An Analysis of Empirical PMF Based Tests For Least Significant Bit Image Steganography

Stark Draper, Prakash Ishwar, David Molnar, Vinod Prabhakaran, Kannan Ramchandran, Daniel Schonberg, and David Wagner*

University of California, Berkeley

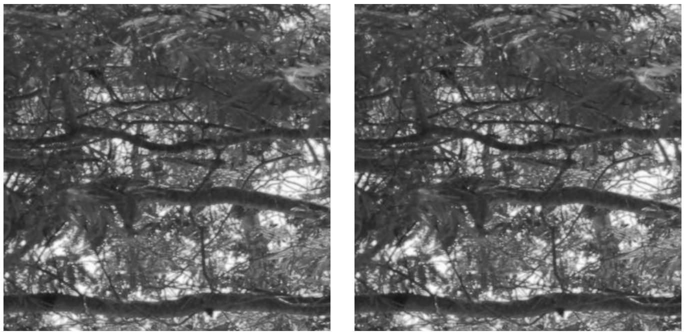
Abstract. We consider here the class of probability mass-function (PMF) based detectors of least significant bit (LSB) embedded steganography. That is, in this paper we investigate the use of frequency counts of pixel intensities as a statistic for tests detecting the presence of hidden messages. We focus on LSB replacement (though we briefly consider LSB matching) embedding as it is a simple technique where the effect on the true PMF of the resulting image can be understood mathematically. We begin our study by considering the existing tests of Westfeld and Pfitzmann [11] and Dabeer et al. [1]. These tests assume that pixel intensities are random values that are independent and identically distributed (i.i.d.). We generalize these tests by considering PMFs of neighboring pixel intensities. We argue that consideration of higher order of correlations provide only diminishing marginal returns, and thus we can make general statements on the value of PMF based detectors. We measure the performance of our tests by calculation of receiver operating curves (ROC) over a corpus of 350 digital images. We then proceed to compare to a non-PMF based test, in particular the RS tests of Fridrich et al [3]. Although our generalized tests outperform existing PMF based predecessors, they are outperformed by the RS tests. This indicates that using PMFs as a statistic for detecting hidden messages is inherently insufficient.

1 Introduction

Steganography is the practice of hiding a message in a covertext. Figure 1 shows an example of least significant bit (LSB) image steganography. While the images appear the same, the LSBs of each pixel in the image on the right have been replaced with a hidden message. Figure 2 depicts a block diagram in which message MSG is encoded into a string of length N (the length of the covertext in pixels). Random variable $\mathbf{M}^N$ denotes the encoded message. A stegoencoder takes $\mathbf{M}^N$, a covertext image denoted $\mathbf{X}^N$, and outputs a stegotext denoted $\mathbf{Y}^N$. The aim of *steganalysis* is to distinguish images with hidden messages from those without. Specifically, given a sample image $\mathbf{z}^N$ from either $\mathbf{X}^N$ or $\mathbf{Y}^N$, we want to decide whether $\mathbf{z}^N$ contains a hidden message or not.

* infohide@taverner.eecs.berkeley.edu. This research was supported by NSF under grant CCR-0325311.

In this paper we consider the class of probability mass function (PMF) based steganography detectors for LSB embedding. We restrict ourselves to PMF-based detectors of LSB embedding because it is possible to develop mathematically how the embedding impacts the PMF. By considering existing tests, and then generalizing new tests, we are able to make statements about the entire class of such detectors. Existing LSB PMF-based tests rely on an i.i.d. assumption about the observation. Our new tests use a first-order memory model in order to capture spatial correlations between neighboring pixels of the covertext. The advantage of considering memory is that it allows for greater separation between the embedded and covertext distributions. It is natural to believe that the greatest advantage in considering memory would be seen by this first-order generalization, and that further order increases would produce diminishing marginal gains. Note that this paper primarily focuses on LSB replacement embedding, though we briefly consider LSB matching.



(a) Normal (b) Embedded

Fig. 1. Example Photograph With LSB Embedding.

Our main contribution is to develop intuitions of how a higher order covertext model affects PMF-based steganalytic performance. To this end, we develop a first order memory generalization of the PMF covertext model and present three new tests based on our model. The first test is a generalization of the χ_2 test proposed by Westfeld and Pfitzmann and used by Provos and Honeyman in StegDetect [11] [8]. This test has the advantages of being simple and intuitive. Our next two tests are generalizations of the memoryless tests of Dabeer et al. [1]. The first of these generalizations is a *blind* test using the Kullback-Leibler (KL) divergence. The test is *blind* in that the test makes no assumptions about the specific the covertext. Our final test is an *informed* test, also based on the KL divergence. The informed test assumes knowledge (possibly only partial) of the covertext statistics. We build on the hypothesis testing framework of Dabeer et al. to obtain our tests. By developing this framework, we are able to show that our blind test is optimal for the class of first order memory models.

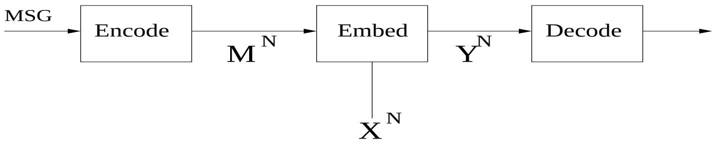


Fig. 2. Embedding message MSG into covertext $\mathbf{X}$ to obtain stegotext $\mathbf{Y}$

Our empirical results suggest that first order memory can lead to significant gains in the case of χ_2 tests, but only minor gains in the case of our blind tests. Since the gains from the consideration of the correlation are only marginal, this suggests that PMF based tests are inherently limited. When further comparing our tests against a non-PMF based test (the RS test of [3]), we see that the non-PMF based test exhibits far superior results, further discouraging the use of PMF based tests.

