



Using JPEG Quantization Tables to Identify Imagery Processed by Software

By

Jesse Kornblum

From the proceedings of

The Digital Forensic Research Conference

DFRWS 2008 USA

Baltimore, MD (Aug 11th - 13th)

DFRWS is dedicated to the sharing of knowledge and ideas about digital forensics research. Ever since it organized the first open workshop devoted to digital forensics in 2001, DFRWS continues to bring academics and practitioners together in an informal environment.

As a non-profit, volunteer organization, DFRWS sponsors technical working groups, annual conferences and challenges to help drive the direction of research and development.

<http://dfrws.org>

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.

and a Nikon Coolpix 7900. The author also examined the images produced by a number of software programs such as libjpeg (libjpeg, 1998), Microsoft Paint, the Gimp (GIMP Team, 2007), Adobe Photoshop (Adobe Systems Incorporated, 2007), and Irfanview (Irfan, 2007). The author also studied images from the camera review web site Digital Photography Review (Askey.Net Consulting Limited, 2007).

The author immediately noted that although some devices always used the same quantization tables, the majority of them used a different set of quantization tables in each image. A further examination of the images allowed the author to classify images into categories: standard tables, extended standard tables, custom fixed tables, and custom adaptive tables.

4.1. Standard tables

Images in this category use scaled versions of the quantization tables published in International JPEG Group standard (libjpeg, 1998), shown in Fig. 1. Because many cameras and programs use these tables, there is no way to determine, based on tables alone, which program or camera created an image.

The base tables shown in Fig. 1 can be scaled using $Q = \{1, 2, \dots, 99\}$ to create 99 separate tables. Scaling with $Q = 0$ would produce grossly unusable images. Scaling with $Q = 100$ would produce a quantization table filled with all ones. Such a table would be indistinguishable from any other base table scaled with $Q = 100$.

Any image using one of the 99 tables defined above is said to be using the standard tables. It is certainly possible that another method could be used to generate one of these tables, but there is no way to distinguish that from the image alone. For example, the author found that some images from particular devices matched an IJG table but others did not. It cannot be determined if the devices are using the IJG tables for some images and not others or the devices are using another method that occasionally produces the same tables as the IJG method.

4.2. Extended standard tables

These images are a special case of the standard tables. They use scaled versions of the IJG tables, but have three tables instead of the two in the standard. The third table is a duplicate of the second. The same methodology used to identify the standard tables can be used to identify extended standard tables.

4.3. Custom fixed tables

Some programs have their own non-IJG quantization tables that do not depend on the image being processed. For example, when Adobe Photoshop saves an image as a JPEG it allows the user to select one of 12 quality settings (different settings are used when saving images "for the web"). The quality setting is used to select one of 12 sets of quantization tables (Adobe Systems Incorporated, 2007).

Some devices use their own custom base quantization table with the IJG scaling method. Regardless, these images

consider a user selected quality factor and the base quantization table. Unlike the images described in the next section, the image itself is not part of the equation.

One of the more frustrating aspects of these devices is that there is no provable method for finding the original tables or scaling method used by an image creation device. The examiner could reverse engineer the device in question, but doing so is not often practical for an illicit imagery investigation. Given an image and its quantization tables, it is possible to compute a base table, T'_b for any assumed value of Q between 1 and 99 and the IJG scaling method. The equation for doing so is shown in Eq. (3). Note that the floor operation used in generating T_s in Eq. (2) means that the values obtained for T'_b may not be equal to the true values of T_b .

$$T'_b[i] = \frac{100 * T_s[i] - 50}{S} \quad (3)$$

The examiner could check each value of T'_b by attempting to use it to scale up to current value of T_s and the values of T_s for other images made by the same image creation device. If the computed value of T'_b is too small to generate the correct value of T_s , then T'_b must be manually increased until it is sufficient to create T_s when scaled. In the event that T'_b is too large for another image, the image was not created using the IJG scaling method and must be considered as a custom adaptive image described in Section 4.4.

The problem with all of the above calculations, however, is that the examiner does not know the true value of Q used for any of the images from the device. The method described in Chandra and Ellis (1999) can be used to estimate a value for Q' , or the quality factor used to scale the IJG standard table to the closest possible value of T_s . That estimate of Q provides a good starting point, but in the end we have a system with many possible solutions. The examiner cannot determine which solution is correct.

The author constructed a program that, given an initial image, accepted a value for Q' from the user. This value was used to compute the T'_b tables. These tables were then scaled to the values of Q from 1 to 99. These tables were then compared to other images generated by the same device. After each image, the tables were adjusted to fit the current image. If the tables no longer fit all of the images, the images were re-categorized as having custom adaptive tables as described in Section 4.4. In general the program was deemed impractical as it did not identify any plausible base tables.

