Steganalysis of QIM-Based Data Hiding using Kernel Density Estimation

Hafiz Malik
ECE Department, Stevens Institute of Technology
Hoboken, NJ 07030
hafiz.malik@stevens.edu

K. P. Subbalakshmi
ECE Department, Stevens Institute of Technology
Hoboken, NJ 07030
ksubbala@stevens.edu

R. Chandramouli
ECE Department, Stevens Institute of Technology
Hoboken, NJ 07030
mouli@stevens.edu

ABSTRACT

This paper presents a novel steganalysis technique to attack quantization index modulation (QIM) steganography. Our method is based on the observation that QIM embedding disturbs neighborhood correlation in the transform domain. We estimate the probability density function (pdf) of this statistical change in a systematic manner using a kernel density estimate (KDE) method. The estimated parametric density model is then used for stego message detection. The impact of the choice of kernels on the estimated density is investigated experimentally. Simulation results evaluated on a large dataset of 6000 quantized images indicate that the proposed method is reliable. The impact of the choice of message embedding parameters on the accuracy of the steganalysis detection is also evaluated. Simulation results show that the proposed method can distinguish between the quantized-cover and the QIM-stego with low false alarm rates (i.e. $P_{fn} \leq 0.03$ and $P_{fp} \leq 0.19$). We demonstrate that the proposed steganalysis scheme can successfully attack steganographic tools like *Jsteg* and *JP Hide and Seek* as well.

Keywords

Steganography, Steganalysis, Quantization Index Modulation, Parameteric Estimation, Kernel Density Estimation, Gamma Density, Skewness, Mode

1. INTRODUCTION

A steganographic information embedding process encodes a message into the cover-object so that the resulting stego-object is perceptually and statistically similar to the cover-object. Rapid proliferation of digital media and the high degree of redundancy in digital representation (despite compression) are some of the motivations for using multimedia data as cover-objects for steganographic applications. There are more than 100 stego softwares available on the Internet ranging from freeware to sophisticated commercial products. Many of the existing stego softwares use least significant bit (LSB) steganography for message embedding. Researchers in the steganographic community have also developed complex and more sophisticated steganographic techniques that are robust to active warden and/or statistical attacks. For example, quantization index modulation (QIM) based data hiding [6] provides flexible trade-off among robustness, capacity, and security of the hidden message. Costa's seminal work [7] provides the theoretical basis of QIM data hiding where the theoretical capacity of the communication with side information over a Gaussian channel was derived. The ideal Costa scheme (ICS) gives a theoretical upper bound on the data hiding capacity under additive white Gaussian noise (AWGN) attack. However, infinite length random codebook requirement makes ICS impractical [9]. A few practical realizations of ICS include, QIM, scalar Costa scheme (SCS), dither modulation (DM)[9], and quantization projection (QP) [13].

Steganalysis refers to the analysis of a given multimedia data (e.g. image, video, audio etc) for the presence of the hidden message with limited or no access to information regarding the embedding algorithm used. Steganalysis techniques may be classified into passive or active depending on whether the aim is to detect the presence or the absence of the hidden message only or to extract the hidden message itself. To date, there appears to have been limited investigation of issues related to steganalysis of QIM steganography. Guillon et al [10] proposed a framework for steganalysis of SCS by modeling QIM steganography as an additive noise channel. Sullivan et al [16] proposed a steganalysis scheme for QIM steganography using supervised learning techniques. Detection performance of machine learning based schemes are limited by several factors. For example, detecting zero-day attack [1], i.e. detecting a stego algorithm not used during the training phase, is not possible. Also, learning based techniques require separate classifier training for each steganographic algorithm. Furthermore, the detection performance depends on the selection of features used to train the classifier and there is no systematic rule for feature selection to achieve desired detection performance [5]. Hence, a steganalyst has limited control on the achievable detector performance.

In this paper we propose a steganalysis technique that does not use a learning based approach. We assume a *stego-only* attack model, that is, the steganalyst have access to the stego image and the message embedding algorithm only. Although we consider the specific example of image steganalysis, even though the proposed method is applicable to other types of data as well. We observe that QIM em-

MM&Sec'07, September 20–21, 2007, Dallas, Texas, USA.

bedding disturbs neighborhood correlation in the transform domain. We estimate the probability density function (pdf) of this statistical change in a systematic manner using a *kernel density estimate* (KDE) method. The estimated parametric density model is then used for stego message detection. Impact of the choice of kernels on the estimated density is investigated. Simulation results evaluated on a large dataset of 6000 quantized images indicates that the proposed method is reliable. Impact of the choice of message embedding parameters on the steganalysis detection accuracy is also evaluated.

The paper is organized as follows: Section 2 highlights irregularities that QIM steganography introduces in the resulting QIM-stego. Details of the randomness mask estimation from the test-image are provided in Section 3.1; a brief overview of kernel density estimation is provided in Section 3.2. Details of nonparametric density estimation from the estimated randomness mask using KDE along with detector design based on the estimated density are provided in Section 3.3. Detection performance of the proposed steganalysis scheme based on the simulation results is discussed in Section 5.1. Future directions and concluding remarks are given in Section 6.

2. QIM STEGANOGRAPHY–*SOME OBSERVATIONS*

A key issue in QIM steganalysis is to the decide if a given test-image is quantized-only ($\mathbf{x}_q$) or quantized with message embedding ($\mathbf{x}_{QIM}$). Some of the experimental observations on the difference between the QIM-stego and quantized-only images are noted below. Firstly, we note that the quantization (with and without message embedding) introduces smoothness in the *pmf* of the resulting quantized image. In order to illustrate this claim, empirical probability mass function(*pmf*) of DCT coefficients of the cover and the corresponding QIM-stego obtained using quantization step-size, $\Delta = \{0.5, 4, 8\}$ are plotted in Fig. 1. It can be observed from Fig. 1 that as Δ increases the empirical *pmf* of the resulting QIM-stego changes from a *super-Gaussian* like *pmf* (e.g. Laplacian *pmf*) to a more Gaussian like *pmf*. Secondly, quantization step-size, Δ, controls the amount of smoothness introduced in the *pmf* of the resulting QIM-stego.

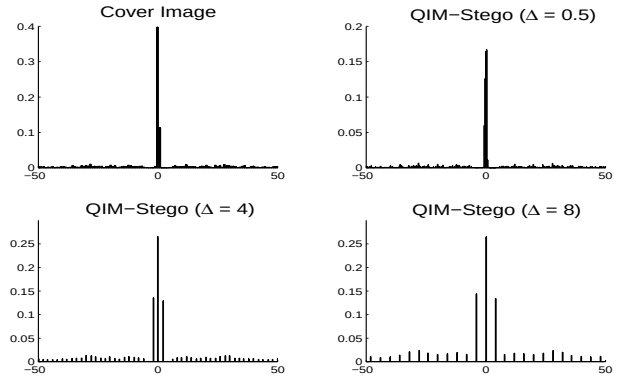

Figure 1: Empirical *pmf* based on histogram of DCT coefficients of the cover (top-left) and quantized DCT coefficients of QIM-stego obtained with $\Delta = \{0.5, 4, 8\}$

Finally, for quantization with message embedding introduces more smoothness than the plain-quantization. To investigate smoothing effect on the cover *pmf* due to quantization further, the empirical *pmf* of the quantized-cover and the QIM-stego are plotted in Fig. 2. It can be observed from Fig. 2 that for same Δ, the QIM introduces more smoothness than the plain-quantization. Moreover, for large Δ, i.e. $\Delta \geq 4$, message embedding using QIM splits the peak of cover *pmf* around zero into three peaks (say $p_{-\Delta}, p_0, p_{\Delta}$ around $-\Delta, 0, \Delta$ respectively), which can be used to distinguish between the quantized-cover and the QIM-stego. However, such visual attack will fail especially when smaller Δ is used for message embedding and/or the cover-image has smooth *pmf*. Learning-based steganalysis techniques have been proposed in the past [16] to distinguish between the quantized-cover and the QIM-stego but as noted earlier, there are some inherent disadvantages with these steganalysis schemes.

