Steganalysis of JPEG images using rich models

Jan Kodovský, Jessica Fridrich

January 23, 2012 / SPIE

BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK

Feature–based steganalysis

Two building blocks

- Feature–space representation of digital images
- Binary classifier trained on examples of cover and stego features

Feature space (image model)

- Statistical descriptor of images (or their noise component)
- Captures dependencies among image coefficients
- Sensitive to stego-modifications, insensitive to image content

Classifier

- Any machine–learning tool can be used (FLD, LR, SVM, NN)
- The choice of the classifier (shapes of decision boundaries, training complexity) and available computing resources inherently influence the feature space design

Current trends in steganalysis

Modern steganography requires more complex feature spaces

- 18 BSM [Avcıbaş,2002], 72 higher–order moments [Farid,2002]
- 23 DCT [Fridrich,2004] $\rightarrow$ 274 PEV [Pevný,2007] $\rightarrow$ 548 CC-PEV
- 324 [Shi,2006] $\rightarrow$ 486 [Chen,2008] – Markov–based features
- 686 SPAM [Pevný,2010] $\rightarrow$ 1234 CDF = SPAM + CC-PEV

Strategies for model enrichment

- Merge existing feature sets together
- Add reference values, include more statistics

increasing dimensionality

Machine learning needs to adapt

- Machine learning should not constrain feature space design
- SVM – accurate, but infeasible in high dimensions
- Ensemble classifier [2011] – scalable w.r.t. dimensionality and the number of training samples

How to build features in JPEG domain
(Without dimensionality constraints)

JPEG domain specifics

- 8×8 blocks of coefficients in different DCT modes
- 64 parallel channels of different statistical properties
- Two types of dependencies: frequency and spatial (intra/inter block)

Model–building guidelines

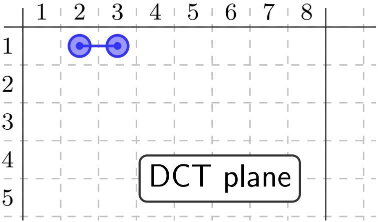
- Capture as many dependencies as possible, proceed systematically
- Learn from previously proposed feature spaces, *e.g.* co-occurrences
- Model individual DCT modes separately $\Rightarrow$ large number of submodels
- Keep submodels well populated – utilize natural symmetries, small T
- Include also integral components – sum over the image, larger T
- Inspiration from spatial domain – BOSS competition
- Diversity, diversity, diversity

$\boxed{\text{Rich Model}}$

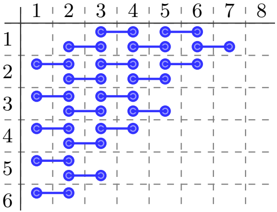
JPEG domain Rich Model (JRM)

1. The first DCT–mode specific model

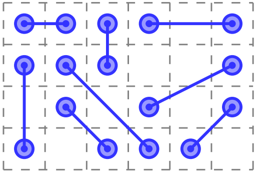
- Absolute values of DCT coefficients
- Selected DCT mode: (1,2)
- Horizontal neighbor
- 2D co-occurrence matrix
- Truncate with $T = 3$
- Dimension = $(T+1)^2 = 16$



2. Cover low-frequencies



3. Extend to other intra-block neighbors



95 submodels

JPEG domain Rich Model (JRM)

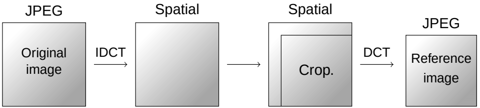
4. Keep scaling up the rich model

- Add inter-block neighbors (horizontal, vertical, diagonal) → 157 submodels
- Repeat everything for <u>differences</u> of DCT coefficients → 628 submodels
  - Horizontal, vertical, and diagonal intra-block differences
  - Horizontal and vertical inter-block differences
- Add <u>integral</u> features (T = 5) → 673 submodels
  - Both inter- and intra-block co-occurrences
  - From both absolute values and differences

