Estimating Steganographic Fisher Information in Real Images

Andrew Ker

adk@comlab.ox.ac.uk

Royal Society University Research Fellow
Oxford University Computing Laboratory

11th Information Hiding Workshop
Darmstadt, 8 June 2009

Estimating Steganographic Fisher Information in Real Images

Outline

- *Background: capacity*
- *Steganographic Fisher Information*
- [*Constructing an estimator*]
- *Illustrative results*

Capacity

The capacity problem is: given fixed

- cover source,
- embedding method,
- limit on “risk” (maximum probability of detection),

what is the largest payload which can safely be embedded?

The square root law says:

the capacity is asymptotically proportional to the square root of the size of the cover.

- Proved for multiple independent covers (Ker, 2007; Ker, 2008).
- Proved for individual Markov chain covers (Filler, Ker, & Fridrich, 2009).
- Verified empirically (Ker, Pevný, Kodovský, & Fridrich, 2008).

Capacity

The capacity problem is: given fixed

- cover source,
- embedding method,
- limit on "risk" (maximum probability of detection),

what is the largest payload which can safely be embedded?

The square root law says:

the capacity is asymptotically proportional to the square root of the size of the cover.

If the cover size is n, the max payload size m follows

$$m \sim r\sqrt{n},$$

where r is the "root rate". Can we determine the root rate?

Fisher Information

We could try to calculate capacity from $\mathrm{D_{KL}}(\mathrm{cover\ images},\ \mathrm{stego\ images})$.

Can we estimate this empirically, from real cover & stego images?

No!
- KL divergence is notoriously difficult to estimate.
- The dimensionality is huge.

Fisher Information

Theorem

If $\{P(\lambda) \mid \lambda \in [0, l]\}$ is a family of distributions (satisfying some regularity conditions), as $\lambda \to 0\,,\; D_{\text{KL}}(P(0) \,\|\, P(\lambda)) \sim \frac{1}{2} I \lambda^2$.

I is the **Fisher Information** for λ.

If $P(\lambda)$ is distribution of images with payload rate λ, then I — the **Steganographic Fisher Information (SFI)** for the family — determines the asymptotic root rate r.

A. Ker. *The Ultimate Steganalysis Benchmark?* Proc. ACM Workshop on Multimedia & Security, 2007.

Fisher Information

Theorem

If $\{P(\lambda) \mid \lambda \in [0, l]\}$ is a family of distributions (satisfying some regularity conditions), as $\lambda \to 0$, $D_{\mathrm{KL}}(P(0) \parallel P(\lambda)) \sim \frac{1}{2} I \lambda^2$.

I is the **Fisher Information** for λ.

If $P(\lambda)$ is distribution of images with payload rate λ, then I — the **Steganographic Fisher Information (SFI)** for the family — determines the asymptotic root rate r.

[SFI must be properly scaled for the embedding efficiency of the embedding and the size of the cover. The scaled version is measured in **symbol nats per bit squared.]**

Fisher Information

Theorem

If $\{P(\lambda) \mid \lambda \in [0, l]\}$ is a family of distributions (satisfying some regularity conditions), as $\lambda \to 0$, $D_{\mathrm{KL}}(P(0) \parallel P(\lambda)) \sim \frac{1}{2} I \lambda^2$.

I is the **Fisher Information** for λ.

If $P(\lambda)$ is distribution of images with payload rate λ, then I — the **Steganographic Fisher Information (SFI)** for the family — determines the asymptotic root rate r.

Thus ***SFI is a measure of evidence*** about the presence of steganography.

Higher SFI corresponds to

- higher KL divergence,
- more accurate detectors,
- lower root rate.

Can we estimate SFI empirically, from real cover & stego images?

Independent pixel groups

The dimensionality of images is still impossibly huge. One solution is:

The independent pixel group model

Model a image as a collection of independent small groups of pixels, e.g. single pixels, pairs, 2x2 blocks, …

This is permissible because almost all steganalysis methods only consider aggregate data from small pixel groups:

Detector	**can be expressed in terms of**
chi-square steganalysis	histogram
sample pairs steganalysis	adjacency matrix
WS steganalysis	frequency of local 3x3 or 5x5 groups
calibrated HCF/COM	frequency of 2x2 or 4x4 groups
most JPEG detectors	frequency of 8x8 or 16x16 blocks
…	…

Related ideas

Q-factor

- Equivalent to Steganographic Fisher Information but unscaled.
- Proposed as a benchmark for **steganalysis** in 2007.
- Focus on steganalysis gives low dimensionality.

