Steganalysis in high dimensions: Fusing classifiers built on random subspaces

Jan Kodovský, Jessica Fridrich

January 25, 2011 / SPIE

BINGHAMTON
UNIVERSITY
STATE UNIVERSITY OF NEW YORK

Motivation

- Modern steganography
  - Minimizing a distortion function in a high dimensional feature space

    Example: HUGO [Pevný-2010] (spatial domain) – 10^7 dimensions
  - Preserving complex models

    Example: Optimized ± 1 embedding (JPEG domain) [Filler-Yesterday]

- Modern approach to steganalysis
  - Needs to follow the suit and capture more and more statistics
  - Cartesian calibration [2009] – doubles dimensionality
  - Merging of existing features together
  - ± 1 embedding $\longrightarrow$ SPAM features (686) [Pevný-2009]
  - YASS algorithm (JPEG domain) $\longrightarrow$ CDF features (1,234) [2010]

Curse of dimensionality

- Growing complexity of training
- Limited training data / no access to the cover source
- Degradation of generalization abilities (overtraining)

  $\Rightarrow$ model assumptions / regularization
- Problems with data / memory management
- Saturation of performance below its potential

Features are *designed* to have low dimensionality

Our goals

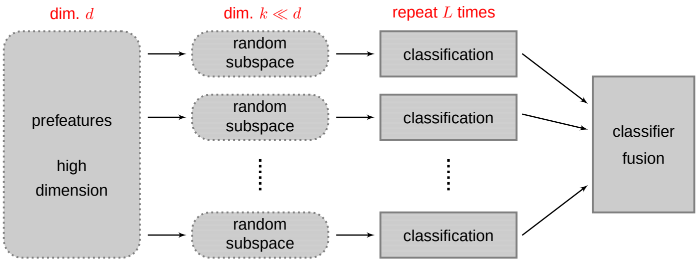
- Challenge the low-dimensional limitation for a feature design
- Replace human design of features with an automatized procedure
- Rethink machine learning approach to steganalysis
- Classify in very high dimensions with low complexity and without compromising the performance
- Improve state-of-the-art steganalysis

What are the options?

1. Apply a classification tool of choice directly
2. Reduce dimensionality and then classify
   - Unsupervised techniques (PCA)
   - Supervised techniques (feature extraction / selection)
   - Can be thought of as part of the feature design
3. Reduce dimensionality and simultaneously classify
   - Minimize an appropriately defined objective function (SVDM)
   - Iterative process with a classification feedback (embedded methods)
4. Ensemble methods
   - Reduce dimensionality randomly and construct a simple classifier
   - Repeat L times and aggregate the individual decisions

The proposed framework

- Step 1 – Form high-dimensional *prefeatures*
  - Capture as many dependencies among cover elements as possible
  - Don't be restricted by a dimensionality
  - Emphasize diversity of individual features
- Step 2 – Classify in high dimensions using an ensemble approach



Specific implementation

- Random subspace = random *selection* (without repetition)

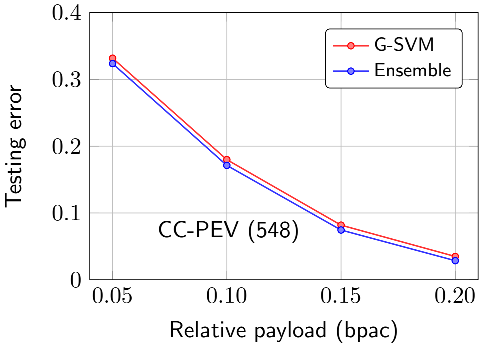
  $\Rightarrow$ The complexity does not depend on the dimensionality d
- Individual classifiers (base learners)
  - Need to be sufficiently diverse (need to make different errors)
  - Weak and unstable classifiers preferable
  - Our choice: Fisher Linear Discriminants (FLDs)
- Fusion = majority voting scheme $\sum_{i=1}^{L}$decision$(i) >$ threshold
- Parameters $k \approx$ 300 – 3000, $L \approx$ 30 – 150

Relation to previous art:

