Evaluation of Random Field Models in Multi-modal Unsupervised Tampering Localization

Paweł Korus & Jiwu Huang

Shenzhen University
College of Information Engineering

IEEE International Workshop on Information Forensics and Security,
4 - 7 Dec 2016, Abu Dhabi
Motivation and Goals (I)

- Reliable forensics should analyze various traces.
- Decision fusion studied in detail for tampering detection\(^1\).
  - Fuzzy logic, Dempster-Shafer theory of evidence.
- In tampering localization - still an open problem\(^2\).
  - Naive pixel-wise application of (even complicated) combination rules.
  - The simplest rules (summation / product) actually yield good performance.
- This naive approach is clearly sub-optimal - example problem:
  - Scale discrepancy: e.g., CFA (8 \times 8) and PRNU (128 \times 128).

---

\(^1\)Fontani et al., TIFS 2013; Barni et al., ICASSP 2012
\(^2\)Ferrara et al., ICMEW 2015; Cozzolino et al. IAP 2013
Motivation and Goals (II)

- Solution: cross-reference results with actual objects.
- Exploiting image content in forensics (before):
  - Manual image segmentation\(^3\).
  - Guided image filtering\(^4\) (feature correlation / structure transfer).
- Problems:
  - How to do reliable image segmentation? How to do it automatically?
  - How to handle object removal?
- Goals of our study:
  - Consider a scenario with mismatched detectors (scale discrepancy).
  - Evaluate random field models with content-dependent potentials.
  - Verify operation for subtle object removal forgeries.
  - Compare standard grid-based and dense CRF models.

---

\(^3\) Barni et al., ISCS 2010; Chierchia et al., ICDSP 2011
\(^4\) Chierchia et al., ICASSP 2014
Individual Detectors

- State-of-the-art CFA detector (small blocks, fine shape)$^5$:
  - Exploits periodicity of resampling artifacts.
  - Compares prediction error of acquired vs. interpolated pixels.
  - GMM-based segmentation into tampered / pristine blocks.
  - Operates on small non-overlapping blocks (best performance for $8 \times 8$ px).

- Photo-response non-uniformity detector (large windows, coarse shape)$^6$:
  - Validates (locally) presence of a known noise signature.
  - Uses a correlation predictor to locally estimate the strength of the signature.
  - Requires relatively large sample (we used overlapping $64 \times 64$ px windows).
  - Tampering probability from Bayesian analysis.

- Both detectors set up to yield same-size tampering probability maps:
  - localization resolution of $8 \times 8$ px image blocks.

---


$^6$Chen et al., TIFS 2008; [https://github.com/pkorus/multiscale-prnu](https://github.com/pkorus/multiscale-prnu)
Standard Combination Rules (I)

- Naive pixel-wise combination rules ($\tau$ - threshold):
  - Sum fusion:
    \[
    t_i = \left( c_i^{(cfa)} + c_i^{(prnu)} \right) / 2 > \tau
    \]
  - Product fusion:
    \[
    t_i = \left( c_i^{(cfa)} c_i^{(prnu)} \right) \left( c_i^{(cfa)} c_i^{(prnu)} + \tilde{c}_i^{(cfa)} \tilde{c}_i^{(prnu)} \right)^{-1} > \tau
    \]
  - Disjunction fusion (two variants of heuristic cleaning):
    \[
    t_i = \left( c_i^{(cfa)} > \tau \right) \lor \left( c_i^{(prnu)} > \tau \right)
    \]
  - Empirical fusion: rule learned from data.

- Heuristic cleaning:
  - For fusion result: morphological opening (disk-shaped SE $15 \times 15$).
  - For individual detectors:
    - CFA - as above;
    - PRNU - disk-shaped SE $31 \times 31$ opening + $19 \times 19$ dilation
Standard Combination Rules (II)

- **(0,0)** sum fusion
- **(0,0)** product fusion
- **(0,0)** empirical fusion
Random Field Models

- Optimization of the following energy function:

\[
E(t) = \frac{1}{|D|} \sum_{d \in D} \sum_{i=1}^{N} \psi_{\tau}(c_i^{(d)} | t_i) + \sum_{i=1}^{N} \sum_{j \in \Xi_i} \phi_p(t_i, t_j)
\]

where:
- \(\psi_{\tau}\) is the unary potential (favors solutions close to observations);
- \(\phi_p\) is a pairwise interaction potential (favors the same decisions among neighbors).

- The pairwise potential has two components:
  - \(\beta_0\) - default interaction strength,
  - \(\beta_1\) - content-dependent interaction strength (based on color similarity).

- We consider two versions:
  - grid CRF - only nearest 8-connected neighborhood,
  - dense CRF - fully connected pairwise field (Gaussian).

- Solvers: graph cuts\(^7\) / iterative mean-field approximations\(^8\).

\(^7\)UGM Toolbox, [http://www.cs.ubc.ca/~schmidtm/Software/UGM.html](http://www.cs.ubc.ca/~schmidtm/Software/UGM.html)

\(^8\)Krähenbühl et al., NIPS 2011
Evaluation Scenario

- Evaluation of localization performance on realistic forgeries:
  - Challenging realistic data set crafted by hand in modern photo editors.
  - 120 images (3 cameras, 1920 × 1080 px uncompressed TIFFs).
  - An extended version is publicly available for research purposes\(^9\).

