

# Improved Detection of LSB Steganography in Grayscale Images

Andrew D. Ker

Oxford University Computing Laboratory, Parks Road, Oxford OX1 3QD, England  
adk@comlab.ox.ac.uk

**Abstract.** We consider methods for answering reliably the question of whether an image contains hidden data; the focus is on grayscale bitmap images and simple LSB steganography. Using a distributed computation network and a library of over 30,000 images we have been carefully evaluating the reliability of various steganalysis methods. The results suggest a number of improvements to the standard techniques, with particular benefits gained by not attempting to estimate the hidden message length. Extensive experimentation shows that the improved methods allow reliable detection of LSB steganography with between 2 and 6 times smaller embedded messages.

## 1 Introduction

Steganography aims to transmit information invisibly, embedded as imperceptible alterations to cover data; steganalysis aims to unmask the presence of such hidden data. Although by no means the most secure method of embedding data in images, LSB steganography tools are now extremely widespread. It is well known that embedding near-to-maximum size messages in images using the LSB technique is quite reliably detectable by statistical analysis [1,2] but that spreading fewer embedded bits around the cover image makes the steganalyst's task much more difficult [3].

In this paper we present improved steganalysis methods, based on the most reliable detectors of thinly-spread LSB steganography presently known [4,5,6], focussing on the case when grayscale bitmaps are used as cover images. They arise as a result of observations from a distributed steganalysis project, undertaken in response to a general call at the 2002 Information Hiding Workshop for thorough evaluation of the reliability of steganalysis techniques. The project uses a network of computers to provide speedy computation of steganalysis statistics over large image libraries, making it easy to see where improvements can arise. An outline of the project, and the first results, can be found in [7].

The aims of this paper are a) to suggest improved steganalysis statistics for LSB steganography, b) to use large image libraries to give experimental evidence of the improvement, and c) to examine closely the upper limits on bit rate which keep LSB steganography undetectable. We do not give theoretical analysis of the improved statistics and in no way claim that they are necessarily optimal; our intention is simply to advance the state of the art.

## 1.1 Scope

We take on the role of an “information security officer”, a hypothetical Warden whose job it is to scrutinise electronic communication. We want to answer the *simple classification question* – whether a given image has hidden data or not – and our work is currently focussed solely on the reliability of steganalysis methods to answer this question. Each steganalysis method will be statistic (a function of the input image) designed to discriminate between the two cases. Thus we are looking for a hypothesis test, where the null hypothesis is that no data is hidden, and the alternative hypothesis is that data is hidden<sup>1</sup>. We have to presuppose a fixed method of embedding data and a fixed length of hidden message, so that both null and alternative hypotheses are *simple* (not depending on an unknown parameter). Then it becomes possible to simulate the distributions taken by steganalysis statistics in both cases.

A good steganalysis statistic would give higher values in the case of hidden data and lower values otherwise; the Warden’s only sensible strategy is to reject the null hypothesis (make a positive diagnosis of steganography) when the statistic exceeds a certain threshold. But in practice the distributions (histograms) of the statistic in the case of null and alternative hypotheses will overlap so there is no threshold which will make the detector work perfectly. Varying the detection threshold plays off the likelihood of false positive results against missed detections (false negative results), and it is the graph of these two probabilities, the Receiver Operating Characteristic (ROC) curve, which fully describes the reliability of a particular statistic against a particular hidden message length.<sup>2</sup>

A key assumption in this paper is that false positive results are considered more serious than missed detections. If most images which come under the scrutiny of the information security officer are innocent it is important that false positives do not swamp true detections. So for the rest of this work we will assume that the Warden requires a detector with a fairly low false positive rate (in the region of 1-10%) and also that the steganographer acts repeatedly, so that even a missed detection rate of 50% is acceptable because eventually they would be caught. We recognise that the numbers involved are fairly arbitrary but it is necessary to start somewhere.

For now we are *not* interested in more advanced analysis of suspect images such as estimates of hidden message length [4,8,5], except in as much as they function as discriminating statistics for the simple classification problem. Such *threshold-free* statistics are popular, but the lack of a detection threshold is illusory because an information security officer would have to know whether

---

<sup>1</sup> Some other authors have reversed the designation of null and alternative hypothesis, but our exposition fits better with the accepted norms of statistics.

<sup>2</sup> Pierre Moulin has pointed out that randomized detectors are optimal, and in the case when the ROC curve is concave can improve performance up to its convex closure. But to exploit this does require a genuinely simple alternative hypothesis and this is not likely to be the case in practice – the Warden does not have advance warning of the amount of hidden data to expect. So for now we ignore this issue, although the reader may wish mentally to take the convex closure of the ROC curves displayed.

to interpret a particular estimated message length as significantly higher than zero or not. A more precise measure of the certainty of a positive diagnosis is the *p-value* of an observation, which can be computed for any type of statistic. Furthermore, we asked in [7] whether statistics designed to estimate the hidden message length were suboptimal for the simple classification problem and we will show here that the answer is yes.

## 1.2 LSB Steganography

Here we consider simple Least Significant Bit (LSB) steganography, long-known to steganographers, in which the hidden message is converted to a stream of bits which replace the LSBs of pixel values in the cover image. When the hidden message contains less bits than the cover image has pixels, we assume that the modifications are spread randomly around the cover image according to a secret key shared with the intended recipient of the stego image. This sort of steganography is only suitable for images stored in bitmap form or losslessly compressed. One should clearly distinguish this method (perhaps best called *LSB replacement*) from an alternative described in [9], where the cover pixel values are randomly incremented or decremented so that the least significant bits match the hidden message (this should perhaps be called *LSB matching*). In the latter case the message is still conveyed using the LSBs of the pixel values of the image, but the simple alteration to the embedding algorithm makes it much harder to detect. None of the methods discussed here will detect this alternative form of steganography, and indeed it is a much more difficult task to do so: a detector for LSB matching in full colour bitmaps is described in [2] but it is ineffective for grayscale covers; another detector which works for full colour images is described in [10] but it is only reliable for very large embedded messages and barely effective for grayscale covers.