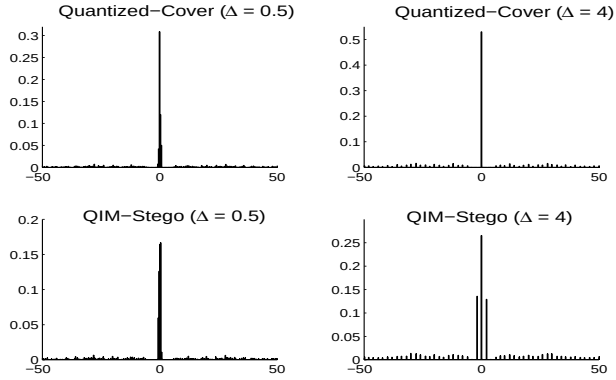

Figure 2: Empirical *pmf* of the quantized-cover (left) and the corresponding QIM-stego (right) both obtained with $\Delta = \{0.5, 4\}$

In order to address limitations of learning-based steganalysis schemes for QIM steganography, a steganalysis scheme based on measure of randomness in the test-image in proposed here. The proposed scheme exploits the fact that message embedding using QIM increases entropy of the resulting stego, that is, the QIM-stego exhibits more randomness than the corresponding quantized-cover, though both quantized images are obtained using same quantization step-size. This fact is illustrated in Fig. 3.

It can be observed from Fig. 3 that the distortion due to QIM embedding is more random than the distortion due to plain-quantization (especially in low-texture regions). It shows that coefficients of the quantized-cover image are more predictable than the corresponding QIM-stego coefficients. The proposed steganalysis scheme exploits this observation to distinguish between the quantized-cover and the QIM-stego. In order to capture irregularities introduced due to message embedding using QIM, a randomness mask based on local similarity is estimated from the test-image. Statistics of the estimated randomness mask is used to distinguish between the cover and the stego.

3. QIM STEGANALYSIS USING KDE

Every steganographic technique introduces statistical and/or perceptual irregularities in the resulting stego-image. In

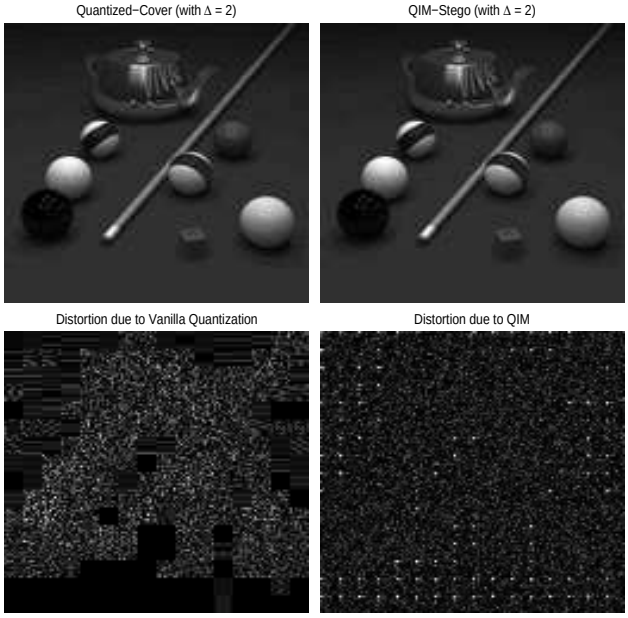

Figure 3: Quantization noise: quantized-cover (left), QIM-stego (right) and associated quantization noise

QIM data hiding, these irregularities manifest themselves as spatial randomness in the quantized DCT coefficients. The randomness mask, $Rc_{\mathbf{x}}$, based on the local similarity is estimated from the test-image (details of $Rc_{\mathbf{x}}$ estimation are provided in Section 3.1). The probability density function (pdf) of the randomness mask is then estimated using a kernel density estimation technique. Simplicity, computational efficiency, and direct dependence of the estimated density on the dataset are the salient features of KDE. First- and higher-order statistics of the estimated density, $\hat{f}_{\mathbf{x}}(x)$, are used to distinguish between the quantized-cover and the QIM-stego. Block diagram of the proposed steganalysis scheme to attack QIM steganography is given in Fig. 4. Details of each processing stage involved in the proposed

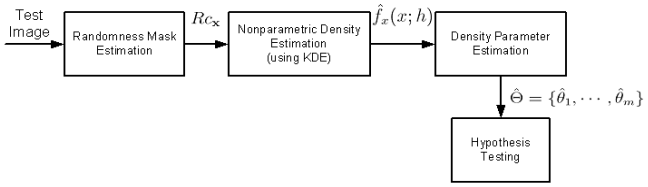

Figure 4: Block diagram of the proposed steganalysis scheme

steganalysis scheme are discussed in the following sections.

3.1 Randomness Mask Estimation

In order to estimate the randomness mask, $Rc_{\mathbf{x}}$, the test-image is segmented into non-overlapping blocks, each of 8x8 pixels. Each block is transformed into DCT domain using the following 2D forward discrete cosine function,

$$x_{k_1,k_2} = \frac{1}{4} G_{k_1} G_{k_2} \sum_{n_1=0}^{7} \sum_{n_2=0}^{7} s_{n_1,n_2} \cos\left(\frac{\pi k_1 (2n_1+1)}{16}\right) \cos\left(\frac{\pi k_2 (2n_2+1)}{16}\right),$$
$$k_1, k_2 = 0, \cdots, 7$$

where

$$G_{k_1}, G_{k_2} = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } k_1 = 0, k_2 = 0 \\ 1 & \text{otherwise} \end{cases}$$

Randomness mask value of coefficient x_{k_1,k_2} in the j^{th} block (or $x^{(j)}_{k_1,k_2}$), $Rc^{(j)}_{k_1,k_2}$, is calculated based on similarity of $x^{(j)}_{k_1,k_2}$ with corresponding coefficients in k-neighboring blocks. Let $x^{(j)}_{NH(k_1,k_2,i)}, i = 1, \cdots, k$ denote the coefficients in k-neighboring blocks. Then the similarity value , $C^{(j)}_{k_1,k_2}$, for coefficient $x^{(j)}_{k_1,k_2}$ is calculated as,

$$C^{(j)}_{k_1,k_2} = \frac{1}{k} \sum_{l=1}^{k} \mathbf{1}_{[x^{(j)}_{k_1,k_2}]} \left(x^{(j)}_{NH(k_1,k_2,l)}\right) \quad (1)$$
$$l = 1, \cdots, k, \quad \text{and} \quad j = 1, \cdots, n$$

and the corresponding randomness mask, $Rc^{(j)}_{k_1,k_2}$, is calculated as,

$$Rc^{(j)}_{k_1,k_2} = 1 - C^{(j)}_{k_1,k_2} \quad (2)$$

where $\mathbf{1}$ is an indicator function, $n = \lfloor \frac{n_1}{8} \rfloor \times \lfloor \frac{n_2}{8} \rfloor$, and $\lfloor x \rfloor$ denotes the largest integer not exceeding x.

