A General Framework for Structural Steganalysis of LSB Replacement

Andrew D. Ker

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

Abstract. There are many detectors for simple Least Significant Bit (LSB) steganography in digital images, the most sensitive of which make use of structural or combinatorial properties of the LSB embedding method. We give a general framework for detection and length estimation of these hidden messages, which potentially makes use of *all* the combinatorial structure. The framework subsumes some previously known structural detectors and suggests novel, more powerful detection algorithms. After presenting the general framework we give a detailed study of one particular novel detector, with experimental evidence that it is more powerful than those previously known, in most cases substantially so. However there are some outstanding issues to be solved for the wider application of the general framework.

1 Introduction

Spatial domain Least Significant Bit (LSB) replacement is a popular and simple type of steganography. It combines high capacity with extreme ease of implementation (see [1] for a 2-line embedding program) and, in digital images, is visually imperceptible. Many of the steganography tools available on the internet use some form of LSB replacement, but in fact it is highly vulnerable to statistical analysis. The literature is replete with such detectors, the most sensitive of which make use of *structural* or combinatorial properties of the LSB algorithm [2,3,4,5,6].

In this paper we present a general framework for structural detectors, which potentially includes *all* the combinatorial properties of LSB replacement, by considering effects of LSB changes on arbitrary groups of samples. As such we will present it as a generalisation of something akin to Sample Pairs Analysis (SPA) [3]. In fact, many previously known structural detectors are special cases of this general framework, although we will only make explicit the connection with the Sample Pairs method. The value of the framework is both to place the older methods into a common context and also to provide new, more powerful, detectors. Because it makes full use of the structural information, in some sense this framework should be the last word on the detection of LSB replacement, although many practical questions remain open.

We give a brief re-presentation of the SPA method, slightly modified in detail and exposition to make the subsequent generalisation work tidily (Sect. 2). The

M. Barni et al. (Eds.): IH 2005, LNCS 3727, pp. 296–311, 2005.

general framework is presented in Sect. 3. The optimal implementation of the method is still unclear, but in Sect. 4 we present a case study of the technique applied to 3-tuples of pixel groups (we call this *Triples Analysis*). Despite some outstanding issues which mean that the method is not applicable to large hidden messages, experimental results show that it provides detection (and length estimation) of LSB steganography which is generally much more sensitive than the previously known methods.

There is still work to be done apply the framework to larger groups of pixels effectively, and to avoid potential problems with very large hidden messages, and we discuss these issues briefly in Sect. 5.

1.1 LSB Steganography and Steganalysis

The LSB embedding method is simple. The secret message consists of a stream of bits, and the cover medium is expressed as a stream of bytes (typically the grayscale or RGB pixel values of a bitmap image). The least significant bits of the cover bytes are overwritten by the secret message. For security, and to spread the stego noise, the cover is usually traversed in a pseudorandom order.

Despite many detectors for LSB steganography it remains of interest because it is one of the few embedding methods simple enough to require no special software [1]. Furthermore it is still possible to use it for secure communication, if the hidden message is kept very short in relation to the capacity of the cover. The aim of the steganalyst must be to refine the detection methods so that reliable detection of smaller messages is possible.

There are, broadly, two approaches to the detection of LSB steganography. One is to use signal processing techniques to extract feature vectors for a learning machine of some sort; literature on this ranges from very simple noise detectors such as [7] to the more sophisticated wavelet methods of [8]. Such detectors are likely to work for a wide range of embedding methods in addition to LSB Replacement, but do not provide any information on the nature of the hidden message and are generally less sensitive than specialised methods.

Other detectors make use of "structural" or combinatorial properties specific to the LSB embedding method. Such detectors appear to have much in common. In each case we assume that a cover image is fixed and a random hidden message, of bit rate p, is embedded. The "Sample Pairs" technique of [3] (discovered independently by this author who called it "Couples"), the "Pairs" method of [4], and the "Difference Histogram" method of [6] all consider pairs of pixels (although they differ in which pairs are selected) and use some macroscopic quantity $F(p)$ which depends on the secret bit rate p; in each case there is a claim or proof that $F(p)$ is a quadratic in p, although key parameters of the quadratic are unknown without the cover image; where necessary, one derives or estimates $F(1)$ by considering maximal embedding, and an assumption about natural cover images provides $F(0)$; now there is enough information to determine the missing parameters and it is possible, given an image with unknown hidden data, to solve for (an estimate of) p.

The method of "RS" [2] is more general in that it uses groups of two or more pixels, but there is a still a quadratically varying quantity, similar algebraic manipulation, and an assumption about cover images sufficient to solve for the length of hidden data. The general framework above is explained in [2], in which similarities between the Sample Pairs method and the RS method are noted[1].

The detection framework we propose here is different: there is still a macroscopic property of images which depends on the length of hidden data, in this case a vector $\mathbf{F}(p)$; we will prove thoroughly how $\mathbf{F}(p)$ depends on p along with some unknown parameters; instead of trying to estimate the latter, we will *invert* the process: given an image we hypothesise a value for p and compute what this would imply for $\mathbf{F}(0)$. The other novelty is a model for cover images (or, more precisely, for the macroscopic properties of cover images). Then we can find the value of p which leads to a value of $\mathbf{F}(0)$ closest to the model: this is the estimator for p. The novel technique is suggested by the detector of [5], and this paper can be seen as a substantially generalised version of that work (which, like almost all other detectors, only considers pairs of pixels). The most important difference between the techniques presented here and most previous detectors is that g-tuples of pixels are considered in full generality. The functional form of $\mathbf{F}(p)$ will be a vector of polynomials of degree g.

This new framework includes the framework of [2] as a special case, and also subsumes the steganalysis methods of [3,4,5,6]. We do not give all the details here; briefly, the connection is made by collapsing the vector $\mathbf{F}(p)$ to a single quantity (by taking a certain linear function of its components, for example). With an appropriate selection of pixel groups and an appropriate linear function of $\mathbf{F}(p)$, each of the LSB steganography detectors in [2,3,4,5,6] are expressible in the new framework: their assumptions about the functional form of the relevant macroscopic property can be justified (and sometimes exposed as approximations), and an equation for the estimate of p derived. In some cases the derived equation is not quite identical to the original, but in each case it is possible to explain why the solutions are approximately equal. The case of RS is particularly interesting, because our new framework predicts a polynomial of degree g when the mask size is g pixles. It can be shown that, if the parameters of the RS method are chosen carefully, the higher coefficients vanish to leave a quadratic equation for p (just as in the standard RS method). However, the collapsing of the vector $\mathbf{F}(p)$ leads to less robust behaviour which our novel detector will avoid.

