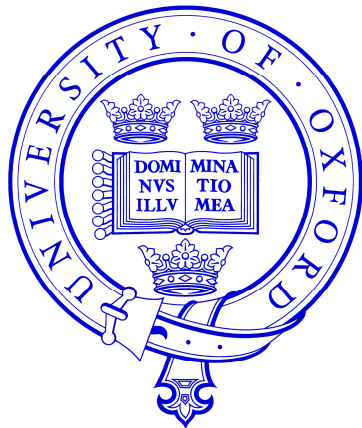


Feature Reduction and Payload Location with WAM Steganalysis



Andrew Ker & Ivans Lubenko

Oxford University Computing Laboratory

contact: adk@comlab.ox.ac.uk

SPIE/IS&T Electronic Imaging, San Jose, CA

19 January 2009

LSB matching (± 1 embedding)

- Host LSBs carry payload, but other bits are also affected.
- Easy to implement, high capacity, visually imperceptible.
- Detectors performance is **poor** and **variable**:

Histogram Characteristic Function (HCF)	Harmsen & Pearlman, 2003, 2004 Ker, 2005 Li <i>et al.</i>, 2008
Analysis of Local Extrema (ALE)	Cancelli <i>et al.</i>, 2007, 2008
Wavelet Higher Order Statistics	Holotyak <i>et al.</i>, 2005
Wavelet Absolute Moments (WAM)	Goljan <i>et al.</i>, 2006

We contribute three things to the development of WAM:

- ☹ *Separate benchmarks for different cover sources*
- ☹ *Feature reduction*
- ☺ *Payload location*

WAM features

The WAM features measure the **predictability of noise residuals**, in the wavelet domain.

1. From input \mathbf{X} , compute 1-level wavelet decomposition:

$$[\mathbf{L}, \mathbf{H}, \mathbf{V}, \mathbf{D}] = \text{DWT}(\mathbf{X})$$

2. The WAM filter gives quasi-Wiener residuals:

$$\mathcal{R}[\mathbf{S}] = \frac{\sigma_0^2 \mathbf{S}}{\sigma_0^2 + \mathbf{v}} \quad (\text{where } \mathbf{v} \text{ is a MAP estimate of local variance based on 4 windows, and } \sigma_0^2 \text{ is the noise variance, here 0.5})$$

3. The 27 WAM features are the absolute central moments of the high-frequency subband residuals:

$$A_m^H = \frac{1}{|\mathbf{X}|} \sum_i \left| \mathcal{R}[\mathbf{H}]_i - \overline{\mathcal{R}[\mathbf{H}]} \right|^m, \quad m = 1, \dots, 9$$

$$A_m^V = \frac{1}{|\mathbf{X}|} \sum_i \left| \mathcal{R}[\mathbf{V}]_i - \overline{\mathcal{R}[\mathbf{V}]} \right|^m, \quad m = 1, \dots, 9$$

$$A_m^D = \frac{1}{|\mathbf{X}|} \sum_i \left| \mathcal{R}[\mathbf{D}]_i - \overline{\mathcal{R}[\mathbf{D}]} \right|^m, \quad m = 1, \dots, 9$$