The $Rc^{(j)}_{k_1,k_2}$ is a nonnegative real valued random variable, and $0 \leq Rc^{(j)}_{k_1,k_2} \leq 1$. When all the neighboring coefficients are quantized to the same value, $Rc^{(j)}_{k_1,k_2} = 0$ implying maximum similarity between the neighbors and the current value, $x^{(j)}_{k_1,k_2}$. Similarly, $Rc^{(j)}_{k_1,k_2} = 1$ implies minimum similarity that corresponds to the case when all coefficients (the current coefficient value and its neighbors) are quantized to k distinct values. To illustrate the notion of randomness mask estimation based on k-neighborhood using Eq. (2); the randomness mask estimation for the selected block (or block of interest (BOI)) using 4-neighborhood is given in Fig. 5.

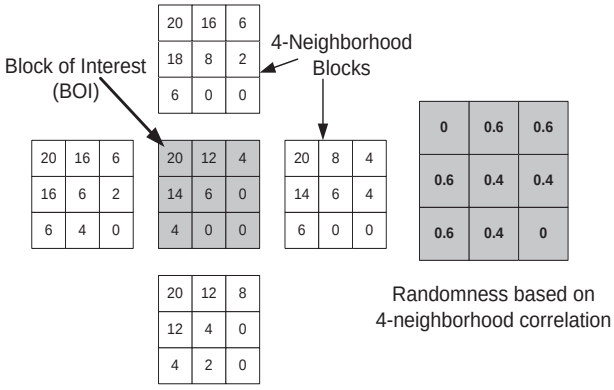

Figure 5: Randomness mask estimation for the selected block based on 4-neighborhood

It is important to note that the estimated randomness mask, $Rc_{\mathbf{x}}$, depends on test-image characteristics, quantization step-size, Δ, used to generate the quantized image, and

the hidden message distribution. According to Eq. (2), a texture rich image would yield lower similarity values than a low-texture image, both quantized using same Δ. Similarly, quantized images generated using smaller Δ will yield randomness mask with higher mean than the quantized-image generated using larger Δ. For a given image, large quantization step-size tends to map neighboring coefficients to fewer distinct quantized value compared to a smaller Δ value. Therefore, it is reasonable to expect that $Rc_{\mathbf{x}}$ estimated from the QIM-stego would have a higher mean value than the randomness mask estimated from the corresponding quantized-cover, both obtained using same Δ.

The two-dimensional $Rc_{\mathbf{x}}$ is then transformed into 64 sequences which are used to estimate the underlying density, $f_{\mathbf{x}}(x)$. The mapping of two-dimensional randomness mask, $Rc_{\mathbf{x}}$, to one-dimensional sequences, $\mathbf{x}_n^{(j)}, i = 0, 1, \cdots, 63$, is illustrated in Fig. 6.

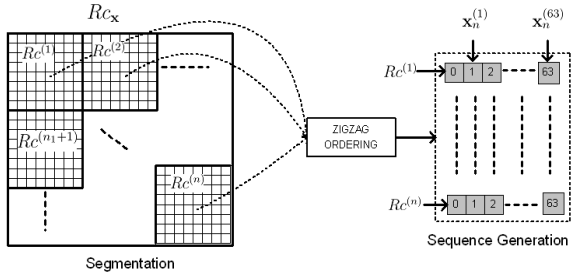

Figure 6: Mapping of two-dimensional randomness mask, $Rc_{k_1,k_2}^{(j)}$, to one-dimensional sequences, $\mathbf{x}_n^{(j)}$

3.2 Kernel Density Estimation

A histogram is the simplest and the most frequently used density estimator. However, it yields non-smooth density estimate depending on the boundaries and width of bins (or bandwidth). Kernel density estimators alleviate some of these problems. In order to remove dependence of the estimated density on the end points of the bins, kernel estimators center a kernel function, $K(x)$, at each data point, x_i. Smooth kernel functions are generally used to obtain a smooth density estimate.

More formally, kernel estimators smooth out the contribution of each observed data point over a local neighborhood of that data point. The contribution of data point x_i to the estimate at an arbitrary point x depends on how apart x_i and x are. The extent of this contribution depends upon the shape of the kernel function, $K(x)$, used and its bandwidth. Let h and x_i denote the bandwidth (or variance) and mean of the kernel $K(x)$ respectively, then the estimated density at any point x can be expressed as,

$$\hat{f}(x;h) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{3}$$

where $\int K(t)dt = 1$ to ensure that $\hat{f}(x)$ integrates to 1.

The kernel function, $K(x)$, is usually chosen to be a smooth unimodal function with a peak at 0 [11, 12]. Even though Gaussian kernel is the most commonly used kernel function for KDE, there are various choices among kernels as listed in Table 1.

Table 1: List of commonly used kernels

Kernel	**K(u)**
Uniform	$\frac{1}{2}I(\mid u \mid \leq 1)$
Triangle	$(1- \mid u \mid) \; I(\mid u \mid \leq 1)$
Epanechnikov	$\frac{3}{4}(1-u^2)I(\mid u \mid \leq 1)$
Quartic	$\frac{15}{16}(1-u^2)^2 I(\mid u \mid \leq 1)$
Triweight	$\frac{35}{32}(1-u^2)^3 I(\mid u \mid \leq 1)$
Gaussian	$\frac{1}{\sqrt{2\pi}} exp(-\frac{1}{2}u^2)$
Laplace	$\frac{1}{2} exp(-\mid u \mid)$
Logistic	$\frac{1}{1+e^{-u}}$

The quality of a kernel estimate depends less on the shape of the kernel $K(x)$ than on the value of its bandwidth h. A suitable bandwidth selection to obtain optimally smooth density estimate from a given dataset is critical. For example, small values of h lead to very spiky estimates (or under-smooth estimates) while larger h values lead to over-smoothing. The mean integrated squared error (MISE) metric is commonly used to choose the optimal bandwidth that minimizes the MISE, that is,

$$h_{opt} = \underset{h}{arg\min} \int_{\mathcal{R}^2} E\{(\hat{f}(x;h) - f(x))^2\}dx \tag{4}$$

where $f(x)$ is the target density.

Note that the MISE is a measure of the average performance of the kernel density estimator as MISE is independent of the actual dataset [11, 12]. One drawback of finding optimal bandwidth h_{opt} using Eq. (4) is that it does not have a closed form solution for an arbitrary target density $f(x)$. Asymptotic approximation of Eq. (4) via a Taylor's series expansion is generally used to find h_{opt} [11, 12, 8, 17, 15]. For more details on optimal bandwidth selection see [11, 12, 15] and references therein.

3.3 Randomness Mask Density Estimation

We use KDE to estimate the probability density of the 64 sequences obtained from the estimated randomness mask, $Rc_{\mathbf{x}}$. Sequence corresponding to DC coefficients of the test-image, $\mathbf{x}_n^{(0)}$, is not used for steganalysis, as DC coefficients of the cover-image are not modified during QIM-based embedding to avoid blocking artifacts in the resulting stego-image.