- [Freund-1999] – Boosting (aggregation of weak classifiers)
- [Breiman-2001] – Random forests (base learners = trees)

Comparison with SVM

- JPEG domain, algorithm nsF5, database of 6500 images
- State-of-the-art feature sets
  - CC-PEV (2×274 = 548) – [Pevný-2007] + Cartesian calibration

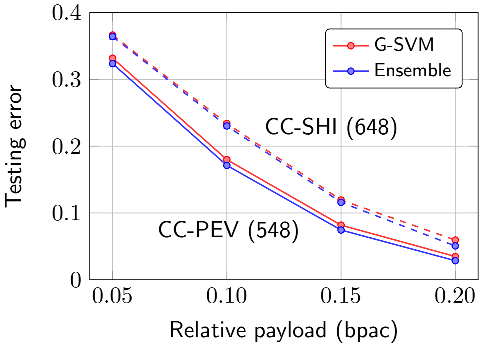


- $k = 400$, $L = 31$
- Ensemble: 70 sec
- G-SVM: 250 sec (3.5 $\times$ longer)

Full training: 8 hrs!

Comparison with SVM

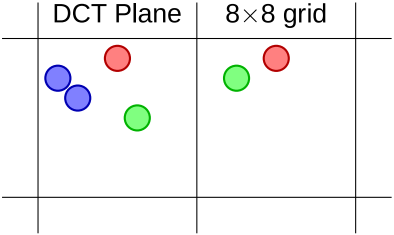
- JPEG domain, algorithm nsF5, database of 6500 images
- State-of-the-art feature sets
  - CC-PEV (2×274 = 548) – [Pevný-2007] + Cartesian calibration
  - CC-SHI (2×324 = 648) – [Shi-2006] + Cartesian calibration



- $k = 400$, $L = 31$
- Ensemble: 70 sec
- G-SVM: 250 sec (3.5 × longer)

Full training: 8 hrs!

Generating high-dimensional prefeatures (in JPEG domain)



intra-block dependencies

inter-block dependencies

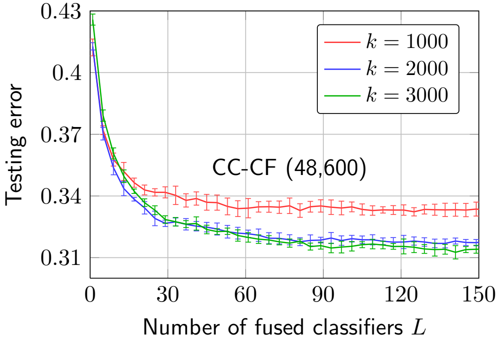
combination of both

- 2D co-occurence matrices
- Driven by mutual information
- N matrices in total
- Truncated to $[-T, T]$
- Cartesian calibration
- Dimension $2 \times N \times (2 \times T + 1)^2$
- $T = 4$, $N = 300 \rightarrow$ dim = 48,600

CC-CF features

Steganalysis of nsF5

- Influence of parameters L and k



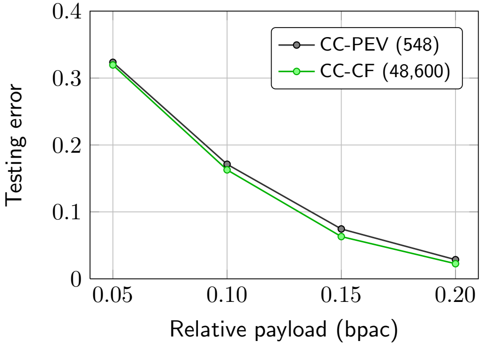
- Payload 0.05 bpac
- $k = 2000$, $L = 149$
  $\rightarrow$ 30 min
- G-SVM: 7.5 hrs
  (15 $\times$ longer)

  Full training > month

- Performance quickly saturates as L grows
- Choice of k is important (1D search may be conducted)

Steganalysis of nsF5

- Can we improve state-of-the-art?