LSB replacement is by no means the best – or even a sensible – steganographic method. However we consider it extremely worthy of study because of its widespread use. A large majority of freely available steganography software makes use of LSB replacement, but there is a more important reason: it can be performed without any special tools at all. Imagine, for example, a steganographer trying to send secrets out of a corporation. If the corporation takes information security seriously then the very presence of any steganographic software on an employee's computer is certain to be noticed and is *prima facie* evidence of wrongdoing, regardless of the undetectability of the actual messages. But a canny steganographer can simply go to a UNIX-style commandline and type

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

to embed a message (backwards) in the LSBs of the pixels in a PGM image (the PGM format is common and there are widely installed commandline tools to convert from JPEG, BMP or other formats, and then back to BMP if necessary

for transmission). This 80 character Perl code is short enough to memorise, and fairly small modifications can be made to spread the embedding around the cover image. The more sophisticated methods of embedding cannot easily be performed without special software<sup>3</sup>. This is why, for now, we focus on LSB replacement.

### 1.3 Pairs, RS, and Couples Steganalysis

We summarise the methods for the detection of LSB steganography on which our later work builds. Nothing in this section is new and details are omitted; the reader is referred to the original papers for a proper explanation of how each statistic works. We re-present the detection statistics of [4,5,6] in a way which emphasises their fundamental similarities. Firstly, all are “threshold-free” statistics which aim to estimate the length of a hidden message, and we assume that the method is used to answer the simple classification problem by accepting the null hypothesis if the estimated length is less than a certain threshold. Pairs Analysis was designed with paletted images in mind, but there is no theoretical reason why it should not work for grayscale images; RS was designed with colour images in mind, although it works by treating each colour component separately and as such is really a grayscale method.

In each case two measurements are made: in this work we will write  $Q(p)$  and  $Q'(p)$  for random variables which are the two measurements when  $2p$  is the amount of embedded data<sup>4</sup>. In each of [4,5,6] either theoretical calculation or experimental evidence shows that the expectations of  $Q(p)$  and  $Q'(p)$  are (precisely or a close approximation to) a quadratic in  $p$ . For a given image with an unknown amount of embedded data (possibly zero) we can observe  $Q(p)$  and  $Q'(p)$ , and also  $Q(1-p)$  and  $Q'(1-p)$  by flipping all LSBs. In each case it is also possible to obtain  $Q(0.5)$  and  $Q'(0.5)$ , either by theoretically derived calculation or by randomizing the LSB plane of the image. Finally, in each of the cases of Pairs, RS and Couples we make the assumption that  $Q(0) = Q'(0)$  – an assumed property of natural images – and the correctness of this assumption is the major factor in the accuracy of the final estimate. The law of large numbers means that the values of the random variables  $Q(p)$  and  $Q'(p)$  will be close to their expectations; there is now sufficient information to solve for the parameter  $p$  ([6] includes detailed calculations). The measurements  $Q$  and  $Q'$  differ for the three methods, although they are not dissimilar.

In Pairs Analysis [5], due to Fridrich *et al*, first *colour cuts* are formed by scanning through and selecting only pixels which fall into each pair of values (0,1), (2,3), and so on. The colour cuts are concatenated to form a single stream,

<sup>3</sup> The exception is LSB matching, which can be done using code not much larger than that above. There is an urgent need for a improved detectors for LSB matching, especially when the embedded message is not of full length or for grayscale covers.

<sup>4</sup>  $p$  is the proportion of pixels with flipped LSBs, which is the expected proportion when  $2p$  pixels are used for steganography because about half of the pixels would have carried the correct LSB already.

a re-ordering of the pixels of the original image. The measure  $Q$  is the *relative homogeneity* of the LSBs of this stream, the proportion of adjacent pixels with the same LSB. The measure  $Q'$  is calculated in the same way except that the pairs of values used to form the colour cuts are the dual pairs (255,0), (1,2), (3,4), etc.

Also due to Fridrich *et al.* is the method of RS [4], also called *dual statistics*. Here the image is sectioned into groups of pixels; the size of the group is variable but in [4] it is either a four-by-one horizontal rectangle, or a two-by-two square. A “mask” is applied to each block – the mask specifies that certain pixels in it should have their LSBs flipped. Each group is classified as *regular*, *singular*, or neither, depending on whether the noise within the pixel group (as measured by the mean absolute value of the differences between adjacent pixels) is increased, decreased, or unchanged after this flipping; we denote the proportion of regular and singular groups as  $R$  and  $S$ . The classification is repeated using the dual form of flipping  $1 \leftrightarrow 2, 3 \leftrightarrow 4, \dots, 255 \leftrightarrow 0$ ; call the proportion of regular and singular groups under the dual flipping  $R'$  and  $S'$ . The two measurements finally used by RS steganalysis are  $Q = R - S$  and  $Q' = R' - S'$ ; under the additional assumption that *both*  $R = R'$  and  $S = S'$  for natural images it becomes possible to derive  $Q(0.5)$  and  $Q'(0.5)$  theoretically rather than resort to experimentation.

The third detection method we consider here is due to Dumitrescu *et al.*; it was presented in [6] where it was called *Sample Pairs Analysis*. The same technique was discovered independently (but not published) by this author and termed *Couples Analysis*. For this paper we use the latter name, partly out of familiarity and partly because “Sample Pairs” could easily be confused with “Pairs”. It is conceptually the simplest method of the three under consideration, and also has the most complete theoretical analysis. We will later show that it is also marginally the most accurate. Consider the set of all horizontally and vertically adjacent pairs of pixels in the image. Let  $E_k$  be the proportion of pairs of pixels which a) differ in value by  $k$  and b) of which the lower of the two values is even.  $O_k$  is the same but with the lower of the two values odd. Suitable measurements are  $Q_i = E_{2i+1}$  and  $Q'_i = O_{2i+1}$ ; in [6] it is shown that  $Q_i$  and  $Q'_i$  (for each  $i$ ) satisfy the properties listed above of  $Q$  and  $Q'$ ; after some analysis the authors suggest using  $Q = \sum_i Q_i$  and  $Q' = \sum_i Q'_i$ . It is also possible to compute  $Q(0.5)$  and  $Q'(0.5)$  exactly, without randomizing the LSB plane.

