Using High-Dimensional Image Models to Perform Highly Undetectable Steganography

Tomáš Pevný[1], Tomáš Filler[2], Patrick Bas[3]

[1]CTU, Prague, Czech Republic
[2] SUNY, Binghamton, USA
[3]Lagis, Lille, France

29th June 2010

Outline

Outline

Steganography

Practical steganography for digital media

- modifies the cover objects to convey the message.
- makes changes as undetectable as possible.

Distortion function

- any function $D : \mathscr{X} \times \mathscr{X} \to [0, \infty]$.
- correlates with detectability.
- is minimized during embedding.

Additive distortion function

$$D(x, y) = \sum_{i=1}^{n} \rho_i |x_i - y_i|$$

- $|x_i - y_i| \leq 1$,
- ρ_i cost of changing one pixel (embedding impact)
- Aditivity implies that embedding changes do not interact.

Separational principle

Theorem[a]

If we want to communicate m bits in n elements (pixels), than the minimal expected distortion is

$$D_{\min}(m, n, \rho) = \sum_{i=1}^{n} p_i \rho_i,$$

where p_i is the probability of changing the ith pixel,

$$p_i = \frac{e^{-\lambda \rho_i}}{1 + e^{-\lambda \rho_i}}.$$

The parameter λ is obtained by solving $\sum_{i=1}^{n} H(p_i) = m$.

[a] J. Fridrich and T. Filler, Practical Methods for Minimizing Embedding Impact in Steganography, 2007

Corollary of the theorem

Corrolary

- Design of the steganographic algorithm boils down to
  - the design of an additive distortion function D, or
  - the setting embedding costs ρ_i.

- Allows to compare additive distortion functions.
- Practical algorithms approaching the distortion bound exists[a].

[a] T. Filler, J. Fridrich, and J. Judas, Minimizing embedding impact in steganography using Trellis-Coded Quantization, 2010

Outline

Designing the distortion function

Distortion function

$$D(\mathbf{x},\mathbf{y}) \quad = \|f(\mathbf{x}) - f(\mathbf{y})\| = \quad \sum_{j=1}^{d} w_j \, |f_j(\mathbf{x}) - f_j(\mathbf{y})|$$

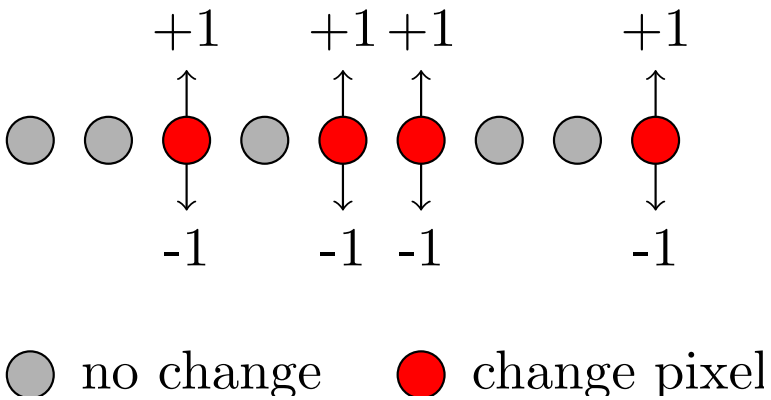
d — number of features

Additive approximation

$$D'(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n} D(\mathbf{x}, y_i\mathbf{x}) |x_i - y_i|$$

n — number of pixels

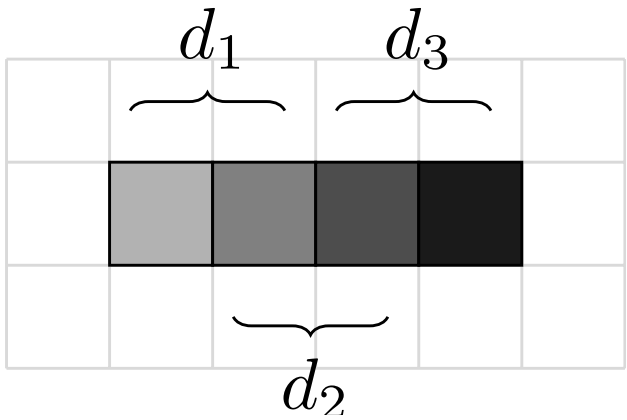
Model Correction



Compensates the suboptimality caused by approximating $D(x, y)$ by $D'(x, y)$.