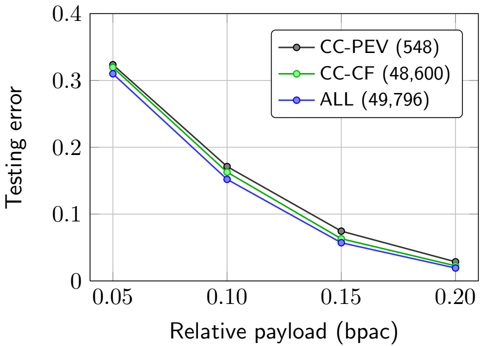


– CC-PEV: G-SVM
– Rest: Ensemble

$k = 2000$, $L = 149$

Steganalysis of nsF5

- Can we improve state-of-the-art?



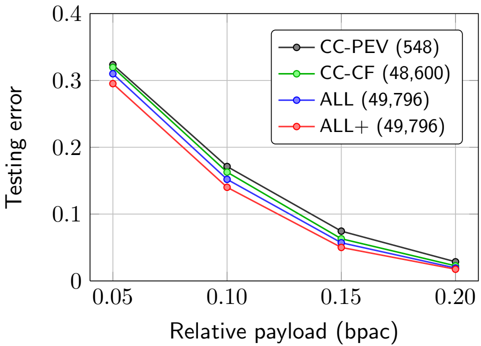
- CC-PEV: G-SVM
- Rest: Ensemble

  $k = 2000,\ L = 149$

- ALL (49,796) = CC-PEV (548) + CC-SHI (648) + CC-CF (48,600)

Steganalysis of nsF5

- Can we improve state-of-the-art?



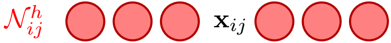
- CC-PEV: G-SVM
- Rest: Ensemble

  $k = 2000,\ L = 149$

- ALL (49,796) = CC-PEV (548) + CC-SHI (648) + CC-CF (48,600)
- ALL+ $=$ ALL with 300/2000 always chosen from CC-PEV

Generating high-dimensional prefeatures (in SPATIAL domain)

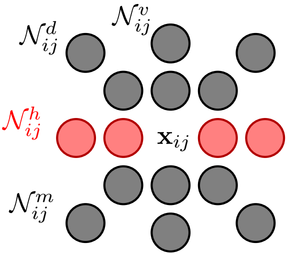
- Modeling the joint distribution of higher order local residuals
- Horizontal residual $\mathbf{H}_{ij} = \mathbf{x}_{ij} - \mathrm{Pred}(\mathcal{N}_{ij}^h)$



Order	$\mathbf{H}_{ij}$
2	$\frac{1}{2}(-\mathbf{x}_{i,j-1} + 2\mathbf{x}_{ij} - \mathbf{x}_{i,j+1})$
3	$\frac{1}{3}(-\mathbf{x}_{i,j-1} + 3\mathbf{x}_{ij} - 3\mathbf{x}_{i,j+1} + \mathbf{x}_{i,j+2})$
4	$\frac{1}{6}(\mathbf{x}_{i,j-2} - 4\mathbf{x}_{i,j-1} + 6\mathbf{x}_{ij} - 4\mathbf{x}_{i,j+1} + \mathbf{x}_{i,j+2})$
5	$\frac{1}{10}(\mathbf{x}_{i,j-2} - 5\mathbf{x}_{i,j-1} + 10\mathbf{x}_{i,j} - 10\mathbf{x}_{i,j+1} + 5\mathbf{x}_{i,j+2} - \mathbf{x}_{i,j+3})$
6	$\frac{1}{20}(-\mathbf{x}_{i,j-3} + 6\mathbf{x}_{i,j-2} - 15\mathbf{x}_{i,j-1} + 20\mathbf{x}_{ij} - 15\mathbf{x}_{i,j+1} + 6\mathbf{x}_{i,j+2} - \mathbf{x}_{i,j+3})$

Generating high-dimensional prefeatures (in SPATIAL domain)