2 An Extensible Presentation of Sample Pairs

There are a number of equivalent ways to present the general framework, and we will do so using terminology somewhat similar to Sample Pairs Analysis in [3].

[1] There is some evidence in [1] that, despite the potentially more general form, the method of RS is slightly inferior to, or at best only as reliable as, that of Sample Pairs. But in any case the performance of the two methods is extremely close, so too for the methods of Pairs and Difference Histogram.

For clarity we will slightly alter some of the notation: we will use throughout caligraphic letters ($\mathcal{X}$) for sets, upper-case letters (X) for random variables, lower-case letters (x) for constants and realisations of random variables, and will make a clear distinction between nonrandom properties of a cover image and random properties of stego images based on that cover image (the randomness coming from the content and location of the hidden payload).

Suppose that a digital image consists of a series of samples $s_1, s_2, \ldots, s_N$ taking values in the range $0 \ldots 2M+1$ (typically $M = 127$). A *sample pair* is a pair (s_i, s_j) for some $1 \leq i \neq j \leq N$. Let $\mathcal{P}$ be a set of sample pairs; in [3] it is all pairs which come from horizontally or vertically adjacent pixels. Write $\mathcal{C}_m$ for the subset of $\mathcal{P}$ consisting of sample pairs where the sample values differ by exactly m after right-shifting by one bit (i.e. dividing by 2). Also write $\mathcal{X}_m$ for the sample pairs of $\mathcal{P}$ which differ in value by m with the higher value even and $\mathcal{Y}_m$ for those which differ by m but with the higher value odd. In this way, $\mathcal{P}$ is partitioned into subsets $\mathcal{C}_m$, $0 \leq m \leq M$, and each $\mathcal{C}_m$ is partitioned into $\mathcal{X}_{2m-1}, \mathcal{X}_{2m}, \mathcal{Y}_{2m}, \mathcal{Y}_{2m+1}$. In [3] $\mathcal{C}_m$ is referred to as a "submultiset", and $\mathcal{X}_m$ and $\mathcal{Y}_m$ as "trace submultisets"; we will use the simpler terminology "trace set" and "trace subset", respectively. Altering LSBs cannot affect which trace set a sample pair lies in, but it can move sample pairs between trace subsets as samples in the pair have their LSB flipped.

Suppose that a random hidden message of length $2pN$, where $0 \leq p \leq 0.5$ is unknown to the detector, is embedded using LSB replacement of a random selection of samples independent of the content of the cover or hidden message[2]. Suppose that a sample pair lies in trace set $\mathcal{C}_m$ with $m > 0$. The probability that neither sample is altered is $(1-p)^2$, the probabilities that either sample is altered is $p(1-p)$, and the probability that both are altered is p^2; each of these events moves the sample pair amongst the trace subsets of $\mathcal{C}_m$ according to the transition diagram Fig. 1 (transitions are labelled with their probabilities). $m = 0$ leads to a special case.

Let c_m, d_m, x_m, y_m be the cardinalities of the sets $\mathcal{C}_m$, $\mathcal{D}_m$, $\mathcal{X}_m$, $\mathcal{Y}_m$ in a particular cover image (they are nonrandom properties of the cover, but unknown to the detector), and let C'_m, D'_m, X'_m, Y'_m be the random variables representing the cardinalities of those sets after LSB embedding of a random message of length $2pN$. We know that $C'_m = c_m$ because $\mathcal{C}_m$ is closed under LSB operations. By considering the probabilistic transition systems in Fig. 1, Dumitrescu *et al.* derive the equations analogous to the following:

$$E\left[p^2 c_m - p(D'_{2m} + 2X'_{2m-1}) + X'_{2m-1}\right] = x_{2m-1}(1-2p)^2 \tag{1}$$

$$E\left[p^2 c_m - p(D'_{2m} + 2Y'_{2m+1}) + Y'_{2m+1}\right] = y_{2m+1}(1-2p)^2 \tag{2}$$

(In [3] the expectation is implicit.) x_m and y_m cannot be observed by a detector which only has access to the stego image but there is a plausible assumption:

$$x_{2m+1} = y_{2m+1} \text{ for all } m; \tag{3}$$

[2] In [3] calculations are performed assuming that pN is the hidden message length. The use of $2pN$ instead makes the following algebra somewhat simpler.

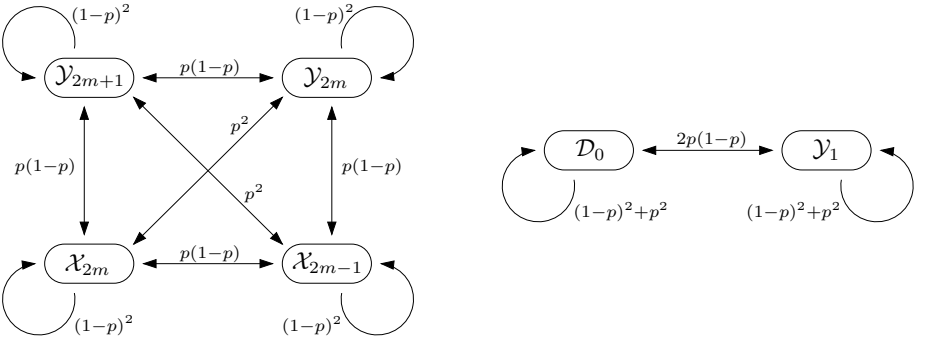

Fig. 1. Transitions between the subsets of $\mathcal{C}_m$, and the probability of each. Left, for all $m \geq 1$. Right, for $m = 0$, where $\mathcal{D}_0 = \mathcal{X}_0 \cup \mathcal{Y}_0$.

in [3] this assumption is cast as an expectation but the quantities involved are nonrandom if the cover image is fixed. It is plausible because sample pairs in a continuous tone image should not have any particular parity structure. Assuming that the observed values from the random variables are close to their expectations, (1), (2), and (3) give enough information to form quadratic equations for p, one for each m. In [3] these equations are summed to give a single quadratic, which is solved for an estimator $\hat{p}$ of p.

