Improved Detection of LSB Steganography in Grayscale Images

Andrew Ker

adk@comlab.ox.ac.uk

Royal Society University Research Fellow at
Oxford University Computing Laboratory

Information Hiding Workshop 2004

Summary

This presentation will tell you about:

1. A project to **evaluate** the reliability of steganalytic algorithms;

2. Some potential pitfalls in this area;

3. Improved steganalysis methods:

   exploiting uncorrelated estimators,

   simplifying, by dropping the message length estimate,

   (applying discriminators to a segmented image);

4. Experimental evidence of improvement.

"Reliability"

The primary aim of an Information Security Officer (Warden) is to perform a reliable hypothesis test:

H_0: No data is hidden in a given image
H_1: Data is hidden (for experiments we posit a fixed amount/proportion)

(as opposed to forming an estimate of the amount of hidden data, or recovering the hidden data)

A steganalysis method is a discriminating statistic for this test; by adjusting the sensitivity of the hypothesis test, false positive (type I error) and false negative (type II error) rates may be traded.

Reliability is a "ROC" curve showing how false positives and false negatives are related.

Distributed Steganalysis Evaluation Project

Applied systematically

Over 200 variants of steganalysis statistics tested so far

Very large image libraries are used

Currently over 90,000 images in total, with more to come

Images come in “sets” with similar characteristics.

Results are produced quickly

Computation performed by a heterogeneous cluster of 7-50 machines

Calculations queued and results stored in a relational database

Currently over 16 million rows of data, will grow to 100+ million

Scope of This Work

Covers

Grayscale bitmaps
(which quite likely were previously subject to JPEG compression)

Embedding method

LSB steganography in the spatial domain using various proportions of evenly-spread pixels

Particular interest in very low embedding rates
(0.01-0.1 secret bits per cover pixel)

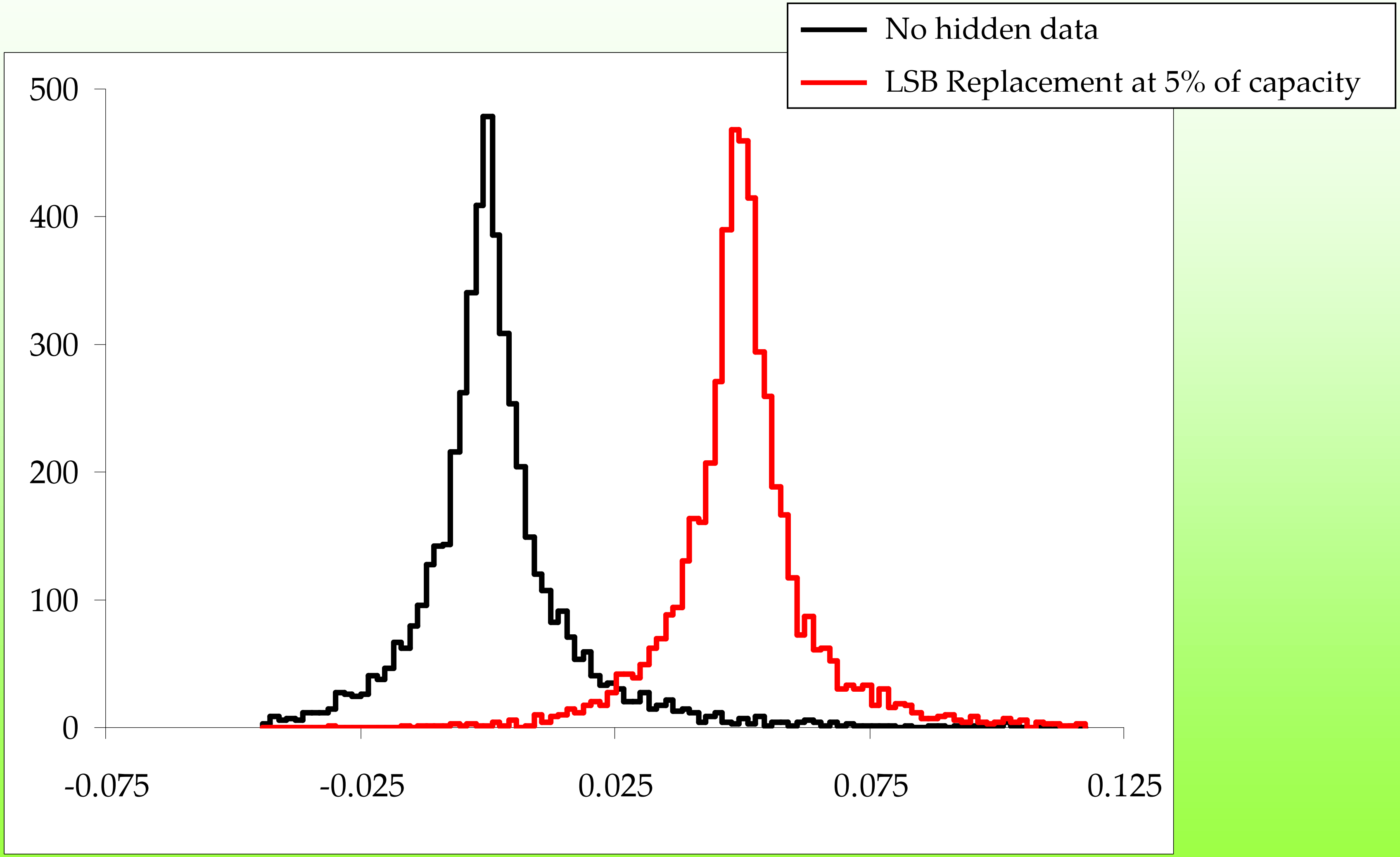
Aiming to improve the closely-related steganalysis statistics

"Pairs" [Fridrich *et al*, SPIE EI'03]
"RS" a.k.a. "dual statistics" [Fridrich *et al*, ACM Workshop '01]
"Sample Pairs" [Dumitrescu *et al*, IHW'02] a.k.a. "Couples"