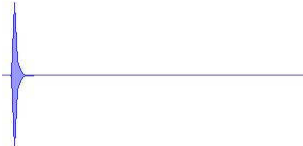
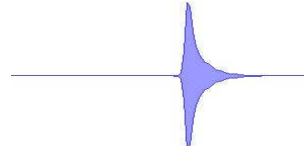
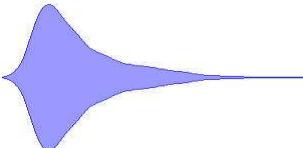
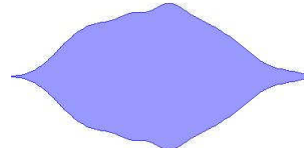
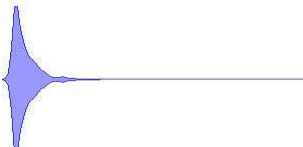
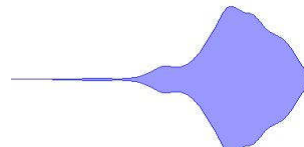
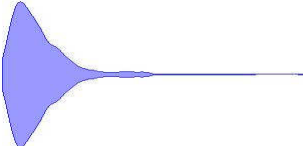
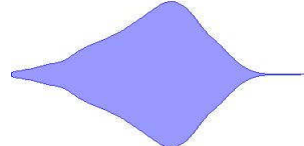
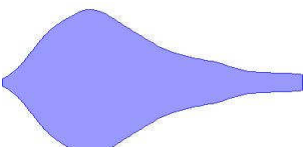
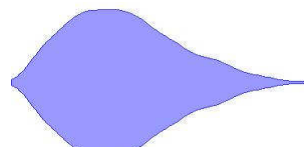
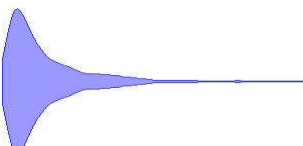
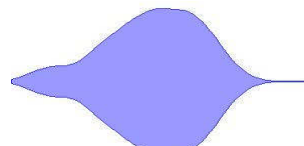
Effect of cover source

We benchmarked the accuracy of WAM steganalysis using three classification engines:

- The original Fisher Linear Discriminator (FLD),
- Multilayer Perceptron, a.k.a. Neural Network (NN),
- Support Vector Machine (SVM),

in **nine** different sets of images.

- *2000 grayscale cover images per set,*
- *all images cropped to 400×300,*
- *payload 0.5bpp (50% max),*
- *benchmarked by minimum of FP+FN, ten-fold cross validation.*

Set	Source	Image noise levels		Classification accuracy (%)		
		in spatial domain	in wavelet domain	FLD	NN	SVM
A	Digital camera <i>never-compressed, pre-processed as grayscale</i>			100	100	100
B	Digital camera <i>never-compressed, pre-processed as colour</i>			69.7	73.4	75.8
C	Various digital cameras <i>never-compressed, unknown pre-processing</i>			80.6	89.2	90.4
D	Photo library CD <i>decompressed JPEGs, quality factor 50</i>			95.5	97.7	97.5
E	Scanned photos <i>downsampled, never-compressed</i>			60.9	64.3	64.7
H	Internet photo sites <i>mixed JPEGs</i>			97.3	98.0	98.1
	...					

Set	Source	Image noise levels		Classification accuracy (%)		
		in spatial domain	in wavelet domain	FLD	NN	SVM
A	Digital camera <i>never-compressed, pre-processed as grayscale</i>			100	100	100
B	Digital camera <i>never-compressed, pre-processed as colour</i>			69.7	73.4	75.8
				significant significant $<$ $<$ <i>(p < 0.01)</i> <i>(p < 0.01)</i>		
C	Various digital cameras <i>never-compressed, unknown pre-processing</i>			80.6	89.2	90.4
				significant $<$ <i>(p < 0.001)</i>		
D	Photo library CD <i>decompressed JPEGs, quality factor 50</i>			95.5	97.7	97.5
				significant $<$ <i>(p < 0.001)</i>		
E	Scanned photos <i>downsampled, never-compressed</i>			60.9	64.3	64.7
				significant $<$ <i>(p < 0.01)</i>		
H	Internet photo sites <i>mixed JPEGs</i>			97.3	98.0	98.1
				significant $<$ <i>(p < 0.01)</i>		

...

Feature reduction

The WAM features cannot be independent: $A_m^H = \frac{1}{|X|} \sum_i |\mathcal{R}[H]_i - \overline{\mathcal{R}[H]}|^m$, etc.

PCA suggests the set of 27 features has only 3-5 independent dimensions.