## 2 Experimental Programme

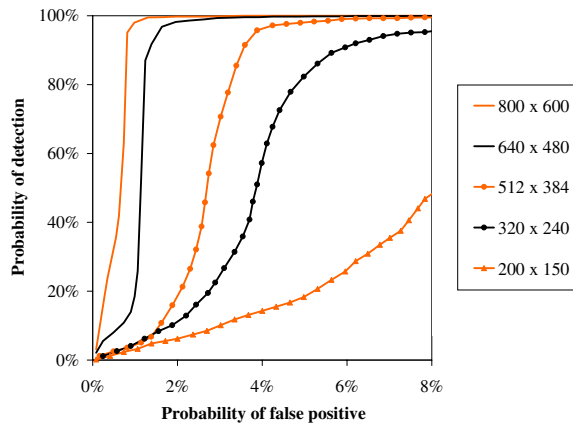
Experimental results come from our distributed steganalysis project (see [7] for some details). In order to evaluate the performance of a particular steganalysis algorithm against a particular method of steganography we need to approximate the distributions of the discriminating statistic in the two cases of absence and presence of hidden data. We do so using a number of large sets of sample images. We also need to repeat with varying amounts of hidden data to establish the level at which steganography becomes detectable. So for each steganography algorithm under consideration, and each steganalysis method being tested, with

a number of message sizes, we compute the discriminating statistic before and after embedding a random message.

Because the number of combinations of steganalysis algorithms (each with a large number of variations), message sizes, and thousands of images to test with the possibility of subjecting them to pre-embedding JPEG compression, is so large we will need millions of computations. This is distributed to network of machines, with the results stored in a relational database. At the time of writing there had been up to 50 machines used at once in the network, and the results database contained over 13 million rows. Results are then extracted, analysed to produce ROC curves for each set of parameters (steganography method, amount of hidden data, steganalysis statistic, image set, etc.) and graphed.

## 2.1 Sample Results

Figure 1 shows some of the results from the database. The chart displayed shows the ROC curves for a small set of 1200 uncompressed images, when 10% LSB steganography (i.e. 0.1 bits per pixel) is used and the images are tested with the standard RS statistic of [4]. The experiment has been repeated with the cover images first resampled down to a number of different sizes, and it is instructive to see what a difference this makes to the reliability of the RS statistic.



**Fig. 1.** ROC curves for a set of 1200 uncompressed images, originally  $1024 \times 768$  but resampled down to a variety of smaller sizes. In each case 10% steganography has been used, and tested against the standard RS statistic of [4]

Compare the curves for the two sets resizes to  $640 \times 480$  and  $320 \times 240$ . It turns out (ROC curve not displayed) that the performance of the RS statistic in the  $640 \times 480$  images when 0.1 bits per pixel steganography is embedded is approximately the same as for the  $320 \times 240$  images when 0.17 bits per pixel LSB steganography is used. This is not contrary to the instinctively obvious fact

that larger images can carry larger messages securely, but it does indicate that the increase is not proportional.

Figure 1 also illustrates the general shape of ROC curves, which tend to fall dramatically when the false positive rate goes below a certain level. Thus it is not often useful to fix a particular false-positive rate and compare different statistics' reliability rates at this point. A more reasonable one-dimensional measure of performance, and one we quote on occasion, is the level of false positives when the threshold is set for 50% reliability. We find that this often serves as a fair summary of the performance. At the end we will focus on an even more particular case, determining the minimum embedding rate for which 50% reliability is achieved with a 5% false positive rate.

When choosing which ROC curves to show we will focus on “interesting” cases – we will choose a steganography embedding rate so that the performance is neither too near perfect (in which case any differences are as likely due to chance as anything else) or too poor (because results of that nature are not interesting). We will also scale the x-axis (false positive rate) so that the graph shows only areas of interest (in particular we will not show false positive rates of more than 10%). The y-axis will always run over reliability rates of 0% to 100%.

## 2.2 Image Sets Used for Testing

In [7] we gave two important examples which warn of some of the difficulties in evaluating steganalysis algorithms. Firstly, we found that cover images which have been JPEG compressed can lead to vastly different reliability of detection, even after the JPEG images were substantially reduced in size in an attempt to “wash out” the compression artifacts. Secondly we found that different resampling methods used to resize sets of images also resulted in different performance when steganalysis methods were tested against them. This makes it clear that there is no such thing as a universally “representative” set of natural images for the purposes of testing steganalysis.

We address this issue in part by obtaining a number of large sets of images and using each set separately, to be sure of covering all image types and also to expose any differences in performance with the eventual aim of explaining them. So in subsequent testing we will use:

**Image Set A:** 2200 simulated uncompressed images, all  $512 \times 512$ . The “simulation” of uncompressed images is performed by taking very large and mildly compressed JPEG files and reducing (in view of the warnings of [7] we have used a mixture of resampling algorithms). These images are “high-quality” in the sense that out-of-focus and incorrectly exposed images have been removed. Since they are uncompressed we will also repeat experiments by pre-compressing them, to measure how much the statistics' reliability depends on this factor.

**Image Set B:** 5000 JPEG images, all sized  $900 \times 600$ . Each is compressed at JPEG quality factor 75. These came from a royalty-free image library purchased by the author. The photographs are of quite good quality in terms of exposure and focus, but they appear to have been scanned in from 35mm film and some show granularity. Some have a small black border.

**Image Set C:** 10000 JPEG images, sizes varying between  $890 \times 560$  and  $1050 \times 691$ . The JPEG compression levels vary from approximately quality factor 50 to 75. These images came from another royalty-free image library, but the quality of pictures is not as good as Set B; some images are blurred or incorrectly exposed.

**Image Set D:** 7500 JPEG images of very variable quality. They were obtained from an open image library which the public may contribute to. Accordingly the images' provenance cannot be verified, but they clearly come from a very wide range of sources. The quality is extremely variable – there are a few blurred, grainy and distorted pictures included. Most of the images are sized between  $800 \times 600$  and  $1024 \times 768$ . The JPEG compression levels are fairly evenly distributed between approximately quality factors 50 and 90.

It will be seen that Image Set A is “difficult” for the steganalyst, in that the statistics' reliability is worse over this set than the others (and this seems the general case for uncompressed images). Set C is the “easiest”. Set D is expected to be difficult because of its heterogeneity. Our image library contains other sets but in the interests of space we do not report results for them.