- Modeling the joint distribution of higher order local residuals
- Horizontal residual $\mathbf{H}_{ij} = \mathbf{x}_{ij} - \mathrm{Pred}(\mathcal{N}_{ij}^{h})$



$$\mathbf{H}_{ij} = \mathbf{x}_{ij} - \mathrm{Pred}(\mathcal{N}_{ij}^{h}) \qquad \mathbf{D}_{ij} = \mathbf{x}_{ij} - \mathrm{Pred}(\mathcal{N}_{ij}^{d})$$
$$\mathbf{V}_{ij} = \mathbf{x}_{ij} - \mathrm{Pred}(\mathcal{N}_{ij}^{v}) \qquad \mathbf{M}_{ij} = \mathbf{x}_{ij} - \mathrm{Pred}(\mathcal{N}_{ij}^{m})$$

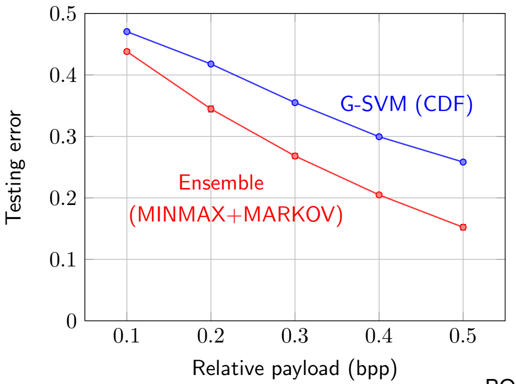
$\mathbf{H}_{ij}, \mathbf{V}_{ij}, \mathbf{D}_{ij}, \mathbf{M}_{ij} \longrightarrow$ MARKOV

$\left.\begin{array}{l} \min\{\mathbf{H}_{ij}, \mathbf{V}_{ij}, \mathbf{D}_{ij}, \mathbf{M}_{ij}\} \\ \max\{\mathbf{H}_{ij}, \mathbf{V}_{ij}, \mathbf{D}_{ij}, \mathbf{M}_{ij}\} \end{array}\right. \longrightarrow$ MINMAX

- 3D co-occurences, dimension $20\times(2\times T+1)^3$ (T = 4 $\to$ dim = 14,580)

Steganalysis of HUGO

- G-SVM $\longrightarrow$ CDF (1,234) $=$ CC-PEV (548) $+$ SPAM (686)
- Ensemble $\longrightarrow$ MINMAX+MARKOV (14,580), $k = 1600$, $L = 51$



BOSSbase (9074 images)
size: 512×512, resized

Summary

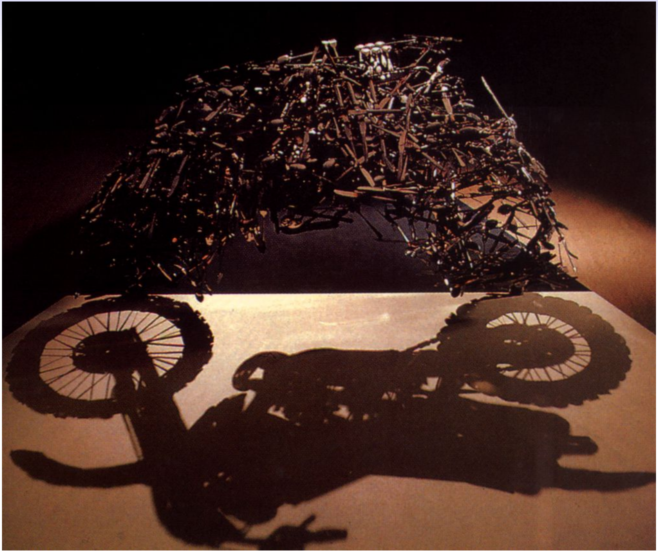
The main contributions for future steganalysis

- High dimensionality doesn't have to be a restriction for the feature design
- Proposed scalable, fast, and simple classification methodology based on ensemble classifiers
- One step further towards automatization of steganalysis
- Showed that state-of-the-art steganalysis can be improved by a large margin

Open problems

- How to design prefeatures?
- How to define random projections?

The power of random projections



Shigeo Fukuda, Lunch With a Helmet On (1987)