Designing steganographic distortion using directional filters

Vojtěch Holub, Jessica Fridrich
Binghamton University, USA

Goal

Advance the state of the art in steganography for digital images in spatial domain.

Design

Embed messages by minimizing additive distortion in the form:

$$D(\mathbf{X},\mathbf{Y}) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \rho_{ij}(\mathbf{X}, Y_{ij})$$

ρ_{ij} is cost of changing pixel X_{ij} to Y_{ij}.

Pixel Costs ρ_{ij}

The main idea is to have small embedding costs in areas where the content is difficult to model in every direction (textures). Smooth regions and clean edges must have large embedding costs.

Coding

Embedding realized using ternary syndrome coding with syndrome-trellis codes (STC's) [1].

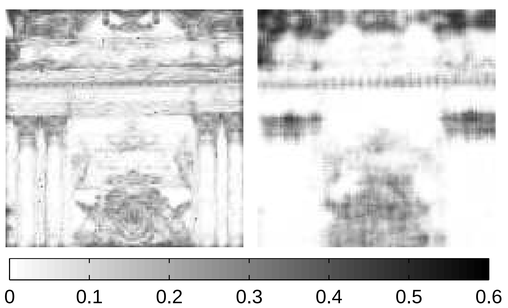
- operate near rate-distortion bound
- computationally efficient, scalable, simple

Relationship to prior art

HUGO [5] minimizes distortion in SPAM feature space; WOW minimizes distortion in a transform domain.

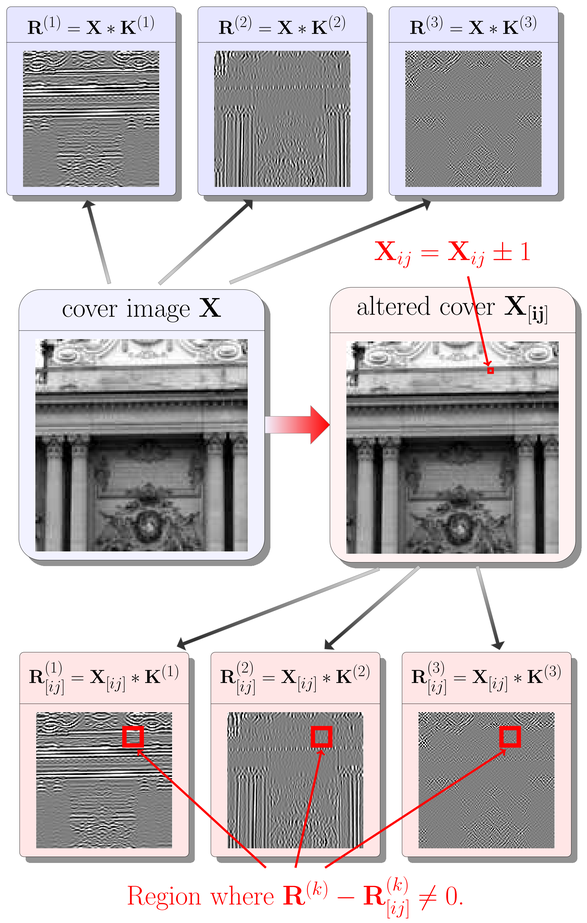
Embedding probabilities

HUGO WOW



Embedding Algorithm

Given cover image $\mathbf{X} \in \{0,\dots,255\}^{M\times N}$,



Region where $\mathbf{R}^{(k)} - \mathbf{R}^{(k)}_{[ij]} \neq 0$.

Assess cover content using a bank of directional filters $\mathbf{K} = \{K^{(1)},\dots,K^{(n)}\}$

Change pixel (ij).

Assess altered cover content in the same manner

Weigh embedding impact with cover coefficients,

$$\xi^{(k)}_{ij} = \sum_{(m,n)\in M\times N} \left|\mathbf{R}^{(k)}_{mn}\right| \cdot \left|\mathbf{R}^{(k)}_{mn} - \mathbf{R}^{(k)}_{[ij]mn}\right|$$

Compute embedding suitabilities

$$\rho_{ij} = \left(\sum_{k=1}^{n} \left|\xi^{(k)}_{ij}\right|^{p}\right)^{-\frac{1}{p}}$$

Aggregate suitabilities into costs using reciprocal Hölder norm

- larger $\xi^{(k)}_{ij} \Rightarrow$ smaller cost ρ_{ij}
- if $\exists k,\ \xi^{(k)}_{ij} = 0 \Rightarrow \rho_{ij} = \infty$
- $p = -1$ worked best $\Rightarrow \rho_{ij} = \sum_{k=1}^{n} \frac{1}{\left|\xi^{(k)}_{ij}\right|}$

Repeat for all pixels (i,j)

$\mathbf{Y} = \mathrm{STC}(\mathbf{m},\mathbf{X},\boldsymbol{\rho})$

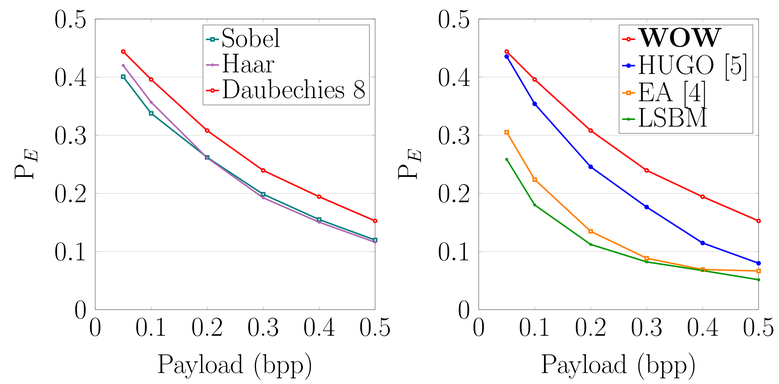
Embed message $\mathbf{m}$ *using ternary STC's [1]*

Experiments

- BOSSbase 1.00 (10,000 images)
- spatial rich model (SRM) [2] with 106 submodels and dimension of 34,761
- ensemble classifier [3] with FLD's as base learners
- $\mathrm{P}_E = \min_{\mathrm{P}_{FA}} \frac{1}{2}(\mathrm{P}_{FA} + \mathrm{P}_{MD})$

Filter banks

- Sobel operator (2 filters)
- 2D Wavelet Haar (3 filters)
- 2D Wavelet Daubechies 8-tap (3 filters)



WOW = proposed method + Daubechies 8-tap filter bank

References

[1] T. Filler, J. Judas, and J. Fridrich. Minimizing additive distortion in steganography using syndrome-trellis codes. *IEEE TIFS*, 6(3):920–935, September 2011.

[2] J. Fridrich and J. Kodovský. Rich models for steganalysis of digital images. *IEEE TIFS*, 7(3):868–882, 2011.

[3] J. Kodovský, J. Fridrich, and V. Holub. Ensemble classifiers for steganalysis of digital media. *IEEE TIFS*, 7(2):432–444, 2012.

[4] W. Luo, F. Huang, and J. Huang. Edge adaptive image steganography based on LSB matching revisited. *IEEE TIFS*, 5(2):201–214, June 2010.

[5] T. Pevný, T. Filler, and P. Bas. Using high-dimensional image models to perform highly undetectable steganography. In *Information Hiding, 12th Int. Conf.*, volume 6387 of *Springer LNCS*, pages 161–177, 2010.