Outline

Features of the model

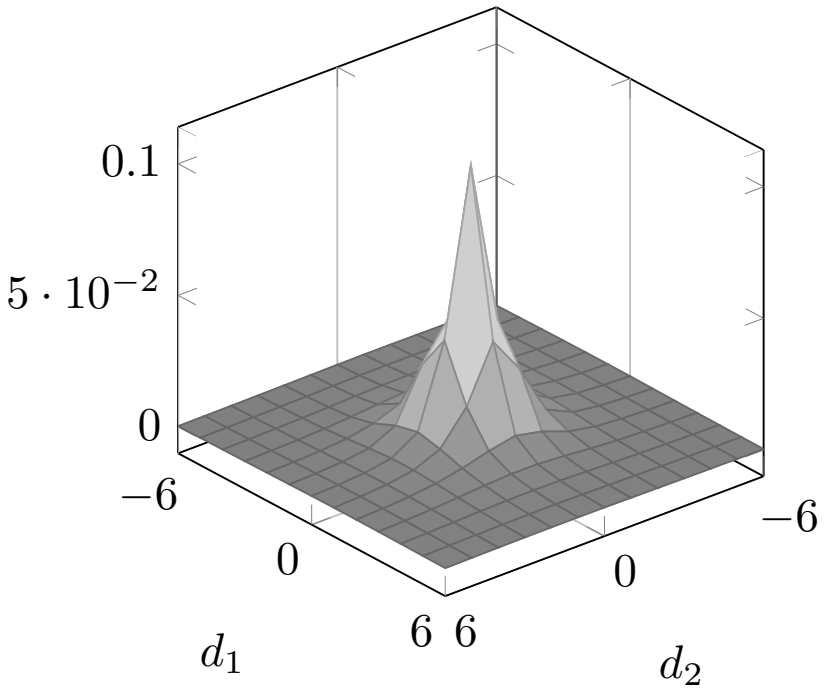


Distortion function

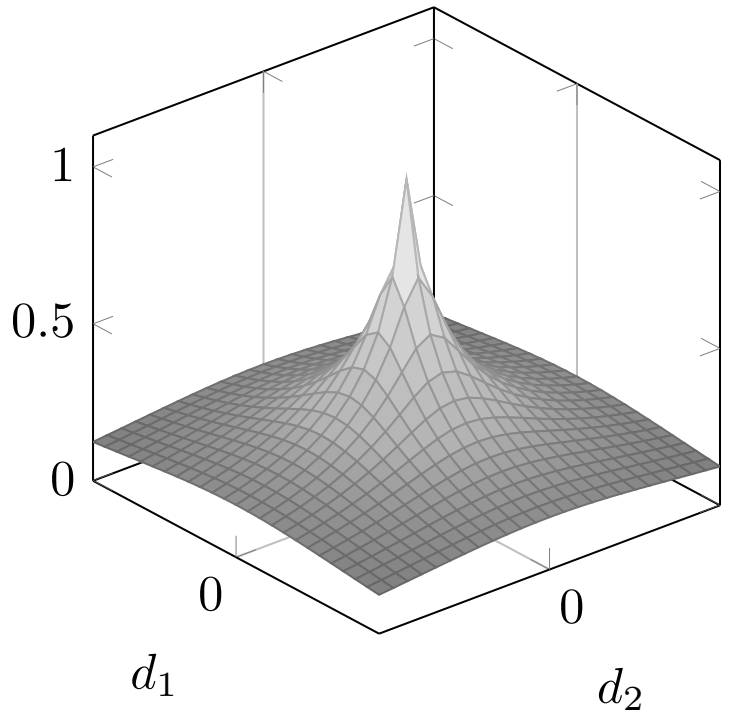
$$D(\mathbf{x},\mathbf{y}) = \sum_{d_1,d_2,d_3=-T}^{T} w_{d_1,d_2,d_3} \left| f_{d_1,d_2,d_3}(\mathbf{x}) - f_{d_1,d_2,d_3}(\mathbf{y}) \right|$$

