Digital Image Authentication from JPEG Headers

Rodrigo E. Gonçalves
Vitor Machado
Wesley T. S. T. Ide
Seminar Series

MC919 - Digital Forensics

Prof. Dr. Anderson Rocha

anderson.rocha@ic.unicamp.br

http://www.ic.unicamp.br/~rocha
Outline

▸ Introduction
▸ JPEG Standard
▸ Camera Signature
▸ Image Authentication
▸ Image Database
▸ Results
▸ Conclusions
▸ Countermeasures
▸ Future studies
▸ References
Introduction
Introduction

- Digital images have been introduced as evidence into courts of law.
- Verifying the integrity of this digital evidence is critical!
Introduction

- Digital forensic techniques have been developed to detect various traces of tampering as region duplication, shadows, etc.

- However, such methods focus on manipulation detection.
Introduction

- On the other hand, authentication is interested in pointing out whether an image is authentic or not.

- Simple or small modifications render the image not authentic anymore.
Introduction

‣ [Mahdian & Saic 2009]
‣ Based on detecting double JPEG compression
‣ Promises to detect any form of image manipulation
‣ Requires fairly sophisticated models and analysis schemes
‣ Vulnerable to simple countermeasures
‣ Can be computationally intensive

‣ [Kee et al. 2011]
‣ Based on JPEG standard, and aspects of the EXIF header format
‣ Computationally efficient
‣ Demands a big database of cameras models
Image Authentication from JPEG Headers
JPEG Standard

- Defines a lossy compression scheme for natural scenes and the JPEG container format.
JPEG Standard

- During the JPEG compression process, the specific quantization tables and Huffman codes needed to decode a JPEG file are embedded into the JPEG header.
Camera Signature

- Approach for detecting the authenticity of JPEG image from its original recording.
- Based on the JPEG header information.
- A camera signature is extracted from a JPEG image header:
  - Dimension
  - Quantization tables
  - Huffman codes
  - Thumbnails
  - EXIF format
Camera Signature

**Image Parameters:**

- First three components of a camera signature:
  - Image dimensions
  - Quantization tables
  - Huffman code
Camera Signature

**Image Parameters:**

- **Image dimensions:**
  - Used to distinguish between cameras with different sensor resolution.
  - In order to compensate the images orientation, the minimum followed by the maximum dimension.
Camera Signature

Image Parameters:

- Quantization tables:

<table>
<thead>
<tr>
<th>image quantization table (Y)</th>
<th>1 1 1 1 1 2 3 3 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 1 1 1 1 1 3 3 3</td>
</tr>
<tr>
<td></td>
<td>1 1 1 1 1 2 3 3 3</td>
</tr>
<tr>
<td></td>
<td>1 1 1 1 1 3 4 4 3</td>
</tr>
<tr>
<td></td>
<td>1 1 2 3 3 5 5 4</td>
</tr>
<tr>
<td></td>
<td>1 2 3 3 4 5 6 5</td>
</tr>
<tr>
<td></td>
<td>2 3 4 4 5 6 6 5</td>
</tr>
<tr>
<td></td>
<td>4 5 5 5 6 5 5 5</td>
</tr>
</tbody>
</table>

© Kee et al.
Camera Signature

**Image Parameters:**

- **Quantization tables:**
  - Specified as an array of 192 values.
  - Each table is specified in column-order and the three tables are specified in the order of luminance (Y), chrominance (Cb) and chrominance (Cr).
Camera Signature

Image Parameters:

- Huffman code:

<table>
<thead>
<tr>
<th>Image Huffman Code DC (Y)</th>
<th>1511111100000000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image Huffman Code DC (Cb)</td>
<td>3111111111100000</td>
</tr>
<tr>
<td>Image Huffman Code DC (Cr)</td>
<td>0000000000000000</td>
</tr>
<tr>
<td>Image Huffman Code AC (Y)</td>
<td>2133243554401125</td>
</tr>
<tr>
<td>Image Huffman Code AC (Cb)</td>
<td>21244347544012119</td>
</tr>
<tr>
<td>Image Huffman Code AC (Cr)</td>
<td>0000000000000000</td>
</tr>
</tbody>
</table>

© Kee et al.
Camera Signature

Image Parameters:

- Huffman code:
  - Six sets of 15 values corresponding to the number of codes of length 1, 2, … 15
  - Each of three channels requires two codes, one for the DC coefficients and one for the AC coefficients
Camera Signature

Image Parameters:

- 284 values were extracted from a full resolution image:
  - 2 of images dimension
  - 192 of quantization
  - 90 of Huffman codes
Camera Signature

Thumbnail Parameters:

▶ Size is typically a few hundred square pixels.

▶ Created by cropping, filtering and downsampling the full-resolution image.

▶ Typically compressed and stored in the header as a JPEG image.
Camera Signature

Thumbnail Parameters:

- 284 values were extracted from the thumbnail image:
  - 2 of images dimension
  - 192 of quantization
  - 90 of Huffman codes

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>194</td>
<td>284</td>
<td>285</td>
<td>286</td>
<td>478</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>568</td>
</tr>
</tbody>
</table>

© R. Gonçalves et al.
Camera Signature

EXIF Metadata Parameters:

| EXIF count | 9 28 2 6 0 7 162 0 |

- Found in the JPEG header.
- Stores information about camera and image.
- According to the EXIF standard, there are five main image file directories (IFDs) into which the metadata is organized.
Camera Signature

**EXIF Metadata Parameters:**

- The authors extract a compact representation by counting the number of entries in each IFD.
- The EXIF standard allows for the creation of additional IFDs.
- The authors count the total number of any additional IFDs and their entries.
Camera Signature

**EXIF Metadata Parameters:**

- Some camera manufacturers customize their metadata in ways that do not conform to the EXIF standard.

