How we learned to stop worrying about content and love the metadata

A metadata-based manipulation de.tector
for the IFS-TC Image Forensics Challenge

Thomas Gloe, Matthias Kirchner, Christian Riess

The entry point to the forensic examination of a digital image is naturally the file it is stored in. The vast majority of practical cases deals with compressed images stored in the JPEG format. JPEG files store a number of mandatory parameters with the file, most importantly quantization tables for decompression. JPEG file information is generally a valuable source of forensic evidence. Different camera models use different compression settings, which usually do not match the settings of image processing software [1]. Re-saving a file in the JPEG format may thus leave conspicuous double-compression artifacts behind. Image processing software also changes internal JPEG file structures [2]. To rule out that such traces can be exploited in the course of the contest, all images were converted to the PNG format (most of them after a considerable amount of downsizing).

Virtually all digital cameras also attach a rich collection of metadata to the file to document details about the image acquisition process. The preferred method to organize and store such metadata is specified in the EXIF standard. EXIF data may contain information about the acquisition device, the acquisition time, camera and compression settings, amongst many others. Such data is often said to be untrustworthy for forensic purposes, as it can be modified easily or stripped off completely [2, 5]. Nevertheless, it undecidedly *does* provide information about an image's history, if available [4]. In fact, we have argued recently that metadata and file structure information are not *per se* as unreliable as commonly assumed [2], and that creating plausible forgeries thereof is highly non-trivial (as long as standard processing and editing software is concerned). Specifically, the presence and order of certain entries typically indicate whether an image was re-saved with image processing software.

Metadata is not exclusive to the JPEG file format. Also the PNG format supports storage of arbitrary metadata (so do many others). The PNG specification suggests so-called PNG text chunks for this purpose.[1] A conversion between image formats does not necessarily remove or modify all metadata. It rather depends on the conversion tool, how individual data segments are treated. Indeed, all image files of the contest did contain metadata. We used a hex editor and `exiftool`[2] to access metadata structures stored along with the image files. The idea was to see whether any auxiliary information was available that could be useful to distinguish between original and manipulated images. We found that most standard EXIF entries of the original JPEG images were not present anymore. In particular information about the source camera and basic camera settings (focal length, etc.) was missing. Also none of the well-known markers of image processing software could be found. Yet we observed a number of differences across different images. All original images in the training set contained PNG text fields '`Datecreate`' and '`Datemodify`', whereas only some of the manipulations did. In addition the vast majority of original images had the '`Color Type`' field set to '`RGB`' (a few exceptions referred to a '`Palette`' type). Many of the manipulated images, on the contrary, had this field set to '`RGB with Alpha`', possibly indicating the use of alpha-masks for image splicing.

The most telltale differences between original and manipulated training images were (missing) references to prior compression settings. While *all* original images contained information about earlier JPEG sampling factors, 390 out of the 450 manipulations did not have a '`jpeg:sampling-factor`' PNG text chunk. The remaining 60 manipulations had the field set to '`1x1,1x1,1x1`' (the three number

[1] `http://www.libpng.org/pub/png/spec/1.2/PNG-Chunks.html`
[2] `http://www.sno.phy.queensu.ca/~phil/exiftool`

