

available at www.sciencedirect.comjournal homepage: www.elsevier.com/locate/diinDigital
Investigation

Using JPEG quantization tables to identify imagery processed by software

Jesse D. Kornblum

Defense Cyber Crime Institute, United States

ABSTRACT

Keywords:

JPEG
Quantization
Digital ballistics
Calvin
Image authentication

The quantization tables used for JPEG compression can also be used to help separate images that have been processed by software from those that have not. This loose classification is sufficient to greatly reduce the number of images an examiner must consider during an investigation. As illicit imagery prosecutions depend on the authenticity of the images involved, this capability is an advantage for forensic examiners. This paper explains how quantization tables work, how they can be used for image source identification, and the implications for computer forensics.

© 2008 Digital Forensic Research Workshop. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Illicit imagery cases became more difficult to prosecute in the United States in the wake of the *Ashcroft v. Free Speech Coalition* Supreme Court decision in 2002 ([Supreme Court of the United States, 2002](#)). In the court's opinion, the prosecution must prove that a real child was harmed by the illicit imagery in order to get a conviction. As such, forensic examiners have been increasingly asked to prove that suspected illicit imagery contains real victims and not computer generated people. On the other hand, the prosecution still only needs to introduce a handful of such images in order to secure a conviction.

One of the more compelling arguments for proving the authenticity of a pictured individual is to show that the image came from a camera and has not been edited. Although a camera could conceivably be used to capture an image of an artificial person (e.g. photographing a computer screen), such an image would hopefully be obviously identifiable.

The science of performing digital ballistics, or matching an image to the individual device that created it, would be an ideal method for finding real pictures to use in legal proceedings. Unfortunately such identifications are not easy. Instead this paper demonstrates how an examiner can match an image back to the type of device that last modified it, either hardware or software.

This paper gives a brief overview of JPEG compression and pays particular attention to the quantization tables used in that process. Those tables control how much information is lost during the compression process. The author has categorized the types of tables and the implications for digital ballistics are discussed. In particular, by eliminating those images that were most likely last processed by a computer program, the examiner is left with fewer images to consider for the remainder of the investigation.

2. JPEG compression

This paper focuses exclusively on images stored in the JPEG Interchange File Format (JFIF) ([Wallace, 1992](#)), a method for storing data compressed with the JPEG standard ([Joint Photographic Experts Group, 1991](#); [Wallace, 1991](#)). The JFIF is the most commonly used format for JPEG data. Throughout the paper, any reference to a JPEG or JPEG file refers to JFIF encoded data.

A JPEG compressed image takes up considerably less space than an uncompressed image. Whereas an uncompressed 640 × 480 pixel 24-bit color image would require 900 kB, a JPEG version of the same image can be compressed to

E-mail address: jesse.kornblum@mantech.com

a mere 150 kB. Converting an image into a JPEG is a six step process that, as a whole, is beyond the scope of this paper (Joint Photographic Experts Group, 1991; Wallace, 1991). First the image is converted from the RGB color space into the YCbCr space, or one based on the brightness and luminance of each pixel. Next the image is downsampled, split into blocks of 8 × 8 pixels, and a discrete cosine transform is applied. The next stage is quantization, where the lossy compression occurs and this paper is focused. Finally an entropy coding (lossless compression) is applied and the image is said to be JPEG compressed.

In the quantization stage, the image creation device must use a table of values known as the quantization tables. Each table has 64 values that range from 0 to 65,535.¹ A lower number means that less data will be discarded in the compression and a higher quality image should result.

Each image has between one and four quantization tables. The most commonly used quantization tables are those published by the Independent JPEG Group (IJG) in 1998 and shown in Fig. 1 (libjpeg, 1998). These tables can be scaled to a quality factor Q. The quality factor allows the image creation device to choose between larger, higher quality images and smaller, lower quality images.

The value of Q can range between 0 and 100 and is used to compute the scaling factor, S, as shown in Eq. (1). Each element i in the scaled table T_s is computed using the ith element in the base table T_b as shown in Eq. (2). All of these computations are done in integer math; there are no decimals (hence the floor function in the equation). Any value of T_s that computes to zero is set to one.