One may ask why we test the spatial-domain LSB steganography method against images which have been stored in JPEG format, especially given the technique of *JPEG compatability analysis* [11]. One reason is that we have found it extremely hard to obtain large sets of images which can be guaranteed never to have undergone compression or other distortions. Furthermore the fact is that most natural images are stored in JPEG format and, just as we are examining LSB steganography for its ease of use and prevalence, we want to test against all likely types of cover image. The casual steganographer may well only have access to JPEG compressed images. Finally, we believe that JPEG compatability analysis can be avoided if simple global operations such as very mild blurring or change of contrast are applied to the JPEG images before LSB embedding.

### 3 Improved Detection Methods and Experimental Evidence

In the main body of this paper we will suggest a number of improved detectors for LSB steganography. In each case we outline a steganalysis method and give some experimental evidence (in the form of ROC curves) of improved reliability. However it is impossible to display the ROC curves of every combination of image set, embedding rate, variation of detection statistic, and so on; we select a representative sample for display and will also comment on the extent to which the improvements hold in general. We begin with a recap of the improvements suggested in [7]; all subsequent results are new research. A table summarising the performance of all the statistics over all Image Sets can be found in Sect. 4.

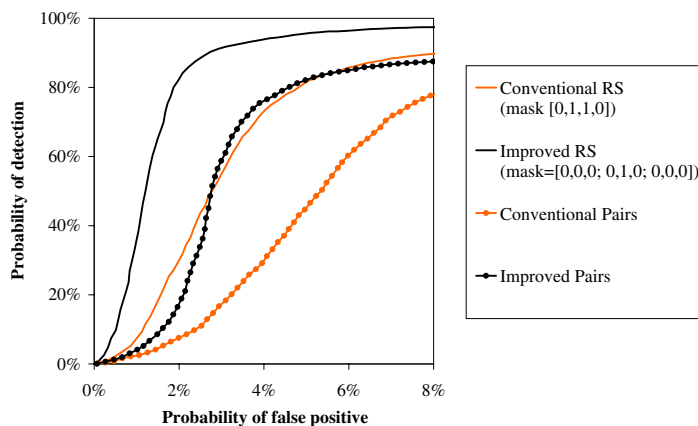
#### 3.1 Improved Pairs & Better RS Masks

RS Steganalysis depends on the particular choice of “mask”, which determines how pixels are grouped and which are flipped during the noise measurements.



In the presentation of [4] the authors mention two masks – the horizontal row  $[0, 1, 1, 0]$  and the square  $[1, 0; 0, 1]$  – without commenting on why they were chosen. In [7] we investigated a number of other masks and found that a performance improvement could be obtained using the square  $[0, 0, 0; 0, 1, 0; 0, 0, 0]$  instead. (The noise measurement used in the RS calculation is extended to two-dimensional masks by summing differences between both all horizontal and all vertical pixel pairs.)

Pairs Analysis was substantially improved by excluding some pixels from the homogeneity measurement, namely those pixels which were not adjacent in the original image. This amounts to splitting the colour cuts into small subsequences of originally adjacent pixels and measuring the homogeneity within those subsequences. The rest of the algorithm is identical to the standard Pairs method (repeating for the alternative pairs of values and solving the same quadratic equation to find an estimate of hidden message length).



**Fig. 2.** ROC curves showing the reliability gained by using the suggested RS “mask” and the Improved Pairs measure. The curves are generated from the 15000 images in Image Sets B and C combined; the hidden message length is 3%

Figure 2 shows the effects of these improvements on the ROC curves in one instance. 15000 JPEG images had 3% steganography embedded: here the false positive rate needed to achieve 50% reliability has reduced from 2.7% to 1.2% when the RS mask is switched from the standard  $[0, 1, 1, 0]$  to the improved  $[0, 0, 0; 0, 1, 0; 0, 0, 0]$ , and the modification to Pairs Analysis has reduced it from 5.3% to 2.7%. Similar improvements are observed across all Image Sets and with all message sizes. In [7] we gave ROC curves showing that in some circumstances the improved Pairs statistic becomes more reliable than the RS method (this is particularly noticeable in the case of uncompressed images, as will be seen in Table 1).

One other minor improvement we mention here, which is not reported in [7], is a simplification of the RS statistic. Recall that the RS message-length estimate

is computed from two measures  $Q = R - S$  and  $Q' = R' - S'$ , where  $R$  and  $R'$  represent the number of *regular* pixel groups under LSB flipping and dual flipping according to the mask, and  $S$  and  $S'$  the *singular* groups. It is easy to see that the results of [4] show that the measures  $R$  and  $R'$  alone suffice to estimate the hidden message length, using the assumption that  $R = R'$  for natural images, so long as one is prepared to determine  $R(0.5)$  by randomizing the LSB plane of the image under consideration. The same applies to  $S$  and  $S'$ . We have found that just using  $R$  and  $R'$  to estimate the hidden message length is actually *more* reliable than the full RS method (this does not apply to  $S$  and  $S'$ , which alone make a very poor detector). This is a surprising result but the improvement is not very substantial and we do not display ROC curves to illustrate it; Table 1 illustrates the incremental advantage sufficiently.

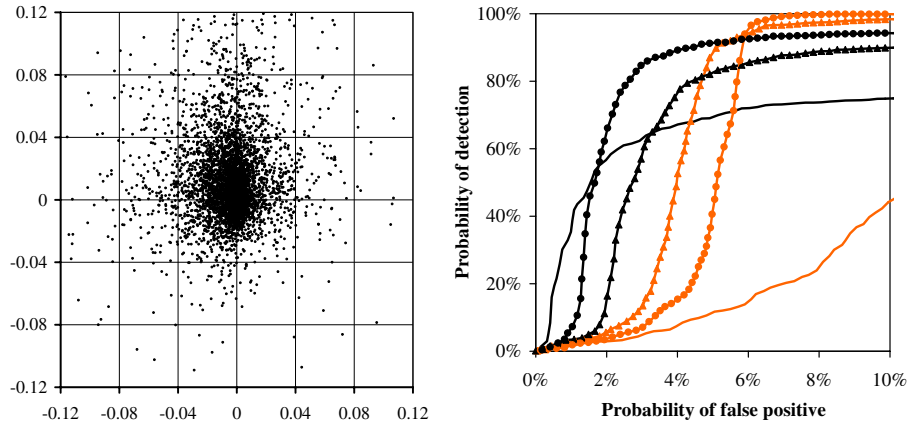
### 3.2 Improving Couples Analysis

As described in [6] Couples Analysis is in fact marginally more reliable than conventional RS steganalysis (see Tab. 1). However the testing performed for that paper was very limited and this may have lead the authors to miss an important feature.

