Steganalysis of QIM Steganography

Hafiz Malik†, *Member, IEEE,* K. P. Subbalakshmi* *Senior Member, IEEE,* and R. Chandramouli* *Senior Member, IEEE*

† 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 a statistical steganalysis method for quantization index modulation (QIM) based steganography. We have shown that, in general, plain-quantization (quantization without message embedding) reduces local-randomness (or increases local-correlation) in the resulting quantized-object and QIM-stego exhibits higher level of local-randomness than the corresponding quantized cover. The local-randomness of the test-image is used to capture traces left behind by quantization (with or without message embedding). We model the distortion due to quantization as a *gamma distribution*. The parameters of this gamma distribution are estimated using maximum likelihood estimators. Distributions of the parameters estimated from the quantized-cover and the QIM-stego images are used to develop a generalized likelihood ratio test (GLRT) to distinguish between the cover and the stego images. Effectiveness of the proposed method is evaluated using a large set (over 35000 images) consisting of test-images obtained using sequential as well as random message embedding. Experimental results show that the proposed scheme can successfully detect QIM-stego images with very low false rates ($P_{fp} < 0.015$ and $P_{fn} < 0.03$) for sequential embedding and false rates ($P_{fn} < 0.038$ and $P_{fp} < 0.013$) for random embedding. In addition, performance comparison with existing state of the art also shows that the proposed method performs significantly superior than the selected methods.**

***Index Terms*—Steganography, Steganalysis, Quantization Index Modulation, Maximum Likelihood Estimation, Generalized Likelihood Ratio Test (GLRT)**

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. Likewise, based on the detection framework used for steganalysis can also be categorized into two main groups: 1) statistical learning based steganalysis, and 2) model based steganalysis.

Statistical learning based steganalysis techniques make no assumption about the statistical properties of the stego-signal. Instead statistics of some features are learnt using a large training data set of cover and stego signals. Some metric that is a function of the feature statistics is used to discriminate between cover and stego. Therefore, statistical learning based steganalysis techniques [2]–[14] consist of two major stages: 1) feature generation and feature selection stage to extract feature vector(s) from the training data set based on some feature selection criteria, and 2) classification stage that uses the extracted feature vectors from test object and a trained classifier for cover-stego discrimination.

Statistical learning based steganalysis techniques generally use supervised learning to train a statistical classifier. First, a k - dimensional feature vector is estimated from the training data set during learning/training phase. Existing statistical learning based methods [2]–[14] select this k-dimensional feature vector either heuristically or using feature selection criterion [15]. Statistical learning based methods rely on statistical features such as image quality metrics [3], histogram [10], the center of mass of the histogram characteristic function [16], characteristic function (CF) moments [9], probability density function (PDF) moments [2], co-occurrence matrix (CM) [8], Markov transition probability matrix (MTPM) [7], and etc. The classifier then learns the best classification rule using the input feature vectors for each steganographic method. These techniques use sophisticated machine learning tools such as linear regressions analysis based on Fisher linear discriminant (FLD), support vector machines (SVM), principal component analysis (PCA), artificial neural networks (ANN), [17] etc. for the classification stage.

Model based steganalysis methods [16], [18]–[26], on the other hand, assume a suitable statistical non-parametric or parametric model for the cover- and the stego-object. Detection statistic is then derived using assumed model. Statistical analysis based on first- and higher-order statistics of the test-image (i.e. mean, variance, skewness, etc.), Chi-square (χ^2) tests, etc. are commonly used to distinguish between the stego- and the cover-object.

Quantization based data hiding schemes [27] are based on

Send correspondence to Hafiz Malik at E-mail: hafiz@umich.edu, Tel.: 1 313 593 5677

Send correspondence to K.P. Subbalakshami at ksubbala@stevens.edu

Send correspondence to R. Chandramouli at mouli@stevens.edu

Costa's seminal work [28] 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 [29]. Practical realizations of ICS include quantization index modulation (QIM) [27], scalar Costa scheme (SCS), dither modulation (DM), and quantization projection (QP), [29]. 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. [30] proposed a framework for steganalysis of SCS by modeling QIM steganography as an additive noise channel. Sullivan et al. [5] proposed a steganalysis scheme for QIM steganography using supervised learning. Zou et al. [6] proposed a supervised learning based steganalysis method which uses 2-dimensional Markov chain of thresholded prediction-error image to capture traces of message embedding. The proposed scheme [6] uses local neighborhood of the current pixel to predict it, and the prediction-error image is generated by subtracting the prediction value from the actual pixel value and then thresholded with a predefined threshold. The feature vector consisting of three empirical transition matrixes of Markov chain along the horizontal, vertical, and diagonal directions is used to train a binary support vector machines (SVM)-based classifier. Similarly, Pevny and Fridrich [13], [31] have also proposed a method to detect of double JPEG compression and a maximum likelihood estimator of the primary quality factor. Detection performance of these statistical learning based methods [5], [6], [13], [31] are 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 [32].

Model based QIM steganalysis techniques [22]–[26], [33]–[37] have been proposed to address limitations of statistical learning based methods. For example, Kim and Bae [37] and Lee at al. [36] have proposed analytical approach using first- and second-order statistical features (i.e., mean and variance) for quantization step size estimation from QIM-stego audio signal subjected to scaling and additive white Gaussian noise attacks. Similarly, Yu and Wang [34] and Wu et al. [35] have also proposed model-based steganalysis methods to estimate secret message length estimation form QIM stego by mathematically modeling QIM embedding distortion as a function of embedding ratio (or secret message length) and used estimated model parameters for secret message length estimation. Low detection performance is one of the limitations of existing QIM-steganalysis schemes [22]–[26], [34]–[37]. For example, QIM steganalysis scheme proposed in [22] degrades significantly for smaller value of quantization step-size (e.g., $\Delta \leq 1$). In addition, relatively high false positive rates is another limitation of this scheme. Similarly, the steganalysis scheme proposed in [26] cannot detect random partial embedding. Major contribution of this paper is to address limitations of existing model-based QIM steganalysis schemes [22]–[26].

The proposed steganalysis scheme uses local-randomness level to capture traces of quantization. The local-randomness mask is estimated by segmenting the test-image into non-overlapping blocks followed by transforming each block into message embedding domain (e.g. DCT domain) then local-randomness mask estimation for each subband using similarity with its neighborhood. The distribution of the local-randomness level, estimated from the test-image, is used to parameterized the quantization process. It has been observed that distribution of local-randomness level depends on two factors: 1) the quantization type (e.g., with or without message embedding) and 2) the cover image texture characteristics (e.g., low or high texture). We have demonstrated that image texture level (low to high texture) also contributes to the underlying local-randomness level. To investigate estimated parameter variations as a function of frequency, we analyzed parameter variations based on both the intra-subband variations, that is, within each frequency subband, and inter-subband variations, that is, across all frequency subbands.

For intra-subband analysis, the underlying density function that captures traces of natural image variations and quantization process is estimated for each frequency subband. For inter-subband analysis, first four moments, $\mu_k, k = 1, \cdots, 4$, are estimated from parameter estimates for all frequency subbands. The distribution of estimated moments are used to estimate the underlying density. The estimated underlying density is used as a *reference density* for a given set, i.e., stego or cover. The *generalized likelihood ratio tests* (GLRT) are used to determine whether the test-image is a quantized-cover or a QIM-stego. The stego-detection decisions from both detectors are fused to obtained the final decision.

Effectiveness of the proposed method is evaluated on a large data set consisting of more than 35000 test-images obtained using plain-quantization, sequential embedding and random embedding. For random embedding, the stego-images are obtained for various embedding rate, $R \in \{10 - 100\}\%$ (i.e., $R\%$ of the coefficients are modified during message embedding process). 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 rates, i.e., positive rates, P_{fp}, and false negative rates, P_{fn}. In particular, the false positives rates are below 0.015 and the false negative rates are below 0.03 for QIM-stego detection. In addition, performance of the proposed method is also compared with the existing state of the art [6], [22]. Performance comparison indicates that the proposed method performs significantly superior than the selected methods [6], [22].

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 require-

ments of QIM-steganalysis and information-theoretic analysis of quantized images are provided in Section II. Details of the local-randomness mask estimation are provided in Section III. Statistical modeling of the local-randomness mask, parameter estimation, and parameter distribution modeling are provided in Section IV. The outline of the proposed steganalysis scheme is provided in Section V. Experimental settings, performance evaluation and comparison are provided in Section VI. Concluding remarks and future directions are discussed in Section VII.

II. Steganalysis of QIM-Steganography

The object of QIM steganalysis is to distinguish between 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).

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

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.

Whereas, 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.

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 $\Delta_k = k\Delta$.