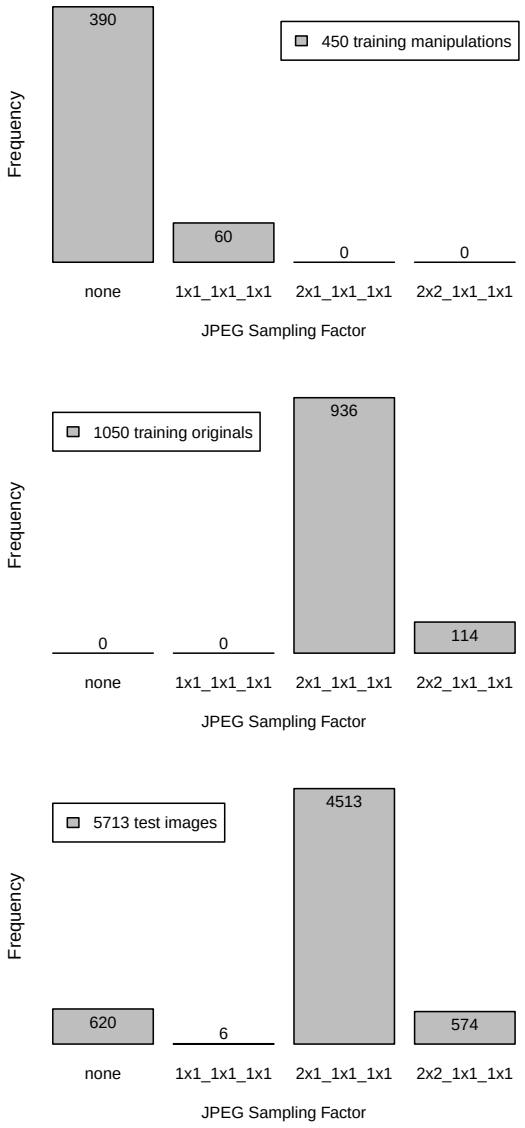

Figure 1: JPEG sampling factor occurrences in the metadata. Histograms from top to bottom: manipulated images in the training set, original images in the training set, images of unknown class in the test set. Manipulated images and original images in the training set belong to disjoint subsets.

combinations refer to luminance and Cb/Cr chroma subsampling settings). Interestingly, *none* of the original images shared these metadata contents. Instead, 936 out of the 1050 samples in the training set named a JPEG sampling factor '`2x1,1x1,1x1`'. The remaining 114 images had '`2x2,1x1,1x1`'. The two uppermost panels in Figure 1 visualize the occurrence of different '`jpeg:sampling-factor`' fields in the original and manipulation training sets. Extending the analysis to the 5713 images in the test set, we encountered the exact same four configurations as in the training set. The bottom panel in Figure 1 reports the specific frequencies. Observe that also the overall order of the individual categories remained the same. Our hypothesis then was formulated in a straight-forward manner:

> *All images in the test set with the* '`jpeg:sampling-factor`' *field set to* '`1x1,1x1,1x1`' *or missing are manipulated. All images in the test set with the* '`jpeg:sampling- factor`' *field set to* '`2x1,1x1,1x1`' or '`2x2,1x1,1x1`' *are original.*

A visual inspection of selected images could not substantially falsify this hypothesis. Many potential manipulations could be confirmed by simply looking at the image. Some less clear cases could be resolved by a comparison with original images of the same or a similar scene in the pile of original training images. Still, a good amount of images that we assumed to be manipulated were not directly

67bd605bfb967e11d6fe481be8e21269

a5af9017a60e5f2206f69486694e3210

a3d2ed17f91f008ea41951795545eba7

a5e5a9a9110a293f63e8cd10e908d489

a30ef361e8eba3e176989dcb7bb95056

a79e69e6960588007f647bcd6a599f66

Figure 2: All six test images with PNG text chunk '`jpeg:sampling-factor`' set to '`1x1,1x1,1x1`'.

3dc169ca1c6fac15eb894ea4070b3354

Figure 3: Original image from the training set, showing the same scene as one of the allegedly manipulated images in Figure 2 ('a3d2ed17f91f008ea41951795545eba7').

identifiable as such. Yet at the same time, none of the potential originals seemed to exhibit strong visual clues that would indicate a manipulation in return.

Figure 2 exemplarily displays all six images in the test dataset with a '`1x1,1x1,1x1`' JPEG sampling factor. The four images in the two left-most columns are unmistakably manipulated. The forgery in the upper right was uncovered through a corresponding original image, depicted in Figure 3 (a white object was added left of the door). The last image in Figure 2 turned out to be not as visually clear.