- Performance metrics: \(F_1\) score, ROC

\[^9\text{http://kt.agh.edu.pl/~korus/downloads/dataset-realistic-tampering/}\]
Evaluation Results

- False positive rate
- True positive rate
- Decision threshold $\tau$
- Average $F_1$ score
- Peak $F_1$ score (grid CRF fusion)
- Peak $F_1$ score (product fusion)
- Peak $F_1$ score (empirical pixel-wise fusion)

<table>
<thead>
<tr>
<th>style</th>
<th>detector</th>
<th>max $F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sum</td>
<td></td>
<td>0.57</td>
</tr>
<tr>
<td>product</td>
<td></td>
<td>0.61</td>
</tr>
<tr>
<td>disjunction (CC)</td>
<td></td>
<td>0.57</td>
</tr>
<tr>
<td>disjunction (IC)</td>
<td></td>
<td>0.60</td>
</tr>
<tr>
<td>empirical</td>
<td></td>
<td>0.61</td>
</tr>
<tr>
<td>grid CRF</td>
<td></td>
<td>0.69</td>
</tr>
<tr>
<td>dense CRF</td>
<td></td>
<td>0.68</td>
</tr>
<tr>
<td>CFA</td>
<td></td>
<td>0.44</td>
</tr>
<tr>
<td>PRNU</td>
<td></td>
<td>0.49</td>
</tr>
</tbody>
</table>
Example Localization Results (best F-score wise)

tampered image

CFA

PRNU

sum fusion

product fusion

empirical fusion

grid CRF

dense CRF

binary disjunction
Example Localization Results (best F-score wise)

tampered image

CFA

PRNU

sum fusion

product fusion

empirical fusion

grid CRF

dense CRF

binary disjunction
Example Localization Results (best F-score wise)

tampered image

CFA

PRNU

sum fusion

product fusion

empirical fusion

grid CRF

dense CRF

binary disjunction
Example Localization Results (best F-score wise)

tampered image  CFA  PRNU

sum fusion  product fusion  empirical fusion

grid CRF  dense CRF  binary disjunction
Example Localization Results (best F-score wise)

tampered image

CFA

PRNU

sum fusion

product fusion

empirical fusion

grid CRF

dense CRF

binary disjunction
Example Localization Results (best F-score wise)

tampered image

CFA

PRNU

sum fusion

product fusion

empirical fusion

grid CRF

dense CRF

binary disjunction
Example Localization Results (best F-score wise)

tampered image

CFA

PRNU

sum fusion

product fusion

empirical fusion

grid CRF

dense CRF

binary disjunction
Example Localization Results (best F-score wise)

tampered image

CFA

PRNU

sum fusion

product fusion

empirical fusion

grid CRF

dense CRF

binary disjunction
Object Insertion vs Object Removal

- Problems with existing evaluation metrics.
  - Oblivion to spatial relationships + collateral damage.
  - Better detection often leads to marginal improvement (bottom) or even deterioration (top) of measurable performance.

- Impact of content guidance on the MRF fusion (see below).

CRF guided by tampered image
\( (F_1 = 0.664 / A = 0.749) \)

CRF guided by original image
\( (F_1 = 0.806 / A = 0.838) \)

CRF with no content guidance
\( (F_1 = 0.812 / A = 0.842) \)

Pixel-wise product fusion
\( (F_1 = 0.844 / A = 0.866) \)

CRF guided by tampered image
\( (F_1 = 0.820 / A = 0.851) \)

CRF guided by original image
\( (F_1 = 0.804 / A = 0.844) \)

CRF with no content guidance
\( (F_1 = 0.805 / A = 0.840) \)

Pixel-wise product fusion
\( (F_1 = 0.799 / A = 0.841) \)
Conclusions & Future Work (I)

- Even naive fusion leads to significant improvement over individual detectors.
- Disjunction fusion has an advantage of customized post-processing.
- Product fusion is not an accurate combination model but it seems to be of limited importance in practice.
- **Scale discrepancy in multi-modal analysis** makes pixel-wise fusion sub-optimal.
- Fusion in localization:
  - not only combination rules matter.
  - need to cross-reference results with image content.
Adoption of neighborhood dependencies further improves performance.

**Content-dependent interactions** are an effective tool to exploit image content - no problems with object removal.

No significant quantitative differences between CRF models, but...

▶ ...existing evaluation protocols and metrics are imperfect,
▶ ...important qualitative differences are obvious.

Future work:

▶ Understanding of actual map utility for forensic analysis (humans in general?).
▶ Better evaluation metrics aware of spatial dependencies and collateral damage.
Limitations and Possible Extensions of the Framework

- Need to test more diverse detectors.
  - Some may work on even larger blocks (e.g., $256 \times 256$ px).
  - Some may not yield tampering probabilities.
  - Some may not work on square blocks at all (segments / super-pixels).

- The current framework does not support compatibility of traces.
  - Should be feasible with proper definition of unary potentials.

- Limited improvement for overconfident detectors.
  - Common for popular Bayesian formulations.
  - Valid traces present in tampered area.
  - To some extent, alleviated by truncated unary potentials.

- Training limitations:
  - Imperfect performance measures.
  - Parameter generalization in diverse conditions, e.g., for varying image size?
  - Is it possible to choose the parameters for each case individually?
Thank You

- Thank you for your attention.
- Discussion?

- Contact: pkorus@agh.edu.pl
- Web: http://kt.agh.edu.pl/~korus/
- Supplement:
  - more details + additional experiments,
- Dataset:
  - 220 images + 3-level GT
- Matlab code:
  - https://github.com/pkorus/multiscale-prnu