tables to determine how much compression is to be applied to an image. Two quantization tables exist, chrominance and luminance, and every time a JPEG file is saved, quantization tables are applied. Each quantization table contains an 8x8 matrix of values that can be assigned a value from 1 to 255. Camera manufacturers define what quantization tables are used in a camera and vary between Make and Model. Social media websites do not retain these values when the images are processed. Instead images are given new quantization tables defined by the social media website when processed. Knowing the quantization tables used by a social networking website can be used to determine which one was used to process and image.

**Camera Model Image Attribution**

The second question, “what camera model was the image captured with?”, involves looking for image identifiers that haven’t been eliminated by Social Media processing. This method is useful for focusing investigations, eliminating potential suspects, producing possible timelines, and even weak multiple image association. The information used for Camera Model Image Attribution is left over metadata, supporting social media documents and Compression Level Analysis (CLA). A basic analysis of metadata and social media documents is not complicated or time-consuming while CLA analysis is more difficult.

Each social media website will remove and add its own metadata. By understanding what information is generated by the social media website, the user can look for any
metadata that has been left over from the original camera. Some social media websites will leave no original metadata while others will leave the original metadata mostly intact.

Social media websites record user usage and store it in an archive. This information is typically only available to law enforcement with a court order. This information will usually contain information such as when the account was logged into and what IP addresses were used. However, some social media websites capture the metadata information of the camera prior to modifying it.

Of the three social media website I looked at, Facebook was the only one that provided image metadata information to the account owner without a court order. This information included the date/time the image was taken, location image was taken, and image capture parameters if the original camera image contain the information. An example of this information is provided in Figure 11.

![Figure 11 Camera Metadata information from a Facebook archive download](image-url)
Specific Camera Image Attribution

The third question, “can the specific camera be matched?”, involves looking primarily at Photo Response Non-uniformity (PRNU). This question is the most difficult because test images must be taken with the specific camera in question. This is often problematic when there are multiple suspected cameras or the evidence camera is inaccessible.

When a large number of suspect phones exist, it is advised that the investigator first perform Camera Model Image Attribution analysis to determine the model of specific camera.

PRNU analysis works best for Specific Camera Image Attribution because it focuses on optic noise which typically varies between cameras of the same model.
CHAPTER V
RESULTS

Overview

A test corpus of 1000 images per camera was used as the database and a smaller independent set of 100 images was used as the mock evidence. In this test, multiple signatures were used to increase the robustness of the methods but can be excessive especially in many investigative applications.

Note: Although some of the methods included are complicated and may be beyond the technical skills of the base investigator, many of the simple tools provided can often be all that is needed to attribute an image. It is also important to note that Social Media processing is a ‘black box’ processing, meaning that there is no published explanation of the process that input images go through. This means that when performing an authentication/attribution analysis, there is a possibility that all possible outputs are not accounted for in testing. Furthermore, it is important to note that the processing may change depending on modifications made by the Social Media Website to the processing algorithm over time. Old assumptions should be verified to ensure they still hold true.

Below, the test pictures were compared, helping determine the answer to the three image attribution questions:

Social Media Website Image Attribution

Metadata from known Social Media Website source images was compared to their known originals and several determinations were made. When an image is uploaded to social media, a majority of its metadata is replaced. Figure 12 illustrates how a section of metadata can change dramatically between the Original and Social Media processed.
The important thing to notice is that each website produces a slightly different signature, providing a simple low-level analysis of Social Media Website Image Attribution. Signatures can be found by uploading exemplar images to each social media website and comparing image before and after.

Images uploaded to Myspace have their original metadata completely replaced. First, you will notice that the APP0 JFIF segment denoted with an ASCII phrase ‘JFIF’. This metadata was added by all social network websites used in this project but was not included with any original images from cellphones. Additionally, in the Myspace uploaded images, several other segments are understandable in ASCII. ‘Copyright International Color Consortium, 2009’ is one of the ASCII segments and was found in all of images uploaded to Myspace regardless of source. Facebook also replaces metadata with the same JFIF and ASCII segments. However, Facebook images can be differentiated from Myspace images by the COM Comment segment found directly proceeding the APP0 JFIF segment with a recognizable ‘*’ ASCII symbol. Myspace images do not contain the COM Comment segment.
Google Plus varies completely in both design and appearance from Myspace and Facebook. Google Plus retains much of the original metadata with distinctive additions, deletion and alterations. Google Plus adds two unusual fields not found in any of the original images. The first unusual field is marked with a 0x9009 tag. This entry appears to be a form of MakerNote information. The meaning of the content is unknown but all images uploaded with Google Plus in this project shared the same 0x9009 entry value.
The second unusual field is an APP1 XMP segment that contains an adobe and w3 URL. This entry was also only found in images uploaded with Google Plus. Several other entries can be used for Social Media Website Image Attribution but they vary depending on the contents of the original metadata. The ExifIFD ImageUniqueID entry appears in all Google Plus uploaded images. The ImageUniqueID values are different for each image and no correlation has been found between the values. The ImageUniqueID entry was also found in original HTC Trophy images. The HTC Trophy ImageUniqueID value entry was changed when uploaded to Google Plus. It is important to understand the original metadata of an image before using the ImageUniqueID to attribute an image to Google Plus.