This paper is organized as follows. We begin by discussing simple PMF image models, define LSB replacement, and outline the hypothesis testing framework in Section 2. In Section 3 we discuss our higher order empirical PMFs and we introduce our covertext model and derive our new tests. Then, in Section 4, we empirically evaluate our tests, the test of Westfeld and Pfitzmann, the tests of Dabeer et al., and RS steganalysis. Finally, in Section 5 we consider LSB matching embedding and provide an extension of our framework to this embedding technique.

2 Background

In this paper, uppercase letters, such as X, refer to random variables, while lowercase letters, such as x, refer to instances of the corresponding random variable. Boldface letters, such as $\mathbf{p}$, refer to either vectors or matrices, with the meaning clear from context. A superscript on a character indicates vector length. For example, $\mathbf{X}^N$ is a vector-valued random variable of length N, $X^N = (x_1, \ldots, x_N)$.

Throughout this work, we assume that images are 8-bit grayscale throughout, with a set of pixel values from the alphabet $\mathcal{A} := \{0, 1, \ldots, 255\}$. Let $\mathcal{P}^0$ be the set of all probability mass functions (PMF) on $\mathcal{A}$. Let $\mathcal{P}'$ be the set of all joint PMFs on $\mathcal{A} \times \mathcal{A}$. Thus

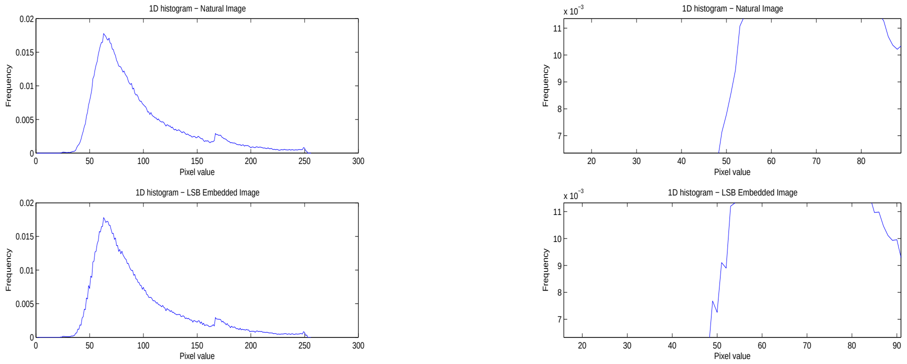
$$\mathcal{P} := \left\{ \mathbf{p} = (p_0, p_1, \ldots, p_{255}) \in \mathbb{R}^{256} : p_i \geq 0, \sum_{i=0}^{255} p_i = 1 \right\}$$

$$\mathcal{P}' := \left\{ \mathbf{p} = \begin{pmatrix} p_{0,0} & \cdots & p_{0,255} \\ \vdots & \ddots & \vdots \\ p_{255,0} & \cdots & p_{255,255} \end{pmatrix} \in \mathbb{R}^{256 \times 256} : p_{i,j} \geq 0, \sum_{i=0}^{255} \sum_{j=0}^{255} p_{i,j} = 1 \right\}$$

We will often speak of the *empirical PMF* $\mathbf{q}(\mathbf{z}^N)$ of a vector $\mathbf{z}^N \in \mathcal{A}^N$. The empirical PMF $\mathbf{q}(\mathbf{z}^N)$ on $\mathcal{A}$ is just the normalized histogram of the pixel values in $\mathbf{z}^N$. Thus $q_i = \frac{n_i}{\sum n_j}$ where n_j is the number of occurrences of value j in $\mathbf{z}^N$, i.e. $n_j = \#\{k \in \{1, \ldots, N\} : z_k = j\}$. Given a matrix $\mathbf{z}^N \in \mathcal{A}^N \times \mathcal{A}^N$, the first order empirical PMF $\mathbf{q}(\mathbf{z}^N)$ on $\mathcal{A} \times \mathcal{A}$ is the normalized histogram of $\frac{(N)(N-1)}{2}$ pairs of pixel values in $\mathbf{z}^N$. We will omit the $\mathbf{z}^N$ and write $\mathbf{q}$ when the meaning is clear from context.

2.1 LSB Replacement

We now proceed to define LSB replacement embedding. While an image with LSB replacement and the original image may be indistinguishable to the human eye, the empirical statistics vary. One can see visual evidence of this by comparing the histograms of pixel values between an original and a modified image; the modified image displays a pronounced "stair-step" effect. Example histograms for LSB replacement in which all pixels of an image have the least significant bit replaced with message data are shown in Figure 3.



(a) 1D Histogram

(b) Zoomed 1D

Fig. 3. Histograms showing effect of LSB embedding at rate 1 in an 8-bit grayscale image. Histograms from the original image are on the top row and histograms from the embedded image are on the bottom row. The plateaus giving the "stair-step" effect its name are pronounced in the embedded image.

When the LSB plane of the pixels are replaced with bits equally likely to be 0 or 1 (true at high rates), the probability of seeing the two pixel intensities whose other bit planes are equivalent are averaged. This is responsible for the "stair-step" character of embedded images. Though less pronounced at rates below one embedded bit per pixel, the effect is still prominent enough to allow detection. Below, we make this notion more precise by first defining LSB embedding of a message and then giving properties that must hold for an embedded image.

