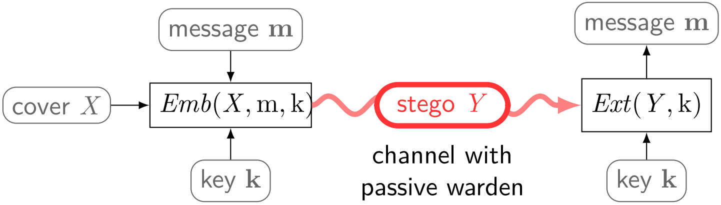
Random Projections of Residuals as an Alternative to Co-occurrences in Steganalysis

Vojtěch Holub, Jessica Fridrich, Tomáš Denemark

BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK

What is steganalysis?

- **Steganography is the art of secret communication**



- **Steganographer's job**
  Modify a cover image to stego image so that it contains a secret message (by flipping LSBs, changing DCT coefficients, ...).
  Goal: make the embedding changes statistically undetectable.
- **Warden's job**: Extract many different image statistics (steganalytic features) and use them in classification by machine learning in order to distinguish between cover and stego images.

History of feature-based steganalysis in spatial domain

- [SPIE 2000] Avcibas, Memon, Sankur: Image quality metrics
- [ICIP 2002] Farid, Lyu: Moments of wavelet coefficients
- [SPIE 2006] Goljan, Holotyak, Fridrich: Wavelet absolute moments (WAM)
- [ICME 2006] Zo, Shi, Su, Xuan: Markov TPM of noise residuals
- [ACM MMSec 2009] Pevný, Bas, Fridrich: 2nd order Markov TPM (SPAM)
- [IH 2011] HUGO BOSS competitors: 4D joint distributions (co-occurrences) of multiple higher-order noise residuals
- [IEEE TIFS 2011] Fridrich, Kodovský, Spatial Rich Model

Spatial Rich Model (SRM)

Given cover/stego image $\mathbf{X} = (x_{ij})$

- **Noise residual** $z_{ij} = x_{ij} - \mathrm{Pred}(\mathcal{N}(x_{ij}))$
  - $\mathrm{Pred}(\mathcal{N}(x_{ij}))\ldots$ pixel predictor on neighborhood $\mathcal{N}$
    - linear filters
    - min/max of several filter outputs
  - z_{ij} has narrower dynamic range
  - better SNR (stego noise to image content)
- **Quantize** $z_{ij} \to r_{ij} = Q_{\mathcal{Q}}(z_{ij})$, $\mathcal{Q} = \{-Tq, -(T-1)q, \ldots, Tq\}$
  - $T\ldots$ truncation threshold
  - $q\ldots$ quantization step
- **Co-occurrence** of 4 adjacent $r_{i,j}$'s = features

Examples of residuals

Linear filters: $\mathbf{Z} = \mathbf{K} * \mathbf{X} - \mathbf{X}$

- Ker–Böhme kernel $\mathbf{K} = \begin{pmatrix} -0.25 & 0.5 & -0.25 \\ 0.5 & 0 & 0.5 \\ -0.25 & 0.5 & -0.25 \end{pmatrix}$
- Local linear predictor $\mathbf{K} = (\frac{1}{2}, 0, \frac{1}{2})$
- Local quadratic predictor $\mathbf{K} = (\frac{1}{3}, 0, 1, -\frac{1}{3})$

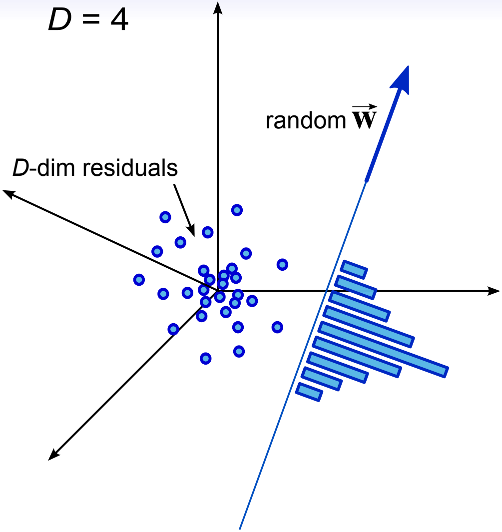
Non-linear filters:

