JPEG QUANTIZATION TABLES FORENSICS: A STATISTICAL APPROACH

Babak Mahdian, Stanislav Saic

Institute of Information Theory and
Automation of the ASCR, Czech Republic
{mahdian,ssaic}@utia.cas.cz

Radim Nedbal

Institute of Computer Science
of the ASCR, Czech Republic
radned@seznam.cz

ABSTRACT

Many digital image forensics techniques extracting various fingerprints are dependent on data on digital images from an unknown environment. As often software modifications leave no appropriate traces in images metadata, critical inconveniences and miscalculations of fingerprints arise. This is the problem addressed in this paper. Modeling information noise in image metadata, we introduce a statistical approach to metadata analysis of images from "unguaranteed" sources. Resulting fingerprints are based on JPEG quantization tables.

Index Terms— Image forensics, image forgery detection, hypergeometric distribution, jpeg quantization tables.

1. INTRODUCTION

One of the typical ways of determining the image integrity is by matching the image being analyzed with its acquisition device via device's fingerprints. Very often the only possible source of the training set for extracting fingerprints are popular photo sharing sites. 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 software (for instance for contrast enhancing or rotating) and by taking into account that many softwares do not leave typical traces in metadata, we may face critical inconveniences and miscalculations of fingerprints. This is the problem addressed in this paper. Modeling information noise in image metadata, we introduce a statistical approach to metadata analysis of images from "unguaranteed" sources.

JPEG photographs 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, in this work, we use QTs as devices fingerprints.

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".

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 [1, 2, 3, 4, 5, 6, 7].

Hany Farid [8, 9] 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 [10] examined several thousand images from various types of image acquisition devices and softwares. This allowed her to categorize the various types of QTs and discuss the implications for image forensics.

3. BASIC NOTATIONS AND PRELIMINARIES

Suppose that any camera is identified as a pair of its make mk and model mk (arbitrary the identification key can be extended to other attributes as size, orientation, format, etc.). Consider a database (DB) that stores above metadata of photographs observed on the Internet. This "ob$\underline{s}$erved" DB is modeled as a ternary relation, denoted S, a subset of the ternary Cartesian product $Cm \times Qt \times U$ of the set Cm of all cameras, the set Qt of all QTs, and the set U of (all potential) camera users. That is to say, a triplet $\langle cm, qt, u\rangle$ is in S iff qt is the QT that has been observed to encode an image taken by the (camera) user u with a camera cm.

Note that S represents noisy information. Indeed, some (unknown amount of) tuples from S result from a software manipulation with some photographs. We assume (Assumption 1) that most software manipulations affect data concerning QTs. Because of this, $\langle cm, qt, u\rangle \in S$ 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 encoded the image that originally have been taken with a camera cm, which, however, had encoded the image 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 $\langle cm, qt \rangle$ is defined to be included in R iff qt is the 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 a camera cm never employs qt to encode an image.

4. IMAGE DATABASE

To create S, we needed to download and process a large number of images. There are a number of popular photo–sharing servers allowing a usable interface to access their photos. To collect a large number of images, Flickr, which is one of the most popular ones, had been chosen. Using the Flickr API, we downloaded two millions images denoted as "original". As aforementioned, the main problem with sources like Flickr is that their images are from an uncontrolled environment. It is not guaranteed that the image really comes from a camera. To minimize the noise rate, we eliminated those images having non–readable metadata or having inconsistency in original and modification dates, inconsistency between width and height in the metadata and actual image's width and height or those having a software tag signifying the traces of some known photo processing software. Furthermore, we eliminated images without 3–channel colors. All these operations, reduced the number of "original" images to 798,924. Our strategy in downloading images was to maximize $|Cm|, |Qt|, |U|$ in DB using non–modified images.

5. A STATISTICAL APPROACH

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 a given QT may be employed by a given camera to encode an image? Indeed, we present an approach based on statistical hypothesis testing that enable 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 tuples from S concerning 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, 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, our test statistic quantifies the noisy information. Last, we determine the upper estimation p of observing a value for a *test statistic* T that is at least as extreme as the value that has been actually observed.

