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 cell phone 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

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 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.

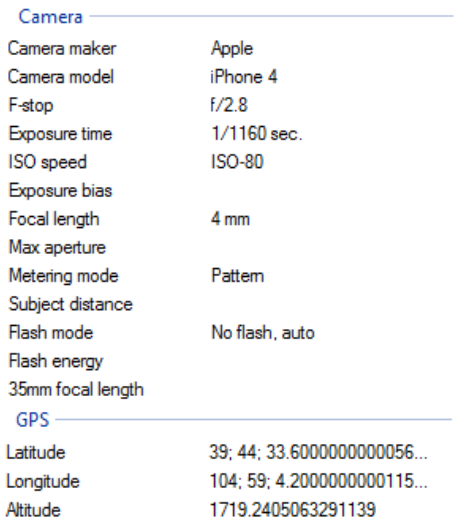


Figure 1 Iphone 4 Metadata

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.

Luminance

1	1	1	1	2	4	5	6
1	1	1	1	2	5	6	5
1	1	1	2	4	5	6	5
1	1	2	2	5	8	8	6
1	2	3	5	6	10	10	7
2	3	5	6	8	10	11	9
4	6	7	8	10	12	12	10
7	9	9	9	11	10	10	9

Chrominance

1	1	2	4	9	9	9	9
1	2	2	6	9	9	9	9
2	2	5	9	9	9	9	9
4	6	9	9	9	9	9	9
9	9	9	9	9	9	9	9
9	9	9	9	9	9	9	9
9	9	9	9	9	9	9	9
9	9	9	9	9	9	9	9

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.

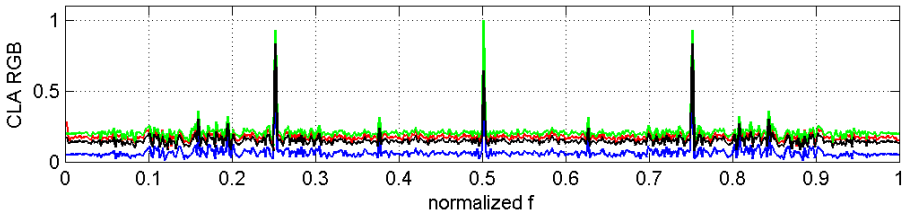


Figure 3 Mytouch 3g CLA

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.

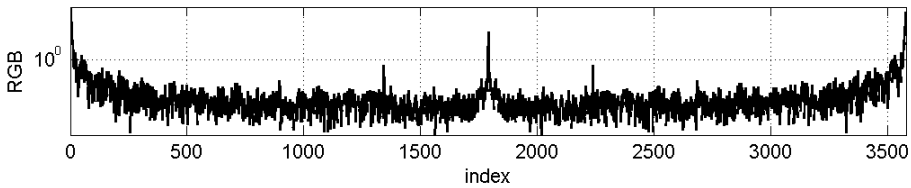


Figure 4 Mytouch 3g CFA

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.

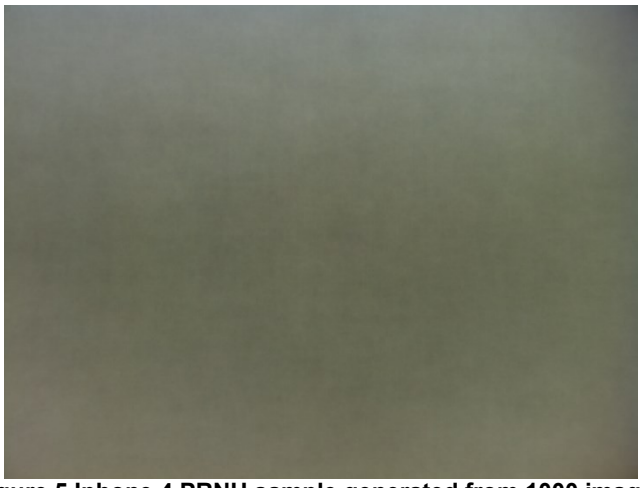


Figure 5 Iphone 4 PRNU sample generated from 1000 images

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.

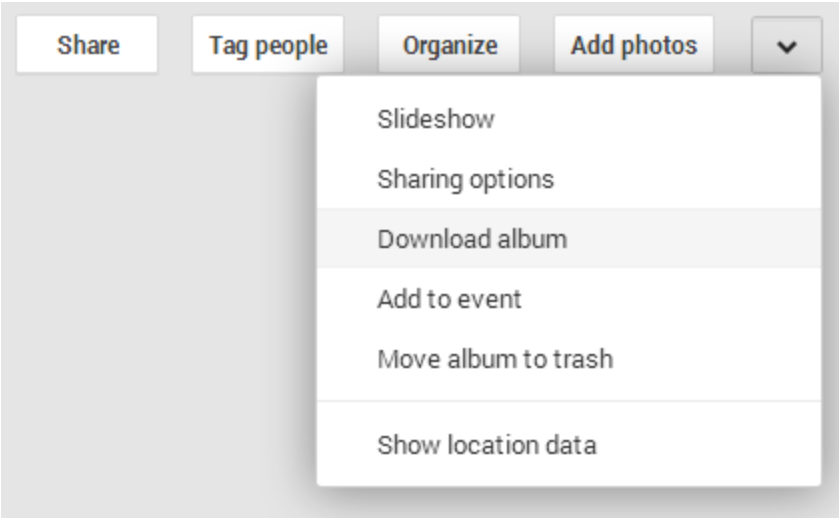


Figure 6 Google Plus Download Album Feature

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.

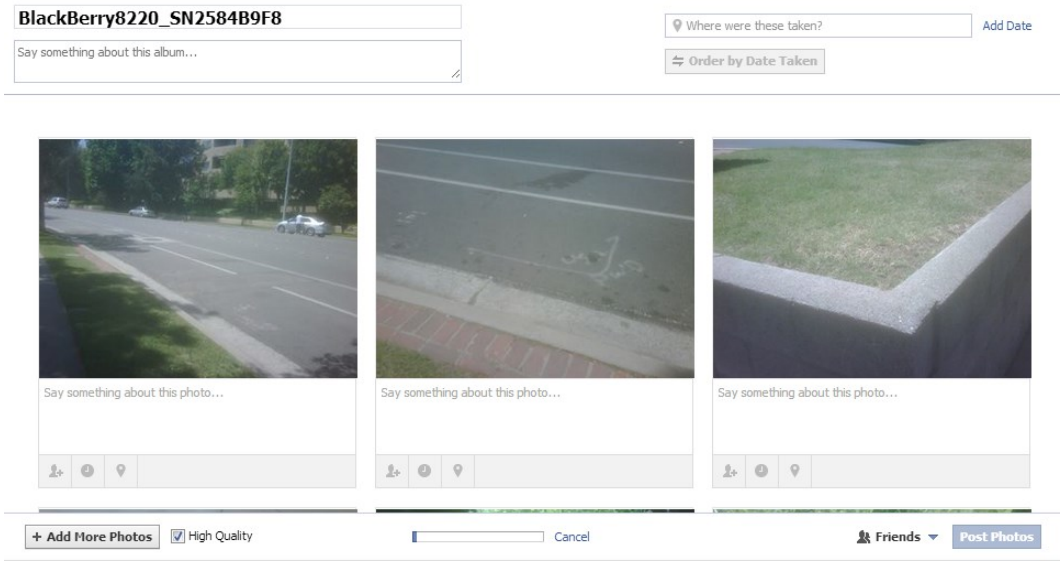


Figure 7 Facebook image upload interface

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.

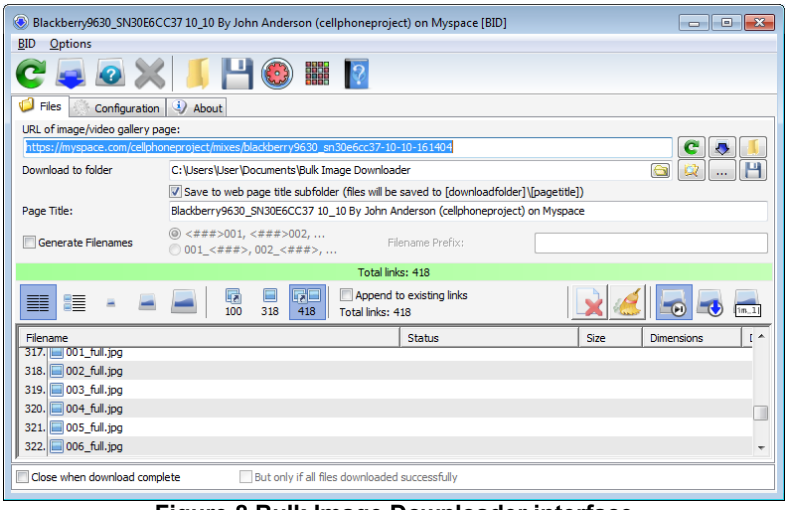


Figure 8 Bulk Image Downloader interface

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.

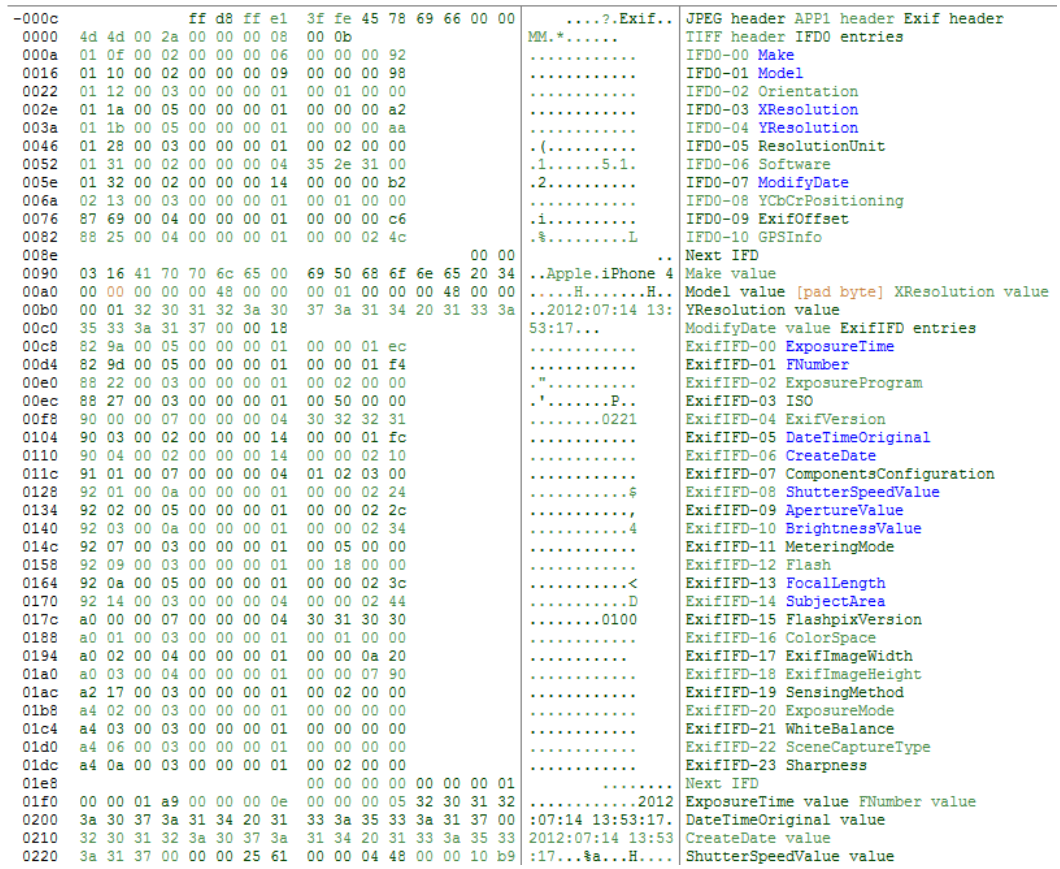


Figure 9 Metadata displayed by ExifToolGUI

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.