The another method of attributing images to a Social Media only works for original images that are missing standard metadata. Some images lack standard metadata tags and Google Plus will add the tags during processing if they are not found in the original image. The most important of these tags is the Software tag because an added tag actually contains a ‘Picasa’ value that can also identify Google Plus. However, if an image already contains one of these standard metadata tags, the information will not be replaced.

The last way to attribute an image to Social Media is through the thumbnail. If an original image does not contain a thumbnail, Google Plus will add one. Also, Google Plus restricts the thumbnail size of an image to width of 160. The height will depend on the aspect ratio of the original thumbnail or full image. Metadata information can be very useful but care should be taken when considering analyzing metadata because metadata can be easily tampered, so other attribution methods should be used to verify.

Social Media websites process images for two reasons: reducing file storage requirements and providing the user with a faster load time. Every new generation of cameras is providing a higher and higher number of pixels making file size and load time
bigger and longer. The simplest way Social Media websites eliminate this issue is to resize the image. Multiple versions of the same image are created to accommodate desktop, cell phone and tablet access. Even more versions are created for album thumbnails, timeline backgrounds, profile pictures, etc. In this test, the highest resolution images were used to facilitate the statistical data comparisons. However, the full resolution images are not always resized. Social Media Websites only resize images when the pass a designated pixel threshold. Castiglione et Al. performed an analysis on the pixel thresholds for Facebook and Google Plus and determined that the threshold was at 2048 x 2048 for Facebook with the High Quality setting and 2048 x 2048 for Google Plus as well. For example, the iPhone 4 original has a resolution of 2592 x 1936. Facebook resized the image to 2048 x 1530 which can identify the Social Media source from the original but not from all other Social Media because Myspace and Google Plus also provided the same dimensions. However, preliminary research done with Tumblr showed iPhone images being resized to 1280 x 956. Dimensional analysis can give clues to the Social Media Image Attribution Social Media website depending on the upload website and original image dimensions. A full chart of dimensional resizing of the test data is provided in Figure 13. Also, in real world applications, non-full resolution images can more easily be traced back to their Social Media source based on their more irregular sizes.
Another way in which many social media websites deal with long image loading times is to change the way the image is loaded all together. A traditional JPEG image is formatted as Baseline and the entire image is encoded in one scan and the image is loaded by the end-user all at once. However, a more efficient method, Progressive, encodes the image in multiple passes allowing users on slower connections to see the image earlier than a Baseline image. The eventual image is the same and the conversion can be done with no loss of quality. Traditionally, cameras take pictures with Baseline encoding and only some Social Media Websites use Progressive. Facebook and Myspace use Progressive encoding while Google Plus and Tumblr use Baseline encoding. It can be difficult to determine whether an image is Baseline or Progressive encoded with traditional tools but there are many freeware tools available that provide this information, such as JPEGSnoop and Exiftool.

Resizing only limits the number of pixels in an image, the true ability of a Social Media Website’s to optimize storage and load time is with lossy JPEG compression. The central tool for this lossy compression is the JPEG Quantization Tables that determine how much compression is applied to the image. Because this step has a large impact on the eventual quality of the image, camera manufacturers and Social Media websites

<table>
<thead>
<tr>
<th>Camera</th>
<th>Original</th>
<th>Google+</th>
<th>Facebook</th>
<th>MySpace</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPhone 3gs</td>
<td>2048 x 1538</td>
<td>2048 x 1536</td>
<td>2048 x 1536</td>
<td>2048 x 1536</td>
</tr>
<tr>
<td>iPhone 4</td>
<td>2592 x 1936</td>
<td>2048 x 1530</td>
<td>2048 x 1530</td>
<td>2048 x 1530</td>
</tr>
<tr>
<td>HTC myTouch 3G</td>
<td>2048 x 1536</td>
<td>2048 x 1536</td>
<td>2048 x 1536</td>
<td>2048 x 1536</td>
</tr>
<tr>
<td>HTC Trophy</td>
<td>2592 x 1944</td>
<td>2048 x 1536</td>
<td>2048 x 1536</td>
<td>2048 x 1536</td>
</tr>
<tr>
<td>HTC EVO 4G</td>
<td>3264 x 1952</td>
<td>2048 x 1225</td>
<td>2048 x 1225</td>
<td>2048 x 1225</td>
</tr>
<tr>
<td>BB Pearl Flip</td>
<td>1600 x 1200</td>
<td>1600 x 1200</td>
<td>1600 x 1200</td>
<td>1600 x 1200</td>
</tr>
<tr>
<td>BB Curve</td>
<td>1600 x 1200</td>
<td>1600 x 1200</td>
<td>1600 x 1200</td>
<td>1600 x 1200</td>
</tr>
<tr>
<td>BB Tour</td>
<td>2048 x 1536</td>
<td>2048 x 1536</td>
<td>2048 x 1536</td>
<td>2048 x 1536</td>
</tr>
<tr>
<td>Moto Cliq</td>
<td>2560 x 1920</td>
<td>2048 x 1536</td>
<td>2048 x 1536</td>
<td>2048 x 1536</td>
</tr>
</tbody>
</table>

Figure 13 Image Dimensions after Social Media resizing
choose different values for the 2 8x8 value tables to best optimize quality vs. file size.

