JPEG Quantization Tables Forensics: A Statistical Approach

Babak Mahdian[1], Stanislav Saic[1], and Radim Nedbal[2]

[1] Institute of Information Theory and Automation of the ASCR, Czech Republic
{mahdian,ssiac}@utia.cas.cz
[2] Institute of Computer Science of the ASCR, Czech Republic
radned@seznam.cz

Abstract. Many digital image forensics techniques using various fingerprints which identify the image source are dependent on data on digital images from an unknown environment. As often software modifications leave no appropriate traces in image metadata, critical miscalculations of fingerprints arise. This is the problem addressed in this paper. Modeling information noise, we introduce a statistical approach for noise-removal in databases consisted of "unguaranteed" images. In this paper, employed fingerprints are based on JPEG quantization tables.

Keywords: Image forensics, image forgery detection, hypergeometric distribution, jpeg quantization tables.

1 Introduction

Verifying the integrity of digital images and detecting the traces of tampering without using any protecting pre–extracted or pre–embedded information have become an important and hot research field of image processing [1, 2].

One of the typical ways of determining the image integrity is by matching the image being analyzed with its acquisition device via device fingerprints. Having no access to a higher number of acquisition devices we have to rely on popular photo sharing sites to get a sufficiently large training set for extracting fingerprints. When using images from such sites, we face a real problem: uncertainty about the image's history. As these images could be processed and re–saved by an editing and image uploading software (for instance for contrast enhancing or down-sizing) and by taking into account that many pieces of software do not leave typical traces in metadata, we may face miscalculations of fingerprints. This is the problem addressed in this paper. Modeling information noise in such databases, we introduce a statistical approach for removal of this noise.

JPEG photographs in photo-sharing sites contain various important metadata. Among others, they are taken by (camera) *users* with *cameras*, which encode them by means of *quantization tables* (QTs). As different image acquisition devices and software editors typically use different JPEG QTs [3, 4], we employ here QTs as device fingerprints.

H. Sako, K. Franke, and S. Saitoh (Eds.): IWCF 2010, LNCS 6540, pp. 150–159, 2011.
© Springer-Verlag Berlin Heidelberg 2011

It should be pointed out that a typical acquisition device uses some family of QTs very often whereas some others very rarely, and, most importantly, there is a (unknown) set of QTs that it uses never. And this is exactly the set that is of a great interest as for the image integrity. Indeed, it is easily seen that it can be exploited to negate it. Also note that a simple threshold based approach to determine the above set is unlikely to be feasible as the QT distribution is far from uniform for most acquisition devices. Therefore we propose a more sophisticated approach that is believed to deal with the non-uniformity of QT distribution effectively. Specifically, it allows to condition a threshold value for determining the above set on the QT distribution.

2 Related Work

There are a number of papers dealing with detection of artifacts brought into the JPEG image by the quantization procedure and corresponding QTs. The artifacts were used to detect the doubly/multiply compressed JPEG images. For example, see [5–12].

For instance, Jan Lukáš and Jessica Fridrich [5] presented a method for estimation of primary quantization matrix from a double compressed JPEG image. The paper presents three different approaches from which the Neural Network classifier based one is the most effective. Tomáš Pevný and Jessica Fridrich [6] proposed a method based on support vector machine classifiers with feature vectors formed by histograms of low–frequency DCT coefficients. Dongdong Fu et al. [7] proposed a statistical model based on Benford's law for the probability distributions of the first digits of the block–DCT and quantized JPEG coefficients. Weiqi Luo et al. [8] proposed a method for detecting recompressed image blocks based on JPEG blocking artifact characteristics. Babak Mahdian and Stanislav Saic [9] proposed a method for detection double compressed JPEG images based on histograms properties of DCT coefficients and support vector machines. Alin C. Popescu and Hany Farid [10] proposed a double JPEG Compression technique by examining the histograms of the DCT coefficients. In [11], Zhenhua Qu et al. formulated the shifted double JPEG compression as a noisy convolutive mixing model to identify whether a given JPEG image has been compressed twice with inconsistent block segmentation.

