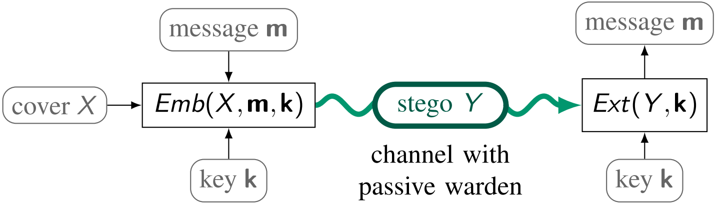
Selection-Channel-Aware Rich Model for Steganalysis of Digital Images

Tomáš Denemark, Vahid Sedighi, Rémi Cogranne, Vojtěch Holub, and Jessica Fridrich

Steganography and 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**: Distinguish between cover and stego images by building a detector. If cover source is known, the best detection is achieved using feature-based steganalysis and machine learning.

Steganography in practice

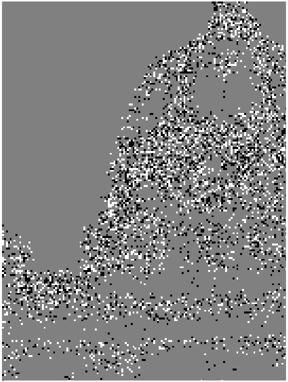
- **Sender**
  Specifies the cost of changing each pixel in the cover, $\rho_{ij} \geq 0$.
  Embeds the message by minimizing the distortion in the form of a sum of costs of all changed pixels, $\sum_{x_{ij} \neq y_{ij}} \rho_{ij}$.
  Problem is equivalent to source coding with a fidelity constraint.

  Can be implemented with syndrome-trellis codes that operate near the rate–distortion bound [Filler 2010].
- **Recepient**
  Extracts the secret message using the parity-check matrix of the shared syndrome-trellis code.

Content-adaptive steganography

- Embedding prefers changing pixels in textured / noisy areas

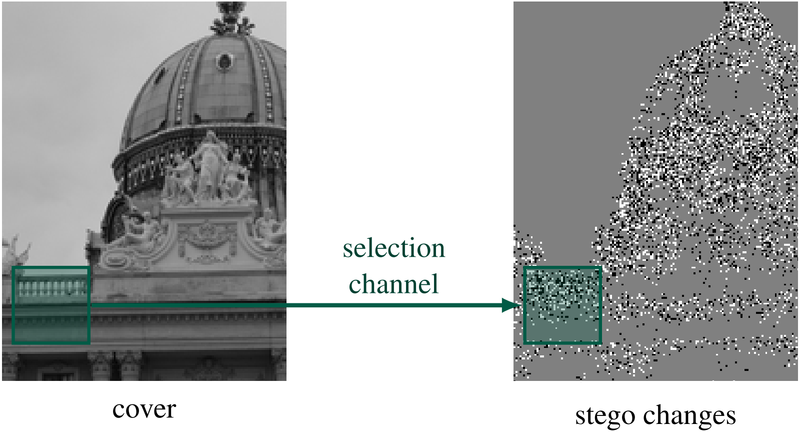
cover



stego changes

Content-adaptive steganography

- Embedding prefers changing pixels in textured / noisy areas



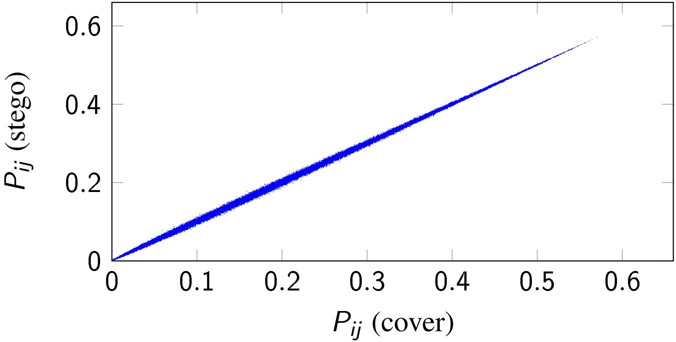
Selection channel

- Formally, the selection channel are the probabilities of changing pixel ij:

