Digital Image Forensics for Device Attribution

Anselmo Ferreira and Anderson Rocha

Reasoning for Complex Data Lab. Institute of Computing State University of Campinas
Summary

1. The problem
2. Moving towards the solution
   2.1 How devices work
   2.2 Understand devices fingerprinting
   2.3 Devices attribution
      2.3.1 State of the art
3. Anti-Forensics approaches
4. Conclusion
The problem

• High availability of electronic devices:
  ▪ They are cheap!

• Our daily documents are now created with these devices:
  ▪ Images.
  ▪ Images created by scanning processes.
  ▪ Videos.
  ▪ Printed Documents.
The problem

• Problem: not only legal documents are created:
  ▪ Terrorist plans.
  ▪ Fake currency.
  ▪ Child Pornography and animal abuse photos.

• How to prove the ownership of these criminal documents?
The problem

Figure 1: Controversial Image posted in Instagram highlights the problem of device source attribution. Extracted from [1].
The problem

Figure 2: one example on how devices can be used for criminal purposes. Extracted from [2].
The problem

Figure 3: Child porn photographs arise the importance of devices attribution. Extracted from [3].
How to move towards the solution

• What is Device Attribution?

“A set of computer vision techniques applied in a digital version of a document, aimed at pointing out which device is the source of the document.”
How to move towards the solution

Figure 4: which printer printed which photo? Extracted from [4].
How to move towards the solution

Figure 5: which device created which document? Extracted from [4].
How to move towards the solution

- Devices Attribution can be done by searching for two kinds of signatures:
  - Intrinsic: inserted by the device in the document.
  - Extrinsic (Blind): given by the analysis of the resulting document.
How to move towards the solution

- Device Attribution involves answering two questions:
  1. Which device model and brand produced a given document?
  2. What specific device produced a given document?
How to move towards the solution

Steps for device attribution:

1. Understand how these devices work.
2. Find unique behavior of each device in the document (e.g., Noise, Texture and Distortions).
3. Describe these behaviors for device attribution.
STEP ONE
HOW THESE DEVICES WORK
HOW LASER PRINTERS WORK

• Data written by a laser beam, which discharges certain places in a drum where ink must be put.

• Positively charged ink is then sticked in the discharged places of the drum.

• The data with ink is spread on the paper by the fuser.
HOW CAMERAS WORK

• The signal (light) is first captured and filtered inside the lens and camera.

• CCD (or CMOS) converts this light into electrical charges in an array of capacitors over a sensor.

• Image processors will improve this analogic-Digital conversion by a series of pixel operations, yielding the final image.

<table>
<thead>
<tr>
<th>Light</th>
<th>Lens</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Lens" /></td>
</tr>
<tr>
<td><img src="image3.png" alt="Light Filters" /></td>
<td><img src="image4.png" alt="CCD" /></td>
</tr>
<tr>
<td><img src="image5.png" alt="Filters" /></td>
<td><img src="image6.png" alt="Mosaicking" /></td>
</tr>
</tbody>
</table>

- Image Stabilization
- Focus
- Demosaicing
- Noise reduction
- Image sharpening
- Compression
HOW SCANNERS WORK

- A mirror reflects light that from the paper in the glass, illuminated by a lamp.
- A set of lens focus the reflected light from the mirror into a sensor (CCD).
- Data in the sensor is digitized via an analog-digital converter (ADC)

http://www.extremetech.com/computing/51612-how-scanners-work/2
STEP TWO
INVESTIGATING DEVICES
FINGERPRINTING
• Common fingerprints:
  ✓ Texture
  ✓ Distortion
  ✓ Noise
  ✓ Imperfections such as dust, scratches, etc.
  ✓ Among others

• They are commonly yielded by devices manufacturing process.
INVESTIGATING DEVICES
FINGERPRINTING

• LASER PRINTERS
  ✓ Electromechanical devices with moving parts that behave differently.
  ✓ Differences can be seen on the halftones of printed material.
  ✓ **Banding** [5]: nonuniform light and dark lines perpendicular to direction in which the paper moves through the printer.
INVESTIGATING DEVICES
FINGERPRINTING

• LASER PRINTERS
  ✓ Attribution process happens on a digitalized (scanned) version of a document.