The remaining 63 sequences are used to estimate the underlying densities using KDE. The KDE package downloaded from [4] supports all the kernels listed in Table 1. We use the *Gaussian kennel*, $K_g(x) = \frac{1}{\sqrt{2\pi}} exp(-\frac{1}{2}u^2)$. Selection of the kernel function, $K_g(x)$, for density estimation, is motivated by the fact that for a given bandwidth, h, the Gaussian kernel yields relatively smoother density estimate, $\hat{f}_x(x)$, than commonly known kernels, such as, logistic, Laplacian, and Epanechnikov (see Fig. 8). Fig. 8 plots the estimated density from sequence $\mathbf{x}_n^{(30)}$ estimated from the quantized *Girl* image (see Fig. 7), using Gaussian, logistic, Laplacial, and Epanechnikov kernels each with bandwidth $h = 0.1$.

Figure 7: Girl image

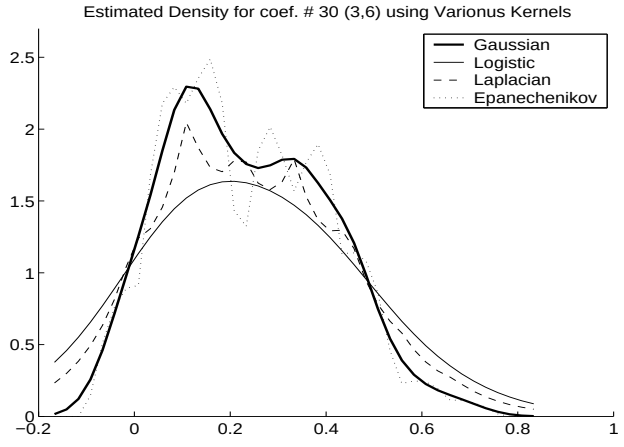

Figure 8: Estimated density from sequence $\mathbf{x}_n^{(30)}$ of $Rc_{\mathbf{x}}$ of quantized *Girl Image* using Gaussian, logistic, Laplacial, and Epanechnikov kernels with $h = 0.1$

Fig. 8 shows that the Gaussian kernel yields relatively smoother density estimate than the other kernels (i.e. logistic, Laplacial, and Epanechnikov). It is important to mention that the estimated density for $Rc_{\mathbf{x}}$, $\hat{f}_{Rc_{\mathbf{x}}}(x;h)$, plotted in Fig. 8 has nonzero value for $x < 0$ though $0 \geq x \leq 1$. Nonzero estimated density values outside [0, 1] can be attributed to the kernel, $K(x)$, and the bandwidth, h, used for density estimation. As kernels used for density estimation (e.g., Gaussian, logistic, Laplacial, and Epanechnikov) plotted in Fig. 8 are continuous and nonzero for $(-\infty, \infty)$ and spread of these kernels is controlled by the bandwidth h. Therefore, though the underlying data is strictly bounded by range [-R, R] but the estimated density using such kernels might have nonzero value $R < x < -R$. The amount of leakage depends on the number of data points around boundaries of the data range, e.g. around 0 and 1 in case of $Rc_{\mathbf{x}}$, shape of the kernel used, and the bandwidth. In order to obtained density estimate which is bounded by the underlying data range one need to use smaller bandwidth h and sharply decaying kernel, e.g. uniform kernel. However, this leakage of the estimated density out side [0, 1] does not contribute to the detection performance of the proposed steganalysis scheme. Therefore, for the rest of the paper we shall assume that the Gaussian kernel is used for density estimation with bandwidth $h = 0.1$ unless otherwise specified.

The estimated density can be used to determine suitable density model $Rc_{\mathbf{x}}$. It has been observed from the estimated densities using 6000 images (3000 quantized-covers and 3000 QIM-stego) that the underlying density can be approximated using a *generalized Gamma distribution* (GGD), that is,

$$f_x(x;\alpha,\beta,\gamma) = x^{\alpha\gamma-1}\frac{\gamma}{\beta^{\alpha\gamma}\Gamma(\alpha)}e^{-(x/\beta)^{\gamma}} \text{ for } \mathbf{x} > 0 \qquad (5)$$

where α, β, and γ are positive real valued parameters of the generalized Gamma distribution. The generalized Gamma distribution includes a wide range of distributions, the Weibull distribution ($\alpha = 1$), the exponential distribution ($\alpha = \gamma = 1$), Gaussian distribution ($\alpha = \frac{1}{2}, \gamma = 2$), Gamma distribution ($\gamma = 1$), etc.

To determine a specific underlying density model, we analyzed natural images with different levels of texture. In this paper, we used uncompressed color image database (UCID) downloaded from [3] for performance evaluation of the proposed steganalysis scheme. Experimental results show that the estimated density from low-texture images is close to the exponential density, whereas density from high-texture images is close to the Gaussian density with peak around 0.5. In addition, images consisting of mixture of uniform and high-texture regions yielded bimodal density estimate. For such images a generalized Gamma distribution is observed to be a good approximation. However, for most of the images in the UCID database [3], the estimated density using KDE was close to Gamma density. Therefore, the randomness mask, $Rc_{\mathbf{x}}$, is modeled by the Gamma density, given by,

$$f_{Rc_x}(x;\alpha,\beta) = x^{\alpha-1}\frac{1}{\beta^{\alpha},\Gamma(\alpha)}e^{-(x/\beta)} \text{ for } \mathbf{x} > 0 \qquad (6)$$

In order to verify the goodness of fit (i.e. closeness of the estimated density to the Gamma distribution), the density estimated from $Rc_{\mathbf{x}}$, $(\hat{f}_{Rc_x(x)})$, is compared with the Gamma distribution function using parameters, $\hat{\alpha}$ and $\hat{\beta}$, computed as likelihood estimates (MLE). Fig. 9 shows the plots of the KDE density estimate of $Rc_{\mathbf{x}}$ estimated from the QIM-stego of the *Girl* image (see Fig. 7) and the Gamma density function corresponding to $\alpha = 4.58$ and $\beta = 0.078$, estimated by MLE. Fig. 9 shows that both densities have approximately same mode (peak) and skewness. Therefore it is reasonable to assume that randomness in the quantized DCT coefficients of images with moderate texture can be approximated by the Gamma distribution.

The estimated density is used to distinguish between the quantized-cover and the QIM-stego images. As discussed

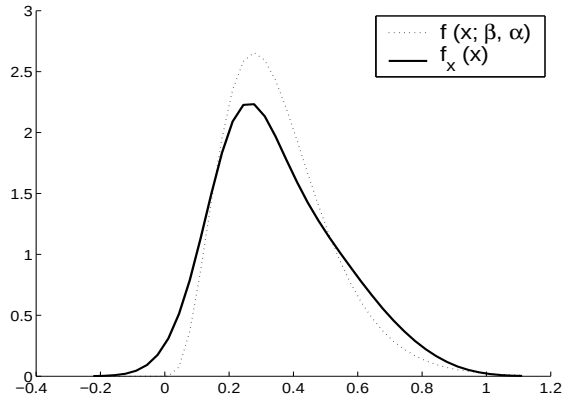

Figure 9: Estimated density using KDE (solid line) and Gamma density function based on estimated parameters $\hat{\alpha} = 4.58$ and $\hat{\beta} = 0.078$ (doted line)

earlier, message embedding using QIM increases randomness in the resulting stego. Therefore, the estimated density from $Rc_{\mathbf{x}_{QIM}}$ (estimated $Rc_{\mathbf{x}}$ from the QIM-stego) is expected to have higher mean value than the corresponding quantized-cover. Statistics of the estimated density can be used to distinguish between the quantized-cover and the QIM-stego. To illustrate this, we generated two quantized images using same Δ, one using plain-quantization and an other using QIM embedding. The randomness mask was estimated from these quantized images. The underlying density was then estimated from the randomness masks using KDE. The estimated densities from the randomness masks of the quantized-cover and the corresponding QIM-stego using the same set of density estimation parameters (i.e. kernel, $K(x)$, and bandwidth, h) are shown in Fig. 10.