Our aim is to extend the sample pairs method to groups of pixels of more than two. For example, consider 3-tuples of adjacent samples, and trace sets $\mathcal{C}_{m,n}$ where the successive sample values differ by m and n, after right-shifting one bit. This does work, giving trace subsets and transition diagrams analogous to Fig. 1. But there are a number of awkward corners. Firstly, special cases proliferate: whereas for sample pairs there is the special transition diagram for $m = 0$, for 3-tuples we reach one special case for $m = 0, n \neq 0$, one for $m \neq 0, n = 0$ and another for $m = n = 0$. For g-tuples with $g \geq 4$ there are even more special cases. Secondly, the ad-hoc process by which Dumitrescu *et al.* derive (1) and (2) is difficult to generalise when the number of trace subsets rises. We will solve these problems by using a slightly modified version of the sample pairs technique.

2.1 A Modified, Extensible, Presentation

To remove the special case at $m = 0$ we have to take slightly finer distinctions in the trace sets and subsets:

$$\begin{aligned}\mathcal{C}_m &= \{(j,k) \in \mathcal{P} \mid \lfloor k/2 \rfloor = \lfloor j/2 \rfloor + m\} \\ \mathcal{E}_m &= \{(j,k) \in \mathcal{P} \mid k = j + m, \text{ with } j \text{ even}\} \\ \mathcal{O}_m &= \{(j,k) \in \mathcal{P} \mid k = j + m, \text{ with } j \text{ odd}\}\end{aligned}$$

with m now able to take negative values. $\mathcal{E}_m$ and $\mathcal{O}_m$ are analogous to $\mathcal{X}_m$ and $\mathcal{Y}_m$ but the new definitions break reflectional symmetry: no longer do we have (j,k) and (k,j) always belonging to the same set. The new transition diagram (probabilities included) is shown in Fig. 2. There are no special cases.

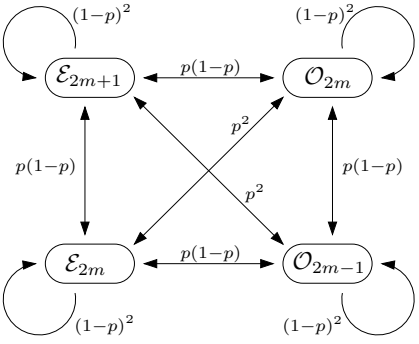

Fig. 2. Transitions between subsets of $\mathcal{C}_m$, in the modified presentation

Consider the random variable E'_{2m}, the cardinality of $\mathcal{E}_{2m}$ after a random message of length $2pN$ is embedded. It is actually the sum of four multinomial distributions, but we can reason about its expectation in an elementary manner. Sample pairs can enter $\mathcal{E}_{2m}$ in four ways: either having been in $\mathcal{E}_{2m}$ before and remaining there (and on average a proportion $(1-p)^2$ of the e_{2m} pairs in this position should remain), having been in $\mathcal{O}_{2m-1}$ before and moving to $\mathcal{E}_{2m}$ ($p(1-p)$ of the o_{2m-1} will do so), having been in $\mathcal{E}_{2m+1}$ ($p(1-p)$ of e_{2m+1} will do so), or having been in $\mathcal{O}_{2m}$ (p^2 of o_{2m} will do so). Thus,

$$E[E'_{2m}] = (1-p)^2 e_{2m} + p(1-p)o_{2m-1} + p(1-p)e_{2m+1} + p^2 o_{2m}.$$

We can repeat this for each of $O'_{2m-1}, E'_{2m+1}, O'_{2m}$ to get four linear equations which we express in vector form as

$$\begin{pmatrix} E[E'_{2m}] \\ E[O'_{2m-1}] \\ E[E'_{2m+1}] \\ E[O'_{2m}] \end{pmatrix} = \begin{pmatrix} (1-p)^2 & p(1-p) & p(1-p) & p^2 \\ p(1-p) & (1-p)^2 & p^2 & p(1-p) \\ p(1-p) & p^2 & (1-p)^2 & p(1-p) \\ p^2 & p(1-p) & p(1-p) & (1-p)^2 \end{pmatrix} \begin{pmatrix} e_{2m} \\ o_{2m-1} \\ e_{2m+1} \\ o_{2m} \end{pmatrix}. \tag{4}$$

The 4-by-4 matrix is the transition matrix of the transition system in Fig. 2 and it is invertible as long as $2p \neq 1$. If we make the assumption that the observed realisations of the random variables e'_{2m}, etc, are close to their expectations, we can form estimators for the unknown cover image quantities e_{2m}, etc.:

$$\begin{pmatrix} \hat{e}_{2m} \\ \hat{o}_{2m-1} \\ \hat{e}_{2m+1} \\ \hat{o}_{2m} \end{pmatrix} = \frac{1}{(1-2p)^2} \begin{pmatrix} (1-p)^2 & -p(1-p) & -p(1-p) & p^2 \\ -p(1-p) & (1-p)^2 & p^2 & -p(1-p) \\ -p(1-p) & p^2 & (1-p)^2 & -p(1-p) \\ p^2 & -p(1-p) & -p(1-p) & (1-p)^2 \end{pmatrix} \begin{pmatrix} e'_{2m} \\ o'_{2m-1} \\ e'_{2m+1} \\ o'_{2m} \end{pmatrix} \tag{5}$$

This has enabled us to hypothesise a value for p and then *undo* the effect of embedding a hidden message of length $2pN$. Certainly we could not expect to recover the cover image itself, but macroscopic properties of the cover, such as the cardinalities of the trace subsets, can be estimated in this way.

At this stage we must use some sort of "model" of cover images. The analogy to the sample pairs method would be $e_{2m+1} = o_{2m+1}$ for each m. Setting $\hat{e}_{2m+1} = \hat{o}_{2m+1}$ and using the relevant components of (5) (with m and with $m+1$) gives

$$(c_m - c_{m+1})p^2 + (e'_{2m+2} + o'_{2m+2} + 2o'_{2m+1} - e'_{2m} - o'_{2m} - 2e'_{2m+1})p + e'_{2m+1} - o'_{2m+1} = 0$$

for each m, which is analogous to Dumitrescu's equation. One can sum all these equations to reach an estimator for p: it is almost identical to the Sample Pairs estimator, the minor difference being due to the split between $\mathcal{C}_{-m}$ and $\mathcal{C}_m$.

Alternatively, we can consider deviations from $\hat{e}_{2m+1} = \hat{o}_{2m+1}$ to be "errors", and solve for p to find the closest image to our model by minimising the sum-square error $\sum(\hat{e}_{2m+1} - \hat{o}_{2m+1})^2$. Treating deviations from the assumptions as errors is a technique described in [5]. We will prefer this paradigm in the generalisation which follows, because it extends to more complex cover assumptions.