Recall that there are a number of alternative measures,  $Q_i = E_{2i+1}$  and  $Q'_i = O_{2i+1}$  for  $i \geq 0$  (where  $E_k$  is the proportion of pairs of pixels which differ by  $k$  and of which the lower is even,  $O_k$  analogously for odd). Let us write  $\hat{p}_i$  for the estimated hidden message length computed using  $Q_i$  and  $Q'_i$ , and  $\hat{p}$  for the estimate described in [6], which uses  $Q = \sum_i Q_i$  and  $Q' = \sum_i Q'_i$ . The authors claim that  $\hat{p}$  is “more robust” than the  $\hat{p}_i$ , a conclusion we generally agree with (although not without reservation as there have been a few circumstances, involving mildly JPEG compressed covers, when  $\hat{p}_1$  was observed to be superior to  $\hat{p}$ ).

However a much more useful fact is that the different estimators  $\hat{p}_i$  are generally uncorrelated. Figure 3, left, shows a scattergram of  $\hat{p}_0$  against  $\hat{p}_1$  generated by the images in Set B (with no embedded data); there is no visible relationship, and the Pearson correlation coefficient is only  $-0.0365$ . Image Sets C and D have similar results; the uncompressed Image Set A gives a higher correlation coefficient of  $0.1743$  but this is still quite a weak relationship. The power of these uncorrelated statistics is that it is much less likely that an image with no hidden data would show up as a false positive for *both* statistics. So we could set thresholds for  $\hat{p}_0$  and  $\hat{p}_1$  and give a positive diagnosis of steganography only if both are exceeded. Furthermore, one need not stop at using two statistics. We also found fairly weak correlation between the other  $\hat{p}_i$  statistics, although the correlation does rise with  $i$ , and the reliability falls. After some experimentation we determined that taking the three values  $\hat{p}_0$ ,  $\hat{p}_1$ , and  $\hat{p}_2$ , and setting the same threshold for each gave the best overall performance. This amounts to using  $\min(\hat{p}_0, \hat{p}_1, \hat{p}_2)$  as the discriminating statistic<sup>5</sup>.

<sup>5</sup> We do not claim that this combination of the  $\hat{p}_i$  is necessarily optimal, merely that it is the best we could find; an interesting direction for further research is to find the best ways to combine *all* of the various detection statistics in the optimal way.



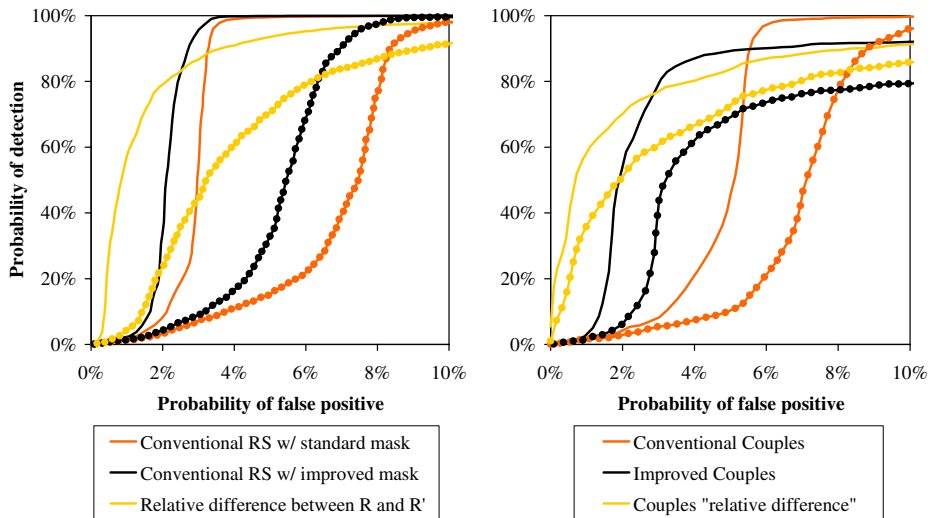
**Fig. 3.** Left, A scattergram plotting the message-length estimates  $\hat{p}_0$  ( $x$ -axis) against  $\hat{p}_1$  ( $y$ -axis). No correlation is evident. Right, ROC curves showing how the reliability of the conventional Couples statistic  $\hat{p}$  varies as the covers are pre-compressed (*shaded lines*), and the improvements gained by using  $\min(\hat{p}_0, \hat{p}_1, \hat{p}_2)$  instead (*black lines*). The covers used are the 2200 uncompressed images in Set A (*unmarked lines*), and the experiment is repeated with the covers pre-compressed using JPEG quality factors of 90 (*lines marked with triangles*) and 50 (*lines marked with circles*). 3% steganography has been used

Figure 3, right, shows the results. The ROC curves are all generated from Image Set A, with the experiment repeated with the covers first JPEG compressed using quality factors of 90 and 50. In the case of uncompressed covers the false positive rate needed to achieve 50% reliability has reduced from 10.7% to 1.5% (a dramatic improvement indeed!). For the mildly JPEG compressed covers it has reduced from 4% to 2.7%, and for the quite heavily compressed quality factor 50 images it has reduced from 5.1% to 1.7%. It is curious that the relative performance of the Couples statistic, as JPEG compression of the covers varies, is exactly reversed by the improved method. Other observations suggest that mildly compressed covers have particular properties which destroy the accuracy of the estimate  $\hat{p}_1$  (but do not affect  $\hat{p}_0$  or  $\hat{p}_2$  nearly as seriously); further research is called for to see if this can be mitigated.

This modified method of Couples Analysis is now substantially more reliable than any of the conventional steganalysis statistics (see Tab. 1) in answering the simple classification question. However the discrimination statistic  $\min(\hat{p}_0, \hat{p}_1, \hat{p}_2)$  is no longer an unbiased estimate of the hidden message length (it will underestimate).