- $\mathbf{Z}^{(1)} = \mathbf{K}^{(1)} * \mathbf{X} - \mathbf{X}$, $\mathbf{Z}^{(2)} = \mathbf{K}^{(2)} * \mathbf{X} - \mathbf{X}$
- $\mathbf{Z}^{(\min)} = \min\{\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}\}$
- $\mathbf{Z}^{(\max)} = \max\{\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}\}$

Limitations of co-occurrences

- Feature dimensionality, $(2T+1)^D$, grows quickly with co-occurrence order D and threshold T
- D and T need to be kept small to have the co-occurrence bins well populated
- Information in the marginals is not utilized
- Dependencies beyond D samples not captured

Alternative descriptor of residuals



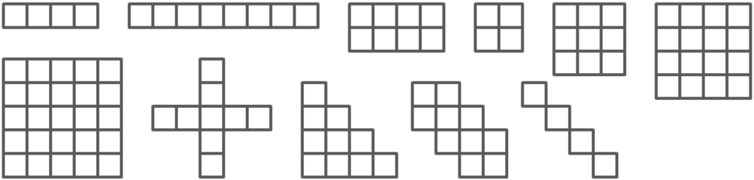
Advantages of projections

- Can capture long-range (and more complex) dependencies among pixels
- Diversification over projection neighborhoods
- Finer quantization (larger threshold T) $\implies$ more info extracted from tails
- Design flexibility by selecting the number of:
  - projection neighborhoods
  - projection vectors
  - quantization bins

Four pillars of the new model

- Projection neighborhoods
- Projection vectors
- Quantizer
- Symmetrization

Projection neighborhoods



Eleven types of projection neighborhoods $\mathcal{P}$ used in the PSRM:
1×4, 1×8, 2×4, 2×2, 3×3, 4×4,
5×5, cross, stairs, thick diagonal, and diagonal

Projection vectors

- For a given projection neighborhood $\mathcal{P}$
- $\mathbf{v} \in \{-2, -1, 0, 1, 2\}^{|\mathcal{P}|}$, generated pseudo-randomly
- $\mathbf{v}$ is mapped to $\mathcal{P}$ in some predefined order

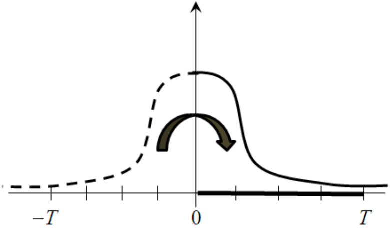
Example

- $\mathcal{P} = 2 \times 2$ square, $\mathbf{v} = (-1, 0, 1, -2)$
- $\mathbf{K}(\mathcal{P}, \mathbf{v}) = \begin{pmatrix} -1 & 0 \\ 1 & -2 \end{pmatrix}$
- $\mathcal{P}(\mathbf{v}) = \mathbf{K} * \mathbf{Z}$ set of all projections

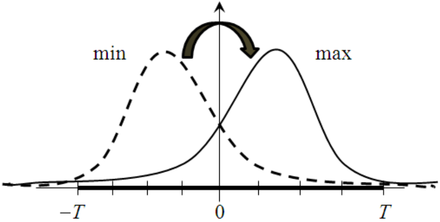
Quantizer

Designed to utilize symmetries of residuals

Linear residual
$\mathbf{Z} = \mathbf{K} * \mathbf{X} - \mathbf{X}$
$|\mathcal{P}(\mathbf{v})|$