• The following approaches are used to detect particular banding of a given printer:
  ✓ Frequency Analysis of Banding (Fourier Spectrum Analysis) for color documents
  ✓ Texture among printed material for text.
  ✓ Etc.
Figure 5: Approaches for Printer Attribution differ when applied to text or color documents.
Figure 6: steps for printer attribution
SCANNERS

- Attribution based on source camera identification (i.e., sensor noise).
- Two-dimensional (2-D) noise pattern of the image is used to identify source scanner.
SCANNERS

Approaches for device attribution:

- Statistical features of the noise
- High-frequency wavelet coefficients of the noise
- Neighborhood prediction errors of the noise
- Investigating how the sensor is moved within the scanner mechanism
- Dust or scratches
STEP THREE- Describing devices fingerprinting for devices attribution
Describing devices fingerprinting for devices attribution

- LASER PRINTERS
Describing devices fingerprinting for devices attribution

- Approaches used for laser printer attribution.
  - halftone-based [6, 7]: applied only in color documents.
  - texture-based [8-13]: applied on text documents.
  - noise-based [14, 15, 16]: applied in both.
Describing devices fingerprinting for devices attribution

The following areas of a document can be used for analysis in laser printer attribution:

- Characters -> applied only in text
- Segmented Areas (Frames) -> applied in text and images
- Whole Document-> applied in text and images
Describing devices fingerprinting for devices attribution

Figure 7: Letter (left) and Frame approach (right).
Laser Printer attribution by Ali et al [8]:

- Applied in letters of text (letter ‘I’)
- The projection (pixel values) are used as fingerprints.
- A machine learning classifier (Gaussian mixture model) is used to recognize these texture patterns for each printer.
Describing devices fingerprinting for devices attribution

Figure 8: Ali et al's approach for laser printer attribution. Extracted from [8].
Laser Printer attribution by Lee et al [14]:

- Applied in color documents (images)
- Printed documents are scanned and converted to CMY color space
- Noise of CMY image is isolated by subtracting the original image CMY and CMY filtered by the Wiener Filter.
Laser Printer attribution by Lee et al [14]:

- Texture information is calculated in this noise image by statistics of five gray level co-occurrence matrices.
- A machine learning classifier is used to recognize these texture patterns.
Describing devices fingerprinting for devices attribution

Figure 9: Lee et al's approach for laser printer attribution. Extracted from [14].
Laser Printer Attribution by Mikkilineni et al [9]

- Applied in text documents.
- Letters “e” are extracted in windows of approximately 180 x 160 pixels.
- Texture descriptors, based on statistics of gray level co-occurrence matrices, are used with machine learning.
Describing devices fingerprinting for devices attribution

Figure 10: Mikkilineni et al's approach for laser printer attribution. Extracted from [9].
Describing devices fingerprinting for devices attribution

- Laser Printer Attribution by Kee and Farid [10]
  - Applied in text documents
  - Letters “e“ are extracted in squared windows.
  - Technique is divided in three steps
    - Pre-processing
    - Printer Profile
    - Ballistics
Laser Printer Attribution by Kee and Farid [10]

- Preprocessing.
  1. A reference character is chosen.
  2. Similar letters are searched, preprocessed by histogram normalization and registered with the reference letter.
Describing devices fingerprinting for devices attribution

- Laser Printer Attribution by Kee and Farid [10]
  - Printer Profile.
  1. Each aligned character in previous step is used as a column of a Matrix $D$.
  2. Principal Component Analysis is performed in $D$.
  3. The final printer consists of both the mean character $\bar{u}$ and the top $p$ eigenvalue eigenvectors yielded by PCA: $e_{ij}, i \in [1, p]$
Describing devices fingerprinting for devices attribution

Laser Printer Attribution by Kee and Farid [10]

- Ballistic.
  1. Suspected characters in vector form, \( c_j \), are first aligned to the reference character.
  2. Each aligned character is then projected onto each basis vector:

\[
\alpha_{ji} = (c_j - \bar{u})e_i \quad (1)
\]
Describing devices fingerprinting for devices attribution

- Laser Printer Attribution by Kee and Farid [10]
  - Ballistic.