### 3.3 Dropping the Message Length Estimate

In [7] we asked whether the use of a statistic designed to estimate the hidden message length could be improved upon, given that we only want to answer the



**Fig. 4.** Left, the effect of switching to the relative difference statistic in RS steganalysis – the ROC curve shown is generated from the 7500 images in Set D and compares the conventional RS statistic (with mask  $[0, 1, 1, 0]$ ), the version with the improved mask  $[0, 0, 0; 0, 1, 0; 0, 0, 0]$ , and using the relative difference between  $R$  and  $R'$  (computed with the mask  $[0, 1, 1, 0]$ ). The experiment is repeated with both 5% (*marked lines*) and 10% (*unmarked lines*) steganography. Right, a similar comparison between conventional Couples Analysis, the improved Couples analysis using  $\min(\hat{p}_0, \hat{p}_1, \hat{p}_2)$ , and finally using  $(Q_0 - Q'_0)/(Q_0 + Q'_0)$ , with both 3% (*marked lines*) and 5% (*unmarked lines*) steganography. The second diagram was generated from the combined 15000 images in Sets B and C

simple question of whether data is hidden or not. We have just seen a statistic which does the latter better at the expense of the former.

Let us return to the most important assumptions which underlie Pairs, RS, and Couples – that  $Q(0) = Q'(0)$  in natural images. A simple and obvious statistic to consider is therefore  $Q - Q'$ , which should be near zero in natural images and (one can show in each of the cases of Pairs, RS and Couples) generally moves away from zero as data is hidden. Unfortunately the magnitudes of  $Q$  and  $Q'$  can differ appreciably between images, usually depending on how noisy the image under consideration is; therefore a more robust measure is the relative difference  $(Q - Q')/(Q + Q')$ . One can compute  $Q$  and  $Q'$  according to any of the methods of Pairs, RS, or Couples. In the case of RS we have found it better to ignore the  $S$  and  $S'$  components and use the relative difference between  $R$  and  $R'$  instead.

These statistics are no longer any use for determining the hidden message length. On the other hand we might hope that, uncluttered by the additional observations and quadratic equation needed to do so, they are a reasonable way to answer the simple classification question.

Figure 4, left, shows the result of switching to the relative difference statistic in the case of both RS and Couples (there is some improvement in doing the same with Pairs, but the results are not so good and we do not show them here). We display the ROC curves for the conventional RS statistic, the version with the better mask, and the relative difference statistic<sup>6</sup>. These curves were generated using Image Set D but similar results are seen across all image sets. We have displayed ROC curves for both 5% and 10% embedded message rates to demonstrate that improvement is evident across a range of embedding levels. At the 5% embedding level the false positive rate at which 50% reliability is achieved has fallen from 7.5% (standard mask) and 5.5% (improved mask) to 3.2% with the relative difference statistic.

The right-hand chart in Fig. 4 shows the improvement as we move from the conventional Couples statistic, to the minimum-of-3 statistics described in the previous section, to the relative difference statistic. In this case we have used the relative difference between  $Q_0$  and  $Q'_0$  – we investigated a number of other statistics based on relative differences between combinations of  $Q_i$ 's but found that  $Q_0$  and  $Q'_0$  was almost always the outstandingly most reliable. The level of improvement is similar to that observed for RS.

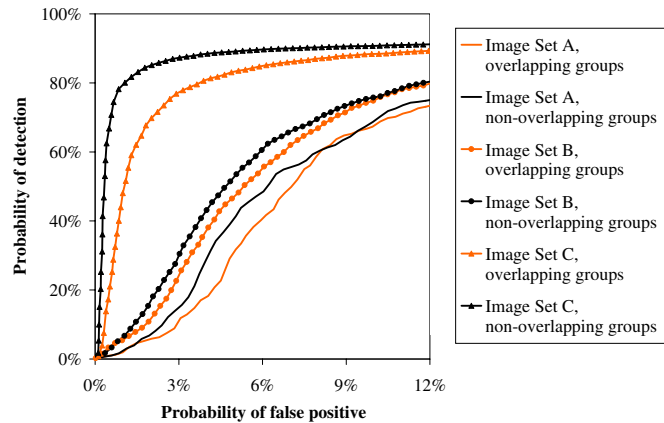
### 3.4 To Overlap or Not to Overlap

Each of the methods of Pairs, RS and Couples involve performing some calculation on pixel groups. For RS the groups are shaped as the mask and the calculation is to see whether noise is increased or reduced after LSB flipping and dual flipping. For Pairs and Couples the groups are simply pairs of pixels adjacent in the image and/or the colour cuts and the calculation is to measure homogeneity (whether the two pixels are equal) or classifying the pair of pixels in one of  $E_k$  or  $O_k$  by measuring their difference. We ask whether the groups should be disjoint or overlap. Since Pairs measures homogeneity is it clear that the groups of pairs must overlap so that every pair of adjacent pixels is considered. The authors of [6] clearly intended the groups in Couples to overlap (“all pairs of two spatially adjacent samples”). It is not clear whether the RS groups used in [4] were intended to overlap.

We firmly expected that using overlapping groups (in any of the methods of Pairs, RS or Couples) would give at best an insignificant improvement over not doing so, since it parallels a result of [7] in which using the same pixels twice was demonstrated to confer no particular advantage. Indeed this is exactly what we found in the case of the statistics which give estimates of the hidden message length. Most surprisingly, the story was quite different for the relative difference statistics: in these cases there was frequently quite a good improvement when using *non-overlapping* groups.

Figure 5 shows some of the advantages of using non-overlapping groups. The ROC curves displayed are for the relative difference between the measures  $R$  and

<sup>6</sup> using the standard mask  $[0, 1, 1, 0]$ ; we have observed that the other masks no longer give improved reliability when the relative difference statistic is used and indeed many are much worse.



**Fig. 5.** ROC curves showing the benefits of non-overlapping groups. The two statistics shown are both computed as the relative difference between  $R$  and  $R'$  (the proportions of regular groups, under the mask  $[0, 1, 1, 0]$ ), but one statistic uses overlapping groups and the other disjoint groups. 2% steganography was used. The experiment is repeated for three sets of images: Image Set A pre-compressed using JPEG quality factor 90, and Image Sets B and C

$R'$  (computed using the mask  $[0, 1, 1, 0]$ ) with overlapping and non-overlapping groups of pixels, with 2% steganography. Since the mask is 4 pixels long the latter has only 1/4 as many groups, but (for a reason as yet unexplained) gives better reliability. The improvement shown for Image Set A (the graph shown is from the case when the images are pre-compressed using JPEG quality factor 90) and Set B is significant but not dramatic. For Image Set C it is more pronounced. Generally, improved reliability is seen with any length of hidden message and with both the RS- and Couples-based relative difference statistics, although the extent of the improvement varies. In uncompressed covers there is little improvement.

