Nonparametric Steganalysis of QIM Steganography using Approximate Entropy

Hafiz Malik†, K. P. Subbalakshmi* and R. Chandramouli*
† Electrical and Computer Engineering Department,
University of Michigan - Dearborn, Dearborn, MI 48128
* Electrical and Computer Engineering Department,
Stevens Institute of Technology, Hoboken, NJ 07030

***Abstract*—This paper proposes an active steganalysis method for quantization index modulation (QIM) based steganography. The proposed nonparametric steganalysis method uses irregularity (or randomness) in the test-image to distinguish between the cover-image and the stego-image. We have shown that plain-quantization (quantization without message embedding) induces regularity in the resulting quantized-object, whereas message embedding using QIM increases irregularity in the resulting QIM-stego. Approximate entropy, an algorithmic entropy measure, is used to quantify irregularity in the test-image. The QIM-stego image is then analyzed to estimate secret message length. To this end, the QIM codebook is estimated from the QIM-stego image using first-order statistics of the image coefficients in the embedding domain. The estimated codebook is then used to estimate secret message. Simulation results show that the proposed scheme can successfully estimate the hidden message from the QIM-stego with very low decoding error probability. For a given cover-object the decoding error probability depends on embedding rate and decreases monotonically, approaching zero as the embedding rate approaches one.**

***Index Terms*—Steganography, Steganalysis, Quantization Index Modulation, Dither Modulation, Entropy, Complexity, Approximate Entropy, Algorithmic Entropy, Message Recovery, Embedding Rate**

I. INTRODUCTION

Steganalysis refers to the act of analyzing a given multimedia data (e.g. images, video, audio etc.) for the presence of hidden messages, with limited or no access to information regarding the embedding algorithm used. Existing steganalysis techniques may be classified into passive- or active-steganalysis [1] depending on whether the aim of the steganalyst is to detect the presence/absence of the hidden message only or to extract the hidden message. Passive steganalysis typically deals with detecting the presence or absence of the hidden message and identifying the steganographic method used for embedding the hidden message. In contrast, the objectives of active steganalysis include one or more of the following: 1) estimation of the embedded message length, 2) estimation of location(s) of the embedded message, 3) estimation of the message embedding key used (if any), 4) extraction of the hidden message, and 5) estimation of parameters of the embedding algorithm.

Quantization based data hiding schemes [2] are based on Costa's seminal work [3] which gives the theoretical capacity of the Gaussian channel by modeling steganography as *communication with side information*. The ideal Costa scheme (ICS) achieves the theoretical upper bound for the capacity of all data hiding schemes under additive white Gaussian noise attack. However, the ICS requires a random codebook of infinite length which makes it impractical [4]. Practical realizations of ICS include quantization index modulation (QIM) [2], scalar Costa scheme (SCS), dither modulation (DM), and quantization projection (QP), [4]. QIM-based data hiding schemes are commonly used for steganography due to their high embedding capacity and controlled embedding distortion-robustness tradeoff.

We now briefly discuss existing QIM steganalysis techniques and set the context of our work. Guillon et al [5] proposed a framework for steganalysis of SCS by modeling QIM steganography as an additive noise channel. Sullivan et al [6] proposed a steganalysis scheme for QIM steganography using supervised learning. Detection performance of the scheme proposed in [6] is constrained by the limitations of learning-based steganalysis, that is, a separate classifier training is required for every new steganographic algorithm, and the detection performance depends on the selection of features used to train the classifier [7]. Work on non-learning based QIM steganalysis techniques include [8] and [9]. The steganalysis scheme proposed in [8] is not applicable for stego-image generated using DM-based embedding, whereas the steganalysis scheme proposed in [9] cannot extract hidden messages and cannot detect random partial embedding. Major contribution of this paper is to address limitations of existing parametric QIM steganalysis schemes. Specifically, we design a nonparametric steganalysis method for the *stego-only attack* scenario, i.e., only the stego-object is available for steganalysis.

Passive QIM steganalysis have seen significant advances in recent times [5], [6], [8]–[11]; active QIM steganalysis, on the other hand, is relatively underdeveloped. Few notable exceptions include Yu and Wang's [12] and Wu et al's [13] methods to estimate secret message length estimation form QIM stego by mathematically modeling QIM embedding

Send correspondence to Hafiz Malik, E-mail: hafiz@umd.umich.edu, Tel.: 1 313 593 5677
Send correspondence to K. P. Subbalakshami, E-mail: ksubbala@stevens.edu
Send correspondence to R. Chandramouli, E-mail: mouli@stevens.edu

distortion as a function of embedding ratio (or secret message length) and use estimated model parameters for secret message length estimation. Similarly, Kim and Bae [14] and Lee at al's [15] have proposed analytical approach using low level statistical features (mean and variance) for quantization step size estimation from QIM-stego audio signal subjected to scaling and additive white Gaussian noise attacks. Pevny and Fridrich [11], [16] have also proposed a method to detect of double JPEG compression and a maximum likelihood estimator of the primary quality factor. The proposed method uses support vector machine classifiers with feature vectors formed by histograms of low-frequency DCT coefficients.

This paper proposes a nonparametric steganalysis scheme for QIM steganography using a measure of randomness (or irregularity) to distinguish between the cover and the stego. First we show that a sequence consisting of QIM-stego image coefficients tends to exhibit higher degree of irregularity (or randomness) than a plain-quantized image. This relative irregularity in finite sequences can be used to distinguish between the cover and the stego images. Information theory offers several measures of entropy such as Shannon's entropy [17], Kolmogorov-Sinai (KS) complexity [18], [19], Lempel-Ziv (LZ) complexity [20], approximate entropy ($ApEn$) [21]–[23], etc. However, the selection of a particular irregularity measure for the test-image depends on 1) the characteristics of the underlying sources generating the cover-image, 2) the size of the test-image, and 3) the knowledge of the cover-image statistics available to the steganalyst. The proposed steganalysis scheme uses $ApEn$ to measure randomness in the test-image. Justification for selecting $ApEn$ over other randomness metrics [17]–[20], is given in Section III. Simulation results for both sequential embedding and random embedding show that the proposed steganalysis technique can distinguish between the cover- and the stego-images with low false positive rates, P_{fp}, and false negative rates, P_{fn}. In particular, the false positives rates are below 0.1 and the false negative rates are below 0.07 for DM-stego and below 0.12 and 0.002 respectively for QIM-stego.

Once the test-image is identified as a QIM-stego image it is analyzed further to estimate the secret message length. The proposed scheme uses first-order statistics to estimate quantization step-size which is then used to estimate secret message length and extract the hidden message. Performance of the proposed active steganalysis is evaluated for various embedding rate, $R \in \{10-100\}\%$ (i.e., $R\%$ of the coefficients are modified during message embedding process).

In this paper, we assume gray-scale cover images of size $N_1 \times N_2$, where $64 \leq N_1, N_2 \leq 512$ and that the embedding is done in the DCT domain. Moveover, a *stego-only attack* scenario is assumed which means that the prior probabilities of the underlying source symbols are not known to the steganalyst.

The rest of the paper is organized as follows. The requirements of QIM-steganalysis are discussed in Section II. Justification of using $ApEn$ to capture randomness in the stego-image is provided in Section III. The outline of the proposed steganalysis scheme along with simulation results for QIM-stego and dither modulation (DM)-stego detection are provided in Section IV. Details of the message estimation algorithm from the QIM-stego image are discussed in Section V. Concluding remarks and future directions are discussed in Section VI.

II. STEGANALYSIS OF QIM STEGANOGRAPHY

A key issue in QIM steganalysis is to distinguish between the following cases:

1) the quantized-cover, $\mathbf{x}_q$, (quantized image obtained using plain-quantization or without message embedding) and the QIM-stego, $\mathbf{x}_{QIM}$, (stego-image obtained using QIM), and
2) the cover, $\mathbf{s}$, and the DM-stego, $\mathbf{x}_{DM}$, (stego obtained using DM).

To design a parametric hypothesis test for stego detection, the probability mass functions of $\mathbf{s}$, $\mathbf{x}_q$, $\mathbf{x}_{QIM}$, and $\mathbf{x}_{DM}$ are required. Let $P_{\mathrm{s}}(s)$, $P_{\mathrm{x}_q}(x)$, $P_{\mathrm{x}_{QIM}}(x)$, and $P_{\mathrm{x}_{DM}}(x)$, denote probability mass function (*pmf*) of coefficients of the cover, quantized cover, QIM-stego, and DM-stego, respectively, in the DCT domain. We assume $\mathbf{s} \in \mathcal{R}$, the set of all real numbers.

A. Quantization, QIM-steganography and DM-steganography

In the case of plain-quantization, the quantizer output, say x_k, is an integer multiple of the quantization step-size, Δ^*, i.e. $x_k = k\Delta^*$. The probability mass function of quantizer output is determined by the unquantized DCT coefficients, $s_i, i = \{1, \cdots, N_k\}$, falling in the range $\mathcal{S}_q(t) \triangleq (t - \frac{k\Delta^*}{2}, t + \frac{k\Delta^*}{2}]$, i.e.,

$$P_{\mathrm{x}_q}(x_k) = \sum_{s_i \in \mathcal{S}_q(t)} P_{\mathrm{s}}(s_i) \tag{1}$$

where N_k is the number of coefficients in the range $\mathcal{S}_q(t)$ and $k \in \mathcal{Z}_+$ where $\mathcal{Z}_+$ denotes the set of all positive integers.

In case of QIM steganography, two identical quantizers are used to encode a binary message sequence, $M \in \{0,1\}^N$, of length N into the host data. Each quantizer is designed with a step-size $\Delta = 2\Delta^*$ and is offset (shifted) from the other by $\Delta/2$. That is, $Q_0(x) = Q_1(x) \pm \Delta/2$, where $Q_0(\cdot)$ and $Q_1(\cdot)$ denote quantizers used to embed message bit '0' and '1' respectively. The difference between plain-quantization and message embedding using QIM is illustrated in Fig. 1.

For QIM with equiprobable message bits, $Pr[m = 0] = Pr[m = 1] = \frac{1}{2}$, the probability of a given output, x_k, can be expressed as,

$$P_{\mathrm{x}_{QIM}}(x_k) = \frac{1}{2} \sum_{s_i \in \mathcal{S}_{QIM}(t)} P_{\mathrm{s}}(s_i) \tag{2}$$

where $\mathcal{S}_{QIM}(t) \triangleq (t - \Delta_k/2, t + \Delta_k/2]$, $x_k = k\Delta$, and $\Delta_k = k\Delta$.

In case of dither modulation, two dither quantizers are used to embed the message bits. A dither quantizer is obtained by adding (or subtracting) a dither value d_u to the quantizer output, x_k, where $\mathbf{d}_u$ is uniformly distributed noise over $[-\Delta/4, \Delta/4]$. Therefore, the quantizer output covers the entire range of the cover-image, unlike in the case of QIM or plain quantization. In this range, $P_{\mathbf{d_u}}(d_u) = 2\epsilon/\Delta$, where ϵ is the

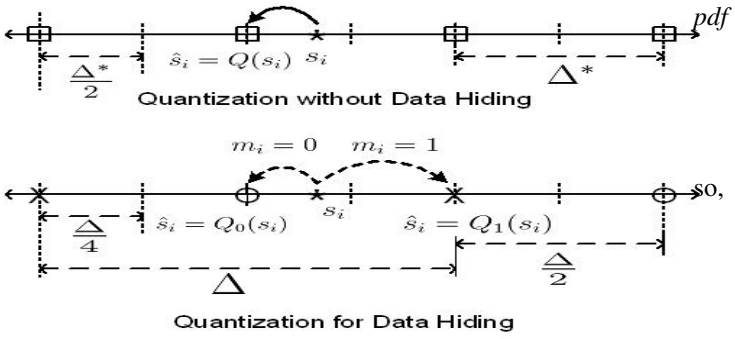

Fig. 1. Shown is the illustration of plain-quantization (in the upper panel) and binary QIM (in the lower panel). The reconstruction grid points corresponding to 'O' and 'X' are used to embed message symbols '0' and '1' respectively, in the lower panel.

granularity of the data. Data hiding based on DM can be expressed as,

$$x_{DM_i}(\mathbf{s}, M) = Q_{m_i}(s_i + d_{u_i}) - d_{u_i},\ i = 0, 1, \cdots\ N-1. \quad (3)$$

x_{DM_i} is generated using one and only one value of d_{u_i}. The theory of subtractive dither (SD) quantization [24], [25] can be used to determine the probability density function *pdf* $P_{DM}(x)$ of $\mathbf{x}_{DM}$. Let $x_{DM_i} = Q_{m_i}(s_i + d_{u_i}) - d_{u_i}$ and quantization error $\varepsilon_i = x_{DM_i} - s_i$. Let us also assume that random variables (rvs) $\mathbf{s}$, M, and $\mathbf{d}_u$ are mutually independent. We use *Schuchmans condition* [24], [25] to determine the *pdf* of $\mathbf{x}_{DM}$ as follows.