For example, we can scale the IJG standard table using Q = 80 by applying Eq. (2) to each element in the table. The resulting values are the scaled quantization tables and are shown in Fig. 2. Note that the numbers in this table are lower than in the standard table, indicating an image compressed with these tables will be of higher quality than ones compressed with the standard table. It should be noted that scaling with Q = 50 does not change the table.

$$S = (Q < 50) ? \frac{5000}{Q} : 200 - 2Q \tag{1}$$

$$T_s[i] = \left\lfloor \frac{S * T_b[i] + 50}{100} \right\rfloor \tag{2}$$

3. Related work

Using quantization tables to identify the origin of digital images is only one method of conducting digital ballistics. A good summary of other methods can be found in Sencar and Memon (2008). The idea of using JPEG quantization tables for digital ballistics was first proposed by Farid (2006). In that report he showed that the quantization tables from 204 images, one per camera at the device’s highest quality setting, were for

¹ In practice these values usually between 0 and 255. Some programmers chose to represent them as 8 bit values, but to be correct 16 bit values should be used.

| | | | | | | | |
|----|----|----|----|-----|-----|-----|-----|
| 16 | 11 | 10 | 16 | 24 | 40 | 51 | 61 |
| 12 | 12 | 14 | 19 | 26 | 58 | 60 | 55 |
| 14 | 13 | 16 | 24 | 40 | 57 | 69 | 56 |
| 14 | 17 | 22 | 29 | 51 | 87 | 80 | 62 |
| 18 | 22 | 37 | 56 | 68 | 109 | 103 | 77 |
| 24 | 35 | 55 | 64 | 81 | 104 | 113 | 92 |
| 49 | 64 | 78 | 87 | 103 | 121 | 120 | 101 |
| 72 | 92 | 95 | 98 | 112 | 100 | 103 | 99 |
| 17 | 18 | 24 | 47 | 99 | 99 | 99 | 99 |
| 18 | 21 | 26 | 66 | 99 | 99 | 99 | 99 |
| 24 | 26 | 56 | 99 | 99 | 99 | 99 | 99 |
| 47 | 66 | 99 | 99 | 99 | 99 | 99 | 99 |
| 99 | 99 | 99 | 99 | 99 | 99 | 99 | 99 |
| 99 | 99 | 99 | 99 | 99 | 99 | 99 | 99 |
| 99 | 99 | 99 | 99 | 99 | 99 | 99 | 99 |
| 99 | 99 | 99 | 99 | 99 | 99 | 99 | 99 |

Fig. 1 – Standard JPEG quantization tables.

the most part different from each other. When overlaps did occur, they were generally among cameras from the same manufacturer. He also demonstrated that the tables used by the digital cameras were different from those used by Adobe Photoshop.

Chandra and Ellis (1999) did some work to determine the base quantization table used to compute the scaled tables found in an existing JPEG image, but they focused on determining an equivalence to the IJG tables, not divining the true tables used. Several papers have been written on recovering the quantization table from previously compressed images (Fan and de Queiroz, 2000, 2003; Neelamani et al., 2006) or doubly compressed images (Lukas and Fridrich, 2003). Many papers and patents have been written regarding quantization table design such as Beretta et al. (1999), Costa and Veiga (2005), Onnasch and Ploger (1994), Wang et al. (2001), and Watson (1993).

4. Classifying JPEGs

For this paper the author examined several thousand images from a wide variety of image creation devices and programs. These devices include a Motorola KRZR K1m, a Canon PowerShot 540, a FujiFilm Finepix A200, a Konica Minolta Dimage Xg,

| | | | | | | | |
|----|----|----|----|----|----|----|----|
| 6 | 4 | 4 | 6 | 10 | 16 | 20 | 24 |
| 5 | 5 | 6 | 8 | 10 | 23 | 24 | 22 |
| 6 | 5 | 6 | 10 | 16 | 23 | 28 | 22 |
| 6 | 7 | 9 | 12 | 20 | 35 | 32 | 25 |
| 7 | 9 | 15 | 22 | 27 | 44 | 41 | 31 |
| 10 | 14 | 22 | 26 | 32 | 42 | 45 | 37 |
| 20 | 26 | 31 | 35 | 41 | 48 | 48 | 40 |
| 29 | 37 | 38 | 39 | 45 | 40 | 41 | 40 |
| 7 | 7 | 10 | 19 | 40 | 40 | 40 | 40 |
| 7 | 8 | 10 | 26 | 40 | 40 | 40 | 40 |
| 10 | 10 | 22 | 40 | 40 | 40 | 40 | 40 |
| 19 | 26 | 40 | 40 | 40 | 40 | 40 | 40 |
| 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 |
| 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 |
| 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 |
| 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 |

Fig. 2 – Standard JPEG quantization tables scaled with Q = 80.

Download English Version:

<https://daneshyari.com/en/article/456536>

Download Persian Version:

<https://daneshyari.com/article/456536>

[Daneshyari.com](https://daneshyari.com)