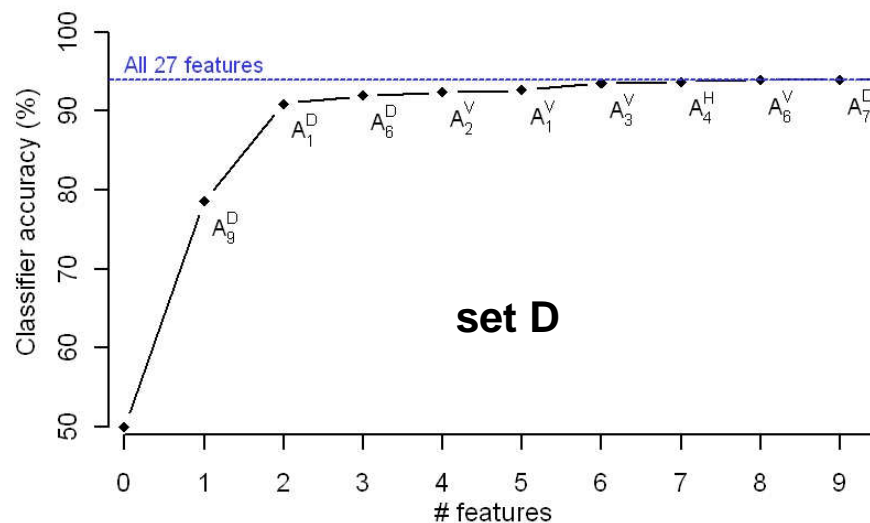
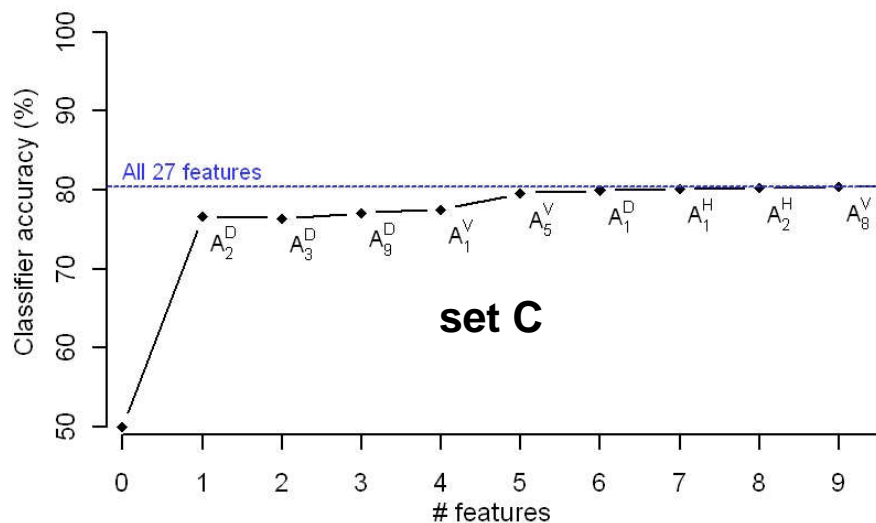
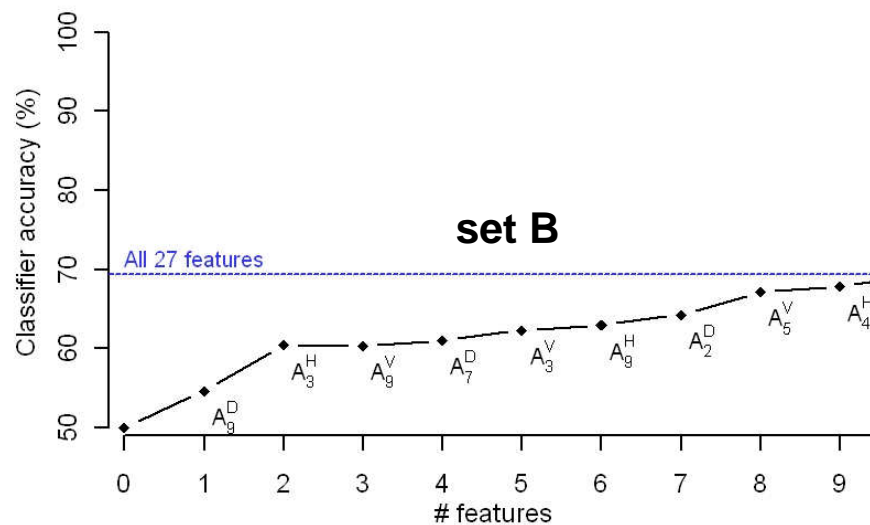
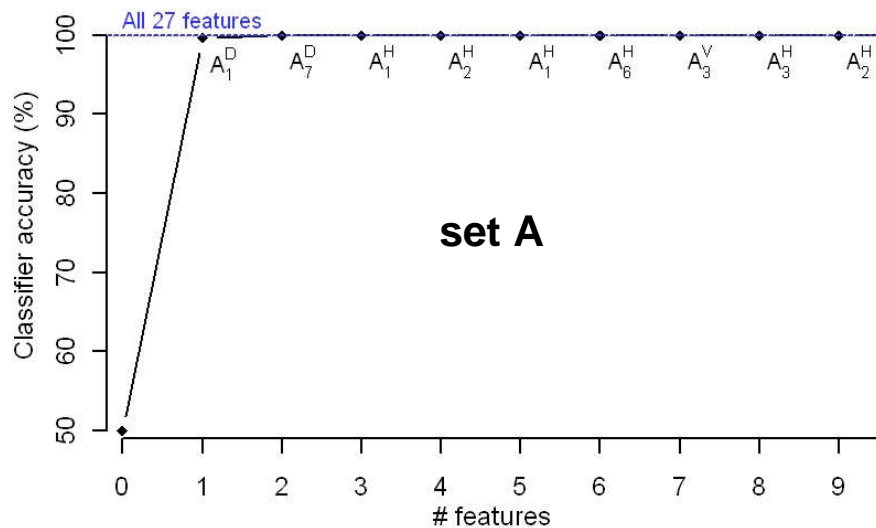
Tried to reduce the feature set using various methods, mainly