The world's smallest steganography software

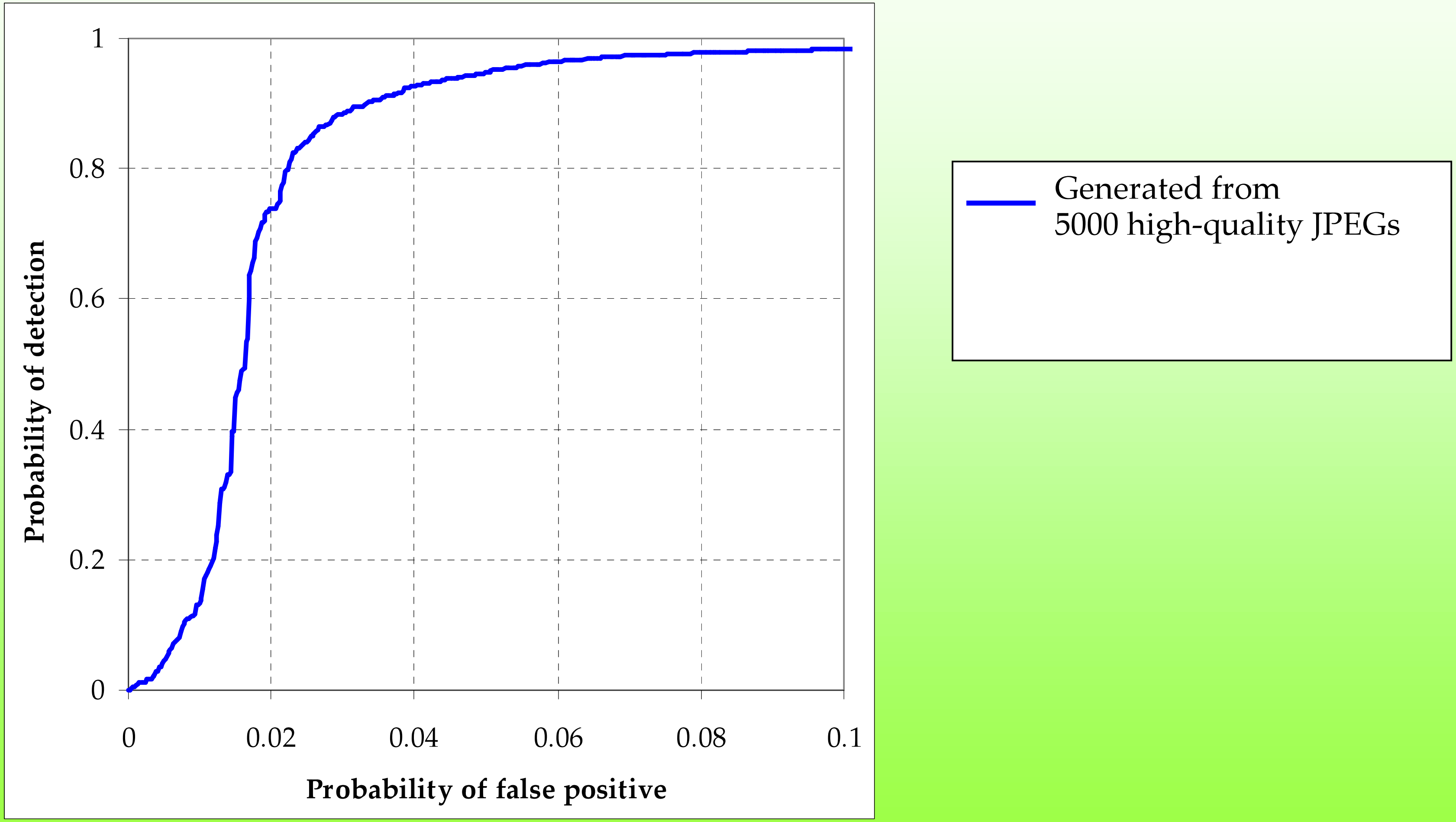
```
perl -n0777e '$_=unpack"b*",$_;split/(\s+)/,<STDIN>,5;
              @_[8]=~s{.}{$&&v254|chop()&v1}ge;print@_'
              <input.pgm >output.pgm stegotext
```

Sample Output: Histograms



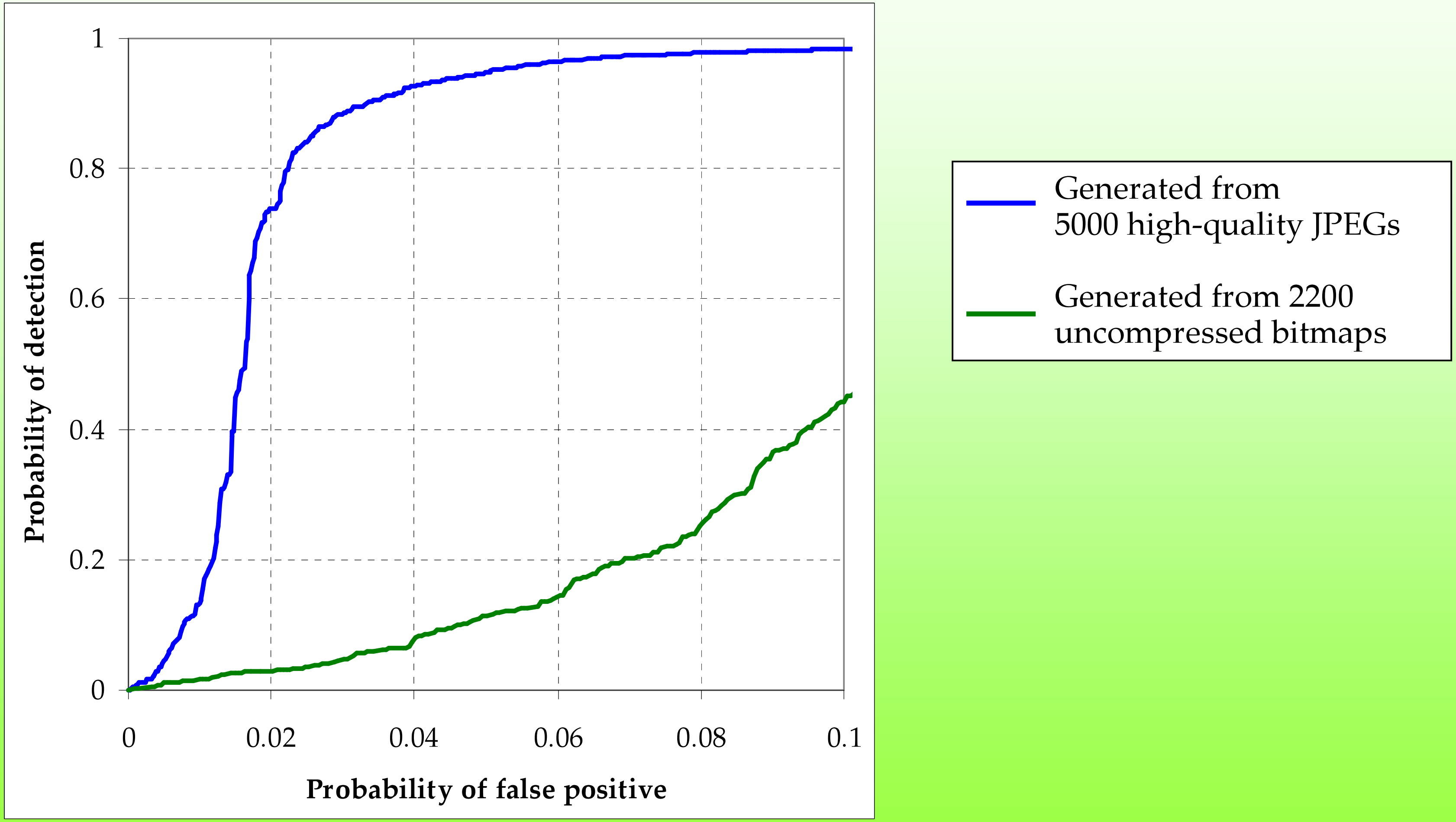
Histograms of the standard "Couples" statistic, generated from 5000 JPEG images

Sample Output: ROC Curves



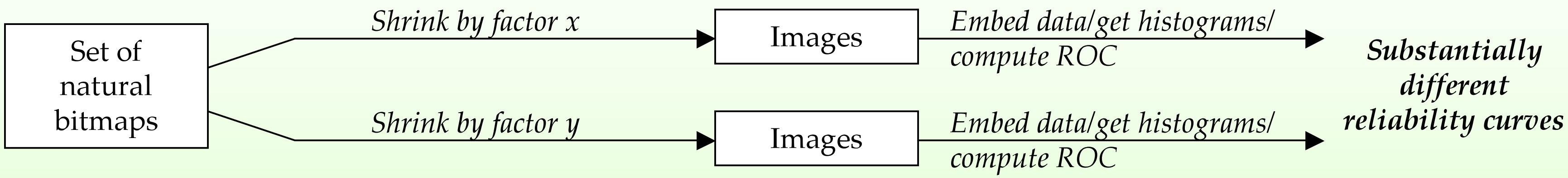
ROC curves for the "Couples" statistic. 5% embedding (0.05bpp).

Sample Output: ROC Curves



ROC curves for the "Couples" statistic. 5% embedding (0.05bpp).