Let $\mathbf{m}^N$ be an encoded message, with $m_i \in \{0, 1, \emptyset\}$. The special $\emptyset$ symbol indicates a location of the message where no embedding is performed. For LSB replacement embedding at *rate* R, we define an encoding function $F_R : \{0,1\}^{NR} \rightarrow$

$\{0, 1, \emptyset\}^N$; this function "expands" an arbitrary message to N symbols with an expected fraction $1 - R$ of $\emptyset$ symbols. Let $\mathbf{x}^N$ be an instance from the covertext distribution $\mathbf{X}^N$. We then define the embedding function $\mathbf{y}^N := E(\mathbf{m}^N, \mathbf{x}^N)$. In the definition of this function below, we define only for the LSB plane, since all other bit planes of $\mathbf{y}^N$ are set equal to those of $\mathbf{x}^N$.

$$LSB(y_i) = \begin{cases} 1 & \text{if } m_i = 1 \\ 0 & \text{if } m_i = 0 \\ x_i & \text{if } m_i = \emptyset \end{cases}$$

If $\mathbf{X}^N$ has PMF $\mathbf{p} \in \mathcal{P}$ and we apply embedding $\mathbf{y}^N = E(\mathbf{m}^N, \mathbf{x}^N)$ at rate R, then $\mathbf{Y}$ will have the PMF $\mathbf{p}_R \in \mathcal{P}_R$ given by

$$\begin{array}{r} p_{R,2l} = (1 - \frac{R}{2})p_{2l} + \frac{R}{2}p_{2l+1} \\ p_{R,2l+1} = \frac{R}{2}p_{2l} + (1 - \frac{R}{2})p_{2l+1} \end{array}, \quad \text{for } l = 0, 1, \ldots, 127 \tag{1}$$

Notice that the larger the embedding rate R, the "smaller" the set of possible PMFs $\mathcal{P}_R$ (as shown by Dabeer et al. [1]). An alternative view of averaging over groups of consecutive pixel intensities is to see that it limits the ratio between those PMF values. The following relationships are derived.

$$\frac{R}{2 - R} \leq \frac{p_{R,2l}}{p_{R,2l+1}} \leq \frac{2 - R}{R}, \quad \text{for } l = 0, 1, \ldots, 127 \tag{2}$$

Thus we see that for all $p_R \in \mathcal{P}_R$ the ratio of pixel pairs is bounded. These bounds explain the "stair-step" effect observed in the histograms, since they limit how much pixels intensities within a group can deviate.

2.2 Hypothesis Testing

We now set up a hypothesis testing framework for tests to detect steganographic embedding. First we define the hypothesis that data is not embedded as $\mathcal{H}_0$ and the hypothesis that data is embedded at rate R as $\mathcal{H}_1$.

$$\mathcal{H}_0 : \mathbf{p} \in \mathcal{P} \setminus \mathcal{P}_R, \quad \mathcal{H}_1 : \mathbf{p} \in \mathcal{P}_R$$

A detector d_N is characterized by the acceptance region $A \subseteq \mathcal{A}^N$:

$$d(\mathbf{z}^N) = \begin{cases} \mathcal{H}_0 & \text{if } \mathbf{z}^N \in A, \\ \mathcal{H}_1 & \text{otherwise.} \end{cases}$$

We denote the false negative probability of a detector by P_1 and the false positive probability by P_2. These probabilities are defined as follows

$$P_1 := \Pr[\mathbf{x}^N \leftarrow \mathbf{X}^N, \mathbf{Y}^N = E(\mathbf{x}^N, \mathbf{m}^N), \mathbf{z}^N \leftarrow \mathbf{Y}^N : d(\mathbf{z}^N) = \mathcal{H}_0]$$
$$P_2 := \Pr[\mathbf{z}^N \leftarrow \mathbf{X}^N : d(\mathbf{z}^N) = \mathcal{H}_1]$$

Notice that the random variable $\mathbf{M}^N$ depends on the message encoding used. For detection to be nontrivial, the distribution over $\mathbf{X}^N$ must not be in the set $\mathcal{P}_R$.

We do not model $\mathbf{X}^N$ as a “random” distribution, thus we do not need to consider a distribution over distributions; instead we assume that the distribution of $\mathbf{X}^N$ is not in $\mathcal{P}_R$. This assumption is one we must make for this class of tests to work, and empirical results indicate that it is reasonable.

Now fix a constant $\lambda > 0$ and consider a sequence of detectors $\{d_1, d_2, \ldots\}$. The value of λ represents the false negative probability we are willing to tolerate, while simultaneously wishing to minimize the false positive probability.

We seek a sequence of detectors that minimizes $\liminf_{N\to\infty} -\frac{1}{N}\log(P_2)$ subject to the constraint $\liminf_{N\to\infty} -\frac{1}{N}\log(P_1) \geq \lambda$. Let $\mathbf{q}$ be the empirical PMF of a sample $\mathbf{z}^N$. Let $D(\mathbf{q}||\mathbf{p})$ be the Kullback-Leibler divergence between PMFs $\mathbf{p}$ and $\mathbf{q}$. Define $D(\mathbf{q}||\mathcal{P}_R) := \min_{\mathbf{p}\in\mathcal{P}_R} D(\mathbf{q}||\mathbf{p})$. We consider a sequence of detectors because by a result of Hoeffding we can specify a test which is asymptotically optimal as N goes to infinity [5]. The optimal test is the following

$$d_{OPT(\lambda)}(\mathbf{q}) = \begin{cases} \mathcal{H}_0 & \text{if } D(\mathbf{q}||\mathcal{P}_R) \geq \lambda, \\ \mathcal{H}_1 & \text{otherwise.} \end{cases}$$

where $\mathbf{q}$ is the empirical PMF derived from a sample $\mathbf{z}^N$.