JRM
(11,255)

---

5. Apply Cartesian calibration → dimension doubles to 22,510



CC-JRM
(22,510)

Ensemble classifier – overview

- Designed to be scalable w.r.t. feature–space dimensionality
- Built as a fusion of many weak classifiers (base learners) built on random subspaces of the original feature space
- Specific implementation choices:
  - Base learner = Fisher Linear Discriminant (FLD)
  - Fusion = majority voting scheme $\sum_i \text{decision}(i) > \text{threshold}$
- All parameters automatically optimized on the training set
- Relationship to prior art: [Breiman-2001] – Random forests
- Fast, comparable accuracy to SVMs
- Detailed description appears in [SPIE, 2011], [TIFS, 2012]
- `http://dde.binghamton.edu/download/ensemble`

Comparison to prior art

Experimental setup

- 6,500 images coming from 22 cameras, resized, JPEG 75
- Ensemble classifier, average testing error over 10 splits

Steganographic methods

- nsF5 – non-shrinkage version of F5 [Westfeld, 2001]
- MBS – model–based steganography [Sallee, 2003]
- YASS – yet another steganographic scheme [Solanki, 2007]
- MME – modified matrix encoding [Kim, 2006]
- BCH – utilizes structured BCH syndrome coding [Sachnev,2009]
- BCHopt – BCH with heuristic optimization [Sachnev, 2009]

Comparison to prior art

Feature sets

- CHEN (486) – Markov features, intra- & inter-block [Chen, 2008]
- CC-CHEN (972) – CHEN features improved by Cartesian calibration
- LIU (216) – differences of abs. values, different calibrations [Liu, 2011]
- CC-PEV (548) – Cartesian–calibrated PEV features [Pevný, 2007]
- CDF (1,234) – CC-PEV expanded by SPAM features [Pevný, 2009]
- CC-C300 (48,600) – driven by mutual information [Kodovský, 2011]
- CF* (7,850) – compact rich model [Kodovský, 2012]
- JRM (11,255) – JPEG domain Rich Model
- CC-JRM (22,510) – Cartesian–calibrated JRM
- J+SRM (35,263) – CC-JRM + Spatial domain Rich Model [under review]

---

http://dde.binghamton.edu/download/feature_extractors

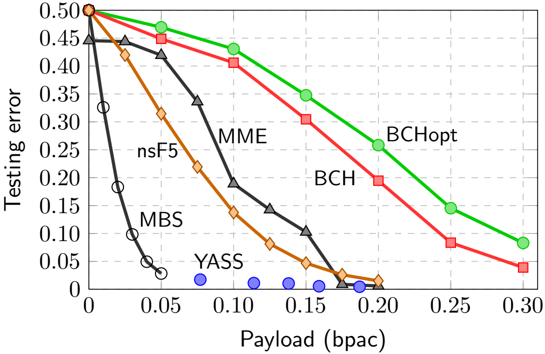
Comparison to prior art

new models

Algorithm	bpac	LIU (216)	CHEN (486)	CC-PEV (548)	CC-CHEN (972)	CDF (1,234)	CF* (7,850)	JRM (11,255)	CC-JRM (22,510)	J+SRM (35,263)	CC-C300 (48,600)
nsF5	0.10	.1732	.3097	.2239	.2470	.2020	.1737	.1782	.1616	.1375	.2207
	0.15	.0706	.2094	.1171	.1393	.0906	.0720	.0793	.0663	.0468	.1127
MBS	0.01	.3826	.4070	.3876	.3962	.3786	.3710	.3478	.3414	.3260	.4038
	0.05	.0812	.1243	.0833	.0946	.0704	.0684	.0427	.0373	.0282	.1176
YASS	0.16	.1793	.2334	.1341	.1476	.0507	.0164	.0210	.0103	.0054	.0370
	0.19	.1301	.1277	.0723	.0876	.0224	.0146	.0165	.0081	.0045	.0350
MME	.10	.2574	.3001	.2613	.2611	.2501	.2466	.2286	.2091	.1891	.3026
	.15	.1677	.2165	.1721	.1735	.1586	.1608	.1404	.1221	.1027	.2299
BCH	0.20	.3087	.3594	.2974	.3124	.2752	.2629	.2707	.2369	.1946	.2958
	0.30	.0862	.1383	.0779	.0889	.0697	.0663	.0715	.0536	.0390	.0912
BCHopt	0.20	.3583	.4032	.3548	.3712	.3368	.3265	.3253	.3030	.2582	.3517
	0.30	.1719	.2400	.1605	.1711	.1356	.1289	.1389	.1102	.0830	.1681