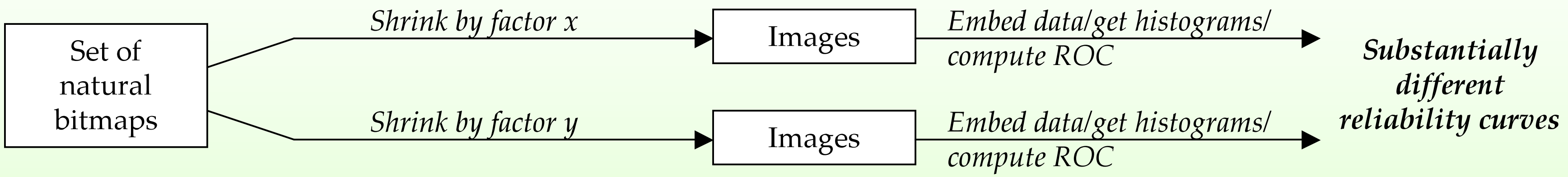
Some Warning Examples



Conclusion

- The size of the cover images affects the reliability of the detector, even for a fixed embedding rate

Some Warning Examples



Conclusion

- The size of the cover images affects the reliability of the detector, even for a fixed embedding rate.

In [Ker, SPIE EI'04] we also showed that

- Whether and how much covers had been previously JPEG compressed affects reliability, sometimes a great deal.
- This effect persists even when the images are quite substantially shrunk after compression.
- Different resampling algorithms in the shrinking process can themselves affect reliability.

Good Methodology for Evaluation

- We have to concede that there is no single "reliability" for a particular detector.
- One should test reliability with more than one large set of cover images.
- It is important to report:
    a. How much data was hidden;
    b. The size of the covers;
    c. Whether they have ever been JPEG compressed, or undergone any other manipulation.
- Take great care in "simulating" uncompressed images.

How does “Couples Analysis” work?

Simulate LSB replacement in proportion $2p$ of pixels by flipping the LSBs of p at random.

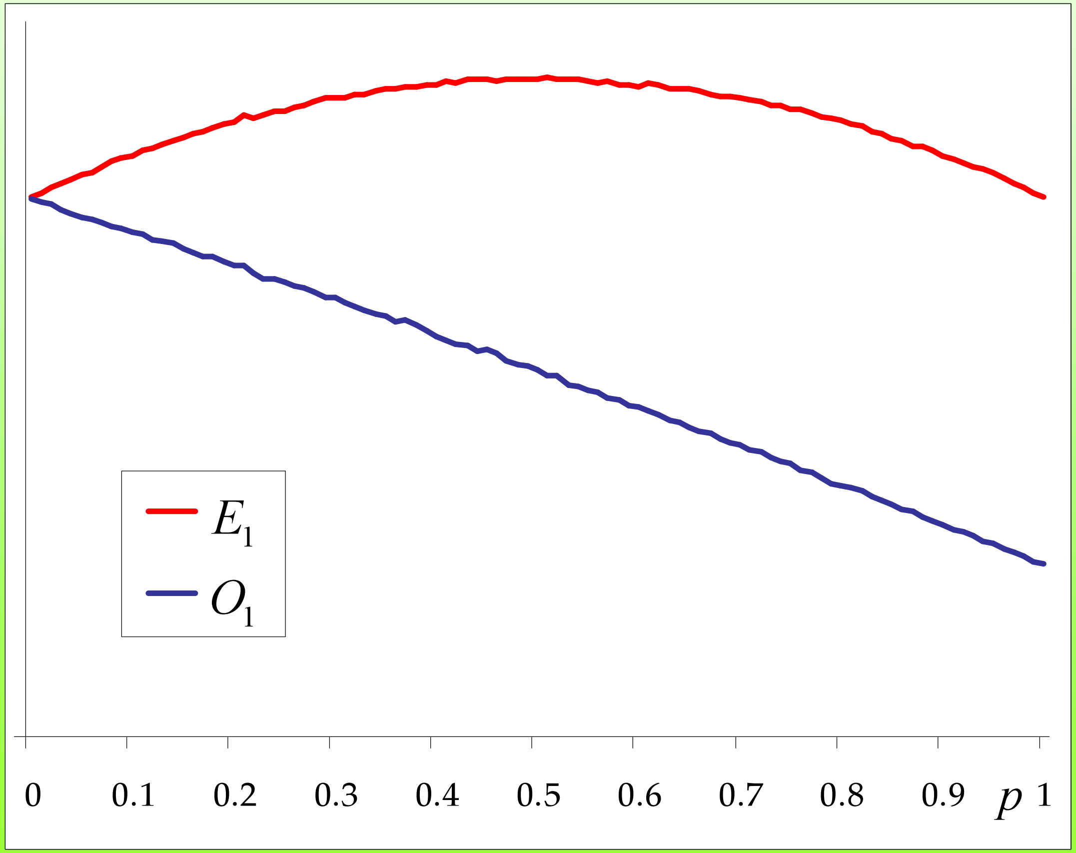
Example cover image:

How does "Couples Analysis" work?

As p varies, compute:

E_i = number of adjacent pixels whose value differs by i, and the lower value is even

O_i = number of adjacent pixels whose value differs by i, and the lower value is odd



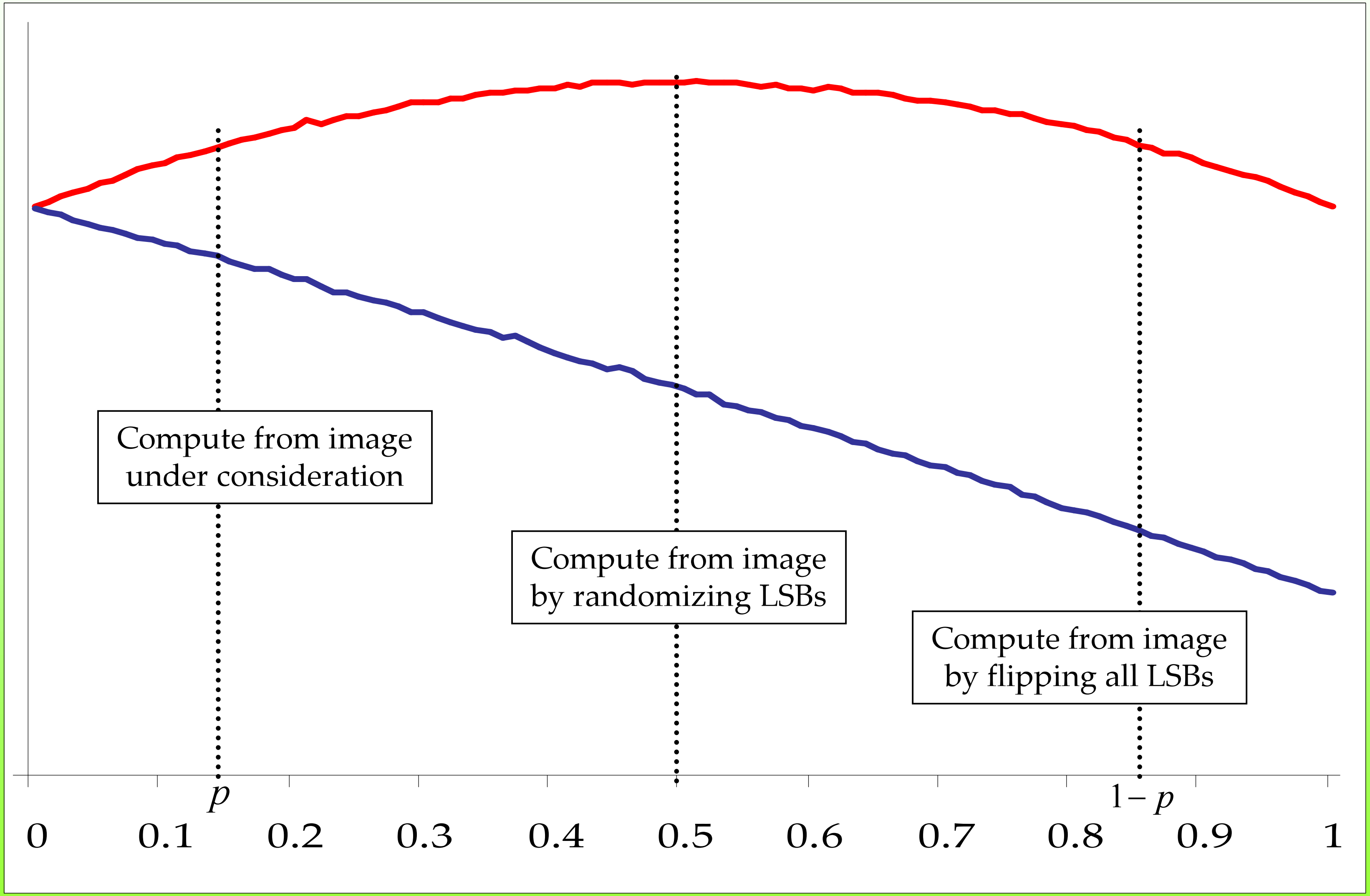
- Both curves quadratic in p
- Meet at p=0

