General-Purpose Image Forensics Using Patch Likelihood under Image Statistical Models

The 7th IEEE International Workshop on Information Forensics and Security

Wei Fan, Kai Wang, and François Cayre

GIPSA-lab, Grenoble, France

18-11-2015
Detecting Image Operations

Has it been previously processed by a certain image operation?

1. Generality
   - Targeted
   - General-purpose

2. Size
   - whole image
   - small image block
Analysis of Current Image Forensics

- Targeted Forensics (*well studied*)
  - Exploit particular artifacts of *specific* image operation
  - Different features for different image operations

- General-Purpose Forensics (*little studied*)
  - Cope with *multiple* image operations
  - Possible to adopt powerful steganalytical features, e.g., SPAM

- Forensic classification on small image blocks
  - Important for revealing *forgery semantics*
  - Image block size ↓ *usually* leads to forensic performance ↓

- X. Qiu *et al.*, “A universal image forensic strategy based on steganalytic model”. In: *Proc. ACM IHMMSec*, 2014, pp. 165-170
Analysis of Current Image Forensics

- Targeted Forensics (*well studied*)
  - Exploit particular artifacts of *specific* image operation
  - Different features for different image operations

- Most current forensic methods are targeted, and few results are reported on small image blocks

1. **Generality**
2. **Classification on small blocks**

- Important for revealing *forgery semantics*
- Image block size ↓ usually leads to forensic performance ↓

---

- X. Qiu *et al.*, “A universal image forensic strategy based on steganalytic model”. In: *Proc. ACM IHMMSec*, 2014, pp. 165-170
Motivation

Question
Given an image block, is it more like a natural, original block or a processed one?

Proposed Solution
Compare the average patch likelihood values calculated under different natural image statistical models.

Gaussian Mixture Model (GMM)

\[ L(\theta|x) = p(x|\theta) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, C_k) \]

D. Zoran and Y. Weiss, “From learning models of natural image patches to whole image restoration”. In: Proc. ICCV. 2011, pp. 479-486
Eigenvectors of GMM Covariance Matrices

\[
\begin{align*}
\pi_1 &= 0.0794 \\
\pi_2 &= 0.0435 \\
\pi_3 &= 0.0421 \\
\pi_4 &= 0.0285
\end{align*}
\]

\[
\begin{align*}
\pi_1 &= 0.0926 \\
\pi_2 &= 0.0358 \\
\pi_3 &= 0.0299 \\
\pi_4 &= 0.0278
\end{align*}
\]

\[
\begin{align*}
\pi_1 &= 0.0267 \\
\pi_2 &= 0.0266 \\
\pi_3 &= 0.0265 \\
\pi_4 &= 0.0263
\end{align*}
\]

D. Zoran and Y. Weiss, “Natural images, Gaussian mixtures and dead leaves”. In: *Proc. NIPS*. 2012, pp. 1736-1744
### Eigenvectors of GMM Covariance Matrices

<table>
<thead>
<tr>
<th>Method</th>
<th>$\pi_1$</th>
<th>$\pi_2$</th>
<th>$\pi_3$</th>
<th>$\pi_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORI</td>
<td>0.0794</td>
<td>0.0435</td>
<td>0.0421</td>
<td>0.0285</td>
</tr>
<tr>
<td>JPG</td>
<td>0.0926</td>
<td>0.0358</td>
<td>0.0299</td>
<td>0.0278</td>
</tr>
<tr>
<td>USM</td>
<td>0.0267</td>
<td>0.0266</td>
<td>0.0265</td>
<td>0.0263</td>
</tr>
</tbody>
</table>

D. Zoran and Y. Weiss, “Natural images, Gaussian mixtures and dead leaves”. In: *Proc. NIPS*. 2012, pp. 1736-1744
Hypothesis Testing

Test

\[
\Lambda(X) = \frac{1}{N} \sum_{i=1}^{N} \log L(\theta_0|x_i) - \frac{1}{N} \sum_{i=1}^{N} \log L(\theta_1|x_i) \geq \eta
\]

- \(x_i\): overlapping patches extracted from image (block) \(X\)
- \(H_0\): \(X\) is original, unprocessed
  \(GMM\) parametrized by \(\theta_0\)
- \(H_1\): \(X\) is processed by a certain image operation
  \(GMM\) parametrized by \(\theta_1\)

Decision Rule

\[
\begin{cases} 
\text{reject } H_0 & \text{if } \Lambda(X) \leq \eta \\
\text{do not reject } H_0 & \text{if } \Lambda(X) > \eta 
\end{cases}
\]
# Image Operations