3 Steganography Tests

Name	Threshold Condition	PMF Type
Memoryless blind	$D(\mathbf{q}\|\|\mathcal{P}_R) \geq \lambda$	Derived
StegDetect	$\chi_2(\mathbf{q}\|\|\mathcal{P}_1) \geq \lambda$	Derived
Memoryless informed	$D(\mathbf{q}\|\|\mathbf{p}_R) - D(\mathbf{q}\|\|\mathbf{p}) \geq \lambda$	Provided
RS Steganalysis	See [3]	Not Applicable

(a) Existing Tests

Name	Threshold Condition	PMF Type
First-order blind	$D(\mathbf{q}\|\|\mathcal{P}_R) \geq \lambda$	Derived
First-order χ_2	$\chi_2(\mathbf{q}\|\|\mathcal{P}_1) \geq \lambda$	Derived
First-order informed	$D(\mathbf{q}\|\|\mathbf{p}_R) - D(\mathbf{q}\|\|\mathbf{p}) \geq \lambda$	Provided

(b) New Tests

Fig. 4. Summary of tests. In the charts above, the middle column gives the threshold condition under which the tests outputs $\mathcal{H}_0$. In the last column in the charts above, an explanation of which type of PMF the input is compared to is given.

In this section we present both existing tests and our extensions of them, a summary is given in Figure 4. Note that each of these tests compares a derived statistic to a threshold value λ. Specifically, in the following tests we compute a statistic $\alpha \in \mathbb{R}$. Given a statistic α, the derived test d_λ works as follows:

$$d_\lambda = \begin{cases} \mathcal{H}_0 & \text{if } \alpha \geq \lambda, \\ \mathcal{H}_1 & \text{otherwise.} \end{cases}$$

Each of the tests below takes the empirical PMF $\mathbf{q}$ as an input. Further note throughout this section we define $0/0 = 1$ for convenience.

3.1 Properties of Neighboring Pixel PMFs

Before proceeding to the tests, we present background on first order memory PMFs. The relation between a first order PMF $\mathbf{p} \in \mathcal{P}'$ and an embedded first order $\mathbf{p}_R \in \mathcal{P}'_R$ is

$$\begin{aligned}
p_{R,2l,2k} &= (1-\tfrac{R}{4})p_{2l,2k} + \tfrac{R}{4}p_{2l,2k+1} + \tfrac{R}{4}p_{2l+1,2k} + \tfrac{R}{4}p_{2l+1,2k} \\
p_{R,2l,2k+1} &= \tfrac{R}{4}p_{2l,2k} + (1-\tfrac{R}{4})p_{2l,2k+1} + \tfrac{R}{4}p_{2l+1,2k} + \tfrac{R}{4}p_{2l+1,2k} \\
p_{R,2l+1,2k} &= \tfrac{R}{4}p_{2l,2k} + \tfrac{R}{4}p_{2l,2k+1} + (1-\tfrac{R}{4})p_{2l+1,2k} + \tfrac{R}{4}p_{2l+1,2k} \\
p_{R,2l+1,2k+1} &= \tfrac{R}{4}p_{2l,2k} + \tfrac{R}{4}p_{2l,2k+1} + \tfrac{R}{4}p_{2l+1,2k} + (1-\tfrac{R}{4})p_{2l+1,2k}
\end{aligned} \qquad (3)$$

$$\text{for } l = 0, 1, \ldots, 127, \quad k = 0, 1, \ldots, 127$$

As in the memoryless case (Eq. (1)), where pixel intensities were considered in groups of two, from Eq. (3) it is clear we now consider pixel intensities in blocks of four (2×2). As with the memoryless case, bounds on the ratios between pixel intensity frequencies can be generated for PMFs residing in $\mathcal{P}_R$.

$$\begin{array}{ll}
\frac{R}{4-3R} \le \frac{p_{R,2l,2k+1}}{p_{R,2l,2k}} \le \frac{4-3R}{R} & \frac{R}{4-3R} \le \frac{p_{R,2l+1,2k}}{p_{R,2l,2k+1}} \le \frac{4-3R}{R} \\
\frac{R}{4-3R} \le \frac{p_{R,2l+1,2k}}{p_{R,2l,2k}} \le \frac{4-3R}{R} & \frac{R}{4-3R} \le \frac{p_{R,2l+1,2k+1}}{p_{R,2l,2k+1}} \le \frac{4-3R}{R} \\
\frac{R}{4-3R} \le \frac{p_{R,2l+1,2k+1}}{p_{R,2l,2k}} \le \frac{4-3R}{R} & \frac{R}{4-3R} \le \frac{p_{R,2l+1,2k+1}}{p_{R,2l+1,2k}} \le \frac{4-3R}{R}
\end{array} \qquad (4)$$

As in the memoryless case, examining histograms of images with high rate LSB embedding gives visual evidence that empirical statistics differ. Figure 5 exhibits the "blockiness" induced by the PMF averaging of high rate embedding.

3.2 Chi-Squared Tests

In order to obtain a PMF for the basis of comparison, χ_2 tests calculate $\mathbf{p}^*$. $\mathbf{p}^*$ is by applying Equations 1 and 3 (respective of PMF dimension) at rate $R = 1$. Once $\mathbf{p}^*$ is generated, a χ_2 distance from the empirical PMF is measured. The formula for a χ_2 distance is

$$\chi_2(\mathbf{p}, \mathbf{q}) = \sum_{i=0}^{255} \frac{(q_i - p_i^*)^2}{p_i^*}$$

In this section, we also discuss the Westfeld and Pfitzmann statistic used by StegDetect here [11] [8]. StegDetect does not explicitly generate a comparison PMF but, as we show, is equivalent to the χ_2 test.

The StegDetect test of Westfeld and Pfitzmann[8] is defined here.

$$\alpha = \sum_{k=0}^{127} \frac{(q_{2k+1} - q_{2k})^2}{q_{2k+1} + q_{2k}}$$

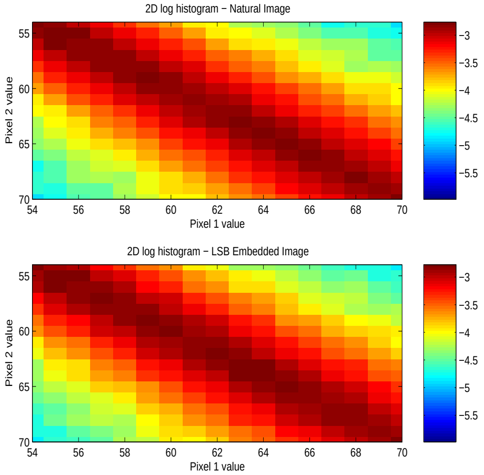


Fig. 5. Two-dimensional histograms of normal and embedded images. Note the "blockiness" effect in the histogram for the embedded image.