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

$$L(x) \quad \triangleq \frac{P_{\mathrm{x}_{QIM}}(x)}{P_{\mathrm{x}_q}(x)} \gtrless \eta \quad \text{(detect QIM-stego)} \tag{3}$$

where the decision threshold, η, can be minimized using Neyman-Pearson rule which maximizes the probability of detection, P_d, for a given probability of false alarm, P_f [38].

Substituting $P_{\mathrm{x}_{QIM}}(x)$ and $P_{\mathrm{x}_q}(x)$ from Eq. (11 & 2) in Eq. (4):

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

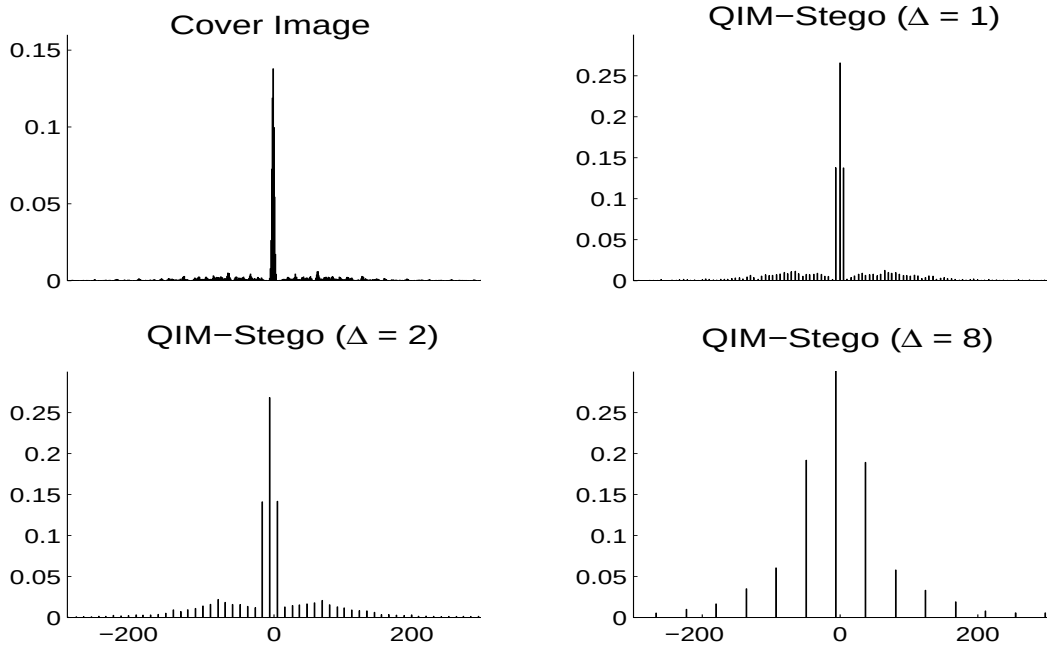
Here, Eq. (4) shows that the likelihood statistic is a function of the cover *pdf*, $P_{\mathrm{s}}(s)$, and under stego-only attack scenario $P_{\mathrm{s}}(s)$ is not available at the stego detector. Therefore, parametric detection based on Neyman-Pearson rule cannot be used directly to detect the QIM-stego image.

An important observation however can be made from Eq. (4) (we have also shown in [25]) that message embedding using QIM introduces smoothness in the *pmf* of the resulting stego-image. To highlight this claim further, we analyzed 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. 1a and shown in Fig. 1b is the comparison of smoothing effect due plain-quantization and QIM.

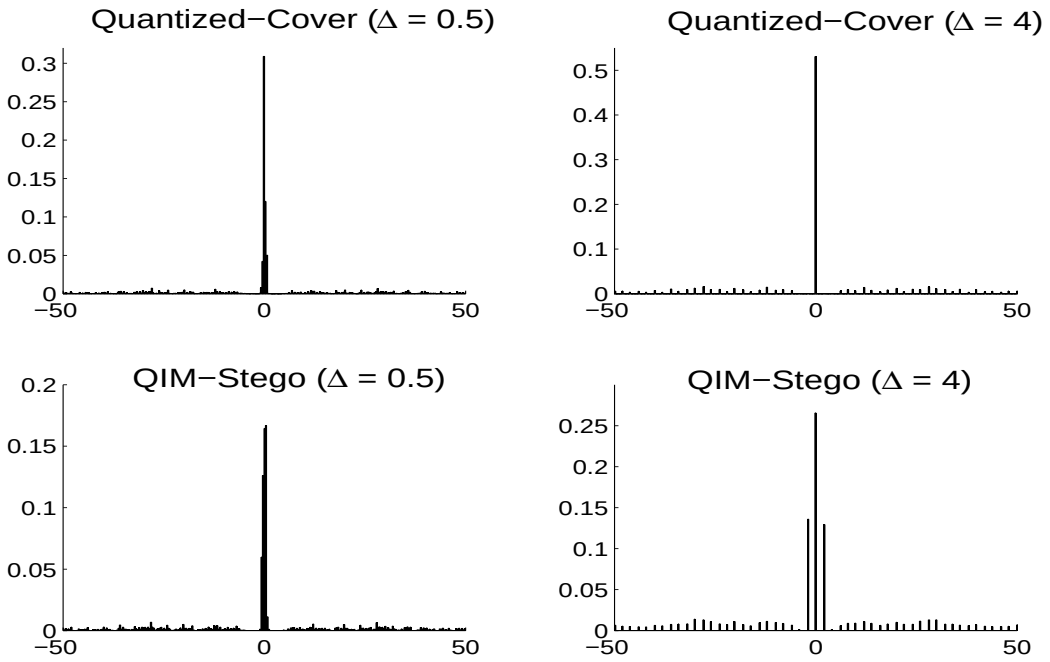
Some of the experimental observations on the difference between the QIM-stego and the quantized-cover images based on their empirical *pmf*s are summarized below.

Firstly, we note that the quantization (with and without message embedding) introduces smoothness in the *pmf* of the resulting quantized images. It can be observed from Fig. 1a that as Δ increases smoothing effect in the *pmf* of the resulting QIM-stego also increases according to the Eq. (4). Secondly, 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. 1b 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 [5] 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 existing steganalysis schemes for QIM steganography [5], [6], [11], [12], [22]–[26], [33]–[37] a parametric steganalysis scheme based on statistical modeling of the local-randomness in the test-image is proposed here. The proposed scheme exploits the local-randomness in the test-image to distinguish between the cover- and the stego-images. More specifically, the proposed scheme exploits the fact that the plain-quantization induces correlation (or reduces randomness) in the resulting quantized-object and message embedding using QIM increases exhibits higher level of correlation (or randomness) that the corresponding



(a) Shown are the empirical *pmf*s of DCT coefficients of the cover (top-left) and quantized DCT coefficients of the QIM-stego obtained with $\Delta = \{0.5, 4, 8\}$(top-right and the bottom-row) [25].



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

Fig. 1: Quantization effect on the distribution of the DCT coefficients

quantized-cover [23]–[25], [39]. To illustrated this fact two quantized images (a quantized-cover and a QIM-stego), $\mathbf{x}_q$ and $\mathbf{x}_{QIM}$ are obtained using same quantization step-size $\Delta = 2$. The randomness masks, $\mathbf{r}_{(q)}$ and $\mathbf{r}_{QIM}$, are estimated from $\mathbf{x}_q$ and $\mathbf{x}_{QIM}$ using local-randomness estimation method discuss in Section III-A. The underlying densities are then estimated for $\mathbf{r}_{(q)}$ and $\mathbf{r}_{QIM}$ using *kernel density estimation* (KDE) method. Shown in Fig. 2a are the plots of the estimated densities for the quantized-cover and the corresponding QIM-stego using the same set of density estimation parameters (i.e., kernel type, $K(x)$, and bandwidth, h).

It can be observed from Fig. 2a that the estimated density for the quantized-cover image , $f_{\mathbf{x}_q}$, exhibits a peak (or mode) around zero which implies that majority of the DCT coefficients of the quantized-cover are highly correlated or exhibit lower level of randomness. Moreover, the estimated density from $\mathbf{x}_q$ exhibits higher skewness than the density estimated from $\mathbf{x}_{QIM}$. Whereas, the estimated density from QIM-stego, $f_{\mathbf{x}_{QIM}}$ exhibits a peak near 0.3 which indicates that the stego coefficients exhibit relatively a higher level of randomness than the quantized-cover. This increase in the randomness in the QIM-stego can be attributed to the randomness in the hidden message. Moreover, the density estimated from $\mathbf{x}_{QIM}$ exhibits lower skewness. These observations confirm the claim that distortion due to message embedding using QIM is relatively less correlated than the distortion due to plain-quantization. This also implies that coefficients of the quantized-cover image are relatively more predictable than the corresponding coefficients in the QIM-stego image. The proposed steganalysis scheme uses relative randomness in the test-image to distinguish between the quantized-cover and the QIM-stego images.

