Information Theoretical Limit of Operation Forensics

Xiaoyu Chu, *Student Member, IEEE*, Yan Chen, *Member, IEEE*, Matthew C. Stamm, *Member, IEEE*, and K. J. Ray Liu, *Fellow, IEEE*

***Abstract*—While more and more forensic techniques have been proposed to detect the processing history of multimedia content, one starts to wonder if there exists a fundamental limit on the capability of forensics. In other words, besides keeping on searching what investigators can do, it is also important to find out the limit of their capability and what they cannot do. In this work, we explore the fundamental limit of operation forensics by proposing an information theoretical framework. Specifically, we consider a general forensic system of estimating operations' hypotheses based on extracted features from the multimedia content. In this system, forensicability is defined as the maximum forensic information that features contain about operations. Then, due to its conceptual similarity with mutual information in information theory, forensicability is measured as the mutual information between features and operations' hypotheses. Such a measurement gives the error probability lower bound of all practical estimators which use these features to detect the operations' hypotheses. Furthermore, it can determine the maximum number of hypotheses that we can theoretically detect. To demonstrate the effectiveness of our proposed information theoretical framework, we apply this framework on a forensic example of detecting the number of JPEG compressions based on DCT coefficient histograms. We conclude that, under typical settings of forensic analysis, the maximum number of JPEG compressions that we can detect using DCT coefficient histogram features is 4. Furthermore, we obtain the optimal strategies for investigators and forgers based on the fundamental measurement of forensicability.**

I. Introduction

Due to the ease of tampering a multimedia file, forensics has gained much attention in the recent decade for providing technical tools to verify the authenticity of multimedia content [1]. Enabled by techniques in existing forensic literature, forensic investigators can not only identify the acquisition environment of multimedia content [2]–[7], but also detect the processing history that the content has gone through after acquisition [8]–[13]. For the purpose of improving the detection performance and identifying more sophisticated manipulations, forensic researchers have always been working on discovering new fingerprints and designing new schemes [14]–[17].

However, as the effort of developing more powerful forensic techniques goes on, evidence has shown difficulties when dealing with complicated manipulation scenarios [16]. One would then wonder if there exists a fundamental limit on forensic capability that can never be exceeded? In other words, what is the limit of investigators' capability? How many manipulation operations that investigators can detect at most? Given this information, we would be able to tell whether the existing technique has achieved the limit. If not, how far can it go? Furthermore, by quantifying the forensic capability, we may also obtain information about how to achieve the capability limit. In addition, given that forgers may manipulate multimedia content to the extent beyond the limit of forensics, special care would be needed for such cases.

There are few works exploring the fundamental limit of forensic capabilities. To the best of our knowledge, the most related work on fundamental limit analysis of forensics was done by Swaminathan *et al.* [18], [19]. They explored the fundamental limit in component forensics by establishing two theoretical frameworks: an estimation framework and a pattern classification framework. Three types of forensic scenarios were defined in each framework regarding how much information investigators have about the components of a camera. Then, fundamental relationships of their forensic performance were derived using the above two theoretical frameworks. Moreover, in the estimation framework, Fisher information was used to obtain the optimal input for semi non-intrusive component forensics. However, these theoretical frameworks were designed for camera identification forensics, and thus they may not be suitable for answering fundamental questions in operation forensics, which focuses on detecting manipulation operations.

In this paper, we explore the fundamental limit of operation forensics by building an information theoretical framework. We consider the forensic scenario of detecting the processing history of given multimedia content. We aim to answer the question of how many operations that investigators can detect, at most? To answer this question, we define *forensicability* as the forensic capability of detecting operations. Unlike the measure of distinguishability proposed in [20], which was based on a simple hypothesis model, our definition is applicable for more general scenarios where multiple operations may happen and many hypotheses can be considered. Given that investigators often use features to estimate process history, in our information theoretical framework, forensicability indicates the maximum forensic information that extracted features can contain about detecting operations. Furthermore, it determines the fundamental limit of forensic detection performance of any scheme based on those features. Then, by introducing a statistical concept of *expected perfect detection*,

This work is supported in part by the NSF grant CCF1320803

Xiaoyu Chu, Yan Chen and K. J. Ray Liu are with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: {cxygrace, yan, kjrliu}@umd.edu).

Matthew C. Stamm is with the Department of Electrical and Computer Engineering, Drexel University, Philadelphia, PA 19104 USA (e-mail: mstamm@coe.drexel.edu).

Fig. 1. Typical process that a multimedia signal may go through when considering forensics.

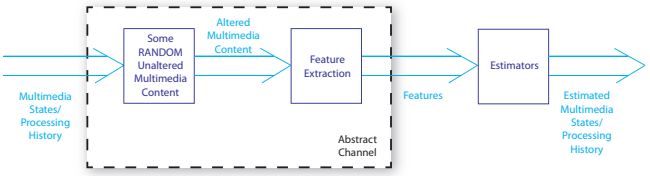


Fig. 2. Abstract channel model in our information theoretical framework.

we are able to use forensicability to determine the maximum number of operations investigators can detect. In addition, the fundamental measure of forensicability provides insights and theoretical support for predicting forgers' behavior and designing optimal forensic schemes.