$$p_{ij} = \frac{e^{-\lambda \rho_{ij}}}{1 + e^{-\lambda \rho_{ij}}},$$

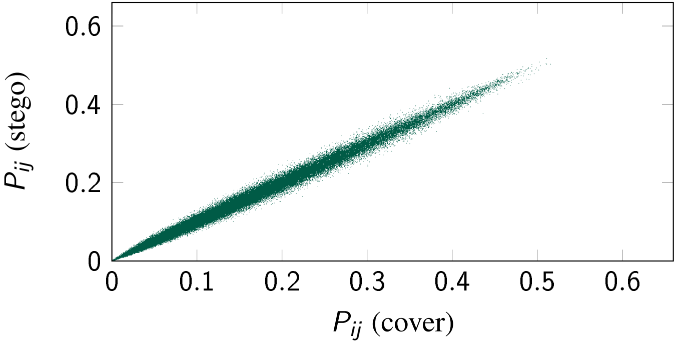
- $\lambda \geq 0$ parameter controlling the payload
  - ρ_{ij} pixel "costs" computed from cover image $\mathbf{x}$
  - costs dictated by content + noise
- Since stego changes are subtle: ρ_{ij} from cover $\approx \rho_{ij}$ from stego image

Selection channel recoverability, WOW



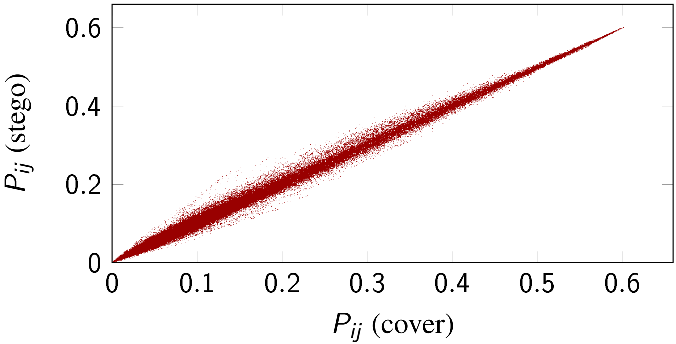
[Holub, IEEE WIFS 2012] Designing Steganographic Distortion Using Directional Filters

Selection channel recoverability, S-UNIWARD



[Holub, EURASIP 2014] Universal Distortion Function for Steganography in an Arbitrary Domain

Selection channel recoverability, HILL



[Li, ICIP 2014] A New Cost Function for Spatial Image Steganography

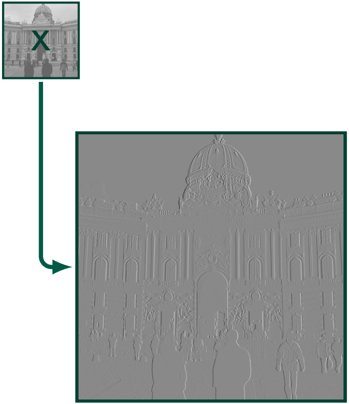
Using Selection Channel for Steganalysis

- [BOSS, IH 2011] no successful attack on HUGO based on approximate knowledge of the selection channel.
- [Schöttle et al., WIFS 2012] improved WS detector for naive content-adaptive LSB replacement.
- [Denemark, SPIE 2014] first successful attack on modern stego scheme that utilized an artifact in selection channel.
- [Tang, ACM IH & MMSec 2014] thresholded SRM – first general purpose attack using selection channel.
- [Denemark, WIFS 2014] maxSRMd2 (this presentation)