Make	Apple
Model	iPhone 4
Orientation	Horizontal (normal)
XResolution	72
YResolution	72
ResolutionUnit	inches
Software	5.1
ModifyDate	2012:07:14 13:53:17
YCbCrPositioning	Centered
	---- **ExifIFD** ----
ExposureTime	1/425
FNumber	2.8
ExposureProgram	Program AE
ISO	80
ExifVersion	0221
DateTimeOriginal	2012:07:14 13:53:17
CreateDate	2012:07:14 13:53:17
ComponentsConfiguratior	Y, Cb, Cr, -
ShutterSpeedValue	1/425
ApertureValue	2.8
BrightnessValue	7.765796703
MeteringMode	Multi-segment
Flash	Auto, Did not fire
FocalLength	3.9 mm
SubjectArea	1295 967 699 696
FlashpixVersion	0100
ColorSpace	sRGB
ExifImageWidth	2592
ExifImageHeight	1936

Figure 10 Further information can be decoded from the imbedded Exif information.

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

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.

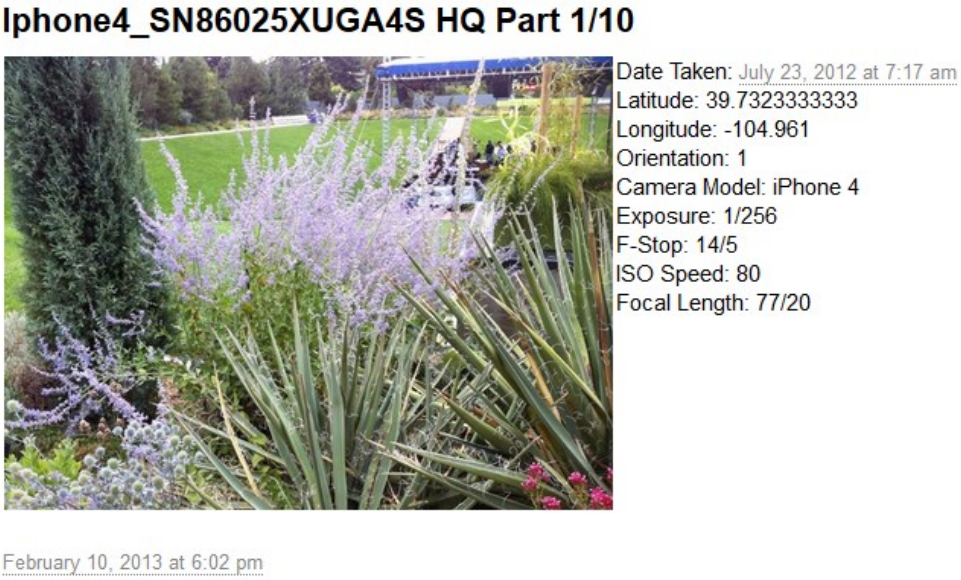


Figure 11 Camera Metadata information from a Facebook archive download

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.

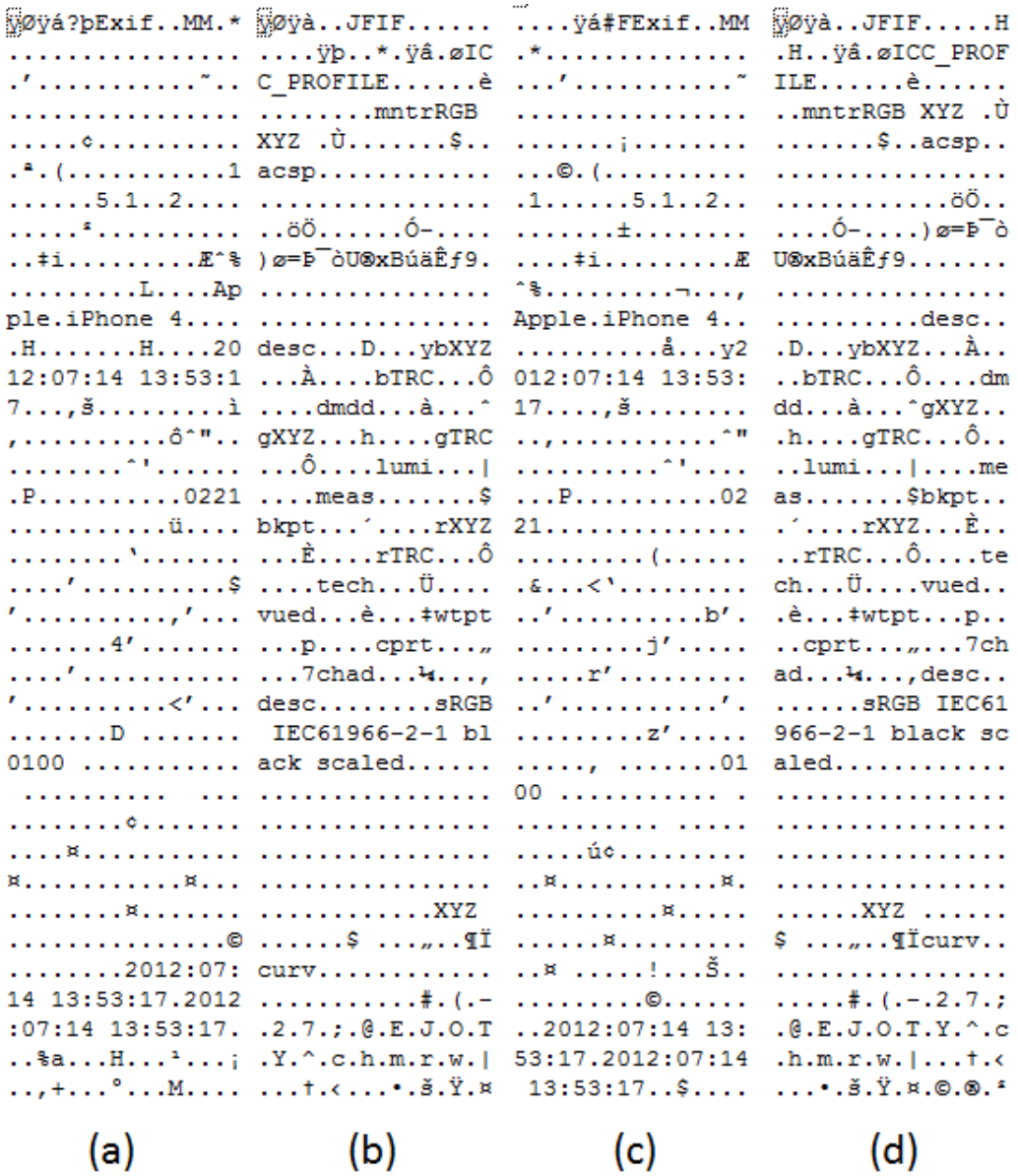


Figure 12 iPhone 4 Metadata after Social Media Website processing

(a) Original (b) Facebook (c) Google Plus (d) Myspace

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.

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

Figure 13 Image Dimensions after Social Media resizing

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

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.

Google Plus Blackberry 8220 Quantization Table

Luminance							
3	2	10	10	10	8	8	8
2	10	10	8	8	8	8	8
10	10	10	10	8	8	8	8
10	10	8	8	8	8	8	8
10	8	8	8	8	10	10	8
8	8	8	8	8	10	12	8
8	8	8	8	10	12	12	8
8	8	8	8	10	8	10	8

Chrominance							
3	4	6	10	12	12	12	12
4	5	6	13	12	2	12	12
6	6	12	13	12	12	12	12
10	13	15	13	12	12	12	12
15	15	13	12	12	12	12	12
15	15	13	12	12	12	12	12
13	12	12	12	12	12	12	12
13	13	13	12	12	12	12	12

Google Plus Blackberry 8220 Thumbnail Quantization Table																
Luminance								Chrominance								
5	3	3	5	7	12	15	18	5	5	7	14	30	30	30	30
4	4	4	6	8	17	18	17	5	6	8	20	30	30	30	30
4	4	5	7	12	17	21	17	7	8	17	30	30	30	30	30
4	5	7	9	15	26	24	19	14	20	30	30	30	30	30	30
5	7	11	17	20	33	31	23	30	30	30	30	30	30	30	30
7	11	17	19	24	31	34	28	30	30	30	30	30	30	30	30
15	19	23	26	31	36	36	30	30	30	30	30	30	30	30	30
22	28	29	29	34	30	31	30	30	30	30	30	30	30	30	30

Google Plus Iphone 4 Quantization Table 3																
Luminance								Chrominance								
3	2	10	10	10	8	8	8	3	4	6	10	13	13	12	12
2	10	10	10	10	8	8	8	4	5	6	12	12	12	12	12
10	10	10	10	8	8	8	8	6	6	12	12	12	12	12	12
10	10	10	8	8	8	8	8	10	13	12	12	12	12	12	12
10	10	8	8	8	10	10	8	13	12	12	12	12	12	12	12
8	8	8	8	8	10	13	8	13	12	12	12	12	12	12	12
8	8	8	8	10	13	13	8	12	12	12	12	12	12	12	12
8	8	8	8	10	8	10	8	12	12	12	12	12	12	12	12

Google Plus Iphone 4 Thumbnail Quantization Table																
Luminance								Chrominance								
5	3	3	5	7	12	15	18	5	5	7	14	30	30	30	30
4	4	4	6	8	17	18	17	5	6	8	20	30	30	30	30
4	4	5	7	12	17	21	17	7	8	17	30	30	30	30	30
4	5	7	9	15	26	24	19	14	20	30	30	30	30	30	30
5	7	11	17	20	33	31	23	30	30	30	30	30	30	30	30
7	11	17	19	24	31	34	28	30	30	30	30	30	30	30	30
15	19	23	26	31	36	36	30	30	30	30	30	30	30	30	30
22	28	29	29	34	30	31	30	30	30	30	30	30	30	30	30