III. Proposed Steganalysis Model

The objective of nonparametric steganalysis system is to detect stego-object reliably without learning characteristics of the stego-objects or using parameters of the underlying steganographic algorithm used for message embedding. The nonparametric steganalysis is done by first estimating feature vector, from the test-object, that captures traces of the message embedding process and then applying binary hypothesis test controlled by a predefined threshold. Performance of steganalysis systems generally is measured in terms of the probability of false-positive, P_{fp}, and the probability of false-negative, P_{fn}. The proposed steganalysis system operates under an additional constraint that only stego-object is available to the steganalyst to a make decision.

The proposed steganalysis scheme exploits the fact that message embedding using QIM introduces randomness in the resulting stego-object. The distribution of local-randomness level in a given image is used to parameterized the quantization process, i.e., quantization with and without message embedding. It has been observed that distribution of local-randomness level depends on the both the quantization type and cover image characteristics. More specifically, image texture level (low to high texture) also contributes to the underlying local-randomness level. To investigate density parameter variations as a function of frequency, intra-subband variations as well as inter-subband variations analysis is performed.

For intra-subband analysis, the underlying density function that captures traces of natural image variations and quantization process is estimated for each frequency subband (or bin). For inter-subband analysis, for each image, first four moments, $\mu_k, k = 1, \cdots, 4$, are computed for parameter estimates for all frequency bins. The distribution of the estimated moments (across the dataset considered) are used to estimate the underlying density to capture traces of natural image variations and quantization process.

The estimated underlying density is used as a reference density for a given set, i.e., stego or cover. The Generalized likelihood ratio tests (GLRT) are used to determine whether the test-image is a quantized cover or QIM-stego. The stego-detection decisions of both detectors are fused to obtained the final decision, D. The block diagram of the proposed steganalysis scheme is given in Fig. 3.

A. Local-Randomness Mask Estimation

The local-randomness mask, $\mathbf{r}$, from the test-image is estimated as,
First, the block-DCT of the test-image is calculated using the following 2-dimensional (2D) forward discrete cosine function,

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

where

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

The local-randomness mask value of coefficient x_{k_1,k_2} in the j^{th} block (e.g. $x^{(j)}_{k_1,k_2}$), $r^{(j)}_{k_1,k_2}$, is then calculated based on local 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 corresponding coefficients of $x^{(j)}_{k_1,k_2}$ (an an arbitrary coefficient in the j^{th} block) in k-neighboring blocks. The similarity value, $C^{(j)}_{k_1,k_2}$, for coefficient $x^{(j)}_{k_1,k_2}$ is calculated as,

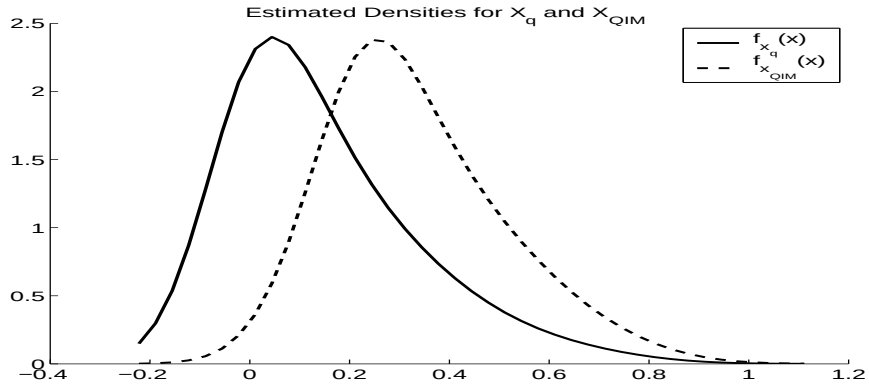
$$C^{(j)}_{k_1,k_2} = \frac{1}{k} \sum_{l=1}^{k} \mathbf{1}_{[x^{(j)}_{k_1,k_2}]} \left(x^{(j)}_{NH(k_1,k_2,l)}\right) \quad l = 1, \cdots, k, \quad \text{and} \quad j = 1, \cdots, n \tag{5}$$

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

$$r^{(j)}_{k_1,k_2} \quad = \quad 1 - C^{(j)}_{k_1,k_2} \tag{6}$$

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 $r^{(j)}_{k_1,k_2}$ is a nonnegative real valued random variable, and $0 \leq r^{(j)}_{k_1,k_2} \leq 1$. It is important to mention that $r^{(j)}_{k_1,k_2} = 0$ implies maximum similarity between the neighbors and the



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

(b) Girl image

Fig. 2: Estimated density plots

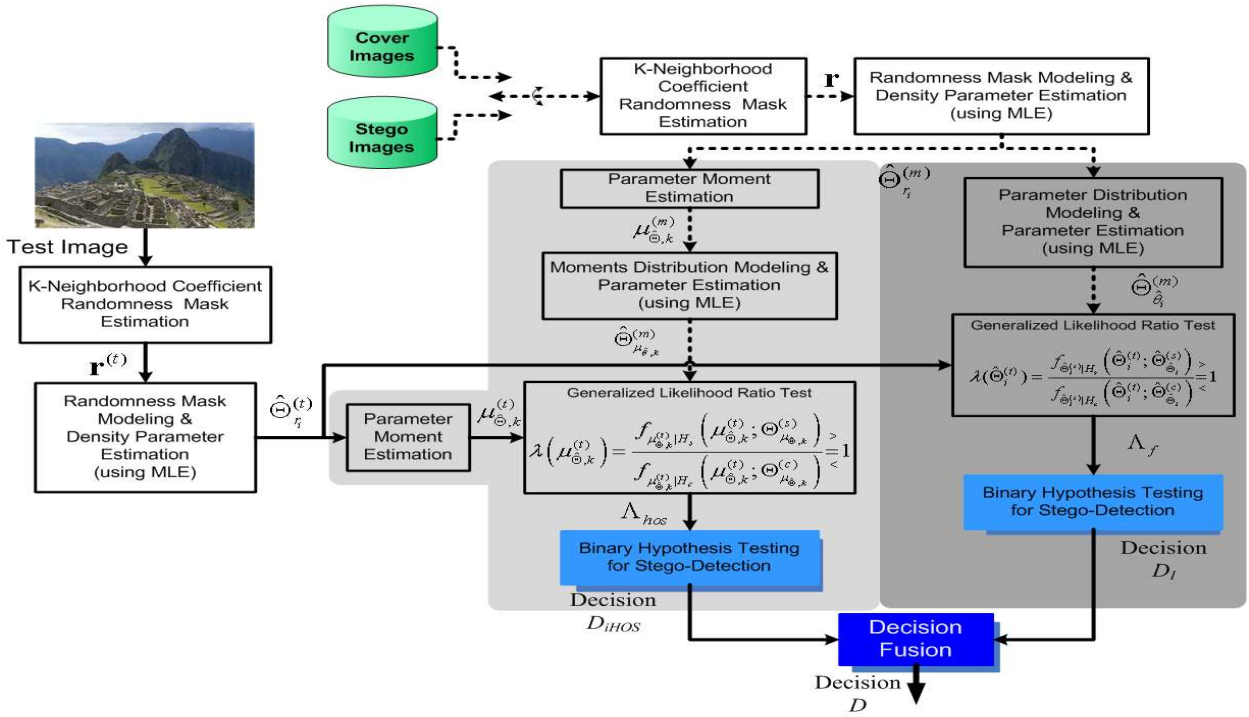


Fig. 3: Schematic diagram of the steganalysis scheme used to attack QIM steganography

current value, $x^{(j)}_{k_1,k_2}$. In this case, all the neighboring coefficients are quantized to the same value. Similarly, $r^{(j)}_{k_1,k_2} = 1$ 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. (6); the randomness mask estimation for the selected block (or block of interest (BOI)) using 8-neighborhood is given in Fig. 4.

Estimated randomness mask, $\mathbf{r}_{k_1,k_2}$, is then mapped to 64 one-dimensional (1D) sequences. The underlying density of the local randomness mask, $f_{\mathrm{r}}(r)$, is then estimated for 64 1D sequences. The mapping of two-dimensional randomness mask, $\mathbf{r}$, to 64 1D sequences, $\mathbf{r}_i, i = 0, 1, \cdots, 63$, is illustrated in Fig. 5.