The remaining of this paper is organized as follows. Section II introduces our information theoretical framework for operation forensics, where forensicability is defined and analyzed for general scenarios. Then, to demonstrate our framework, we apply it to the forensic problem of multiple JPEG compression detection in Section III. In this section, specific models for DCT coefficient histogram features are proposed to derive the expression of forensicability in this example. Then, Section IV performs all experiments corresponding to the theoretical analysis in Section III. Among these experimental results, we obtain the maximum number of JPEG compressions one can detect using DCT coefficient histograms. In addition, the best strategies for investigators and forgers are also analyzed in this section. Lastly, Section V concludes our work.

II. Information Theoretical Framework

In this section, we introduce our information theoretical framework for general operation forensic systems. Under this model, we define the capability of investigators as forensicability, which determines the lower bound of estimation error probability and helps us answer the question of when we cannot detect any more operations.

A. Channel between Multimedia States and Features

Let us consider the process of a typical forensic analysis shown in Fig. 1. Unaltered multimedia content may go through some processing before investigators obtain it. In order to identify the processing history that the obtained multimedia content went through, investigators extract features from the content. Based on the extracted features, specific estimators are proposed to finally estimate the processing history.

During this process, it is often assumed that there are a finite number of hypotheses on processing histories that the multimedia content may go through. Investigators determine which hypothesis actually happened based on the analysis of extracted features. For example, to detect if the multimedia content was edited by a certain operation, like contrast enhancement [10], resizing [8] or compression [21], simple hypothesis test was used to distinguish the unaltered multimedia content and the content edited by the certain operation. In another example of detecting the number of compressions, the hypotheses would include single compression, double compression, triple compression and so on. In this work, processing history hypotheses considered in a certain forensic analysis are denoted as *multimedia states*. Then, investigators' goal is to distinguish multimedia states based on extracted features.

Given the discussion above, we reformulate the forensic system in a different way such that the relationship between multimedia states and features can be emphasized. As it is shown in Fig. 2, in this new formulation, the multimedia state is the input to the system. When a certain multimedia state is applied on unaltered multimedia content, features can be extracted from the processed multimedia content. Then, estimators will be applied on these features to estimate the input multimedia state.

By exploring fundamental limits in operation forensics, we want to answer "what is the maximum information about multimedia states that investigators can obtain from the extracted features?" In other words, we are concerning the fundamental relationships between multimedia states and features, regardless of specific detectors or estimators that investigators may use to make final decisions. This motivates us to abstract all processes between multimedia states and features as a channel. Within this channel, the unaltered multimedia content can be any particular content, and it is modeled as a random variable. As a result, the relationship between multimedia states and features becomes stochastic instead of deterministic.

To demonstrate our abstract channel and further explain the relationship between multimedia states and features, let us consider an example of detecting the number of JPEG compressions using the DCT coefficients feature. As it is shown in Fig. 3, the multimedia state is the number of JPEG compressions from 1 to M. The feature is DCT coefficient histogram represented in a vector. Fig. 4 illustrates the mapping between multimedia states and features in this example. Specifically, with the same number of compressions applied, different images result in different DCT coefficient histograms, which we call them a histogram set. When we detect double compressions, we are distinguishing single compression, $X = 1$, and double compression, $X = 2$. Given the distinctive fingerprints for single compression and double compression, the DCT coefficient histogram sets resulted from these two inputs can be well separated after some post-processing [11]. Thus, for $M = 2$, classification schemes can be used to distinguish the input according to the output. However, as the number of compressions considered in the system increases, more overlapping between different histogram sets may occur, which will affect the accuracy of the detection. Finally, at a certain point, we cannot distinguish all inputs and we say that we have reached our limit of detecting multiple compressions. Detailed modeling and analysis will be discussed in Section III.

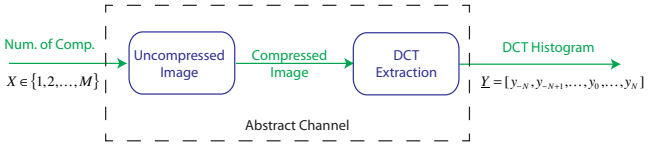


Fig. 3. Channel model for the example of multiple compression detection forensics.

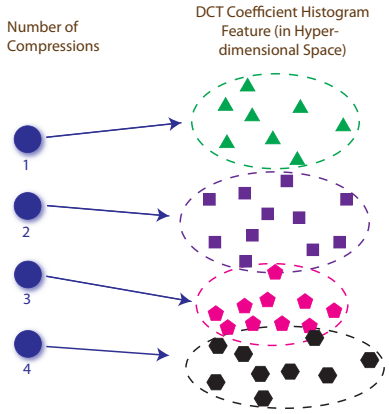


Fig. 4. An illustration of the mapping between multimedia states and features in the example of multiple compression detection.

B. Forensicability

Given the channel model built up between multimedia states and features, we are ready to define forensicability for operation forensics. Let us consider the general abstract channel proposed in our information theoretical framework, as it is simplified in Fig. 5. Let $X \in \{1, 2, \ldots, M\}$ denote the input of the channel, i.e., the multimedia state considered in a forensic analysis. Let $\underline{Y}$ denote the output of the channel, which is a vector containing features that examined by investigators. After obtaining feature $\underline{Y}$, investigators design estimators based on their statistics to estimate X. We define *forensicability* in this forensic system as the maximum information that features contain about multimedia states, regardless of any specific estimators used afterward. It is well known that, in a channel model, mutual information implies the reduction in uncertainty of input due to the knowledge of output. Thus, given the similarity between these two concepts, we define forensicability as follows.