The pairs of measures

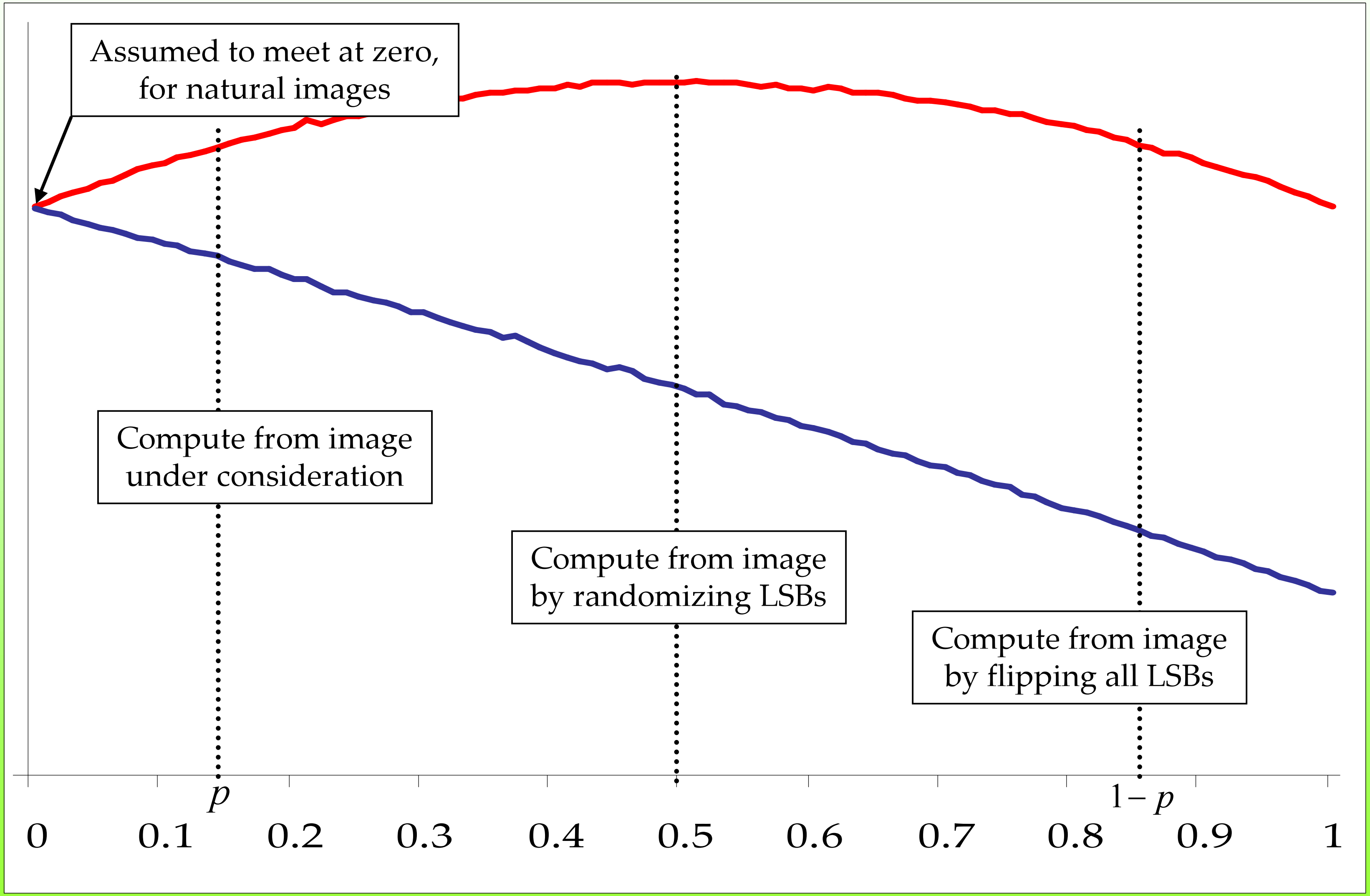
$$E_3 \;\&\; O_3$$
$$E_5 \;\&\; O_5$$
$$\vdots$$
$$\sum_{odd\ i} E_i \;\&\; \sum_{odd\ i} O_i$$

all have the same properties.

How does “Couples Analysis” work?



How does “Couples Analysis” work?



Choice of Discriminators

Unlike Pairs and RS, Couples has a number of estimators for the proportion of hidden data:

$\hat{p}_0$ from E_1 and O_1

$\hat{p}_1$ from E_3 and O_3

$\hat{p}_2$ from E_5 and O_5

$\vdots$

$\hat{p}$ from $\sum_{odd\ i} E_i$ and $\sum_{odd\ i} O_i$

The last one is used in [Dumitrescu *et al*, IHW'02]