Maximum Mean Discrepancy (MMD)

- Another information-theoretic measure of evidence
- Proposed as a benchmark for steganography by Pevný & Fridrich at the last Information Hiding Workshop.
- Focus on features means moderate dimensionality.

Estimating SFI

For groups of n pixels the suitably scaled SFI can be derived as

$$SFI = \frac{1}{2ne^2}\left[\sum_{\boldsymbol{x}\in\mathcal{X}^n} Q(\boldsymbol{x})^2 P(\boldsymbol{x})^{-1} - n^2\right]$$

where

e is the embedding efficiency (payload bits per change).

$P(\boldsymbol{x})$ is the probability of observing the group $\boldsymbol{x}$ in cover objects.

$Q(\boldsymbol{x})$ is the probability of observing the group $\boldsymbol{x}$ in stego objects where exactly one element was changed by embedding.

We can estimate it by

- *finding a large corpus,*
- *computing the empirical histogram of pixel groups,*
- *plugging in the empirical histogram for $P(x)$,*
- *deriving $Q(x)$ using the embedding function.*

Implementation

Computing a histogram of pixel groups is not always easy:

group size n $\longrightarrow$ *potentially* 256^n *histogram bins*

(assuming 8-bit greyscale images).

Solution:

- Red-black trees to store partial histograms.
- Shuffle-merging of partial histograms.

Given an embedding function,

- adjoining of $Q(\boldsymbol{x})$ to each $P(\boldsymbol{x})$ entry by binary search.

Real images

We use a corpus of cover images:

- 2118 never-compressed images,
- about 4.5M pixels each,
- taken with the same digital camera,
- saturated images excluded,
- some denoising in conversion from RAW to greyscale bitmap.

Re-using each image in four orientations, the total evidence base is approx. $4 \cdot 10^{10}$ groups.

Nontrivial computational demands:

- histograms (up to 3x3 pixel groups) total 630GB in size,
- took 6 CPU weeks to compute SFI, using a small cluster of 12 machines.

Compare embedding functions

Embedding methods cause different types of distortion and have varying embedding efficiencies.

Which is better?

Recall, lower SFI corresponds to

- lower asymptotic KLD,
- less evidence of steganography,
- higher root rate,

i.e. better embedding.

Compare embedding functions

Embedding methods cause different types of distortion and have varying embedding efficiencies.

Which is better?

	SFI estimate for		
Group shape	LSB replacement	LSB matching	2LSB replacement
□	0.000207	0.0000275	0.000332
□□	0.0968	0.0309	0.159
□□□	0.330	0.155	1.24
□□□□	0.973	0.563	12.5
□□□□□	2.02	1.21	153
□□□□□□	3.13	1.86	209

Pixel difference

Many steganalysis methods consider only pixel **difference,** essentially discarding DC information of each pixel group.

Some preserve parity information.

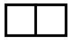
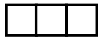
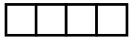
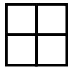
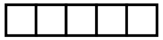
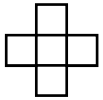
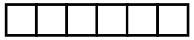
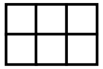
Is this a sensible choice?

	SFI estimate for		
Embedding Function	raw pixels	difference + (mod 2) information	difference only
LSB replacement	0.0968	0.0831	0.0233
LSB matching	0.0309	0.0233	0.0233

pixel group shape □□ in all cases

Compare pixel groups

What shape pixel group carries the most evidence of LSB replacement embedding?

Group shape	SFI estimate
	0.000207
	0.0968
	0.330
	0.973
	0.470
	2.02
	0.497
	3.13
	1.37

Conclusions

Steganographic Fisher Information is a measure of ***evidence***.

Its empirical estimation allows some fundamental comparisons:

- of embedding methods,
- of cover sources,
- of detector limitations imposed by considering different types of pixel group.

So far, we learned lessons about:

- ***relative security of LSB/2LSB replacement & LSB matching embedding,***
- ***need to preserve pixel parity if reducing to pixel difference,***
- ***relative evidence in pixel groups of different shapes.***

We can only go as far as groups of 8 or maybe 9 pixels, unless the image corpus is enormous. We need find a better estimator for SFI.