Definition 1: In operation forensics, where features are used to identify multimedia states, *forensicability* of using feature $\underline{Y}$ towards identifying multimedia state X, denoted as $F(X;\underline{Y})$, is defined as the mutual information between X and $\underline{Y}$, i.e.,

$$F(X;\underline{Y}) \triangleq I(X;\underline{Y}). \tag{1}$$

Forensicability of an operation forensic system implies the maximum forensic information that features contain about multimedia states. More importantly, it determines the best performance investigators can obtain by examining these features through all possible estimators. We demonstrate this significance in the following theorem.

Theorem 1: Consider any estimator of the multimedia state $\hat{X}$ such that $X \to \underline{Y} \to \hat{X}$ is a Markov Chain, i.e., the value of $\hat{X}$ depends only on $\underline{Y}$ and not on X. Let $P_e = \mathbb{P}(X \neq \hat{X})$ denote the error probability. If the estimator is better than a random decision where $\hat{X}$ is uniformly and randomly drawn from the set of X, i.e., $P_e \leq \frac{M-1}{M}$, then we have

$$P_e \geq P_e^0, \tag{2}$$

where P_e^0 is the lower bound of error probability. It is unique and satisfies the following equation

$$H(P_e^0) + P_e^0 \log_2(M-1) = H(X) - F(X;\underline{Y}). \tag{3}$$

Proof: This theorem can be proved by modifying the derivation of Fano's inequality in [22] as follows. The Fano's inequality in [22] is stated as

$$H(P_e) + P_e \log_2 |\mathcal{X}| \geq H(X|\underline{Y}), \tag{4}$$

where $|\mathcal{X}|$ is the cardinality of the input X. We slightly tighten a step in its derivation, $H(X|\hat{X}, E=1) \leq P_e \log_2 |\mathcal{X}|$, to $H(X|\hat{X}, E=1) \leq P_e \log_2(|\mathcal{X}|-1)$, where $E = \mathbf{1}(\hat{X} \neq X)$ is an error random variable and $\mathbf{1}(\cdot)$ is an indicator function. Then, given that $|\mathcal{X}| = M$, the following modified inequality can be obtained,

$$H(P_e) + P_e \log_2(M-1) \geq H(X|\underline{Y}). \tag{5}$$

Next, we examine the derivative of the left hand side of above inequality with respect to P_e,

$$\frac{\partial\big(H(P_e) + P_e \log_2(M-1)\big)}{\partial P_e} = \log_2\left(\frac{1-P_e}{P_e}(M-1)\right) \geq 0. \tag{6}$$

The last step holds because $P_e \leq \frac{M-1}{M}$. Therefore, the left hand side of (5) is an increasing function of P_e. Then, the minimum of P_e can be obtained by solving the equality of (5). Hence, we have, $P_e \geq P_e^0$, where P_e^0 is the unique solution of the following equation,

$$H(P_e^0) + P_e^0 \log_2(M-1) = H(X|\underline{Y}) = H(X) - I(X;\underline{Y}). \tag{7}$$

■

The lower bound P_e^0 can be achieved if and only if all of the following conditions are satisfied.

1) $H(E|\hat{X}) = H(E)$, i.e., E and $\hat{X}$ are independent. Furthermore, it can be easily proved that the independence between E and $\hat{X}$ implies that the error probability for each given estimated result is the same, i.e., $\mathbb{P}(X \neq i|\hat{X}=i) = \mathbb{P}(X \neq j|\hat{X}=j), \forall 1 \leq i,j \leq M$. For the specific setting of this work, it indicates that multimedia states are equally hard to be correctly identified.
2) $H(X|\hat{X}, E=1) = \log_2(M-1)$, which implies that no information can be inferred from a known missed detection towards finding the correct one. For the specific

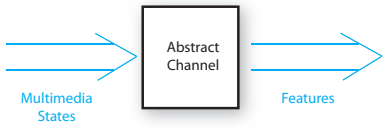


Fig. 5. Abstract channel between multimedia states and features in the information theoretical framework for operation forensics.

setting of this work, this condition means that, given a wrong estimated multimedia state, probabilities of the true multimedia state being any other multimedia states are the same.

3) $I(X;\hat{X}) = I(X;\underline{Y})$, i.e., $X \to \hat{X} \to \underline{Y}$ is also a Markov chain. This implies that, the estimated input contains all information that the real input has about the channel output. For the specific setting of this work, it means that the distribution of features given an estimated multimedia state will not change if the real multimedia state is also known.

In addition, with the assumption of uniform prior for X, which is commonly used in forensic analysis, the error probability lower bound will be only dependent on forensicability:

$$H(P_e^0) + P_e^0 \log_2(M-1) = \log_2 M - F(X;\underline{Y}). \quad (8)$$

Note that, while uniform priors are adopted in this paper, cases with non-uniform priors can be similarly handled by using the initial equation (3) instead of (8).