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

Facebook Blackberry 8220 Quantization Table 1																
Luminance								Chrominance								
9	6	5	9	13	22	28	33	9	10	13	25	53	53	53	53
6	6	8	10	14	31	32	30	10	11	14	36	53	53	53	53
8	7	9	13	22	31	37	30	13	14	30	53	53	53	53	53
8	9	12	16	28	47	43	33	25	36	53	53	53	53	53	53
10	12	20	30	37	59	56	42	53	53	53	53	53	53	53	53
13	19	30	35	44	56	61	50	53	53	53	53	53	53	53	53
26	35	42	47	56	65	65	55	53	53	53	53	53	53	53	53
39	50	51	53	60	54	56	53	53	53	53	53	53	53	53	53

Facebook Blackberry 8220 Quantization Table 2																
Luminance								Chrominance								
7	5	4	7	10	17	21	26	7	8	10	20	42	42	42	42
5	5	6	8	11	24	25	23	8	9	11	28	42	42	42	42
6	5	7	10	17	24	29	24	10	11	24	42	42	42	42	42
6	7	9	12	21	37	34	26	20	28	42	42	42	42	42	42
8	9	16	24	29	46	43	32	42	42	42	42	42	42	42	42
10	15	23	27	34	44	47	39	42	42	42	42	42	42	42	42
21	27	33	37	43	51	50	42	42	42	42	42	42	42	42	42
30	39	40	41	47	42	43	42	42	42	42	42	42	42	42	42

Facebook Blackberry 8220 Quantization Table 3																
Luminance								Chrominance								
9	6	6	9	14	23	30	35	10	10	14	27	57	57	57	57
7	7	8	11	15	34	35	32	10	12	15	38	57	57	57	57
8	8	9	14	23	33	40	32	14	15	32	57	57	57	57	57
8	10	13	17	30	50	46	36	27	38	57	57	57	57	57	57
10	13	21	32	39	63	60	45	57	57	57	57	57	57	57	57
14	20	32	37	47	60	66	53	57	57	57	57	57	57	57	57
28	37	45	50	60	70	70	59	57	57	57	57	57	57	57	57
42	53	55	57	65	58	60	57	57	57	57	57	57	57	57	57

Test Image																
Luminance								Chrominance								
7	5	4	7	10	17	21	26	7	8	10	20	42	42	42	42
5	5	6	8	11	24	25	23	8	9	11	28	42	42	42	42
6	5	7	10	17	24	29	24	10	11	24	42	42	42	42	42
6	7	9	12	21	37	34	26	20	28	42	42	42	42	42	42
8	9	16	24	29	46	43	32	42	42	42	42	42	42	42	42
10	15	23	27	34	44	47	39	42	42	42	42	42	42	42	42
21	27	33	37	43	51	50	42	42	42	42	42	42	42	42	42
30	39	40	41	47	42	43	42	42	42	42	42	42	42	42	42

Figure 15 Images uploaded to Facebook have several Quantization Tables choices

Myspace Blackberry 8220																
Luminance								Chrominance								
5	3	3	5	7	12	15	18	5	5	7	14	30	30	30	30
4	4	4	6	8	17	18	17	5	6	8	20	30	30	30	30
4	4	5	7	12	17	21	17	7	8	17	30	30	30	30	30
4	5	7	9	15	26	24	19	14	20	30	30	30	30	30	30
5	7	11	17	20	33	31	23	30	30	30	30	30	30	30	30
7	11	17	19	24	31	34	28	30	30	30	30	30	30	30	30
15	19	23	26	31	36	36	30	30	30	30	30	30	30	30	30
22	28	29	29	34	30	31	30	30	30	30	30	30	30	30	30

Myspace Motorola Cliq																
Luminance								Chrominance								
5	3	3	5	7	12	15	18	5	5	7	14	30	30	30	30
4	4	4	6	8	17	18	17	5	6	8	20	30	30	30	30
4	4	5	7	12	17	21	17	7	8	17	30	30	30	30	30
4	5	7	9	15	26	24	19	14	20	30	30	30	30	30	30
5	7	11	17	20	33	31	23	30	30	30	30	30	30	30	30
7	11	17	19	24	31	34	28	30	30	30	30	30	30	30	30
15	19	23	26	31	36	36	30	30	30	30	30	30	30	30	30
22	28	29	29	34	30	31	30	30	30	30	30	30	30	30	30

Myspace Apple Iphone 4																
Luminance								Chrominance								
5	3	3	5	7	12	15	18	5	5	7	14	30	30	30	30
4	4	4	6	8	17	18	17	5	6	8	20	30	30	30	30
4	4	5	7	12	17	21	17	7	8	17	30	30	30	30	30
4	5	7	9	15	26	24	19	14	20	30	30	30	30	30	30
5	7	11	17	20	33	31	23	30	30	30	30	30	30	30	30
7	11	17	19	24	31	34	28	30	30	30	30	30	30	30	30
15	19	23	26	31	36	36	30	30	30	30	30	30	30	30	30
22	28	29	29	34	30	31	30	30	30	30	30	30	30	30	30

Figure 16 Images uploaded to Myspace share the same Quantization Tables

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.

(a)

0 -> FFD8 = JPEG Start [0]
2 -> FFE1 = APP
1B0 -> FFDB = Quantization Table
1F5 -> FFDB = Quantization Table
23A -> FFC4 = Huffman Table
25B -> FFC4 = Huffman Table
312 -> FFC4 = Huffman Table
333 -> FFC4 = Huffman Table
3EA -> FFC0 = Baseline DCT
3FD -> FFDA = Start of Scan (SOS)
53046 -> FFD9 = JPEG End [53046]

(b)

0 -> FFD8 = JPEG Start [0]
C14 -> FFDB = Quantization Table
C59 -> FFDB = Quantization Table
C9E -> FFC2 = Progressive DCT
CB1 -> FFC4 = Huffman Table
CCD -> FFC4 = Huffman Table
CE6 -> FFC4 = Huffman Table
CFF -> FFDA = Start of Scan (SOS)
4BD8 -> FFC4 = Huffman Table
4C01 -> FFDA = Start of Scan (SOS)
7BB7 -> FFC4 = Huffman Table
7BDC -> FFDA = Start of Scan (SOS)
7C7D -> FFC4 = Huffman Table
7CA0 -> FFDA = Start of Scan (SOS)
7E2A -> FFC4 = Huffman Table
7E55 -> FFDA = Start of Scan (SOS)
8219 -> FFC4 = Huffman Table
8244 -> FFDA = Start of Scan (SOS)
1033C -> FFDA = Start of Scan (SOS)
11952 -> FFC4 = Huffman Table
11975 -> FFDA = Start of Scan (SOS)
13074 -> FFC4 = Huffman Table
13096 -> FFDA = Start of Scan (SOS)
14CE1 -> FFC4 = Huffman Table
14D0A -> FFDA = Start of Scan (SOS)
266B8 -> FFD9 = JPEG End [266B8]

(c)

0 -> FFD8 = JPEG Start [0]
C0E -> FFDB = Quantization Table
C53 -> FFDB = Quantization Table
C98 -> FFC2 = Progressive DCT
CAB -> FFC4 = Huffman Table
CC7 -> FFC4 = Huffman Table
CE1 -> FFDA = Start of Scan (SOS)
53AE -> FFC4 = Huffman Table
53DA -> FFDA = Start of Scan (SOS)
9D68 -> FFC4 = Huffman Table
9D8D -> FFDA = Start of Scan (SOS)
A112 -> FFC4 = Huffman Table
A138 -> FFDA = Start of Scan (SOS)
A75C -> FFC4 = Huffman Table
A78A -> FFDA = Start of Scan (SOS)
B353 -> FFC4 = Huffman Table
B37D -> FFDA = Start of Scan (SOS)
1578C -> FFDA = Start of Scan (SOS)
16DA1 -> FFC4 = Huffman Table
16DC3 -> FFDA = Start of Scan (SOS)
1901C -> FFC4 = Huffman Table
1903F -> FFDA = Start of Scan (SOS)
1B71D -> FFC4 = Huffman Table
1B745 -> FFDA = Start of Scan (SOS)
37B5A -> FFD9 = JPEG End [37B5A]

(d)

0 -> FFD8 = JPEG Start [0]
14 -> FFE1 = APP
262 -> FFD8 = JPEG Start [262]
276 -> FFDB = Quantization Table
2BB -> FFDB = Quantization Table
300 -> FFC0 = Baseline DCT
313 -> FFC4 = Huffman Table
330 -> FFC4 = Huffman Table
368 -> FFC4 = Huffman Table
383 -> FFC4 = Huffman Table
3A2 -> FFDA = Start of Scan (SOS)
13FA -> FFD9 = JPEG End [13FB]
13FC -> FFE1 = APP
1520 -> FFDB = Quantization Table
15A6 -> FFC0 = Baseline DCT
15B9 -> FFC4 = Huffman Table
15D7 -> FFC4 = Huffman Table
1623 -> FFC4 = Huffman Table
163F -> FFC4 = Huffman Table
1673 -> FFDA = Start of Scan (SOS)
40E96 -> FFD9 = JPEG End [40E96]

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.