Though these values are not always unique from camera to camera, in traditional camera authentication, Quantization Tables can be used to eliminate a camera as a ‘suspect’ if it is unable to produce the Quantization Tables in question. Cameras have traditionally provided static Quantization Tables, meaning that images taken with the same camera with the same quality settings will always have the same Quantization Tables. However, Social Media websites do not follow the same model. Google Plus provides several Quantization Tables values but only one set of Thumbnail Quantization Tables regardless of the capture source. In Figure 14, the Google Plus Blackberry 8220 image and the Iphone 4 image have different Quantization Tables but identical Thumbnail Quantization Tables. Facebook provides several possible tables which presumably are applied according to image content. Figure 15 shows several possible Facebook Blackberry 8220 Quantization Tables with Quantization Table 2 matching the Blackberry 8220 test image. Though Quantization Tables may vary from image to image, the limited number of tables used by Facebook allows for Social Media Website Image Attribution. Myspace provides a single Quantization Table for all images. Figure 16 shows several Myspace images sharing the same Quantization Tables. Each of the image sources have different features but they can each be used for Social Media Website Image Attribution as long as a trend can be determined.
<table>
<thead>
<tr>
<th>Luminance</th>
<th>Chrominance</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 3 3 5 7 12 15 18</td>
<td>5 5 7 14 30 30 30 30</td>
</tr>
<tr>
<td>4 4 4 6 8 17 18 17</td>
<td>5 6 8 20 30 30 30 30</td>
</tr>
<tr>
<td>4 4 5 7 12 17 21 17</td>
<td>5 6 8 20 30 30 30 30</td>
</tr>
<tr>
<td>4 5 7 9 15 12 26 19</td>
<td>14 20 30 30 30 30 30 30</td>
</tr>
<tr>
<td>5 7 11 17 20 33 31 23</td>
<td>30 30 30 30 30 30 30 30</td>
</tr>
<tr>
<td>7 11 17 19 24 31 34 28</td>
<td>30 30 30 30 30 30 30 30</td>
</tr>
<tr>
<td>15 19 23 26 31 36 36 30</td>
<td>30 30 30 30 30 30 30 30</td>
</tr>
<tr>
<td>22 28 29 29 34 30 31 30</td>
<td>30 30 30 30 30 30 30 30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Luminance</th>
<th>Chrominance</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0</td>
</tr>
</tbody>
</table>

Google Plus Blackberry 8220 Thumbnail Quantization Table

Google Plus Iphone 4 Quantization Table 3

Google Plus Iphone 4 Thumbnail Quantization Table

Figure 14 Images uploaded to Google Plus has several similar Quantization tables but only one thumbnail Quantization table

Facebook Blackberry 8220 Quantization Table 1
Figure 15 Images uploaded to Facebook have several Quantization Tables choices
Some modifications by Social Media websites actually change the overall file structure of the image. In Figure 17, Facebook and Myspace processing dramatically changes the structure of the image. The major reason for the huge variation in file structure is due to changing the image from Baseline to Progressive. A Baseline image requires less Huffman Tables and only one start of scan. A Progressive image breaks up the image into multiple scans which is more conducive for internet/mobile viewing. We noted earlier that the conversion from Baseline to Progressive is lossless and this is because unlike Quantization Tables that are inherently lossy, the Huffman Tables that are added are a lossless transform. Facebook and Myspace processed images are both progressive but Figure 17 shows Facebook processed images have one more Huffman Table than Myspace processed images. Figure 17 also shows that Google Plus remains Baseline and the structure changes primarily due to the addition of a thumbnail image. The file structure analysis of the eight other camera sources is provided in Appendix A.
The structure of each of the images after social media processing maintains the social media specific structure.
Figure 17 File Structure after social media processing

(a) Blackberry 8220 Original  
(b) Facebook Processed  
(c) Myspace Processed  
(d) Google Plus Processed
Camera Model Image Attribution

Camera Model Image Attribution follows a similar model to Social Media Website Image Attribution. Analysis always starts with the simplest techniques and gradually gets more technically and computationally complex. Camera Model Image Attribution will involve metadata, supporting social media documents and Compression Level Analysis (CLA).

The first information to look at is the metadata. In Social Media Website Image Attribution, the focus of the analysis was on the information added by the Social Media processing. In Camera Model Image Attribution, the focus of the analysis is on the information that was left behind from before the Social Media processing. In the previous section it was noted that most of the original information is removed from the image. This is typically the case with most sources. Facebook and Myspace are among the sources that eliminate camera information in the metadata. Google Plus and Tumblr retain some camera information allowing for attribution. Images from Facebook and Myspace will provide no metadata information useful for Camera Model Image Attribution. However, it is fairly easy to determine the camera source of Google Plus and Tumblr images because part of the information retained is the Camera Make and Model. Original images that contain Software entries can also linked to the Google Plus or Tumblr with the particular operating system that is used by the phone. Original images that contain any type of time entries such as ModifyDate, DateTimeOriginal or CreateDate can be used to attribute if any original time information is known. However, Google Plus and Tumblr were the only websites tested that retained time information. It again should be noted that metadata can be easily tampered with but other attribution methods can be used to verify.