$f^{\rightarrow}_{d_1,d_2,d_3}$ — # of differences (d_1, d_2, d_3) between neighboring pixels

Setting the weights



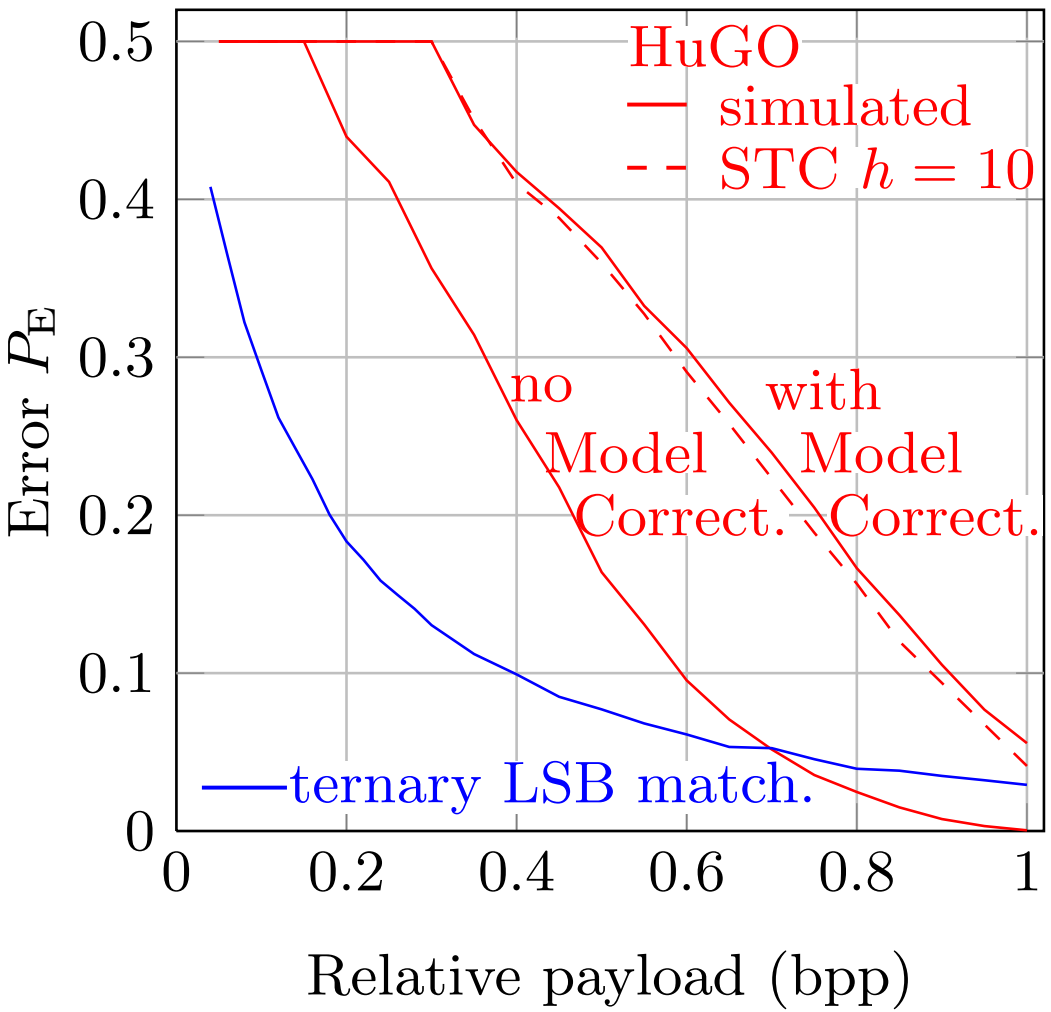
Mean of $\mathbf{C}^{\mathbf{X},\rightarrow}_{d_1 d_2}$ feature