- forward selection,
- backward selection,

for each cover set separately.

→ different features for each set of covers!

Feature reduction



Feature reduction

The WAM features cannot be independent: $A_m^H = \frac{1}{|X|} \sum_i |\mathcal{R}[H]_i - \overline{\mathcal{R}[H]}|^m$, etc.

PCA suggests the set of 27 features has only 3-5 independent dimensions.

Tried to reduce the feature set using various methods, mainly

- forward selection,
- backward selection,

for each cover set separately. → *different features for each set of covers!*

Using FLD, tested all combinations of four features, ranked by aggregate score over all cover sets. → *best selection was $\{A_1^D, A_1^H, A_1^V, A_6^D\}$.*

Set	Source	Image noise levels		27 features 4 features		
		in spatial domain	in wavelet domain	FLD	NN	SVM
A	Digital camera <i>never-compressed, pre-processed as grayscale</i>			100 100	100	100 100
B	Digital camera <i>never-compressed, pre-processed as colour</i>			69.7 62.7	73.4	75.8 67.6
C	Various digital cameras <i>never-compressed, unknown pre-processing</i>			80.6 76.2	89.2	90.4 83.2
D	Photo library CD <i>decompressed JPEGs, quality factor 50</i>			95.5 92.1	97.7	97.5 94.3
E	Scanned photos <i>downsampled, never-compressed</i>			60.9 55.5	64.3	64.7 57.1
H	Internet photo sites <i>mixed JPEGs</i>			97.3 91.0	98.0	98.1 93.5

...

Pooled steganalysis

Suppose the steganalyst has N stego objects which contain *different payloads* placed in the *same locations* in *different covers*. There are plausible scenarios in which this could happen.

Can we find the payload locations, which should be more noisy than the others?