The mentioned methods are directly dependent on quantization tables and mostly they need them to be different in the image acquisition device and the software. Unfortunately, when leaving the lab conditions and applying these methods in real–life conditions, typically they produce considerably higher false positive rates than which are reported in papers [1]. Furthermore they efficiency is high only under limited conditions. These drawbacks mostly are caused by high variety of real–life image textures and properties (size and frequency of uniform regions, etc.).

To our best knowledge, there are two papers directly analyzing the potential of QTs to separate images that have been processed by software from those that have not.

Hany Farid [4, 13] using a database of one million images analyzed the potential of JPEG QTs to become a tool for source identification. He found out that while the JPEG QTs are not unique, they are effective at narrowing the source of an image to a single camera make and model or to a small set of possible cameras. His approach was based on equivalence classes.

Jesse D. Kornblum [3] examined several thousand images from various types of image acquisition devices and software. This allowed him to categorize the various types of QTs and discuss the implications for image forensics.

By comparison, being based on sophisticated statistical considerations, our approach aims at minimizing sensitivity to the data noise. As a result, in this sense, we provide a more robust tool for source identification.

3 Basics of JPEG Compression

Typically, the image is first converted from RGB to YCbCr, consisting of one luminance component (Y), and two chrominance components (Cb and Cr). Mostly, the resolution of the chroma components are reduced, usually by a factor of two [14]. Then, each component is split into adjacent blocks of 8 × 8 pixels. Blocks values are shifted from unsigned to signed integers. Each block of each of the Y, Cb, and Cr components undergoes a discrete cosine transform (DCT). Let $f(x, y)$ denotes a 8 × 8 block. Its DCT is:

$$\begin{aligned} F(u,v) = {} & \frac{1}{4}C(u)C(v) \\ & \sum_{x=0}^{7}\sum_{y=0}^{7} f(x,y) cos\frac{(2x+1)u\pi}{16} cos\frac{(2y+1)v\pi}{16}, \end{aligned} \tag{1}$$

where

$$\begin{aligned} & (u, v \in \{0 \cdots 7\}); \\ & C(u), C(v) = 1/\sqrt{2} \quad \text{for} \quad u, v = 0; \\ & C(u), C(v) = 1 \qquad \text{otherwise.} \end{aligned} \tag{2}$$

In the next step, all 64 $F(u, v)$ coefficients are quantized. Then, the resulting data for all blocks is entropy compressed typically using a variant of Huffman encoding (there also can be other forms of entropy coding like arithmetic, etc.). The quantization step is performed in conjunction with a 64–element quantization matrix, $Q(u, v)$. Quantization is a many–to–one mapping. Thus it is a lossy operation. Quantization is defined as division of each DCT coefficient by its corresponding quantizer step size defined in the quantization matrix, followed by rounding to the nearest integer:

$$F^Q(u,v) = round(\frac{F(u,v)}{Q(u,v)}), \quad u, v \in \{0 \cdots 7\} \tag{3}$$

Generally, the JPEG quantization matrix is designed by taking the visual response to luminance variations into account, as a small variation in intensity is more visible in low spatial frequency regions than in high spatial frequency regions. Typically, JPEG images contain one quantization table for the luminance band and one common quantization table for chrominance bands [14]. These tables are saved in the jpeg files header.

4 Basic Notations and Preliminaries

Suppose that any camera is determined uniquely by a pair of attributes make mk and model md[1]. Given a ternary relation, denoted S, a subset of the ternary Cartesian product $Cm \times Qt \times U$ of the set Cm of all the cameras, the set Qt of all QTs, and the set U of (all the potential) camera end users, we interpret a triplet $\langle cm, qt, u \rangle$ from S as the QT qt that has been OBSERVED to encode an image taken by the (camera) user u with a camera cm.

Assuming that S stores data from the Internet for instance, it provides noisy information. Indeed, some (unknown amount of) triplets from S result from a software manipulation of some photographs. Here, we assume (cf. Assumption 2) that most (or typical) software manipulations affect QTs. Accordingly, $\langle cm, qt, u \rangle \in S$ does not necessarily entail that qt is the QT that has been employed by a camera cm to encode an image taken by the (camera) user u. In fact, qt might be the QT employed by a software application to encode the image that was taken originally with a camera cm, which, however, had encoded the image originally by means of a QT different from qt. This is the noise inherent in S.