Many Social Media websites store account usage statistics. Facebook, Google Plus, and Myspace all provide options for law enforcement to obtain information with the
proper court order. (Facebook) (Google) (Myspace) However, none of these resources divulge what sort of information is provided to law enforcement. The only glimpse into the content contained in these reports comes from Facebook’s user account archive download feature that provides a plethora of information including message history, login locations, IP Addresses, Searches, and more. What is interesting for Image Attribution, is that they provide photo metadata. When Facebook eliminates the original metadata information, it actually keeps a portion of the information that can be downloaded in an archive by the owner of the account or by law enforcement. An example of this information is provided in Figure 11. Depending on the image source, different information is included but in all of the tests performed, the camera model was always included. The complete collection of Facebook extracted social media metadata is included in Appendix B.

Compression Level Analysis (CLA) utilizes the distinctive shape of the second derivative in JPEG compressed images to match an unknown image to a known source. The traditional method involves comparing the shape of CLA graphs of two signals. This comparison can be seen in Figure 18. The first graph, (a), is an unknown image from Facebook. Two comparison images from Facebook are shown: (b) is a Iphone 4 image and (c) is a Blackberry 9630 image after Facebook processing. Visual comparison shows that (a) and (c) are the same image. However, sometimes the Social Media images are more difficult to match visually. Some visual comparisons look like Figure 19 instead. The first graph, (a), is an unknown image from Myspace. Two comparison images from Myspace are shown: (b) is a Iphone 3gs image and (c) is a Iphone 4 image after Myspace processing. Visual comparison is inconclusive. (a) and (b) happen to both be Iphone 3gs images from Myspace. Comparing several database images and unknown images (from the same source) provides a higher chance of getting a good comparison.
Another way of comparing Compression Level Analysis values is by computing correlation coefficients. Figure 20 shows the correlation coefficients of comparisons between a set of 1000 database images and an evidence file for each camera. Results were improved slightly in Figure 21 by comparing the entire set of 100 images per evidence camera to database bounds files that were created with 1000s images per camera.

*Figure 18 Visual Analysis of the CLA graphs: (a) unknown image (b) Iphone 4 (c) Blackberry 9630*
Figure 19 Visual Analysis of the CLA graphs:
(a) unknown image (b) Iphone 3gs (c) Iphone 4

Figure 20 Correlation Coefficient values of the evidence images compared directly to evidence images
### Evidence

<table>
<thead>
<tr>
<th>MS CLA</th>
<th>BlackBerry 8220</th>
<th>BlackBerry 8330</th>
<th>BlackBerry 9630</th>
<th>HTC EVO 4G</th>
<th>HTC Trophy</th>
<th>Iphone 3gs</th>
<th>Iphone 4</th>
<th>Moto Cliq</th>
<th>myTouch 3g</th>
</tr>
</thead>
<tbody>
<tr>
<td>BlackBerry 8220</td>
<td>8</td>
<td>38</td>
<td>21</td>
<td>68</td>
<td>34</td>
<td>40</td>
<td>35</td>
<td>47</td>
<td>61</td>
</tr>
<tr>
<td>BlackBerry 8330</td>
<td>38</td>
<td>21</td>
<td>39</td>
<td>58</td>
<td>42</td>
<td>31</td>
<td>21</td>
<td>41</td>
<td>72</td>
</tr>
<tr>
<td>BlackBerry 9630</td>
<td>35</td>
<td>26</td>
<td>12</td>
<td>74</td>
<td>32</td>
<td>44</td>
<td>23</td>
<td>29</td>
<td>63</td>
</tr>
<tr>
<td>HTC EVO 4G</td>
<td>6</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>HTC Trophy</td>
<td>80</td>
<td>42</td>
<td>71</td>
<td>91</td>
<td>38</td>
<td>77</td>
<td>27</td>
<td>54</td>
<td>95</td>
</tr>
<tr>
<td>Iphone 3gs</td>
<td>21</td>
<td>6</td>
<td>24</td>
<td>36</td>
<td>36</td>
<td>2</td>
<td>3</td>
<td>16</td>
<td>29</td>
</tr>
<tr>
<td>Iphone 4</td>
<td>94</td>
<td>83</td>
<td>93</td>
<td>99</td>
<td>33</td>
<td>92</td>
<td>5</td>
<td>40</td>
<td>99</td>
</tr>
<tr>
<td>Moto Cliq</td>
<td>14</td>
<td>5</td>
<td>9</td>
<td>33</td>
<td>11</td>
<td>23</td>
<td>3</td>
<td>2</td>
<td>69</td>
</tr>
<tr>
<td>myTouch 3g</td>
<td>21</td>
<td>78</td>
<td>7</td>
<td>99</td>
<td>82</td>
<td>82</td>
<td>87</td>
<td>96</td>
<td>6</td>
</tr>
</tbody>
</table>

*Figure 21 Number of images in the evidence set that exceeded the database bounds*

### Specific Camera Image Attribution

Specific Camera Image Attribution follows all the methods utilized in Camera Image Attribution with the addition of Photo Response Non Uniformity. Methods in Camera Image Attribution can be used to narrow down the possible cameras down to a single or a few camera models.