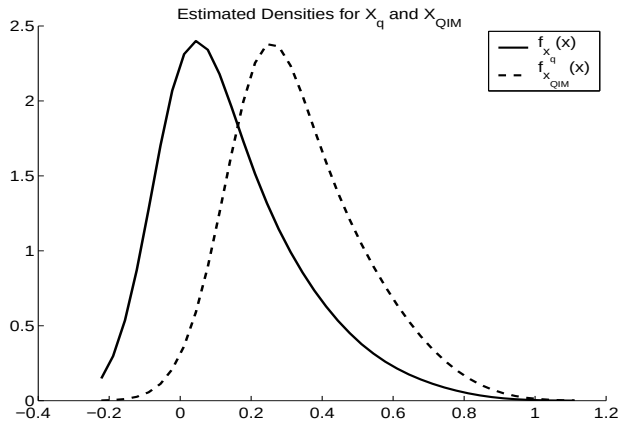

Figure 10: Estimated density plots obtained from $Rc_{\mathbf{x}}$ corresponding to the quantized-cover (*Girl* image) and the QIM-stego (both obtained using $\Delta = 4$)

Fig. 10 shows that the estimated density for the quantized-cover image has a peak (or mode) at zero which implies that the DCT coefficients of the quantized-cover are highly correlated. This observation also agrees with the fact that there exist high local spatial correlation in *Girl* image. Also, the estimated density from $\mathbf{x}_q$ exhibits higher skewness than the density estimated from $\mathbf{x}_{QIM}$. The estimated density from QIM-stego exhibits a peak near 0.3 which indicates that the stego coefficients have relatively higher level of randomness compare to the quantized-cover. This increase in the randomness in the QIM-stego can be attributed to the randomness in the hidden message M. Moreover, the density estimated from $\mathbf{x}_{QIM}$ exhibits lower skewness.

We propose a nonparametric hypothesis test to distinguish between the quantized-cover and the QIM-stego based on the estimated density function. Nonparametric tests are known to be robust to uncertainties in the assumptions about underlying probability distributions. To this end, the proposed scheme uses *skewness* and the *mode* of the estimated density to distinguish between the quantized cover and the QIM-stego. The underlying density parameters, that is, α and β are estimated by fitting $Rc_{\mathbf{x}}$ to Gamma density using MLE. The skewness, Skew, of the underlying density is calculated using $\hat{\alpha}$ and $\hat{\beta}$ estimated using the MLE as,

$$\text{Skew} = \frac{2}{\sqrt{\hat{\alpha}}} \tag{7}$$

and mode, Mode, of the underlying Gamma density is calculated as,

$$\text{Mode} = (\hat{\alpha} - 1)\hat{\beta} \tag{8}$$

Estimated statistics of the underlying density is then used to distinguish between the quantized-cover and the QIM-stego. Following binary hypothesis test is used to distinguish between the $\mathbf{x}_q$ and $\mathbf{x}_{QIM}$,

$$\begin{aligned} &\textbf{Decide} \quad \text{Stego} \quad \text{if } \tau_1 \geq \text{Mode} \leq \tau_2 \; OR \;\; \text{Skew} \leq \tau_3 \\ &\textbf{Decide} \quad \text{Cover} \qquad\qquad\qquad Otherwise \end{aligned} \tag{9}$$

where τ_1,τ_1, and τ_2 are positive real valued constants.

Here decision thresholds τ_1 and τ_2 determine the detection performance of the detector. It has been observed based on simulation results for 6000 quantized images that decision thresholds $\tau_1 = 0.20$, $\tau_2 = 0.45$, and $\tau_3 = 1.1$ yield low false alarm rates, i.e., $P_{fp} < .19$ and $P_{fn} < 0.03$ for $\Delta = 4$. Simulation results presented in this paper are compiled using decision threshold values $\tau_1 = 0.20$, $\tau_2 = 0.45$, and $\tau_3 = 1.1$.

4. SIMULATION RESULTS

The performance of the proposed steganalysis scheme is evaluated over UICD database [3]. The UCID database contains around 1400 natural images of almost all textures. Simulation results presented here are based on first 1000 of the UICD database. These 1000 images were resize to 256x256 pixels each and the transformed to gray scale for message embedding and steganalysis. Six thousand quantized images (3000 quantized-cover and 3000 QIM-stego) were generated by quantizing 1000 natural images using uniform quantizer with quantization step-size $\Delta = \{0.5, 2.0, 4.0\}$. Each QIM-stego image was generated by embedding 40KB random binary message with equally probable message symbols. These six thousand quantized images were tested using the proposed steganalysis scheme. Each test-image was processed to generate the randomness mask, $Rc_{\mathbf{x}}$. The underlying density was estimated using KDE. Mode and skewness of the estimated $Rc_{\mathbf{x}}$ were used to distinguish between the quantized-cover and the QIM-stego. In order to illustrate the effect of QIM embedding, scatter plots of the estimated