The memoryless χ_2 test is defined as follows.

$$\alpha = \chi_2(\mathbf{p}^*, \mathbf{q}) = \sum_{i=0}^{255} \frac{(q_i - p_i^*)^2}{p_i^*}$$

$$p_{2k}^* = p_{2k+1}^* = \frac{q_{2k+1} + q_{2k}}{2}, \quad \text{for } k = 0, 1, \ldots, 127$$

Finally, the first order χ_2 test is as follows.

$$\alpha = \chi_2(\mathbf{p}^*, \mathbf{q}) = \sum_{i=0}^{255} \sum_{j=0}^{255} \frac{(q_{i,j} - p_{i,j}^*)^2}{p_{i,j}^*}$$

$$p_{2l,2k}^* = p_{2l+1,2k}^* = p_{2l,2k+1}^* = p_{2l+1,2k+1}^* = (q_{2l,2k} + q_{2l+1,2k} + q_{2l,2k+1} + q_{2l+1,2k+1})/4$$
$$\text{for } l = 0, 1, \ldots, 127, \quad k = 0, 1, \ldots, 127$$

As can be seen above, the first order χ_2 test is the natural extension of its memoryless χ_2 counterpart. All three tests have the advantage of being computationally simple and easy to implement. Since these tests do not consider rate, however, they suffer when lower embedding rates are employed.

The StegDetect and the memoryless χ_2 test are in fact the same test, as proven mathematically below.

$$\chi_2(\mathbf{p}^*, \mathbf{q}) = \sum_{i=0}^{255} \frac{(q_i - p_i^*)^2}{p_i^*} = \sum_{k=0}^{127} \left(\frac{(q_{2k} - p_{2k}^*)^2}{p_{2k}^*} + \frac{(q_{2k+1} - p_{2k+1}^*)^2}{p_{2k+1}^*} \right) \quad (5)$$

$$= \sum_{k=0}^{127} \frac{(q_{2k} - \frac{q_{2k}+q_{2k+1}}{2})^2 + (q_{2k+1} - \frac{q_{2k}+q_{2k+1}}{2})^2}{\frac{q_{2k}+q_{2k+1}}{2}} \tag{6}$$

$$= \sum_{k=0}^{127} \frac{(q_{2k} - q_{2k+1})^2 + (q_{2k+1} - q_{2k})^2}{2(q_{2k} + q_{2k+1})} = \sum_{k=0}^{127} \frac{(q_{2k} - q_{2k+1})^2}{q_{2k+1} + q_{2k}} \tag{7}$$

3.3 Blind Tests

Unlike the test above, our "blind" tests (blind of the covertext PMF) require knowledge of the rate R. Following our hypothesis testing framework, this test finds the distance between the observed PMF and the set of rate R embedded PMFs $\mathcal{P}_R^i$. Our memoryless Blind tests uses the statistic $\alpha = D(\mathbf{q}||\mathcal{P}_R)$ while our first order Blind test uses $\alpha = D(\mathbf{q}||\mathcal{P}'_R)$.

The difficulty of computing these statistics efficiently lies in finding the PMF $\mathbf{p} \in \mathcal{P}_R$ (or $\mathbf{p} \in \mathcal{P}'_R$) that is closest to the observed PMF in $\mathcal{P}_R$ or $\mathcal{P}_R$ respectively. For the memoryless case, Dabeer et al. gave an algorithm for calculating the PMF $\mathbf{p}^*$ that minimizes this distance [1]. Pseudocode is given in Figure 6. Intuitively, this algorithm checks each set of PMF values to see if they violate Eq. (2). In the case that the conditions are not violated, the values of $\mathbf{p}^*$ are set equal to $\mathbf{q}$. If the conditions are violated though, the ratio of the two values are set equal to the violated bound while their sum is held constant.

For the first order memory test, the algorithmics become more involved. Let $\mathbf{q}$ denote the observed first order PMF. Pseudocode for computing $\mathbf{p}^*$ is shown in Figure 7.

```
Algorithm 3.1: FindPStarMemoryless(q)

for $l \leftarrow 0$ to 127
if $\frac{q_{2l+1}}{q_{2l}} > \frac{2-R}{R}$
  then $p^*_{2l} = \frac{R}{2}(q_{2l} + q_{2l+1}) and p^*_{2l+1} = (1 - \frac{R}{2})(q_{2l} + q_{2l+1})$
  else if $\frac{q_{2l+1}}{q_{2l}} < \frac{R}{2-R}$
  then $p^*_{2l} = (1 - \frac{R}{2})(q_{2l} + q_{2l+1}) and p^*_{2l+1} = \frac{R}{2}(q_{2l} + q_{2l+1})$
  else $p^*_{2l} = q_{2l}$ and $p^*_{2l+1} = q_{2l+1}$
```

Fig. 6. Algorithm for computing $\mathbf{p}^*$ for memoryless sources.

In this code, if a 2×2 block of pixels does not violate the conditions of Eq. (4), then the block is transferred to the $\mathbf{p}^*$. If the conditions of Eq. (4) are violated, then the values of $\mathbf{q}$ are scaled so that the block satisfies those conditions while maintaining the same summation and relative scale. By following this procedure, the distance between the observed PMF $\mathbf{q}$ and the set $\mathcal{P}_R$ is found.