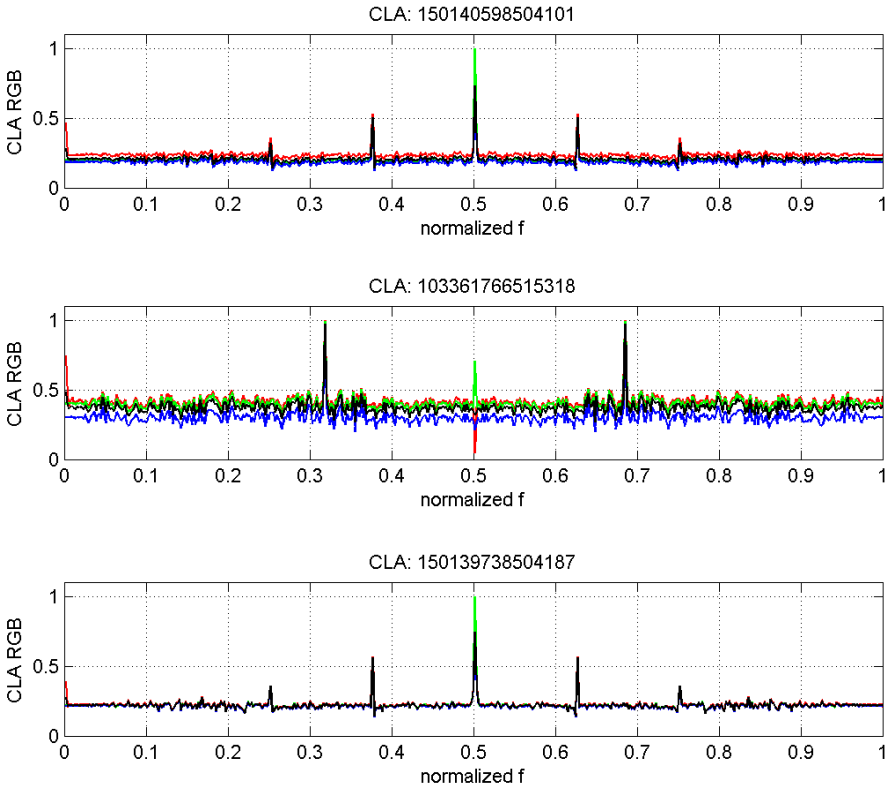


Figure 18 Visual Analysis of the CLA graphs:
(a) unknown image (b) Iphone 4 (c) Blackberry 9630

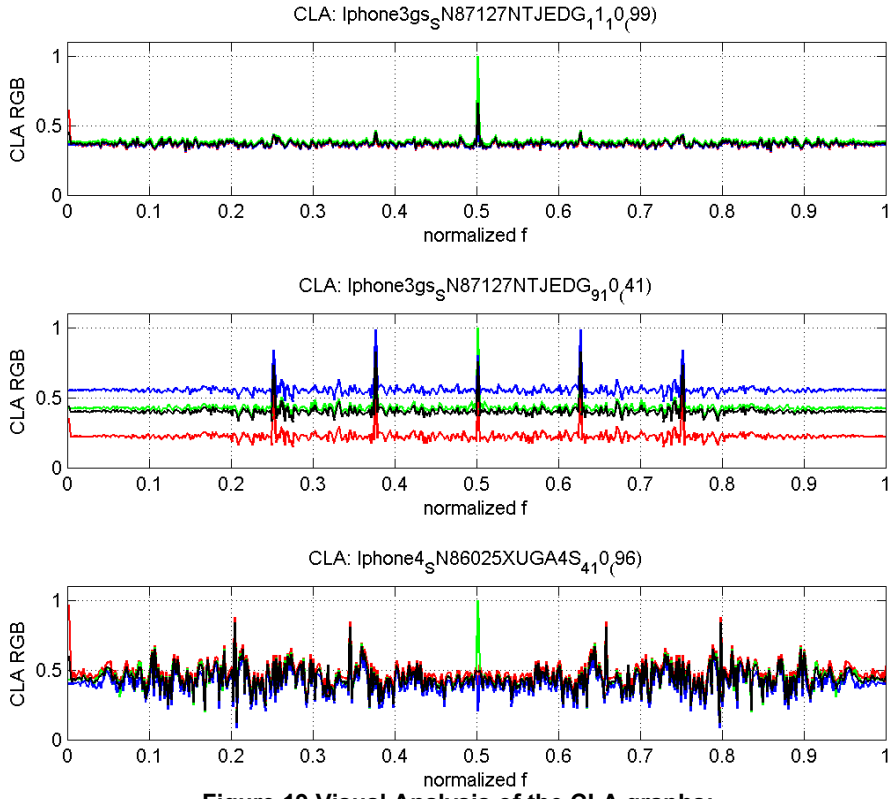


Figure 19 Visual Analysis of the CLA graphs:
(a) unknown image (b) Iphone 3gs (c) Iphone 4

Evidence

Database

MS CLA (10^{-4})	BlackBerry 8220	BlackBerry 8330	BlackBerry 9630	HTC EVO 4G	HTC Trophy	Iphone 3gs	Iphone 4	Moto Cliq	myTouch 3g
BlackBerry 8220	8.137924	1.035192	4.128476	0.160340	0.170629	0.410947	0.130122	0.387368	5.104571
BlackBerry 8330	0.802366	0.480355	0.637146	1.402276	0.447099	1.067850	0.280445	0.015096	0.575658
BlackBerry 9630	-0.520764	-1.985024	4.328719	1.195647	-0.820402	0.560466	0.398243	0.479906	3.734199
HTC EVO 4G	2.009978	0.857954	2.781525	5.272244	1.379014	2.927852	1.194785	1.736767	0.722307
HTC Trophy	1.615223	0.517807	1.348160	0.685387	0.625574	-0.544331	0.307442	-0.844984	-0.956594
Iphone 3gs	-0.862377	-0.106058	-0.308055	3.025248	0.519058	2.738466	0.175664	2.455793	1.916648
Iphone 4	0.146709	-0.487489	0.779164	1.296052	2.894473	0.553082	1.845296	-0.255366	0.556869
Moto Cliq	0.939129	-0.947985	3.068425	0.374696	0.680593	0.541077	-1.434051	0.426081	1.344976
myTouch 3g	-0.576551	-2.116025	5.637958	1.310072	1.240868	2.149378	1.463614	1.092350	6.064369

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

Evidence

Bounds

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

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.

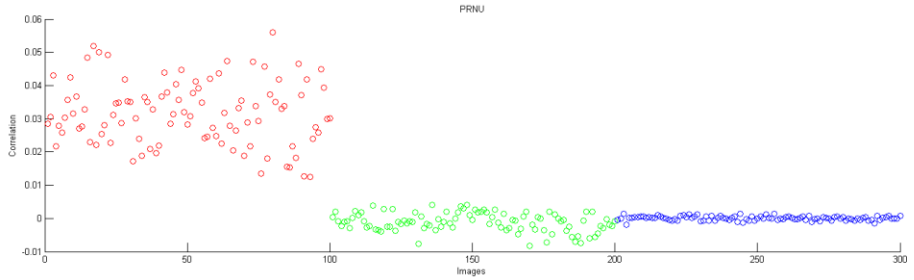


Figure 22 Facebook PRNU Comparison: Blackberry 8220 Test Image vs. PRNU sample: (red) Blackberry 8220 (green) Blackberry 8330 (blue) Blackberry 9630

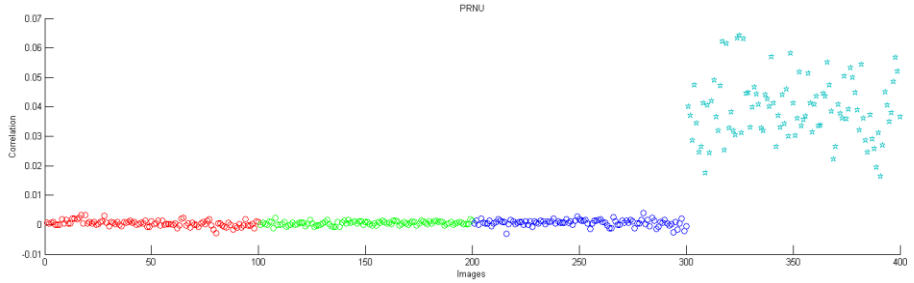


Figure 23 Google Plus PRNU Comparison: Iphone 4 Test Image vs. PRNU sample: (red) HTC Trophy (green) Iphone 3gs (blue) Iphone 4 Alt (Aqua) Iphone 4

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

Amerini, Irene, et al. "Analysis of denoising filters for photo response non uniformity noise extraction in source camera identification." *16th International Conference on Digital Signal Processing*. Santorini-Hellas, Greece, 2009. 1-7.

Castiglione, Aniello, Giuseppe Cattaneo and Alfredo De Santis. "A Forensic Analysis of Images on Online Social Networks." *Third International Conference on Intelligent Networking and Collaborative Systems*. Fukuoka, Japan: IEEE, 2011. 679-684.

Facebook. *Facebook Information for Law Enforcement Authorities*. n.d. <https://www.facebook.com/safety/groups/law/guidelines/>.

Google. *Google Transparency Report*. n.d. <http://www.google.com/transparencyreport/userdatarequests/>.

Helenek, Katherine. "Facebook ®: Do You Leave a Trace? A Forensic Analysis of Facebook ® Artifacts." *AAFS 64th Annual Scientific Meeting*. Atlanta: American Academy of Forensic Science, 2012. 148.

Lukas, Jan, Jessica Fridrich and Miroslav Goljan. "Digital Camera Identification From Sensor Pattern Noise." *IEEE Transactions on Information Forensics and Security* 1.2 (2006): 205-214.

Myspace. *Myspace Law Enforcement Guidelines*. n.d. <https://www.askmyspace.com/t5/Legal-Policy/Law-Enforcement-Guidelines/ba-p/38505>.

Popescu, Alin C. and Hany Farid. "Exposing digital forgeries in color filter array interpolated images." *IEEE Transactions on Signal Processing* 53.10 (2005): 3948-3959.

—. "Statistical Tools for Digital Forensics." *6th International Workshop on Information Hiding*. Toronto, Canada, 2004. 128-147.