To represent noise-free information, we introduce a virtual, binary relation, denoted R, that is a subset of the binary Cartesian product $Cm \times Qt$. A pair $\langle cm, qt \rangle$ from R is interpreted as the QT qt that (in reality, which is unknown) is employed to encode some image taken by a camera cm. The other way round, $\langle cm, qt \rangle \notin R$ entails that cm never employs qt to encode an image.

5 Image Database

To create S, we needed to download and process a large number of images. Keeping at disposition a variety of popular photo–sharing servers from which photos can by downloaded, we have opted for Flickr, one of the most popular photo sharing sites. We downloaded two millions images labeled as "original." Nevertheless, as has been pointed out, Flickr, in fact, is an "uncontrolled arena." In other words, there is no guarantee that the image at hand is coming from the camera model information as indicated by metadata. To minimize the noise, as discussed in the previous section, we have discarded images with illegible metadata, a software tag signifying traces of some known photo processing software, or inconsistencies in original and modification dates or between quoted and actual width and height. Also, we discarded images without 3–channel colors. All

[1] In general, other sets of appropriate attributes can be considered, e.g., size, orientation, format, etc.

these operations, reduced the number of ”original” images to 798,924. Our strategy was to maximize $|Cm|, |Qt|, |U|$ and, at the same time, to minimize the noise in S.

We believe that the frequency of original images in this remaining set is much higher than noise (images processed by software).

6 A Statistical Approach for Noise Removal

In general, the question arises: Given observed information, represented as S, what can be concluded about reality, represented as R? Specifically, given S, can we objectively quantify a “confidence” that the QT as found in the image file may correspond to the one used by the camera upon capturing? Indeed, we present an approach based on statistical hypothesis testing that enables to make a lower estimation of this confidence.

In brief, we utilize a statistical analysis of information noise inherent in S. Noisy information generally is contained in any set of tuples from S. Specifically, given a “testing” pair $t_0 = \langle cm_0, qt_0 \rangle$, our default position is that all the triplets from S containing attribute values cm_0 and qt_0 represent noisy information only. Accordingly, we set out the null hypothesis

H_0: “t_0 is not included in R”

and introduce a *test statistic* $\mathcal{T}$, which, in general, is a numerical summary of S that reduces S to a set of values that can be used to perform the hypothesis test. Specifically, $\mathcal{T}$ quantifies the noisy information. Last, we determine the upper estimation p of observing a value v for $\mathcal{T}$ that is at least as extreme as the value that has been actually observed. That is, the probability of observing v or some more extreme value for T is no more than p.

The test statistic is defined as the mapping $T\colon Cm \times Qt \longrightarrow \mathbb{N}_0$ that maps each pair $\langle cm, qt \rangle$ from the binary Cartesian product $Cm \times Qt$ to the cardinality (a value from the set of nonnegative integers, denoted $\mathbb{N}_0$) of the set of all and only those users who, in accordance with S, have taken some image with a camera cm that has encoded it by means of qt. In symbols:

$$T(cm, qt) = \text{card}\{u \mid \langle cm, qt, u \rangle \in S\} \tag{4}$$

for any pair $\langle cm, qt \rangle$ from $Cm \times Qt$.

The rationale behind using the above test statistic is based on the ASSUMPTION OF PROPORTIONALITY:

Assumption 1. Given a pair $\langle cm, qt \rangle$ of a camera cm and a QT qt, the amount of noisy information in S concerning $\langle cm, qt \rangle$ is directly proportional to the number of (observed) distinct users (in S) who have taken an image encoded by means of qt with cm.

Speaking in broad terms, we conclude that the number of these users is too big to be attributed exclusively to an information noise if the number exceeds a specified significance level. To determine this significance level, we introduce mappings in terms of which we define the exact *sampling distribution* of T. It will be seen that, under the above and the undermentioned assumptions, this exact sampling distribution of T is the *hypergeometric distribution* that is relative to an appropriate set of cameras.

Observe that H_0 implies that any image that, in accordance with its metadata, has been taken with a camera cm_0 and encoded by means of qt_0 must in fact have been modified with a software application. Moreover, consider the following assumption of software manipulations.

Assumption 2. Software manipulations usually do not change image metadata concerning a camera.