Choice of Discriminators

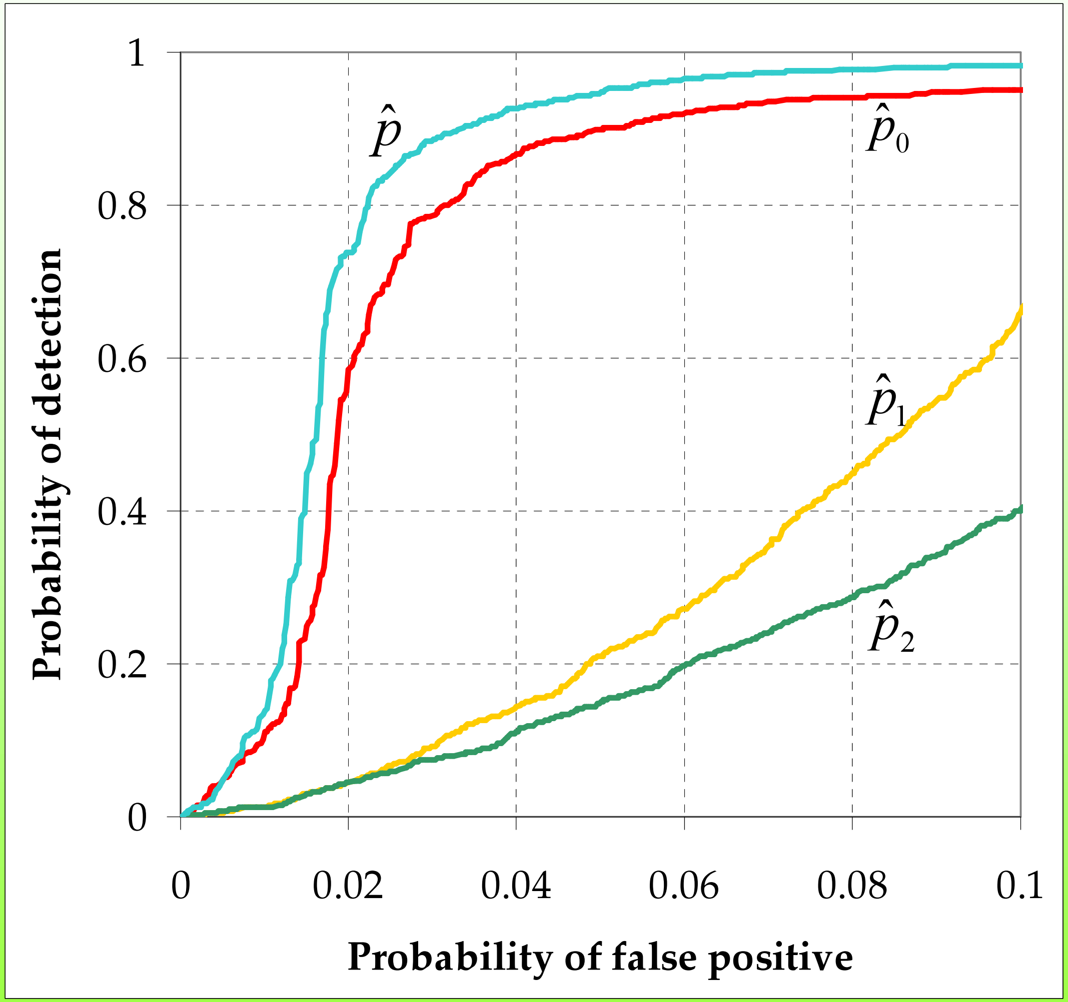
$\hat{p}_0$ from E_1 and O_1

$\hat{p}_1$ from E_3 and O_3

$\hat{p}_2$ from E_5 and O_5

$\vdots$

$\hat{p}$ from $\sum_{odd\ i} E_i$ and $\sum_{odd\ i} O_i$



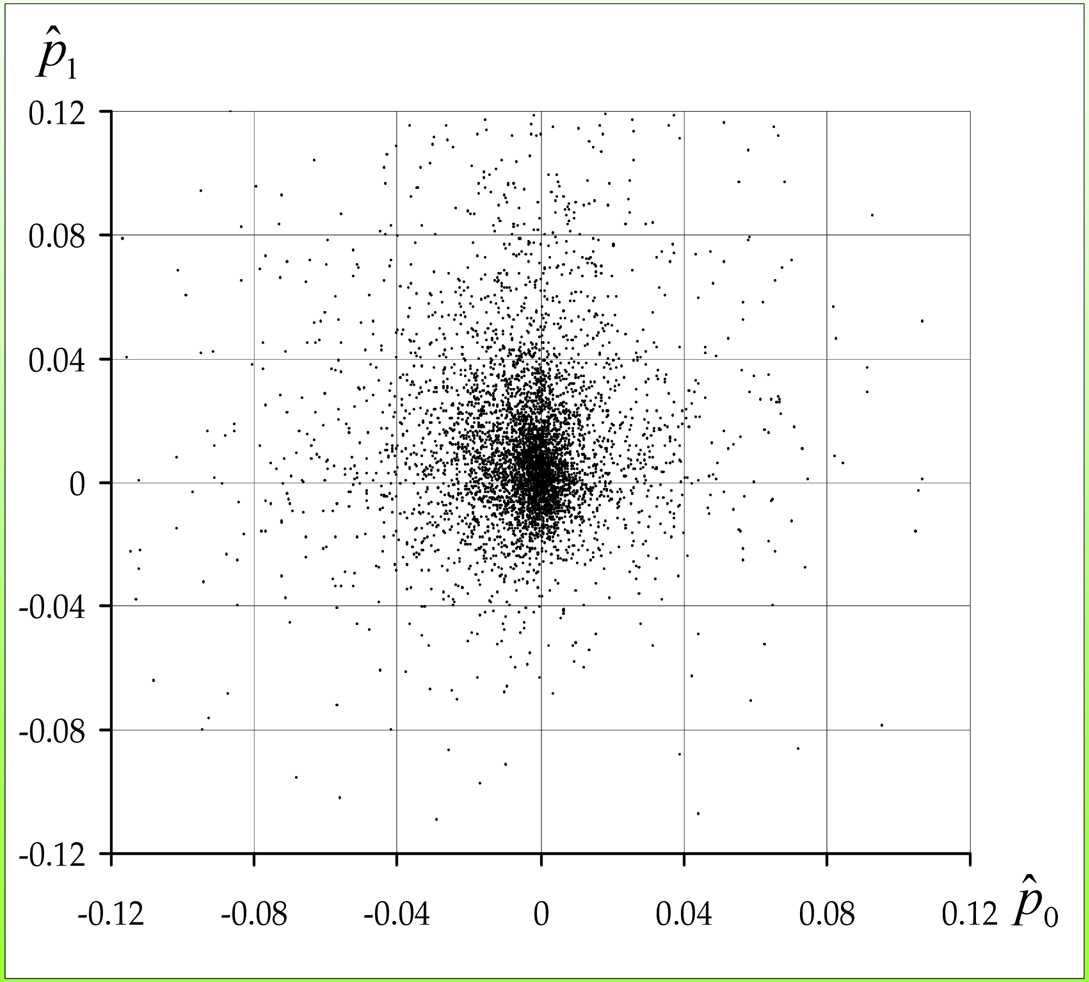
ROC curves generated from 5000 JPEG images of high quality. 5% embedding (0.05bpp).

Estimators are Uncorrelated

We observe that the estimators $\hat{p}_i$ are very loosely correlated.

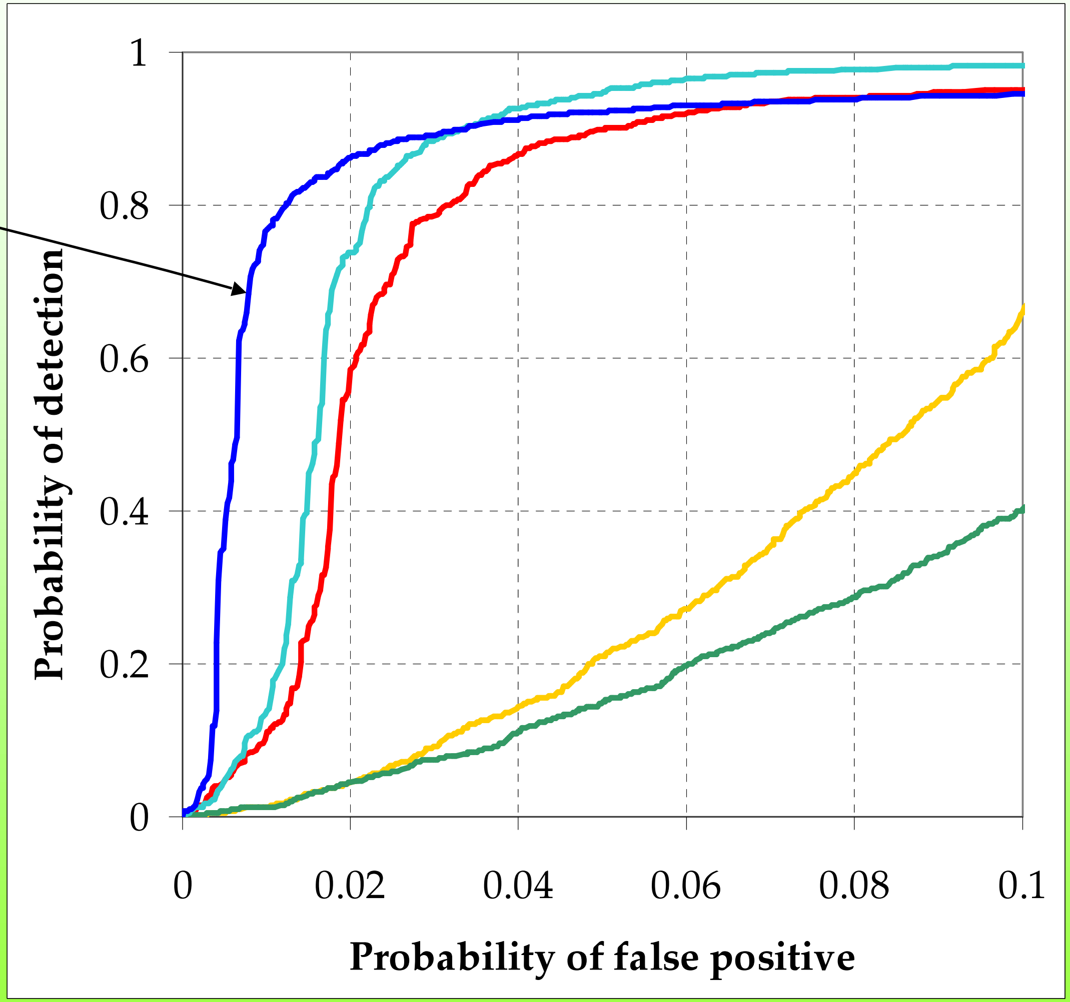
Scattergram shows $\hat{p}_0$ & $\hat{p}_1$ when no data embedded in 5000 high-quality JPEG images; the correlation coefficient is **-0.036**