WAM residuals live in a transform domain: we need to take them back to the spatial domain.

WAM residuals

1. From input \mathbf{X} , compute 1-level wavelet decomposition:

$$[\mathbf{L}, \mathbf{H}, \mathbf{V}, \mathbf{D}] = \text{DWT}(\mathbf{X})$$

2. The WAM filter gives quasi-Wiener residuals:

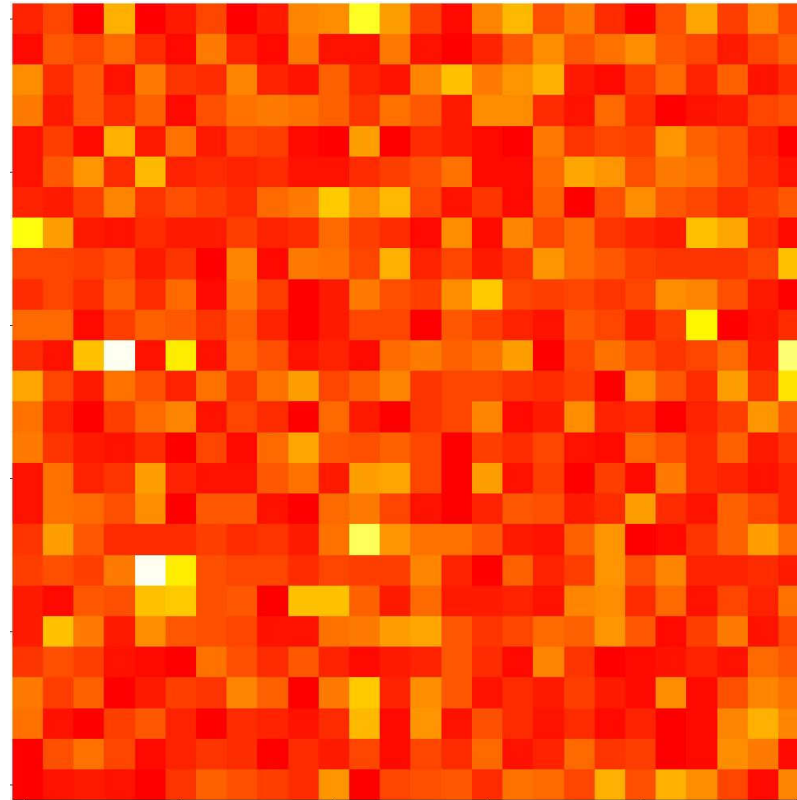
$$\mathcal{R}[\mathbf{S}] = \frac{\sigma_0^2 \mathbf{S}}{\sigma_0^2 + \mathbf{v}} \quad \left(\begin{array}{l} \text{where } \mathbf{v} \text{ is a MAP estimate of local variance based on 4 windows,} \\ \text{and } \sigma_0^2 \text{ is the noise variance, here 0.5} \end{array} \right)$$

3'. Transform filtered residuals back to spatial domain:

$$\mathbf{X}' = \text{DWT}^{-1}(\mathbf{0}, \mathcal{R}[\mathbf{H}], \mathcal{R}[\mathbf{V}], \mathcal{R}[\mathbf{D}])$$

We expect higher absolute residuals in locations containing payload.