- High dimension is not sufficient
- Steganalysis benefits from cross-domain models
- Cartesian calibration helps even in high dimensions
- Diverse and compact rich models deliver best results

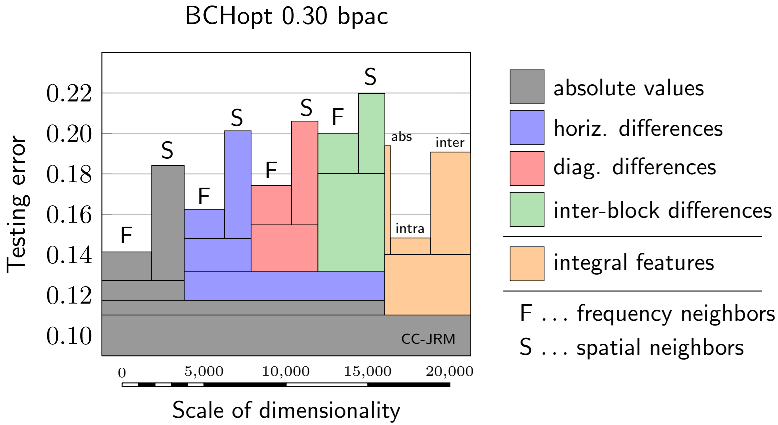
Comparison of stego methods



J+SRM
(35,263)

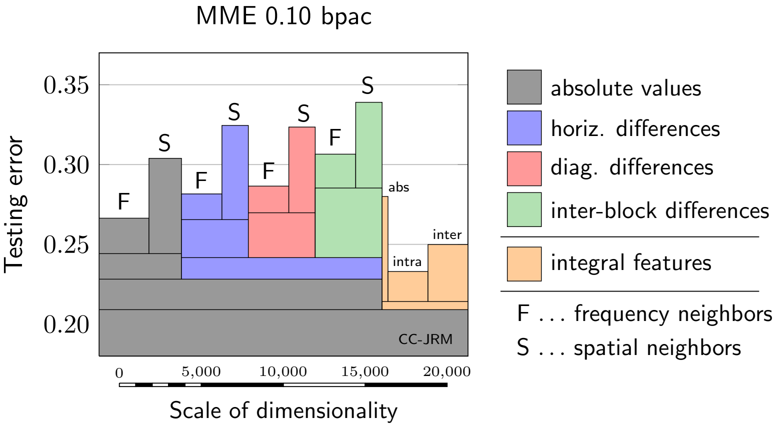
- MBS and YASS are by far the least secure
- Side-informed schemes (MME, BCH, BCHopt) perform better
- Jumps in MME due to its suboptimal coding