Theorem 1. ***[Schuchmans Condition]***
*In an SD quantizing system with step size Δ, the total error is statistically independent of the system input for arbitrary input distributions if and only if the characteristic function (*cf*) of the dither, CF_d, satisfies the condition*

$$CF_d\left(\frac{k}{\Delta}\right) = 0 \qquad \vee k \in \mathcal{Z}_+ \quad (4)$$

Furthermore, the total error will be uniformly distributed for arbitrary input distributions if and only if this condition holds.

Proof: Proof of this theorem can be found in [24], [25]. ■

As the dither vector, $\mathbf{d}_u$, is uniformly distributed over $[-\Delta/4, \Delta/4]$ for DM-steganography, and the corresponding *cf* is a *sinc* function defined as,

$$CF_d(u) = sinc(u) \triangleq \frac{\sin(\pi u \Delta/2)}{\pi u \Delta/2} \quad (5)$$

which satisfies *Schuchmans condition*, i.e. $CF_d\left(\frac{k}{\Delta}\right) = 0 \quad \vee k \in \mathcal{Z}_+$, the resulting quantization error, ε, is uniformly distributed over $[-\Delta/2, \Delta/2]$ and statistically independent of $\mathbf{s}$. Now to determine $P_{\mathrm{x}_{DM}}(x)$, consider the following model for DM-steganography,

$$\mathbf{x}_{DM} = \mathbf{s} + \varepsilon. \quad (6)$$

In this case, $P_{\mathrm{x}_{DM}}(x)$ can be obtained by simply convolving *pdf* of ε, $P_\varepsilon(x)$, and $P_{\mathbf{s}}(x)$, where $P_\varepsilon(x)$ is defined as,

$$P_\varepsilon(x) = \begin{cases} \frac{\epsilon}{2\Delta} & |x| < \Delta/4 \\ \frac{\epsilon}{4\Delta} & |x| = \Delta/4 \\ 0 & otherwise, \end{cases}$$

so,

$$P_{\mathrm{x}_{DM}}(x_i) = (P_\varepsilon \star P_{\mathbf{s}})(x) \quad (7)$$

$$= \frac{\epsilon}{\Delta} \sum_{k=-\Delta/4}^{\Delta/4} P_{\mathbf{s}}(s_{i-k}) \quad (8)$$

where, $\star$ denotes convolution operation.

Using the *pmf* (resp. *pdf*) of the output of the QIM (resp. DM) quantizer, a likelihood ratio test (***LRT***) can be set up for stego detection. The ***LRT*** can be expressed as,

$$L(x) \triangleq \frac{P_{\mathrm{x}_{QIM}}(x)}{P_{\mathrm{x}_q}(x)} \gtrless \tau \quad (\text{ detect QIM-stego}) \quad (9)$$

$$\triangleq \frac{P_{\mathrm{x}_{DM}}(x)}{P_{\mathrm{s}}(x)} \gtrless \tau \quad (\text{detect DM-stego}) \quad (10)$$

where the decision threshold, τ, can be minimized using Neyman-Pearson rule which maximizes the probability of detection, P_d, for a given probability of false alarm, P_f [26].

Substituting $P_{\mathrm{x}_{QIM}}(x)$ and $P_{\mathrm{x}_q}(x)$ from Eq. (1 & 2) in Eq. (4):

$$L(x) = \prod_{i=1}^{N} \left(\frac{\frac{1}{2} \sum_{s \in (x_i - \Delta/2, x_i + \Delta/2]} P_s(s)}{\sum_{s \in (x_i - \Delta/4, x_i + \Delta/4]} P_s(s)} \right) \quad (11)$$

Eq. (11) shows that the likelihood statistic is a function of the cover *pdf*, $P_{\mathrm{s}}(s)$, and under stego-only attack scenario $P_{\mathrm{s}}(s)$ is not available at the stego detector. Therefore, parametric detection based on Neyman-Pearson rule cannot be used to detect the QIM-stego image.

Similarly, to detect DM-stego, we obtain the likelihood ratio by substituting $P_{\mathrm{x}_{DM}}(x)$ from Eq. (8) in Eq. (10):

$$L(x) = \prod_{i=1}^{N} \left(\frac{\frac{\epsilon}{\Delta} \sum_{k=-\Delta/4}^{\Delta/4} P_{\mathbf{s}}(s_{i-k})}{P_s(s)} \right) \quad (12)$$

Eq. (12) shows that the likelihood statistic is also a function of the cover *pdf*, $P_{\mathrm{s}}(s)$, therefore parametric detector cannot be used for DM-stego detection either. An important observation however can be made from Eq. (11 & 12) that message embedding using QIM or DM introduces smoothness in the *pmf* of the resulting stego image. To highlight this claim further, we analyze the empirical *pmf*s (obtained using histograms) of the quantized-cover and the QIM-stego images. The empirical *pmf*s of DCT coefficients of the QIM-stego for $\Delta = \{0.5, 4, 8\}$ are shown in Fig. 2. Shown in Fig. 3 the comparison of smoothing effect due plain-quantization and QIM. Some of the experimental observations on the difference between the QIM-stego and the quantized-cover images based on their empirical *pmf*s are summarized below.

Firstly, we note that the quantization (with and without messageembedding) introduces smoothness in the *pmf* of the resulting quantized images. It can be observed from Fig. 2 that as Δ increases smoothing effect in the *pmf* of the resulting QIM-stego also increases according to the Eq. (11). Secondly,

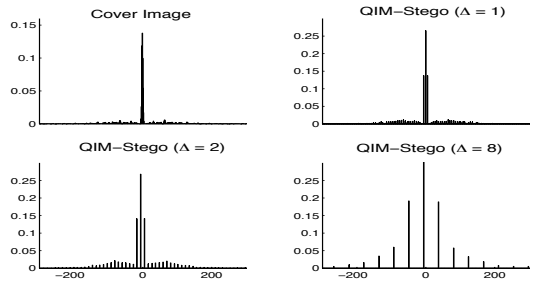

Fig. 2. Shown are the empirical *pmf*s (based on histogram) of DCT coefficients of the cover (top-left) and quantized DCT coefficients of the QIM-stego obtained with Δ = {0.5, 4, 8}(top-right and the bottom-row)

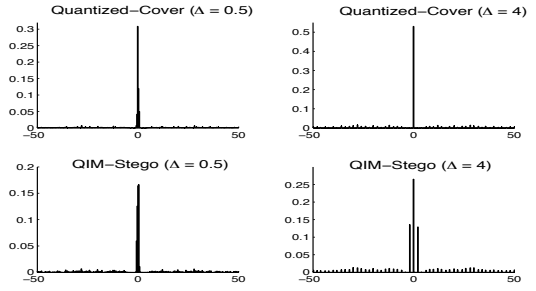

Fig. 3. Shown are empirical *pmf*s of the quantized-cover (top-row) and the corresponding QIM-stego (bottom-row)

the quantizer step-size, Δ, controls the amount of smoothness introduced in the *pmf* of the quantized-image. Finally, quantization with message embedding (e.g. QIM) introduces more smoothness than plain-quantization. It can be observed from Fig. 3 that for the same value of Δ, the QIM introduces more smoothness than plain-quantization. Moreover, for large Δ ($\Delta \geq 4$) message embedding using QIM splits the peak around zero in the cover pmf into three peaks (e.g. peaks $P_{-\Delta}, P_0$, and P_{Δ} around $-\Delta, 0$, and Δ respectively), which can be used to distinguish between the quantized-cover and the QIM-stego. However, such visual attacks might not guarantee consistent results especially when QIM-stego is generated using smaller quantization step-size or the the cover-image has smoothly varying *pmf*. Relative smoothness in the *pmf* of the test-image can be used to distinguish between the cover and the stego. Learning-based steganalysis techniques have been proposed in the past [6] to distinguish between the quantized-cover and the QIM-stego, but as noted earlier, there are some inherent disadvantages with these steganalysis schemes.

To address the limitations of learning-based steganalysis schemes for QIM steganography, a nonparametric steganalysis scheme based on measure of randomness in the test-image is proposed here. The proposed scheme exploits relative randomness in the test-image to distinguish between the cover- and the stego-images.

Theorem 2. *If $\mathbf{x}_q \triangleq Q(\mathbf{s})$ is a quantized sequence obtained using plain-quantization (uniform quantization without message embedding) then,*

$$H\left(Q(\mathbf{s})\right) \leq H(\mathbf{s}) \tag{13}$$

where $H(x)$ is Shannon's entropy of rv x.

Proof: The proof of this theorem is given in Appendix A. ■

It is interesting to note that Theorem 1 gives similar interpretation of a ν-bit quantization of a continuous random variable, $\mathbf{s}$, in terms of entropy as shown in [27] (see p. 229), which states that entropy of an ν-bit quantization of $\mathbf{s}$ can be approximated as $h(\mathbf{s}) + \nu$, where $h(\mathbf{s})$ denotes differential entropy of a continuous random variable $\mathbf{s}$.

Theorem 3. *If $\mathbf{x}_{QIM} \triangleq Q_{QIM}(\mathbf{s}, \mathbf{m})$ is a quantized sequence obtained using QIM (uniform quantization with message embedding) and $\mathbf{x}_q \triangleq Q(\mathbf{s})$ is a quantized sequence obtained using plain quantization (uniform quantization without message embedding) then,*

$$H\left(\mathbf{x}_{QIM}\right) \geq H\left(\mathbf{x}_q\right), \tag{14}$$

*Proof: Let $\mathbf{s} = \{s_i\}_{i=1}^{N}$ be a real valued random sequence to be quantized with associated (*pdf*) P_s. A uniform quantizer $Q_{N_0}(s)$ is defined as partition $\Lambda = \Delta_k = [t_k, t_{k+1}), t_{k+1} > t_k, k = \{1, \cdots, N_0\}$ where N_0 is the number of equilength partitions, and a reconstruction codebook $\hat{x}_k$ defined as $Q(s) = \hat{x}_k, s \in \Delta_k$. Let $\mathbf{x}_q = \{\hat{x}_k\}_{k=1}^{N_0}$ be the plain-quantizer output. Similarity, let $\mathbf{x}_{QIM}$ be the quantized sequence obtained by embedding a binary message $\mathbf{m} \in \{0,1\}^N$ (with $Pr[m=0] = Pr[m=1] = \frac{1}{2}$) independent of $\mathbf{s}$, using QIM quantizer with partition length Δ.*

The mutual information between the continuous random variable $\mathbf{s}$ and the corresponding discrete random sequence, $\mathbf{x}_q = Q(\mathbf{s})$ obtained using plain-quantization can be expressed in the following two forms,

$$\begin{aligned} I\left(\mathbf{s}, Q(\mathbf{s})\right) &= H\left(Q(\mathbf{s})\right) - H\left(Q(\mathbf{s})|\mathbf{s}\right) && (15)\\ &= h\left(\mathbf{s}\right) - h\left(\mathbf{s}|Q(\mathbf{s})\right) && (16)\end{aligned}$$

where H denotes Shannon's entropy and h the differential entropy. Since $Q(\mathbf{s})$ is a deterministic function of $\mathbf{s}$, $H\left(Q(\mathbf{s})|\mathbf{s}\right) = 0$, hence self-information of the plain quantizer output can be expressed as,

$$H\left(Q(\mathbf{s})\right) = h\left(\mathbf{s}\right) - h\left(\mathbf{s}|Q(\mathbf{s})\right), \tag{17}$$

Similarly, mutual information between continuous random variable $\mathbf{s}$ and the corresponding discrete random sequence, $\mathbf{x}_{QIM} = Q_{QIM}(\mathbf{s}, \mathbf{m})$ obtained using QIM can be expressed in the following two forms,

$$\begin{aligned} I\left(\mathbf{s}, Q_{QIM}(\mathbf{s}, \mathbf{m})\right) &= H\left(Q_{QIM}(\mathbf{s})\right) \\ &\quad -H\left(Q_{QIM}(\mathbf{s})|\mathbf{s}\right) && (18)\\ &= h\left(\mathbf{s}\right) - h\left(\mathbf{s}|Q_{QIM}(\mathbf{s})\right) && (19)\end{aligned}$$