IV. Modeling Local Randomness Mask

It has been observed through extensive experimentation that the estimated local-randomness mask for a given frequency band i, $\mathbf{r}_i$, where $i = 1, \cdots, 64$ can be modeled using the *gamma distribution*. The gamma distribution can be parameterized in terms of a shape parameter γ and scale parameter β, where $\gamma > 0$ and scale parameter $\beta > 0$. The *probability density function* (pdf) of a Gamma-distributed random variable $r_{i,j}$ can be expressed as,

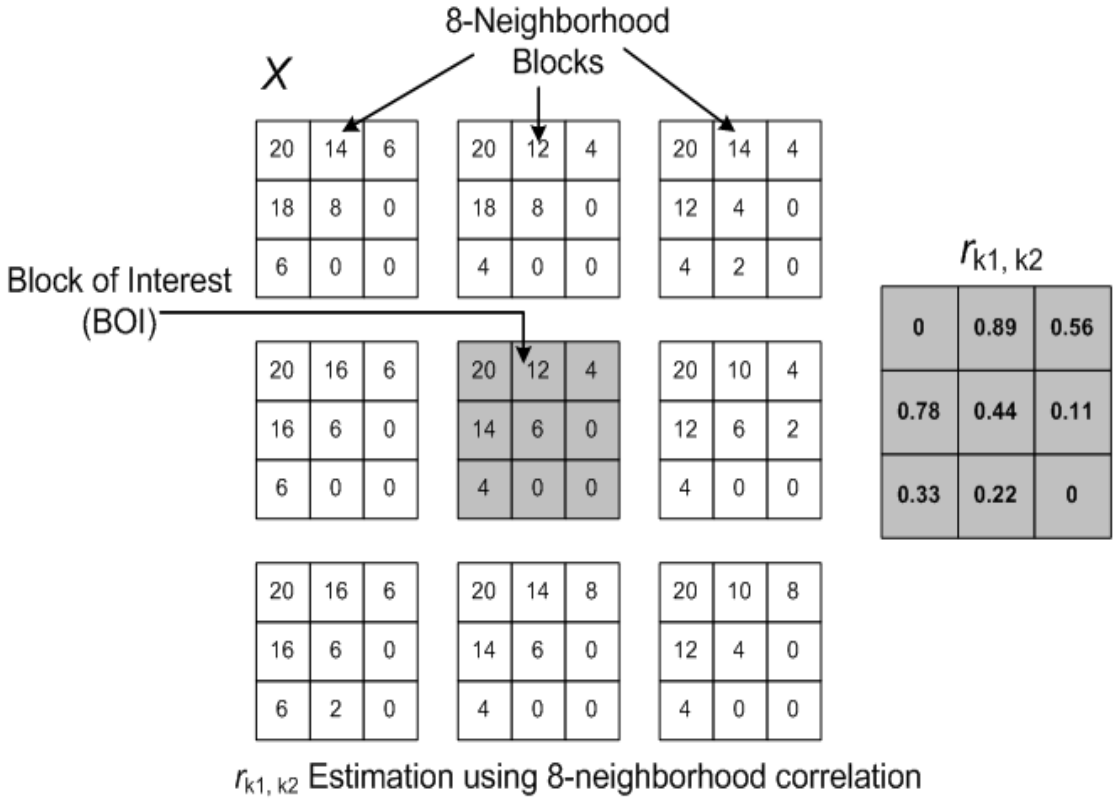


Fig. 4: Local-randomness mask estimation for the selected block based on 8-neighborhood

$$f_{r_{i,j}}(\mathbf{r};\beta_i,\gamma_i)=\frac{r_{i,j}^{\gamma_i-1}}{\beta_i^{\gamma_i}\Gamma(\gamma_i)}e^{-\frac{r_{i,j}}{\beta_i}}\text{for } r_{i,j}>0\text{and}\beta_i,\ \gamma_i>0 \tag{7}$$

Goodness of fit of the assumed model for local-randomness fit, $\mathbf{r}_i$, is evaluated by comparing the theoretical pdf and the *cumulative distribution function* (cdf) of the assumed gamma distribution with estimated parameters β_i and γ_i and empirical pdf and cdf estimated from $\mathbf{r}_i$. Shown in the top panel of Fig. 6a are plots of theoretical (doted line) and empirical (thick line) pdfs and shown in the bottom panel of Fig. 6a are the theoretical (doted line) and empirical (thick line) cdfs of the $\mathbf{r}_{18}$ estimated from image #3 of the image data set downloaded [40]. Here empirical pdf is estimated using density estimation with *epanechnikov* kernel [22].

It can be observed from Fig. 6a that both the empirical pdf and cdf estimated from $\mathbf{r}_{18}$ fit reasonably well in the assumed theoretical gamma pdf and cdf, respectively. Goodness of the fit has also been evaluated using the probability-probability (p-p) plot comparison. Shown in Fig. 6b is the p-p plot of the theoretical and empirical densities of the $\mathbf{r}_{18}$ estimated from image #3 of the image data set downloaded [40].

A. Gamma Density Parameter Estimation

The gamma density parameters β and γ, defined in Eq. (7), are estimated using maximum likelihood (ML) estimation. Let us assume that $\mathbf{r}_i$ is consist of n independent samples $\mathbf{r}_i = \{r_i\}_{j=1}^n$. The likelihood density function can be expressed as,

$$l(\beta_i,\gamma_i)=\Pi_{j=1}^n f_{r_{i,j}}(r(i,j);\beta_i,\gamma_i)=\Pi_{j=1}^n\frac{r_{i,j}^{\gamma_i-1}}{\beta_i^{\gamma_i}\Gamma(\gamma_i)}e^{-\frac{r_{i,j}}{\beta_i}} \tag{8}$$

The log-likelihood is given as,

$$\begin{aligned} L(\beta_i,\gamma_i) &= \log\ l(\beta_i,\gamma_i) && (9)\\ &= (\gamma_i-1)\sum_{j=1}^n\log(r_{i,j})-n\log(\Gamma(\gamma_i)) && (10)\\ &\quad -n\gamma_i\log(\beta_i)-\frac{1}{\beta_i}\sum_{j=1}^n(r_{i,j}) \end{aligned}$$

By differentiating the log-likelihood function with respect to β_i and setting it to zero yields the maximum likelihood estimator of the β_i, that is,

$$\hat{\beta}_i=\frac{1}{\gamma_i n}\sum_{j=1}^n r_{i,j} \tag{11}$$

Substituting Eq. (11) into the log-likelihood function (in Eq. (9)) results,

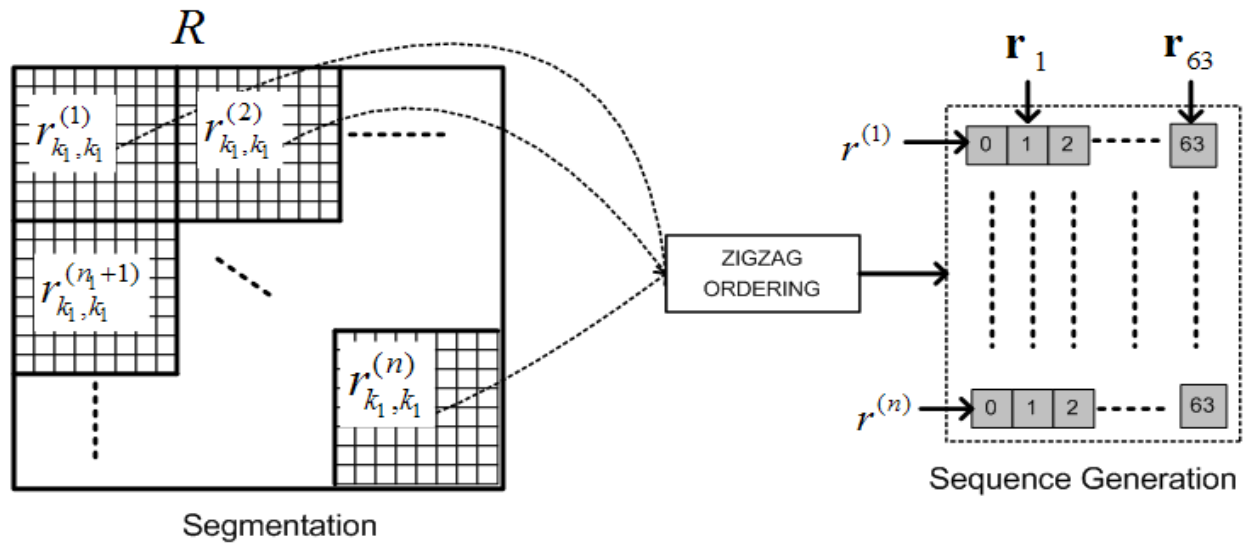
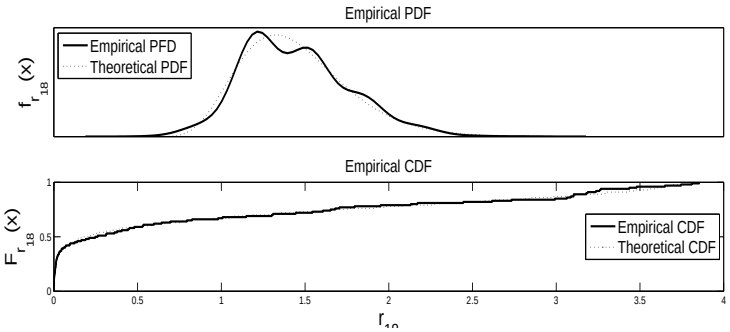
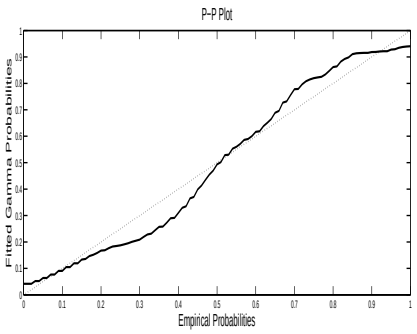