$\hat{p}_0$ & $\hat{p}_1$ form independent discriminators



Improved Couples Discriminator

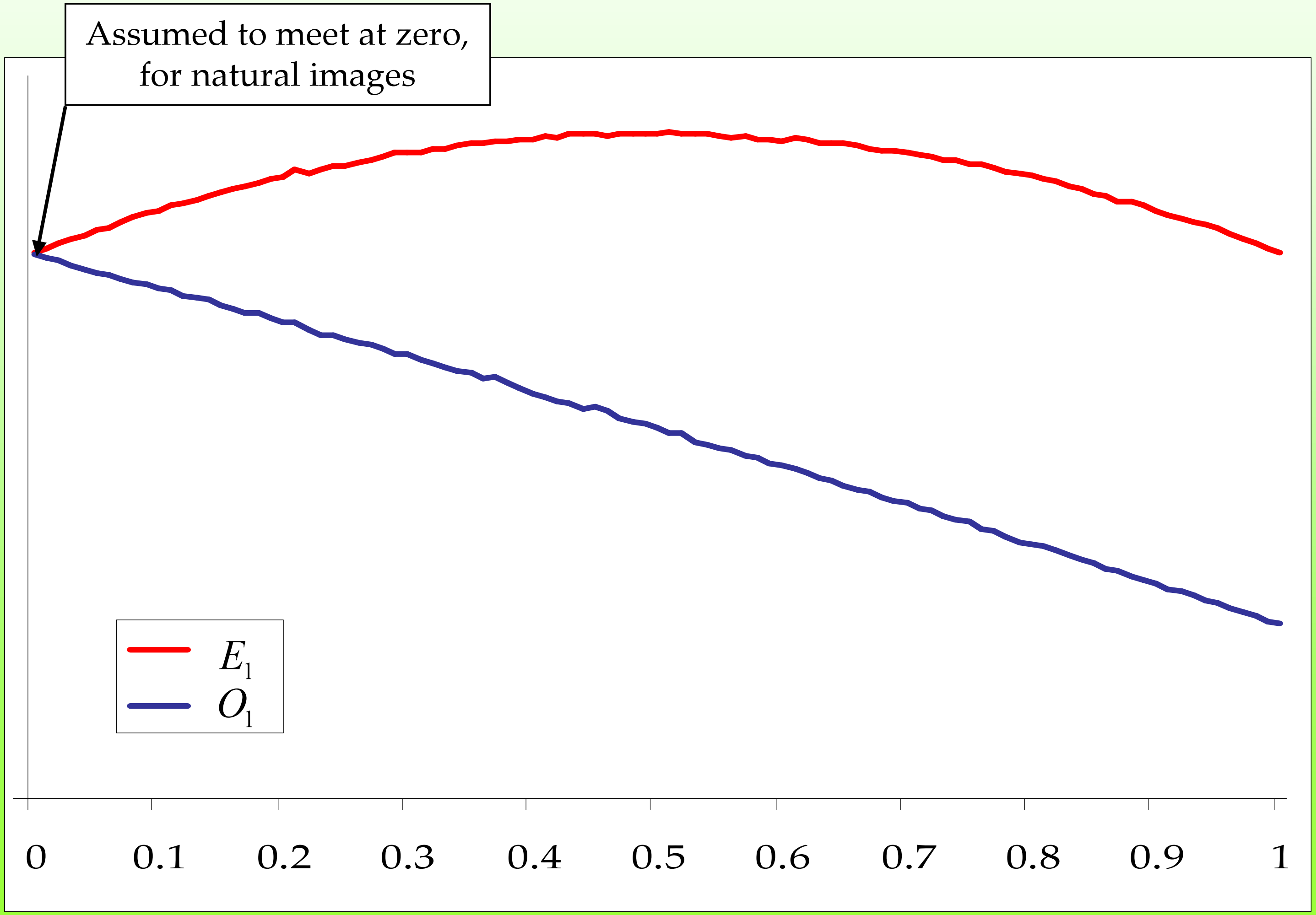
$\min(\hat{p}_0, \hat{p}_1, \hat{p}_2)$



ROC curves generated from 5000 JPEG images of high quality. 5% embedding (0.05bpp).