$$w(d_1, d_2) = \left[\sqrt{d_1^2 + d_2^2} + \sigma\right]^{-\gamma}$$

Outline

Detectability by Spam

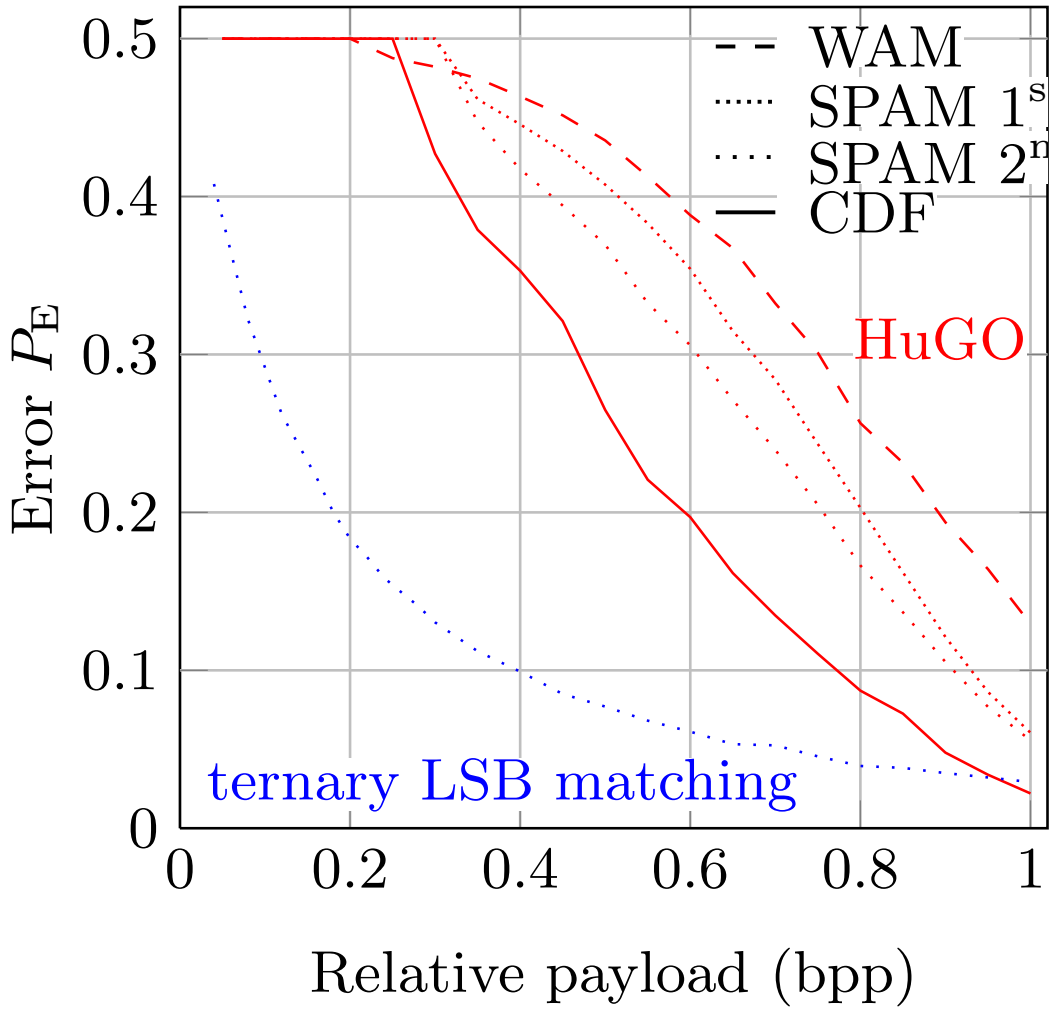


fixed size 512×512 images

$$P_E = \min \frac{1}{2}\left(P_{Fp} + P_{Fn}\right)$$

SVMs with Gaussian kernel.

Detectability by feature sets



HuGO, where did you hide the message?