Essentially, this assumption states that any image that, in accordance with its metadata, has been taken with a camera, say cm, indeed, has been taken with that camera. Consequently, $T(cm, qt)$ is interpreted as the number of all distinct users from S who have taken an image with a camera cm, which, IN ACCORDANCE WITH S, has encoded the image by means of qt. However, taking into account possible software manipulations, qt, in the reality that is OUT OF ACCORD WITH S, might have been employed by a software application used by a user to modify the image in question. Specifically, provided that H_0 is true, $T(cm_0, qt_0)$ is interpreted as the number of all distinct users from S who have taken an image with a camera cm_0, but, CONTRARY TO S, not cm_0 but a software application, used by a user to modify the image, has encoded the image by means of qt_0.

Next, C denoting a subset of Cm, we introduce the following mappings:

$$G\colon\ Cm \longrightarrow \mathbb{N}_0, \tag{5}$$

$$N\colon\ 2^{Cm} \longrightarrow \mathbb{N}_0, \tag{6}$$

$$n\colon\ Qt \times 2^{Cm} \longrightarrow \mathbb{N}_0. \tag{7}$$

defined by the following respective rules:

$$G(cm) = \operatorname{card}\{u \mid \langle cm, qt, u\rangle \in S\}, \tag{8}$$

$$N(C) = \sum_{cm\in C} G(cm), \tag{9}$$

$$n(qt, C) = \sum_{cm\in C} T(cm, qt). \tag{10}$$

$G(cm)$ is interpreted as the number of all (observed) distinct users (i.e., from S) who have taken an image with a camera cm. Accordingly, $N(C)$ is the summation of these numbers (of all distinct users from S) for all cameras from the set C. That is, each user is added in $N(C)$ k-times if he or she has taken images with k distinct respective cameras from C. Last, $n(qt, C)$ is the summation of the numbers of all (observed) distinct users (from S) who have taken an image with a respective camera from the set C, whereas the image is encoded by means of

qt: either, in accordance with S, the camera, or, out of accord with S, a software application, used by a user to modify the image, has employed qt to encoded the image. That is, each user is added in $n(qt, C)$ k-times if he or she has taken images with k distinct respective cameras from C, whereas the image is encoded by means of qt.

Specifically, suppose a set C including only cameras that never employ qt_0 to encode an image. Then $n(qt_0, C)$ is interpreted as the summation of the numbers of all (observed) distinct users (from S) who have taken an image with a camera from the set C, whereas the image is encoded by means of qt_0 by a software application, used by a user to modify the image. Moreover, suppose that the camera cm_0 is included in C. Indeed, in accordance with H_0, cm_0 is a camera that never employs qt_0 to encode an image.

Note that, for large S, $G(cm)$ is proportional to the number of all images taken with a given camera cm. In particular, CONSIDERING ONLY IMAGES TAKEN WITH CAMERAS FROM C, the $G(cm_0)$ to $N(C)$ ratio, $\frac{G(cm_0)}{N(C)}$, is interpreted as the probability that an image has been taken with a camera cm_0 (by a user, say u_1). Similarly, $G(cm_0) - 1$ to $N(C)$ ratio, $\frac{G(cm_0)-1}{N(C)}$, could by interpreted as the probability that an image has been taken with a camera cm_0 by a user, say u_2, different from the user u_1. However, observe that this interpretation is correct only if the following assumption is adopted.

Assumption 3. Given any camera cm from C and a set U_{cm} of users who have taken an image with a camera cm, the probability p_u that an image has been taken by a user u is (approximately) equal to $\frac{1}{G(cm)}$ for any user u from U_{cm}.

Then, disregarding all images that have been taken by considered users u_1, u_2 with respective cameras (identified as cm_0), $1 - \frac{G(cm_0)-2}{N(C)}$, is interpreted as a probability that an image has been taken with a camera from C (but distinct from cm_0) by a user, say u'_1. To put it another way, knowing that a given image has not been taken with a camera $cm_0$by a user u_1 or u_2, the probability the image has been taken (by any user) with a camera from C (but distinct from cm_0) is equal to $1 - \frac{G(cm_0)-2}{N(C)}$. In general, continuing the above train of thoughts, $\frac{G(cm_0)-k}{N(C)-\ell}$ is interpreted as a probability that, disregarding images that have been taken by any of $k+\ell$ considered users with respective cameras from C, an image has been taken with a camera cm_0. $1 - \frac{G(cm_0)-k}{N(C)-\ell}$ is interpreted analogously. Now the following proposition is clear upon reflection.