3. the reconstruction error between a test character $c_j$ and the new basis representation is computed to determine the suitability of the printer profile

$$r_j = \bar{u} + \sum_{i=1}^{p} \alpha_{ji} \tilde{e}_i$$
Describing devices fingerprinting for devices attribution

- Laser Printer Attribution by Kee and Farid [10]
  - Ballistic.
  4. The error between the test character \((c_j)\) and the reconstructed one \((r_j)\) must be minimum for a the source printer.

\[ E_j = \sqrt{(\bar{c}_j - \bar{r}_j)^T(\bar{c}_j - \bar{r}_j)} \] (3)

5. Again, this is done per printer profile (i.e., per printer)
Laser Printer Attribution by Choi et al [16]

- Applied in color documents (images)
- Based on statistics of Discrete Wavelet Transform from color bands
- 39 statistical features are extracted from the HH Discrete Wavelet Transform sub-band per image.
Describing devices fingerprinting for devices attribution

Figure 11: Printer noise signatures seen from HH sub-band of DWT from the same image printed with 4 different printers. Extracted from [16].
Describing devices fingerprinting for devices attribution

Figure 12: Choi et al's approach for laser printer attribution. Extracted from [16].
Describing devices fingerprinting for devices attribution

OUR SOLUTION!
Describing devices fingerprinting for devices attribution

Our solution:

- We propose the multiscale and multidirectional texture analysis inside printed material.
- Our analysis can work on images, texts or both.
- It can be applied in the whole document, in letters or frames.
Describing devices fingerprinting for devices attribution

Figure 14: Microscope analysis of printed text shows that inside printed material there are multidirectional and multiscale texture patterns.
Our Contributions

1. A multidirectional GLCM approach
2. A multidirectional/multiscale GLCM approach
3. A convolutional gradient multidirectional and multiscale texture filter
Our Contributions

4. More realistic dataset
5. Investigation on chunks of documents (frames)
6. Dimensionality reduction approach
Describing devices fingerprinting for devices attribution

1. Multidirectional GLCM approach

- GLCMs: 2d histograms aimed at describing the neighborhood of pixels of a given image in a given direction and offset.

- A series of statistics are calculated from these matrices and are used for image description.

- Our solution is using more directions (eight) in the GLCM approach.
1. Multidirectional GLCM approach

Figure 15: (Left) original approach yields 4 GLCMs (Right) our proposed approach yields 8 GLCMs.
Describing devices fingerprinting for devices attribution

1. Multidirectional GLCM approach

   - At each direction (GLCM), 22 statistics are calculated

   - A $22 \times 8 = 176$ dimensional feature vector is used to identify the texture of a given printer
2. Multidirectional/Multiscale GLCM approach

- We used the gaussian image decomposition in the last approach

- We use four scales: the original, two downscales and one up-scale.

- At each scale, 176 statistical features are extracted as in the previous approach.
Describing devices fingerprinting for devices attribution

Figure 16: Our multidirectional and multiscale GLCM approach.
3. A convolutional gradient multidirectional and multiscale texture filter

- Textures on almost flat areas (with small gradient value) are intentionally generated by the printer firmware.

- Our third approach for laser printer attribution is a filter, called Convolution Texture Gradient Filter (CTGF).

- It relies on the analysis of texture at these low-level gradient areas.
Describing devices fingerprinting for devices attribution

3. A convolutional gradient multidirectional and multiscale texture filter

- CTGF learns a set of \( n \times n \) pixel patterns (texture) in low-gradient areas that appear more frequently in a given printer, but not in others.

- Given a printed document, seven transformations are applied to find the printer signature.
Describing devices fingerprinting for devices attribution

Figure 17: Our Convolutional Gradient Filter Approach
Step 1: Negative

- As a pre-processing step, the image pixels in $S$ are inverted.
- Values close to zero will mean white pixels and 255, black pixels.
- This is made for convenience in the algorithm operations and yields a negative image $N$. 
Describing devices fingerprinting for devices attribution

- **Step 2: Crop borders**
  - Used to eliminate scanning noise at the image borders, generated by external light, folding,
  - The negative image N is cropped, eliminating 6% of pixels in each border.
  - New matrix: $R$. 
Step 3.1: Convolution with ones.