<table>
<thead>
<tr>
<th>ORI</th>
<th>no image processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>GF</td>
<td><em>Gaussian filtering</em> with window size $3 \times 3$, and standard deviation $0.5$ to generate the filter kernel</td>
</tr>
<tr>
<td>JPG</td>
<td><em>JPEG compression</em> with quality factor $90$</td>
</tr>
<tr>
<td>MF</td>
<td><em>median filtering</em> with window size $3 \times 3$</td>
</tr>
<tr>
<td>RS</td>
<td><em>resampling</em> with bicubic interpolation to scale the image to $80%$ of its original size</td>
</tr>
<tr>
<td>USM</td>
<td><em>unsharp masking</em> with window size $3 \times 3$, and parameter $0.5$ for the Laplacian filter to generate the sharpening filter kernel</td>
</tr>
<tr>
<td>WGN</td>
<td><em>white Gaussian noise addition</em> with standard deviation $2$</td>
</tr>
</tbody>
</table>

- 6 image operations, each of which is with one fixed parameter setting
Image Datasets

1. GFTR: 2457 images of size $512 \times 512$ for **training**
   - SPAM (686-dimensional), 2457 samples (whole image or block)
   - GMM (200 components), $\sim 1.2$ million extracted $8 \times 8$ patches

2. GFTE: 2448 images of size $512 \times 512$ for **testing**
   - whole image ($512 \times 512$), 2448 samples for each image operation
   - image block ($32 \times 32$, $16 \times 16$), $2448 \times 10$ samples for each image operation

---

- ftp://firewall.teleco.uvigo.es:27244/DS_01_UTFI.zip
- ftp://lesc.dinfo.unifi.it/pub/Public/JPEGloc/dataset/
## Experimental Results

<table>
<thead>
<tr>
<th></th>
<th>GF</th>
<th>JPG</th>
<th>MF</th>
<th>RS</th>
<th>USM</th>
<th>WGN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>512 × 512</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPAM-based</td>
<td>99.86</td>
<td>98.20</td>
<td>99.94</td>
<td>96.45</td>
<td>99.73</td>
<td>98.53</td>
</tr>
<tr>
<td>Proposed-S</td>
<td>99.10</td>
<td>97.28</td>
<td>95.69</td>
<td>92.61</td>
<td>99.73</td>
<td>99.45</td>
</tr>
<tr>
<td>Proposed-T</td>
<td>99.82</td>
<td>99.49</td>
<td>99.31</td>
<td>92.67</td>
<td>99.73</td>
<td>99.80</td>
</tr>
<tr>
<td><strong>32 × 32</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPAM-based</td>
<td>99.35</td>
<td>94.18</td>
<td>99.43</td>
<td>89.23</td>
<td>98.76</td>
<td>95.04</td>
</tr>
<tr>
<td>Proposed-S</td>
<td>97.69</td>
<td>95.83</td>
<td>93.81</td>
<td>90.96</td>
<td>99.22</td>
<td>95.50</td>
</tr>
<tr>
<td>Proposed-T</td>
<td>97.73</td>
<td>96.04</td>
<td>93.99</td>
<td>90.96</td>
<td>99.21</td>
<td>97.55</td>
</tr>
<tr>
<td><strong>16 × 16</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPAM-based</td>
<td>98.38</td>
<td>88.00</td>
<td>99.26</td>
<td>78.21</td>
<td>97.82</td>
<td>91.20</td>
</tr>
<tr>
<td>Proposed-S</td>
<td>97.27</td>
<td>94.27</td>
<td>92.88</td>
<td>89.70</td>
<td>98.59</td>
<td>95.58</td>
</tr>
<tr>
<td>Proposed-T</td>
<td>97.37</td>
<td>94.68</td>
<td>93.01</td>
<td>89.72</td>
<td>98.59</td>
<td>95.66</td>
</tr>
</tbody>
</table>