Photo Response Non Uniformity requires that a set of test images be taken with the suspect camera(s) for analysis which may not be available in some investigations. Minimizing the number of suspect cameras with the Camera Image Attribution method can make camera requests more reasonable and capturing test data less time consuming. Even though acquiring the source camera can often be difficult, PRNU is a very valuable tool because it can tell the exact camera which took the picture in question. In Figure 22, a comparison of three PRNU samples to an unknown image easily identified the matching camera as a Blackberry 8220. In Figure 23, an iPhone 4 test image is compared against four PRNU samples including two different iPhone 4 PRNU samples. The correlation difference between the alternate iPhone 4 and true iPhone 4 clearly shows how PRNU can still determine Specific Camera Image Attribution.
after Social Media processing. A complete collection of PRNU comparison graphs and further analysis is included in Appendix C.

Figure 22 Facebook PRNU Comparison: Blackberry 8220 Test Image vs. PRNU sample: (red) Blackberry 8220 (green) Blackberry 8330 (blue) Blackberry 9630

Figure 23 Google Plus PRNU Comparison: Iphone 4 Test Image vs. PRNU sample: (red) HTC Trophy (green) Iphone 3gs (blue) Iphone 4 Alt (Aqua) Iphone 4
CHAPTER VI
CONCLUSION

In this paper, I proposed three methods of determining image attribution when different information is required. This research leveraged the immense research in image authentication but also incorporated Social Media specific fingerprints and information to overcome the processing that makes traditional image authentication techniques difficult. Three Image Attribution Methods were proposed and tested with an image test corpus. Some of the Image Attribution techniques provided high certainty while others provided less certainty. The combined techniques in each method provided sufficient certainty. However, Image Attribution much like Image Authentication will always be vulnerable to forgeries. By utilizing several methods, image forgery is much more difficult. As Social Media and image processing continues to change and evolve, this research should be built upon to deal with changing processing standards.
CHAPTER VII
FURTHER RESEARCH

Research into Image Authentication has exploded in the last few years. This thesis utilized image dimensions, JPEG structure, metadata information, quantization tables, supporting social media documents, compression level analysis, and photo response non-uniformity. Further analysis that could add further fingerprints include: Discrete Cosine Transform Analysis and JPEG Compression Error Level Analysis. These techniques could be applied to the Camera Model Image Attribution to further confirm the identity of a source camera model. Better Photo Response Non-Uniformity samples could be utilized using a more advanced denoising filter in the PRNU process as suggested by Amerini et Al. It is the author’s hope that this system can, in the future, be enhanced with further identifiers and the results better quantified into more unified probabilistic terms.
BIBLIOGRAPHY


APPENDIX A
FILE STRUCTURE ANALYSIS

Figure 24 File Structure after social media processing
(a) Blackberry 8330 Original (b) Facebook Processed (c) Myspace Processed (d) Google Plus Processed
Figure 25 File Structure after social media processing

(a) Blackberry 9630 Original
(b) Facebook Processed
(c) Myspace Processed
(d) Google Plus Processed
Figure 26 File Structure after social media processing

(a) HTC PC36100 Original
(b) Facebook Processed
(c) Myspace Processed
(d) Google Plus Processed