C. Expected Perfect Detection

While the lower bound of error probability gives fundamental limit on estimators' performance, we also want to answer the question of "when cannot we detect any more operations?" For example, in the multiple compression detection problem discussed earlier, we may want to know how many compressions we can detect at most. To answer these questions, we need a criterion to make decisions on whether we can or cannot detect more. One may suggest to check the equality of $F(X;\underline{Y}) \leq H(X)$. If equality holds, then there exists some estimator which can distinguish all considered multimedia states with zero error probability. Otherwise, it implies that not all multimedia states can be distinguished with zero error probability by any estimator. However, theoretically, this equality will never hold as long as the channel is not perfect. In other words, the error probability can never be zero and perfect detection never exists in theory. Therefore, the question becomes "how small should the error probability be so that we can still consider it as a perfect detection?"

Such a question leads us to examine the relationship between theoretical and experimental results. Given a rare incident, i.e., the probability that this incident happens tends to zero, it is very likely that we will not observe it in real experiments. Therefore, if the theoretical error probability is small enough, then we may not see the occurrence of error within a limited number of observations. Inspired by this idea, we reformulate the process of experimental testing as follows.

Given an image that may belong to any multimedia state considered in the analysis, there is probability P_e that the image will be misidentified. When we experimentally evaluate the performance of a detector on a database, we go through the following steps. First, an image is picked from a database containing images of all possible multimedia states. Then the detection scheme is applied on this image to obtain an estimated multimedia state. Lastly, by comparing the estimated multimedia state with the ground truth, we know whether the detection was correct or not. Given that nothing is known until the last step, each image is treated equally during estimation. By iterating these steps for every image in the database, the experimental error probability can be calculated as the total number of misclassifications divided by the size of the database. This process can be considered as a sequential process, where each time an image is randomly picked and its multimedia state is estimated by a detector, whose theoretical detection error probability is P_e. Then, by definition of P_e, for each individual detection, the tested image has probability P_e of being misidentified and probability $1 - P_e$ of being correctly detected. From this formulation, we can see an analogy between the process of experimental testing and a Bernoulli process.

Motivated by the discussion above, we model each sample in the testing database as an independent and identical Bernoulli random variable with probability P_e of missed detection. It is well known in probability theory that, the expected time of the first occurrence of missed detection happens at $1/P_e$. In other words, if the experimental database only has $S < 1/P_e$ samples, then the missed detection may not occur in expected sense, where the expectation is taken among all databases with the same size S. Thus, we propose the definition of *expected perfect detection* as follows.

Definition 2: Given an experimental database of size S, the *expected perfect detection* happens if and only if the theoretical error probability satisfies $P_e < 1/S$.

Based on this definition, a simple corollary below can give us the criterion to determine when we cannot detect any longer.

Corollary 1: For an experimental database of size S, if the lower bound of error probability obtained from (3) satisfies $P_e^0 > 1/S$, then no expected perfect detection can be obtained for any estimators.

We note that all above analysis is based on the law of large number. Experimentally, we find that the size of the database needs to be at the order of thousands for the expected perfect detection argument being hold. Fortunately, most experimental databases used in forensic analysis satisfy this condition.

III. Information Theoretical Modeling for JPEG Compression Forensics

To demonstrate the effectiveness of our proposed framework for operation forensics, we use the multiple JPEG compressions detection forensics as an example [23].

A. Background on JPEG Compression Forensics

An image's JPEG compression history is forensically important because it helps investigators to identify the image's acquisition process and detect possible manipulations [24], [25]. Specifically, by estimating the quantization table of a singly compressed image, one can identify the model of the camera that captured the image [24]. Furthermore, when a forger manipulates a JPEG image and re-saves it in the same format, double JPEG compression fingerprints may left in the image [11], [25]–[28]. The more times the JPEG image is manipulated, the more times of JPEG compressions it may go through. Thus, detecting the number of JPEG compressions that an image has gone through can help investigators to

understand how much the image has been tampered. However, as the number of JPEG compressions increases, the multiple compression fingerprints become less distinguishable [16], [17]. So a natural question would be "how many JPEG compressions can we detect, at most?"

Before applying our information theoretical model to answer this question, let us first review the typical process of a JPEG compression. When JPEG compressing an image, block-wise DCT transform is first applied on the pixel domain to obtain coefficients in DCT domain. Then, these coefficients are quantized and encoded by an entropy coder to get the JPEG data file. Whenever the image is edited or processed, decompression is needed, which follows the reverse procedure of compression. During decompression, the quantized DCT coefficients cannot be recovered. Thus, by examining the difference of DCT coefficients between uncompressed and compressed images, one can observe important fingerprints of JPEG compression. Furthermore, multiple JPEG compressions can also be detected by examining these coefficients.

Let D_0 denote a coefficient of a certain DCT subband of an uncompressed image. We use the Laplacian model to characterize the distribution of D_0 [29], where

$$\mathbb{P}(D_0 = \rho) = \frac{\lambda}{2} e^{-\lambda|\rho|} \triangleq f_\lambda(\rho), \quad \rho \in \mathbb{R}. \tag{9}$$

During JPEG compression, let a_1 be the quantization step used in this subband, and D_1 denote the DCT coefficient after compression, then

$$D_1 = \text{round}\left(\frac{D_0}{a_1}\right) \cdot a_1. \tag{10}$$

Thus, D_1 has a discrete distribution of

$$\begin{aligned} \mathbb{P}(D_1 = l_1 a_1) &= \int_{(l_1-1/2)a_1}^{(l_1+1/2)a_1} f_\lambda(\rho) d\rho, \quad l_1 \in \mathbb{Z}, \\ &= \begin{cases} 1 - e^{-\lambda a_1/2}, & \text{if } l_1 = 0, \\ e^{-\lambda|l_1 a_1|} \sinh\left(\frac{\lambda a_1}{2}\right), & \text{if } l_1 \neq 0. \end{cases} \end{aligned} \tag{11}$$

By examining the DCT coefficient histogram, investigators can detect whether the image is singly compressed or not. Furthermore, quantization step sizes can also be estimated if the image is detected as a singly compressed one [24].