Experiment 1 – systematic merging



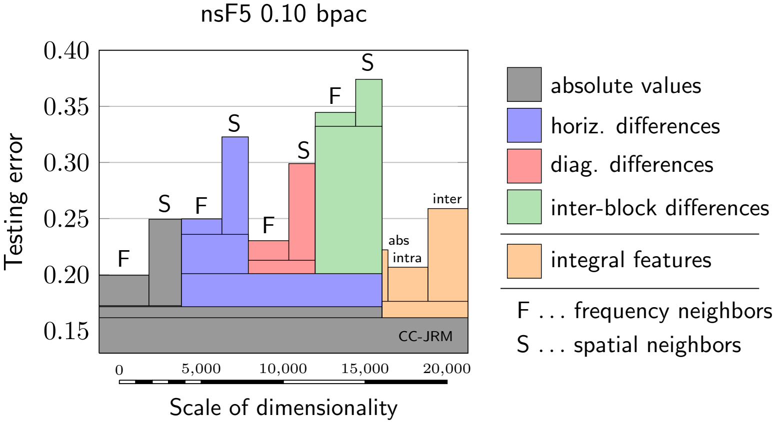
- Width of each bar is proportional to the model dimensionality
- Reveals what types of features are effective against a given scheme

Experiment 1 – systematic merging



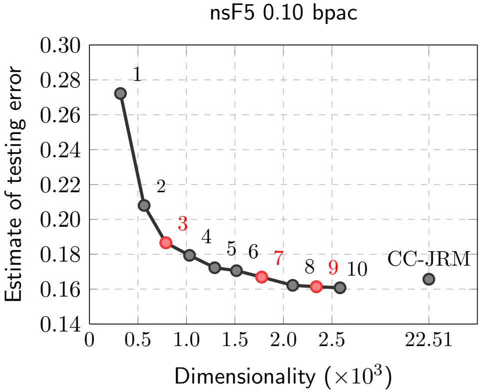
- Width of each bar is proportional to the model dimensionality
- Reveals what types of features are effective against a given scheme

Experiment 1 – systematic merging



- Width of each bar is proportional to the model dimensionality
- Reveals what types of features are effective against a given scheme

Experiment 2 – forward feature selection

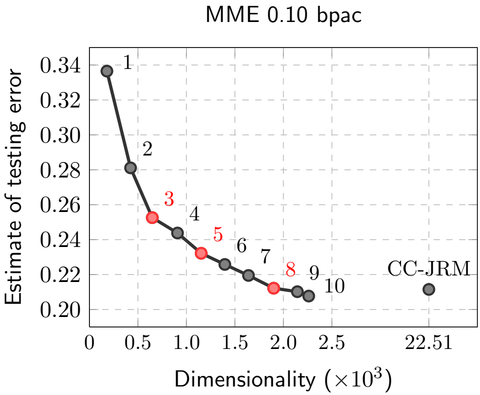


1. F (abs. values)
2. F (integral)
3. S (abs. values) ← reference
4. F (integral)
5. S (differences)
6. F (abs. values)
7. F (differences) ← reference

F . . . frequency neighbors
S . . . spatial neighbors

- Greedy minimization of the testing error estimate (2 × 51 submodels)
- Add the submodel that best complements those already selected
- Red corresponds to reference submodels from Cartesian calibration

Experiment 2 – forward feature selection



1. F (integral)
2. S (integral)
3. F (abs. values) ← reference
4. F (differences)
5. S (differences) ← reference
6. F (integral)
7. F (itnegral)

F . . . frequency neighbors
S . . . spatial neighbors

- Greedy minimization of the testing error estimate (2 × 51 submodels)
- Add the submodel that best complements those already selected
- Red corresponds to reference submodels from Cartesian calibration

Conclusion

Summary

- Steganalysis using rich models and scalable machine learning improves previous approaches
- CC-JRM is universally effective rich model for JPEG domain
- For a fixed steganographic channel, dimensionality of CC-JRM can be drastically reduced
- Merging with Spatial Rich Model further improves steganalysis
- Calibration helps even in high-dimensional spaces

Open problems

- Bottleneck of steganalysis becomes feature extraction
- Robustness of rich models w.r.t. cover–source mismatch

Resources

- Ensemble: `http://dde.binghamton.edu/download/ensemble`
- Features: `http://dde.binghamton.edu/download/feature_extractors`