<table>
<thead>
<tr>
<th>0 -&gt; FFDB = JPEG Start [0]</th>
<th>0 -&gt; FFDB = JPEG Start [0]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 -&gt; FFE1 = APP</td>
<td>14 -&gt; FFE1 = APP</td>
</tr>
<tr>
<td>2C4 -&gt; FFDB = Quantization Table</td>
<td>366 -&gt; FFDB = Quantization Table</td>
</tr>
<tr>
<td>34A -&gt; FC0 = Baseline DCT</td>
<td>3AB -&gt; FC0 = Baseline DCT</td>
</tr>
<tr>
<td>35D -&gt; FFC4 = Huffman Table</td>
<td>3F0 -&gt; FFC0 = Baseline DCT</td>
</tr>
<tr>
<td>501 -&gt; FFDA = Start of Scan (SOS)</td>
<td>3B0 -&gt; FFDA = Start of Scan (SOS)</td>
</tr>
<tr>
<td>63C3 -&gt; FF9 = JPEG End [63C4]</td>
<td>3B4 -&gt; FFDA = Start of Scan (SOS)</td>
</tr>
<tr>
<td>63C5 -&gt; FFDB = Quantization Table</td>
<td>3B6 -&gt; FFDA = Start of Scan (SOS)</td>
</tr>
<tr>
<td>644B -&gt; FFC0 = Baseline DCT</td>
<td>3B9 -&gt; FFDA = Start of Scan (SOS)</td>
</tr>
<tr>
<td>645E -&gt; FFC4 = Huffman Table</td>
<td>18A0D -&gt; FFC4 = Huffman Table</td>
</tr>
<tr>
<td>6602 -&gt; FFDA = Start of Scan (SOS)</td>
<td>18A82 -&gt; FFDA = Start of Scan (SOS)</td>
</tr>
<tr>
<td>DD412 -&gt; FF9 = JPEG End [DD412]</td>
<td>1BD27 -&gt; FFC4 = Huffman Table</td>
</tr>
<tr>
<td></td>
<td>1BD4F -&gt; FFDA = Start of Scan (SOS)</td>
</tr>
<tr>
<td></td>
<td>37E36 -&gt; FF9 = JPEG End [37E36]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 -&gt; FFDB = JPEG Start [0]</td>
<td>0 -&gt; FFDB = JPEG Start [0]</td>
<td>0 -&gt; FFDB = JPEG Start [0]</td>
<td>0 -&gt; FFDB = JPEG Start [0]</td>
</tr>
<tr>
<td>C0E -&gt; FFDB = Quantization Table</td>
<td>C53 -&gt; FFDB = Quantization Table</td>
<td>C98 -&gt; FFC2 = Progressive DCT</td>
<td>C9E -&gt; FFC4 = Huffman Table</td>
</tr>
<tr>
<td>C98 -&gt; FFC2 = Progressive DCT</td>
<td>CC8 -&gt; FFC4 = Huffman Table</td>
<td>CE2 -&gt; FFDA = Start of Scan (SOS)</td>
<td>CC8 -&gt; FFC4 = Huffman Table</td>
</tr>
<tr>
<td>CB1 -&gt; FFC4 = Huffman Table</td>
<td>CE2 -&gt; FFDA = Start of Scan (SOS)</td>
<td>6989 -&gt; FFC4 = Huffman Table</td>
<td>CE2 -&gt; FFDA = Start of Scan (SOS)</td>
</tr>
<tr>
<td>C66 -&gt; FFC4 = Huffman Table</td>
<td>69E2 -&gt; FFDA = Start of Scan (SOS)</td>
<td>69E2 -&gt; FFDA = Start of Scan (SOS)</td>
<td>69E2 -&gt; FFDA = Start of Scan (SOS)</td>
</tr>
<tr>
<td>CFF -&gt; FFDA = Start of Scan (SOS)</td>
<td>984E -&gt; FFC4 = Huffman Table</td>
<td>9873 -&gt; FFDA = Start of Scan (SOS)</td>
<td>9873 -&gt; FFDA = Start of Scan (SOS)</td>
</tr>
<tr>
<td>551D -&gt; FFC4 = Huffman Table</td>
<td>A56A -&gt; FFC4 = Huffman Table</td>
<td>A56A -&gt; FFC4 = Huffman Table</td>
<td>A56A -&gt; FFC4 = Huffman Table</td>
</tr>
<tr>
<td>5544 -&gt; FFDA = Start of Scan (SOS)</td>
<td>A58F -&gt; FFDA = Start of Scan (SOS)</td>
<td>A58F -&gt; FFDA = Start of Scan (SOS)</td>
<td>A58F -&gt; FFDA = Start of Scan (SOS)</td>
</tr>
<tr>
<td>6851 -&gt; FFC4 = Huffman Table</td>
<td>AE6A -&gt; FFC4 = Huffman Table</td>
<td>AE6A -&gt; FFC4 = Huffman Table</td>
<td>AE6A -&gt; FFC4 = Huffman Table</td>
</tr>
<tr>
<td>6874 -&gt; FFDA = Start of Scan (SOS)</td>
<td>AF11 -&gt; FFDA = Start of Scan (SOS)</td>
<td>AF11 -&gt; FFDA = Start of Scan (SOS)</td>
<td>AF11 -&gt; FFDA = Start of Scan (SOS)</td>
</tr>
<tr>
<td>69C1 -&gt; FFC4 = Huffman Table</td>
<td>AFF1 -&gt; FFC4 = Huffman Table</td>
<td>AFF1 -&gt; FFC4 = Huffman Table</td>
<td>AFF1 -&gt; FFC4 = Huffman Table</td>
</tr>
<tr>
<td>69E3 -&gt; FFDA = Start of Scan (SOS)</td>
<td>B019 -&gt; FFDA = Start of Scan (SOS)</td>
<td>B019 -&gt; FFDA = Start of Scan (SOS)</td>
<td>B019 -&gt; FFDA = Start of Scan (SOS)</td>
</tr>
<tr>
<td>6AFF -&gt; FFC4 = Huffman Table</td>
<td>131BA -&gt; FFDA = Start of Scan (SOS)</td>
<td>131BA -&gt; FFDA = Start of Scan (SOS)</td>
<td>131BA -&gt; FFDA = Start of Scan (SOS)</td>
</tr>
<tr>
<td>6817 -&gt; FFDA = Start of Scan (SOS)</td>
<td>14EC2 -&gt; FFC4 = Huffman Table</td>
<td>14EC2 -&gt; FFC4 = Huffman Table</td>
<td>14EC2 -&gt; FFC4 = Huffman Table</td>
</tr>
<tr>
<td>6B26 -&gt; FFC4 = Huffman Table</td>
<td>14EE4 -&gt; FFDA = Start of Scan (SOS)</td>
<td>14EE4 -&gt; FFDA = Start of Scan (SOS)</td>
<td>14EE4 -&gt; FFDA = Start of Scan (SOS)</td>
</tr>
<tr>
<td>6B48 -&gt; FFDA = Start of Scan (SOS)</td>
<td>18A0D -&gt; FFC4 = Huffman Table</td>
<td>18A0D -&gt; FFC4 = Huffman Table</td>
<td>18A0D -&gt; FFC4 = Huffman Table</td>
</tr>
<tr>
<td>9753 -&gt; FFDA = Start of Scan (SOS)</td>
<td>18A82 -&gt; FFDA = Start of Scan (SOS)</td>
<td>18A82 -&gt; FFDA = Start of Scan (SOS)</td>
<td>18A82 -&gt; FFDA = Start of Scan (SOS)</td>
</tr>
<tr>
<td>B450 -&gt; FFC4 = Huffman Table</td>
<td>1BD27 -&gt; FFC4 = Huffman Table</td>
<td>1BD27 -&gt; FFC4 = Huffman Table</td>
<td>1BD27 -&gt; FFC4 = Huffman Table</td>
</tr>
<tr>
<td>B471 -&gt; FFDA = Start of Scan (SOS)</td>
<td>1BD4F -&gt; FFDA = Start of Scan (SOS)</td>
<td>1BD4F -&gt; FFDA = Start of Scan (SOS)</td>
<td>1BD4F -&gt; FFDA = Start of Scan (SOS)</td>
</tr>
<tr>
<td>D518 -&gt; FFC4 = Huffman Table</td>
<td>37E36 -&gt; FF9 = JPEG End [37E36]</td>
<td>37E36 -&gt; FF9 = JPEG End [37E36]</td>
<td>37E36 -&gt; FF9 = JPEG End [37E36]</td>
</tr>
<tr>
<td>D539 -&gt; FFDA = Start of Scan (SOS)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EFB1 -&gt; FFC4 = Huffman Table</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EFDC -&gt; FFDA = Start of Scan (SOS)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1C1C3 -&gt; FF9 = JPEG End [1C1C3]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 27 File Structure after social media processing
(a) HTC Trophy Original (b) Facebook Processed (c) Myspace Processed (d) Google Plus Processed
**Figure 28 File Structure after social media processing**