Min-max residual
$\mathbf{Z}^{(\mathrm{min})} = \min\{\mathbf{Z}^{(\mathrm{h})}, \mathbf{Z}^{(\mathrm{v})}\}$
$\mathbf{Z}^{(\mathrm{max})} = \max\{\mathbf{Z}^{(\mathrm{h})}, \mathbf{Z}^{(\mathrm{v})}\}$
$-\mathcal{P}_{\mathrm{min}}(\mathbf{v}) \cup \mathcal{P}_{\mathrm{max}}(\mathbf{v})$



Symmetrization

- Symmetries of natural images: isotropy and non-directionality
- To obtain more robust statistics, for a given projection kernel $\mathbf{K}(\mathcal{P}, \mathbf{v})$, we combine projections obtained using
  - its mirror versions, $\overleftrightarrow{\mathbf{K}}$, $\mathbf{K}\updownarrow$
  - its transpose and rotation by $180°$, $\mathbf{K}^T$,$\mathbf{K}^{\circlearrowleft}$
  - always adds up to 8 symmetries

Example

- $\mathbf{K}(\mathcal{P}, \mathbf{v}) = \begin{pmatrix} -1 & 0 \\ 1 & -2 \end{pmatrix}$
- $\overleftrightarrow{\mathbf{K}}(\mathcal{P}, \mathbf{v}) = \begin{pmatrix} 0 & -1 \\ -2 & 1 \end{pmatrix}$ $\qquad \mathbf{K}\updownarrow(\mathcal{P}, \mathbf{v}) = \begin{pmatrix} 1 & -2 \\ -1 & 0 \end{pmatrix}$
- $\mathbf{K}^T(\mathcal{P}, \mathbf{v}) = \begin{pmatrix} -1 & 1 \\ 0 & -2 \end{pmatrix}$ $\qquad \mathbf{K}^{\circlearrowleft}(\mathcal{P}, \mathbf{v}) = \begin{pmatrix} -2 & 1 \\ 0 & -1 \end{pmatrix}$

Dimensionality

- $3 \le N_P \le 11 \dots$ number of projection neighborhoods
- $N_{\mathbf{v}} \dots$ number of projection vectors per neighborhood
- $N_{\mathbf{Z}} = 39 \dots$ number of residuals
- $T = 4 \dots$ number of quantization bins

$$2(T+1)N_{\mathbf{Z}}N_{\mathbf{v}}N_{\mathcal{P}}$$

Experimental setup

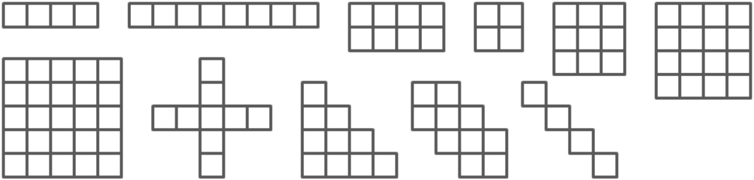
- **Database:** BOSSbase 1.01 with 10,000 512×512 grayscale images
- **Classifier:** Ensemble with Fisher linear discriminant base learners
  - thresholds set to minimize total average error under equal priors:

$$P_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \frac{1}{2}(P_{\mathrm{FA}} + P_{\mathrm{MD}})$$

- **Performance evaluation:**
  - $E_{\mathrm{OOB}} \ldots$ Out-Of-Bag (OOB) testing error estimate
- **Stego algorithms:**
  - HUGO (Pevný et al., IH 2010) with switch "--T 255" turned on.
  - WOW, Wavelet Obtained Weights (Holub et al., WIFS 2012)
  - Both content adaptive, use STCs to minimize embedding distortion

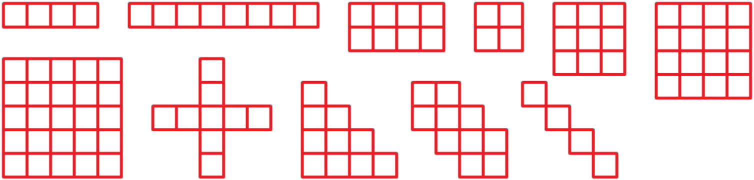
Neighborhood groups

Trade-off between number of neighborhoods and number of projection vectors per neighborhood for **fixed dimensionality**



