Luca Debiasi\textsuperscript{1}, Elisabet Leitet\textsuperscript{2}, Kristin Norell\textsuperscript{3}, Theodoros Tachos\textsuperscript{2} and Andreas Uhl\textsuperscript{1}

\textsuperscript{1}Multimedia Signal Processing and Security Lab, University of Salzburg, Austria

\textsuperscript{2}Swedish National Forensic Centre, Swedish Police Authority, Linköping, Sweden

\textsuperscript{3}Formerly Swedish National Forensic Centre, Swedish Police Authority, Linköping, Sweden

May 3, 2019
1 Introduction and Motivation

2 Data Sets and Experimental Setup

3 Results and Discussion

4 Conclusion
Criminal Case Overview

- Criminal case investigated by the Swedish Police Authority
- Swedish National Forensic Centre (NFC) was consulted for victim identification in large collection of images
- Offender communicated with adolescents and exchanged images
- Goal: Finding potential victims in images
Motivation

- Typically perpetrators have an urge of documenting their criminal actions
- Ideal case for investigators: suspect’s camera is available
- If it is not, source based clustering of images could be useful to ...
  - ... identify the number of victims
  - ... find images taken of the same victim
  - ... evaluate number of perpetrators
- Due to the large amount of images manual processing is unfeasible
- Screening of the data is needed

→ NFC suggested a PRNU-based clustering approach on a per-source-basis to facilitate search for victims
PRNU-based Source Camera Clustering

- Screening of data is solved by partitioning the data set under investigation using a clustering algorithm.
- PRNU noise estimate is extracted for all images.
- How to measure quality of the clustering? → Cluster validity assessment.
- Criteria for optimal cluster solution: Compactness and Separation.
- Internal and external cluster validity indices (CVIs).
Data Sets

**Criminal case data set**: images extracted from the suspect’s computer.

<table>
<thead>
<tr>
<th>Image Size</th>
<th>Exam. Imgs.</th>
<th>Imgs. EXIF (%)</th>
<th># Cameras EXIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\geq 256 \times 256$</td>
<td>3078</td>
<td>2097 (~68%)</td>
<td>60</td>
</tr>
<tr>
<td>$\geq 512 \times 512$</td>
<td>1961</td>
<td>1006 (~51%)</td>
<td>47</td>
</tr>
<tr>
<td>$\geq 1024 \times 1024$</td>
<td>851</td>
<td>765 (~90%)</td>
<td>35</td>
</tr>
</tbody>
</table>

**Table**: Data set properties.

**Figure**: Distribution of image resolutions and ISO sensitivity.
Dresden Image Database [1]

- Built for development and benchmarking of camera-based digital forensic techniques
- 74 different digital cameras drawn from 25 different models
- Contains similar scenes captured with different camera models
- 16960 “natural” images in JPG format acquired with different parameters (focal length, ISO sensitivity, exposure, ...)
Experimental Setup

PRNU Extraction

- Different sized patches from the image center: $256 \times 256$, $512 \times 512$ and $1024 \times 1024$ pixels.
- Increased PRNU size leads to less images being available.

Source camera clustering techniques

- **Agglomerative clustering:**
  Blind Camera Fingerprinting and Image Clustering (BCF)[2]

- **Hierarchical clustering:**
  Unsupervised Clustering of Digital Images (UCDI)[3],
  Fast Image Clustering (FICL)[4]

- **Spectral clustering:**
  Multi-Class Spectral Clustering (MCSC) [5]
Experimental Setup (cont.)

Cluster validity indices

- **External**: Homogeneity (HOM), Completeness (COM), Adjusted Mutual Information (AMI), Adjusted Rand Index (ARI)
- **Internal**: Davies-Bouldin Index (DBI), Silhouette Index (SI), Calinski-Harabasz Index (CHI), Dunn Index (DI)

Experiment 1: Clustering of Dresden Image DB

- Evaluation of clustering algorithms and CVIs performance with known ground truth

Experiment 2: Criminal Case Dataset

- Interpretation of clustering outcome and CVI scores using previously gained insights
- Computation of external CVIs based on EXIF information
Figure: Number of obtained clusters for the Dresden Image DB.
Figure: External (a-b) and internal (c-d) CVI scores.
Figure: Number of obtained clusters for the criminal case dataset.
Figure: External EXIF-based (a-b) and internal (c-d) CVI scores.
Discussion and Recommendations

Challenges of criminal case data:

- Large image set with uneven distribution among origins
- Unknown acquisition conditions and image post-processings
- Varying acquisition parameters (e.g. ISO) and image resolutions
- Trade-off PRNU size vs. number of images available for investigation
- Internal CVIs very unreliable

Possible solutions:

- Almost no literature with solutions:
  - scaling/cropping [6][7], denoising/recompression/demosaicing [8]

Recommendations for similar scenarios:

- Use SI as cluster validation index (most reliable)
- Alternative: Use EXIF information in conjunction with AMI or ARI (adjusted for chance) → EXIF is easily manipulated/removed
- FICL recommended for screening purposes (high cluster number)
Examination of a large data set of still images from a suspect’s computer

Goal: Organise the images by their source camera(s) using PRNU-based clustering techniques

Evaluation of clustering outcome by means of different internal and external cluster validity indices (CVIs)

Clustering of ground truth data (Dresden Image DB) revealed problems with clustering techniques and CVIs

Clustering of criminal case data yielded uncertain results

We gave recommendations on the most reliable clustering techniques and CVIs in this scenario

Challenging data set raised many questions and issues for future work:
... How robust is PRNU when used for real world forensic data?
... How can clustering quality assessment be reliably performed?


Thank you for your attention!