APPENDIX A

FILE STRUCTURE ANALYSIS

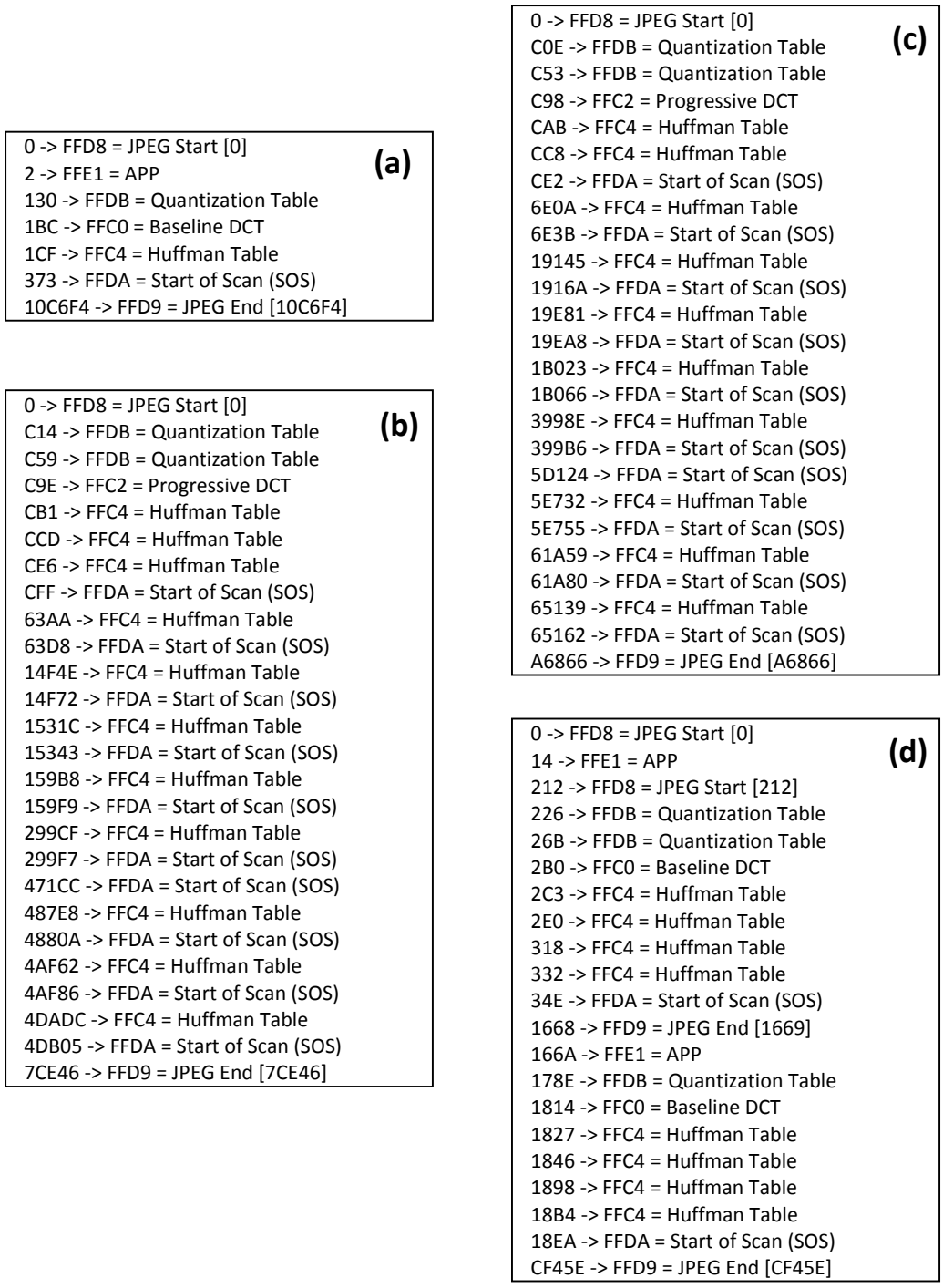


Figure 24 File Structure after social media processing
(a) Blackberry 8330 Original (b) Facebook Processed (c) Myspace Processed (d) Google Plus Processed

(a)

0 -> FFD8 = JPEG Start [0]
2 -> FFE1 = APP
19E -> FFDB = Quantization Table
224 -> FFC4 = Huffman Table
3C8 -> FFC0 = Baseline DCT
3DB -> FFDA = Start of Scan (SOS)
8AC21 -> FFD9 = JPEG End [8AC21]

(b)

0 -> FFD8 = JPEG Start [0]
C14 -> FFDB = Quantization Table
C59 -> FFDB = Quantization Table
C9E -> FFC2 = Progressive DCT
CB1 -> FFC4 = Huffman Table
CCD -> FFC4 = Huffman Table
CE7 -> FFC4 = Huffman Table
D01 -> FFDA = Start of Scan (SOS)
7B25 -> FFC4 = Huffman Table
7B52 -> FFDA = Start of Scan (SOS)
EB13 -> FFC4 = Huffman Table
EB38 -> FFDA = Start of Scan (SOS)
FE2C -> FFC4 = Huffman Table
FE51 -> FFDA = Start of Scan (SOS)
10805 -> FFC4 = Huffman Table
1083E -> FFDA = Start of Scan (SOS)
14009 -> FFC4 = Huffman Table
14032 -> FFDA = Start of Scan (SOS)
2BC47 -> FFDA = Start of Scan (SOS)
2E07F -> FFC4 = Huffman Table
2E0A7 -> FFDA = Start of Scan (SOS)
34912 -> FFC4 = Huffman Table
34936 -> FFDA = Start of Scan (SOS)
39FF4 -> FFC4 = Huffman Table
3A01C -> FFDA = Start of Scan (SOS)
73A93 -> FFD9 = JPEG End [73A93]

(c)

0 -> FFD8 = JPEG Start [0]
C0E -> FFDB = Quantization Table
C53 -> FFDB = Quantization Table
C98 -> FFC2 = Progressive DCT
CAB -> FFC4 = Huffman Table
CC7 -> FFC4 = Huffman Table
CE1 -> FFDA = Start of Scan (SOS)
7AAF -> FFC4 = Huffman Table
7ADD -> FFDA = Start of Scan (SOS)
E6B0 -> FFC4 = Huffman Table
E6D4 -> FFDA = Start of Scan (SOS)
F5EE -> FFC4 = Huffman Table
F613 -> FFDA = Start of Scan (SOS)
FD01 -> FFC4 = Huffman Table
FD3A -> FFDA = Start of Scan (SOS)
134BF -> FFC4 = Huffman Table
134E8 -> FFDA = Start of Scan (SOS)
29193 -> FFDA = Start of Scan (SOS)
2B5C5 -> FFC4 = Huffman Table
2B5E9 -> FFDA = Start of Scan (SOS)
3156E -> FFC4 = Huffman Table
31591 -> FFDA = Start of Scan (SOS)
36886 -> FFC4 = Huffman Table
368AD -> FFDA = Start of Scan (SOS)
713D5 -> FFD9 = JPEG End [713D5]

(d)

0 -> FFD8 = JPEG Start [0]
14 -> FFE1 = APP
212 -> FFD8 = JPEG Start [212]
226 -> FFDB = Quantization Table
26B -> FFDB = Quantization Table
2B0 -> FFC0 = Baseline DCT
2C3 -> FFC4 = Huffman Table
2E0 -> FFC4 = Huffman Table
318 -> FFC4 = Huffman Table
332 -> FFC4 = Huffman Table
34E -> FFDA = Start of Scan (SOS)
1668 -> FFD9 = JPEG End [1669]
166A -> FFE1 = APP
178E -> FFDB = Quantization Table
1814 -> FFC0 = Baseline DCT
1827 -> FFC4 = Huffman Table
1846 -> FFC4 = Huffman Table
1898 -> FFC4 = Huffman Table
18B4 -> FFC4 = Huffman Table
18EA -> FFDA = Start of Scan (SOS)
CF45E -> FFD9 = JPEG End [CF45E]

Figure 25 File Structure after social media processing
(a) Blackberry 9630 Original (b) Facebook Processed (c) Myspace Processed (d) Google Plus Processed

(a)

0 -> FFD8 = JPEG Start [0]
2 -> FFE1 = APP
2C2 -> FFD8 = JPEG Start [2C2]
2C4 -> FFDB = Quantization Table
34A -> FFC0 = Baseline DCT
35D -> FFC4 = Huffman Table
501 -> FFDA = Start of Scan (SOS)
63C3 -> FFD9 = JPEG End [63C4]
63C5 -> FFDB = Quantization Table
644B -> FFC0 = Baseline DCT
645E -> FFC4 = Huffman Table
6602 -> FFDA = Start of Scan (SOS)
DD412 -> FFD9 = JPEG End [DD412]

(b)

0 -> FFD8 = JPEG Start [0]
C14 -> FFDB = Quantization Table
C59 -> FFDB = Quantization Table
C9E -> FFC2 = Progressive DCT
CB1 -> FFC4 = Huffman Table
CCD -> FFC4 = Huffman Table
CE6 -> FFC4 = Huffman Table
CFF -> FFDA = Start of Scan (SOS)
551D -> FFC4 = Huffman Table
5544 -> FFDA = Start of Scan (SOS)
6851 -> FFC4 = Huffman Table
6874 -> FFDA = Start of Scan (SOS)
69C1 -> FFC4 = Huffman Table
69E3 -> FFDA = Start of Scan (SOS)
6AFF -> FFC4 = Huffman Table
6B17 -> FFDA = Start of Scan (SOS)
6B26 -> FFC4 = Huffman Table
6B4B -> FFDA = Start of Scan (SOS)
9753 -> FFDA = Start of Scan (SOS)
B450 -> FFC4 = Huffman Table
B471 -> FFDA = Start of Scan (SOS)
D518 -> FFC4 = Huffman Table
D539 -> FFDA = Start of Scan (SOS)
EFB1 -> FFC4 = Huffman Table
EFDC -> FFDA = Start of Scan (SOS)
1C1C3 -> FFD9 = JPEG End [1C1C3]

(c)

0 -> FFD8 = JPEG Start [0]
C0E -> FFDB = Quantization Table
C53 -> FFDB = Quantization Table
C98 -> FFC2 = Progressive DCT
CAB -> FFC4 = Huffman Table
CC8 -> FFC4 = Huffman Table
CE2 -> FFDA = Start of Scan (SOS)
69B9 -> FFC4 = Huffman Table
69E2 -> FFDA = Start of Scan (SOS)
984E -> FFC4 = Huffman Table
9873 -> FFDA = Start of Scan (SOS)
A56A -> FFC4 = Huffman Table
A58F -> FFDA = Start of Scan (SOS)
AEEA -> FFC4 = Huffman Table
AF11 -> FFDA = Start of Scan (SOS)
AFF1 -> FFC4 = Huffman Table
B019 -> FFDA = Start of Scan (SOS)
131BA -> FFDA = Start of Scan (SOS)
14EC2 -> FFC4 = Huffman Table
14EE4 -> FFDA = Start of Scan (SOS)
18A0D -> FFC4 = Huffman Table
18A32 -> FFDA = Start of Scan (SOS)
1BD27 -> FFC4 = Huffman Table
1BD4F -> FFDA = Start of Scan (SOS)
37E36 -> FFD9 = JPEG End [37E36]

(d)

0 -> FFD8 = JPEG Start [0]
14 -> FFE1 = APP
352 -> FFD8 = JPEG Start [352]
366 -> FFDB = Quantization Table
3AB -> FFDB = Quantization Table
3F0 -> FFC0 = Baseline DCT
403 -> FFC4 = Huffman Table
421 -> FFC4 = Huffman Table
462 -> FFC4 = Huffman Table
47D -> FFC4 = Huffman Table
4A0 -> FFDA = Start of Scan (SOS)
132E -> FFD9 = JPEG End [132F]
1330 -> FFE1 = APP
1454 -> FFDB = Quantization Table
14DA -> FFC0 = Baseline DCT
14ED -> FFC4 = Huffman Table
150C -> FFC4 = Huffman Table
1555 -> FFC4 = Huffman Table
1571 -> FFC4 = Huffman Table
15A0 -> FFDA = Start of Scan (SOS)
43612 -> FFD9 = JPEG End [43612]

Figure 26 File Structure after social media processing
(a) HTC PC36100 Original (b) Facebook Processed (c) Myspace Processed (d) Google Plus Processed

(a)