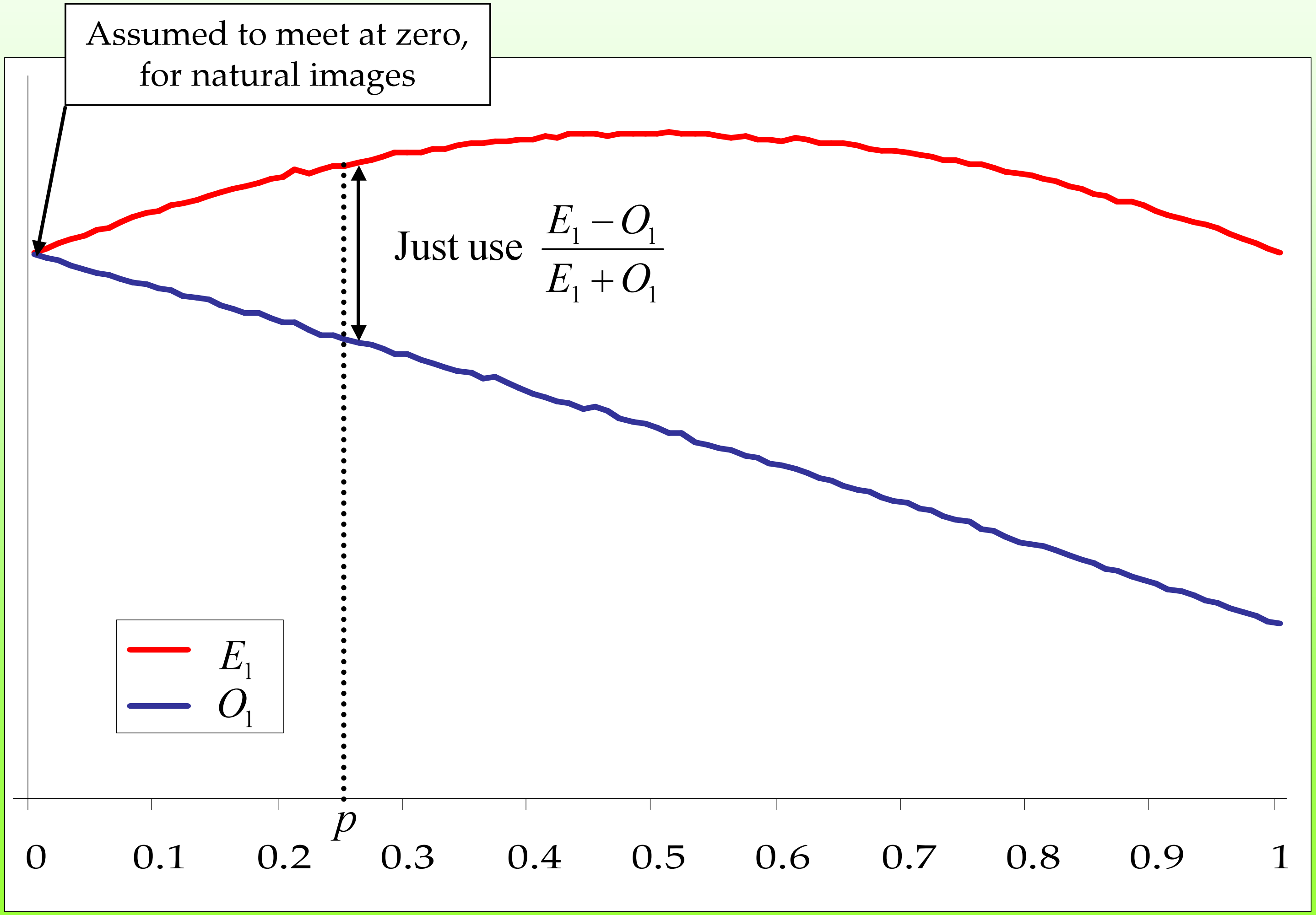
Dropping the Message-Length Estimate

There is a much simpler sign that data has been embedded, which does not involve solving a quadratic equation:

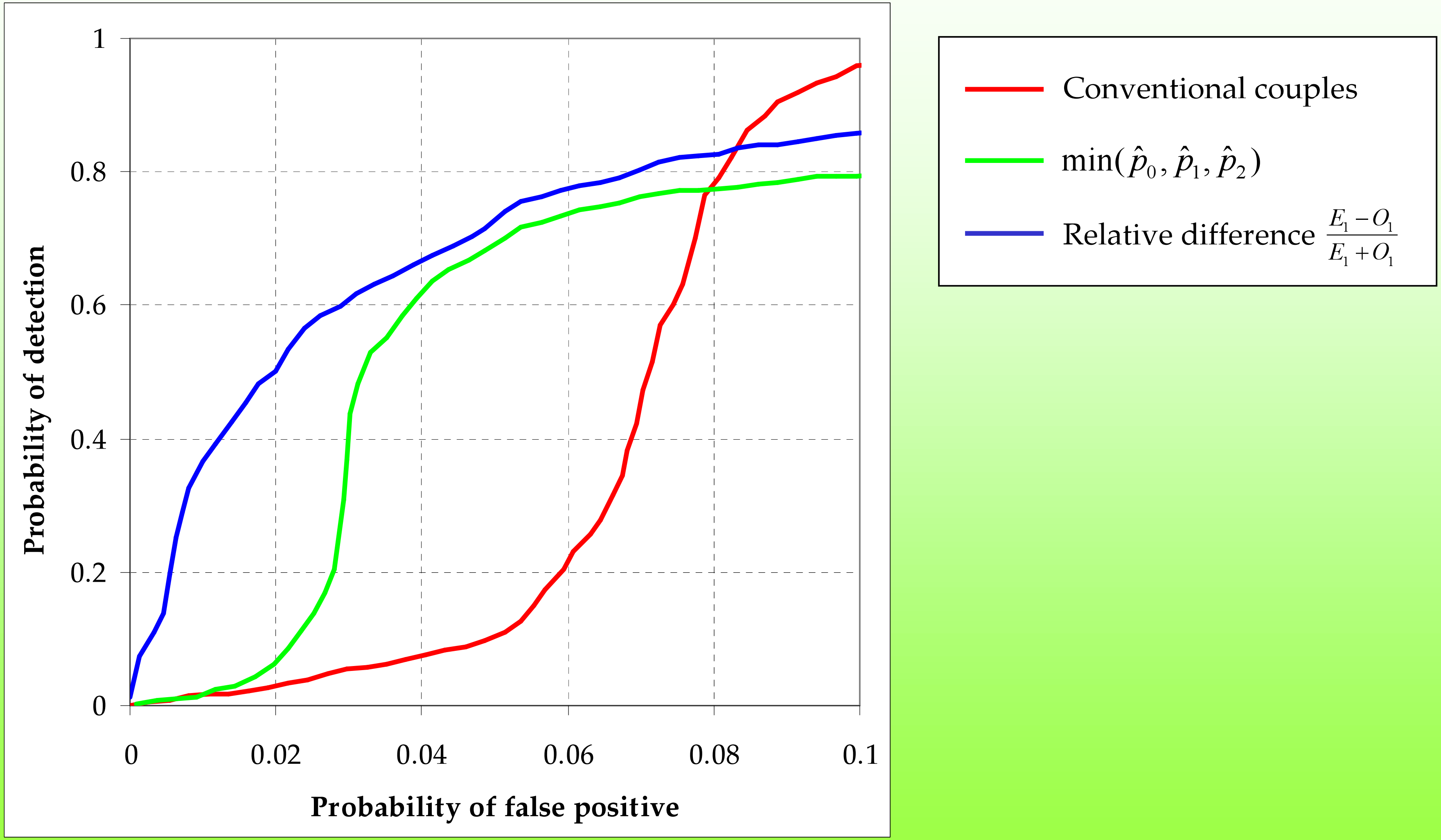


Dropping the Message-Length Estimate

There is a much simpler sign that data has been embedded, which does not involve solving a quadratic equation:



Dropping the Message-Length Estimate



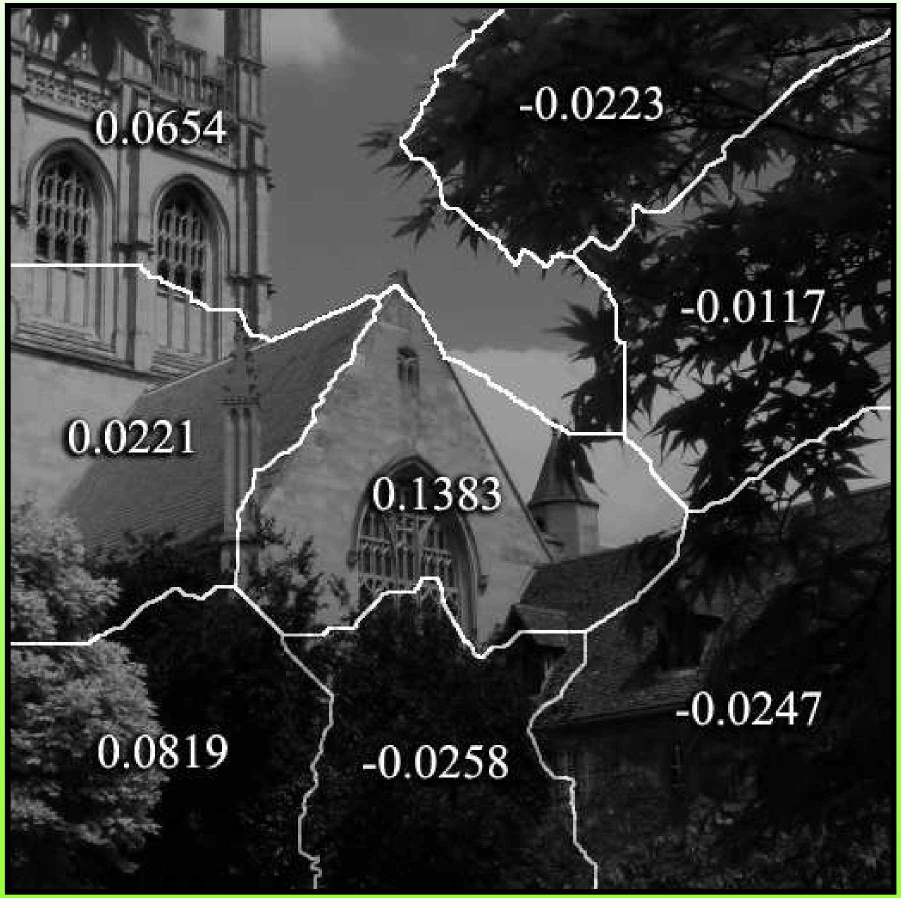
ROC curves generated from 15000 mixed JPEG images, 3% embedding.