Spatial Rich Model (SRM)

cover $\mathbf{x}$

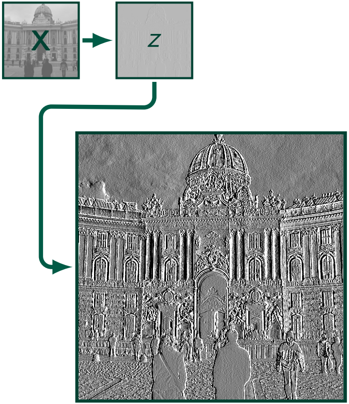
Spatial Rich Model (SRM)



noise residual z

- $z_{ij} = x_{i,j} - \mathrm{Pred}(\mathscr{N}(x_{ij}))$
- $\mathrm{Pred}(\mathscr{N}(x_{ij}))$ ... pixel predictor on neighborhood $\mathscr{N}$
- linear and min/max filters
- z_{ij} has narrower dynamic range
- better SNR (stego noise to image content)

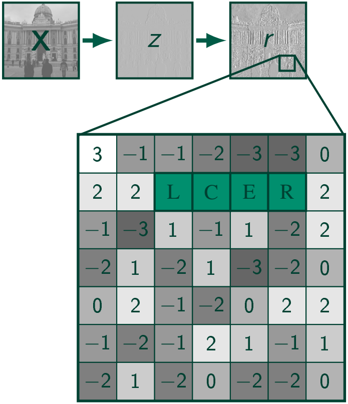
Spatial Rich Model (SRM)



quantized residual r

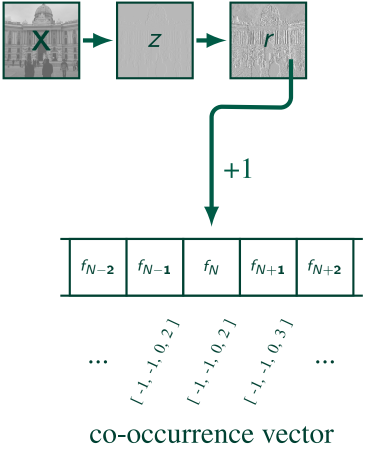
- $z_{ij} \to r_{ij} = Q_{\mathscr{Q}}(z_{ij})$
- $\mathscr{Q} = \{-Tq, -(T-1)q, \ldots, Tq\}$
- T ... truncation threshold
- q ... quantization step (SRM uses $q = 1, 1.5, 2$)

Spatial Rich Model (SRM)



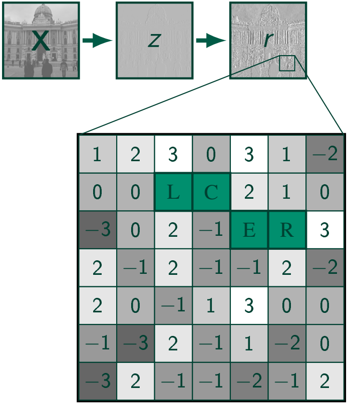
- collect quartets of values
- horizontal and vertical directions

Spatial Rich Model (SRM)



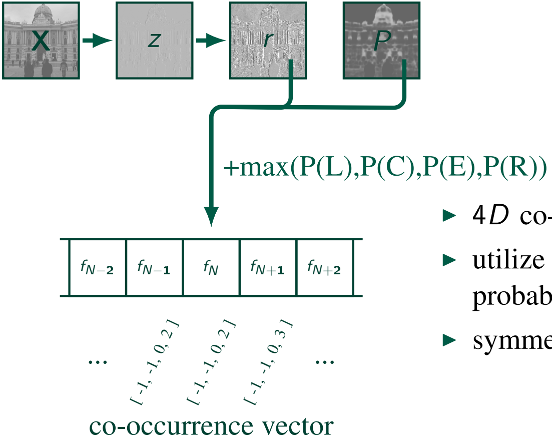
- $4D$ co-occurrence matrix
- symmetrization