Note that we have not used an assumption that $e_{2m} = o_{2m}$. Although as plausible as $e_{2m+1} = o_{2m+1}$ it is not helpful in estimating p. It is easy to check, using (4), that when $e_{2m} = o_{2m}$ the same holds for stego images too. Therefore it does not provide any discrimination between cover and stego images.

3 Generalised Framework

We now generalise by considering g-tuples of sample values, for arbitrary g. The same overall method will be used: determination of the probabilities of transition between trace subsets, hypothesising a value for p, inverting the formula to express the cardinalities of the cover image trace subsets in terms of those of the stego image, using a model for cover images, and solving for p. Further investigation is needed to decide how optimally to apply the last step.

Suppose that a set of g-tuples of sample values $\mathcal{T}$ is selected (e.g. the intensities of all horizontal rows of g adjacent pixels). The trace sets and subsets are:

$$\begin{aligned}
\mathcal{C}_{m_1,\ldots,m_{g-1}} &= \{(s_1,\ldots,s_g) \in \mathcal{T} \mid \lfloor s_{i+1}/2 \rfloor = \lfloor s_i/2 \rfloor + m_i \text{ for each } 1 \le i < g\} \\
\mathcal{E}_{m_1,\ldots,m_{g-1}} &= \{(s_1,\ldots,s_g) \in \mathcal{T} \mid s_{i+1} = s_i + m_i, \text{ with } s_1 \text{ even}\} \\
\mathcal{O}_{m_1,\ldots,m_{g-1}} &= \{(s_1,\ldots,s_g) \in \mathcal{T} \mid s_{i+1} = s_i + m_i, \text{ with } s_1 \text{ odd}\}
\end{aligned}$$

Changing LSBs of samples cannot affect which of the trace sets the tuples inhabit, but they are moved between the trace subsets according to a 2^g-state transition process. It is convenient to write $\mathcal{A}_{0,m_1,\ldots,m_{g-1}}$ for $\mathcal{E}_{m_1,\ldots,m_{g-1}}$ and $\mathcal{A}_{1,m_1,\ldots,m_{g-1}}$ for $\mathcal{O}_{m_1,\ldots,m_{g-1}}$, and to abbreviate the subscripts using sequence notation (we write $\boldsymbol{s}$ for a sequence of integers and $\boldsymbol{s.t}$ for concatenation). We write $P(\mathcal{A}_{\boldsymbol{s}}, \mathcal{A}_{\boldsymbol{t}})$ for the probability of transition between $\mathcal{A}_{\boldsymbol{s}}$ and $\mathcal{A}_{\boldsymbol{t}}$.

We specify the trace subsets each set $\mathcal{C}_{m_1,\ldots,m_{g-1}}$ is divided into using induction on g, as follows. The base case is $g = 1$: there is a single trace set $\mathcal{C}$ (all individual samples) divided into two subsets $\mathcal{A}_0$ and $\mathcal{A}_1$, and $P(\mathcal{A}_0, \mathcal{A}_0) = P(\mathcal{A}_1, \mathcal{A}_1) = 1 - p$, $P(\mathcal{A}_0, \mathcal{A}_1) = P(\mathcal{A}_1, \mathcal{A}_0) = p$. If $\mathcal{C}_{\boldsymbol{t}}$ divides into trace subsets $\mathcal{A}_{\boldsymbol{s_1}}, \ldots, \mathcal{A}_{\boldsymbol{s_n}}$ then $\mathcal{C}_{\boldsymbol{t}.k}$ divides into

$$\mathcal{A}_{\boldsymbol{s_1}.(2k+\alpha_1)}, \ldots, \mathcal{A}_{\boldsymbol{s_n}.(2k+\alpha_n)}, \quad \mathcal{A}_{\boldsymbol{s_1}.(2k+\alpha_1+1)}, \ldots, \mathcal{A}_{\boldsymbol{s_n}.(2k+\alpha_n+1)}$$

where α_i is zero if the sum of components in $\boldsymbol{s_i}$ is even, and minus one otherwise. The transition probabilities are given by

$$\begin{aligned} P(\mathcal{A}_{\boldsymbol{s_i}.(2k+\alpha_i)}, \mathcal{A}_{\boldsymbol{s_j}.(2k+\alpha_j)}) &= (1-p)P(\mathcal{A}_{\boldsymbol{s_i}}, \mathcal{A}_{\boldsymbol{s_j}}) \\ P(\mathcal{A}_{\boldsymbol{s_i}.(2k+\alpha_i+1)}, \mathcal{A}_{\boldsymbol{s_j}.(2k+\alpha_j)}) &= pP(\mathcal{A}_{\boldsymbol{s_i}}, \mathcal{A}_{\boldsymbol{s_j}}) \\ P(\mathcal{A}_{\boldsymbol{s_i}.(2k+\alpha_i)}, \mathcal{A}_{\boldsymbol{s_j}.(2k+\alpha_j+1)}) &= pP(\mathcal{A}_{\boldsymbol{s_i}}, \mathcal{A}_{\boldsymbol{s_j}}) \\ P(\mathcal{A}_{\boldsymbol{s_i}.(2k+\alpha_i+1)}, \mathcal{A}_{\boldsymbol{s_j}.(2k+\alpha_j+1)}) &= (1-p)P(\mathcal{A}_{\boldsymbol{s_i}}, \mathcal{A}_{\boldsymbol{s_j}}) \end{aligned}$$

If the trace subsets are considered in the order given, the transition probabilities can be expressed concisely as matrices. The matrices, and their inverses, are g-fold Kronecker tensor products:

$$T_1 = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} \qquad T_1^{-1} = \frac{1}{1-2p}\begin{pmatrix} 1-p & -p \\ -p & 1-p \end{pmatrix}$$

$$T_{g+1} = \left(\begin{array}{c|c} (1-p)T_g & pT_g \\ \hline pT_g & (1-p)T_g \end{array}\right) \qquad T_{g+1}^{-1} = \frac{1}{1-2p}\left(\begin{array}{c|c} (1-p)T_g^{-1} & -pT_g^{-1} \\ \hline -pT_g^{-1} & (1-p)T_g^{-1} \end{array}\right)$$