In this case, $Q_{QIM}(\mathbf{s})$ is a not a deterministic function of $\mathbf{s}$, therefore $H\left(Q_{QIM}(\mathbf{s}, \mathbf{m})|\mathbf{s}\right) \neq 0$, hence self-information the QIM quantizer output can be expressed as,

$$\begin{aligned} H\left(Q_{QIM}(\mathbf{s}, \mathbf{m})\right) &= h\left(\mathbf{s}\right) - h\left(\mathbf{s}|Q_{QIM}(\mathbf{s}, \mathbf{m})\right) \\ &\quad +H\left(Q_{QIM}(\mathbf{s}, \mathbf{m})|\mathbf{s}\right), && (20)\end{aligned}$$

Subtracting Eq. (17) from Eq. (20) we obtain,

$$\begin{aligned} H\left(Q_{QIM}(\mathbf{s},\mathbf{m})\right) - H\left(Q(\mathbf{s})\right) &= h\left(\mathbf{s}|Q(\mathbf{s})\right) && (21)\\ &\quad -h\left(\mathbf{s}|Q_{QIM}(\mathbf{s},\mathbf{m})\right)\\ &\quad +H\left(Q_{QIM}(\mathbf{s},\mathbf{m})|\mathbf{s}\right)\\ &\overset{(a)}{=} h\left(\mathbf{s}\right) - h\left(Q(\mathbf{s})\right) && (22)\\ &\quad -h\left(\mathbf{s}|Q_{QIM}(\mathbf{s},\mathbf{m})\right)\\ &\quad +H\left(Q_{QIM}(\mathbf{s},\mathbf{m})|\mathbf{s}\right)\\ &\overset{(b)}{\approx} -h\left(Q(\mathbf{s})\right) && (23)\\ &\quad +H\left(Q_{QIM}(\mathbf{s},\mathbf{m})\right) \end{aligned}$$

where (a) follows from the fact that $h\left(\mathbf{s}|Q(\mathbf{s})\right) = h\left(Q(\mathbf{s})|\mathbf{s}\right) + h\left(\mathbf{s}\right) - h\left(Q(\mathbf{s})\right)$*, since* $Q(\mathbf{s})$ *is a deterministic function of* $\mathbf{s}$*,* $h\left(Q(\mathbf{s})|\mathbf{s}\right) = 0$ *and (b) from the fact that*

$$\begin{aligned} h\left(\mathbf{s}|Q_{QIM}(\mathbf{s},\mathbf{m})\right) &= h\left(Q_{QIM}(\mathbf{s},\mathbf{m})|\mathbf{s}\right) + h(\mathbf{s}) && (24)\\ &\quad -h\left(Q_{QIM}(\mathbf{s},\mathbf{m})\right)\\ &\overset{(c)}{\approx} H\left(Q_{QIM}(\mathbf{s},\mathbf{m})|\mathbf{s}\right) + h(\mathbf{s}) && (25)\\ &\quad -H\left(Q_{QIM}(\mathbf{s},\mathbf{m})\right) \end{aligned}$$

where (c) follows from the fact that $h\left(Q_{QIM}(\mathbf{s},\mathbf{m})|\mathbf{s}\right) \approx H\left(Q_{QIM}(\mathbf{s},\mathbf{m})|\mathbf{s}\right) + \log(\Delta)$ *and* $h\left(Q_{QIM}(\mathbf{s},\mathbf{m})\right) \approx H\left(Q_{QIM}(\mathbf{s},\mathbf{m})\right) + \log(\Delta)$.

Since, $h\left(Q(\mathbf{s})\right) \leq 0$ *(differential entropy of discrete r.v. can be consider* ≤ 0 *([27] Ch. 9, pp. 229)), and* $H\left(Q_{QIM}(\mathbf{s},\mathbf{m})\right) \geq 0$

$\Rightarrow$

$H\left(Q_{QIM}(\mathbf{s},\mathbf{m})\right) \geq H\left(Q(\mathbf{s})\right)$ ■

This fact is illustrated in Fig. 4. It can be observed from Fig. 4 that the distortion due to message embedding using QIM is relatively more irregular (random) than the distortion due to plain-quantization (especially in low-texture regions). This implies that coefficients of the quantized-cover image are relatively more predictable (regular) than the corresponding coefficients in the QIM-stego image. The proposed steganalysis scheme uses relative irregularity in the test-image to distinguish between the cover, $(\mathbf{s}, \mathbf{x}_q)$, and the stego, $(\mathbf{x}_{QIM}, \mathbf{x}_{DM})$, images.

The proposed schemes uses $ApEn$ to access randomness in the test-image. The next section provides motivation for using $ApEn$ to capture irregularity in the test-image along with a brief overview of other irregularity measures such as Shannon's entropy [17], Kolmogorov-Sinai (KS) complexity [18], [19], Lempel-Ziv (LZ) complexity [20], etc.

III. Why Approximate Entropy?

The proposed steganalysis scheme uses irregularity in the test-image to attack QIM steganography [1]. Entropy measuring tools in the information theory literature such as Shannon's entropy [17], Kolmogorov-Sinai (KS) complexity [18], [19], Lempel-Ziv (LZ) complexity [20], approximate entropy ($ApEn$) [21]–[23], etc. can be used to measure irregularity in the test-image. However, selection of a particular irregularity measure in the test-image depends on 1) the characteristics of the underlying sources generating the cover-image, 2) the size of the test-image, and 3) whether the cover-image statistics is available to the steganalyst or not. Therefore, entropy measures presented in [17]–[23] cannot be used blindly to quantify the irregularity of the time-series generated from the test-image.

[1]for rest of the paper QIM steganography means message embedding using QIM or DM unless otherwise specified

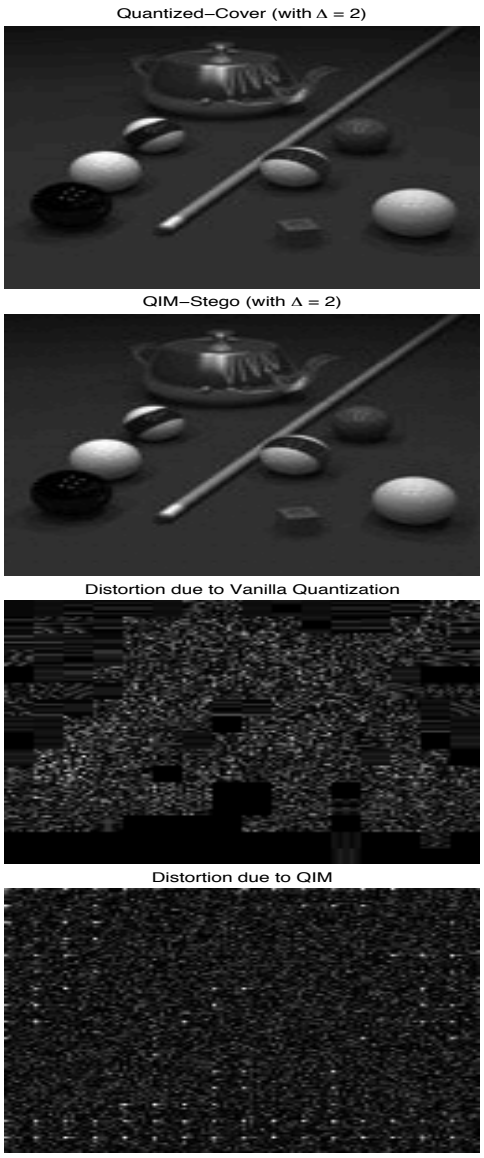

Fig. 4. Illustration of quantization noise: quantized-cover and quantization noise (left); QIM-stego and the corresponding quantization noise (right)

For example, KS complexity is an algorithmic measure [18], [19] which uses rate of information generation to classify deterministic dynamical systems. But the KS complexity methods fail to quantify time-series representing output of a stochastic or mixed processes [21], [22]. Moreover, the KS complexity is very sensitive to small amount of noise or outliers. These inabilities of KS complexity to quantify irregularity in stochastic processes or noisy data can be attributed to its non-statistical framework used to calculate complexity in the time-series. Therefore, the application of KS complexity

to practical time-series like the DCT coefficient of the test-image, will only evaluate noise not the properties of the underlying sources. In addition, KS complexity requires large amount of data (theoretically infinite sequence) to converge [28]. Therefore KS complexity cannot be used to quantify smaller sequences, generated using test-images, based on their estimated KS complexities.

Shannon proposed entropy as a measure of randomness (or irregularity [17]) in the output of a probabilistic source that generates an infinite sequence of symbols. Entropy characterizes the irregularity of a given source by the probabilities of symbols and blocks of symbols. Shannon's probabilistic entropy [17] requires prior probabilities of the underlying source symbols or block of symbols to estimate irregularity in a given sequence. However it cannot be used in our case, as we assume *stego-only attack* scenario where probabilities of the symbols and block of symbols are not available to the steganalyst.

Pincus proposed an algorithmic entropy method, known as approximate entropy (ApEn) in [21]–[23] to measure irregularity (or complexity) in the finite sequences when prior probabilities of symbols and blocks of symbols are not known. The $ApEn$ makes no prior assumption on the sequence of symbols or the source generating it. The $ApEn$ is motivated by Shannon's information-theoretic entropy of a Markov process rather than by the conditional complexity of algorithmic information theory [21]–[23]. The $ApEn$ is very useful in discriminating finite sequences based on their relative irregularity. The $ApEn$ is a statistical tool designed to quantify irregularity in the time-series [21]–[23]. Mathematically, $ApEn$ is a natural information theoretical parameter, i.e. the rate of entropy, for an approximating Markov chain to a process [22], [29]. The $ApEn$ provides both noise filtering and artifacts suppression capabilities through suitable filtering threshold selection [22]. In addition, despite algorithmic similarities, the $ApEn$ is not an approximate value of the KS entropy [18], [19] rather it is a family of statistics parameterized by the filtering threshold, r, and embedding dimension, ξ [21], [22], [30]. The salient features of the $ApEn$ make it an attractive candidate to access irregularity in the real-world practical finite or periodic sequences:

- $ApEn$ is an algorithmic entropy measure,
- its robustness to the noise as long as noise is below a specified filtering threshold,
- it is applicable to short sequences, for example, it is possible to estimate regularity with good confidence level using only a few hundred points,
- a change in the estimated $ApEn$ corresponds to change in the complexity of the underlying process, and
- $ApEn$ allows a direct computable alternative to severely noncomputable approaches like KS complexity,

The proposed steganalysis scheme uses $ApEn$ to estimate irregularity in the test-image. An algorithm to calculate $ApEn$ from a finite-length sequence and its mathematical interpretation are discussed next.

A. Approximate Entropy Estimation

Approximate entropy is a *regularity statistic* that quantifies irregularity or fluctuations in a time-series, $\{x\}_1^n$, where n is the number of observations of the time-series. The $ApEn$ reflects the likelihood that blocks of length ξ that are close together remain close together for blocks augmented by one position in the following observations. A time-series containing many repetitive patterns (e.g. a regular sequence) exhibits a relatively small $ApEn$ value, whereas a time-series consisting of less predictable patterns (or a more irregular sequence) exhibits higher $ApEn$ value. A detailed description of the algorithm for computing $ApEn$ and its statistical properties can be found in [21]–[23], [31]–[33] and references therein.

Definition of ApEn: Consider a time-series sequence, $\{x\}_1^n$, consisting of n measurements equally spaced in time i.e. $x_1,\ x_2,\ \cdots,\ x_n$. For a fixed-positive integer ξ and a positive real number r, consider embedding vectors $\mathbf{u}_{(1)},\ \mathbf{u}_{(2)},\ \cdots,\ \mathbf{u}_{(n-\xi+1)}$ in $\mathcal{R}^{\xi}$, where $\mathbf{u}_{(i)} = [x_i,\ x_{i+1},\ \cdots,\ x_{i+\xi-1}]$. Let us define the correlation measure, $C_i^{\xi}(r)$, for every $i,\ 1 \leq\ i\ \leq\ n-\xi+1$,

$$C_i^{\xi}(r) = \frac{[(\#\text{ of vectors } j \leq\ (n-\xi+1)) \mid (d(\mathbf{u}_i, \mathbf{u}_j) \leq\ r)]}{n-\xi+1} \tag{26}$$

where $d(\mathbf{u}_i, \mathbf{u}_j)$ is the L_∞ norm between vectors $\mathbf{u}_i$ and $\mathbf{u}_j$, which can be expressed as,

$$d(\mathbf{u}_i, \mathbf{u}_j) = \underset{k=1,\cdots,\xi}{max} |\,(u(i+k-1) - u(j+k-1))\,| \tag{27}$$

here the quantity $C_i^{\xi}(r)$ is a fraction of patterns of length ξ that resemble the pattern of the same length that begins at index i. In other words, $C_i^{\xi}(r)$ measures the regularity (or frequency) of patterns similar to a given pattern of window length ξ and a tolerance r.