4.4. Custom adaptive tables

These images do not conform to the IJG standard. In addition, they may change, either in part or as a whole, between images created by the same device using the same settings. They may also have constants in the tables; values that do not change regardless of the quality setting or image being processed.

For example, the author examined 21 pictures captured with a Fuji Finepix A200 camera. Of these pictures, eight images had identical quantization tables; four pictures shared one set of tables but had differences in the other two tables. The remaining 13 images all had unique quantization tables. What struck the author as odd, however, was that for all 21 images the first value in each of the three quantization tables

was four. The other values in the tables ranged from 1 to 24; a wide range. The author could not devise a set of base quantization tables that would include such a wide variation of values but keep one member of the tables constant. The author has hypothesized that this camera uses one constant value in the each table but scales the remainder of them.

It should be noted that the camera's manufacturer, the Fuji Xerox Company Limited, holds several patents regarding 'image creation apparatuses.' These include at least one that describes creating custom quantization tables based on the image being processed (Yokose, 2005). In that particular patent, a base quantization table is modified depending not only on the standard scaling method, but also on the resolution of the original uncompressed image.

5. Using quantization tables for ballistics

An examiner can encounter hundreds of thousands of images in the course of a single investigation. As noted above, it may be difficult to prosecute an offender using images that have been retouched by a computer. Given the large volume of images, it would benefit an examiner to only consider those images that could be used for prosecution, and thus only consider images that have not been altered.

JPEG quantization tables can be used for digital ballistics to identify and eliminate from consideration those images that most likely have been altered by a computer. That is, those images whose quantization tables are the most likely to have been generated by software can be eliminated from the investigation.

This method may have some false positives, or images that were not modified by a computer but are still eliminated from the investigation. But given the scale of such investigations and how few images are needed for a successful prosecution, a few false positives are acceptable.

There will be some special cases, however, where the quantization tables indicate the image was last modified by software but it has other indicators that it originally came from a camera. For example, the image could contain a complete set of EXIF data from a known camera or color signatures of real skin. In this case the examiner could use the quantization tables as part of a larger system to evaluate images.

Ideally, the examiner could use the JPEG quantization tables to determine exactly what kind of device created each image and categorize the images accordingly. The program JPEGsnoop aims to do exactly this (Hass, 2008). The program comes with a database of tables that be compared against

input files. Unfortunately, however, JPEGsnoop assumes that each camera can use only one quantization table. The use of custom adaptive tables, however, means that programs like JPEGsnoop would need to hold an unwieldy number of tables to be practical. Worse, some tables may be used by several devices, including both cameras and software programs, rendering the database inaccurate when attempting to determining an image's origin.

6. Calvin

The author has developed a software library called Calvin to help programmers use quantization tables for digital ballistics. The goal of Calvin is to identify those images who cannot be guaranteed to have been created by a real camera. For our purposes this means any image that could have been last processed by software. The program was named in honor of Dr. Calvin Goddard, the inventor of forensic ballistics (Federal Bureau of Investigation, 2003).

The Calvin library is able to display the quantization tables from existing images and determine if a new table is in a set of known tables. By default the library contains the standard tables, extended standard tables, and the tables used by Adobe Photoshop. Additional tables can be loaded from a configuration file. The library can be used to generate these configuration files from existing images.

6.1. Display mode

The user may wish to add more quantization tables to the set used by Calvin. The program can extract and display the quantization table for any image. For ease of use, the output is presented in the library's configuration file format. The standard tables scaled with $Q=80$ (shown earlier in Fig. 2) are shown as a configuration file entry in Fig. 3.

6.2. Comparison mode

An examiner can also use the Calvin library to compare the quantization table from an unknown image to the set of known tables. The user presents the library with an unknown file and is told whether or not the quantization table is contained in the known set. Presuming that the set of known signatures contains only the tables used by software programs, a negative response means that the image in question was possibly created by a hardware device. It could have been created by a program that uses tables not in the set of knowns.

```
/home/user/img/known/zoey-krzr-0127.jpg
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, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
```

Fig. 3 – Sample Calvin configuration file entry.

Conversely, a positive response from the library means that the image was *most likely* last modified by a program. It could have been created by a hardware device that uses the same tables as a known software program.

It is also possible that a real picture that has been processed by a software package, for example, cropped, would get a positive response from Calvin. Such a program would most likely be recompressed with the quantization tables used by the program, not the original quantization tables that the hardware device wrote into the image.