These tests have two significant advantages over the simple χ_2 tests presented earlier. First, they are able to exploit knowledge of the rate R. This allows for strong performance at all rates. Second, the KL divergence is a more complete "distance" then the χ_2 metric.

```
**Algorithm 3.2:** FindPStarMemory($\mathbf{q}$):
for $l \leftarrow 0$ to $127$
for $k \leftarrow 0$ to $127$
    $Q \leftarrow \{q_{2l,2k}, q_{2l,2k+1}, q_{2l+1,2k}, q_{2l+1,2k+1}\}$
    $\{w_1, w_2, w_3, w_4\} \leftarrow SortHighToLow(Q)$
    if $\frac{w_1}{w_4} > \frac{4-3R}{R}$ then
        $\tilde{w}_i \leftarrow 4\left(\frac{w_i - w_4}{w_1 - w_4}\right)\left(\frac{1-R}{R}\right) + 1$
        $\hat{w}_i \leftarrow \tilde{w}_i \frac{\sum_{j=1}^{4} w_j}{\sum_{j=1}^{4} \tilde{w}_j}$
        $\tilde{W} \leftarrow \{w_1, w_2, w_3, w_4\}$
        $\{p^*_{2l,2k}, p^*_{2l,2k+1}, p^*_{2l+1,2k}, p^*_{2l+1,2k+1}\} \leftarrow InvertOrderOfSortOperation(\tilde{W})$
    else $\{p^*_{2l,2k}, p^*_{2l,2k+1}, p^*_{2l+1,2k}, p^*_{2l+1,2k+1}\} \leftarrow \{q_{2l,2k}, q_{2l,2k+1}, q_{2l+1,2k}, q_{2l+1,2k+1}\}$
```

Fig. 7. Algorithm for computing $\mathbf{p}^*$ for first-order sources.

3.4 Informed Tests

In addition to the empirical PMF and target rate, "informed" tests take in information on the covertext PMF $\mathbf{p}$. These tests compare the KL distances from both the given distribution and the rate R expected embedded distribution. For the memoryless case, the test takes in an input PMF $\mathbf{p}$ and observed PMF $\mathbf{q}$ and then runs as follows.

- Calculate the PMF $\mathbf{p}_R \in \mathcal{P}_R$ which would result if we applied rate R embedding (with a random message) to a covertext with PMF $\mathbf{p} \in \mathcal{P}$. This can be done according to Equation 1.
- Calculate $\alpha = D(\mathbf{q}||\mathbf{p}_R) - D(\mathbf{q}||\mathbf{p})$.

The first order memory case can be described similarly to the memoryless case. The test takes in an input PMF $\mathbf{p}$ and observed PMF $\mathbf{q}$ and then runs as follows.

- Calculate the PMF $\mathbf{p}_R \in \mathcal{P}'_R$ which would result if we applied rate R embedding (with a random message) to a covertext with PMF $\mathbf{p} \in \mathcal{P}'$. This can be done according to Equation 3.
- Calculate $\alpha = D(\mathbf{q}||\mathbf{p}_R) - D(\mathbf{q}||\mathbf{p})$.

Note that these tests opperate even if $\mathbf{p}$ is only an approximation to the true distribution of the covertext. The tests are robust to small errors in $\mathbf{p}$; however, the quality of the test is limited by the quality of the supplied distribution. Careful choice of distribution is crucial when using an informed test.

4 Empirical Results for LSB Replacement

We calculated these tests on a set of 350 images chosen from a larger corpus of 2977 Portable Network Graphics (png) images obtained from Sullivan et al. [2]. We embed at 3 rates; $R = 0.05$, $R = 0.5$, and $R = 1.0$. The messages were generated randomly according to a Bernoulli(1/2) i.i.d. random process. We ran each of the tests given in Section 3 on the embedded images.

Based on this data, receiver operating curves (ROC) were generated by calculating false negative and positive rates from a variety of thresholds. We provide the resulting ROCs in Figure 9. In the ROC plots, the results for the χ_2 tests (on the left) and the blind tests (on the right) at Rates 0.5 and 0.05 are presented. The memoryless tests are plotted with solid lines while the memory model tests are plotted with dashed lines. The $R = 0.05$ results are always higher in the plot. For every test on the rate $R = 1$ embedded data, the ROC curves are nearly ideal. Since these would be difficult to see in the plots, they have been omitted.

As expected, StegDetect and the memoryless χ_2 tests resulted in the same parameter value. As a result, the StegDetect lines are omitted below. Further, we see that the memory version of the χ_2 square test outperforms the memoryless version at rate $R = 0.5$. Since $R = 0.05$ is such a low rate both tests perform very poorly, though moderate performance gain from the memory test is observed.

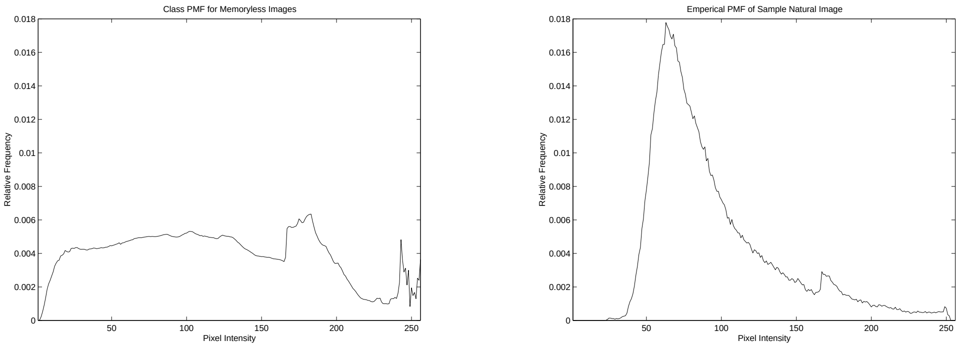
The memory blind test outperforms the memoryless blind test at rate $R = 0.5$. Here, the improvement in performance is less significant though, and they perform nearly identically for large false negative rates. As with the χ_2 tests, both blind tests perform nearly identically for $R = 0.05$, with the memoryless test slightly outperforming the memory test. Both tests perform poorly though due to the low rate. Note that at each rate, the blind tests outperform the χ_2 tests. This agrees with intuitions, since the χ_2 tests assume rate $R = 1$ embedding.