- Textures with \( n \times n \) neighbor pixels contained in \( R \) are then represented by their sum and maximum gradient between the central pixel and its neighbors.
- The convolution of \( R \) with an \( n \times n \) matrix of ones results in the matrix of textures sums \( C \).
Describing devices fingerprinting for devices attribution

- Step 3.1: Convolution with ones

\[ C = R \star \begin{bmatrix} 1 & \ldots & 1 \\ \ldots & \ldots & \ldots \\ 1 & \ldots & 1 \end{bmatrix} \]
Describing devices fingerprinting for devices attribution

- **Step 3.2: Gradient(R)**
  - In parallel with the previous step, a gradient between a pixel and its $n^2-1$ neighbors is calculated.
  - The maximum gradient is used.
  - A new matrix, $G$ is created.
Step 3.2: Gradient Filter

- Textures with the gradients of interest are filtered.
- \( g_{\text{low}} \) and \( g_{\text{high}} \) define the range of gradient values that are valuable for printer signature.
- Such parameters are selected from a training set of documents per suspected printer for maximum results on the learning process.
Describing devices fingerprinting for devices attribution

- **Step 3.2: Gradient Filter**
  - The matrix $T$ of texture codes (sums) is then created by filtering textures that are in the defined range.

$$T = \begin{bmatrix} t_{1,1} & \ldots & t_{1,c-2} \\ \vdots & \ddots & \vdots \\ t_{r-2,1} & \ldots & t_{r-2,c-2} \end{bmatrix}, \text{ where } \begin{cases} t_{i,j} = c_{i,j} & \text{if } g_{\text{low}} \leq g_{i,j} \leq g_{\text{max}} \\ t_{i,j} = 0 & \text{otherwise} \end{cases}$$ (5)
Describing devices fingerprinting for devices attribution

- **Step 4: Histogram**
  - Counting the number of positions for each texture in T from one to $255 \times n^2$ generates the histogram vector $H$ with $255 \times n^2$ dimensions.

\[ H = \text{hist}(T, 1 : 255 \times n^2) \]
Describing devices fingerprinting for devices attribution

- **Step 5: MinMax.**
  - The final feature vector \( V \) is generated by applying a Min-Max normalization on \( H \), scaling the components to the interval \([0:1]\).

  \[
  V_{ij} = \frac{H_{ij} - u}{v - u} \quad u = \min(H), \quad v = \max(H) \quad (7)
  \]

  - As the final feature vector is a histogram of sums of pixels, it has \(255 \times n^2\) dimensions, where \(n\) is the dimension of a squared sliding window used to calculate the texture.
Figure 18: Filtered textures using the proposed CTGF in printed text (top) and images (bottom) are different in different printers.
Figure 19: Printer signatures from different printers using the proposed CTGF.
4. More realistic dataset
   - Databases used in prior works are limited in some way because:
     - These databases always consider fonts of same size and style.
     - They are composed by text or figures.
     - They expect that the scanned documents are available in high resolutions.
4. More realistic dataset

- 120 documents were printed on ten laser printers in standard resolutions, yielding 1,184 documents.
- These are Wikipedia documents with one, two or three pages and contain different letter sizes, fonts and figures.
Describing devices fingerprinting for devices attribution

4. More realistic dataset
   ▪ Scanned by a Canon 8800 Scanner at a 600 dpi resolution and saved in TIFF format.
   ▪ Separated by two factors: Language (English or Portuguese) and Figures (With or Without).
Describing devices fingerprinting for devices attribution

Table 1: number of documents per printer used in our proposed dataset.

<table>
<thead>
<tr>
<th>#</th>
<th>Printer ID</th>
<th>Manufacturer</th>
<th>Laser Printer Model</th>
<th>Number of Printed Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B4070</td>
<td>Brother</td>
<td>HL-4070CDW</td>
<td>120</td>
</tr>
<tr>
<td>2</td>
<td>C1150</td>
<td>Canon</td>
<td>D1150</td>
<td>116</td>
</tr>
<tr>
<td>3</td>
<td>C3240</td>
<td>Canon</td>
<td>MF3240</td>
<td>120</td>
</tr>
<tr>
<td>4</td>
<td>C4370</td>
<td>Canon</td>
<td>MF4370DN</td>
<td>120</td>
</tr>
<tr>
<td>5</td>
<td>H1518</td>
<td>Hewlett Packard</td>
<td>CP1518</td>
<td>120</td>
</tr>
<tr>
<td>6</td>
<td>H225A</td>
<td>Hewlett Packard</td>
<td>CP2025A</td>
<td>119</td>
</tr>
<tr>
<td>7</td>
<td>H225B</td>
<td>Hewlett Packard</td>
<td>CP2025B</td>
<td>110</td>
</tr>
<tr>
<td>8</td>
<td>LE260</td>
<td>Lexmark</td>
<td>E260DN</td>
<td>119</td>
</tr>
<tr>
<td>9</td>
<td>OC330</td>
<td>OKI Data</td>
<td>C330DN</td>
<td>120</td>
</tr>
<tr>
<td>10</td>
<td>SC315</td>
<td>Samsung</td>
<td>CLP-315</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td><strong>1,184</strong></td>
</tr>
</tbody>
</table>
5. Investigation on chunks of documents (frames)

