Phase-Aware Projection Model for Steganalysis of JPEG Images

Vojtěch Holub and Jessica Fridrich
JPEG Steganography: Modify certain DCT coefficients of the image by \( \pm 1 \) to communicate the message.
JPEG Steganography: Modify certain DCT coefficients of the image by ±1 to communicate the message.

Steganalysis: Distinguish between cover and stego images by building a detector. If cover source is known and the steganographic scheme is not faulty, the best detection is achieved using feature-based steganalysis and machine learning.
Phase-Aware Projection Model for Steganalysis of JPEG Images
JPEG vs. spatial domain steganalysis

- **Spatial domain steganalysis:**
  - Analyzes dependencies among noise residuals.
  - Adjacent noise residuals put into a 4D co-occurrence \( \Rightarrow \) treated as a stationary signal.

- **JPEG domain steganalysis:**
  - Analyzes dependencies among quantized DCT coefficients.
  - DCT coefficients extracted from \( 8 \times 8 \) blocks, each mode uses a different DCT base and is quantized differently \( \Rightarrow \) non-stationarity.
JPEG vs. spatial domain steganalysis

- Spatial domain steganalysis:
  - Analyzes dependencies among noise residuals.
  - Adjacent noise residuals put into a 4D co-occurrence ⇒ treated as a stationary signal.

- JPEG domain steganalysis:
  - Analyzes dependencies among quantized DCT coefficients.
  - DCT coefficients extracted from 8 × 8 blocks, each mode uses a different DCT base and is quantized differently ⇒ non-stationarity.

Fact

After JPEG is decompressed into spatial domain, image pixels are non-stationary.
Which coefficients have the same statistics

JPEG (DCT) domain

8 x 8 block
Which coefficients have the same statistics

JPEG (DCT) domain

8 × 8 block
Which coefficients have the same statistics

JPEG (DCT) domain

8 × 8 block
Projection Spatial Rich Model [Holub, 2013]

- Originally designed for spatial domain steganalysis.
- First extract multiple residuals using 39 different linear and non-linear (min-max) filters.
- Residuals are convolved with normalized random projection kernels \( \Pi \in \mathbb{R}^{s_1 \times s_2}, s_1, s_2 \in \{1, \ldots, 8\} \).
- A histogram is built from the projection values for each residual and projection kernel.

The histogram is built from all projection values (all locations) \( \implies \) implicit assumption that residuals at all locations have identical statistics.

**Can it be improved?**
Phase-aware projections

Residual domain of decompressed JPEG

8 × 8 block
Phase-aware projections

Residual domain of decompressed JPEG

Histogram built from absolute values of projections – symmetries.
Results on linear and minmax residuals

- Detection error $E_{OOB}$
- J-UNIWARD at 0.4 bpnzac, BOSSbase 1.01, QF 75
- Two of PSRMQ3 submodels (linear and non-linear)

<table>
<thead>
<tr>
<th>Feature type</th>
<th>'spam14h' &amp; 'spam14v'</th>
<th>'minmax41'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>$\nu = 110$ dim 660</td>
<td>$\nu = 1000$ dim 6000</td>
</tr>
<tr>
<td>Phase-aware</td>
<td>0.2587</td>
<td>0.2034</td>
</tr>
<tr>
<td>Phase-aware symmetrized</td>
<td>0.2576</td>
<td>0.1536</td>
</tr>
<tr>
<td>Phase-aware symmetrized</td>
<td>0.2292</td>
<td>0.1582</td>
</tr>
</tbody>
</table>
PHARM features

- Merger of 7 SPAM residuals (7 linear filters)

\[
\begin{pmatrix}
-1 & 1 \\
1 & -3 & 3 & -1
\end{pmatrix}
\begin{pmatrix}
1 \\
-3 & 3 \\
-1
\end{pmatrix}
\begin{pmatrix}
1 & 1 \\
-1 & -1
\end{pmatrix}
\begin{pmatrix}
1 & 1 \\
-1 & 1
\end{pmatrix}
\begin{pmatrix}
1 & -1 \\
-1 & 1
\end{pmatrix}
\]

- These 7 filters were obtained by a forward feature-selection algorithm using the $\overline{E}_{OOB}$ estimate of the detection error from 25 prediction kernels.

- All PHARM parameters were optimized with respect to detection of J-UNIWARD
  
  - $\nu$ - number of random projections per residual
  - $s$ - maximal size of the random projection matrix
  - $T$ - number of histogram bins
  - $q$ - quantization (width of histogram bins) – depends on JPEG quality factor
PHARM in numbers

- Total dimensionality: $7 \cdot T \cdot \nu = 7 \cdot 2 \cdot 900 = 12,600$ (dimensionality of PSRMQ3 is 12,870)
- Quantization $q = \frac{65}{4} - \frac{3}{20} QF$ (QF 75: $q = 5$, QF 95: $q = 2$)
- Extraction time of $512 \times 512$ grayscale image, Intel i7 2 GHz laptop:

<table>
<thead>
<tr>
<th>Feature set</th>
<th>PHARM</th>
<th>DCTR</th>
<th>JRM</th>
<th>SRMQ1</th>
<th>PSRMQ3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensionality</td>
<td>12,600</td>
<td>8,000</td>
<td>22,510</td>
<td>12,753</td>
<td>12,870</td>
</tr>
<tr>
<td>Extraction time (s)</td>
<td>4.2</td>
<td>0.6</td>
<td>4.5</td>
<td>1.3</td>
<td>640</td>
</tr>
</tbody>
</table>
PHARM vs. JPEG steganography

J-UNIWARD [Holub, 2013]

Phase-Aware Projection Model for Steganalysis of JPEG Images
PHARM vs. JPEG steganography

UED [Guo, 2014]

Phase-Aware Projection Model for Steganalysis of JPEG Images
PHARM vs. JPEG steganography

nsF5 [Westfeld, 2001, Fridrich, 2007]

Phase-Aware Projection Model for Steganalysis of JPEG Images
PHARM vs. JPEG steganography

SI-UNIWARD [Holub, 2013]

Phase-Aware Projection Model for Steganalysis of JPEG Images
Conclusion

Currently, the most reliable detection of modern JPEG stego schemes (J-UNIWARD, UED) is achieved by spatial domain steganalysis (PSRMQ3) – counterintuitive.

Utilizing the knowledge of properties of decompressed JPEGs can further improve the detection.

General approach using ‘phase-aware’ features is proposed.


PHARM achieves superior detection of J-UNIWARD and UED with greatly reduced computational complexity over PSRM.

Source code in Matlab and C++/MEX available at http://dde.binghamton.edu/download/feature_extractors/