The approximate entropy, $ApEn(\xi, r, n)$, of a sequence $\{x\}_1^n$, with parameters ξ, r, and n is defined as,

$$ApEn(\xi, r, n) = \left[\Phi^{\xi}(r) - \Phi^{\xi+1}(r)\right], \tag{28}$$

where

$$\Phi^{\xi}(r) = \frac{\sum_{i=1}^{n-\xi+1} \ln\ C_i^{\xi}(r)}{n-\xi+1} \tag{29}$$

and,

$$\Phi^{\xi}(r) - \Phi^{\xi+1}(r) = E_i\{\log\left(Pr\left[(\alpha \leq\ r) \mid (\beta \leq\ r)\right]\right)\} \tag{30}$$

where $\alpha = |u(j+\xi) - u(i+\xi)|$, $\beta = |u(j+k) - u(i+k)|$, $k = 0, 1, \cdots, \xi-1$, E_i denotes average over i, and $Pr[\cdot|\cdot]$ is conditional probability.

The $ApEn(\xi, r, n)(\cdot)$ measures the logarithmic frequency with which blocks of length ξ that are close together remain close together for blocks augmented by one position. A smaller value of $ApEn$ implies regularity in the time-series, that is, similar patterns are highly predictable from additional similar measurements. Whereas, a large value of $ApEn$ indicates that the underlying time-series is highly irregular. For a given application $ApEn(\xi, r, n)$ should be considered as a family of statistics and for time-series comparisons a fixed set of values of ξ and r should be used.

IV. STEGANALYSIS USING *ApEn*

We used a measure of irregularity in the test-image to decide if the given image is stego or not. Irregularity in the test-image is measured in terms of estimated $ApEn$ from the test-image. To calculate $ApEn$ from the test-image ($\mathbf{S}$, $\mathbf{x}_q$, $\mathbf{x}_{QIM}$, or $\mathbf{x}_{DM}$) using the $ApEn(\xi, r, n)$ algorithm outlined in Section III-A, the test-image must be transformed into finite sequences. To this end, the test-image is segmented into non-overlapping blocks, of 8x8 pixels, and the two-dimensional ($2D$) DCT for each block is calculated. Each block in the DCT domain is then converted into a one-dimensional ($1D$) vector using zigzag ordering (commonly used during baseline JPEG compression [34]). These $1D$ blocks of the test-image are used to generate 64 sequences, $\mathbf{x}_n^i, i = \{0, \cdots, 63\}$, each of length n. Here $n = \lfloor \frac{N_1}{8} \rfloor \times \lfloor \frac{N_2}{8} \rfloor$ where $\lfloor x \rfloor$ denote the largest integer not exceeding x. Fig. 5 illustrates the finite-length sequence generation process from the test-image.

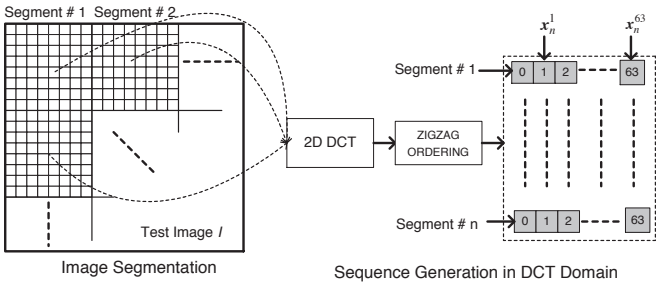

Fig. 5. Finite-length sequence generation from the test-image

The resulting finite sequences are then analyzed to estimate randomness in the test-image. To estimate randomness (or irregularity) in the test-image, finite sequences, $\mathbf{x}_n^i, i = \{1, \cdots, 63\}$ are analyzed using Eq. (28) which generates a 63-dimensional vector of $ApEn$ estimates, i.e.,

$$ApEn_i = ApEn(\mathbf{x}_n^i, \xi, r, n), i = 1, \cdots, 63 \qquad (31)$$

This vector, $\mathbf{ApEn}$, represents the randomness in the test-image and is used to distinguish between the cover- and the stego-image.

A. Steganalysis of QIM-stego

To investigate the effects of message embedding using QIM on the irregularity of the resulting QIM-stego image, the $ApEn(\xi, r, n)$ is calculated from $\mathbf{S}$, $\mathbf{x}_q$ and $\mathbf{x}_{QIM}$. To this end, two quantized images (e.g. $\mathbf{x}_q$ and $\mathbf{x}_{QIM}$) were generated from an uncompressed cover-image, $\mathbf{S}$, of size 256x256 using uniform quantizers with $\Delta = 2$ and $\Delta^* = 1$. To obtain quantized images, we used image number 47 of the image database downloaded from [35] as a cover-image. The cover-image was resized to 256x256 and converted to gray-scale. To embed binary message into the gray-scale cover-image using QIM, the cover-image was first segmented into non-overlapping blocks, each of 8x8 pixels and then the $2D$ DCT transform was applied to each block followed by message embedding using QIM. A 64 KB binary message was embedded in the cover-image using binary QIM which yielded the QIM-stego image. Similarly, the corresponding quantized-cover image was obtained. Both the quantized-cover and the QIM-stego images were then transformed into 64 $1D$ sequences each. The $ApEn$ was estimated from these $1D$ sequences generated from the AC coefficients (in the DCT domain) of $\mathbf{S}$, $\mathbf{x}_q$ and $\mathbf{x}_{QIM}$, with parameter settings $\xi = 4$ and $r = 0.1 \times \sigma_x$. Fig. 6 shows plots of the estimated $ApEn$ from $\mathbf{S}$, $\mathbf{x}_q$, and $\mathbf{x}_{QIM}$ in DCT domain. In Fig. 6 the horizontal axis represents the sequence number (AC coefficients number) and the vertical axis represents the estimated $ApEn$ (or level of randomness in each sequence).

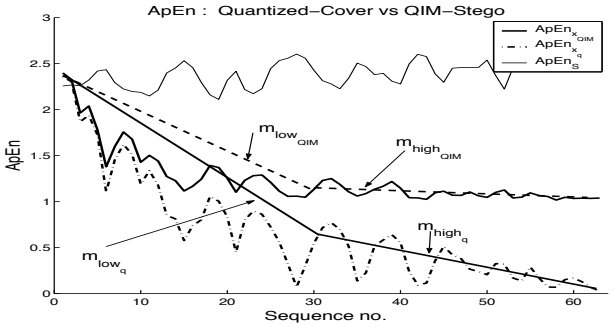

Fig. 6. Plots of the estimated $ApEn$ from $\mathbf{S}$, $\mathbf{x}_q$, and $\mathbf{x}_{QIM}$ in DCT domain

Following observations can be made from Fig. 6:

- The estimated $ApEn$ from $\mathbf{S}$ remains approximately constant for all unquantized images sequences which implies that all unquantized images exhibit approximately same level of randomness.
- In general, the estimated $ApEn$ from $\mathbf{x}_q$ and $\mathbf{x}_{QIM}$ decreases from low to high frequency. Here low and high frequency correspond to sequence number 1 to 32 and sequence number 32 to 63, respectively.
- For both quantized-images, i.e. $\mathbf{x}_q$ and $\mathbf{x}_{QIM}$, the estimated $ApEn$ decreases at a higher rate in the low frequency-coefficients than in the high frequency-coefficients.
- The estimated $ApEn$ from $\mathbf{x}_{QIM}$ has lower gradient in both frequency regions than the estimated $ApEn$ from $\mathbf{x}_q$.
- Let m_{low} and m_{high} denote gradient of the estimated $ApEn$ in low and high frequency-coefficients respectively, and change in the gradient, δm, of the estimated $ApEn$. Then δm is given by

$$\delta m = (m_{low} - m_{high})/m_{low} \times 100 \qquad (32)$$

  The value of δm for the quantized-cover is well below 50% (36% to be exact) and δm is well above 50% (85% to be exact) for the QIM-stego.
- For QIM-stego, the estimated $ApEn$ is approximately constant for high frequency coefficients.
- The estimated $ApEn$ from the QIM-stego is higher than the $ApEn$ estimated from the quantized-cover in high frequency coefficients which implies that in high frequency coefficients the QIM-stego is relatively more irregular than the corresponding quantized-cover. This higher $ApEn$ value in the QIM-stego compared to the

corresponding quantized-cover can be attributed to the randomness in the embedded message M.

These observations indicate that variation in the gradient of the estimated $ApEn$ from low to high frequency coefficients along with $ApEn$ value in the high frequency coefficients can be used to distinguish between the quantized-cover and the QIM-stego. The proposed steganalysis scheme however uses relative change in the gradient , δm, from low to high frequency-coefficients to detect QIM-stego image. A schematic diagram of the proposed steganalysis scheme against QIM-steganography is given in Fig. 7.

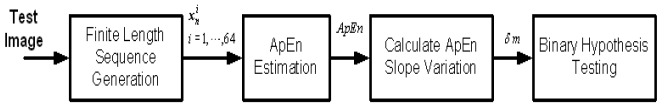

Fig. 7. Schematic diagram of the proposed steganalysis scheme to distinguish between the quantized-cover and the QIM-stego

B. Experimental Results for QIM-Stego

Detection performance of the proposed steganalysis scheme to detect QIM steganography was tested for the following message embedding strategies,

- ***Sequential Embedding*** (SE): In this case, for each DCT block, same set of AC coefficients is selected for message embedding. In addition, for sequential embedding we considered the following two cases,
  1) *all frequency embedding* (AFE),where the message, M, is embedded into all AC coefficients of 8x8 blocks in DCT domain using QIM, and
  2) *mid-frequency embedding* (MFE), where the message M is embedded into AC coefficient number 5 to 32 of zigzag scanned 8x8 blocks in the DCT domain. The MFE is commonly used to introduce lower embedding distortion at the cost of embedding capacity but without compromising robustness of the embedded message.
- ***Random Embedding*** (RE): In this case, for each DCT block, a set of AC coefficients is selected randomly for message embedding. For random embedding, embedding rates $R \in \{0.3, 0.5, 0.7, 0.9, 1.0\}$ are considered for random AC coefficient selection for message embedding.

1) Sequential Embedding: Detection performance of the proposed steganalysis scheme for QIM steganography is evaluated in terms of false rates, that is, false positive rate, P_{fp}, and false negative rate, P_{fn}. Image s from the *Uncompressed Colour Image Database* (UCID) [35] was used to evaluate performance of the proposed steganalysis scheme for QIM-stego detection. This image database [35] contains 1338 uncompressed color images, however results presented in this paper are based on gray-scale versions of the first 1000 images of the database [35]. Moreover, these 1000 images of the database [35] were resized to 256x256. Two-thousand QIM-stego images were obtained by sequentially embedding 2000 random messages into first 1000 images of the database using QIM with $\Delta = 2.0$ (1000 QIM-stego images using AFE and 1000 QIM-stego images using MFE). Similarly, 1000 quantized-images were obtained by quantizing these 1000 gray-scale images using $\Delta^* = 1$. The proposed steganalysis scheme was then applied to the resulting 3000 quantized images (1000 QIM-stego using AFE, 1000 QIM-stego using MFE, and 1000 quantized-cover). Shown in Table I is detection performance the proposed steganalysis scheme, these simulation results are generated with decision threshold $\delta m = 50\%$ and estimation parameters $\xi = 2$, $r = 0.1 \times \sigma_x$, and $n = 1024$. It is important to mention that, simulation results for MFE listed in the Table I are generated using abrupt changes in the estimated $ApEn$ from the test image at the interfaces of message carrying coefficients, i.e., finite sequence no. 5 ($\mathbf{x}_n^5$) and finite sequence no. 32 ($\mathbf{x}_n^{32}$) was used for QIM-stego detection.

TABLE I
STEGO DETECTION PERFORMANCE: ***Sequential Embedding***

	$\mathbf{X}_q$ *vs* $\mathbf{X}_{QIM}$		$\mathbf{S}$ *vs* $\mathbf{X}_{DM}$	
Error	*AFE*	*MFE*	*AFE*	*MFE*
P_{fp}	0.12	0.08	0.1	0.05
P_{fn}	0.002	0.001	0.07	0.01

It can be observed that Table I that the proposed non-parametric steganalysis scheme can successfully distinguish between the quantized-cover and the QIM-stego images with relatively low false rates, e.g., $P_{fp} < 0.12$, $P_{fn} < 0.02$. In addition, MFE embedding is relatively less secure (here security of an embedding algorithm is measured in terms of detection rate) that the AFE embedding, though MFE embedding introduces less distortion than AFE case. This is mainly because, detector for MFE is using different detection criterion and is exploiting prior knowledge about embedding algorithm.