Fig. 5: Illustration of mapping of 2D randomness mask, $r^{(j)}_{k_1,k_2}$, to 1D sequences, $\mathbf{r}_i$



(a) Shown in the top panel are the plots of theoretical pdf (doted line) and empirical pdf (thick line) and shown in the bottom panel are the theoretical cdf (doted line) and empirical cdf (thick line) of the $\mathbf{r}_{18}$ estimated from image #3 of the UCID image data set [40].



(b) Shown is the p-p plot of theoretical and empirical densities of the $\mathbf{r}_{18}$ estimated from image #3 of the UCID image data set.

Fig. 6: Goodness of fit comparison of randomness-mark distribution.

$$L(\beta_i,\gamma_i) = (\gamma_i-1)\sum_{j=1}^{n}\log(r_{i,j}) - n\log(\Gamma(\gamma_i)) - n\gamma_i\log(\frac{1}{\gamma_i n}\sum_{j=1}^{n} r_{i,j}) - \gamma_i n \quad (12)$$

By differentiating the modified log-likelihood function (Eq. 12) with respect to γ_i and setting it to zero yields the maximum likelihood estimator of the γ_i parameter,

$$\psi_0(\gamma_i) = \log\gamma_i + \frac{1}{n}\sum_{j=1}^{n}\log(r_{i,j}) - \log\left(\frac{1}{n}\sum_{j=1}^{n} r_{i,j}\right) \quad (13)$$

where $\psi_0(x)$ is *diagamma function* which denotes the first-order derivative of the $\log\Gamma(\mathbf{x})$, that is, $\psi_0(x) = \frac{\Gamma'(x)}{\Gamma(x)}$.

There is no closed-form solution for γ_i, therefore, we employed Newton's algorithm [41] to obtain an estimate of γ_i. An initial value of γ_i can be found either using the method of moments or using the approximation of *digamma function*, that is,

$$\ln(\gamma_i) - \psi_0(\gamma_i) \approx \frac{1}{2\gamma_o}\left(1+\frac{1}{6\gamma_i+1}\right) \quad (14)$$

From here on, let us denote estimates of β_i and γ_i by $\hat{\beta}_i$ and $\hat{\gamma}_i$, respectively.

The objective of parameterizing the underlying density of $\mathbf{r}$ is to distinguish between the quantized-cover and the QIM-stego. Let the subset of test-images consisting of quantized-cover images be denoted by a superscript $'c'$ and the underlying distribution is parameterized by $\hat{\theta}_i^{(c)} = \{\hat{\beta}_i^{(c)}, \hat{\gamma}_i^{(c)}\}$. Similarly, the subset consisting of QIM-stego images be denoted by a superscript $'s'$ and the underlying distribution is parameterized by $\hat{\theta}_i^{(s)} = \{\hat{\beta}_i^{(s)}, \hat{\gamma}_i^{(s)}\}$.

B. Parameter Distribution Modeling

A likelihood ratio test (LRT) can be used to distinguish between the cover and the stego images, if accurate estimates

of the parameters of the underlying density of both classes, e.g., $\hat{\Theta}_i^{(c)}$ and $\hat{\Theta}_i^{(s)}$, are known to the steganalyst. However, under stego-only attack model, only the test-image is available to the steganalyst to determine whether it carries a hidden message or not. Therefore, under stego-only attack model, estimates of only one class are available. To get around this obstacle, we rely on learning the *reference density* of the estimated parameter of both the cover and the stego images. To this end, we analyzed both the cover- and stego-image data sets to estimate density parameters, $\hat{\Theta}_{i,j}^{(m)}$, where $j = 1, \cdots, N$ and $m \in \{c, s\}$, here N is the number of images in each set (stego and cover). For each set, the distribution of the estimated parameters are used to estimate the underlying densities, that is, $f_{\hat{\beta}_i^{(m)}}(x)$ and $f_{\hat{\gamma}_i^{(m)}}(x)$. It has been observed that for both the cover and stego sets, the estimated parameters also obey the gamma density, e.g., $f_{\hat{\beta}_i^{(m)}}(x) \sim Gamma(\Theta_{\hat{\beta}_i}^{(m)})$ and $f_{\hat{\gamma}_i^{(m)}}(x) \sim Gamma(\Theta_{\hat{\gamma}_i}^{(m)})$.

The goodness of fit of the assumed model for parameters distribution is evaluated by comparing the theoretical pdfs and cdfs of the assumed gamma distribution with estimated parameters $\Theta_{\hat{\beta}_i}^{(m)}$ and $\Theta_{\hat{\gamma}_i}^{(m)}$ and the empirical pdf and cdf estimated from parameters. Shown in the top panel of Fig. 7a are the plots of the theoretical (doted line) and the empirical (continuous line) pdfs and shown in the bottom panel of the Fig. 7a are the theoretical (doted line) and the empirical (thick line) cdfs of the $\beta_{\hat{\gamma}_{18}}^{(s)}$ estimated from stego data set consisting of 400 images. The empirical pdf here is estimated using density estimation with *epanechnikov* kernel [22].

It can be observed from Fig. 7a that the empirical pdf and cdf estimated from $\beta_{\hat{\gamma}_{18}}^{(s)}$ fit reasonably well in the assumed theoretical gamma pdf and cdf, respectively. Goodness of the fit of the assumed distribution model is also evaluated using the p-p plot comparison. Shown in Fig. 7b is the p-p plot of the theoretical and the empirical densities of $\beta_{\hat{\gamma}_{18}}^{(s)}$ estimated from 400 stego images. It can be observed from Fig. 7b that the empirical pdf matches reasonably well with the assumed gamma pdf.

The reference parameters are estimated from the parameters densities, $f_{\hat{\beta}_i^{(m)}}(x)$ and $f_{\hat{\gamma}_i^{(m)}}(x)$, using MLE framework discussed in the previous Section IV-A. The generalized likelihood ratio test (GLRT) is then applied to the estimated parameters from the test-image and the reference parameters to determine whether the test-image is a quantized-cover or a QIM-stego.

V. Stego-Detection

For stego-detection both the intera-subband and the inter-suband parameter estimation are considered.

A. Stego-Detection using Intra-Subband Parameter Estimation

First step in stego-detection using intra-subband is to segment the test-image using 8×8 non-overlapping blocks, transform it to DCT domain, estimate the local-randomness mask and map the 2D mask, $\mathbf{r}$ to 64 1D sequences and estimate the underlying density parameter for each frequency bin $\hat{\beta}_i^{(t)}$, and $\hat{\gamma}_i^{(t)}$, by fitting the estimated local randomness masks, $\mathbf{r}_i, i = 0, \cdots, 64$ from the test-image. Here, superscript $'t'$ denotes test-image, where $t \in \{c, s\}$.

To distinguish between the quantized-cover and the QIM-stego, for each parameter a binary hypothesis test based on the generalized likelihood ratio test (GLRT) [38] is used for each frequency-bin, $i = 0, \cdots, 63$ and each parameter, that is,

$$\lambda(\hat{\beta}_i^{(t)}) = \frac{f_{\hat{\beta}_i^{(t)}|H_s}(\hat{\beta}_i^{(t)}; \hat{\Theta}_{\hat{\beta}_i}^{(s)})}{f_{\hat{\beta}_i^{(t)}|H_c}(\hat{\beta}_i^{(t)}; \hat{\Theta}_{\hat{\beta}_i}^{(c)})}, \tag{15}$$

and

$$\lambda(\hat{\gamma}_i^{(t)}) = \frac{f_{\hat{\gamma}_i^{(t)}|H_s}(\hat{\gamma}_i^{(t)}; \hat{\Theta}_{\hat{\gamma}_i}^{(s)})}{f_{\hat{\gamma}_i^{(t)}|H_c}(\hat{\gamma}_i^{(t)}; \hat{\Theta}_{\hat{\gamma}_i}^{(c)})}, \tag{16}$$

where *generalized* is used here to indicate that without exact knowledge of the parameter values, as we are using ML estimates of these parameters estimated from the data from two subsets, i.e. test-, cover-, and QIM-stego- images.