- Here we propose to investigate the laser printer signatures in segmented areas of the document.
- Useful when only parts of documents are available.
- The frames are rectangular areas of the image with sufficient printed material.
5. Investigation on chunks of documents (frames)

- The document (A4 paper) is divided in a matrix of frames with five columns by six rows of about 900 by 980 pixels.
- We state that the minimum accepted ratio between dark pixels (black and dark grey) and blank ones (blank and light gray) should be 0.02.
Describing devices fingerprinting for devices attribution
Describing devices fingerprinting for devices attribution

Figure 21: Experiments show that the best gradient values for the proposed CTGF filter is (1,32).
Describing devices fingerprinting for devices attribution

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy Statistics on Crossfolding 5x2 Experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
</tr>
<tr>
<td>CTGF_MDMS_F</td>
<td>98.97</td>
</tr>
<tr>
<td>GLCM_MDMS_F</td>
<td>97.10</td>
</tr>
<tr>
<td>GLCM_MDMS_C</td>
<td>96.63</td>
</tr>
<tr>
<td>GLCM_MD_F</td>
<td>95.40</td>
</tr>
<tr>
<td>GLCM_MD_C</td>
<td>95.78</td>
</tr>
<tr>
<td>HOG_C [36]</td>
<td>94.42</td>
</tr>
<tr>
<td>LBP_F [33]</td>
<td>94.22</td>
</tr>
<tr>
<td>CTGF_3x3_F</td>
<td>93.17</td>
</tr>
<tr>
<td>GLCM_C [3,4]</td>
<td>92.24</td>
</tr>
<tr>
<td>CTGF_MDMS_D</td>
<td>88.70</td>
</tr>
<tr>
<td>GLCM_F [3,4]</td>
<td>91.82</td>
</tr>
<tr>
<td>LBP_C [33]</td>
<td>88.16</td>
</tr>
<tr>
<td>GLCM_MD_D</td>
<td>84.32</td>
</tr>
<tr>
<td>GLCM_MDMS_D</td>
<td>86.34</td>
</tr>
<tr>
<td>LBP_D [33]</td>
<td>86.17</td>
</tr>
<tr>
<td>CTGF_5x5_F</td>
<td>85.32</td>
</tr>
<tr>
<td>RECONSTRUCT_ERROR_C [7]</td>
<td>81.79</td>
</tr>
<tr>
<td>CTGF_7x7_F</td>
<td>80.89</td>
</tr>
<tr>
<td>CTGF_3x3_D</td>
<td>82.29</td>
</tr>
<tr>
<td>GLCM_D [3,4]</td>
<td>78.75</td>
</tr>
<tr>
<td>CTGF_5x5_D</td>
<td>77.74</td>
</tr>
<tr>
<td>HOG_D [36]</td>
<td>78.00</td>
</tr>
<tr>
<td>CTGF_7x7_D</td>
<td>73.52</td>
</tr>
<tr>
<td>HOG_F [36]</td>
<td>72.74</td>
</tr>
<tr>
<td>NOISE_STATS_C [18]</td>
<td>67.29</td>
</tr>
<tr>
<td>NOISE_STATS_F [18]</td>
<td>40.48</td>
</tr>
<tr>
<td>NOISE_STATS_D [18]</td>
<td>37.27</td>
</tr>
<tr>
<td>DWT_STATS_D [22]</td>
<td>33.00</td>
</tr>
<tr>
<td>DWT_STATS_F [22]</td>
<td>32.08</td>
</tr>
<tr>
<td>DWT_STATS_C [22]</td>
<td>24.96</td>
</tr>
</tbody>
</table>

Table 2: Laser printer attribution experiments comparing our proposed technique against the state the art in characters (c), frames (f) and documents (d).
# Describing devices fingerprinting for devices attribution