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:

$$T(cm, qt) = \mathrm{card}\{u \mid \langle cm, qt, u \rangle \in S\} \tag{1}$$

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 of an amount of noisy information concerning a given camera and an amount of (observed) distinct users who have taken an image with that camera. Speaking in broad terms, we conclude that an amount of these users is too big to be attributed exclusively to an information noise if the amount 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 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 1. *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, in fact 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. Taking into account possible software manipulations, $T(cm, qt)$ is interpreted as the number of all distinct users from S who have taken an image with a camera cm, 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 encode the image. 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, whereas the image is encoded by means of qt_0, which, contrary to S, has not been employed by a camera cm_0 but by a software application, used by a user to modify the image.

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

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

$$N\colon 2^{Cm} \longrightarrow \mathbb{N}_0 \ , \tag{3}$$

$$n\colon Qt \times 2^{Cm} \longrightarrow \mathbb{N}_0 \tag{4}$$

defined by the following respective rules:

$$G(cm) = \operatorname{card}\{u \mid \langle cm, qt, u\rangle \in S\} \ , \tag{5}$$

$$N(C) = \sum_{cm \in C} G(cm) \ , \tag{6}$$

$$n(qt, C) = \sum_{cm \in C} T(cm, qt) \ . \tag{7}$$

$G(cm)$ 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 2. *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*

$$F\colon \mathbb{N}_0 \times Cm \times Qt \times 2^{Cm} \longrightarrow \langle 0, 1\rangle \tag{8}$$

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

$$h(x; n, G, N) = \frac{\binom{G}{x}\binom{N-G}{n-x}}{\binom{N}{n}} \ , \tag{9}$$

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 employe qt_0 to encode an image:

$$C \subseteq \{cm \mid \langle cm, qt_0\rangle \notin R\} \ . \tag{10}$$

Most importantly, note that

$$p = 1 - F\big(T(cm_0, qt_0), cm_0, qt_0, C\big) \tag{11}$$

is the *p-value* that is interpreted as the probability of obtaining a test statistic at least as extreme as 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 (10) 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* (10) *results in an upper estimation of the p-value.*

To see the assertion of the corollary, observe that failing to fulfill (10) increases $n(qt_0, C)$ defined by (7) but, due to

Assumption 1, 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 (10) holds. Now it is immediate that failing to fulfill (10) overvalues p (defined as (11)) the probability estimation that the null hypothesis will be rejected incorrectly.

6. 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 softwares 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.

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

7. DISCUSSION

It is apparent that despite QTs cannot uniquely identify the source effectively, 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*.'

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

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

8. REFERENCES

[1] J. Fridrich and J. Lukas, "Estimation of primary quantization matrix in double compressed jpeg images," in *Proceedings of DFRWS*, Cleveland, OH, USA, August 2003, vol. 2.

[2] J. Fridrich and T. Pevny, "Detection of double–compression for applications in steganography," *IEEE Transactions on Information Security and Forensics*, vol. 3, no. 2, pp. 247–258, June 2008.

[3] Dongdong Fu, Yun Q. Shi, and Wei Su, "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.

[4] Weiqi Luo, Zhenhua Qu, Jiwu Huang, and Guoping Qiu, "A novel method for detecting cropped and recompressed image block," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Honolulu, HI, USA, April 2007, vol. 2, pp. 217–220.

[5] Babak Mahdian and Stanislav Saic, "Detecting double compressed jpeg images," in *The 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP-09)*, London, UK, December 2009.

[6] A. C. Popescu, *Statistical Tools for Digital Image Forensics*, Ph.D. thesis, Ph.D. dissertation. Department of Computer Science, Dartmouth College, Hanover, NH, 2005.

[7] Zhenhua Qu, Weiqi Luo, and Jiwu Huang, "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.

[8] H. Farid, "Digital image ballistics from JPEG quantization," Tech. Rep. TR2006-583, Department of Computer Science, Dartmouth College, 2006.

[9] H. Farid, "Digital image ballistics from JPEG quantization: A followup study," Tech. Rep. TR2008-638, Department of Computer Science, Dartmouth College, 2008.

[10] Jesse D. Kornblum, "Using jpeg quantization tables to identify imagery processed by software," in *Proceedings of the Digital Forensic Workshop*, August 2008, pp. 21–25.