The decision between the quantized-cover and the QIM-stego is made according to the following rule

$$\lambda(\hat{\beta}_i^{(t)}) \underset{H_s}{\overset{H_c}{\lessgtr}} 1. \tag{17}$$

and

$$\lambda(\hat{\gamma}_i^{(t)}) \underset{H_s}{\overset{H_c}{\lessgtr}} 1. \tag{18}$$

For each frequency-bin, a joint-hypothesis test, Λ_i is obtained for both parameters by fusing $\lambda(\hat{\beta}_i^{(t)})$ and $\lambda(\hat{\gamma}_i^{(t)})$ using the *OR-logic*, i.e., $\Lambda_i = \lambda(\hat{\beta}_i^{(t)}) + \lambda(\hat{\gamma}_i^{(t)})$, where '$+$' denotes the *OR-logic operator*.

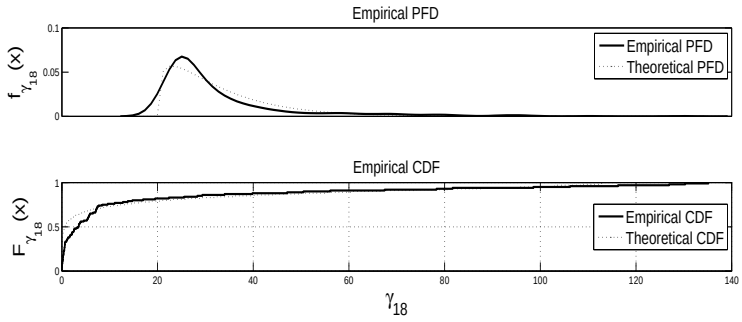
Following binary hypothesis testing based on majority voting of the joint-hypothesis is used to declare stego-detection decision, D_I,

$$D_I = \begin{cases} 1 \quad (i.e.,\ \mathbf{x}_t = \mathbf{x}_{QIM}) & \text{if } \sum_{i=1}^{63} \mathbf{1}_{\Lambda_i} \geq \eta_I \\ 0 \quad (i.e., \mathbf{x}_t = \mathbf{x}_q) & \text{otherwise} \end{cases}$$

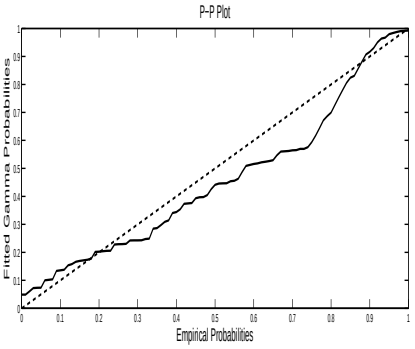
where $\mathbf{1}$ is an indicator function and η_I is predefined decision threshold for frequency-dependent based stego-detector.

B. Stego-Detection using Inter-Subband Estimation Parameters

For the inter-subband based stego-detection case, higher-order statistics (HOS) of the estimated parameters for all frequency bins is used. First step in stego-detection using inter-subband higher-order statistics (iHOS) is to segment the test-image using 8×8 non-overlapping blocks, transform it to DCT domain, estimate the local-randomness mask and map the 2D mask, $\mathbf{r}$ to 64 1D sequences estimate underlying density parameter from test-images by fitting the estimated local randomness masks, $\mathbf{r}_i, i = 0, \cdots, 64$. Next step is to compute higher-order moments from estimated density parameter sequences, i.e., $\hat{\beta}_i^{(t)}$, and $\hat{\gamma}_i^{(t)}, i = 0, \cdots, 63$. More specifically, first four moments, that is, mean $\mu_{\Theta,1}^{(t)}$, variance



(a) Shown in the top panel are the plots of the theoretical pdf (doted line) and the empirical pdf (continuous line) and shown in the bottom panel are the theoretical cdf (doted line) and the empirical cdf (thick line) of the $\beta^{(s)}_{\hat{\gamma}_{18}}$ estimated from 400 stego images.



(b) Shown is the p-p plot of theoretical and empirical densities of the $\beta^{(s)}_{\hat{\gamma}_{18}}$ estimated from 400 stego images

Fig. 7: Goodness of fit comparison of the estimated parameter distribution

$\mu^{(t)}_{\Theta,2}$, skewness $\mu^{(t)}_{\Theta,3}$, and kurtosis $\mu^{(t)}_{\Theta,4}$ are computed for both estimated parameters, that is, $\hat{\beta}^{(t)}_i$ and $\hat{\gamma}^{(t)}_i$.

To learn reference density parameters, we again analyzed the cover and the stego data sets to estimate density parameters, $\hat{\Theta}^{(m)}_{i,j}$, where $j = 1, \cdots, N$ and $m \in \{c, s\}$, here N is the number of images each set (stego and cover). For each image, first four moments are computed from $\hat{\Theta}^{(m)}_{i,j}$, that is, $\mu^{(m)}_{\beta,k,j} = \Phi(\hat{\beta}^{(m)}_{i,j}, k)$ and $\mu^{(m)}_{\gamma,k,j} = \Phi(\hat{\gamma}^{(m)}_{i,j}, k)$ where $\Phi(\cdot, k)$ is k^{th} moment computing function. Distributions of the estimated moments are used to estimate the underlying density parameters by fitting the estimated moments to a gamma distribution. The reference density parameters of first four moments are estimated from the parameters densities using MLE framework discussed in the previous Section IV-A. The GLRT is then applied to the moments estimated from the test-image, $\mu^{(t)}_{\Theta,k}$, and the reference parameters to determine whether the test-image is a quantized-cover or a QIM-stego.

To distinguish between the quantized-cover and the QIM-stego using iHOS, for each moment $\mu^{(t)}_{\hat{\beta},k}$ and $\mu^{(t)}_{\hat{\gamma},k}$ where $k = 1, \cdots, 4$ a binary hypothesis test based on the GLRT, that is,

$$\lambda(\mu^{(t)}_{\hat{\beta},k}) = \frac{f_{\mu_{\hat{\beta},k}|H_s}(\mu^{(t)}_{\hat{\beta},k}; \hat{\Theta}^{(s)}_{\mu_{\hat{\beta}},k})}{f_{\mu_{\hat{\beta},k}|H_c}(\mu^{(t)}_{\hat{\beta},k}; \hat{\Theta}^{(c)}_{\mu_{\hat{\beta}},k})}, \tag{19}$$

and

$$\lambda(\mu^{(t)}_{\hat{\gamma},k}) = \frac{f_{\mu_{\hat{\gamma},k}|H_s}(\mu^{(t)}_{\hat{\gamma},k}; \hat{\Theta}^{(s)}_{\mu_{\hat{\gamma},k}})}{f_{\mu_{\hat{\gamma},k}|H_c}(\mu^{(t)}_{\hat{\gamma},k}; \hat{\Theta}^{(c)}_{\mu_{\hat{\gamma},k}})}. \tag{20}$$

The decision between the quantized-cover and the QIM-stego is made according to the following decision rules,

$$\lambda(\mu^{(t)}_{\hat{\beta},k}) \underset{H_s}{\overset{H_c}{\gtrless}} 1. \tag{21}$$

and

$$\lambda(\mu^{(t)}_{\hat{\gamma},k}) \underset{H_s}{\overset{H_c}{\gtrless}} 1. \tag{22}$$

For each moment, a joint-hypothesis test, Λ_i is obtained by fusing $\lambda(\hat{\beta}^{(t)}_i)$ and $\lambda(\hat{\gamma}^{(t)}_i)$ using the *OR-logic*, i.e., $\Lambda_k = \lambda(\mu^{(t)}_{\hat{\beta},k}) + \lambda(\mu^{(t)}_{\hat{\gamma},k})$.

Following binary hypothesis testing based on majority voting of joint-hypothesis is used to declare stego-detection decision, D_{iHOS},

$$D_{iHOS} = \begin{cases} 1 \quad (i.e., \mathbf{x}_t = \mathbf{x}_{QIM}) & \text{if } \sum_{k=1}^{4} \mathbf{1}_{\Lambda_k} \geq \eta_{hos} \\ 0 \quad (i.e., \mathbf{x}_t = \mathbf{x}_q) & \text{otherwise} \end{cases}$$

where η_{iHOS} is predefined decision threshold for HOS based stego detector.

C. Decision Fusion

For stego-detection, the the outputs of both stego detectors, that is, D_I and D_{iHOS}, are fused to generate the final stego-detection decision. To achieve this goal, the *OR-logic* based decision fusion rule is used, that is, $D_{stego} = D_I + D_{iHOS}$.