2) Random Embedding: Similarly, to evaluate performance of the proposed scheme to attached QIM stego generated using random embedding, first 200 images of the *Uncompressed Colour Image Database* (UCID) [35] was used. The selected 200 images of the database [35] were resized to 256x256. One thousand QIM-stego images were obtained by embedding 200 random messages using QIM with $\Delta = 4.0$ and embedding rate $R \in \{0.3, 0.5, 0.7, 0.9, 1.0\}$, here QIM-stego images were generated by randomly selecting $R\%$ AC coefficients of the input image for message embedding and the remaining $(1 - R)\%$ coefficients were quantized using plain-quantizer (without message embedding) with $\Delta^* = 2$. Similarly, 200 quantized-images were obtained by quantizing selected 200 gray-scale images using plain-quantizer with $\Delta^* = 2$. The proposed steganalysis scheme was then applied to the resulting 1200 quantized images (1000 QIM-stego using RE, 200 quantized-cover using plain-quantizer). Shown in Table III is detection performance the proposed steganalysis scheme for various embedding rates. These simulation results are generated with decision threshold $\delta m = 40\%$,

$var\{ApEn(x_{high})\} \leq 0.01$ (here $var\{ApEn(x_{high})\}$ denotes variance of estimated $ApEn$ from sequence number 32 to 63) and $ApEn$ estimation parameters $\xi = 4$, $r = 0.2 \times \sigma_x$, and $n = 1024$.

TABLE II
QIM-STEGO DETECTION PERFORMANCE: ***Random Embedding***

	Embedding Rate R				
Error	*0.3*	*0.5*	*0.7*	*0.9*	*1.0*
P_{fp}	0.2	0.2	0.2	0.2	0.2
P_{fn}	0.60	0.5	0.22	0.04	0.003

It can be observed from Table II that false negative rates P_{fn} improves gradually as embedding rate, R, increases. In addition, in case of RE, lower embedding is relatively more secure than the higher embedding rate. It is also worth mentioning that for embedding rate $R < 1$, random embedding is more secure than sequential embedding; consider, for example, MFE and $R = 0.5$ in case of random embedding, false negative rates in case of RE is much higher than then MFE. This is mainly because that in case of RE, detector is not exploiting any knowledge about the either embedding algorithm or characterization of test image which is being exploited for MFE detection.

C. Steganalysis of the DM-Stego

To detect DM-stego based on irregularity in the test-image, the $ApEn(\xi, r, n)$ is calculated from the finite sequences obtained from $\mathbf{s}$, and $\mathbf{x}_{DM}$. The DM-stego was generated by segmenting the cover-image into non-overlapping blocks, each of 8x8 pixels, followed by $2D$ DCT transform. The DM-stego image, $\mathbf{x}_{DM}$, was obtained by embedding a binary message and with $\Delta = 2$, and a dither vector $\mathbf{d}_u \sim \mathcal{U}(0, 2^2/12)$. Shown in Fig. 8 are the plots of the estimated $ApEn$ from the gray-scale cover-image, $\mathbf{s}$, (image number 47 of the database downloaded from [35]) and the corresponding $\mathbf{x}_{DM}$ (in DCT domain) with $ApEn$ parameters, $\xi = 4$, $r = 0.1 \times \sigma_x$, and $n = 1024$.

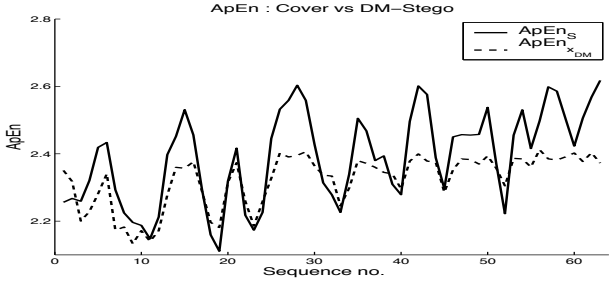

Fig. 8. Plots of the estimated $ApEn$ from $\mathbf{S}$, and $\mathbf{x}_{DM}$ in DCT domain

It can be observed from Fig. 8 that message embedding using DM reduces variance of the estimated $ApEn$, that is, Var$\{\mathbf{ApEn}_s\} >$ Var$\{\mathbf{ApEn}_{x_{DM}}\}$, where Var$\{\mathbf{x}\}$ denotes variance of sequence $\mathbf{x}$. We have observed through simulation results that reduction in the variance of the estimated $ApEn$ from the DM-stego is a function of the cover-image characteristics and quantization step-size used for message embedding. Therefore, variance of the estimated **ApEn** from the test-image cannot be used to distinguish between the cover and the DM-stego, since we consider a blind steganalysis scheme where the steganalyzer has no prior information about the host image or stego parameters. In addition, we have also observed that DM steganography actually increases variance of the DM-stego coefficients. To amplify the difference between the estimated $ApEn_s$ and $ApEn_{x_{DM}}$ from $\mathbf{S}$ and $\mathbf{x}_{DM}$ respectively, we normalized the estimated $ApEn$ vector from the test-image by its variance, i.e., $nApEn_x = ApEn_x/\sigma_x^2$.

The estimated *normalized* $ApEn$, $nApEn$, vector still cannot be used to distinguish between the cover and the DM-stego, as still only one vector is available to the steganalyst to determine whether the test-image is a cove image or a DM-stego. To resolve this issue, a second test-image (say DM^2-stego) is generated by reprocessing the test-image. The reprocessed test-image is obtained by encoding an arbitrary message $\acute{M}$ using DM with an arbitrary dither vector $\acute{\mathbf{d}}_u$ and an arbitrary step-size, $\acute{\Delta}$. It has been observed that the estimated $nApEn$ vectors from the DM$^{(2)}$-stego and the test-image are very close in 63-dimensional space if the test-image is a DM-stego image and are far apart otherwise. To illustrate this claim, we estimated $nApEn$ vectors from $\mathbf{S}$, $\mathbf{x}_{DM}$, and $\mathbf{x}_{DM^{(2)}}$. Shown in Fig. 9 are the plots of the estimated $nApEn$ vectors from $\mathbf{S}$, $\mathbf{x}_{DM}$, and $\mathbf{x}_{DM^{(2)}}$.

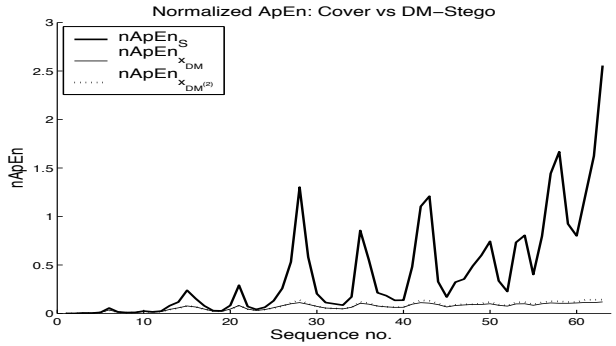

Fig. 9. Shown are the plots of normalized $ApEn$ ($nApEn$) estimated from $\mathbf{S}$, $\mathbf{x}_{DM}$, and $\mathbf{x}_{DM(2)}$.

It can be observed from Fig. 9 that the estimated $nApEn$ vectors from the DM$^{(2)}$-stego and the DM-stego are very close, and the estimated $nApEn$ vectors from the cover and the DM-stego are far apart. This observation reveals that the distance between the estimated $nApEn$ vectors from the test-image and its corresponding reprocessed version (i.e. DM$^{(2)}$-stego) can be used to distinguish between the cover and the DM-stego. A simple binary hypothesis based on the distance between the estimated $nApEn$ vectors estimated from the test-image and DM$^{(2)}$-stego can be used to detect DM-stego.

The proposed steganalysis method to detect DM-stego is summarized as follows:

1) the test-image is reprocessed to obtained DM$^{(2)}$-stego by embedding an arbitrary message, $\acute{M}$, using DM with arbitrary parameters $\acute{\mathbf{d}}_u$ and $\acute{\Delta}$.
2) The $nApEn$ vectors are estimated from both the test-image and the corresponding DM$^{(2)}$-stego.

3) The Euclidian distance, D, between the estimated $nApEn$ vectors from the test-image and the DM$^{(2)}$-stego, defined as,

$$D = \sqrt{\sum_{i=1}^{63} \left(nApEn_i^{(t)} - nApEn_i^{(DM^{(2)})} \right)^2}, \quad (33)$$

is then used to distinguish between the cover and the DM-stego. Here, $nApEn^{(t)}$ and $nApEn^{(DM^{(2)})}$ denote estimated normalized $ApEn$ vectors estimated from the test-image and the corresponding DM$^{(2)}$-stego image, respectively.

Schematic diagram of the proposed steganalysis scheme to distinguish between the cover and the DM-stego is given in Fig. 10.

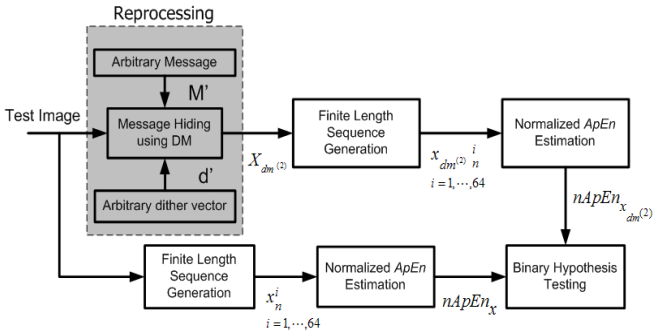

Fig. 10. Schematic diagram of the proposed steganalysis scheme to distinguish between the cover and the DM-stego

D. Experimental Results for DM-Stego

Detection performance of the proposed scheme for DM steganography is also tested for sequential as well as random embedding.

1) Sequential Embedding: Detection performance of the proposed steganalysis scheme to attack DM-stego is also evaluated for the same image database [35] which was used to evaluate performance of the QIM-stego detection. Two-thousand DM-stego images were obtained by sequentially embedding 2000 random messages into the first 1000 images of the database [35] using DM with $\Delta = 2.0$ and an independent and uniformly distributed dither vectors $\mathbf{d_u}$. Here, again these 1000 images were resized to 256x256 and transformed to gray-scale for message embedding. The proposed steganalysis scheme was then applied to the resulting 3000 test-images (2000 DM-stego images and 1000 cover-images in the DCT domain). During the detection process, each test-image was reprocessed to obtain the corresponding DM$^{(2)}$-stego image by embedding an independent message $\acute{M}$ using randomly selected quantization step-size $\acute{\Delta} \in \{1.0, 5.0\}$, and an independent dither vector $\acute{\mathbf{d}}_u$ into all AC coefficients. The $nApEn$ vectors were estimated from each test-image and its corresponding DM$^{(2)}$-stego image using $ApEn$ parameter settings, $\xi = 2$ and $r = 0.1 \times \sigma_x$. Shown in Table I are experimental results for 3000 test-images analyzed using proposed scheme. Simulation results to detect DM-stego listed in Table I are based on quantization step-size $\Delta = 2$ and decision threshold $Th = 2.0$. In addition, in case of MFE, abrupt jump around interfaces of modified coefficients, i.e., finite sequence no. 5 ($\mathbf{x}_n^5$) and finite sequence no. 32 ($\mathbf{x}_n^{32}$) was used for DM-stego detection.

It can be observed that Table I that the proposed non-parametric steganalysis scheme can successfully distinguish between the quantized-cover and the DM-stego images with relatively low false rates, e.g., $P_{fp} < 0.1$, $P_{fn} < 0.07$, and MFE embedding is relatively less secure that the AFE embedding. This is mainly because, detector for MFE is using different detection criterion and is exploiting prior knowledge about embedding algorithm.