0 -> FFD8 = JPEG Start [0]
2 -> FFE1 = APP
1282 -> FFD8 = JPEG Start [1282]
1284 -> FFDB = Quantization Table
12C9 -> FFDB = Quantization Table
130E -> FFC0 = Baseline DCT
1321 -> FFC4 = Huffman Table
1342 -> FFC4 = Huffman Table
13F9 -> FFC4 = Huffman Table
141A -> FFC4 = Huffman Table
14D1 -> FFDA = Start of Scan (SOS)
4BAF -> FFD9 = JPEG End [4BB0]
4BB2 -> FFE1 = APP
55F9 -> FFDB = Quantization Table
563E -> FFDB = Quantization Table
5683 -> FFC0 = Baseline DCT
5696 -> FFC4 = Huffman Table
56B7 -> FFC4 = Huffman Table
576E -> FFC4 = Huffman Table
578F -> FFC4 = Huffman Table
5846 -> FFDA = Start of Scan (SOS)
177FE0 -> FFD9 = JPEG End [177FE0]

(b)

0 -> FFD8 = JPEG Start [0]
C14 -> FFDB = Quantization Table
C59 -> FFDB = Quantization Table
C9E -> FFC2 = Progressive DCT
CB1 -> FFC4 = Huffman Table
CCE -> FFC4 = Huffman Table
CE8 -> FFC4 = Huffman Table
D02 -> FFDA = Start of Scan (SOS)
8BE0 -> FFC4 = Huffman Table
8C0F -> FFDA = Start of Scan (SOS)
16B62 -> FFC4 = Huffman Table
16B89 -> FFDA = Start of Scan (SOS)
1730A -> FFC4 = Huffman Table
17332 -> FFDA = Start of Scan (SOS)
175FB -> FFC4 = Huffman Table
1763C -> FFDA = Start of Scan (SOS)
239DA -> FFC4 = Huffman Table
23A04 -> FFDA = Start of Scan (SOS)
43AFD -> FFDA = Start of Scan (SOS)
45F38 -> FFC4 = Huffman Table
45F5B -> FFDA = Start of Scan (SOS)
49E76 -> FFC4 = Huffman Table
49E98 -> FFDA = Start of Scan (SOS)
4C944 -> FFC4 = Huffman Table
4C96C -> FFDA = Start of Scan (SOS)
90884 -> FFD9 = JPEG End [90884]

(c)

0 -> FFD8 = JPEG Start [0]
C0E -> FFDB = Quantization Table
C53 -> FFDB = Quantization Table
C98 -> FFC2 = Progressive DCT
CAB -> FFC4 = Huffman Table
CC8 -> FFC4 = Huffman Table
CE2 -> FFDA = Start of Scan (SOS)
9D53 -> FFC4 = Huffman Table
9D83 -> FFDA = Start of Scan (SOS)
1E7DC -> FFC4 = Huffman Table
1E806 -> FFDA = Start of Scan (SOS)
1FFEC -> FFC4 = Huffman Table
20014 -> FFDA = Start of Scan (SOS)
20A89 -> FFC4 = Huffman Table
20AD0 -> FFDA = Start of Scan (SOS)
3AE11 -> FFC4 = Huffman Table
3AE3A -> FFDA = Start of Scan (SOS)
65A9D -> FFDA = Start of Scan (SOS)
67EC8 -> FFC4 = Huffman Table
67EEC -> FFDA = Start of Scan (SOS)
6D08F -> FFC4 = Huffman Table
6D0B2 -> FFDA = Start of Scan (SOS)
7112A -> FFC4 = Huffman Table
71152 -> FFDA = Start of Scan (SOS)
B6349 -> FFD9 = JPEG End [B6349]

(d)

0 -> FFD8 = JPEG Start [0]
14 -> FFE1 = APP
1292 -> FFD8 = JPEG Start [1292]
12A6 -> FFDB = Quantization Table
12EB -> FFDB = Quantization Table
1330 -> FFC0 = Baseline DCT
1343 -> FFC4 = Huffman Table
1361 -> FFC4 = Huffman Table
13A1 -> FFC4 = Huffman Table
13BC -> FFC4 = Huffman Table
13DC -> FFDA = Start of Scan (SOS)
2804 -> FFD9 = JPEG End [2805]
2806 -> FFE1 = APP
2A27 -> FFDB = Quantization Table
2AAD -> FFC0 = Baseline DCT
2AC0 -> FFC4 = Huffman Table
2ADF -> FFC4 = Huffman Table
2B37 -> FFC4 = Huffman Table
2B53 -> FFC4 = Huffman Table
2B8B -> FFDA = Start of Scan (SOS)
D8E2D -> FFD9 = JPEG End [D8E2D]

Figure 27 File Structure after social media processing
(a) HTC Trophy Original (b) Facebook Processed (c) Myspace Processed (d) Google Plus Processed

(a)

0 -> FFD8 = JPEG Start [0]
2 -> FFE1 = APP
1D6 -> FFD8 = JPEG Start [1D6]
1D8 -> FFDB = Quantization Table
25E -> FFC0 = Baseline DCT
271 -> FFC4 = Huffman Table
415 -> FFDA = Start of Scan (SOS)
F0B -> FFD9 = JPEG End [F0C]
F0D -> FFDB = Quantization Table
F93 -> FFC0 = Baseline DCT
FA6 -> FFC4 = Huffman Table
114A -> FFDA = Start of Scan (SOS)
F4928 -> FFD9 = JPEG End [F4928]

(b)

0 -> FFD8 = JPEG Start [0]
C14 -> FFDB = Quantization Table
C59 -> FFDB = Quantization Table
C9E -> FFC2 = Progressive DCT
CB1 -> FFC4 = Huffman Table
CCD -> FFC4 = Huffman Table
CE6 -> FFC4 = Huffman Table
CFF -> FFDA = Start of Scan (SOS)
7FC5 -> FFC4 = Huffman Table
7FF3 -> FFDA = Start of Scan (SOS)
FE9A -> FFC4 = Huffman Table
FEC0 -> FFDA = Start of Scan (SOS)
1045E -> FFC4 = Huffman Table
10487 -> FFDA = Start of Scan (SOS)
10680 -> FFC4 = Huffman Table
106B9 -> FFDA = Start of Scan (SOS)
12A26 -> FFC4 = Huffman Table
12A4F -> FFDA = Start of Scan (SOS)
264E8 -> FFDA = Start of Scan (SOS)
2893B -> FFC4 = Huffman Table
2895D -> FFDA = Start of Scan (SOS)
2C561 -> FFC4 = Huffman Table
2C582 -> FFDA = Start of Scan (SOS)
2E3BD -> FFC4 = Huffman Table
2E3E5 -> FFDA = Start of Scan (SOS)
6163A -> FFD9 = JPEG End [6163A]

(c)

0 -> FFD8 = JPEG Start [0]
C0E -> FFDB = Quantization Table
C53 -> FFDB = Quantization Table
C98 -> FFC2 = Progressive DCT
CAB -> FFC4 = Huffman Table
CC8 -> FFC4 = Huffman Table
CE2 -> FFDA = Start of Scan (SOS)
8EAC -> FFC4 = Huffman Table
8ED9 -> FFDA = Start of Scan (SOS)
1736D -> FFC4 = Huffman Table
17394 -> FFDA = Start of Scan (SOS)
18763 -> FFC4 = Huffman Table
18789 -> FFDA = Start of Scan (SOS)
18DA9 -> FFC4 = Huffman Table
18DE6 -> FFDA = Start of Scan (SOS)
200A9 -> FFC4 = Huffman Table
200D2 -> FFDA = Start of Scan (SOS)
3D2DD -> FFDA = Start of Scan (SOS)
3F70E -> FFC4 = Huffman Table
3F732 -> FFDA = Start of Scan (SOS)
445C0 -> FFC4 = Huffman Table
445E2 -> FFDA = Start of Scan (SOS)
47B61 -> FFC4 = Huffman Table
47B89 -> FFDA = Start of Scan (SOS)
7A480 -> FFD9 = JPEG End [7A480]

(d)

0 -> FFD8 = JPEG Start [0]
14 -> FFE1 = APP
254 -> FFD8 = JPEG Start [254]
268 -> FFDB = Quantization Table
2AD -> FFDB = Quantization Table
2F2 -> FFC0 = Baseline DCT
305 -> FFC4 = Huffman Table
322 -> FFC4 = Huffman Table
35B -> FFC4 = Huffman Table
375 -> FFC4 = Huffman Table
394 -> FFDA = Start of Scan (SOS)
13E4 -> FFD9 = JPEG End [13E5]
13E6 -> FFE1 = APP
150A -> FFDB = Quantization Table
1590 -> FFC0 = Baseline DCT
15A3 -> FFC4 = Huffman Table
15C2 -> FFC4 = Huffman Table
1615 -> FFC4 = Huffman Table
1631 -> FFC4 = Huffman Table
1669 -> FFDA = Start of Scan (SOS)
8D0BD -> FFD9 = JPEG End [8D0BD]

Figure 28 File Structure after social media processing
(a) Moto Cliq Original (b) Facebook Processed (c) Myspace Processed (d) Google Plus Processed

(a)

0 -> FFD8 = JPEG Start [0]
2 -> FFE1 = APP
29E -> FFD8 = JPEG Start [29E]
2A0 -> FFDB = Quantization Table
2E5 -> FFDB = 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 -> FFD9 = JPEG End [2217]
221A -> FFDB = Quantization Table
22A0 -> FFC0 = Baseline DCT
22B3 -> FFC4 = Huffman Table
2457 -> FFDA = Start of Scan (SOS)
17668B -> FFD9 = JPEG End [17668B]

(b)

0 -> FFD8 = JPEG Start [0]
C14 -> FFDB = Quantization Table
C59 -> FFDB = Quantization Table
C9E -> FFC2 = Progressive DCT
CB1 -> FFC4 = Huffman Table
CCE -> FFC4 = Huffman Table
CE8 -> FFC4 = Huffman Table
D02 -> FFDA = Start of Scan (SOS)
AC4A -> FFC4 = Huffman Table
AC76 -> FFDA = Start of Scan (SOS)
224BD -> FFC4 = Huffman Table
224E5 -> FFDA = Start of Scan (SOS)
22D75 -> FFC4 = Huffman Table
22D9F -> FFDA = Start of Scan (SOS)
23A7A -> FFC4 = Huffman Table
23AB8 -> FFDA = Start of Scan (SOS)
39A4F -> FFC4 = Huffman Table
39A77 -> FFDA = Start of Scan (SOS)
64081 -> FFDA = Start of Scan (SOS)
66492 -> FFC4 = Huffman Table
664B6 -> FFDA = Start of Scan (SOS)
69525 -> FFC4 = Huffman Table
6954A -> FFDA = Start of Scan (SOS)
6CE9C -> FFC4 = Huffman Table
6CEC4 -> FFDA = Start of Scan (SOS)
B1C48 -> FFD9 = JPEG End [B1C48]

(c)

0 -> FFD8 = JPEG Start [0]
C0E -> FFDB = Quantization Table
C53 -> FFDB = Quantization Table
C98 -> FFC2 = 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 -> FFD9 = JPEG End [A80D2]

(d)