VI. Experimental Results for QIM-Stego

Detection performance of the proposed steganalysis scheme to detect QIM steganography is 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. To this end, the message, M, is embedded into all AC coefficients of 8×8 blocks in DCT domain using QIM.
- ***Random Embedding*** (RE): In this case, a set of AC coefficients is selected randomly from each block (in DCT domain) for message embedding. For random embedding, embedding rates $R \in \{0.1, 0.3, 0.5, 0.7, 0.9, 1.0\}$ are considered.

A. Sequential Embedding

Detection performance of the proposed steganalysis scheme for QIM steganography is evaluated in terms of false rates, that is, false positive rates, P_{fp}, and false negative rates, P_{fn}.

The *Uncompressed Colour Image Database* (UCID) [40] is used to evaluate performance of the proposed steganalysis scheme for QIM-stego detection. This image database [40] contains 1338 uncompressed color images, results presented here are based on resized (to 256x256 pixels) and gray-scale versions of the first eight hundred (800) images of the database [40]. Eight-thousand QIM-stego images were obtained by sequentially embedding 8000 random messages into first 800 images of the database using QIM with $\Delta = \{0.25, .26, .5, 1, 1.7, 2, 3.8, 4, 6.9, 8\}$. Similarly, 8000 quantized-cover images were obtained by quantizing first 800 images with $\Delta^* = \Delta/2$. The resulting 16000 quantized images (8000 QIM-stego and 8000 quantized-cover) are analyzed using proposed steganalysis scheme. Shown in Table I is the average detection performance the proposed three stego detectors averaged over all quantization-stego sizes. The simulation results presented in Table Table I are generated with decision thresholds $\eta_I = 32$ and $\eta_{iHOS} = 2$ and gamma density parameters are estimated using MLE framework discussed in Section IV-A with confidence interval of 95%.

It can be observed that Table I that the proposed steganalysis scheme can successfully distinguish between the quantized-cover and the QIM-stego images with relatively low false rates, e.g., $P_{fp} < 0.015$, $P_{fn} < 0.03$. Moreover, detection performance is independent of the quantization step-size used. It can also be observed from Table I that decision fusion based detector significantly improves the detection performance irrespective of quantization step-size.

B. Random Embedding

Effectiveness of the proposed scheme for random embedding is also evaluated using first 880 images of the UCID [40]. Over sixteen thousand QIM-stego images (16320 to be exact) were obtained by embedding 16320 random messages using binary QIM with $\Delta = \{2, 3.8, 4, 8\}$ and embedding rate $R \in \{0.1, 0.3, 0.5, 0.7, 0.9, 1.0\}$. The 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 $\Delta^* = \Delta/2$. Similarly, 3520 quantized-images were obtained by quantizing selected 880 gray-scale images using plain-quantizer with $\Delta^* = \Delta/2$. The resulting 19840 quantized images (880*4*6 = 16320 QIM-stego images using RE and 880*4 = 3520 quantized-cover images using plain-quantizer) are analyzed using proposed scheme. Shown in Table II is the average detection performance the proposed three stego detectors for various embedding rates averaged over all quantization step-size. The simulation results presented in Table II are generated with decision thresholds $\eta_I = 32$ and $\eta_{iHOS} = 2$ and gamma density parameters are estimated using MLE framework discussed in Section IV-A with confidence interval of 95%.

It can be observed from Table II that for all quantization step-size false negative rates P_{fn} improves gradually as embedding rate, R, increases for all three types of detectors and fusion based detection significantly improves detector performance. It can also be observed from Table III that for all quantization step-size fusion based detection improves false positive rates P_{fp} performance.

TABLE II: **QIM-Stego Detection Performance: *Random Embedding*** P_{fn}

	Embedding Rate R					
Error	*0.1*	*0.3*	*0.5*	*0.7*	*0.9*	*1.0*
Quantization Step-Size $\Delta = 2$						
Intra-Subband Analysis-based Detection						
P_{fn}	0.265	0.2235	0.175	0.157	0.091	0.046
Inter-Subband Analysis-based Detection						
P_{fn}	0.0603	0.053	0.016	0.003	0	0
Fusion-based Detection						
P_{fn}	0.0382	0.01	0.002	0	0	0
Quantization Step-Size $\Delta = 3.8$						
Intra-Subband Analysis-based Detection						
P_{fn}	0.296	0.243	0.21	0.165	0.102	0.068
Inter-Subband Analysis-based Detection						
P_{fn}	0.034	0.01	0.01	0	0	0
Fusion-based Detection						
P_{fn}	0.0294	0.006	0.006	0	0	0
Quantization Step-Size $\Delta = 4$						
Intra-Subband Analysis-based Detection						
P_{fn}	0.291	0.249	0.215	0.147	0.059	0.035
Inter-Subband Analysis-based Detection						
P_{fn}	0.065	0.025	0.01	0.002	0	0
Fusion-based Detection						
P_{fn}	0.034	0.019	0.009	0	0	0
Quantization Step-Size $\Delta = 8$						
Intra-Subband Analysis-based Detection						
P_{fn}	0.309	0.259	0.231	0.178	0.068	0.038
Inter-Subband Analysis-based Detection						
P_{fn}	0.034	0.01	0.006	0.002	0	0
Fusion-based Detection						
P_{fn}	0.003	0.003	0.003	0.002	0	0

TABLE III: **Stego Detection Performance: *Random Embedding*** P_{fn}

	Quantization Step-size Δ			
False Rates	2.0	3.8	4.0	8.0
Intra-Subband Analysis-based Detection				
P_{fp}	0.297	0.357	0.353	0.320
Inter-Subband Analysis-based Detection				
P_{fp}	0.026	024	0.021	0.024
Fusion-based Detection				
P_{fp}	0.013	0.006	0.004	0.024

C. Performance Comparison with Existing Stat of the Art

In our last experiment, we compared the performance of the proposed method against existing state of the art [6], [22]. More specifically, we compared the performance of the proposed method against Zou et al. method [6] and Malik et al. method [22]. To this end, we generated a set of 1200 quantized-cover images by quantizing first 400 images of the UCID [40] with quantization-step size $\Delta^* = \{0.5, 1, 2\}$. Similarly, we generated another set of 1200 QIM-stego images by embedding 1200 random messages in the first 400 images of the UCID [40] using $\Delta = \{1, 2, 4\}$. Here stego images are generated by embedding random messages at a rate of 1 bit per sample. The stego detection performance of the proposed method and the selected two methods [6], [22] are tested using quantized-cover and QIM-stego sets.

For Zou et als' method [6], a two class support vector machines (SVM) classifier is trained with the "*radial basis*

TABLE I: **Stego Detection Performance:** ***Sequential Embedding***

$\mathbf{X}_q$ *vs* $\mathbf{X}_{QIM}$										
Quantization Step-size Δ										
False Rates	0.25	0.26	0.5	1.0	1.7	2.0	3.8	4.0	6.9	8.0
Intra-Subband Analysis-based Detection										
P_{fp}	0.087	0.29	0.29	0.3	0.094	0.183	0.136	0.245	0.135	0.241
P_{fn}	0.073	0.083	0.1	0	0.005	0.023	0.014	0.01	0	0
Inter-Subband Analysis-based Detection										
P_{fp}	0.003	0.04	0.034	0.016	0.018	0.027	0.03	0.09	0	0.02
P_{fn}	0.03	0	0.003	0	0	0.001	0	0	0	0
Decision Fusion-based Detection										
P_{fp}	0.003	0	0.015	0.008	0	0.001	.001	0.06	0	0
P_{fn}	0.03	0	0.003	0	0	0	0	0	0	0

function kernel". The Matlab SVM code downloaded from LIBSVM [42] is used to achieve this goal. For training and testing we randomly divided the test images (both the cover and the stego) into two sets and use one for training and the other for testing. Detection performance is obtained using 5-fold cross-validation method. Similarly, for Malik et als' method [22] the for decision thresholds, $\tau_i, i = 1, 2, 3$ are set to $\tau_1 = 0.2$, $\tau_2 = 0.45$ and $\tau_3 = 1.1$. Detection performance measured in terms of false rates and accuracy, $Acc = (2-(P_{fp}+P_{fn}))/2 \times 100$, for three methods is shown in Table .