Given a trace set $\mathcal{C}_{\boldsymbol{t}}$ divided into trace subsets $\mathcal{A}_{\boldsymbol{s_1}}, \ldots, \mathcal{A}_{\boldsymbol{s_n}}$, write a_i for the size of $\mathcal{A}_{\boldsymbol{s_i}}$ in the cover image, and A'_i for the size of the subset under random embedding of a message of length $2pN$. Just as with (4) we have

$$E[\mathbf{A}'] = T_g\mathbf{a}.$$

Observing the stego image we can count the realisations of the random variables A_i, say a'_i. Assuming that the realisations are close to their expectations, we can form estimators for the unknown values a_i:

$$\hat{\mathbf{a}} = T_g^{-1}\mathbf{a}'.$$

Note that T_g depends on p. Finally, we need a model for cover images, the analogue of (3); this may include $e_{\boldsymbol{s}} = o_{\boldsymbol{s}}$, although (as in Sect. 2) not all $\boldsymbol{s}$ will provide a useful discrimination between cover and stego images. We estimate p by finding the value which makes our estimate of $\hat{\mathbf{a}}$ the closest fit to the model. How best to do this depends on the cover image assumptions, but the technique of [5] (in which deviations are treated as errors and the sum-square error minimised) should generally be applicable.

4 Case Study: $g = 3$

The case $g = 1$ is degenerate. We have seen that the case $g = 2$ is very similar to the Sample Pairs method or the more robust modification of [5], depending on the cover image model used. We now consider the case $g = 3$, which we call *Triples Analysis* (by analogy with our name for SPA, "Couples"), showing in detail how the general framework can be used for steganalysis. Experimental

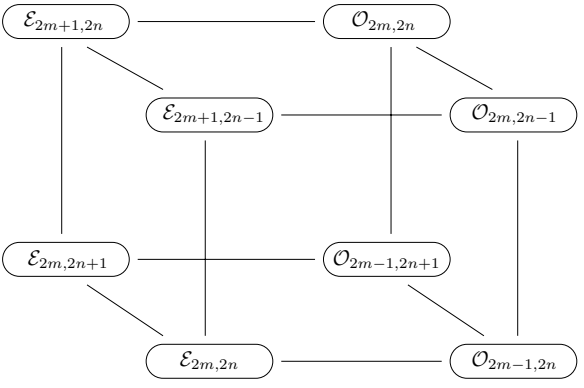

Fig. 3. The 8 trace subsets of $\mathcal{C}_{m,n}$. Subsets connected by an edge are related by the flipping of the LSB of exactly one sample in the 3-tuple.

results are included to demonstrate that the extension to 3-tuples provides a substantially more sensitive detector.

Fix a trace set $\mathcal{C}_{m,n}$; it is divided into 8 trace subsets. The full transition diagram contains a lot of information and we do not display all of it. Instead, in Fig. 3 we show how 3-tuples are moved amongst the trace subsets when a single sample has the LSB altered. In general, the probability of transition from one trace subset to another is $p^i(1-p)^{3-i}$, where i is the length of the shortest path between them in Fig. 3. If the trace subsets are enumerated in the order $\mathcal{E}_{2m,2n}$, $\mathcal{O}_{2m-1,2n}$, $\mathcal{E}_{2m+1,2n-1}$, $\mathcal{O}_{2m,2n-1}$, $\mathcal{E}_{2m,2n+1}$, $\mathcal{O}_{2m-1,2n+1}$, $\mathcal{E}_{2m+1,2n}$, $\mathcal{O}_{2m,2n}$ then the transition matrix is

$$T_3 = \begin{pmatrix}
(1-p)^3 & p(1-p)^2 & p(1-p)^2 & p^2(1-p) & p(1-p)^2 & p^2(1-p) & p^2(1-p) & p^3 \\
p(1-p)^2 & (1-p)^3 & p^2(1-p) & p(1-p)^2 & p^2(1-p) & p(1-p)^2 & p^3 & p^2(1-p) \\
p(1-p)^2 & p^2(1-p) & (1-p)^3 & p(1-p)^2 & p^2(1-p) & p^3 & p(1-p)^2 & p^2(1-p) \\
p^2(1-p) & p(1-p)^2 & p(1-p)^2 & (1-p)^3 & p^3 & p^2(1-p) & p^2(1-p) & p(1-p)^2 \\
p(1-p)^2 & p^2(1-p) & p^2(1-p) & p^3 & (1-p)^3 & p(1-p)^2 & p(1-p)^2 & p^2(1-p) \\
p^2(1-p) & p(1-p)^2 & p^3 & p^2(1-p) & p(1-p)^2 & (1-p)^3 & p^2(1-p) & p(1-p)^2 \\
p^2(1-p) & p^3 & p(1-p)^2 & p^2(1-p) & p(1-p)^2 & p^2(1-p) & (1-p)^3 & p(1-p)^2 \\
p^3 & p^2(1-p) & p^2(1-p) & p(1-p)^2 & p^2(1-p) & p(1-p)^2 & p(1-p)^2 & (1-p)^3
\end{pmatrix}$$

The inverse of T_3 consists of third order rational polynomials in p. A very convenient substitution is $q = 1/(1-2p)$; then we have (after some simplification)

$$T_3^{-1} = \frac{1}{8}\begin{pmatrix}
(1+q)^3 & (1-q)(1+q)^2 & (1-q)(1+q)^2 & (1-q)^2(1+q) & \cdots \\
(1-q)(1+q)^2 & (1+q)^3 & (1-q)^2(1+q) & (1-q)(1+q)^2 & \cdots \\
(1-q)(1+q)^2 & (1-q)^2(1+q) & (1+q)^3 & (1-q)(1+q)^2 & \cdots \\
(1-q)^2(1+q) & (1-q)(1+q)^2 & (1-q)(1+q)^2 & (1+q)^3 & \cdots \\
(1-q)(1+q)^2 & (1-q)^2(1+q) & (1-q)^2(1+q) & (1-q)^3 & \cdots \\
(1-q)^2(1+q) & (1-q)(1+q)^2 & (1-q)^3 & (1-q)^2(1+q) & \cdots \\
(1-q)^2(1+q) & (1-q)^3 & (1-q)(1+q)^2 & (1-q)^2(1+q) & \cdots \\
(1-q)^3 & (1-q)^2(1+q) & (1-q)^2(1+q) & (1-q)(1+q)^2 & \cdots
\end{pmatrix}. \qquad (6)$$