0 -> FFD8 = JPEG Start [0]
14 -> FFE1 = APP
310 -> FFD8 = JPEG Start [310]
324 -> FFDB = Quantization Table
369 -> FFDB = Quantization Table
3AE -> FFC0 = Baseline DCT
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 -> FFC0 = Baseline DCT
1A65 -> FFC4 = 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 29 File Structure after social media processing
(a) Iphone 3gs Original (b) Facebook Processed (c) Myspace Processed (d) Google Plus Processed

(a)

0 -> FFD8 = JPEG Start [0]
2 -> FFE1 = APP
380 -> FFD8 = JPEG Start [380]
382 -> FFDB = Quantization Table
3C7 -> FFDB = Quantization Table
40C -> FFC0 = Baseline DCT
41F -> FFC4 = Huffman Table
440 -> FFC4 = Huffman Table
4F7 -> FFC4 = Huffman Table
518 -> FFC4 = Huffman Table
5CF -> FFDA = Start of Scan (SOS)
1EEC -> FFD9 = JPEG End [1EED]
2002 -> FFDB = Quantization Table
2088 -> FFC0 = Baseline DCT
209B -> FFC4 = Huffman Table
223F -> FFDA = Start of Scan (SOS)
2166C1 -> FFD9 = JPEG End [2166C1]

(b)

0 -> FFD8 = JPEG Start [0]
C14 -> FFDB = Quantization Table
C59 -> FFDB = Quantization Table
C9E -> FFC2 = Progressive DCT
CB1 -> FFC4 = Huffman Table
CCE -> FFC4 = Huffman Table
CE7 -> FFC4 = Huffman Table
D00 -> FFDA = Start of Scan (SOS)
7FE9 -> FFC4 = Huffman Table
8017 -> FFDA = Start of Scan (SOS)
122A5 -> FFC4 = Huffman Table
122C8 -> FFDA = Start of Scan (SOS)
12347 -> FFC4 = Huffman Table
1236B -> FFDA = Start of Scan (SOS)
12429 -> FFC4 = Huffman Table
12462 -> FFDA = Start of Scan (SOS)
15EA7 -> FFC4 = Huffman Table
15ED1 -> FFDA = Start of Scan (SOS)
310A4 -> FFDA = Start of Scan (SOS)
334B3 -> FFC4 = Huffman Table
334D6 -> FFDA = Start of Scan (SOS)
343C6 -> FFC4 = Huffman Table
343E9 -> FFDA = Start of Scan (SOS)
34F51 -> FFC4 = Huffman Table
34F7A -> FFDA = Start of Scan (SOS)
72120 -> FFD9 = JPEG End [72120]

(c)

0 -> FFD8 = JPEG Start [0]
C0E -> FFDB = Quantization Table
C53 -> FFDB = Quantization Table
C98 -> FFC2 = Progressive DCT
CAB -> FFC4 = Huffman Table
CC9 -> FFC4 = Huffman Table
CE2 -> FFDA = Start of Scan (SOS)
9366 -> FFC4 = Huffman Table
9394 -> FFDA = Start of Scan (SOS)
1B557 -> FFC4 = Huffman Table
1B57B -> FFDA = Start of Scan (SOS)
1B8B1 -> FFC4 = Huffman Table
1B8D7 -> FFDA = Start of Scan (SOS)
1BB83 -> FFC4 = Huffman Table
1BBC2 -> FFDA = Start of Scan (SOS)
2A18B -> FFC4 = Huffman Table
2A1B4 -> FFDA = Start of Scan (SOS)
52480 -> FFDA = Start of Scan (SOS)
548AD -> FFC4 = Huffman Table
548CF -> FFDA = Start of Scan (SOS)
56854 -> FFC4 = Huffman Table
56877 -> FFDA = Start of Scan (SOS)
58209 -> FFC4 = Huffman Table
58232 -> FFDA = Start of Scan (SOS)
9FB7A -> FFD9 = JPEG End [9FB7A]

(d)

0 -> FFD8 = JPEG Start [0]
14 -> FFE1 = APP
3FE -> FFD8 = JPEG Start [3FE]
412 -> FFDB = Quantization Table
457 -> FFDB = Quantization Table
49C -> FFC0 = Baseline DCT
4AF -> FFC4 = Huffman Table
4CC -> FFC4 = Huffman Table
50B -> FFC4 = Huffman Table
525 -> FFC4 = Huffman Table
546 -> FFDA = Start of Scan (SOS)
166F -> FFD9 = JPEG End [1670]
1672 -> FFE1 = APP
1796 -> FFDB = Quantization Table
181C -> FFC0 = Baseline DCT
182F -> FFC4 = Huffman Table
184E -> FFC4 = Huffman Table
18A3 -> FFC4 = Huffman Table
18BE -> FFC4 = Huffman Table
18EA -> FFDA = Start of Scan (SOS)
B1C5E -> FFD9 = JPEG End [B1C5E]

Figure 30 File Structure after social media processing
(a) Iphone 4 Original (b) Facebook Processed (c) Myspace Processed (d) Google Plus Processed

(a)

0 -> FFD8 = JPEG Start [0]
2 -> FFE1 = APP
21E -> FFD8 = JPEG Start [21E]
220 -> FFDB = Quantization Table
2A6 -> FFC0 = Baseline DCT
2B9 -> FFC4 = Huffman Table
45D -> FFDA = Start of Scan (SOS)
30BF -> FFD9 = JPEG End [30C0]
30C1 -> FFDB = Quantization Table
3147 -> FFC0 = Baseline DCT
315A -> FFC4 = Huffman Table
32FE -> FFDA = Start of Scan (SOS)
B5B92 -> FFD9 = JPEG End [B5B92]

(b)

0 -> FFD8 = JPEG Start [0]
14 -> FFE1 = APP
2BC -> FFD8 = JPEG Start [2BC]
2D0 -> FFDB = Quantization Table
315 -> FFDB = Quantization Table
35A -> FFC0 = Baseline DCT
36D -> FFC4 = Huffman Table
38A -> FFC4 = Huffman Table
3C3 -> FFC4 = Huffman Table
3DE -> FFC4 = Huffman Table
3FE -> FFDA = Start of Scan (SOS)
137B -> FFD9 = JPEG End [137C]
137E -> FFE1 = APP
14A2 -> FFDB = Quantization Table
1528 -> FFC0 = Baseline DCT
153B -> FFC4 = Huffman Table
1559 -> FFC4 = Huffman Table
15A2 -> FFC4 = Huffman Table
15BE -> FFC4 = Huffman Table
15FA -> FFDA = Start of Scan (SOS)
93B6E -> FFD9 = JPEG End [93B6E]

(c)

0 -> FFD8 = JPEG Start [0]
C0E -> FFDB = Quantization Table
C53 -> FFDB = Quantization Table
C98 -> FFC2 = Progressive DCT
CAB -> FFC4 = Huffman Table
CC8 -> FFC4 = Huffman Table
CE2 -> FFDA = Start of Scan (SOS)
9C7B -> FFC4 = Huffman Table
9CA5 -> FFDA = Start of Scan (SOS)
1C5CF -> FFC4 = Huffman Table
1C5F4 -> FFDA = Start of Scan (SOS)
1D763 -> FFC4 = Huffman Table
1D788 -> FFDA = Start of Scan (SOS)
1EB6D -> FFC4 = Huffman Table
1EB9F -> FFDA = Start of Scan (SOS)
2805A -> FFC4 = Huffman Table
28082 -> FFDA = Start of Scan (SOS)
463B3 -> FFDA = Start of Scan (SOS)
487E1 -> FFC4 = Huffman Table
48806 -> FFDA = Start of Scan (SOS)
4EB9E -> FFC4 = Huffman Table
4EBC1 -> FFDA = Start of Scan (SOS)
54F53 -> FFC4 = Huffman Table
54F7A -> FFDA = Start of Scan (SOS)
9181D -> FFD9 = JPEG End [9181D]

(d)

0 -> FFD8 = JPEG Start [0]
C14 -> FFDB = Quantization Table
C59 -> FFDB = Quantization Table
C9E -> FFC2 = Progressive DCT
CB1 -> FFC4 = Huffman Table
CCD -> FFC4 = Huffman Table
CE7 -> FFC4 = Huffman Table
D01 -> FFDA = Start of Scan (SOS)
8C7C -> FFC4 = Huffman Table
8CA5 -> FFDA = Start of Scan (SOS)
15469 -> FFC4 = Huffman Table
1548E -> FFDA = Start of Scan (SOS)
15AF6 -> FFC4 = Huffman Table
15B1A -> FFDA = Start of Scan (SOS)
16334 -> FFC4 = Huffman Table
1635E -> FFDA = Start of Scan (SOS)
19516 -> FFC4 = Huffman Table
1953F -> FFDA = Start of Scan (SOS)
31F7F -> FFDA = Start of Scan (SOS)
3439D -> FFC4 = Huffman Table
343C0 -> FFDA = Start of Scan (SOS)
36FB8 -> FFC4 = Huffman Table
36FDA -> FFDA = Start of Scan (SOS)
39CCD -> FFC4 = Huffman Table
39CF5 -> FFDA = Start of Scan (SOS)
60650 -> FFD9 = JPEG End [60650]

(e)

0 -> FFD8 = JPEG Start [0]
14 -> FFE1 = APP
230 -> FFD8 = JPEG Start [230]
232 -> FFDB = Quantization Table
2B8 -> FFC0 = Baseline DCT
2CB -> FFC4 = Huffman Table
46F -> FFDA = Start of Scan (SOS)
30D1 -> FFD9 = JPEG End [30D2]
30D3 -> FFDB = Quantization Table
3118 -> FFDB = Quantization Table
315D -> FFC0 = Baseline DCT
3170 -> FFC4 = Huffman Table
318E -> FFC4 = Huffman Table
31D1 -> FFC4 = Huffman Table
31ED -> FFC4 = Huffman Table
321D -> FFDA = Start of Scan (SOS)
51768 -> FFD9 = JPEG End [51768]

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 8:55 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: -118.141155556
Camera Model: PC36100
ISO Speed: 114
Focal Length: 457/100

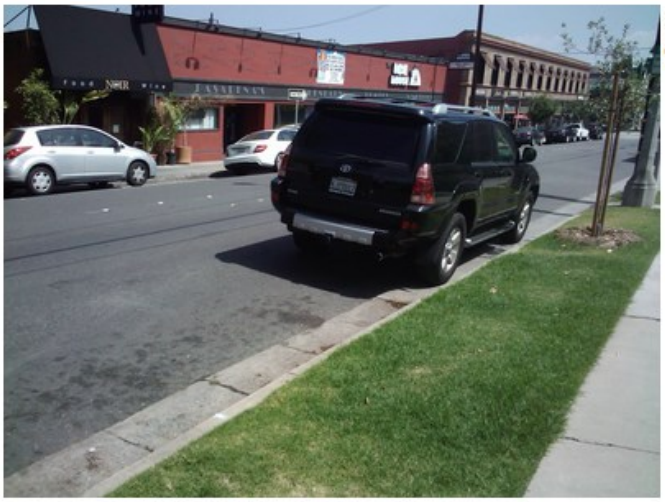
June 8, 2013 at 7:11 pm

HtcTrophy_SNHT16JM600679 HQ 1/10

Date Taken: April 27, 2013 at 6:19 am
Orientation: 1
Camera Model: mwp6985
ISO Speed: 100