TABLE IV: Detection Performance Comparison

	Quantization Step-size Δ		
False Rates	$\Delta = 1$	$\Delta = 2$	$\Delta = 4$
Zuo et als' Method [6]			
P_{fp}	0.659	0.476	0.206
P_{fp}	0.31	0.322	0.309
Acc(%)	51.53	60.12	74.34
Malik et als' Method [22]			
P_{fp}	0.63	0.46	0.19
P_{fp}	0.041	0.034	0.03
Acc(%)	67.95	73.5	89.5
Proposed Method: Intra-Subband Analysis-Based Detection			
P_{fp}	0.3	0.183	0.245
P_{fp}	0	0.023	0.01
Acc(%)	85	89.7	87.45
Proposed Method: Inter-Subband Analysis-Based Detection			
P_{fp}	0.016	0.027	0.09
P_{fp}	0	0.001	0
Acc(%)	99.2	98.6	95.5
Proposed Method: Fusion-Based Detection			
P_{fp}	0.008	0.001	0.06
P_{fp}	0	0	0
Acc(%)	99.6	99.95	97.0

It can be observed from Table IV the proposed scheme even with Intra-Subband analysis-based detection (the worst performing case of the proposed method) outperforms the selected method [6], [22] for all three quantization step-size values.

VII. Conclusion

In this paper, a new statistical modeling-based framework is proposed to characterize traces of quantization. The proposed scheme models the traces of quantization using gamma distribution and use it to develop a parametric steganalysis technique to attack QIM steganography. We have shown that quantization (with and without message embedding) reduces local-randomness in the resulting quantized signal and the QIM-stego exhibits a higher level of randomness than the corresponding quantized-cover. The proposed scheme uses local-randomness based to capture traces of message embedding. The maximum likelihood estimation is used to estimate parameters of the underlying distribution. Distribution of the estimated parameters is modeled using gamma distribution. The GLRT based parametric classifier is used to distinguish between the QIM-stego and the quantized-cover. Effectiveness of the proposed method is tested on a large data set, consisting of more than 35000 images, for both the sequential and random message embedding. Simulation results show that the proposed method can successfully distinguish between the quantized-cover and the QIM-stego with very low false alarm rates, i.e., ($P_{fp} < 0.015$ and $P_{fn} < 0.03$) for sequential embedding and false rates ($P_{fn} < 0.038$ and $P_{fp} < 0.013$) for random embedding. Detection performance comparison with existing state of the art [6], [22] indicates that the proposed scheme performs significantely better than the selected method.

References

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

[2] S. Lyu and H. Farid, "Steganalysis using higher-order image statistics," *IEEE Trans. Information Forensics Security*, vol. 1, no. 1, pp. 111 – 119, Mar. 2006.

[3] I. Avcibas, N. Memon, and B. Sankur, "Steganalysis using image quality metrics," *IEEE Trans. Image Processing*, vol. 12, no. 2, pp. 221 – 229, Feb. 2003.

[4] M. Celik, G. Sharma, and A. Tekalp, "Universal image steganalysis using rate-distortion curves," in *SPIE, Electronic Imaging, Security, Steganography, Watermarking of Multimedia Contents VI*, E. J. Delp and P. W. Wong, Eds., San Jose, CA, 2004, IS&T/SPIE, vol. 5306, pp. 467 – 476.

[5] 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.

[6] D. Zou, Y. Shi, W. Su, and G. Xuan, "Steganalysis based on markov model of thresholded prediction-error image," IEEE, 2006, Int. Conf. on Multimedia and Expo.

[7] D. Fu, Y. Shi, D. Zou, and G. Xuan, "Jpeg steganalysis using empirical transition matrix in block dct domain," in *IEEE 8th Workshop on Multimedia Signal Processing*, Victoria, BC, Canada, Oct. 2006, IEEE, pp. 310 – 313.

[8] Q. Liu, A. Sung, and M. Qiao, "Neighboring joint density-based jpeg steganalysis," *ACM Trans. Information Forensics Security*, vol. 2, no. 3, pp. article 16, Feb. 2011.

[9] G. Xuan, Y. Shi, J. Gao, D. Zou, C. Yang, Z. Zhang, P. Chai, C. Chen, and W. Chen, "Steganalysis based on multiple features formed by statistical moments of wavelet characteristic functions," in *7th Int. Workshop Information Hiding*, S. Katzenbeisser M. Barni, J. Herrera and F. Perez-Gonzalez, Eds., Barcelona, Spain, June 2005, vol. 3727 of *Lecture Notes in Computer Science*, pp. 262 – 277, Berlin: Springer-Verlag.

[10] T. Pevny and J. Fridrich, "Merging markov and dct features for multi-class jpeg steganalysis," in *SPIE, Electronic Imaging, Security, Steganography, Watermarking of Multimedia Contents IX*, E. J. Delp and P. W. Wong, Eds., San Jose, CA, Feb. 2007, IS&T/SPIE, vol. 6505, pp. 1 – 14.

[11] K. Sullivan, O. Dabeer, U. Madhow, B. Manjunath, and S. Chandrasekaran, "Llrt based detection of lsb hiding," in *IEEE Int. Conf. Image Processing (ICIP)*, 2003, vol. 1, pp. 497–500.

[12] K. Solanki, N. Jacobsen, S. Chandrasekaran, U. Madhow, and B. Manjunath, "High volume data hiding in images: Introducing perceptual criteria into quantization based embedding," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2002.

[13] T. Pevny and J. 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.

[14] H. Farid, "Detecting hidden messages using higher-order statistical models," in *IEEE Int. Conf. on Image Processing (ICIP)*, Rochester, NY, 2002.

[15] Y. Wang and P. Moulin, "Optimized feature extraction for learning-based image steganalysis," *IEEE Trans. on Information Forensics and Security*, vol. 1, no. 2, pp. 31 – 45, 2006.

[16] J. Harmsen and W. Pearlman, "Steganalysis of additive noise modelable information hiding," in *SPIE, Electronic Imaging, Security, Steganography, Watermarking of Multimedia Contents V*, E. J. Delp and P. W. Wong, Eds., Santa Clara, CA, Jan. 2003, IS&T/SPIE, vol. 5020, pp. 131 – 142.

[17] R. Duda, D. Stork, and P. Hart, *Pattren Classification*, John Wiley & Sons,, 2000.

[18] J. Fridrich and M. Goljan, "Digital image steganography using stochastic modeling," in *SPIE, Electronic Imaging, Security, Steganography, Watermarking of Multimedia Contents V*, E. J. Delp and P. W. Wong, Eds., Santa Clara, CA, 2003, IS&T/SPIE, vol. 5020, pp. 191 – 202.

[19] J. Fridrich, M. Goljan, and D. Hogea, "Attacking the outguess," in *ACM Workshop on Multimedia and Security 2002*, Juan-les-Pins, France, Dec. 2002, ACM.

[20] J. Fridrich, M. Goljan, and D. Hogea, "Steganalysis of jpeg images: Breaking the f5 algorithm," in *5th Int. Workshop Information Hiding*, S. Katzenbeisser M. Barni, J. Herrera and F. Perez-Gonzalez, Eds. 2002, vol. 2578 of *Lecture Notes in Computer Science*, pp. 310 – 323, Springer Berlin / Heidelberg.

[21] J. Fridrich, T. Holotyak, and D. Soukal, "Maximum likelihood estimation of secret message length embedded using $+/-k$ steganography in spatial domain," in *SPIE, Electronic Imaging, Security, Steganography, Watermarking of Multimedia Contents VII*, E. J. Delp and P. W. Wong, Eds., Santa Clara, CA, Jan. 2005, IS&T/SPIE, vol. 5681, pp. 595 – 606.

[22] 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).

[23] H. Malik, "Statistical modeling of footprints of qim steganography," in *Proceedings of the 10th ACM workshop on Multimedia and security (MM&Sec'08)*, Oxford, UK, August 2008, ACM, pp. 149 – 158, ACM.

[24] H. Malik, "Statistical modeling of footprints of qim steganography," in *IEEE Int. Conference on Multimedia and Expo, 2010*. IEEE, July 2010, pp. 1484 – 1492, IEEE.

[25] H. Malik, K.P. Subbalakshami, and R. Chandramouli, "Nonparametric steganalysis of qim steganography using approximate entropy," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 418–431, April 2012.

[26] 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*.

[27] 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.

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

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

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

[31] T. Pevny and J. 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.

[32] 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.

[33] S. Trivedi and R. Chandramouli, "Secret key estimation in sequential steganalysis," *IEEE Transactions on Signal Processing: Supplement on Secure Media*, vol. 53, no. 2, pp. 746–757, February 2005.

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

[35] Q. Wu, W. Li, and X. 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.

[36] K. Lee, D. Kim, T. Kim, and K. 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.

[37] S. Kim and K. 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.

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

[39] H. Malik, "Steganalysis of qim steganography using approximate entropy," Tech. Rep., University of Michigan - Dearborn, September 2011.

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

[41] C. Kelley, *Solving Nonlinear Equations with Newton's Method*, SIAM, 2003.

[42] C. Chang and C. Lin, "Libsvm: A library for support vector machines," http://www.csie.ntu.edu.tw/ cjlin/libsvm, 2001.