Co-occurrences in maxSRMd2



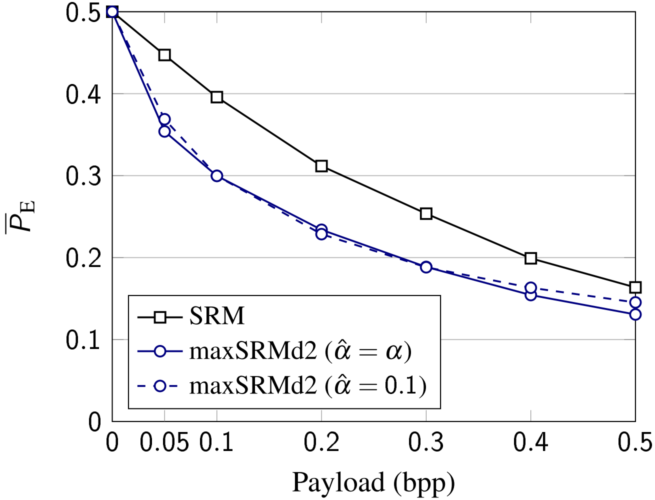
- collect quartets of values
- horizontal and vertical directions
- twice as many symmetries

Co-occurrences in maxSRMd2

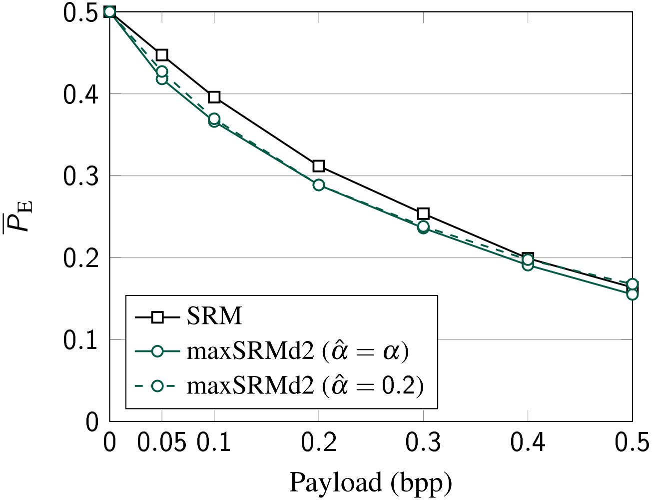


- $4D$ co-occurrence matrix
- utilize embedding probabilities
- symmetrization

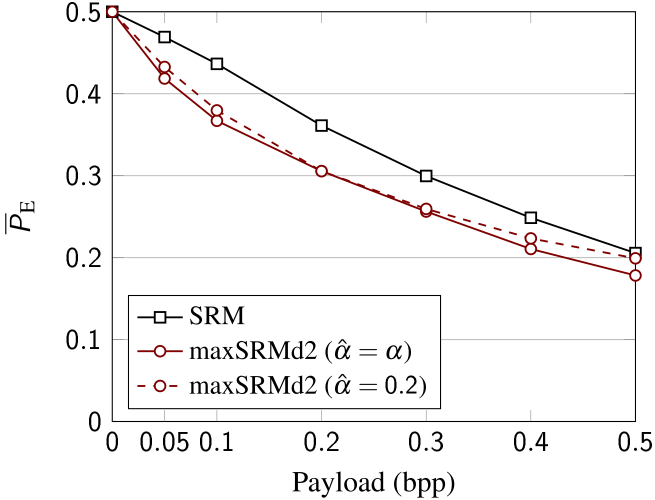
Detection gain w.r.t. SRM (WOW)



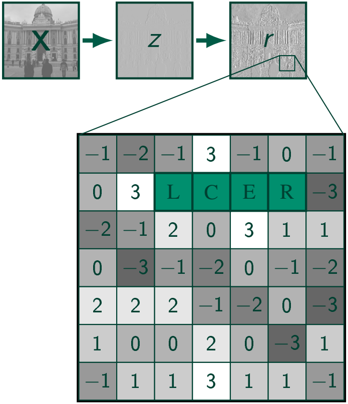
Detection gain w.r.t. SRM (S-UNIWARD)