2) Random Embedding: To evaluate performance of the proposed steganalysis scheme for *random embedding* case for DM, first 200 images of the (UCID) [35] was used. The selected 200 images of the database [35] were resized to 256x256. One thousand DM-stego images were obtained by embedding 200 random messages using DM method discussed in Section II with $\Delta = 4.0$ and and an independent and uniformly distributed dither vectors $\mathbf{d_u}$. Here, DM-stego images were generated by randomly selecting $R\%$ AC coefficients of the input image for message embedding and the remaining $(1-R)\%$ coefficients remained unaltered. During the detection process, each test-image was reprocessed to obtain the corresponding DM$^{(2)}$-stego image by embedding an independent message $\acute{M}$ using randomly selected quantization step-size $\acute{\Delta} \in \{1.0, 8.0\}$, and an independent dither vector $\acute{\mathbf{d}}_u$. The $nApEn$ vectors were estimated from each test-image and its corresponding DM$^{(2)}$-stego image using $ApEn$ parameter settings, $\xi = 4$ and $r = 0.2 \times \sigma_x$. The proposed steganalysis scheme was tested for 1200 test images (1000 DM-stego using RE and 200 cover images). Shown in Table III is detection performance the proposed steganalysis scheme for various embedding rates i.e. $R \in \{0.3, 0.5, 0.7, 0.9, 1.0\}$. Simulation results shown in Table III are based on quantization step-size $\Delta = 4.0$ and decision threshold $Th = 0.5$.

TABLE III
DM-STEGO DETECTION PERFORMANCE: *Random Embedding*

	Embedding Rate R				
Error	*0.3*	*0.5*	*0.7*	*0.9*	*1.0*
P_{fp}	0.29	0.22	0.18	0.13	0.12
P_{fn}	0.34	0.15	0.010	0.005	0.001

It can be observed from Table III that the proposed scheme false negative rates P_{fn} improves gradually as embedding rate, R, increases. In addition, in case of RE, lower embedding is relatively more secure than the larger embedding. It is also worth mentioning that for embedding rate $R < 1$, random embedding is more secure than sequential embedding.

E. Discussion

Experimental results listed in Table I show that the proposed nonparametric steganalysis scheme can successfully distinguish between the quantized-cover (cover) and the QIM-stego (DM-stego) images with relatively low false rates, e.g., $P_{fp} < 0.12$, $P_{fn} < 0.07$. We also note that the mid-frequency

embedding (MFE) is less secure than the all-frequency embedding (AFE). This is an interesting observation, as a stego-image obtained using MFE carries approximately one-half of message embedded into the stego image obtained using the AFE. Moreover, MFE introduces less embedding distortion than AFE. Simulation results presented in Table I contradict the fact that for a given data hiding scheme, a smaller payload and/or lower embedding distortion provides better security than a larger payload and/or higher embedding distortion.

The explanation of this effect is as follows. In the case of MFE, additional knowledge about embedding algorithm and characterization of the test image (in terms of $ApEn$) was exploited which contributed to superior detection performance. As in case of MFE, only mid-frequency coefficients are modified during message embedding process, therefore, coefficients of the resulting stego-image can be classified into two classes, say C1 and C2. Let coefficients which are modified during message embedding process, e.g., finite sequence no. 5 ($\mathbf{x}_n^5$) to 32 ($\mathbf{x}_n^{32}$), belong to class C1 and the remaining sequences to class C2. Sequences belonging to class C1 exhibit higher level of randomness than the sequences from class C2. Therefore, change in the randomness level from ($\mathbf{x}_n^4$) to ($\mathbf{x}_n^5$) and ($\mathbf{x}_n^{32}$) to ($\mathbf{x}_n^{33}$) can be used to distinguish between the cover and the stego. This observation is illustrated in Fig. 11.

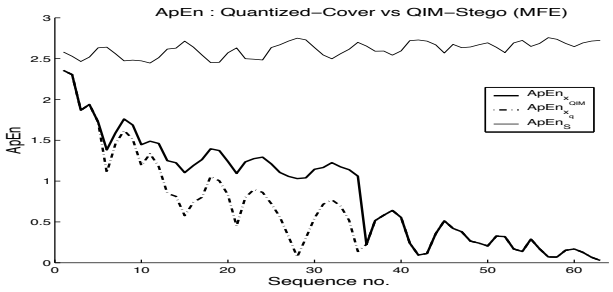

Fig. 11. Plots of the estimated $ApEn$ from the cover, quantize-cover, and the QIM-stego obtained using MFE

It can be observed from Fig. 11 that there is an abrupt change in the estimated $ApEn$ from ($\mathbf{x}_n^{32}$) to ($\mathbf{x}_n^{33}$). Therefore, in case of MFE, an abrupt change in the estimated irregularity in the sequences from C1 and C2 contribute to better detection in case of MFE than AFE. In contrast, when AFE is used there is no abrupt change in the estimated $ApEn$ vector although, there is still enough distinction between the estimated $ApEns$ from the $\mathbf{x}_{QIM}$ and the $\mathbf{x}_q$ to distinguish between the stego and cover images.

Experimental results for random embedding shown in Tables II and III shows that false negative rates P_{fn} for QIM as well as DM improve gradually as embedding rate, R, increases. In addition, in case of RE, lower embedding yields better security (measured in terms of detection rate) than the larger embedding and embedding rate $R < 1$, random embedding is more secure than sequential embedding, for example, MFE and $R = 0.5$ in case of random embedding, false negative rate in case of RE is much higher than then MFE. This is mainly because that in case of RE, detector does not exploit any knowledge about the either the embedding algorithm or about the characterization of test image which is being exploited for MFE detection.

V. Message Recovery

This section will provide details of the proposed active steganalysis framework. This active steganalysis framework is applicable to QIM-stego images only. Once the test-image is identified as a QIM-stego, next step is to estimate the hidden message $\hat{M}$, the secret key $K_{\hat{M}}$, and the hidden message length $L_{\hat{M}}$, from the QIM-stego. The proposed message recovery process consists of two stages: 1) Codebook estimation stage, and 2) Message decoding stage.

The proposed codebook estimation stage uses the first-order statistics of the QIM-stego image to estimate the quantization step-size, $\hat{\Delta}$, used for message embedding. The estimated step-size is then used to decode the hidden message from the QIM-stego. It is important to mention that the detection performance of the message recovery from the QIM-stego directly depends on the accuracy of the estimated Δ.

The proposed message recovery method assumes that the QIM-stego is obtained by embedding a plain-text rather than an encrypted message (or cipher-text) using binary QIM in DCT domain. Moreover, no permutation is applied to the selected image coefficients during the message embedding process. Let the embedding rate, $0 \leq R \leq 1$, represents the fraction of image coefficients used during message embedding. As, binary QIM encodes one bit of information in each processed coefficient, therefore, a gray-scale image of size $N_1 \times N_2$ can carry up to $N = 63 \times \lfloor \frac{N_1}{8} \rfloor \lfloor \frac{N_2}{8} \rfloor$ bits at an embedding rate of 1 bit per pixel (bpp) (assuming DC coefficients are not modified during message embedding process). Shown in Fig.12 is the schematic diagram of the proposed message recovery scheme. The next few sections outline details of the proposed message recovery algorithm.

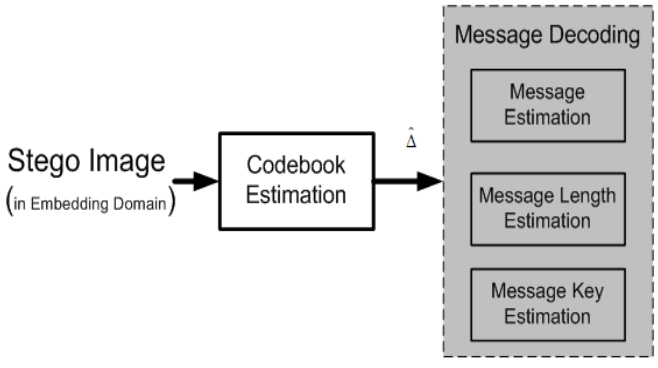

Fig. 12. Schematic diagram of the proposed message recovery scheme

A. Codebook or Δ Estimation

To estimate quantization step-size, Δ, from the QIM-stego, a sequence, x_1^N, is generated by selecting AC coefficients of the QIM-stego obtained by segmenting the QIM-stego into 8x8 non-overlapping blocks and transforming them into the DCT domain. The sequence x_1^N is then analyzed to estimate the

underlying quantization step-size. Salient steps of the proposed Δ–estimation algorithm are outlined below:

Step1: A sorted sequence $s_{x_1}{}^N = sort(x_1^N)$ is obtained from x_1^N by sorting x_1^N in ascending order, i.e., $x_i \leq x_{i+1}, i = \{1, \cdots, N-1\}$

Step2: The first-difference of the sorted sequence, δx_1^N, is calculated as, $\delta x_i = s_{xi+1} - s_{xi}, i = \{1, 2, \cdots N-1\}$

Step3: The following observations can be made on δx_1^N:

1) A run of consecutive zeros in δx_1^N indicate same quantization bin Bn_i (or a reconstruction grid point), $i = \{1, 2, \cdots, N_b\}$ where N_b denotes total number of bins in x_1^N, i.e., $1 \leq N_b \leq N$.
2) These N_b quantization bins, $Bn_i, i = \{1, \cdots, N_b\}$, give N_b distinct reconstruction grid points, i.e., let $Bn_i = k_i\Delta, i = 1, \cdots, N_b$, where $k_i \neq k_j, \forall\ i \neq j$, and $\{k_i,\ k_j\} \in \mathcal{Z}_+$, here $\mathcal{Z}_+$ denote a set of positive integers.
3) The number of coefficients in each bin gives the bin count, Bc_i, of the corresponding quantization bin, that is,
$Bc_i = \sum_{j=1}^{N} \mathbf{1}_{[Bn_i]}(x_j)$,
where $\mathbf{1}$ is the indicator function.
4) The first-difference of the sequence consisting of quantization bin candidates, $\mathbf{Bn}$, yields integer multiples of Δ i.e. $\delta Bn = Bn_{i+1} - Bn_i = t_1\Delta, \forall i$, where $t_1 \in \mathcal{Z}_+$.

Step4: A sequence consisting of candidate values of Δ, D_{list}, is obtained from $\mathbf{Bn}$ and $\delta\mathbf{Bn}$ by sorting them in ascending order and removing repeated entries (if any), i.e., $D_{list} = remove(sort(\mathbf{Bn} : \delta\mathbf{Bn}))$, where $D_{\text{list}(i)} < D_{\text{list}(i+1)}, \forall i$ and $D_{\text{list}(i)} < D_{\text{list}(j)}, \forall i\ \&\ j$.

Step5: A score vector $\mathbf{W}$ based on weighted sum of multiplicity count, m_c, and bin count b_c, of the corresponding entries of D_{list} is calculated,

$$w_i = \alpha_1 \cdot b_{ci} + \alpha_2 \cdot m_{ci} \tag{34}$$

where weighting coefficients α_1 and α_2 are positive real numbers such that $\alpha_1 + \alpha_2 = 1$, here multiplicity count, m_{ci}, and bin count, b_{ci}, values of i^{th} entry in the D_{list} are defined as,

1) multiplicity count, m_{ci}, gives the number of entries in D_{list} that are integer multiples of $D_{\text{list}(i)}, i = \{1, 2, \cdots, 2N_b\}$, and m_{ci} is calculated as,

$$mc_i = \sum_{i=1}^{2N_b} \mathbf{1}_{[Z_+]}(q_i) \tag{35}$$

where q_i is defined as,

$$q_i = \frac{D_{\text{list}(j)}}{D_{\text{list}(i)}}, j = \{i+1, i+2, \cdots, 2N_b - 1\},\ \text{and}$$

2) the bin count, b_{ci} gives the number of coefficients in $\mathbf{x}$ that are integer multiples of $D_{\text{list}(i)}$.

Step6: Entry corresponding to the highest weighted sum score in $\mathbf{W}$ is selected as an estimate of the quantization step-size, $\hat{\Delta}$. For example, if i^* represents the index of the entry in $\mathbf{W}$ with the maximum count, i.e., $w_{i^*} = \max(\mathbf{W})$ then an estimate of Δ is given as $\hat{\Delta} = D_{\text{list}(i^*)}$.

B. Experimental Results for Δ Estimation

In order to evaluate the performance of the proposed Δ–estimation algorithm, we applied the proposed algorithm to 4032 QIM-stego images, 256×256 pixels each. These QIM-stego images were obtained by embedding 438 random messages in the first 64 images of the UCID [35] using binary QIM with quantization step-size $0 < \Delta \leq 2$ in DCT domain. Embedding rates were in the range $0.1 \leq\ R \leq 1$. Average (ave) and standard deviation (std) of the estimated step-size, $\hat{\Delta}$, for different embedding rates along with true Δ used during message embedding are listed in Table IV. Estimated $\hat{\Delta}$ listed in Table IV were obtained using weighting coefficients $\alpha_1 = 0.25$ and $\alpha_2 = 0.75$.