The prior distributions we chose were obtained by averaging all the PMFs within our image corpus. Unfortunately, because our corpus is large, the resulting prior PMF did not match any single image well. In Figure 8 we show histograms of the prior distribution used for the informed test and the histogram of a particular image file to illustrate these differences. As a result of the choice of prior distribution we made (though it was the natural choice), each informed tests performed very poorly at every rate. Each time, the test results were nearly equivalent to random guessing. This suggests that proper choice of prior distribution for the informed test is essential.

Finally, we comment on the results of RS steganalysis. Our results showed RS steganalysis performing well in all cases, showing ideal or near ideal performance. We were able to draw two tentative conclusions from this result. First, the RS test has an advantage by trying to estimate the rate at which the embedding occurred as opposed to a parameter off which to make a decision. Second, there seems to be an advantage in working off the true data instead of the PMF as a summary statistic. This would suggest that a better model for steganalysis would be to include even higher order relationships, as the RS test does.

5 LSB Matching

Above, we have focused exclusively on LSB replacement embedding. An alternate form of LSB embedding is LSB matching [7]. In this section we briefly describe LSB matching and describe some results of our steganalysis framework against LSB matching on a corpus of digital camera images. As before, we encode a



(a) (b)

Fig. 8. Comparison of PMF used for informed test (a) and a sample PMF for a natural image (b). As can be seen in the image on the left, by averaging over the entire corpus of images a roughly flat PMF is achieved, significantly differing from the empirical PMF of any actual image.

message MSG into $\mathbf{m}^N$ at rate R with $m_i \in \{0, 1, \emptyset\}$. LSB matching is defined as follows:

$$y_i = \begin{cases} x_i \pm 1 & \text{if } LSB(y_i) \neq m_i \\ x_i & \text{if } LSB(y_i) = m_i \\ x_i & \text{if } m_i = \emptyset \end{cases}$$

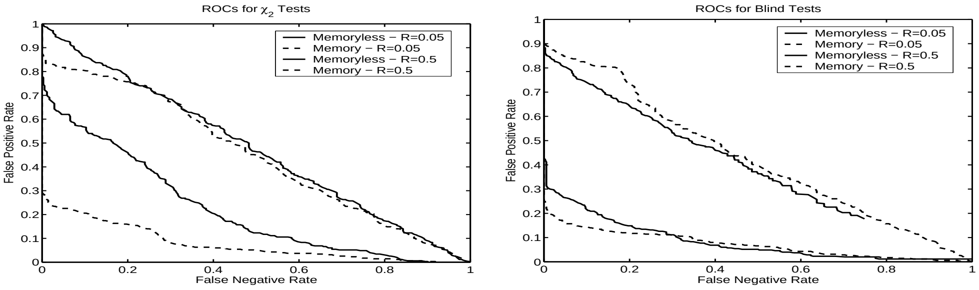
Whenever x_i is not one of the extreme values (i.e., $x_i \in \{0, 255\}$) the choice between addition and subtraction is made at random (each option being equally likely).

The effect of LSB matching on the PMF of an image contrasts with that of LSB replacement. Instead of the "stair-step" effect induced by replacement, LSB matching has a smoothing effect. Ignoring the "edge-effects" due to the embedding rule at the extreme values of the PMF, we can write $\mathbf{p}_R$ as a filtered version of $\mathbf{p}$.

$$\begin{array}{cc} \text{For memoryless PMFs} & \mathbf{p}_R = \mathbf{p} * \mathbf{f} \\ \text{For neighboring pixel PMFs} & \mathbf{p}_R = \mathbf{p} * \mathbf{f} * \mathbf{f}^T \\ \text{where} & \mathbf{f} = (R/4, 1 - R/2, R/4) \end{array}$$

In these equations, the symbol $*$ denotes convolution and the superscript $\mathbf{f}^T$ indicates transpose. Since embedding is done on a pixel by pixel basis, the neighboring pixel PMF equation consists of a convolution on each dimension. As can be seen by examining this filter, the filter is a moving average filter and thus has a smoothing effect. As with LSB replacement, as the rate increases the space of possible PMFs shrinks.

It is clear that simply applying the tests above designed for LSB replacement will be sub-optimal. In Figure 10 we demonstrate the results of applying our tests to 50 digital camera images. We note that the RS steganalysis offers no performance advantage at rate 1 and performs slightly worse than our tests at



(a) ROC Curves, χ_2 Tests

(b) ROC Curves, Blind Tests

Fig. 9. ROC Curves for our tests applied to LSB replacement, over 350 digital camera images. In each of these plots, the rate $R = 0.5$ are plotted in a darker shade while the rate $R = 0.05$ are plotted in a lighter shade (and are consistently higher in the plot). In addition, memoryless tests are plotted with solid lines while memory tests are plotted in dashed lines. Results for all other tests are omitted for clarity.

rate 0.5. This is because RS steganalysis takes advantage of structural properties that are present only with LSB replacement and not LSB matching. Ker gives a more detailed overview of these properties [6]. Developing optimal tests for LSB matching (particularly in the blind case) is a part of ongoing work.