Proposition 1 (Sampling distribution of test statistic). Consider a mapping

$$F\colon \mathbb{N}_0 \times Cm \times Qt \times 2^{Cm} \longrightarrow \langle 0, 1\rangle \tag{11}$$

that coincides with the hypergeometric (cumulative) distribution function, whose *probability mass function* is defined as follows:

$$h(x; n, G, N) = \frac{\binom{G}{x}\binom{N-G}{n-x}}{\binom{N}{n}} \ , \tag{12}$$

where, by abuse of notation,

$$n = n(qt, C) \ , \qquad G = G(cm) \ , \qquad N = N(C) \ .$$

Then $F(x, cm_0, qt_0, C)$ is the sampling (discrete cumulative) distribution of $T(cm_0, qt_0)$ under H_0 if C includes cm_0 and only those cameras that never employ qt_0 to encode an image:

$$C \subseteq \{cm \mid \langle cm, qt_0 \rangle \notin R\} \ . \tag{13}$$

Most importantly, note that

$$p = 1 - F\big(T(cm_0, qt_0), cm_0, qt_0, C\big) \tag{14}$$

is the *p-value* that is interpreted as the probability of obtaining a test statistic at least as extreme as $\mathrm{T}(mk_0,qt_0)$, which is uniquely determined by S (i.e., the observed data), assuming that the null hypothesis H_0 is true. It presents the probability of incorrectly rejecting H_0.

Finally, we discuss an important subtlety of the condition (13) imposed on C in the above proposition. In fact, this condition is hard to fulfill as R is unknown. However, the following corollary is easily verified:

Corollary 1. Failing to fulfill (13) results in an upper estimation of the p-value.

To see the assertion of the corollary, observe that failing to fulfill (13) increases $n(qt_0, C)$ defined by (10) but, due to Assumption 2, affects neither $G(cm)$ for any cm from C and thus nor $N(C)$. It follows from properties of the hypergeometric distribution that its (cumulative) distribution function is inversely proportional to n for fixed but arbitrary x, G, and N. Consequently, for cm_0, qt_0 and fixed but arbitrary x, $F(x, cm_0, qt_0, C)$ has a global maximum at C if (13) holds. Now it is immediate that failing to fulfill (13) OVERVALUES p (defined as (14)) the probability estimation that the null hypothesis will be rejected incorrectly.

Note that rejecting H_0 entails accepting the alternative hypothesis, namely that t_0 is included in R, which means that qt_0 may be employed by a camera cm_0 to encode an image. Consequently, the confidence of accepting the alternative hypothesis can be quantified by the value $1 - p$ with rigorous interpretation, and the presented statistical approach results in a lower estimation of this value.

7 Experimental Results

We have carried out an experiment on 1000 randomly selected JPEG non-modified images taken by 10 cameras (100 images per camera) to demonstrate the efficiency of the proposed approach. For every image, we have repeated a statistical test procedure with the significance level (the probability of mistakenly rejecting the null hypothesis) set to 1%. All the cameras from S has been included in the set C, the parameter of the sampling distribution of test statistic T. Consequently, in accordance with the corollary, we have obtained a rather coarse upper estimation of the p-value, the probability of incorrectly rejecting H_0.

Results are shown in Tab. 1. The column denoted by "*orig*" refers to non-modified (i.e., original) images. Modified images have been simulated by re-saving original images so that randomly selected QTs (randomly for each image) used by popular software like Adobe Photoshop and GIMP (the column denoted by "*misc*") have been employed to encode them. Respective numbers of non-rejecting H_0 are shown, i.e., the numbers of images, whose originality cant be confirmed statistically from our data.

Table 1. Data in each cell are obtained using 100 JPEG images

cm	*size*	*orig*	*misc*
Canon EOS 20D	3504 × 2336	0	37
Canon EOS 50D	4752 × 3168	0	34
Canon PowerShot A75	2048 × 1536	0	28
Konica KD-400Z	2304 × 1704	0	16
Nikon Coolpix P80	1280 × 960	0	14
Nikon E990	2048 × 1536	5	10
Olympus C740UZ	2048 × 1536	2	7
Olympus X450	2048 × 1536	1	23
Panasonic DMC-LX2	3840 × 2160	0	19
Sony DSC-W40	2816 × 2112	2	16

8 Discussion

Denoising a DB (database) of QTs of JPEG images from "unguaranteed" sources is a complex task. Cameras and pieces of software often have complicate and unpredictable behavior. Many cameras compute QTs on the fly (based on the scene). Furthermore, a huge number of cameras and software employ standard IJG QTs. There also are devices using a particular set of QTs very widely and another set of QTs very rarely.