TABLE IV
ESTIMATED STEP-SIZE $\hat{\Delta}$ AT DIFFERENT R

	Embedding Rate R					
	1	**0.8**	**0.6**	**0.5**	**0.3**	**0.1**
True Δ	*Average Estimated Step-Size* $\hat{\Delta}$ (std)					
0.25	0.25(0)	0.25(0)	0.25(0)	0.25(0)	0.25(0)	0.25(0)
0.37	0.37(0)	0.37(0)	0.37(0)	0.37(0)	0.37(0)	0.37(0)
0.5	0.5(0)	0.5(0)	0.5(0)	0.5(0)	0.5(0)	0.5(0)
0.62	0.62(0)	0.62(0)	0.62(0)	0.62(0)	0.62(0)	0.62(0)
0.75	0.75(0)	0.75(0)	0.75(0)	0.75(0)	0.75(0)	0.75(0)
1.0	1.0(0)	1.0(0)	1.0(0)	1.0(0)	1.0(0)	1.0(0)
1.25	1.25(0)	1.25(0)	1.25(0)	1.25(0)	1.25(0)	1.25(0)
1.5	1.5(0)	1.5(0)	1.5(0)	1.5(0)	1.5(0)	1.5(0)
2.0	2.0(0)	2.0(0)	2.0(0)	2.0(0)	2.0(0)	2.0(0)

It can be observed from Table IV that the proposed Δ estimation algorithm has successfully estimated Δ from QIM-stego images carrying messages of different lengths. Simulation results listed in Table IV also reveal that the proposed scheme is insensitive to the quantization step-size, Δ. Moreover, it has also been observed through simulations that the proposed Δ–estimation algorithm occasionally fails to estimate accurate Δ from the QIM-stego images of size 256×256 obtained by encoding message at $R < 0.1$. But, it is observed through extensive simulations that the proposed algorithm always estimates Δ accurately, when applied to QIM-stego-images of size $N_1 \times N_2 \geq 512 \times 512$, obtained using same QIM settings as for the QIM-stego images of size 256×256, but embedding rates as low as 0.05 bpp. This would indicate that to estimate Δ accurately, the QIM-stego should carry at least T quantized coefficients. The value of the constant T depends on the cover-image texture. We noted that for the image database downloaded from [35] T was 6000.

C. Message Decoding

The estimated quantization step-size, $\hat{\Delta}$, is then used to estimate the hidden message, $\hat{M}$, the embedding key $\hat{K}_M$, and message length $\hat{L}_M$. The hidden message length can be estimated by simply calculating the number of stego coefficients which are integer multiples of $\hat{\Delta}$. The hidden message length, $\hat{L}_M$, can be calculated as,

$$L_{\hat{M}} = \sum_{i}^{N} \mathbf{1}_{[k\hat{\Delta},\ k \in \mathcal{Z}_+]}(x_i) \tag{36}$$

Similarly, indices of the stego coefficients corresponding to integer multiples of $\hat{\Delta}$ give an estimate of the $\hat{K}_M$.

Given that the steganalyzer has given no a priori knowledge about the cover-image and the hidden message, determining which quantizer was actually used to map message symbol 1 (or 0) can only be resolved by trial and error. Therefore, there is an uncertainty in deciding whether reconstruction grid points corresponding to odd multiples of Δ was used to encode message symbols '0' or even. To resolve this uncertainty, the proposed scheme decodes two messages, one for each quantizer selection possibility. Let the first estimated message, say $\hat{M}_0$, correspond to that obtained by decoding reconstruction grid points corresponding to odd multiples of Δ as message symbol '0', and the second estimated message, $\hat{M}_1$, for '1'. Here one extracted message, say $\hat{M}_0$, will have decoding bit error rate, $P_e \to 0$ as $R \to 1$, whereas for $\hat{M}_1$, $P_e \to 1$ as $R \to 1$. The choice resulting in a "meaningful" message is declared as the correct choice.

D. Experimental Results for Message Recovery

The proposed message recovery procedure was tested for 600 QIM-stego images each of 256x256 pixels. These QIM-stego images were obtained by embedding 600 random messages in the first 10 images of the UCID database [35] using binary QIM with data embedding rates in the region $0.1 \leq R \leq 1$ and step-size, Δ, equal to 2. The average decoding bit error rate along with its first-standard deviation spread in the estimated message from these 600 QIM-stego images is plotted in Fig. 13.

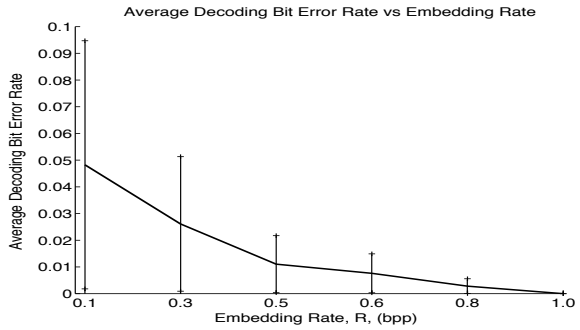

Fig. 13. Average decoding bit error rate as a function of embedding rate, R

It has been observed experimentally that the average decoding bit error rate P_e depends 1) the cover-image statistics, and 2) the embedding rate R.

Experiments show that for a given $0.1 \leq R \leq 1$ and Δ, the QIM-stego images obtained from low-texture images exhibit higher P_e than the QIM-stego images obtained from rich-texture cover-images. The higher decoding error in the low-texture QIM-stego images can be attributed to what we call the *natural binning to zero*, that is, unquantized DCT coefficients are naturally rounded to zero. This is mainly due to the fact that low-texture images exhibit relatively large number of AC coefficients with value close to zero. These *naturally quantized coefficients* contribute to the decoding bit error as the decoder falsely identifies such quantized *coefficients* as message carriers. In addition, simulation results to investigate detection performance as a function of embedding rate revealed that *natural binning to zero* induced decoding error approaches 0 as R approaches 1; this claim is illustrated in Figs. 14 and 15 .

Shown in the top panel of Fig. 14 and 15 are the locations (white colored pixels) used for message embedding. Shown in the central panel are the estimated message locations (white colored pixels) using the proposed message decoding method from the stego-images. Shown in the bottom panel are the errors (white colored pixels) in the estimated message. Messages were embedded in these images using 8x8 non-overlapping blocks in DCT domain. It can be observed from Figs. 14 and 15 that in general decoding bit error probability decreases as message embedding rate increases. This is because as by increasing the number of message carrying coefficients will simultaneously reduce the number of coefficients susceptible to *natural binning to zero*, hence lower decoding error. Moreover, we have also observed that the decoding bit error probability depends on the stego-image texture, that is, low-texture image, e.g. ***Girl*** image, exhibits higher decoding bit error rate than the rich-texture stego-image, e.g. ***Spring*** image. More specifically, plain (or low activity) regions in the stego-image are the major source of decoding bit error.

Finally, to investigate the effect of image statistics on the decoding error due to *natural binning to zero* in the estimated message two images were used, the ***Girl*** image (a low-texture image) and ***Spring*** image (rich-texture image). Shown in the Table V is the decoding error rates for the these images. These results are generated with the embedding rate $R \in \{0.1, 0.3, 0.5\}$ and message embedding parameter $\Delta = 2$.

TABLE V
DECODING ERROR DUE TO *Natural Binning to Zero*

Image	*Embedding Rate R*		
	0.1	**0.3**	**0.5**
Error in the Estimated Message			
Girl	19.0×10^{-3}	5.5×10^{-3}	1.3×10^{-3}
Spring	5.9×10^{-3}	2.4×10^{-3}	0.46×10^{-3}

Error rate in the estimated message listed in Table V shows that *Girl-stego* exhibits higher decoding error compared to *Spring-stego* for all embedding rates. These simulation results indicate that it is easier to steganalyze QIM-stego images that use either one or all of the following: (1) large block sizes, (2) high embedding rate or (3) schemes that do not include zero as one of the quantization grid point for message embedding.

VI. CONCLUSION

This paper presents a novel nonparametric steganalysis scheme to detect QIM-based data hiding. The proposed steganalysis scheme is not learning based therefore capable of addressing limitations of learning-based steganalysis schemes. The proposed scheme uses normalized irregularity in the test-image, as measured by the approximate entropy, to distinguish between the quantized-cover and the QIM-stego images. Experimental results presented show that the proposed steganalysis scheme can successfully distinguish between the cover and the stego with low false rates $P_{fp} < 0.12$ and $P_{fp} < 0.002$ (in case of QIM embedding). In addition, the QIM-stego image

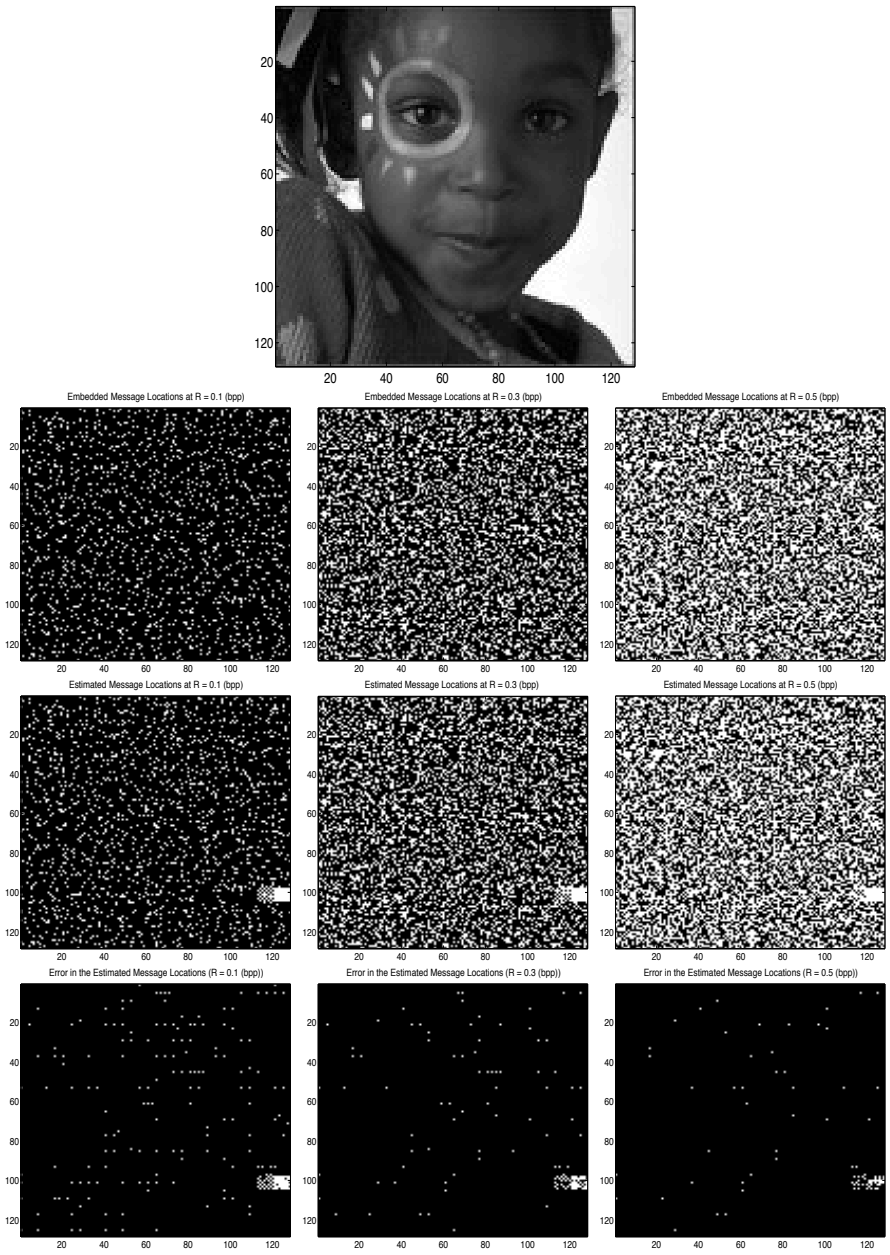

Fig. 14. Error in the estimated message location due to natural binning for different embedding rates: embedded message locations (first row), estimated message locations (second row), and error in the estimated message locations due to natural binning (bottom row)

is analyzed further to estimate quantization step-size which is then used to recover the the location, length and the actual hidden message. Simulation results for message decoding show that the proposed message recovery method discussed in this paper is capable of detecting and decoding the embedded message with very low decoding error probability $P_e < 0.1$. Currently we are investigating the performance of the proposed steganalysis scheme to detect stego images carrying smaller messages embedded using non-sequential embedding.