Splitting into Segments

Using the standard RS method this image, which has no hidden data, estimates an embedding rate of 6.5%.

Splitting into Segments

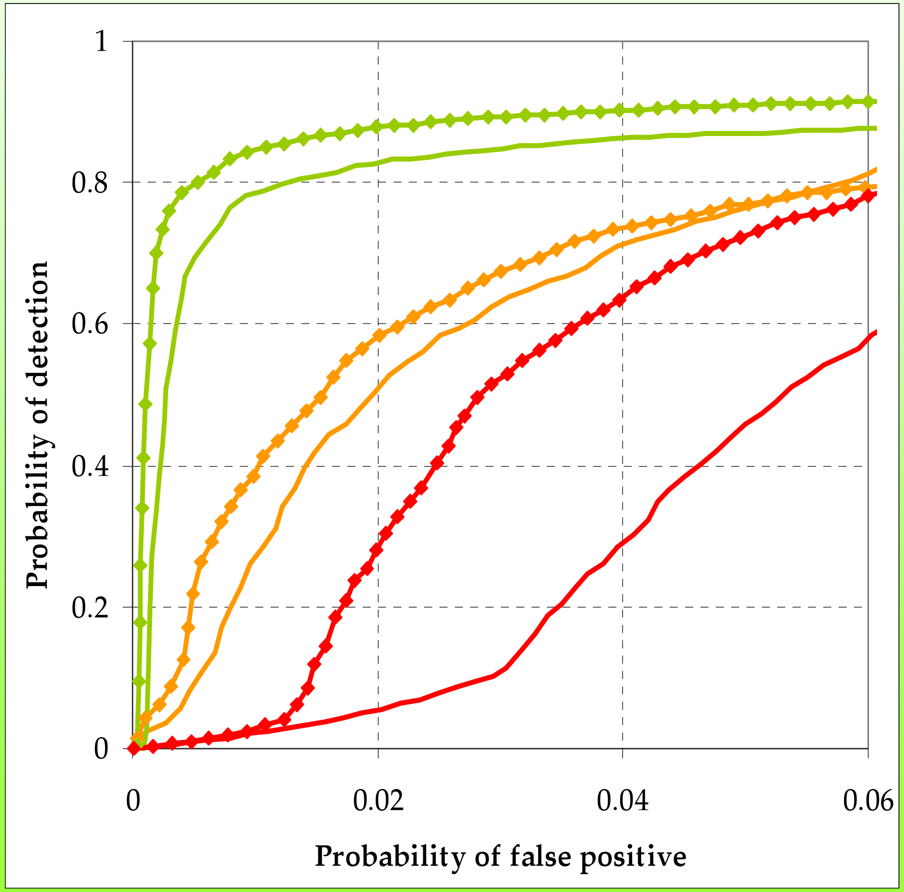
Segment the image using the technique in [Felzenszwalb & Huttenlocher, IEEE CVPR '98] and compute the RS statistic for each segment.



Taking the median gives a more robust estimate, in this case of 0.5%.

Result of Segmenting

Segmenting is a "bolt on" which can be added to any other estimator. Here, to the modified RS method which computes the relative difference between *R* and *R'* (analogous to E_1 and O_1).



ROC curves from three image sets. 3% embedding.

- 10000 low quality JPEGs
- 5000 high quality JPEGs
- 7500 very mixed JPEGs

Marked curves are the segmenting versions (taking the 30% percentile of per-segment statistics)

Experimental Evidence of Improvements

We have computed very many ROC curves which depend on:

- which cover image set was used;
- (if not JPEG compressed already) how much JPEG pre-compression applied;
- how much data was hidden;
- which detection statistic is used as a discriminator.

There are too many curves. The database of statistic computations is 4.3Gb!

… How to display all this data?

We make an arbitrary decision that a "reliable" statistic is one which makes false positive errors at less than 5% when false negatives are 50%.

For each statistic and image set display the lowest embedding rate at which this reliability is achieved.

Lowest embedding rate for which 50% false negatives achieved with no more than 5% false positives:

Conventional Pairs	[Fridrich *et al*, SPIE EI'03]
Conventional RS	[Fridrich *et al*, ACM Workshop '01]
Conventional Couples	[Dumitrescu *et al*, IHW'02]
RS w/ optimal mask	[Ker, SPIE EI'04]
Improved Pairs	[Ker, SPIE EI'04]
Improved Couples $\min(\hat{p}_0, \hat{p}_1, \hat{p}_2)$	Presented here
Relative difference of E_1 & O_1 *(using non-overlapping pixel groups)*	Presented here
Relative difference of R, R' *(using optimal mask and non-overlapping pixel groups and segmenting the image into 6-12 groups, taking 30th percentile of the per-segment statistics)*	Presented here

Lowest embedding rate for which 50% false negatives achieved with no more than 5% false positives:

	2200 bitmaps
Conventional Pairs	10%
Conventional RS	11%
Conventional Couples	9%
RS w/ optimal mask	10%
Improved Pairs	8%
Improved Couples $\min(\hat{p}_0, \hat{p}_1, \hat{p}_2)$	**3.2%**
Relative difference of E_1 & O_1 *(using non-overlapping pixel groups)*	8.5%
Relative difference of *R, R'* *(using optimal mask and non-overlapping pixel groups and segmenting the image into 6-12 groups, taking 30th percentile of the per-segment statistics)*	--

Lowest embedding rate for which 50% false negatives achieved with no more than 5% false positives:

	2200 bitmaps + *JPEG compression*	
	none	*q.f. 50*
Conventional Pairs	10%	6%
Conventional RS	11%	5.5%
Conventional Couples	9%	5%
RS w/ optimal mask	10%	5%
Improved Pairs	8%	2.8%
Improved Couples $\min(\hat{p}_0, \hat{p}_1, \hat{p}_2)$	**3.2%**	1.8%
Relative difference of E_1 & O_1 *(using non-overlapping pixel groups)*	8.5%	**0.8%**
Relative difference of R, R' *(using optimal mask and non-overlapping pixel groups and segmenting the image into 6-12 groups, taking 30th percentile of the per-segment statistics)*	--	--

Lowest embedding rate for which 50% false negatives achieved with no more than 5% false positives:

	2200 bitmaps + *JPEG compression*		5000 JPEGs *(high quality)*	10000 JPEGs *(low quality)*	7500 JPEGs *(very mixed)*
	none	*q.f. 50*			
Conventional Pairs	10%	6%	4%	1.8%	7%
Conventional RS	11%	5.5%	2.8%	1.6%	7%
Conventional Couples	9%	5%	3%	1.4%	6.5%
RS w/ optimal mask	10%	5%	2.2%	1.2%	5.5%
Improved Pairs	8%	2.8%	3%	1.2%	5%
Improved Couples $\min(\hat{p}_0, \hat{p}_1, \hat{p}_2)$	**3.2%**	1.8%	2%	3.8%	3.6%
Relative difference of E_1 & O_1 *(using non-overlapping pixel groups)*	8.5%	**0.8%**	2.4%	0.6%	2.8%
Relative difference of R, R' *(using optimal mask and non-overlapping pixel groups and segmenting the image into 6-12 groups, taking 30th percentile of the per-segment statistics)*	--	--	**1.4%**	**0.5%**	**2.0%**

The End