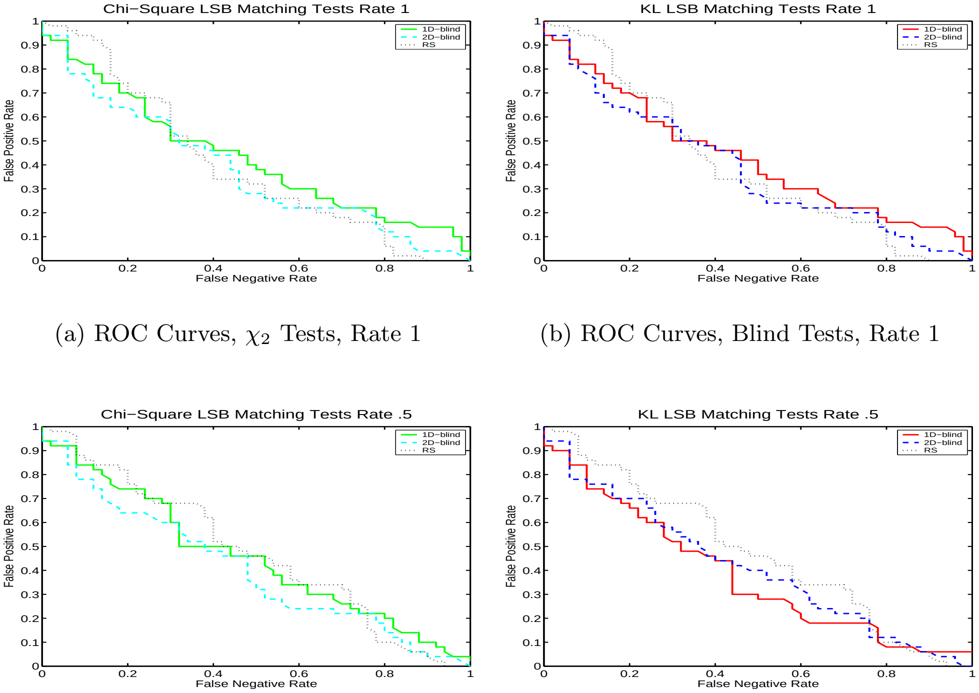
6 Related Work

Early work in detection of LSB embedding was done by Westfeld and Pfitzmann, who proposed a χ_2 test for detection [11]. Later, Westfeld introduced a generalization of the χ_2 test that succeeds even at low embedding rates; this test works by "hashing" parts of the image into different combinations, then running a test on each individual combination [10]. Westfeld also pointed out the possibility of estimating the *length* of embedded data, which our tests do not provide.

Provos and Honeyman created the StegDetect system and searched millions of pictures from the Internet looking for steganographic embedding using the Westfeld and Pfitzmann test [8]. Dabeer et al. introduced the hypothesis testing framework and proved the KL divergence test is the optimal LSB detector for memoryless covertext distributions [1]. Fridrich et al. analyzed the RS test, which takes into account spatial correlations of individual pixels [3]. Fridrich and Goljan later proposed another method based on local estimators that has similar performance to the RS test, but admits a cleaner theoretical derivation [4].

7 Conclusions and Future Directions

The gains achieved by considering the first order memory model over a memoryless assumption suggest that the extensions to higher-order memory models



(a) ROC Curves, χ_2 Tests, Rate 1

(b) ROC Curves, Blind Tests, Rate 1

(c) ROC Curves, χ_2 Tests, Rate 0.5

(d) ROC Curves, Blind Tests, Rate 0.5

Fig. 10. ROC Curves for our tests on LSB matching, over 50 digital camera images. All embedding was performed at rate 1 and rate 0.5. Memoryless tests are plotted with solid lines while memory tests are plotted in dashed lines. The RS test curve for each rate is included on all graphs.

for images would be of only limited value. Although offering the advantage of a rigorous mathematical framework, PMF based steganalysis seems inferior to RS steganography, at least on the class of test images we have considered. One area to explore would be other cover text models, such as the graphical Wainwright-Simoncelli-Willsky wavelet tree model [9]. Such a model may offer advantages in generating sufficient statistics for steganography since the model is not restricted to pixel intensity frequency counts.

8 Acknowledgments

We would like to thank Martin Wainwright for his valuable discussions on this topic. We also thank Jessica Fridrich, Miroslav Goljan, Andrew Ker, the other attendees, and the anonymous reviewers of the Information Hiding Workshop for feedback and suggestions. Finally, we thank the UCSB Vision Group for graciously allowing us to use their digital camera image set.

References

1. O. Dabeer, K. Sullivan, U. Madhow, S. Chandrasekaran, and B. S. Manjunath. Detection of hiding in the least significant bit. In *IEEE Trans. on Signal Processing*, volume 52, pages 3046–3058, Oct. 2004.
2. UCSB Vision Group; Sullivan et al., 2004. Collection of 2977 digital camera images.
3. J. Fridrich and M. Goljan. Practical steganalysis of digital images - state of the art. In *Proceedings of SPIE*, volume 4675, 2002.
4. J. Fridrich and M. Goljan. On estimation of secret message length in LSB steganography in spatial domain. In *SPIE*, 2004.
5. W. Hoeffding. Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.*, 36:369–408, 1965.
6. A. Ker. A general framework for structural steganalysis of LSB replacement. In *IHW 2005*, 2005.
7. A. Ker. Steganalysis of LSB matching in grayscale images. *IEEE Signal Processing Letters*, 12(6), June 2005.
8. N. Provos and P. Honeyman. Hide and seek: An introduction to stegangography. *IEEE Security & Privacy Magazine*, May 2003.
9. M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky. Random cascades on wavelet trees and their use in analyzing and modeling natural images. *Applied Computational and Harmonic Analysis*, 11:89–123, 2001.
10. A. Westfeld. Detecting low embedding rates. In *IHW 2002*, 2002.
11. A. Westfeld and A. Pfitzmann. Attacks on steganographic systems. In *IHW 1999*, 1999.