### 3.5 Reducing Outliers by Segmenting

The final improvement we suggest is still work-in-progress. Our aim is to mitigate the sometimes alarming outliers in the null distributions, natural images which have a large *bias* (estimated hidden message length when there is no hidden message). We have observed that very large bias sometimes occurs in certain textures in an image when the rest of the image is quite normal – the overall bias comes out too high due to the influence of this abnormal texture. This differs from the situation when LSB steganography is present, where one expects to see a higher message-length estimate in all parts, assuming that the LSB flipping has been spread randomly over the whole image.

We have tried segmenting images according to their texture content and computing the discrimination statistics for each segment, then discarding outliers by taking the median (or a similar centrally weighted measure) of the values for

each segment. The picture on the top left of Fig. 6 is a good example. Under the standard RS statistic this image (one of Set A) has a bias of 0.0651, by no means the most extreme outlier in the data sets but still a substantial error. We segment the image according to content and compute the bias for each segment; the results are displayed in the picture on the top right of Fig. 6. The median of the biases for each segment is 0.0052, a much smaller error.

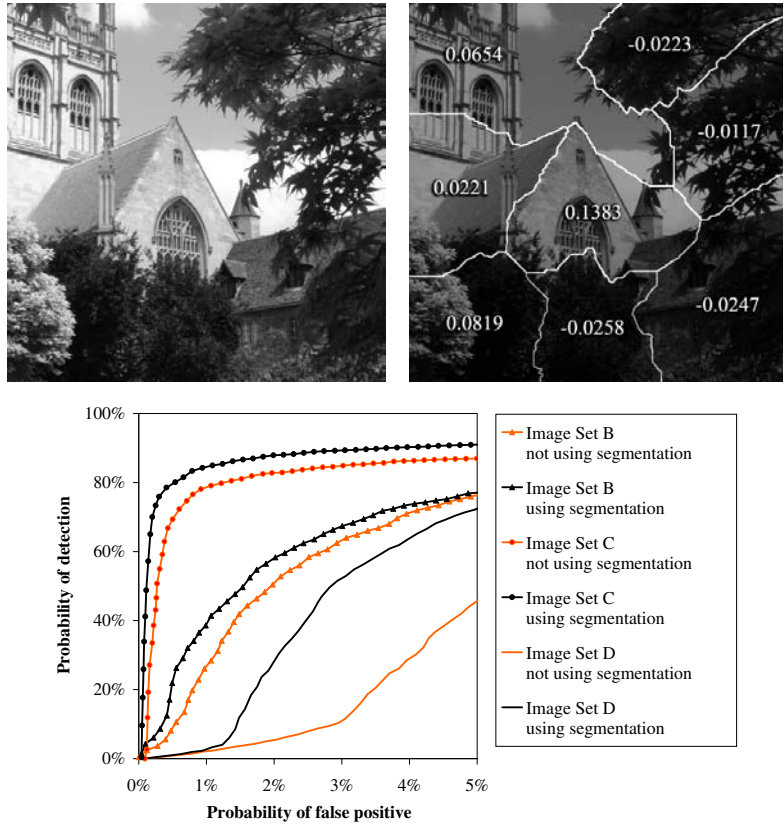
To perform the segmentation we chose the method of [12], partly because it avoids oversegmentation in highly detailed images without human intervention, and partly because an implementation is available for download. As can be seen from the pictures in Fig. 6 its choice of segments is sometimes rather surprising, but it does seem to separate different textures quite well. We make no claim that this segmenting method is in any way optimal for steganalysis purposes (indeed one might hope to perform segmentation according to the steganalysis statistics themselves); the results here are intended as a springboard for further research into the issue.

Segmentation is not particularly fast so we restricted our attention to adding segmentation to the best-performing statistics found so far (non-overlapping relative difference between  $R$  and  $R'$  or  $Q_0$  and  $Q'_0$ ). We adjusted the segmentation parameters so that most images were segmented into 6-12 segments and found that lower false positive rates were given by taking roughly the 30th percentile out of the statistics computed for individual segments (this biases the results low, trading worse reliability at high false positives for better reliability at low false positives – precisely the sort of trade we want to make).

The graph in Fig. 6 shows the benefits of using segmentation, comparing the relative difference between the non-overlapping versions of  $R$  and  $R'$  statistic with and without segmentation. 3% steganography was embedded in Image Sets B, C and D. The improvement in the case of Set C is particularly good, with the false positive rate needed to achieve 50% reliability dropping from 0.26% to less than 0.08%. When segmentation was added to the Couples  $Q_0$  and  $Q'_0$  relative difference statistic there was also an improvement, but not as much (ROC curves not shown). We hope to improve more on these results after further investigation.

## 4 Summary of Results and Conclusions

We conclude with a summary of the improvements made by these new detection statistics. It is necessary to simplify, so we have used a definition of “reliable” detection as meaning 5% false positives and at most 50% missed detections (we recognise that these figures are arbitrary but they are in keeping with the philosophy that false positives are more severe than false negatives – such a detector would be reasonably useful for an Information Security Officer who would only make a definite diagnosis of steganography after seeing a number of positive results coming from the same person). We measured the lowest level of steganography for which such reliability is attained by each statistic, repeating for each Image Set, and also subjected the covers of Image Set A to JPEG compression at mild (quality factor 90), moderate (75) and strong (50) levels so



**Fig. 6.** Above, the results of segmentation. Below, ROC curves showing the benefits; the statistic used is the non-overlapping version of the relative difference between  $R$  and  $R'$ , as computed using the mask  $[0, 1, 1, 0]$ . The segmenting statistic takes the 30th percentile of the estimates for each segment. 3% steganography was embedded

as to examine this factor. Table 1 contains the results, starting with conventional methods, then listing the improved versions of Pairs and RS suggested in [7] and then displaying the statistics suggested in this paper. The segmenting statistic was not tested against Image Set A because initial results showed no likelihood of improvement.