# Experimental Results

Simple threshold: $\eta = 0$

<table>
<thead>
<tr>
<th></th>
<th>GF</th>
<th>JPG</th>
<th>MF</th>
<th>RS</th>
<th>USM</th>
<th>WGN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>512 × 512</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPAM-based</td>
<td>99.86</td>
<td>98.20</td>
<td>99.94</td>
<td>96.45</td>
<td>99.73</td>
<td>98.53</td>
</tr>
<tr>
<td>Proposed-S</td>
<td>99.10</td>
<td>97.28</td>
<td>95.69</td>
<td>92.61</td>
<td>99.73</td>
<td>99.45</td>
</tr>
<tr>
<td>Proposed-T</td>
<td>99.82</td>
<td>99.49</td>
<td>99.31</td>
<td>92.67</td>
<td>99.73</td>
<td>99.80</td>
</tr>
<tr>
<td><strong>32 × 32</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPAM-based</td>
<td>99.35</td>
<td>94.18</td>
<td>99.43</td>
<td>89.23</td>
<td>98.76</td>
<td>95.04</td>
</tr>
<tr>
<td>Proposed-S</td>
<td>97.69</td>
<td>95.83</td>
<td>93.81</td>
<td>90.96</td>
<td>99.22</td>
<td>95.50</td>
</tr>
<tr>
<td>Proposed-T</td>
<td>97.73</td>
<td>96.04</td>
<td>93.99</td>
<td>90.96</td>
<td>99.21</td>
<td>97.55</td>
</tr>
<tr>
<td><strong>16 × 16</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPAM-based</td>
<td>98.38</td>
<td>88.00</td>
<td>99.26</td>
<td>78.21</td>
<td>97.82</td>
<td>91.20</td>
</tr>
<tr>
<td>Proposed-S</td>
<td>97.27</td>
<td>94.27</td>
<td>92.88</td>
<td>89.70</td>
<td>98.59</td>
<td>95.58</td>
</tr>
<tr>
<td>Proposed-T</td>
<td>97.37</td>
<td>94.68</td>
<td>93.01</td>
<td>89.72</td>
<td>98.59</td>
<td>95.66</td>
</tr>
</tbody>
</table>

Trained threshold $\eta$ on GFTR dataset

## Experimental Results

<table>
<thead>
<tr>
<th></th>
<th>GF</th>
<th>JPG</th>
<th>MF</th>
<th>RS</th>
<th>USM</th>
<th>WGN</th>
</tr>
</thead>
<tbody>
<tr>
<td>512 × 512</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPAM-based</td>
<td>99.86</td>
<td>98.20</td>
<td>99.94</td>
<td>96.45</td>
<td>99.73</td>
<td>98.53</td>
</tr>
<tr>
<td>Proposed-S</td>
<td>99.10</td>
<td>97.28</td>
<td>95.69</td>
<td>92.61</td>
<td>99.73</td>
<td>99.45</td>
</tr>
<tr>
<td>Proposed-T</td>
<td>99.82</td>
<td>99.49</td>
<td>99.31</td>
<td>92.67</td>
<td>99.73</td>
<td>99.80</td>
</tr>
<tr>
<td>32 × 32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPAM-based</td>
<td>99.35</td>
<td>94.18</td>
<td>99.43</td>
<td>89.23</td>
<td>98.76</td>
<td>95.04</td>
</tr>
<tr>
<td>Proposed-S</td>
<td>97.69</td>
<td>95.83</td>
<td>93.81</td>
<td>90.96</td>
<td>99.22</td>
<td>95.50</td>
</tr>
<tr>
<td>Proposed-T</td>
<td>97.73</td>
<td>96.04</td>
<td>93.99</td>
<td>90.96</td>
<td>99.21</td>
<td>97.55</td>
</tr>
<tr>
<td>16 × 16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPAM-based</td>
<td>98.38</td>
<td>88.00</td>
<td>99.26</td>
<td>78.21</td>
<td>97.82</td>
<td>91.20</td>
</tr>
<tr>
<td>Proposed-S</td>
<td>97.27</td>
<td>94.27</td>
<td>92.88</td>
<td>89.70</td>
<td>98.59</td>
<td>95.58</td>
</tr>
<tr>
<td>Proposed-T</td>
<td>97.37</td>
<td>94.68</td>
<td>93.01</td>
<td>89.72</td>
<td>98.59</td>
<td>95.66</td>
</tr>
</tbody>
</table>

- At least comparable to the SPAM feature
- Especially advantageous on small blocks

---

Fine-Grained Image Tampering Localization

Fine-Grained Image Tampering Localization

Fine-Grained Image Tampering Localization

Fine-Grained Image Tampering Localization

ORI

Forgery (with RS)

SPAM-based

Proposed

Conclusions

1. A general-purpose framework for image forensics

- Comparison of average patch likelihood values calculated under different image models
- At least comparable performance compared with the SPAM feature
- Conceptually simplicity, no handcrafted feature extraction, and easiness to be extended

Perspectives

- Multi-class classification
- More image operations with more parameters
- Richer natural image statistical models
Thank you for your attention!

Q & A