## Table 4: Laser printer attribution experiments comparing f-measures.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CTGF_MDMS_F</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GLCM_MDMS_F</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GLCM_MDMS_C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GLCM_MD_D</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>LBP_F [33]</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>HOG_C [36]</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CTGF_3x3_F</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GLCM_C [3,4]</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CTGF_MDMS_D</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GLCM_F [3,4]</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>RECONST_ERROR_C [7]</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CTGF_3x3_D</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NOISE_STATS_C [18]</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DWT_STATS_D [22]</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Legend:**
- **1** = Row method is statistically better than column method
- **0** = Row method is statistically equivalent to column method
- **-1** = Row method is statistically worse than column method
Describing devices fingerprinting for devices attribution

- Scanners
Scanners

- As scanners have almost similar function than cameras, their identification are based on sensor pattern noise [17-18] or sensor imperfections [19].
Describing devices fingerprinting for devices attribution

- Scanner attribution by Khanna et al [18]
  - Solution based on texture of pattern noise in the row and column direction.
  - Technique divided in two steps:
    - Reference Pattern construction
    - Noise Correlation
Describing devices fingerprinting for devices attribution

- Scanner attribution by Khanna et al [18]
  - Reference Pattern Construction
    - For each Printer, a series of scanned documents are used to construct a reference Pattern that will identify the scanner.
      - Firstly, every image $I_k$ has its noise extracted by subtracting it from the denoised image.
      - The Filter used is the Wiener Filter.
Describing devices fingerprinting for devices attribution

- Scanner attribution by Khanna et al [18]
  - Reference Pattern Construction

\[ I_{noise}^k = I^k - I_{denoised}^k \]
Describing devices fingerprinting for devices attribution

- Scanner attribution by Khanna et al [18]
  - Reference Pattern Construction
    - The 2D reference Pattern is the mean of each noise from K images.

\[
\tilde{I}_{noise}^{array}(i,j) = \frac{1}{K} \sum_{k=1}^{K} I_{noise}^{k}(i,j) \quad (7)
\]

86
Describing devices fingerprinting for devices attribution

Figure 22: Reference pattern extracted from scanned images proposed by Khanna et al [18].
Describing devices fingerprinting for devices attribution

- Scanner attribution by Khanna et al [18]
  - Noise Correlation
    - To identify the source scanner of a given document, its noise is compared to each reference pattern of x-th printer by correlation.
    - Higher correlation wins!

\[
C(I_{\text{noise}}^{\text{suspect}}, I_{\text{noise}}^{\text{array}}) = \frac{(I_{\text{noise}}^{\text{suspect}} - I_{\text{noise}}^{\text{suspect}}) \cdot (I_{\text{noise}}^{\text{array}} - I_{\text{noise}}^{\text{array}})}{||I_{\text{noise}}^{\text{suspect}}|| - ||I_{\text{noise}}^{\text{suspect}}|| \cdot ||I_{\text{noise}}^{\text{array}} - I_{\text{noise}}^{\text{array}}||} \quad (8)
\]
Figure 23: Correlation based scanner attribution by Khanna et al [18].
Describing devices fingerprinting for devices attribution

- Scanner attribution by Khanna et al [18]
  - Another extension to this work is assuming the mean of lines and columns of noise as two vectors.

\[
\tilde{I}_{\text{noise}}^l(1, i) = \frac{1}{N} \sum_{j=1}^{N} \tilde{I}_{\text{noise}}^\text{array}(i, j) \quad (9)
\]

\[
\tilde{I}_{\text{noise}}^c(1, j) = \frac{1}{M} \sum_{i=1}^{M} \tilde{I}_{\text{noise}}^\text{array}(i, j) \quad (10)
\]
Describing devices fingerprinting for devices attribution

Scanner attribution by Khanna et al [18]

From these vector, a set of features are extracted to identify the source scanner:

1. Mean of $\tilde{I}_l^{\text{noise}}$ and $\tilde{I}_c^{\text{noise}}$
2. Correlation between each line of $\tilde{I}_l^{\text{noise}}$ and $\tilde{I}_s^{\text{suspected}}$
3. Correlation between each column of $\tilde{I}_c^{\text{noise}}$ and $\tilde{I}_s^{\text{suspected}}$
4. Statistics over each vector.
Describing devices fingerprinting for devices attribution

- Scanner attribution by Gou et al [19]
  - Scanning noise taken from multiple perspectives:
    - image denoising
    - wavelet analysis,
    - neighborhood prediction
  - Statistical noise features is taken from each of them.
Describing devices fingerprinting for devices attribution