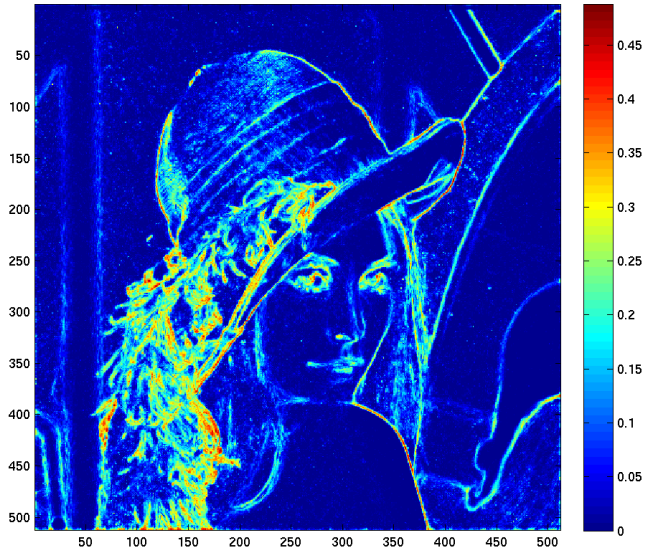


Fig: 0.25 bits per pixel

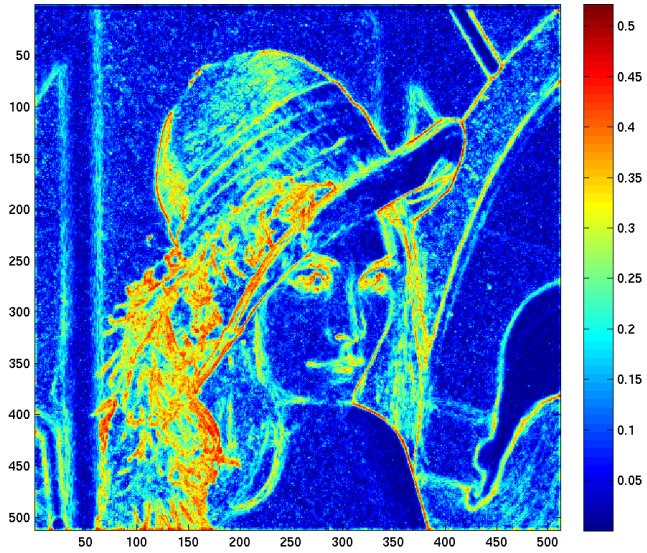


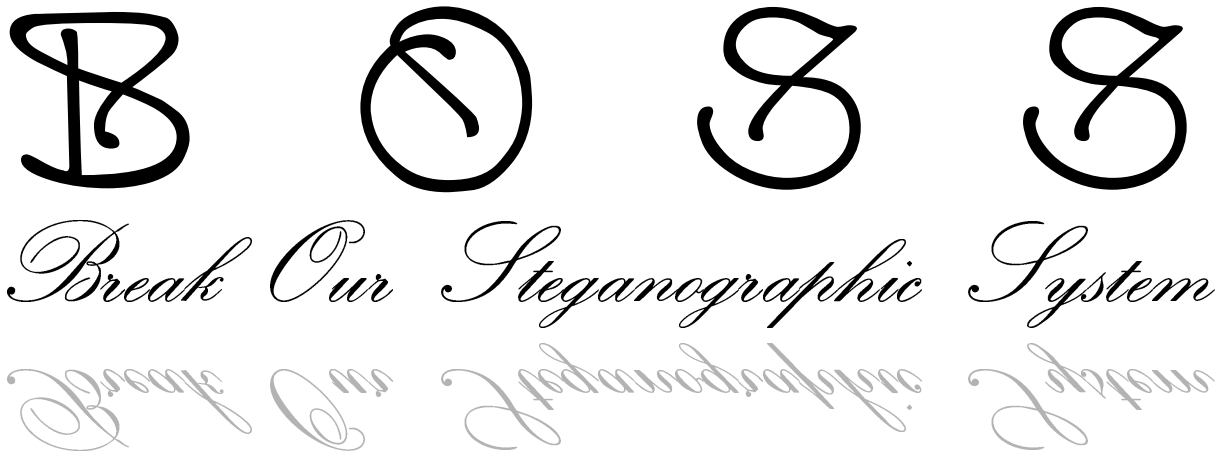
Fig: 0.50 bits per pixel

Outline

Conclusion

- We presented a methodology to design a steganographic algorithm by applying state of the art principles:
  - separate distortion function from coding
  - use of high-dimensional model (10^7 features).
- The practical realization, HuGO allows the embedder to hide $7\times$ longer message than LSB matching at the same level of security.

Do you want be the BOSS?



Steganalytic challenge is coming up in June 2010!
1000 images, 500 with a hidden message
Guess which ones!

http://boss.gipsa-lab.grenoble-inp.fr