When recompressing this singly compressed image using quantization step of $a_2, a_2 \neq a_1$, in the examined subband, let D_2 denote the DCT coefficient after two compressions, then we have

$$D_2 = \text{round}\left(\frac{D_1}{a_2}\right) \cdot a_2 = \text{round}\left(\text{round}\left(\frac{D_0}{a_1}\right) \cdot \frac{a_1}{a_2}\right) \cdot a_2, \tag{12}$$

and

$$\mathbb{P}(D_2 = l_2 a_2) = \sum_{(l_2-\frac{1}{2})a_2 \leq l_1 a_1 < (l_2+\frac{1}{2})a_2} \mathbb{P}(D_1 = l_1 a_1), \quad l_2 \in \mathbb{Z}. \tag{13}$$

Due to the effect of double quantization, the histogram of D_2 will present periodic characteristics, either periodic peaks or periodic zeros. Then, by examining the Fourier transform of the histogram, investigators can distinguish between singly compressed images and doubly compressed images [11], [25], [26].

B. DCT Coefficients Feature Model

Given that the histogram of DCT coefficients is a commonly used feature to detect JPEG compressions, in this example, we examine the fundamental limit of using DCT coefficient histograms to detect multiple JPEG compressions. We note that, other features used to detect JPEG compressions can be analyzed by similar approaches. As it is shown in Fig. 3, we consider an abstract channel where the input $X \in \{1, 2, ..., M\}$ is the number of JPEG compressions and the output $\underline{Y}$ is the DCT coefficient histogram written in a vector form.

To demonstrate the relationship between X and $\underline{Y}$, we take one subband as an illustration. We use λ to denote the parameter of the Laplace distribution of the coefficient D_0 in this subband when it is not compressed (9). Let $\mathcal{Q}_M = (q_1, q_2, ..., q_M)$ denote all possible quantization step sizes that may be used for this subband during compressions. Since in multiple compression detection forensics, the given image is a JPEG image and investigators try to detect how many compressions have been done before this last one, we keep the last compressions the same for all hypotheses. Without loss of generality, we take q_M as the quantization step size used in the last compression for all hypotheses. Then, if there are actually m applications of JPEG compressions, the DCT coefficient should have been quantized by step sizes $\{q_{M-m+1}, q_{M-m+2} ..., q_M\}$ in order. Let D_m denote the DCT coefficients if m times of JPEG compressions are applied. By following the analysis in (12), we have,

$$D_m = \text{round}\left(... \text{round}\left(\text{round}\left(\frac{D_0}{q_{M-m+1}}\right) \times \frac{q_{M-m+1}}{q_{M-m+2}} \right) \right) \times q_M. \tag{14}$$

Given this equation and (9), we can derive the distribution of D_m, which only has nonzero values at integer multiples of q_M. Let vector $\underline{v}_m(\lambda, \mathcal{Q}_M)$ denote this theoretical distribution, with each element $v_{n,m}(\lambda, \mathcal{Q}_M)$ representing the nonzero probability mass function $\mathbb{P}(D_m = nq_M)$, then

$$\underline{v}_m(\lambda, \mathcal{Q}_M) = \big[\mathbb{P}(D_m = -Nq_M), ..., \mathbb{P}(D_m = Nq_M)\big]. \tag{15}$$

In reality, however, we may not observe the theoretical distribution from the DCT histogram due to the model mismatch and/or the rounding and truncation in the compression and decompression. Instead, the normalized DCT coefficient histogram that we observe may be a noisy version of the theoretical distribution. Let random variable $\underline{Y}_m(\lambda, \mathcal{Q}_M)$ denote the observed normalized histogram if m applications of JPEG compressions were applied, i.e.,

$$\underline{Y}_m(\lambda, \mathcal{Q}_M) = [B_m(-Nq_M), ...B_m(Nq_M)], \tag{16}$$

where $B_m(nq_M), -N \leq n \leq N$, denotes the normalized histogram bin at location nq_M when m times of compressions happened. Then, by assuming that the observation noise, denoted by $\underline{W}$, is an additive noise, we have

$$\underline{Y}_m(\lambda, \mathcal{Q}_M) = \underline{v}_m(\lambda, \mathcal{Q}_M) + \underline{W}. \tag{17}$$

Let random variable $\underline{V}(\lambda, \mathcal{Q}_M) \in \{\underline{v}_1(\lambda, \mathcal{Q}_M), \underline{v}_2(\lambda, \mathcal{Q}_M), \ldots, \underline{v}_M(\lambda, \mathcal{Q}_M)\}$ denote the

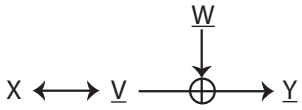


Fig. 6. Abstract channel inner structure for the model in Fig. 3.

theoretical distribution of DCT coefficients. Then, for a certain subband, given a fixed λ and $\mathcal{Q}_M$, the abstract channel in Fig. 3 can be depicted as the diagram in Fig. 6. Specifically, for each hypothesis on the number of JPEG compression X, it dictates a theoretical distribution on DCT coefficients $\underline{V}$, which can be calculated by (9) and (14). But due to the observation noise $\underline{W}$, the obtained normalized DCT coefficient histogram is $\underline{Y}$ in (17).