Scanner attribution by Gou et al [19]

1. Image denoising:
   - The noise is extracted from the image by subtracting the image denoised from the scanned image.
   - The noise suffers $\log_2$ image transform.
   - Two statistics are taken from this transform image: the mean and the standard deviation.
Describing devices fingerprinting for devices attribution

Figure 24: Image denoise approach for scanner attribution by Gou et al [19].
2. Wavelet Analysis

- This analysis is done in a normalized version of the image.
- Three high-frequency subbands HH, HL, LH from wavelet decomposition are used for analysis. Some statistical measures are then calculated:
  - the mean of coefficients.
  - standard deviation of coefficients.
  - goodness of Gaussian fitting.
Figure 25: Wavelet statistical analysis approach for scanner attribution by Gou et al [19].
3. Neighborhood prediction
   - The scanned image is first normalized
   - Smooth areas are found with gradient and intensity thresholds.
   - At each region, its center pixel value is predicted using a linear model on its eight neighbors.
   - Absolute prediction errors are calculated
   - Mean and standard deviation as features
3. Neighborhood prediction

- Absolute prediction errors are calculated
- Mean and standard deviation as features
- This process is done in each color channel
- A total of $30 + 18 + 12 = 60$ statistical is extracted by the whole technique.
Figure 26: Neighborhood prediction approach for scanner attribution by Gou et al [19].
Describing devices fingerprinting for devices attribution

- Scanner Attribution by Dirik et al [20]
  - Uses traces of dust, dirt, and scratches over scanner platen on scanned images.
  - Technique divided in two steps:
    - Dust and scratch reference construction per scanner
    - Source scanner identification
Dust and scratch reference construction

- scanned images are first filtered with a high pass filter.
- A model is searched on the filtered image through normalized cross correlation (NCC).
- Regions that yield high NCC are deemed to be dust and scratch locations.
Figure 27: Dust/scratch model for high pass filtered scanned image. Extracted from [20].
Describing devices fingerprinting for devices attribution

1. Dust and scratch reference construction
   - Two different scans of completely black background are sufficient.
   - Black scans: just dust and scratch positions are detected.
   - Dust and scratches in two images are not aligned due to the vertical and horizontal scanner head position shifts.
1. Dust and scratch reference construction
   - Matching of images through cross correlation.
   - Scanner dust and scratch reference: Hadamard product of the scanned images correctly aligned.
Describing devices fingerprinting for devices attribution

2. Source scanner identification
   ▪ Given a suspected document, its scratches and dust are detected.
   ▪ They are then correlated with scanners templates as shown before.
   ▪ High correlation wins!
Figure 28: Cross-correlation results for scanner identification. (Left) the image dust/scratch positions are matched with the scanner dust/scratch template. (Right) there is no matching. Extracted from [20].
Proposed techniques so far are vulnerable to attacks [20]:
- Signature removal.
- Signature replacement (spoofing).
Printers

- OCR + Another Printer: remove any extrinsic or intrinsic signature of laser printer in a text document.
- Printed halftone images could be scanned, converted to continuous tone, and then reprinted after using standard watermark attacks to remove any embedded watermark.
Scanners

- A process called flatfielding can replace the sensor noise.

- This operation includes the subtraction of light and dark patterns from the image.

- Then, these same patterns can be replaced by others from another sensor.
Conclusion

- Device Attribution is a hot research field in forensics
  - Criminal investigations.
  - Documents Authentication.

- New device technologies are coming! how to deal with this?

- How to deal with anti-forensics?
Conclusion

- New devices (and challenges) are coming!

Figure 29: Which 3D printer printed this gun? extracted from [21].
Conclusion

- New devices (and challenges) are coming!

Figure 30: Which 3D printer printed this mask? extracted from [22].
Conclusion

- New devices (and challenges) are coming!

Secretly Record Your Friends and Enemies With Handy Pocket Camera Drone
Because camera phones weren't enough.

BY MOLLY MULSHINE | 1/21 11:30AM

Do you ever feel frustrated that you can't keep tabs on your significant other at all times? Nervous that everyone is hanging out without you? Curious about what the heck your neighbors are doing over there?

Thankfully, there's a Pocket Drone currently being funded on Kickstarter that will solve these problems and more. It only takes 20 seconds to unpack and launch. Then, you

Figure 31: Which photos this drone took? Extracted from[23].
References