Neighborhood groups

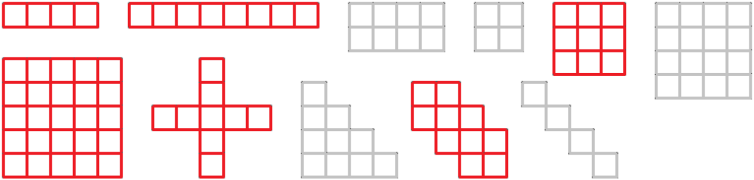
Trade-off between number of neighborhoods and number of projection vectors per neighborhood for **fixed dimensionality**



1. **All (11):** Maximal diversity across projection neighborhoods

Neighborhood groups

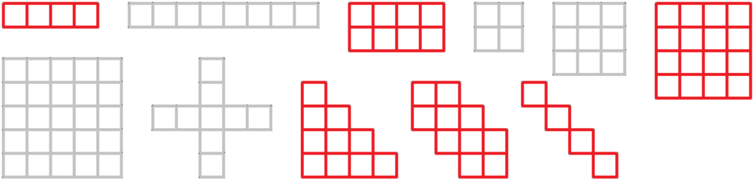
Trade-off between number of neighborhoods and number of projection vectors per neighborhood for **fixed dimensionality**



1. **All (11):** Maximal diversity across projection neighborhoods
2. **Diverse (6):** Good diversity for smaller number of neighborhoods

Neighborhood groups

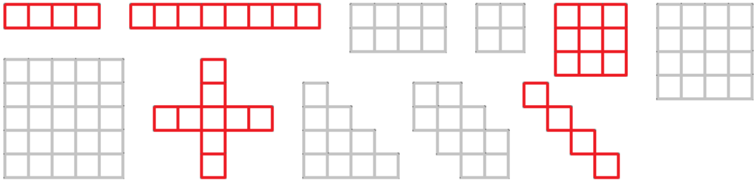
Trade-off between number of neighborhoods and number of projection vectors per neighborhood for **fixed dimensionality**



1. **All (11):** Maximal diversity across projection neighborhoods
2. **Diverse (6):** Good diversity for smaller number of neighborhoods
3. **Subsets of 4×4 (6):** Neighborhoods of size 4

Neighborhood groups

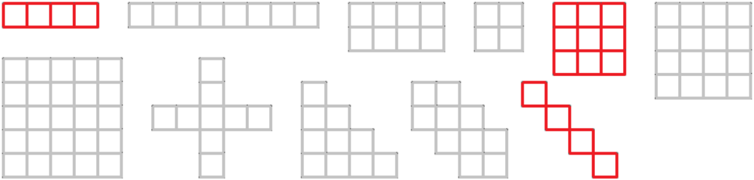
Trade-off between number of neighborhoods and number of projection vectors per neighborhood for **fixed dimensionality**



1. **All (11):** Maximal diversity across projection neighborhoods
2. **Diverse (6):** Good diversity for smaller number of neighborhoods
3. **Subsets of 4×4 (6):** Neighborhoods of size 4
4. **Sparse (5):** Sparse neighborhoods plus 3×3

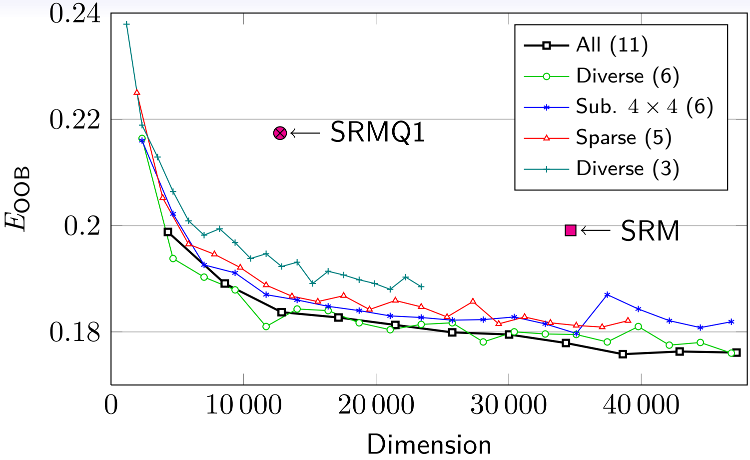
Neighborhood groups

Trade-off between number of neighborhoods and number of projection vectors per neighborhood for **fixed dimensionality**