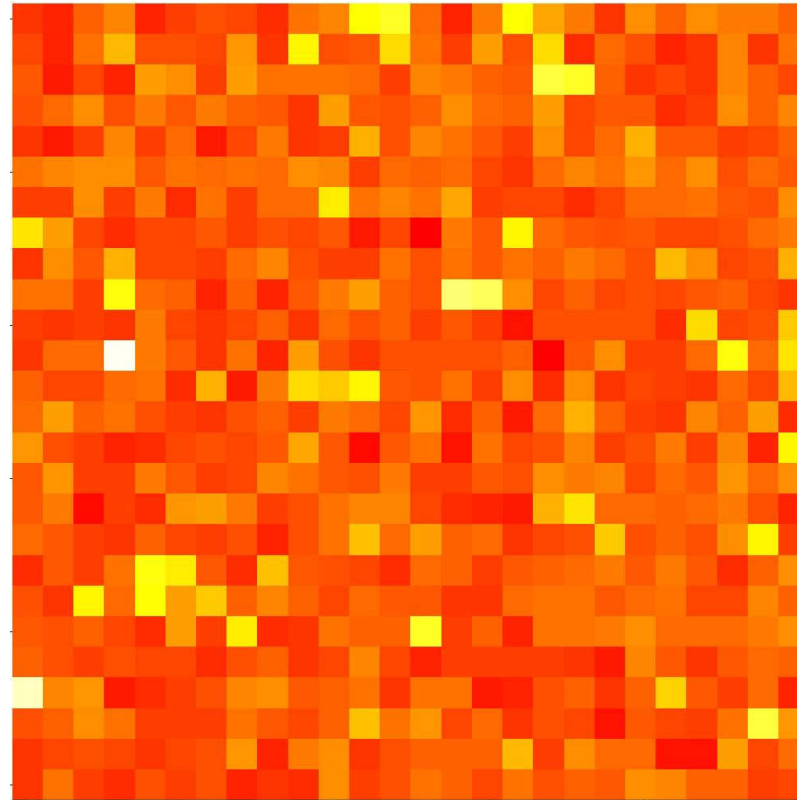
Experimental results



low high

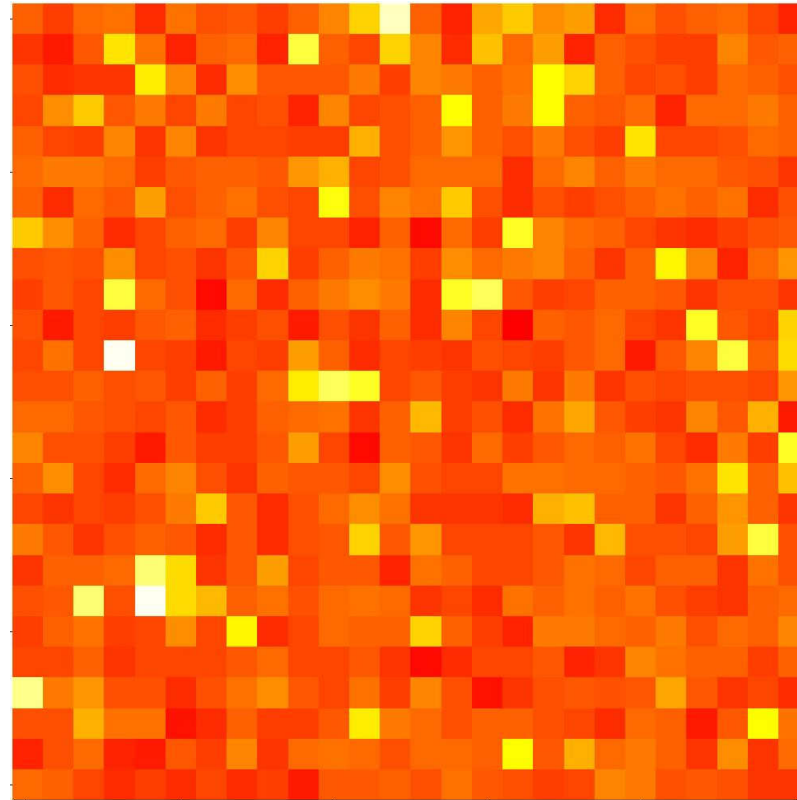
25x25 region, absolute residuals at each pixel , 1 stego image with 10% payload