Eventually, we decided to submit a first complete test run according to our hypothesis. A 100 % accuracy proved it true. To get an impression whether our findings would generalize in any way to real-world forensic investigations (ignoring that images are typically stored in the JPEG format) or whether they are a product of specific artifacts in the contest's database, we converted original JPEG images from the Dresden Image Database [3] to the PNG format. Table 1 summarizes and compares the metadata entries of three typical cases. The two left-most columns represent manipulated images from the contest, the right-most column corresponds to one of our original images, acquired with a Casio EX-Z150 digital camera. We used ImageMagick's `convert` with standard settings for format conversion.[3] Differences between the images are highlighted in red. Observe that the general structure of metadata

[3] `http://www.imagemagick.org`

Table 1: PNG chunks in two selected test images in comparison to an original image of the 'Dresden Image Database' after conversion to the PNG format with Image Magick's `convert`. Differences are highlighted in red.

chunk	description	entry	testdata		Casio image
			a011dd03bb8ec05cd9894e53c1e2dd6a	67bd605bfb967e11d6fe481be8e21269	Casio_EX-Z150_0_4978
IHDR	image header	width	1024	1024	3264
		height	645	768	2448
		bit depth	8	8	8
		color type	6	2	2
		compression method	0	0	0
		filter method	0	0	0
		interlace method	0	0	0
gAMA	image gamma		2.2	2.2	2.2
sRGB	standard RGB color space		perceptual	perceptual	perceptual
cHRM	primary chromaticities	white point x	0.31269	0.31270	0.31270
		white point y	0.32899	0.32900	0.32900
		red x	0.63999	0.64000	0.64000
		red y	0.33000	0.33000	0.33000
		green x	0.21000	0.30000	0.30000
		green y	0.71000	0.60000	0.60000
		blue x	0.14999	0.15000	0.15000
		blue y	0.05999	0.06000	0.06000
bKGD	background color		255 255 255	255 255 255	
pHYs	physical pixel dimensions	pixels per unit x	9449	7086	2834
		pixels per unit y	9449	7086	2834
		unit specifier	meter	meter	meter
IDAT	image data				
tEXt	textual information	date:create	2013-05-01T17:07:36-03:00	2013-05-01T17:07:36-03:00	2013-09-30T10:45:41+02:00
tEXt	textual information	date:modify	2013-04-16T16:29:15-03:00	2013-04-16T15:42:00-03:00	2013-09-30T10:45:41+02:00
tEXt	textual information	jpeg:colorspace	NA	2	2
tEXt	textual information	jpeg:sampling-factor	NA	1x1,1x1,1x1	2x1,1x1,1x1
IEND	image trailer				

chunks is largely the same. In accordance to the training set, our original image has a '`2x2,1x1,1x1`' JPEG sampling factor where the manipulated images display the reported peculiarities. Interestingly, PNG images converted with Gimp 2.8.2 or Apple Preview 6.0.1 exhibit a different selection and order of chunks (not reported in the table). This suggests similar distinctive features as they are present in JPEG files [2], and we surmise that `convert` was used to prepare the images for the contest. We plan to further investigate these so far unreported differences in our future work. By and large, we do believe, however, that already our present findings alone re-emphasize that metadata is a forensically relevant part of a digital image file, independent of its format. Hence, forensic investigators and image forgers alike must not ignore this data.

References

[1] H. Farid. *Digital Image Ballistics from JPEG Quantization: A Followup Study*. Tech. rep. TR2008-638. Hanover, NH: Department of Computer Science, Dartmouth College, 2008.

[2] T. Gloe. "Forensic Analysis of Ordered Data Structures on the Example of JPEG Files". In: *IEEE Workshop on Information Forensics and Security (WIFS)*. 2012, pp. 139–144.

[3] T. Gloe and R. Böhme. "The Dresden Image Database for Benchmarking Digital Image Forensics". In: *Journal of Digital Forensic Practice* 3.2–4 (2010), pp. 150–159.

[4] E. Kee, M. K. Johnson, and H. Farid. "Digital Image Authentication from JPEG Headers". In: *IEEE Transactions on Information Forensics and Security* 6.3 (Sept. 2011), pp. 1066–1075.

[5] H. T. Sencar and N. Memon. "Overview of State-of-the-art in Digital Image Forensics". In: *Algorithms, Architectures and Information Systems Security*. Ed. by B. B. Bhattacharya, S. Sur-Kolay, S. C. Nandy, and A. Bagchi. Vol. 3. Statistical Science and Interdisciplinary Research. World Scientific Press, 2008. Chap. 15, pp. 325–348.