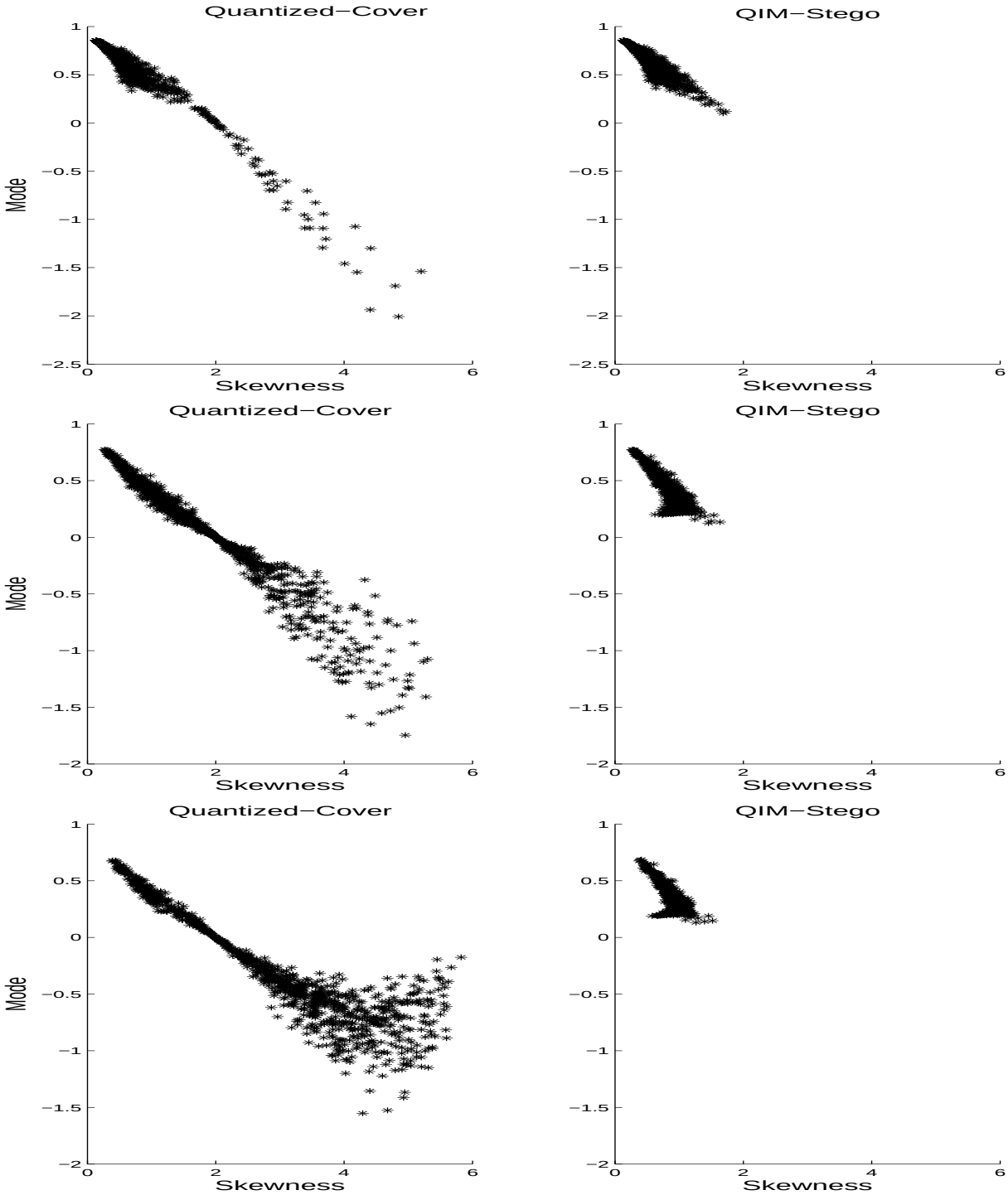

Figure 11: Scatter plots of mode verses skewness of the estimated randomness mask from the quantized-cover (left column) and the QIM-stego (right column) for $\Delta = \{0.5, 2, 4\}$ (from top to bottom)

mode and skewness of the underlying density for the quantized and the corresponding stego for $\Delta = \{0.5, 2, 4\}$ (from top to bottom) are shown in Fig. 11.

We observe the following from Fig. 11:

- Mode of Rc_{x} estimated from stego images lies between 0.1 and 0.7, i.e. $\mathrm{Mode}_{\mathrm{QIM}} \in (0.1, 0.7)$, range of estimated mode shrinks as Δ increases, and for $\Delta \geq 4$ $\mathrm{Mode}_{\mathrm{QIM}} \in (0.2, 0.5)$.
- Skewness of Rc_{x} estimated from stego-images lies between 0 and 1.2, i.e $\mathrm{Skew}_{\mathrm{QIM}} \in (0, 1.2)$ and as Δ increases the range of estimated skewness reduces slightly.
- Mode of Rc_{x} estimated from quantized-cover images falls between -1.5 and 0.7, $\mathrm{Mode}_{\mathrm{q}} \in (-1.5, 0.7)$ and as Δ increases, the estimated mode start shifting towards -1.
- Skewness of the estimated density from quantized-cover images falls between 0 and 6, $\mathrm{Skew}_{\mathrm{q}} \in (0, 6)$ and as Δ increases, the estimated skewness start shifting towards 6.
- With high probability, estimated $\mathrm{Mode}_{\mathrm{QIM}} \in (0.2, 0.5)$ and $\mathrm{Skew}_{\mathrm{QIM}} \in (0, 1.2)$ from stego-image and this probability approaches 1 for $\Delta \geq 4$
- Probability of estimated $\mathrm{Mode}_{\mathrm{q}} \in (0.2, 0.5)$ and $\mathrm{Skew}_{\mathrm{q}} \in (0, 1.2)$ from the quantized-cover is close to 1 for smaller Δ and start decreasing as Δ increases.

Therefore false rates, i.e., P_{fn}, and P_{fn}, decreases as Δ increases. Fig. 11 also shows that for smaller Δ the proposed scheme yields high false positive rate, as $\Pr\{\mathrm{Mode}_{\mathrm{q}} \in (0.2, 0.5) \bigcap \mathrm{Skew}_{\mathrm{q}} \in (0, 1.2)\} \approx 0.9$. However, two clusters start to emerge as Δ increases and therefore the false alarm decreases. Detection performance of the proposed steganalysis scheme in terms false rate, i.e. P_{fp} and P_{fn}, for decision thresholds $\tau_1 = 0.2$, $\tau_1 = 0.45$ and $\tau_3 = 1.1$ are listed in Table 2.

Table 2: Detection performance as function of quantization step-size

	Quantization Step-Size Δ		
	$\Delta = 0.5$	$\Delta = 2$	$\Delta = 4$
P_{fp}	0.91	0.46	0.19
P_{fn}	4.5×10^{-2}	3.4×10^{-2}	3.0×10^{-2}

Simulation results presented in Table 2 shows that the proposed steganalysis scheme can detect QIM-stego, with low false rates, when the hidden message is embedded using $\Delta \geq 4$. Moreover, to ensure security of the hidden message, smaller Δ, i.e. $\Delta \leq 0.5$ should be used. But message embedded using smaller Δ cannot combat *active warden attack*. An active warden deliberately alters every object (stego or cover) that Alice and Bob are exchanging, to foil covert communication between them. In order to combat active warden attack, message should be embedded using larger Δ. Therefore, at a given capacity, security of the hidden message using QIM is achieved at the cost of robustness and vice versa. On the other hand, if both stronger security (against steganalyst) and robustness (against active warden attack) of the hidden message are desirable then embedding capacity is compromised to achieve these goals.

5. ATTACKING JSTEG

The proposed steganalysis scheme was also used to attack Jsteg [2], a freeware that hides messages in baseline JPEG compressed images. Jsteg embeds messages in the GIF image by replacing the LSB of the quantized run-length coded DCT coefficients with the secret message, during JPEG compression process. To attack Jsteg, quantized AC coefficients (after run-length coding) of the JPEG test-image are used to estimate the randomness mask, $Rc_{\mathbf{x}}$. Since for most JPEG compressed images only the low- and mid-frequency coefficients survive for the entropy coding stage, the randomness mask is estimated from these AC coefficients. The underlying density of the estimated randomness mask is used to distinguish between the JPEG-cover and the JEPG-stego.

To illustrate effect of message embedding using Jsteg on the randomness in the JPEG stego image, a JPEG stego was generated using Jsteg steganographic tool (available at [2]) by embedding 7KB message in the 512x212 gray scale ***Lenna*** image using quality factor $q = 100$. Estimated densities using proposed steganalysis scheme from the JPEG stego (stego obtained using Jsteg) and the corresponding JPEG cover are shown in Fig. 12.