- These customizations yield errors when parsing the metadata.

- The authors consider these errors to be a feature of camera design and therefore count the total number of parser errors.
Camera Signature

EXIF Metadata Parameters:

- 8 values were extracted from the metadata:
  - 5 entry counts from the standard IFDs
  - 1 for the number of additional IFDs
  - 1 for the number of entries of additional IFDs
  - 1 for the number of parser errors
Image Authentication
Image Authentication

- Authors extracted signature from the image.
- Photo alteration is detected by comparing the signature with a database of known authentic camera signatures.
Image Authentication

- An image’s EXIF metadata can be relatively easily edited, so it may seem peculiar to rely on it for forensic purposes.

- Modifying the content of any existing EXIF field will not affect the extracted EXIF counts nor the extracted signature.
Image Authentication

- It is possible to change the camera make and model fields in an attempt to conceal the camera source.

- This can be detected when the extracted signature is found to be inconsistent with the make and model.
Image Authentication

- If the number of EXIF fields are increased or decreased, then the EXIF count will be inconsistent with the expected signature.
- Although an image’s EXIF metadata can be edited, it still provides useful information for forensic and ballistic analysis.
Image Database
Image Database

10 millions of images collected from Flickr.com.

- Removed images based on some criterias:
  - Not tagged as "original" by Flickr
  - Not with 3-channel color JPEG
  - Duplicated (MD5 hashes)
  - Creation and modification date different
Image Database

- With metadata of some software (e.g. Adobe Photoshop)
- Resized by Flickr
- Resolution different of native camera (analysed by Mechanical Turk)

Reduced to 2.2 million images...
Image Database

- Entries with less than 25 images having the same paired make, model, and signature were removed.

Reduced to 1.3 million images!
Image Database

Final numbers:

- 1.3 million of images
- 9,163 cameras with distinct configurations
- 33 camera manufacturers
- 773 camera and cellphone models
- Average of 12 different signatures for each camera make and model
Results
Results

- Number of images for each camera configuration:
Results

- 9,163 different camera configurations were separated in equivalence classes.
- An equivalence class of size n means that n cameras of different make and model share the same signature.
Results

- Distributions of equivalence class sizes were based on the distinctiveness of the signature using each of 3 parameters.
Results

- Signatures consisting of image parameters only:

© Kee et al.
Results

- Thumbnail parameters only:
Results

- EXIF parameters only:
Results

- Image + thumbnail parameters:

© Kee et al.
Results

- All parameters:
Results

- Summary of equivalence classes:

<table>
<thead>
<tr>
<th></th>
<th>Equivalence Class Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>image</td>
<td>12.9%</td>
</tr>
<tr>
<td>thumbnail</td>
<td>1.1%</td>
</tr>
<tr>
<td>EXIF</td>
<td>8.8%</td>
</tr>
<tr>
<td>image + thumbnail</td>
<td>24.9%</td>
</tr>
<tr>
<td>all</td>
<td>69.1%</td>
</tr>
</tbody>
</table>

© Kee et al.
Conclusions
Conclusions

‣ Separately, the image, thumbnail and EXIF parameters are not particularly distinct, but when combined, they provide a highly distinct signature.

‣ Although the thumbnail parameters are the least distinct, their addition to the overall signature is significant.
Conclusions

- After that, an analysis was made comparing the relative distributions of the individual parameters.
Conclusions

- Relative distribution of all configurations:
Conclusions

- EXIF count if most unique, followed by image dimensions, image quantization table (Y channel), and then thumbnail quantization table (Y channel).

- Least distinct parameters are the Huffman tables and the thumbnail dimensions.
Conclusions

- Each equivalence class consists of cameras from the same manufacturer and series. The only exceptions are the following four equivalence classes of size two:
  - Casio EX-Z60 | Canon Powershot SX120 IS
  - Nikon Coolpix P90 | Panasonic DMC-FZ18
  - Panasonic DMC-TZ5 | Nikon Coolpix S52
  - Panasonic DMC-ZS7 | Nikon Coolpix S630
Conclusions

- An equivalence class of size greater than 1 implies a non distinct signature (ambiguity in identifying the camera and model).

- Some group of camera models (e.g. Sony DSC series) are likely to be in the same equivalence class.

- Although there is an ambiguity, the signature still significantly constrains the identity of the camera make and model.
Conclusions

- The signatures from Adobe Photoshop were compared with the 9,163 camera signatures (only the image and thumbnail quantization tables and Huffman codes were used for comparison).
- No overlap was found between Photoshop and camera signatures.
Countermeasures
Countermeasures

‣ Forgers could conceal their traces of tampering by extracting the signature of a camera, modifying the image, and then re-saving the image with the appropriate parameters.

‣ The analysis is also vulnerable to a standard re-broadcast attack in which a digital image is manipulated, printed, and re-photographed.
Future studies
Future studies

▶ Facing the challenge of new cameras and cellphones being constantly released by acquiring signatures from a wide variety of devices

▶ Device authentication [Lukáš et al. 2006] to compare an image of unknown origin to a database of known cameras (can be computationally demanding)

▶ Reducing the complexity by focusing only on relevant cameras
References
References


Thank You!

Obrigado!