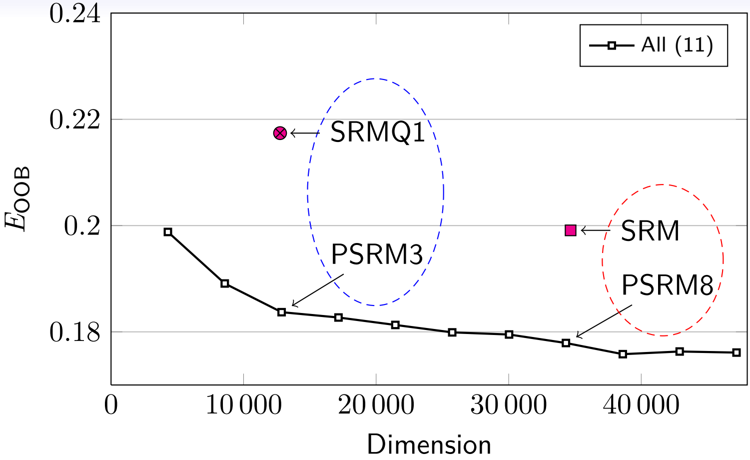
1. **All (11):** Maximal diversity across projection neighborhoods
2. **Diverse (6):** Good diversity for smaller number of neighborhoods
3. **Subsets of 4×4 (6):** Neighborhoods of size 4
4. **Sparse (5):** Sparse neighborhoods plus 3×3
5. **Diverse (3):** Maximum number of projection vectors

Results across neighborhood groups



Detection error E_{OOB} as a function of PSRM dimensionality for five combinations of projection neighborhoods. Tested on WOW 0.4 bpp.

Results across neighborhood groups



Detection error E_{OOB} as a function of PSRM dimensionality for five combinations of projection neighborhoods. Tested on WOW 0.4 bpp.

PSRM on WOW and HUGO

bpp (dim)	WOW SRM 34,671	WOW PSRM8 34,320	WOW SRMQ1 12,753	WOW PSRM3 12,870	HUGO SRM 34,671	HUGO PSRM8 34,320	HUGO SRMQ1 12,753	HUGO PSRM3 12,870
0.05	44.72	44.38	45.76	45.12	43.55	42.38	44.75	42.95
0.1	39.58	38.15	41.32	39.07	36.51	35.13	37.55	35.51
0.2	31.17	29.14	33.16	29.68	25.42	24.44	26.76	24.64
0.3	25.36	22.53	26.91	23.08	17.92	16.48	19.30	17.13
0.4	19.91	17.79	21.74	18.37	12.78	11.64	13.37	12.09
0.5	16.36	13.87	17.59	14.26	8.56	8.20	9.43	8.40

PSRM's average detection gain over all payloads

1.9 %	**2.8** %	**1.1** %	**1.7** %

Summary

- Proposed **PSRM**, Projection Spatial Rich Model
- Keeps the same residuals as SRM but represents them differently
- Instead of co-occurrences, we form first-order statistics of projections of residual groups on random directions
- Pros:
  - Diversification over projection neighborhoods boosts detection
  - Markedly better detection for highly content-adaptive steganography (WOW)
  - Accuracy of SRM reached with 5–7 times smaller dimension
  - Design flexibility (dimensionality vs. accuracy trade-off)
- Cons:
  - Computational complexity (for PSRM8 over 10,000 convolutions and histograms must be computed)
- Matlab and C++ extractors available at `http://dde.binghamton.edu/download/feature_extractors`