(a) Moto Cliq Original (b) Facebook Processed (c) Myspace Processed (d) Google Plus Processed
**Figure 29 File Structure after social media processing**

(a) Iphone 3gs Original

0 -> FFD8 = JPEG Start [0]
29E -> FFD8 = JPEG Start [29E]
2A0 -> FFD8 = Quantization Table
2E5 -> FFD8 = Quantization Table
32A -> FFC0 = Baseline DCT
33D -> FFC4 = Huffman Table
35E -> FFC4 = Huffman Table
415 -> FFC4 = Huffman Table
436 -> FFC4 = Huffman Table
4ED -> FFDA = Start of Scan (SOS)
2216 -> FFDB = Quantization Table
22A0 -> FFDB = Huffman Table
2283 -> FFDC = Huffman Table
2457 -> FFDA = Start of Scan (SOS)
17668B -> FFDB = Huffman Table

(b) Facebook Processed

0 -> FFD8 = JPEG Start [0]

C0E -> FFDB = Quantization Table
C53 -> FFDB = Quantization Table
C98 -> FFDC = Progressive DCT
CAB -> FFC4 = Huffman Table
CC8 -> FFC4 = Huffman Table
CE2 -> FFDA = Start of Scan (SOS)
A1D8 -> FFDC = Huffman Table
A206 -> FFDA = Start of Scan (SOS)
1FC78 -> FFDC = Huffman Table
1FC9E -> FFDA = Start of Scan (SOS)
20357 -> FFDC = Huffman Table
2037F -> FFDA = Start of Scan (SOS)
20DBE -> FFDC = Huffman Table
20DFB -> FFDA = Start of Scan (SOS)
32655 -> FFDC = Huffman Table
3267D -> FFDA = Start of Scan (SOS)
593F3 -> FFDA = Start of Scan (SOS)
5B825 -> FFDC = Huffman Table
5B847 -> FFDA = Start of Scan (SOS)
5E2EF -> FFDC = Huffman Table
5E311 -> FFDA = Start of Scan (SOS)
614AB -> FFDC = Huffman Table
614D2 -> FFDA = Start of Scan (SOS)
A80D2 -> FFDC = Huffman Table

(c) Myspace Processed

0 -> FFD8 = JPEG Start [0]

C0E -> FFDB = Quantization Table
C53 -> FFDB = Quantization Table
C98 -> FFDC = Progressive DCT
CAB -> FFC4 = Huffman Table
CC8 -> FFC4 = Huffman Table
CE2 -> FFDA = Start of Scan (SOS)
A1D8 -> FFC4 = Huffman Table
A206 -> FFDA = Start of Scan (SOS)
1FC78 -> FFC4 = Huffman Table
1FC9E -> FFDA = Start of Scan (SOS)
20357 -> FFC4 = Huffman Table
2037F -> FFDA = Start of Scan (SOS)
20DBE -> FFC4 = Huffman Table
20DFB -> FFDA = Start of Scan (SOS)
32655 -> FFC4 = Huffman Table
3267D -> FFDA = Start of Scan (SOS)
593F3 -> FFDA = Start of Scan (SOS)
5B825 -> FFC4 = Huffman Table
5B847 -> FFDA = Start of Scan (SOS)
5E2EF -> FFC4 = Huffman Table
5E311 -> FFDA = Start of Scan (SOS)
614AB -> FFC4 = Huffman Table
614D2 -> FFDA = Start of Scan (SOS)
A80D2 -> FFDC = Huffman Table

(d) Google Plus Processed

0 -> FFD8 = JPEG Start [0]