7. Conclusion

The author has demonstrated how JPEG quantization tables can be used for digital ballistics to eliminate images that could not be used in a prosecution for illicit imagery. The methodology is not perfect, but given the large number of available images and the small number needed in court, it should be sufficient. A more elegant solution, however, would be to combine this kind of digital ballistic information with other metadata from an image. Other factors, such as the presence or absence of EXIF data, signatures of known programs, and color signatures of real skin, could reduce the examiner's workload even more. In the meantime, however, using JPEG quantization tables for digital ballistics is a big step forward for examiners and should improve their productivity and success.

Acknowledgments

The author would like to thank the people who provided both pictures and support during this research: Rik Farrow, Joe Lewthwaite, Brian Martin, Jennifer Reichwein, and Peiter "Mudge" Zatkó. Invaluable technical support was provided by Robert J. Hansen. Extra special thanks to S-.

REFERENCES

- Adobe Systems Incorporated. Adobe photoshop. CS3 ed.; 2007.
- Askey.Net Consulting Limited. Digital photography review, <<http://www.dpreview.com/>>; June 2007.
- Beretta Giordano, Bhaskaran Vasudev, Konstantinides Konstantinos, Natarajan Balas K. US Patent 5,883,979: method for selecting JPEG quantization tables for low bandwidth applications; 1999.
- Chandra Surendar, Ellis Carla Schlatter. JPEG compression metric as a quality aware image transcoding. In: Proceedings of the 2nd USENIX symposium on internet technologies & systems; 1999. p. 81-92.
- Costa LF, Veiga ACP. A design of JPEG quantization table using genetic algorithms. In: Proceedings of the ACIT signal and image processing; 2005.
- Fan Zhigang, de Queiroz Ricardo. Maximum likelihood estimation of JPEG quantization table in the identification of bitmap compression history. IEEE Transactions on Image Processing 2000;1:948-51.
- Fan Zhigang, de Queiroz Ricardo. Identification of bitmap compression history: JPEG detection and quantizer estimation. IEEE Transactions on Image Processing 2003;12(2).
- Farid Hany. Digital image ballistics from JPEG quantization. Technical Report TR2006-583, Department of Computer Science, Dartmouth College; 2006.
- Federal Bureau of Investigation. The birth of the FBI's technical laboratory, <<http://www.fbi.gov/hq/lab/labdedication/labstory.htm>>; 2003.
- The GIMP Team. GNU image manipulation program. 2.2.15 ed., <<http://gimp.org/>>; 2007.
- Hass Calvin. JPEGsnoop. 1.2.0 ed., <<http://www.impulseadventure.com/photo/jpeg-snoop.html>>; 2008.
- Irfan Skiljan. IrfanView. 4.0 ed.; 2007.
- Joint Photographic Experts Group. Information technology - digital compression and coding of continuous-tone still images: requirements and guidelines. ISO/IEC 10918-1:1994; 1991.
- JPEG Group. libjpeg. 6b ed., <<http://www.ijg.org/>>; 1998.
- Lukas Jáň, Fridrich Jessica. Estimation of primary quantization matrix in double compressed JPEG images. In: Proceedings of the 2003 digital forensic research workshop, SUNY, Binghamton; 2003.
- Microsoft Corporation. Microsoft Paint overview.
- Neelamani Ramesh (Neelsh), de Queiroz Ricardo, Fan Zhigang, Dash Sanjeeb, Richard Baraniuk G. JPEG compression history estimation for color images. IEEE Transactions on Image Processing June 2006;15(6).
- Onnasch Prause, Ploger. Quantization table design for JPEG compression of angiocardiographic images. Computers in Cardiology 1994.
- Sencar Husrev T, Memon Nasir. Overview of state-of-the-art in digital image forensics. Part of Indian Statistical Institute Platinum Jubilee Monograph series 'Statistical Science and Interdisciplinary Research'; 2008.
- Supreme Court of the United States. Ashcroft v. Free Speech Coalition; 2002. case 00-795.
- Wallace Gregory K. The JPEG still picture compression standard. IEEE Transactions on Consumer Electronics 1991;38(1):18-34.
- Wallace Gregory K. JPEG file interchange format. C-Cube Microsystems September 1992.
- Wang Ching-Yang, Lee Shiuh-Ming, Chang Long-Wen. Designing JPEG quantization tables based on human visual system. Signal Processing: Image Communication 2001;16(5):501-6.
- Watson Andrew B. DCT quantization matrices visually optimized for individual images. In: Proceedings of the society for optical engineering; 1993. p. 202-16.
- Yokose Taro. US Patent 6,968,090: image coding apparatus and method; 2005.

Jesse D. Kornblum is a Research and Development Engineer for the Defense Cyber Crime Institute. A contractor with the ManTech International Corporation, his research focuses on computer forensics and computer security. He has authored and maintains a number of computer forensics tools including foremost, md5deep and ssdeep. When choosing sodas, Mr. Kornblum prefers cane sugar to high fructose corn syrup.