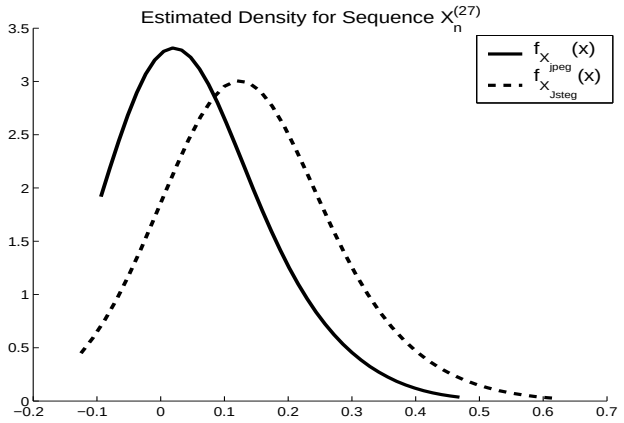

Figure 12: Estimated density plots from the stego image obtained using Jsteg tool and the corresponding JPEG cover image

It can be observed from Fig. 12 that the estimate density from the JPEG cover exhibits peak around zero, whereas density estimated from the JPEG stego has its peak between 0.2 and 0.5. In addition, estimated densities from these JPEG images also have different skewness values. Therefore, first- and third-moment of the estimated density can be used to attack Jsteg [2]. However, we have also observed through extensive simulations that some times the estimated densities from the JPEG cover and the corresponding JPEG stego, i.e., $\hat{f}_{Rc_{\mathbf{x}}}(x)$ and $\hat{f}_{Rc_{JSTEG}}(x)$, are hard to differentiate, especially when the JPEG stego image is carrying small hidden message. As size of the hidden message directly depends on,

- texture of the cover image
- quality factor, Q, used for compression, and
- the cover image size.

We have observed that low-texture images of size less than 128x128 pixels when compressed using quality factor,

$Q \leq 75$ can carry very small message size, e.g. message size less than $4KB$. As a results first- and third-moment of the estimated density from such JPEG stego image are approximately same to first- and third-moment of the estimated density from the corresponding JPEG cover image. Therefore, the proposed steganalysis fails to detect such JPEG stego images.

This fact is illustrated in Fig. 14 where estimated density from Fruits JPEG image of 128x128 pixels (see Fig. 13), obtained using $Q = 50$, and the corresponding JPEG stego image obtained using Jsteg tool [2]. It can be observed from Fig. 14 that it is hard to distinguish between the JPEG cover and the JPEG stego based on the mode and skewness of the underlying density.

Figure 13: Fruits image

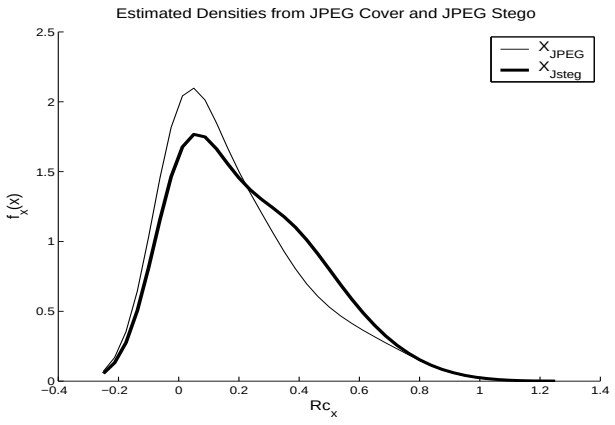

Figure 14: Estimated density plots from the JPEG cover image obtained with $Q = 50$ and the corresponding JPEG stego obtained by embedding $3K$B message using Jsteg tool with $Q = 50$

To handel such cases, that is, detect JPEG stego images, carrying small hidden messages, obtained using Jsteg with $50 \geq Q \leq 75$, the proposed steganalysis scheme in the previous section 5.1 is modified slightly. To this end, the JPEG test-image is recompressed using Jsteg to generate the corresponding JPEG$^{(2)}$ stego image, $\mathbf{x}_{JSTEG}$. The recompressed JPEG$^{(2)}$ stego image is generated by embedding an arbitrary message, $\acute{M}$, using Jsteg with quality factor Q. The recompression stage involves, decompression of the JPEG test-image followed by JPEG compression using Jsteg. An arbitrary message, $\acute{M}$, is embedded during JPEG compression using Jsteg with quality factor Q. Randomness mask is then estimated from the test-image and JPEG$^{(2)}$ stego image. Estimated randomness masks from these JPEG images, e.g. $Rc_{\mathbf{x}}$ and Rc_{JSTEG}, is used to estimate the underlying densities using KDE, e.g., $\hat{f}_{Rc_{\mathbf{x}}}(x)$ and $\hat{f}_{Rc_{JSTEG}}(x)$. The Kullback-Leblier (KL) distance between the estimated densities, e.g., $\hat{f}_{Rc_{\mathbf{x}}}(x)$ and $\hat{f}_{Rc_{JSTEG}}(x)$, is used to distinguish between the JPEG cover and the JPEG stego image. The KL distance between probability mass functions, $f_{x1}(x)$ and $f_{x2}(x)$ is defined as,

$$D\left(\hat{f}_{x1}(x) \parallel \hat{f}_{x2}(x)\right) = \sum_{x \in \S} f_{x1}(x) \log_2 \left(\frac{f_{x1}(x)}{f_{x2}(x)}\right) \quad (10)$$

$$= E_{f_{x1}}\left(\log_2\left(\frac{f_{x1}(x)}{f_{x2}(x)}\right)\right) \quad (11)$$

where $E_{f_{x1}}$ denote expectation over $f_{x1}(x)$.

It is expected that if the test-image is a JPEG stego-image obtained using Jsteg, than an arbitrary message embedding using Jsteg again would cause a small change in the level of randomness in the resulting $\mathbf{x}_{JSTEG}$ hence small variation in the the underlying density estimated from x_{JSTEG}. Therefore, the KL distance between the estimated densities would be relatively smaller than the KL distance between the estimated density from the JPEG cover image and the density estimated from the corresponding JPEG stego image. This notion is illustrated in 15.

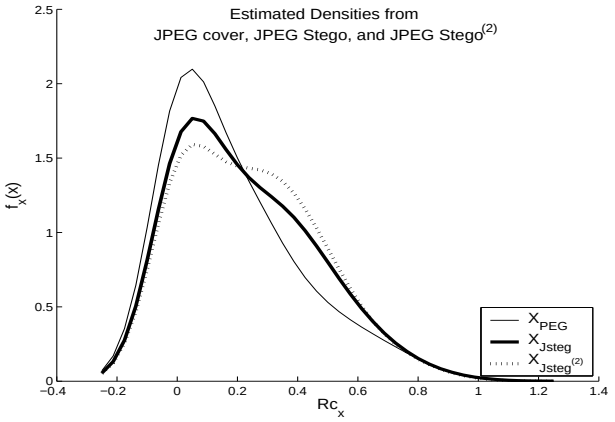

Figure 15: Estimated density plots from the JPEG cover (obtained with $Q = 50$). JPEG stego and JPEG stego$^{(2)}$ both obtained by embedding $3K$B message using Jsteg tool with $Q = 50$

It can be observed from Fig. 15 that the densities estimated from the JPEG stego and JPEG stego$^{(2)}$ (stego image obtained by embedding an arbitrary message in the JPEG stego using Jsteg with $Q = 50$), $\hat{f}_{Rc_{JSTEG}}(x)$ and $\hat{f}_{Rc_{JSTEG^{(2)}}}(x)$, have relatively smaller KL distance (e.g. 0.17 bits/sample) than the KL distance between the densities estimated from the JPEG cover and the JPEG stego, $D\left(\hat{f}_{Rc_{\mathbf{x}}}(x) \parallel \hat{f}_{Rc_{JSTEG}}(x)\right)$ which is 0.85 bits/sample. The KL distance between the estimated density based on the randomness mask from the JPEG test-image and the corresponding JPEG$^{(2)}$ stego image, obtained by embedding

an arbitrary message in the JPEG test-image using Jsteg, can be used to distinguish between the JPEG cover and the JPEG stego image.