References

- **İ. Avcıbaş, N. D. Memon and B. Sankur**. Image steganalysis with binary similarity measures. Proc. IEEE, International Conference on Image Processing, ICIP volume 3, pages 645–648, Rochester, NY, September 22–25, 2002.
- **L. Breiman**. Random forests. Machine Learning, 45:5–32, October 2001.
- **C. Chen and Y. Q. Shi**. JPEG image steganalysis utilizing both intrablock and interblock correlations. Circuits and Systems, ISCAS, IEEE International Symposium on, pages 3029–3032, May 2008.
- **H. Farid and L. Siwei**. Detecting hidden messages using higher-order statistics and support vector machines. Information Hiding, 5th International Workshop, volume 2578 of LNCS, pages 340–354, Noordwijkerhout, The Netherlands, October 7–9, 2002. Springer-Verlag, New York.
- **J. Fridrich**. Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes. Information Hiding, 6th International Workshop, volume 3200 of LNCS, pages 67–81, Toronto, Canada, May 23–25, 2004. Springer-Verlag, New York.
- **Y. Kim, Z. Duric, and D. Richards**. Modified matrix encoding technique for minimal distortion steganography. Information Hiding, 8th International Workshop, volume 4437 of LNCS, pages 314–327, Alexandria, VA, July 10–12, 2006. Springer-Verlag, New York.
- **J. Kodovský, J. Fridrich, and V. Holub**. Ensemble classifiers for steganalysis of digital media. IEEE Transactions on Information Forensics and Security, 2012. To appear.
- **J. Kodovský and J. Fridrich**. Steganalysis of JPEG images using rich models. Proc. SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics of Multimedia XIV, San Francisco, CA, January 22–26, 2012.
- **J. Kodovský and J. Fridrich**. Steganalysis in high dimensions: Fusing classifiers built on random subspaces. Proc. SPIE, Electronic Imaging, Media Watermarking, Security and Forensics of Multimedia XIII, volume 7880, pages OL 1–13, San Francisco, CA, January 23–26, 2011.
- **Q. Liu**. Steganalysis of DCT-embedding based adaptive steganography and YASS. Proc. of the 13th ACM Multimedia & Security Workshop, pages 77–86, Niagara Falls, NY, September 29–30, 2011.
- **T. Pevný and J. Fridrich**. Merging Markov and DCT features for multi-class JPEG steganalysis. Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX, volume 6505, pages 3 1–3 14, San Jose, CA, January 29–February 1, 2007.

References

- **T. Pevný, P. Bas, and J. Fridrich**. Steganalysis by subtractive pixel adjacency matrix. IEEE Transactions on Information Forensics and Security, 5(2):215–224, June 2010.
- **V. Sachnev, H. J. Kim, and R. Zhang**. Less detectable JPEG steganography method based on heuristic optimization and BCH syndrome coding. Proc. of the 11th ACM Multimedia & Security Workshop, pages 131–140, Princeton, NJ, September 7–8, 2009.
- **P. Sallee**. Model-based steganography. Digital Watermarking, 2nd International Workshop, volume 2939 of LNCS, pages 154–167, Seoul, Korea, October 20–22, 2003. Springer-Verlag, New York.
- **Y. Q. Shi, C. Chen, and W. Chen**. A Markov process based approach to effective attacking JPEG steganography. Information Hiding, 8th International Workshop, volume 4437 of LNCS, pages 249–264, Alexandria, VA, July 10–12, 2006. Springer-Verlag, New York.
- **K. Solanki, A. Sarkar, and B. S. Manjunath**. YASS: Yet another steganographic scheme that resists blind steganalysis. Information Hiding, 9th International Workshop, volume 4567 of LNCS, pages 16–31, Saint Malo, France, June 11–13, 2007. Springer-Verlag, New York.
- **A. Westfeld**. High capacity despite better steganalysis (F5 – a steganographic algorithm). Information Hiding, 4th International Workshop, volume 2137 of LNCS, pages 289–302, Pittsburgh, PA, April 25–27, 2001. Springer-Verlag, New York.