Experimental results



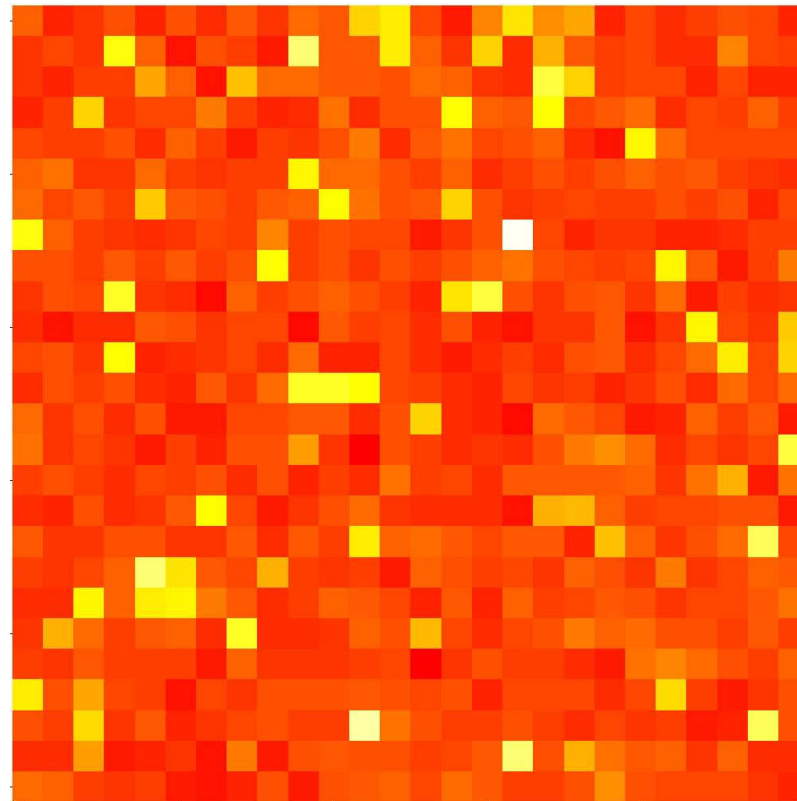
25x25 region, average absolute residuals at each pixel, 10 stego images with 10% payload

Experimental results



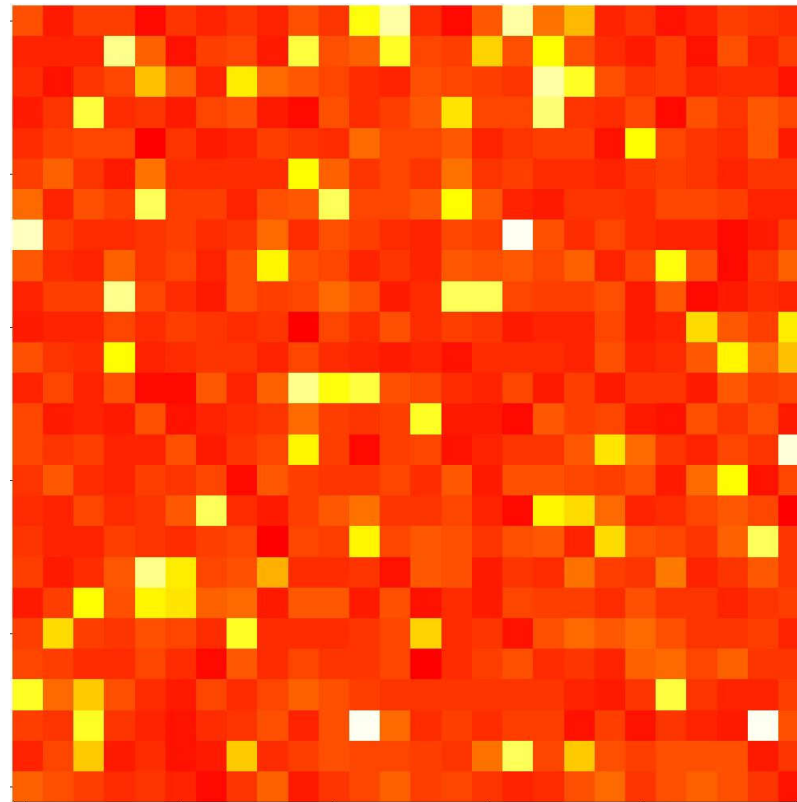
25x25 region, average absolute residuals at each pixel, 20 stego images with 10% payload