C. Forensicability for JPEG Compression Forensics

Based on our information theoretical framework, forensicability of using DCT histogram to detect multiple JPEG compressions is

$$F_{\lambda,\mathcal{Q}_M}(X;\underline{Y}) = I\big(X;\underline{Y}(\lambda,\mathcal{Q}_M)\big). \tag{18}$$

To calculate forensicability, we first assume that the observation noise on different histogram bins are independent with each other, then the variance of $\underline{W}$ is a diagonal matrix. Furthermore, based on experimental results, which will be shown in Section IV, we use the multivariate Gaussian distribution to model the observation noise as follows

$$\underline{W}(\lambda,\mathcal{Q}_M) \sim \mathcal{N}\Big(\underline{d}, diag\big(\beta \underline{V}^{2\alpha}(\lambda,\mathcal{Q}_M)\big)\Big), \tag{19}$$

where $\underline{d}$, $\beta > 0$ and $\alpha > 0$ are constant parameters, which will be estimated later. We note that, in our model, the variance of observation noise, Var($\underline{W}$), is proportional to the signal $\underline{V}$ that the noise is added on. This is because that the model mismatch and the rounding and truncation effect in the compression and decompression are more obvious on significant histogram bins.

In this example, we consider the case where we have no biased information on how many compressions that the image might have gone through, i.e., X has equal probability of being any value in $\{1,2,\ldots,M\}$. Then, given (17), (18) and (19), we can derive the forensicability of using DCT histogram to detect multiple JPEG compressions as the following expression

$$F_{\lambda,\mathcal{Q}_M}(X;\underline{Y}) = \log_2 M - \frac{1}{M}\sum_{m=1}^{M}\mathbb{E}\bigg[\log_2\sum_{j=1}^{M} exp\Big(\Phi_j^m(\underline{V})\Big)\bigg], \tag{20}$$

where

$$\Phi_j^m(\underline{V}) = \sum_{n=-N}^{N}\bigg[\alpha\ln\frac{v_{n,m}}{v_{n,j}} - \frac{(Y_n - v_{n,j})^2}{2\beta v_{n,j}^{2\alpha}} + \frac{(Y_n - v_{n,m})^2}{2\beta v_{n,m}^{2\alpha}}\bigg]. \tag{21}$$

Note that the right hand side expression in (20) and (21) still depend on λ and $\mathcal{Q}_M$. We remove these dependencies from variables in the sequel to simplify the expression. It is also noticed from (20) and (21) that forensicability does not depend on the constant mean $\underline{d}$ of the observation noise. This is because that any constant deviation of the output can be directly subtracted from input without any effect on the channel performance.

Before calculating forensicability, we need to estimate parameters β and α in the variance of observation noise (19). Based on (17) and (19), we apply maximum likelihood estimator to obtain the optimal β and α. Given that $\underline{d}$ has no effect on forensicability, we first derive the estimator for $\underline{d} = \underline{0}$. Let $Y_{\lambda_i,n,m}$ denote the n^{th} histogram bin of the i^{th} image (whose Laplace parameter is λ_i) after m times of compressions. Then, the optimal β and α are

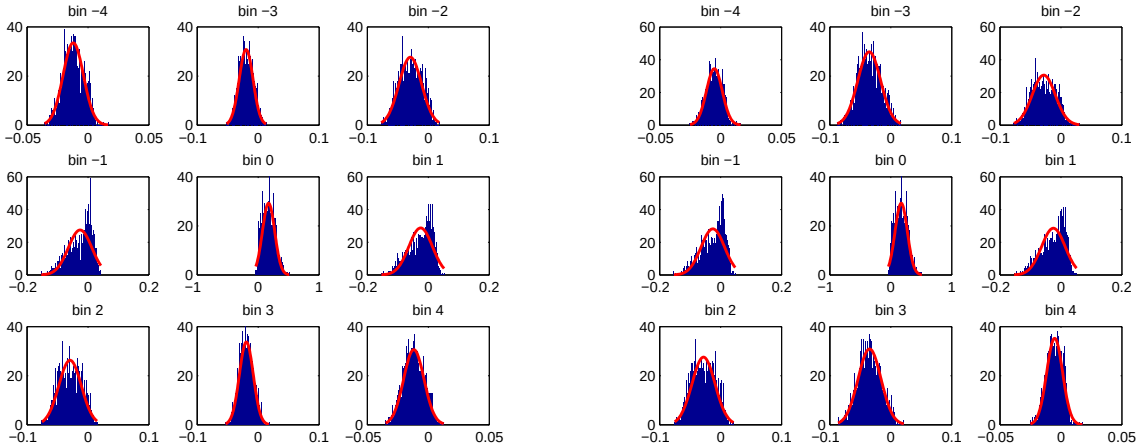
$$(\hat{\beta},\hat{\alpha}) = \underset{\beta>0,\alpha>0}{\arg\max}\log\sum_{i=1}^{K}\sum_{n=-N}^{N}\sum_{m=1}^{M}\mathbb{P}(Y_{\lambda_i,n,m} = y_{\lambda_i,n,m}). \tag{22}$$