May 31, 2013 at 7:37 pm

MotorolaCliq_SNJ116NJ376L HQ 2/10



Date Taken: May 24, 2013 at 8:08 am
Camera Model: MB200

June 1, 2013 at 12:07 pm

mytouch3g_SNHT05SPB00016 HQ 1/10

Date Taken: October 4, 2012 at 12:12 am
Latitude: 34.1430638889
Longitude: -118.126144444
Camera Model: T-Mobile myTouch 3G
Focal Length: 372/100

February 14, 2013 at 12:34 am

Iphone3gs_SN87127NTJEDG HQ 2/10

Date Taken: November 3, 2012 at 8:31 am
Orientation: 1
Camera Model: iPhone 3GS
Exposure: 1/1805
F-Stop: 14/5
ISO Speed: 64
Focal Length: 77/20

February 12, 2013 at 5:29 am

Iphone4_SN86025XUGA4S HQ Part 1/10

Date Taken: July 23, 2012 at 7:17 am
Latitude: 39.7323333333
Longitude: -104.961
Orientation: 1
Camera Model: iPhone 4
Exposure: 1/256
F-Stop: 14/5
ISO Speed: 80
Focal Length: 77/20

February 10, 2013 at 6:02 pm

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 again 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.

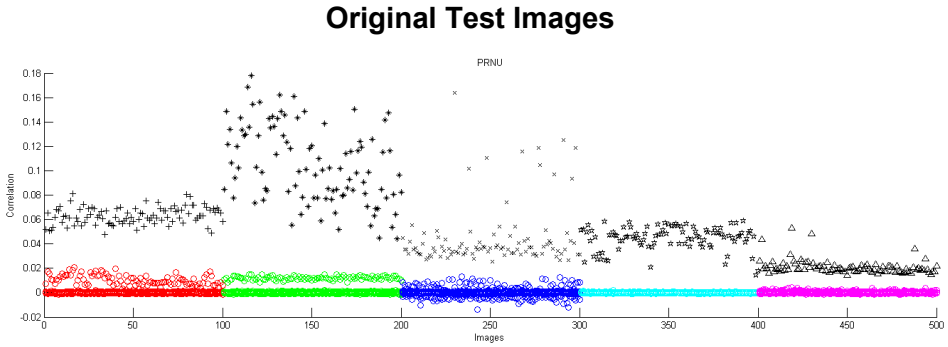


Figure 32 PRNU Samples (Left to Right): BB Pearl Flip, BB Curve, BB Tour, HTC EVO 4g, and HTC Trophy

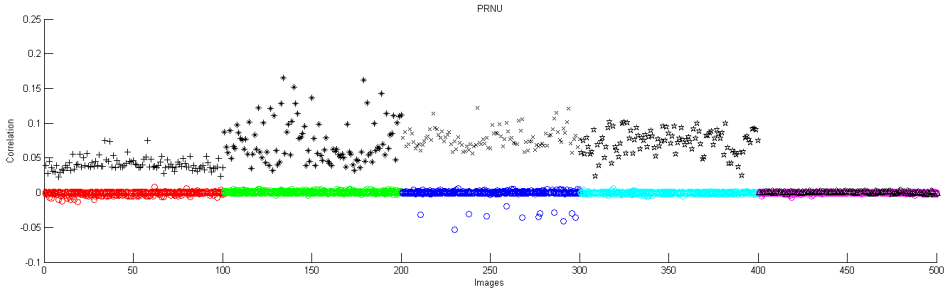


Figure 33 PRNU Samples (Left to Right): Moto Cliq, Mytouch 3g, Iphone 3gs, Iphone 4, Iphone 4 Alt

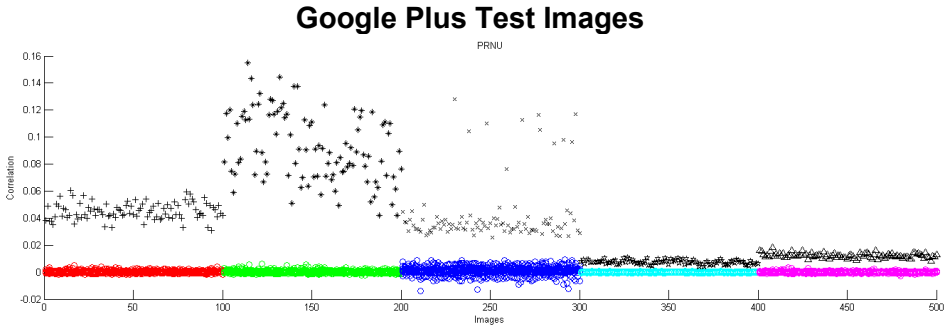


Figure 34 PRNU Samples (Left to Right): BB Pearl Flip, BB Curve, BB Tour, HTC EVO 4g, and HTC Trophy

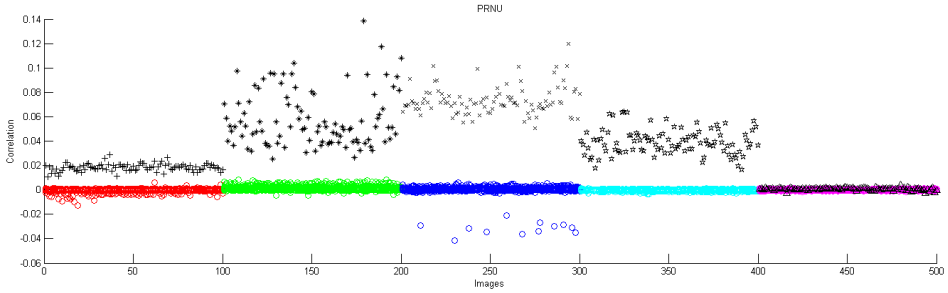


Figure 35 PRNU Samples (Left to Right): Moto Cliq, Mytouch 3g, Iphone 3gs, Iphone 4, Iphone 4 Alt

Facebook Test Images

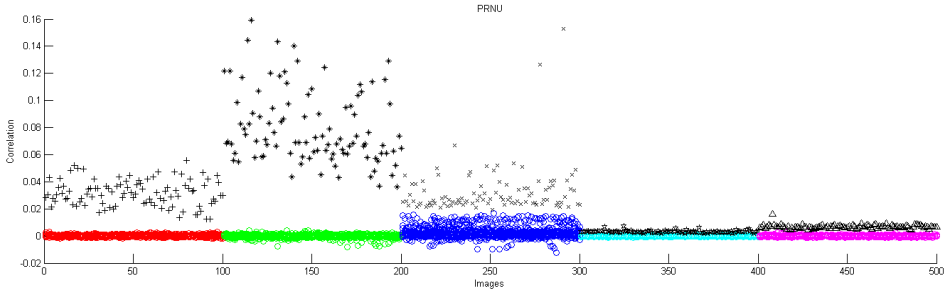


Figure 36 PRNU Samples (Left to Right): BB Pearl Flip, BB Curve, BB Tour, HTC EVO 4g, and HTC Trophy

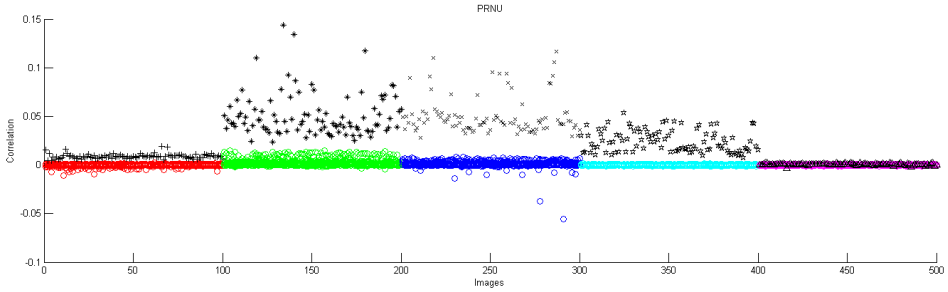


Figure 37 PRNU Samples (Left to Right): Moto Cliq, Mytouch 3g, Iphone 3gs, Iphone 4, Iphone 4 Alt

Myspace Test Images

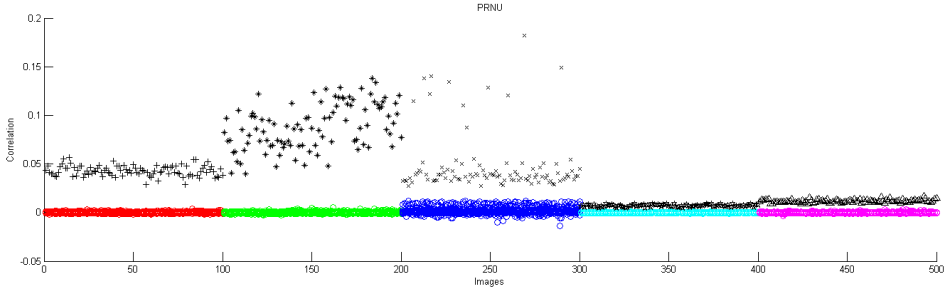


Figure 38 PRNU Samples (Left to Right): BB Pearl Flip, BB Curve, BB Tour, HTC EVO 4g, and HTC Trophy

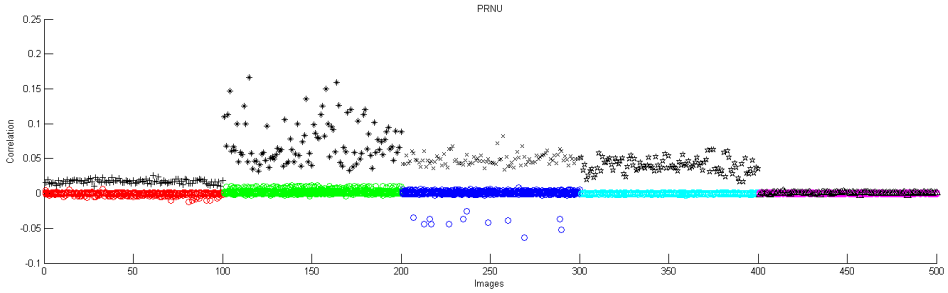


Figure 39 PRNU Samples (Left to Right): Moto Cliq, Mytouch 3g, Iphone 3gs, Iphone 4, Iphone 4 Alt