(Only half of T_3^{-1} is displayed, but the rest can be deduced by rotational symmetry). Given a stego image we consider each trace set $\mathcal{C}_{m,n}$ in turn and count the trace subsets to make a vector $\mathbf{x}'$. Then we can hypothesise a value of p and form estimates for the sizes of the trace subsets of the cover image using:

$$\hat{\mathbf{x}} = T_3^{-1}\mathbf{x}'.$$

In the case $g = 2$ there was just one property which we assumed that cover images have: $e_{2m+1} = o_{2m+1}$ for each m. In the case $g = 3$ there is an analogous property, which we will refer to as *parity symmetry*:

$$e_{m,n} = o_{m,n}$$

for each m and n. However, there are also some other plausible symmetries which might enrich our cover image model. One is *order symmetry*:

$$e_{m,n} = e_{n,m}$$

for each m and n (similarly $o_{m,n} = o_{n,m}$), and another is *reflectional symmetry*:

$$e_{m,n} = \begin{cases} e_{-n,-m}, & \text{if } m+n \text{ is even} \\ o_{-n,-m}, & \text{if } m+n \text{ is odd} \end{cases}$$

(and similarly for $o_{m,n}$ with even and odd swapped). Between them, the assumptions of order and reflectional symmetry state that pixels within groups can be considered in any order without changing the size of the trace subsets.

Recall, from Sect. 2, that some cover assumptions may not distinguish covers from stego images: this lead us to discard $e_{2m} = o_{2m}$ when $g = 2$. Here, it is routine to check that parity symmetry, if true for covers, is also true for stego images when m and n are both even, or $m = n$, and that order symmetry, if true for covers, is also true for stego images when *either* m *or* n is even, or $m = n$. Finally, reflectional symmetry never gives discrimination between covers and stego images.

Consider just one case of parity symmetry, $e_{2m+1,2n+1} = o_{2m+1,2n+1}$. To use the generalised framework to make an estimate of p, we compute "error terms" for each m and n, $\epsilon_{m,n} = \hat{e}_{2m+1,2n+1} - \hat{o}_{2m+1,2n+1}$. Then we find the value of p which minimises the sum-square of the errors. First, write

$$\begin{aligned}
d_0 &= e'_{2m+1,2n+1} - o'_{2m+1,2n+1} \\
d_1 &= e'_{2m+1,2n+2} + e'_{2m,2n+2} + o'_{2m,2n+1} - o'_{2m+1,2n} - o'_{2m+2,2n} - e'_{2m+2,2n+1} \\
d_2 &= e'_{2m,2n+3} + o'_{2m-1,2n+2} + o'_{2m,2n+2} - o'_{2m+2,2n-1} - e'_{2m+2,2n} - e'_{2m+3,2n} \\
d_3 &= o'_{2m-1,2n+3} - e'_{2m+3,2n-1}
\end{aligned}$$

(each d_i also depends on m and n, but in the interests of readibility we will leave these parameters implicit.) Then, using (6) and gathering similar terms,

$$\begin{aligned}
\epsilon_{m,n} &= \tfrac{1}{8}(d_0(1+q)^3 + d_1(1-q)(1+q)^2 + d_2(1-q)^2(1+q) + d_3(1-q)^3) \\
&= \tfrac{1}{8}((d_0 + d_1 + d_2 + d_3) + q(3d_0 + d_1 - d_2 - 3d_3) + \\
&\quad q^2(3d_0 - d_1 - d_2 + 3d_3) + q^3(d_0 - d_1 + d_2 - d_3))
\end{aligned}$$

We will find the value of q to minimise $S(q) = \sum_{m,n} \epsilon_{m,n}^2$. Writing

$$\begin{aligned} c_0 &= d_0 + d_1 + d_2 + d_3, & c_1 &= 3d_0 + d_1 - d_2 - 3d_3, \\ c_2 &= 3d_0 - d_1 - d_2 + 3d_3, & c_3 &= d_0 - d_1 + d_2 - d_3, \end{aligned}$$

(again leaving m and n implicit) we have

$$S(q) = \tfrac{1}{64} \textstyle\sum_{m,n} c_0^2 + q(2c_0c_1) + q^2(2c_0c_2 + c_1^2) + q^3(2c_0c_3 + 2c_1c_2) + q^4(c_2^2 + 2c_1c_3) + q^5(2c_2c_3) + q^6(c_3^2). \quad (7)$$

so that

$$S'(q) = \tfrac{1}{64} \textstyle\sum_{m,n} 2c_0c_1 + q(4c_0c_2 + 2c_1^2) + q^2(6c_0c_3 + 6c_1c_2) + q^3(4c_2^2 + 8c_1c_3) + q^4(10c_2c_3) + q^5(6c_3^2). \quad (8)$$

(To include other instances of parity symmetry or order symmetry in the cover model, the above calculations are repeated with the appropriate $\epsilon_{m,n}$ and included in the sum.) There will always be at least one real root of the quintic (8), but it could lead to up to 5 roots for q. We can discard implausible roots inside the range $(-10, 10/11)$ (because these would give obviously wrong estimates of p outside $(-0.05, 0.55)$) and substitute the remaining roots back into (7) to determine the location of the minimum $\hat{q}$. Finally, $\hat{p} = \frac{1}{2}(1 - \frac{1}{\hat{q}})$.

4.1 Experimental Results

Triples Analysis was implemented and widely tested. We comment on some implementation choices: we used all triples of horizontally adjacent pixels for $\mathcal{T}$ (some basic experiments indicate that it makes negligible difference if vertical groups are also included). For colour images the red, green and blue components were initially considered separately, and the trace subsets for each channel added together. After some initial experiments we found that it was marginally most accurate to use *only* the assumption $e_{2m+1,2n+1} = o_{2m+1,2n+1}$, although further research is needed to determine why this would be the case. Finally, we used only m, n in the range -5 to $+5$, because the trace sets are very small outside of this range.

Further initial experiments also indicate that the Triples estimator has a flaw: when the hidden message is long, the estimator gives wildly inaccurate results. In fact this can be explained theoretically, but for reasons of space we do not do so here. It is not a substantial problem: we can "screen" the method by first applying the standard SPA estimate, and proceeding to the Triples estimate only when the SPA estimate is below, say, 0.5[3]. In any case, our main interest is in the difficult case of detection when p is small.

[3] Here and hereafter we use p in the more usual way, to represent the proportionate length of the hidden message (previously this quantity was called $2p$). For screening, it seemed best to use a modification of the detector in [5] (which, as published, contains a few bugs) and proceed to the Triples estimate only for $\hat{p} < 0.3$.