Detection gain w.r.t. SRM (HILL)

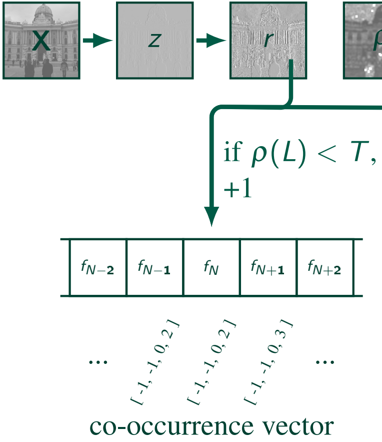


Co-occurrences in thresholded SRM (tSRM)



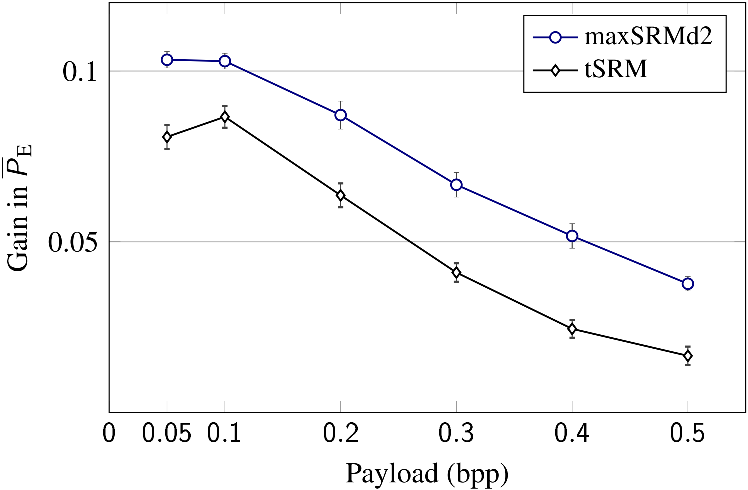
- collect quartets of values
- horizontal and vertical directions

Co-occurrences in thresholded SRM (tSRM)

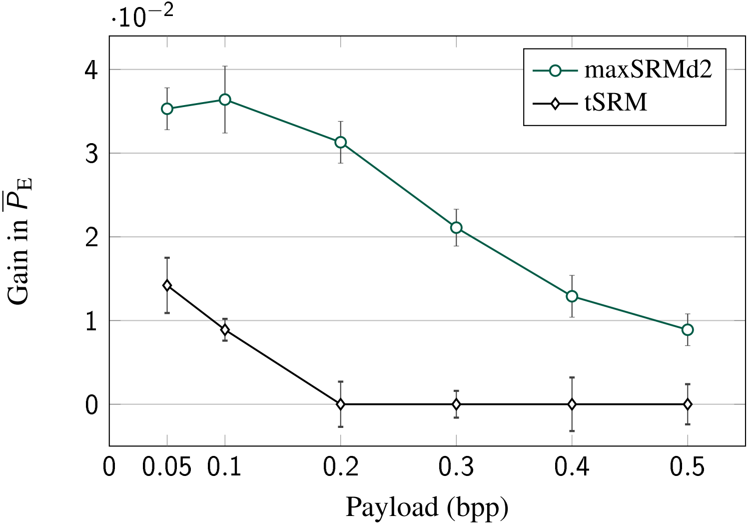


- $4D$ co-occurrence matrix
- utilize only some values
- symmetrization

Comparison between maxSRMd2 and tSRM (WOW)



Comparison between maxSRM and tSRM (S-UNIWARD)



Summary

- maxSRM is a general-purpose feature set capable of utilizing the selection channel for detection of content-adaptive steganography
- Overly content-adaptive embedding hurts security (WOW)
- When designing steganography, selection-channel attacks need to be considered
  - often, improvement w.r.t. SRM leads to bigger loss w.r.t. maxSRM
- Matlab code available from `http:\\dde.binghamton.edu\download`