Many pieces of software modify images (for instance, enhance the contrast or rotate the image) without leaving any traces in image's JPEG file metadata. This makes difficult the task of obtaining acquisition devices's fingerprints from a DB consisting of images coming from an uncontrolled environment.

It is apparent that QTs cannot uniquely identify the source effectively. Despite this they provide valuable information, supplemental in the forgery detection task. This has been shown in the previous section.

We point out that our results are affected by the C parameter in the aforementioned fashion. In particular, a careful selection of cameras (to be included in C) based on an appropriate heuristics, is supposed to improve results remarkably. Specifically, it is expected to lower the probability of incorrectly rejecting H_0 concerning QTs of non-modified images, resulting in lower values in the column referred to as "*orig.*'

The approach presented is general and can straightforwardly be applies to other features forming devices fingerprints.

Acknowledgments. This work has been supported by the Czech Science Foundation under the projects No. GACR 102/08/0470 and GACR P202/10/P509. This work has been supported by the Institutional Research Plan AV0Z10300504 "Computer Science for the Information Society: Models, algorithms, Applications" and Institute for Theoretical Computer Science project no. 1M0021620808 (1M0545).

References

1. Mahdian, B., Saic, S.: A bibliography on blind methods for identifying image forgery. Signal Processing: Image Communication 25, 389–399 (2010)
2. Farid, H.: A survey of image forgery detection. IEEE Signal Processing Magazine 2(26), 16–25 (2009)
3. Kornblum, J.D.: Using jpeg quantization tables to identify imagery processed by software. In: Proceedings of the Digital Forensic Workshop, August 2008, pp. 21–25 (2008)
4. Farid, H.: Digital image ballistics from JPEG quantization, Tech. Rep. TR2006-583, Department of Computer Science, Dartmouth College (2006)
5. Fridrich, J., Lukas, J.: Estimation of primary quantization matrix in double compressed jpeg images. In: Proceedings of DFRWS, Cleveland, OH, USA, August 2003, vol. 2 (2003)
6. Fridrich, J., Pevny, T.: Detection of double–compression for applications in steganography. IEEE Transactions on Information Security and Forensics 3(2), 247–258 (2008)
7. Fu, D., Shi, Y.Q., Su, W.: A generalized benford's law for jpeg coefficients and its applications in image forensics. In: SPIE Electronic Imaging: Security, Steganography, and Watermarking of Multimedia Contents, San Jose, CA, USA (January 2007)
8. Luo, W., Qu, Z., Huang, J., Qiu, G.: A novel method for detecting cropped and recompressed image block. In: IEEE International Conference on Acoustics, Speech and Signal Processing, Honolulu, HI, USA, vol. 2, pp. 217–220 (April 2007)
9. Mahdian, B., Saic, S.: Detecting double compressed jpeg images. In: The 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009), London, UK (December 2009)
10. Popescu, A.C.: Statistical Tools for Digital Image Forensics, Ph.D. thesis, Ph.D. dissertation. Department of Computer Science, Dartmouth College, Hanover, NH (2005)
11. Qu, Z., Luo, W., Huang, J.: A convolutive mixing model for shifted double jpeg compression with application to passive image authentication. In: IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, USA, April 2008, pp. 4244–1483 (2008)
12. Kihara, M., Fujiyoshi, M., Wan, Q.T., Kiya, H.: Image tamper detection using mathematical morphology. In: ICIP, vol. (6), pp. 101–104 (2007)
13. Farid, H.: Digital image ballistics from JPEG quantization: A followup study, Tech. Rep. TR2008-638, Department of Computer Science, Dartmouth College (2008)
14. Pennebaker, W.B., Mitchell, J.L.: JPEG Still Image Data Compression Standard. Kluwer Academic Publishers, Norwell (1992)