Triples Analysis was compared with the standard methods of RS[4] and SPA. LSB Steganography was simulated on a number of sets of cover images, and detection statistics computed. To avoid overspecialisation, and in view of the wide variation in results depending on the cover image type noted in [1], we used a variety of sets of cover images:

Bitmap images: 3000 uncompressed bitmaps downloaded from `http://photogallery.nrcs.usda.gov`, very high resolution images apparently scanned from film, reduced in size to approximately 640×450.

JPEG images: three digital stock photo libraries: one of 5000 "high-quality" images, stored at quality factor 75, all sized 900×600; one of 10000 "medium-quality" images, stored at quality factors between 50 and 75, of similar size; and one of 20000 smaller "low-quality" images, 640×400, quality factor 58.

The experiments were repeated with each set of covers separately, and also with the bitmap images subject to JPEG compression prior to use as a bitmap cover image (to examine in isolation the effect of JPEG compression). Note that we have restricted our experiments to colour covers.

4.2 Reliability as an Estimator

We used the methods of RS, Sample Pairs and Triples to estimate the value of p, for each image in each set. This was repeated with the true value of p varying from 0 to 1, at intervals of 0.05. As with RS and Sample Pairs, the Triples estimator is approximately unbiased: for example, over the set of 3000 bitmaps the average error was observed to be between -0.016 and 0.009, depending on the true value of p.

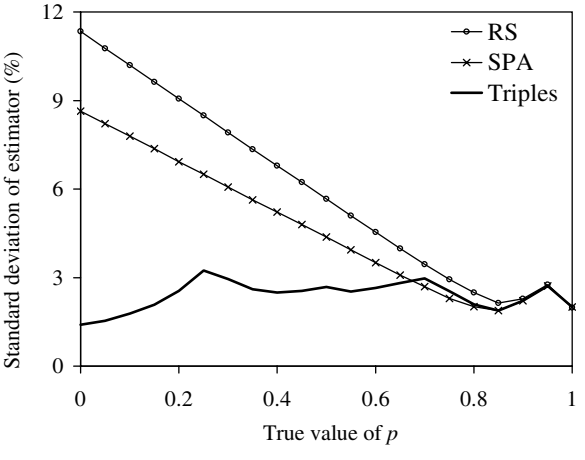

Fig. 4. Standard deviation of estimators, observed from a set of 3000 cover images subject to JPEG compression at quality factor 75, as p varies.

[4] The RS "mask" used was the standard $[0, 1, 1, 0]$, from [2].

Table 1. Standard deviation[5] of estimators ($\times 10^2$) when the true value of p is zero

Detector	3000 Uncompressed Bitmaps				Other sets of JPEG images		
	Unaltered	JPEG q.f.			5000 high quality	10000 med. quality	20000 low quality
		90	75	50			
RS	2.67	10.94	11.38	10.64	3.63	4.34	9.51
SPA	2.56	8.56	8.64	7.87	2.62	3.31	7.17
LSM	2.96	3.90	2.71	2.49	1.29	1.41	2.63
Triples	2.36	2.08	1.40	1.20	0.55	0.35	1.35

We compare the estimators by their sample standard deviation: one graph is shown in Fig. 4. We observe that, for this set of cover images, the Triples estimator is very substantially more accurate than the RS or SPA estimators in the case of small hidden messages. Indeed, it is surprising quite how unreliably the standard estimators performed – this is due to JPEG compression artefacts which cause the RS or SPA cover image assumptions to fail, whereas the Triples method treats errors in the cover assumption more robustly. Note that the poor performance of RS and SPA is mitigated as the hidden message length increases.

Rather than repeat such charts for every set of cover images, we merely compare the estimators when the true value of p is zero, i.e. for cover images. This gives a reasonable summary of relative performance, because the graph shapes are broadly similar in all cases, with the methods converging to similar performance near $p = 1$, and because small hidden messages are of particular interest (their detection being difficult). Table 1 shows this information for each cover image set, also including the robust modified SPA estimator of [5].

In the case of uncompressed bitmaps the Triples estimator is somewhat more accurate than RS or SPA. In the case of JPEG compressed covers it is very substantially more accurate: whereas the RS and SPA methods lose accuracy because of compression artefacts, the Triples method actually gains accuracy. Our conclusion is that the Triples method is more reliable and very much more robust to artefacts in the cover images.

4.3 Reliability as a Discriminator

A different question is how well the detector discriminates between the case $p = 0$ and $p = p_1$, for fixed p_1. In [1] it is shown that the discrimination problem is not necessarily optimally solved by an estimator for p. Following the methodology of [1] we should use a discriminator which shows how well an image under analysis matches the cover assumptions. This would simply be $S(0)$. However, $S(0)$ does not make a good discriminator because cover images vary (a lot!) in how well they meet the cover assumptions. A better discriminator is $S(0)/S(\hat{p})$ – this

[5] According to [9] the sample standard deviation is not a consistent estimator for measuring magnitude of error; subsequent experiments using a robust scale estimator yielded comparable results.

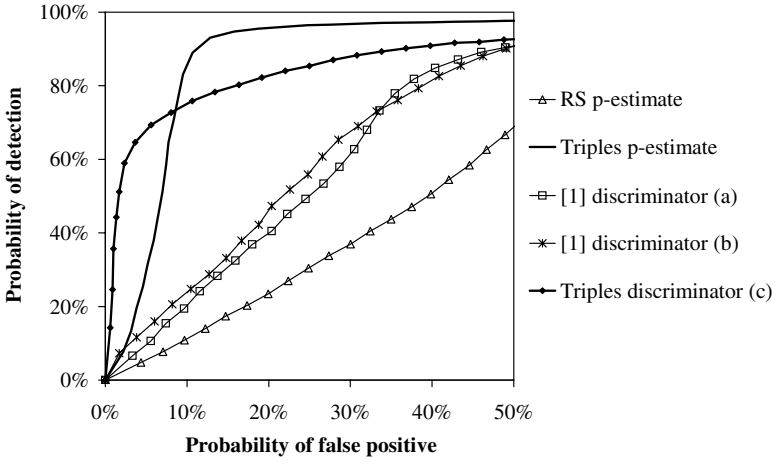

Fig. 5. Receiver Operating Characteristic curves, observed for the set of 3000 cover images subject to JPEG compression at quality factor 75. Random messages of length 2% of the maximum have been embedded. Detectors shown are the RS and Triples estimator of p, two discriminators from [1], and the Triples quantity $S(0)/S(\hat{p})$.

statistic should be near 1 for cover images and higher for stego images. We emphasise that this value is not merely testing whether $\hat{p} = 0$: it is a measure of *how certain* we are that $p \neq 0$; when the function S has a low gradient near $\hat{p}$ we should have correspondingly lower confidence in the estimate, and this is reflected in the quotient discriminator[6].