According to Karush-Kuhn-Tucker conditions, we have

$$\begin{cases} \displaystyle\sum_{i=1}^{K}\sum_{n=-N}^{N}\sum_{m=1}^{M}(y_{\lambda_i,n,m} - v_{\lambda_i,n,m})^2\ln v_{\lambda_i,n,m}\Big(\frac{1}{v_{\lambda_i,n,m}}\Big)^{2\hat{\alpha}} \\ \qquad\qquad = \hat{\beta}\displaystyle\sum_{i=1}^{K}\sum_{n=-N}^{N}\sum_{m=1}^{M} v_{\lambda_i,n,m}, & (23) \\ \displaystyle\sum_{i=1}^{K}\sum_{n=-N}^{N}\sum_{m=1}^{M}\frac{(y_{\lambda_i,n,m} - v_{\lambda_i,n,m})^2}{v_{\lambda_i,n,m}^{2\hat{\alpha}}} = \hat{\beta}K(2N+1)M. & (24) \end{cases}$$

Given that the theoretical distribution $v_{\lambda_i,n,m} \in [0,1]$, the left hand side of (23) is monotonically increase with $\hat{\alpha}$. Then $\hat{\alpha}$ can be approximated for any given $\hat{\beta}$. In addition, from (24), $\hat{\beta}$ can be derived for any fixed $\hat{\alpha}$. Thus, an iterative algorithm can be used to obtain the optimal $\hat{\beta}$ and $\hat{\alpha}$ from (23) and (24). For $\underline{d} \neq \underline{0}$ cases, similar estimators can be derived with $y_{\lambda_i,n,m}$ substituted by $y_{\lambda_i,n,m} - d_n$, where $d_n, n \in [-N,N]$, is the n^{th} element in $\underline{d}$.

Lastly, we note that, as the first work proposing and calculating forensicability in operation forensics, JPEG compression forensics has been chosen as it is a well studied problem in literature. Furthermore, the existing model of DCT coefficient histograms has helped us simplify the analysis of channel characteristics. Nevertheless, similar approaches can be applied to other forensic problems to find their fundamental limit of forensicability. For example, in contrast enhancement detection [10], the input of the channel is either unaltered, i.e., $X = 0$, or contrast enhanced, i.e., $X = 1$. The extracted feature can be taken as the high frequency component of the image pixel histogram. Then, similar approaches can be applied to model the relationship between features and multimedia states. Forensicability can also be calculated to imply the best performance one can possibly obtain. Furthermore, by comparing the forensicability of contrast enhancement detection and those of other detections, such as resizing detection [8], one can find which manipulation is fundamentally easier to be detected. In addition, our framework may also be used to explore the fundamental limit of detecting the order of manipulation operations [15]. In this case, multimedia states would be any combinations of considered operations, and features can be built by concatenating all useful features for distinguishing the order of these operations.



(a) (b)

Fig. 7. Histograms of observation noise and their estimated Gaussian distributions (plotted in red lines) on different histogram bins for (a) single compressed images with quantization step size of 6 in the examined subband and (b) doubly compressed images with quantization step size of 6 then 7 in the examined subband. Bin i means that the observation noise on normalized histogram bin $B(iq_{last})$ is examined, where q_{last} denotes the last quantization step size.

IV. Data-Driven Results and Analysis

In this section we provide experimental support for our proposed framework and calculate the forensicability for JPEG compression forensics. From analyzing forensicability, we are able to answer how many JPEG compressions, at most, that investigators can detect. Furthermore, we also examine the effect of compression quality factors and different DCT subbands on forensicability in order to provide guidance of strategies for both investigators and forgers.

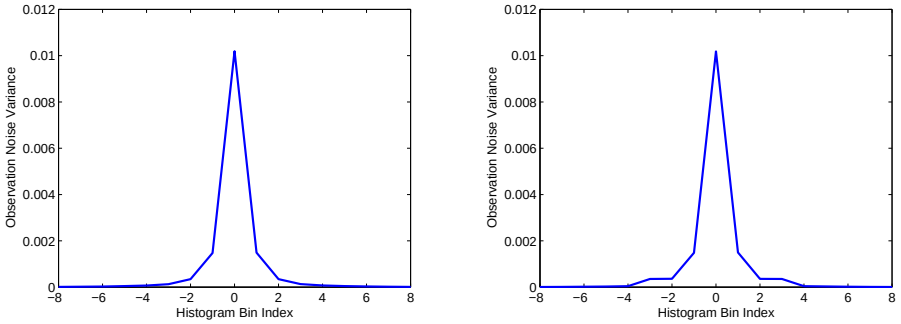
A. Verification of Observation Noise Model

To support our proposed observation noise model in (19), we conduct an experiment to examine the difference between observed normalized histograms and their theoretical distributions. Our test images are generated from the 1338 uncompressed images from UCID database [30]. We first create the 1338 singly compressed images by JPEG compressing the uncompressed images using quality factor of 80. We examine the $(2, 3)$ subband, where the corresponding quantization step size is 6. Double compressed images are also examined for verification, where we obtain these test images by double JPEG compressing the uncompressed 1338 images using quality factors 80 and then 75. The corresponding quantization step sizes for the examined subband are 6 and 7 respectively. The observed normalized histograms are obtained directly from these two sets of compressed images. We calculate the theoretical distributions for singly compressed images and doubly compressed images based on their uncompressed versions. Specifically, for each of the 1338 images, we first estimate the Laplace parameter λ based on the DCT coefficients of the uncompressed image. Then the theoretical distribution is calculated according to (14) and (15) for given λ and quantization step sizes. Observation noise is calculated by subtracting the theoretical distributions from the observed normalized histograms.