Experimental results



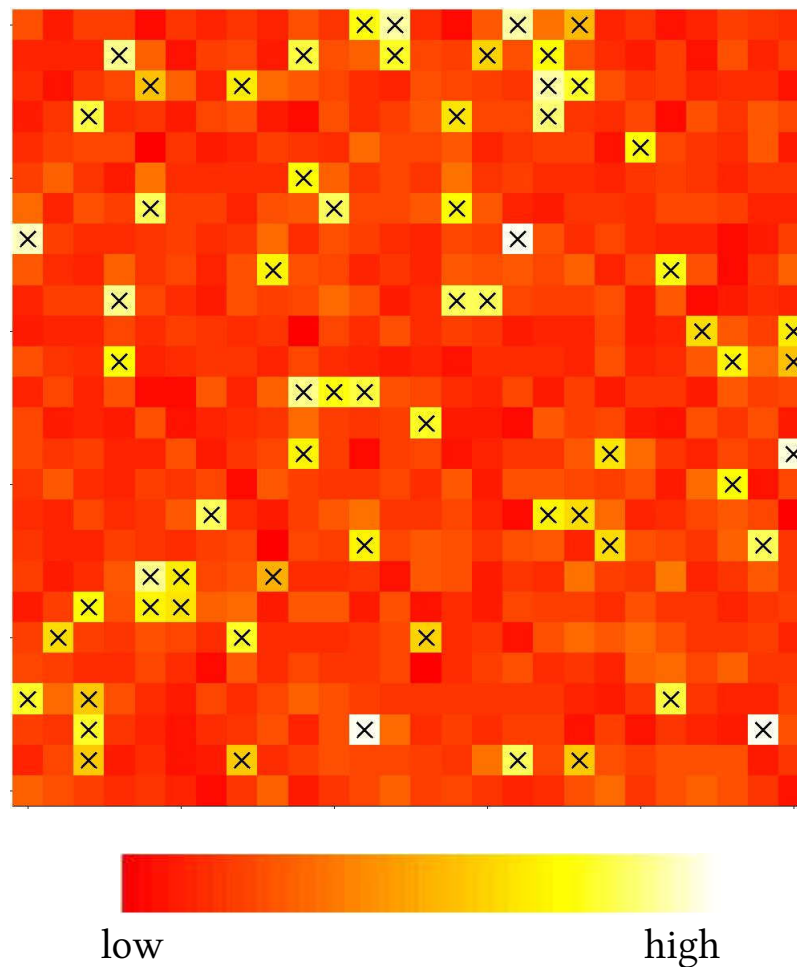
25x25 region, average absolute residuals at each pixel, 50 stego images with 10% payload

Experimental results



25x25 region, average absolute residuals at each pixel, 100 stego images with 10% payload

Experimental results



25x25 region, average absolute residuals at each pixel, 100 stego images with 10% payload
× = payload locations

Experimental results

Payload can be located accurately with enough images:

# stego images	Payload location accuracy (%)			
	Set A	Set B	Set C	Set D
10	84.3	53.6	74.7	64.8
100	99.8	64.8	97.6	93.4
1000	100	82.5	100	100

Conclusions

- Tested WAM features with a three classification engines in nine cover sets.
Moreover, we can measure the statistical significance of differences.
 - *everyone should do this!*
- Just like other LSB matching detectors, WAM works very well sometimes, and its feature set can be reduced with little loss in power.
But we cannot predict when it will work and when it will not, and the reduced feature set depends on unknown cover properties.
 - *an avenue for further research.*
- Converting WAM residuals to spatial domain, and averaging, allows us to estimate payload location, given enough stego images with payload in the same locations.
This demonstrates why steganographic embedding keys must not be re-used.