APPENDIX

Proof of Theorem 2

Proof: Let $\mathbf{s} = \{s_i\}_{i=1}^{N}$ be a real valued random sequence to be quantized. Let P_s denotes its probability density function (*pdf*). A quantizer $Q_{N_0}(s)$ is defined as a partition $\Lambda = \Delta_k = [t_k, t_{k+1}), t_{k+1} > t_k, k = \{1, \cdots, N_0\}$ where N_0 is the number of partitions, and a reconstruction codebook $\hat{x}_k$ as $Q(s) = \hat{x}_k$ if $s \in \Delta_k$. Let us assume, without loss of generality, that $\hat{x}_k$ are distinct and the corresponding *pmf* of indexed quantizer output points is denoted by $p_k = Pr\{Q(s) = \hat{x_k}\} = Pr(s \in \Delta_k)$.

Let us calculate randomness of $Q_{N_0}(\mathbf{s}) = \hat{\mathbf{x}}_{N_0} = \{\hat{x}_k\}_{k=1}^{N_0}$, using Shannon's entropy, H, as,

$$H_{N_0} \triangleq H\left(Q(\mathbf{s})\right) = -\sum_{k=1}^{N_0} p_k \log(p_k) \tag{37}$$

Now obtain the (N_0+1)-partition quantizer, $Q_{N_0+1}(\cdot)$, by dividing one partition in $Q_{N_0}(\cdot)$, say Δ_j, into two partitions

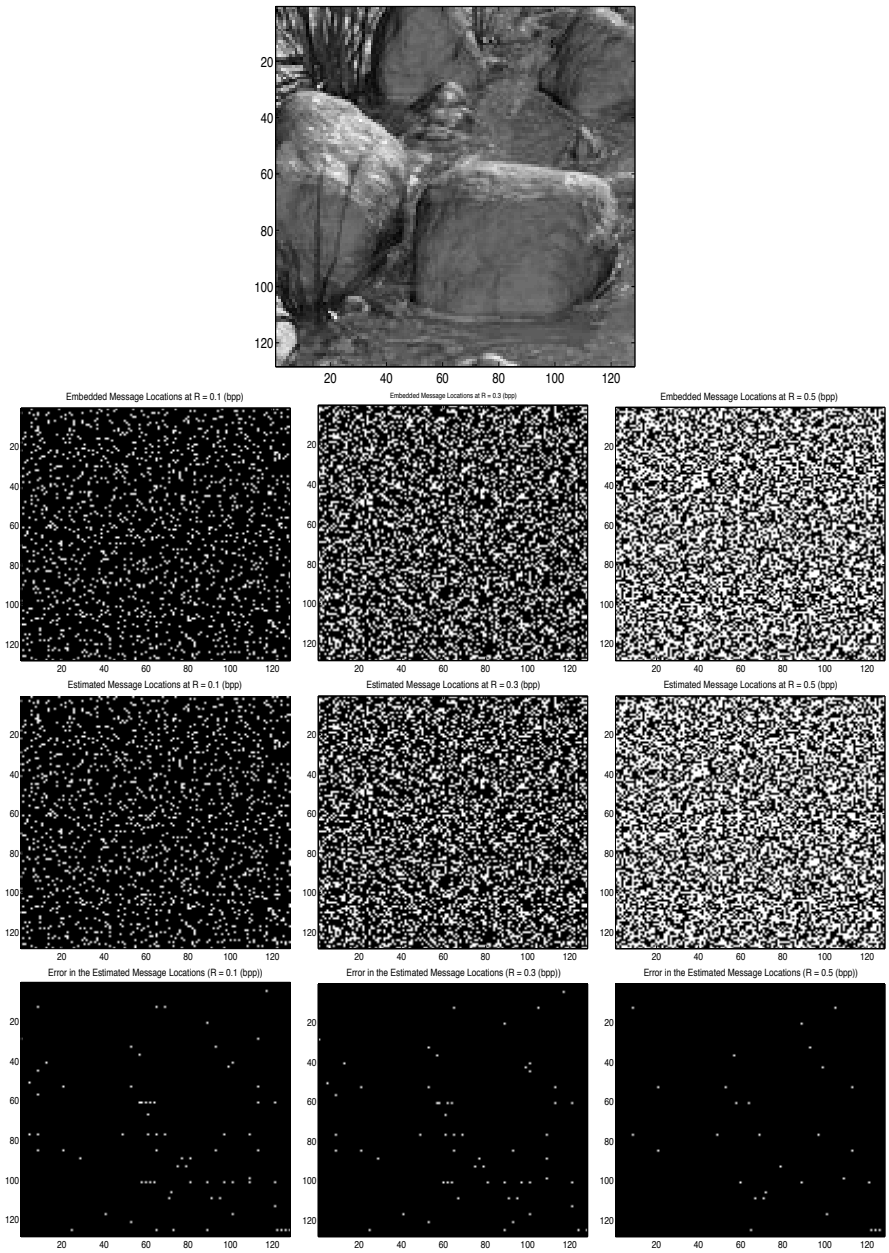

Fig. 15. Error in the estimated message location due to natural binning for different embedding rates: embedded message locations (first row), estimated message locations (second row), and error in the estimated message locations due to natural binning (bottom row)

Δ_{j1} and Δ_{j2}. Let p_α and p_β be the probabilities of the quantizer output points Δ_{j1} and Δ_{j2}, respectively, where,

$$p_\alpha + p_\beta = p_j \tag{38}$$

Shannon's entropy of the quantized signal obtained using an (N_0+1)-partition quantizer, i.e. $Q_{N_0+1}(\mathbf{s}) = \hat{\mathbf{x}}_{N_0} = \{\hat{x}_k\}_{k=1}^{N_0+1}$, can be expressed as,

$$H_{N_0+1} = H_{N_0} + p_j \log(p_j) - p_\alpha \log p_\alpha - p_\beta \log p_\beta \tag{39}$$

Let $\mu \triangleq \frac{p_\alpha}{p_j}, 0 \leq \mu < 1$, and $\bar{\mu} \triangleq 1-\mu$. The last three terms on the right hand side (RHS) of Eq.(39) can be expressed as,

$$\begin{aligned} f(\mu) &= p_j \log(p_j) - \mu p_j \log \mu p_j \qquad (40)\\ &\quad -\bar{\mu} p_j \log \bar{\mu} p_j \\ &= p_j \log(p_j) - \mu p_j [\log \mu + \log p_j] \qquad (41)\\ &\quad -\bar{\mu} p_j [\log \bar{\mu} + \log p_j] \\ &= -\left(\mu \log \mu + (1-\mu)\log(1-\mu)\right) p_j \\ &= H(\mu) p_j \qquad (42) \end{aligned}$$

where, $H(\mu) \triangleq -\left(\mu \log \mu + (1-\mu)\log(1-\mu)\right)$. Therefore, H_{N_0+1} can be expressed as,

$$\begin{aligned} H_{N_0+1} &= H_{N_0} + f(\mu) \qquad (43)\\ H_{N_0+1} &= H_{N_0} + H(\mu) p_j \qquad (44) \end{aligned}$$

As $f(\mu) \geq 0$,

$$H_{N_0+1} \geq H_{N_0} \Rightarrow H_{2\times N_0} \geq H_{N_0} \tag{45}$$

And, as $N_0 \rightarrow \infty$ we obtain an unquantized random sequence. ■

REFERENCES

[1] R. Chandramouli, "A mathematical framework for active steganalysis," *ACM Multimedia Systems*, vol. 9, no. 3, pp. 303–311, September 2003.

[2] B Chen and G Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Information Theory*, vol. 47, no. 4, May 2001.

[3] M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

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

[5] P. Guillon, T. Furon, and P. Duhamel, "Applied public-key steganography," in *Proc. IS&T/SPIE*, 2002, pp. 38–49.

[6] 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)*, 2004, vol. 2, pp. 1165–1168.

[7] R. Chandramouli and K. Subbalakshmi, "Current trends in steganalysis: A critical survey," in *IEEE Int. Conf. on Control, Automation, Robotics and Vision (ICARCV)*, December 2004, vol. 2, pp. 964–967.

[8] H. Malik, K. Subbalakshmi, and R. Chandramouli, "Nonparametric steganalysis of qim-based data hiding using kernel density estimation," Dallas, Texas, USA, September 2007, ACM, 9th Workshop on Multimedia & Security (MM&Sec 2007).

[9] H. Malik, K. Subbalakshmi, and R. Chandramouli, "Nonparametric steganalysis of qim data hiding using approximate entropy," San Jose, CA, USA, January 2008, IS&T/SPIE, vol. 6819 of *Security, Steganography, and Watermarking of Multimedia Content X*.

[10] Luis Perez-Freire, Pedro Comesana-Alfaro, and Fernando Perez-Gonzalez, "Detection in quantization-based watermarking: performance and security issues," in *Security, Steganography, and Watermarking of Multimedia Contents VII*, Edward J. Delp III; Ping W. Wong, Ed., 2005, vol. 5681, pp. 721–733.

[11] Tomas Pevny and Jessica Fridrich, "Detection of double-compression in jpeg images for applications in steganography," *IEEE Trans. on Info. Forensics and Security*, vol. 3, no. 2, pp. 247–258, 2008.

[12] Xiao-Yi Yu and Aiming Wang, "Detection of quantization data hiding," in *Int. Conf. on Multimedia Information Networking and Security (MINES '09)*, December 2009, pp. 45–47.

[13] Qinxia Wu, Weiping Li, and Xiao Yi Yu, "Revisit steganalysis on qim-based data hiding," in *Fifth Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP'09)*, 2009, pp. 929–932.

[14] Siho Kim and Keunsung Bae, "Estimation of quantization step size against amplitude modification attack in scalar quantization-based audio watermarking," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP'06)*, May 2006, vol. V, pp. 389 – 392.

[15] Taejeong Kim Kiryung Lee, Dong Sik Kim and Kyung Ae Moon, "Em estimation of scale factor for quantization-based audio watermarking," in *Digital Watermarking*, 2004, vol. 2939 of *Lecture Notes in Computer Science*, pp. 316 – 327.

[16] Tomas Pevny and Jessica Fridrich, "Estimation of primary quantization matrix for steganalysis of double-compressed jpeg images," in *Proc. SPIE, Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, 2008.

[17] C. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423 & 623–656, 1948.

[18] A. Kolmogorov, "A new metric invariant of transitive automorphisms of lebesgue spaces," *Dokl. Akad. Nauk*, vol. SSSR119, no. 5, pp. 861–864, 1958.

[19] Y. Sinai, "On the concept of entropy for a dynamical system," *Dokl. Akad. Nauk*, vol. SSSR124, pp. 768–771, 1959.

[20] A. Lempel and J. Ziv, "On the complexity of finite sequences," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 75–81, 1976.

[21] S. Pincus, "Approximate entropy as a measure of system complexity," *Proc. Natl. Acad. Sci. USA*, vol. 88, pp. 2297–2301, March 1991.

[22] S. Pincus and A. Goldberger, "Physiological time-series analysis: What does regularity quantify?," *Am. J. Physiol (Heart Circ Physiol)*, vol. 266, no. 4 Pt 2.

[23] S. Pincus, "Approximate entropy as a complexity measure," *CHAOS*, vol. 5, no. 1, pp. 110–117, 1995.

[24] L. Schuchman, "Dither signals and their effect on quantization noise," *IEEE Trans. Commun. Technol.*, vol. COM-12, pp. 162165, December 1964.

[25] R.A. Wannamaker, *The Mathematical Theory of Dithered Quantization*, Ph.D. thesis, Dept. of Applied Mathematics, Univ. of Waterloo, Waterloo, ON, Canada, June 1997.

[26] H. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, Berlin, Germany, 2^{nd} edition, 1994.

[27] T. Covet and J. Thomas, *Elements of information theory*, Wiley-Interscience, New York, NY, USA, 1991.

[28] D. Ornstien and B Weiss, "How sampling reveals a process," *Ann. of Prob.*, vol. 18, pp. 905–930, 1990.

[29] S. Pincus, "Approximating markov chains," *Proc. Natl. Acad. Sci. USA*, vol. 89, pp. 4432–4436, 1992.

[30] S. Pincus and R. Kalman, "Not all (possibly) "random" sequences are created equal," *Proceedings of the National Academy of Science USA*, vol. 94, pp. 3513–3518, 1997.

[31] S. Pincus and L. Goldberger, "Physiological time-series analysis: What does regularity quantify," *American Physiological Society*, vol. (Modeling in Physiology), pp. H1648–H1656, 1994.

[32] S. Pincus and W. Huang, "Approximate entropy: Statistical properties and applications," *Communications in Statistics: Theory and Methods*, vol. 21, no. 11, pp. 3061–3077, 1992.

[33] S. Pincus and B. Singer, "Randomness and degree of irregularity," *Proceedings of the National Academy of Science USA*, vol. 93, pp. 2083–2088, 1996.

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

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