Block diagram of the steganalysis scheme use to distinguish between the JPEG cover and the JPEG stego image is given in Fig. 16.

5.1 Experimental Results

Detection performance of the proposed steganalysis scheme to attach Jsteg steganographic tool [2] is evaluated for a datased of 3000 JPEG images. First 500 images of the UCID image database [3] were used to generate these 3000 JPEG images. The dataset used for performance evaluation of the proposed steganalysis scheme consists of 50% JPEG cover images and rest 50% JPEG stego. The JPEG cover images were generated by compressing 500 uncompressed images using baseline JPEG [14] with quality factor $Q = \{100, 75, 50\}$. Whereas, JPEG stego images were generated by compressing 500 uncompressed images using Jsteg tool with $Q = \{100, 75, 50\}$. A random message was into each JPEG stego image.

During testing phase, each test-image (JPEG cover or JPEG stego) was recompressed to generate corresponding JPEG$^{(2)}$ stego image using Jsteg tool. The JPEG$^{(2)}$ stego image, $\mathbf{X}_{JSTEG}$, was generated by first decompressing the JPEG test-image followed by recompression using Jsteg with associated quality factor Q. An arbitrary message M was embedded into each test-image to generate $\mathbf{X}_{JSTEG}$. Randomness mask was then estimated from both the test-image and the corresponding $\mathbf{X}_{JSTEG}$ using run-length coded AC coefficients. Estimated randomness masks from the test-image and the JPEG$^{(2)}$ stego image, e.g. $Rc_{\mathbf{x}}$ and Rc_{JSTEG}, were used to estimate the underlying densities using KDE. The KL distance between $\hat{f}_{Rc_{\mathbf{x}}}(x)$ (estimated from the JPEG test-image) and $\hat{f}_{Rc_{JSTEG}}(x)$ (estimated from the the JPEG$^{(2)}$ stego image), that is, $D\left(\hat{f}_{Rc_{\mathbf{x}}}(x) \parallel \hat{f}_{Rc_{JSTEG}}(x)\right)$, was compared against decision threshold, τ_4, to distinguish between the JPEG cover and the JPEG stego.

Detection performance of the proposed steganalysis scheme with these experimental settings and decision threshold $\tau_4 = 0.5$ is given in Table 3.

Table 3: Detection performance as function of quantization step-size

	Quality Factor Q		
	$Q = 50$	$Q = 75$	$Q = 100$
P_{fp}	0.192	0.192	0.182
P_{fn}	3.6×10^{-2}	3.6×10^{-2}	2.2×10^{-2}

6. CONCLUSION

This paper presents steganalysis scheme for QIM-based data hiding. The proposed steganalysis scheme is non-learning based therefore can address limitations of learning-based steganalysis schemes. We have shown that QIM based steganography increases randomness in the resulting QIM-stego. The proposed steganalysis scheme therefore uses measure of randomness in the DCT coefficients of the test-image to distinguish between the quantized-cover and the QIM-stego. Simulation results show that the proposed steganalysis scheme can detect the QIM-stego with low false negative rate. In addition, detection performance of the proposed scheme depends on quantization-step size used to generate the quantized test-image.

The proposed steganalysis scheme is also extended to attack Jsteg steganographic tool [2]. Experimental results to evaluated performance of the proposed steganalysis scheme shows that it can successfully distinguish between the JPEG cover and the JPEG stego obtained with quality factor as low as 50 with low false rates.

7. REFERENCES

[1] http://en.wikipedia.org/wiki/Zero-Day_Attack.

[2] J. korejwa: Jsteg. available at ftp://ftp.funet.fi/pub/crypt/steganography/.

[3] Ucid: An uncompressed colour image database. available at http://www-users.aston.ac.uk/ schaefeg/datasets/UCID/ucid.html.

[4] Wafo: Wave analysis for fatigue and oceanography. available at http://www.maths.lth.se/matstat/wafo/.

[5] R. Chandramouli and K. Subbalakshmi. Current trends in steganalysis: A critical survey. In *IEEE Int. Conf. on Control, Automation, Robotics and Vision, ICARCV*, volume 2, pages 964–967, December 2004.

[6] B. Chen and G. Wornell. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Trans. Information Theory*, 47(4), May 2001.

[7] M. Costa. Writing on dirty paper. *IEEE Transactions on Information Theory*, 29(3):439–441, May 1983.

[8] T. Duong. *Plug-in Bandwidth Selectors for Bivariate Kernel Density Estimation*. PhD thesis, Department of Mathematics and Statistics, the University of Western Australia, 2002.

[9] J. Eggers and B. Girod. *Informed Watermarking*. Kluwer Academic Publisher, 2002.

[10] P. Guillon, T. Furon, and P. Duhamel. Applied public-key steganography. In *Proc. IS&T/SPIE*, pages 38–49, 2002.

[11] P. V. Kerm. Adaptive kernel density estimation. *Stata Journal*, 3(2), June 2003.

[12] P. V. Kerm. Adaptive kernel density estimation. Technical report, 2003.

[13] F. Perez-Gonzalez, F. Balado, and J. R. Hernandez. Performance analysis of existing and new methods for data hiding with known-host information in additive channels. *IEEE Transaction on Signal Processing*, 51(4), April 2003.

[14] K. Sayood. *Introduction to Data Compression*. Morgan Kaufmann, 2^{nd} edition, 2000.

[15] D. W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons,, 1992.

[16] K. Sullivan, Z. Bi, U. Madhow, S. Chandrasekaran, and B. Manjunath. Steganalysis of quantization index modulation data hiding. In *IEEE Int. Conf. Image Processing (ICIP)*, volume 2, pages 1165–1168, 2004.

[17] X. Zhang, M. L. King, and R. J. Hyndman. Bandwidth selection for multivariate kernel density estimation using mcmc. Monash Econometrics and Business Statistics Working Papers 9/04, Monash University, Department of Econometrics and Business Statistics, 2004.

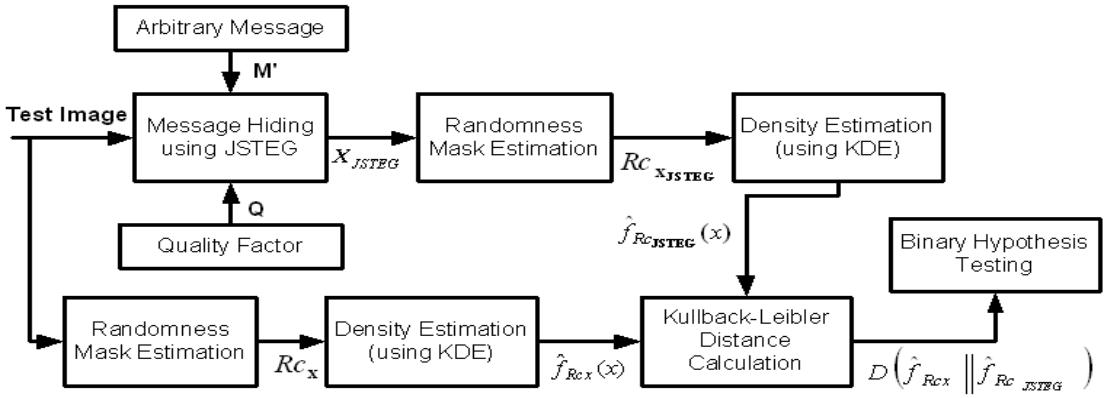

Figure 16: Block diagram of the steganalysis scheme used to attack Jsteg steganographic tool