14 -> FFE1 = APP
310 -> FFD8 = JPEG Start [310]
324 -> FFDB = Quantization Table
369 -> FFD8 = Quantization Table
3AE -> FFDC = Huffman Table
3C1 -> FFC4 = Huffman Table
3DD -> FFC4 = Huffman Table
41A -> FFC4 = Huffman Table
435 -> FFC4 = Huffman Table
456 -> FFDA = Start of Scan (SOS)
18A6 -> FFD9 = JPEG End [18A7]
18A8 -> FFE1 = APP
19CC -> FFDB = Quantization Table
1A52 -> FFDC = Huffman Table
1A65 -> FFDC = Huffman Table
1A84 -> FFC4 = Huffman Table
1AD2 -> FFC4 = Huffman Table
1AEE -> FFC4 = Huffman Table
1B23 -> FFDA = Start of Scan (SOS)
104F8D -> FFD9 = JPEG End [104F8D]
Figure 30 File Structure after social media processing

(a) Iphone 4 Original (b) Facebook Processed (c) Myspace Processed (d) Google Plus Processed
Figure 31 File Structure after social media processing

(a) myTouch Original
(b) Facebook Processed
(c) Myspace Processed
(d) Google Plus Processed
(e) Tumblr Processed
APPENDIX B
FACEBOOK EXTRACTED METADATA INFORMATION

BlackBerry8220_SN2584B9F8 HQ 2/10

Date Taken: April 19, 2011 at 6:29 pm
Orientation: 1
Camera Model: BlackBerry 8220
Exposure: 0/1

June 3, 2013 at 8:56 am

Blackberry8330_SN3042C00D HQ 6/10

Date Taken: September 5, 2010 at 2:02 pm
Orientation: 1
Camera Model: BlackBerry 8330

June 8, 2013 at 10:45 am
Blackberry9630_SN30E6CC37 HQ 10/10

Date Taken: June 2, 2013 at 8:30 am
Orientation: 1
Camera Model: BlackBerry 9630
Exposure: 0/1

June 2, 2013 at 10:16 pm

HTCPC36100_SNHT0CVHL05640 HQ 3/10

Date Taken: May 18, 2013 at 10:51 am
Latitude: 34.1277583333
Longitude: -118.14115556
Camera Model: PC36100
ISO Speed: 114
Focal Length: 457/100

June 6, 2013 at 7:11 pm
HtcTrophy_SNHT16JM600679 HQ 1/10

MotorolaCliqu_SNJ16NJ376L HQ 2/10
APPENDIX C

PRNU COMPARISON GRAPHS

The PRNU comparison graphs are organized by test image source. Each graph shows several PRNU samples on the X axis. Incorrect comparisons are represented by color circles (e.g. BB Pearl Flip PRNU Sample vs. iPhone 4 test image). Correct comparisons are represented by black non-circle symbols (e.g. BB Pearl Flip PRNU Sample vs. BB Pearl Flip test image). Since no test images were included to compare against the Alternate Iphone 4 image, Iphone 4 Alt PRNU sample vs. Iphone 4 test image was used for the correct comparison.

These graphs illustrate several properties of the test images and the social media processing including the effect of social media processing on PRNU, the effect of camera sources on PRNU, and the variation between various social media processing.

The effect of social media processing on PRNU can be seen by comparing the original test image results against social media test image results. There are higher correlations between PRNU samples and the original test images than social media processed test images. This is to be expected because of the social media processing. However, the correlations between PRNU samples and social media processed test images is still high enough to differentiate positive PRNU matches.

The effect of camera sources on PRNU can be seen by comparing the PRNU sample correlations against one another within the same test image sets. Two different properties represent themselves, overall correlation and correlation variance. The HTC Trophy has a relatively low overall correlation while the Blackberry Pearl Flip has a relatively high overall correlation. The HTC Trophy has a relatively low variance in correlation while the Blackberry Curve has a relatively high variance in correlation.

The variation between various social media processing can be seen by comparing the various social media test image results against one another. It is clear that original
test images provide more correlation than any of the social media processed test images. However, the fact that each set of social media processed images provides different levels of correlation shows that each website is applying different processing to the images.

**Original Test Images**

![Original Test Images](image)

Figure 32 PRNU Samples (Left to Right): BB Pearl Flip, BB Curve, BB Tour, HTC EVO 4g, and HTC Trophy

**Google Plus Test Images**

![Google Plus Test Images](image)

Figure 34 PRNU Samples (Left to Right): BB Pearl Flip, BB Curve, BB Tour, HTC EVO 4g, and HTC Trophy
Figure 35 PRNU Samples (Left to Right): Moto Cliq, Mytouch 3g, Iphone 3gs, Iphone 4, Iphone 4 Alt

Facebook Test Images

Figure 36 PRNU Samples (Left to Right): BB Pearl Flip, BB Curve, BB Tour, HTC EVO 4g, and HTC Trophy

MySpace Test Images

Figure 37 PRNU Samples (Left to Right): Moto Cliq, Mytouch 3g, Iphone 3gs, Iphone 4, Iphone 4 Alt

Figure 38 PRNU Samples (Left to Right): BB Pearl Flip, BB Curve, BB Tour, HTC EVO 4g, and HTC Trophy
Figure 39 PRNU Samples (Left to Right): Moto Cliq, Mytouch 3g, Iphone 3gs, Iphone 4, Iphone 4 Alt