The table shows, amongst other things, that relative difference statistics (computed using non-overlapping groups) are much better than the alternatives, and that segmenting images is a promising direction for further work. The exception is for the only set of uncompressed images, when the improved version of Couples Analysis performs the best (and note that none of the optimal statistics is capable of estimating the hidden message length). For each Image Set the best-performing statistic will reliably detect LSB steganography at between 2 and 6 times lower rates than the conventional methods, and there is a



**Table 1.** Final results. The table shows the lowest bit rate (in bits per pixel) of LSB steganography which can be “reliably” detected by the various methods and for each image set. Here we take reliable detection to mean 50% or higher probability of detection when the false positive rate is 5%. Entries in the table higher than 0.04 are accurate to 0.005; entries between 0.01 and 0.04 are accurate to 0.002, and entries below 0.01 are accurate to 0.001

Statistic	Image Set A w/compression				Image Set B	Image Set C	Image Set D
	None	q.f. 90	q.f. 75	q.f. 50			
Conventional Pairs	0.100	0.085	0.060	0.060	0.040	0.018	0.070
Conventional RS	0.110	0.045	0.050	0.055	0.028	0.016	0.070
Conventional Couples	0.090	0.040	0.050	0.050	0.030	0.014	0.065
RS with optimal mask	0.100	0.038	0.045	0.050	0.022	0.012	0.055
Improved Pairs	0.080	0.050	0.030	0.028	0.030	0.012	0.050
RS $R$ only	0.105	0.040	0.040	0.050	0.026	0.014	0.060
Improved Couples	<b>0.032</b>	0.030	0.020	0.018	0.020	0.038	0.036
Relative difference of $R, R'$	0.065	0.026	0.022	0.022	0.022	0.012	0.036
*Relative difference of $R, R'$	0.065	<b>0.022</b>	0.018	0.020	0.020	0.006	0.032
Couples $Q_1$ relative difference	0.085	0.030	0.016	0.012	0.028	0.009	0.034
*Couples $Q_1$ relative difference	0.085	0.028	<b>0.012</b>	<b>0.008</b>	0.024	0.006	0.028
*Relative difference of $R, R'$ with segmenting					<b>0.014</b>	<b>0.005</b>	<b>0.020</b>

\* indicates alternative versions using non-overlapping groups

suggestion that the best improvements come from the most highly compressed images. Since some improvement has been observed across all Image Sets we can be confident that the new statistics are genuinely and significantly more reliable.

Also important to note is the vast difference in reliability as the statistics are tested across the different Image Sets. One should therefore view the improved bit rates as relative to the conventional ones. We have already commented that image size makes a difference to steganalysis reliability, but it is clear that JPEG compression does too and there may be other factors as yet uncovered. Thus it is impossible to say that there is a definitive “safe” bit rate, below which steganography cannot be detected. It would appear, though, that a steganographer who chooses their covers carefully can still transmit quite a lot of hidden data (and this paper excludes any discussion of adaptive techniques for choosing where to embed, let alone methods other than LSB). This also suggests an interesting line of future research, where suspect images are classified in some way so that the best statistic for that particular type of image can be used. The issue is so complex that a learning machine may be necessary.

In conclusion, we have suggested a number of improved methods for deciding whether a grayscale bitmap contains LSB steganography or not. Thanks to the distributed steganalysis project we are able to give extensive experimental evidence of the extent of the improvement. This depends entirely on the particular weight one gives to false positive or negative results, but we have shown that when the aim is to reduce false positives (and when a fairly arbitrary definition of “reliable” is made) the new statistics allow reliable detection of between 2 and 6 times less embedded data than the previously best methods. In most cases, however, we have not tried to give a theoretical explanation of why the improvement occurs – our new methods are heuristic and there is no claim of optimality. We hope that the results presented here will stimulate research to this end.

### Acknowledgements

The author is a Royal Society University Research Fellow. Some of the work presented here was done while a Junior Research Fellow at University College, Oxford with additional funding from DSTL. The author is grateful to Sue Haines and Rob Thompson at DSTL for conversations and suggestions.

### References

1. Westfeld, A., Pfitzmann, A.: Attacks on steganographic systems. In: Proc. Information Hiding Workshop. Volume 1768 of Springer LNCS. (1999) 61–76
2. Westfeld, A.: Detecting low embedding rates. In: Proc. Information Hiding Workshop. Volume 2578 of Springer LNCS. (2002) 324–339
3. Chandramouli, R., Memon, N.: Analysis of LSB based image steganography. In: Proc. IEEE International Conference on Image Processing. (2001) 1019–1022
4. Fridrich, J., Goljan, M., Du, R.: Reliable detection of LSB steganography in color and grayscale images. Proc. ACM Workshop on Multimedia and Security (2001) 27–30
5. Fridrich, J., Goljan, M., Soukal, D.: Higher-order statistical steganalysis of palette images. In Delp III, E.J., Wong, P.W., eds.: Security and Watermarking of Multimedia Contents V. Volume 5020 of Proc. SPIE. (2003) 178–190
6. Dumitrescu, S., Wu, X., Wang, Z.: Detection of LSB steganography via sample pair analysis. In: Proc. Information Hiding Workshop. Volume 2578 of Springer LNCS. (2002) 355–372
7. Ker, A.: Quantitative evaluation of Pairs and RS steganalysis. In Delp III, E.J., Wong, P.W., eds.: Security, Steganography, and Watermarking of Multimedia Contents VI. Volume 5306 of Proc. SPIE. (2004) 83–97
8. Fridrich, J., Goljan, M.: Practical steganalysis of digital images – state of the art. In Delp III, E.J., Wong, P.W., eds.: Security and Watermarking of Multimedia Contents IV. Volume 4675 of Proc. SPIE. (2002) 1–13
9. Sharp, T.: An implementation of key-based digital signal steganography. In: Proc. Information Hiding Workshop. Volume 2137 of Springer LNCS. (2001) 13–26
10. Harmsen, J., Pearlman, W.: Higher-order statistical steganalysis of palette images. In Delp III, E.J., Wong, P.W., eds.: Security and Watermarking of Multimedia Contents V. Volume 5020 of Proc. SPIE. (2003) 131–142

11. Fridrich, J., Goljan, M., Du, R.: Steganalysis based on JPEG compatability. In Tescher, A.G., Vasudev, B., Bove, Jr, V.M., eds.: Multimedia Systems and Applications IV. Volume 4518 of Proc. SPIE. (2002) 275–280
12. Felzenszwalb, P.F., Huttenlocher, D.P.: Image segmentation using local variation. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (1998) 98–104