Performance of the discriminators, in the case $p_1 = 0.02$ and for one particular set of cover images, is displayed in Fig. 5. We have included the RS and Triples estimators of p, along with three discriminators which do not estimate p:

(a) from [1]: compute the estimate for p by applying the standard SPA calculation separately to each trace set $\mathcal{C}_i$, call it p_i; take the minimum of p_0, p_1 and p_2.
(b) from [1]: compute the relative difference of x_1' and y_1', these quantities as defined by the standard SPA method [3].
(c) novel: the ratio $S(0)/S(\hat{p})$ (including all four useful cover assumptions).

The embedding rate of 0.02 is below reliable detectability by the standard methods. In this case the Triples method is superior, and the discriminator which does not estimate p has even better performance at low false positive rates.

Again, it is impossible to include such graphs for each cover set and each value of p_1. Instead, we follow [1] by showing the lowest value of p_1 for which a certain (fairly arbitrary) level of reliability is achieved. This data is shown

[6] Some initial experiments suggested that, for discrimination, it was best to include all three useful cases of parity symmetry, and the one useful case of order symmetry, in the computation of S. This is in contrast to the case of simple estimation and some further work should be undertaken to investigate this.

Table 2. The lowest embedding rate p_1 for which "reliable" discrimination from $p = 0$ is achieved. Here, "reliable" is taken to mean a false positive rate of 5% and a false negative rate of 50%. Figures above 0.1 are accurate to 0.01; figures between 0.01 and 0.1 are accurate to 0.002; figures below 0.01 are accurate to 0.001.

Detector	3000 Uncompressed Bitmaps				Other sets of JPEG images		
	Unaltered	JPEG q.f.			5000 high quality	10000 med. quality	20000 low quality
		90	75	50			
RS	0.054	0.26	0.28	0.27	0.072	0.080	0.22
SPA	0.052	0.21	0.22	0.20	0.052	0.058	0.17
LSM	0.062	0.098	0.072	0.060	0.024	0.024	0.060
Triples	**0.042**	**0.040**	**0.026**	**0.018**	**0.010**	**0.005**	**0.016**
(a)	**0.028**	0.090	0.068	0.052	0.022	0.020	0.050
(b)	0.12	0.13	0.064	0.046	0.022	0.012	0.032
(c)	0.054	**0.016**	**0.012**	**0.018**	**0.006**	**0.003**	**0.009**

in Tab. 2. The Triples method is vastly superior in every case except for simple uncompressed covers, in which case discriminator (a) from [1] is most sensitive. This suggests that further improvements to the final stage of the Triples method may be possible, if a number of uncorrelated estimates for p can be produced.

5 Conclusions and Further Work

We have described a general framework for steganalysis of LSB Replacement, which can consider arbitrary tuples of pixels. It involves a new paradigm for detection, in which the effects of embedding a message of known length can be inverted, and a cover image model against which a best fit is found. The framework can include many of the previously known detectors, although we have not, in this paper, given the mathematical details of such relationships.

To demonstrate that the framework is worthwhile, we have tested one case, called Triples Analysis, which is a generalisation of the Sample Pairs/Couples method to include 3-tuples of pixels. It is necessary to screen the Triples method by first applying a standard estimator, because of inaccurate results when the hidden message length is high. A range of experiments verify that this makes for a reliable detector and estimator of hidden messages, performing somewhat better than the standard detectors on uncompressed covers, and very much better on images where the cover has artefacts. We conclude that it is a more robust detector, less prone to floods of false positive results caused by the cover type. (Although for reasons of space we have not included further experimental results, we observed that Triples Analysis maintains superior performance when the cover images are JPEGs which have been reduced in size – even when the reduction is as much as a factor of 5.)

Although the general framework uses all the structure of the LSB embedding method, it does not close the book on LSB detectors. We should apply it to tuples of pixels larger than 3, in the hope that even further improvements will result. However there are problems: increasing the group size g divides the set of tuples $\mathcal{T}$ into ever-smaller trace sets, and the assumption that random variables are close to their expectations causes errors when the law of large numbers cannot be relied upon (indeed, Triples Analysis already suffers with poor performance on very small images). It will be necessary to combine some of the trace sets, but to do so in a way which does not reduce the method back to the case of lower g. Further work is needed to identify how best to do this, how to produce the results at the final stage (i.e. whether minimising the sum-square error is optimal), and whether there are better ways to select $\mathcal{T}$ than simply horizontal rows of pixels.

Finally, we might hope to use denoising techniques to determine further information about the cover image, combining the best attributes of the structural detectors with those based on signal processing.

Acknowledgements

The author is a Royal Society University Research Fellow.

References

1. Ker, A.: Improved detection of LSB steganography in grayscale images. In: Proc. 6th Information Hiding Workshop. Volume 3200 of Springer LNCS. (2004) 97–115
2. 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
3. Dumitrescu, S., Wu, X., Wang, Z.: Detection of LSB steganography via sample pair analysis. In: Proc. 5th Information Hiding Workshop. Volume 2578 of Springer LNCS. (2002) 355–372
4. 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
5. Lu, P., Luo, X., Tang, Q., Shen, L.: An improved sample pairs method for detection of LSB embedding. In: Proc. 6th Information Hiding Workshop. Volume 3200 of Springer LNCS. (2004) 116–127
6. Zhang, T., Ping, X.: A new approach to reliable detection of LSB steganography in natural images. Signal Processing **83** (2003) 2085–2093
7. Harmsen, J., Pearlman, W.: Steganalysis of additive noise modelable information hiding. In Delp III, E.J., Wong, P.W., eds.: Security and Watermarking of Multimedia Contents V. Volume 5020 of Proc. SPIE. (2003) 131–142
8. Lyu, S., Farid, H.: Steganalysis using colour wavelet statistics and one-class vector support machines. In Delp III, E.J., Wong, P.W., eds.: Security, Steganography, and Watermarking of Multimedia Contents VI. Volume 5306 of Proc. SPIE. (2004) 35–45
9. Böhme, R.: Assessment of steganalytic methods using multiple regression models. In: Proc. 7th Information Hiding Workshop. Springer LNCS (2005)