Fig. 7 plots the histograms of observation noise and their estimated Gaussian distributions for different histogram bin locations for both singly compressed images and doubly compressed images. From these results, we can see that Gaussian distributions can well approximate the distributions of the observation noise for most of cases. Furthermore, the mean of the histograms does not change much between singly compressed images and doubly compressed images. This gives support on our constant mean model of the observation noise.

Fig. 8 plots the variance of observation noise for different histogram bin locations for both singly compressed images and doubly compressed images. Given the discussion in Section III-A, the DCT coefficient distribution of singly compressed image is quantized Laplace distribution. Although different images have different Laplace parameters and their DCT coefficient distributions may be different, these distributions share a common shape of having a central peak at zero and decreasing fast as the absolute value of the variable increases. The observation noise variance of singly compressed images exhibits similar characteristics as it is plotted in Fig. 8(a). Furthermore, for double compressed images where the second quality factor is lower than the first one, double compression fingerprints of periodic peaks will be presented in DCT coefficient histograms. Similar fluctuation of the observation noise variance is shown in Fig. 8(b). Therefore, both figures in Fig. 8 show that the variance of observation noise changes in the similar way as the value of theoretical distribution changes. In other words, these experimental results show that the variance of observation noise is proportional to the theoretical distribution. This validates the proposed variance model of the observation noise in (19). Furthermore, instead of using a linear model, an exponential proportionality principle is adopted in the variance model to make it more general.

We note that there may be more accurate but complicated models for the observation noise. We use the model in (19) as a tradeoff between the accuracy of modeling and the complexity



(a) (b)

Fig. 8. Variance of observation noise versus histogram bin index for (a) single compressed images with quantization step size of 6 in the examined subband; and (b) doubly compressed images with quantization step size of 6 then 7 in the examined subband.

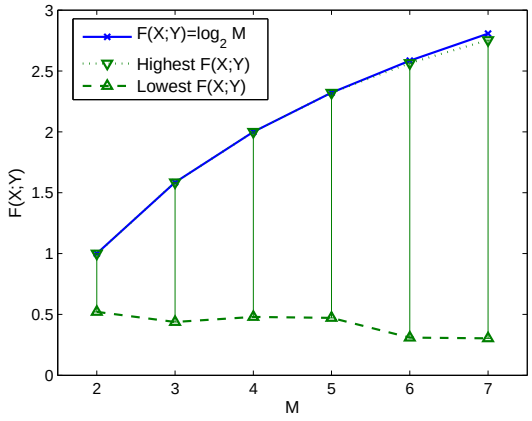


Fig. 9. The reachable forensicabilities of different compression quality factors $\mathcal{Q}_M$ and the upper bound of forensicability for different M's.

of analysis.

B. Forensicability Calculation

In order to calculate forensicability, we first estimate parameters β and α from (23) and (24). We use the normalized DCT coefficient histograms of singly compressed images and their corresponding theoretical distributions obtained from last subsection to estimate. Due to the nonzero mean of observation noise, we subtract this mean from the observed normalized histograms before using them in (23) and (24). Then, we exclude insignificant histogram bins due to the severe noise effect on those small histogram bins. Specifically, we use those normalized histogram bins whose theoretical probabilities are equal or greater than 5×10^{-4}. This results in total 36298 histogram bins used for estimation. The estimated parameters' values are

$$\hat{\beta} = 0.0494, \quad \hat{\alpha} = 0.744. \tag{25}$$

Given β and α, forensicability of multiple JPEG compression forensics can be obtained from (20) and (21). Since (20) is not a closed form and we cannot calculate the precise value, we use Monte Carlo simulation to approximate the result. This is a commonly used method in information theoretic analysis [31]. We demonstrate the results for subband $(2, 3)$, where we take a typical value of $\lambda = 0.1$. We find that the quantization step size in this subband changes from 1 to 14 when varying the JPEG compression quality factor from 50 to 100. By excluding the trivial cases where one quantization interval is an integer multiple of another, we choose the candidate quantization step sizes as

$$\{5, 6, 7, 8, 9, 11, 13\}. \tag{26}$$

Then, for each M, we randomly select values from this candidate set to construct $\mathcal{Q}_M$, under the constraint that two adjacent elements are not equal.

For each different $\mathcal{Q}_M$, $F_{\lambda,\mathcal{Q}_M}(X;\underline{Y})$ is estimated by Monte Carlo averaging and plotted in Fig. 9. The green lines with triangle ending points show the range of all possible forensicabilities at each M for different $\mathcal{Q}_M$'s. As we can expect, quantization step sizes play an important role in determining forensicabilities. We will analyze this effect in later sections. In Fig. 9, we also plot the line of $F_{\lambda,\mathcal{Q}_M}(X;\underline{Y}) = \log_2 M$, which is the upper bound of forensicability for uniform priors, indicating perfect detection. Despite variations of forensicabilities for different $\mathcal{Q}_M$'s, the gap between the highest reachable forensicability and its upper bound becomes more obvious when M increases. This indicates that, as M increases, even when we encounter the scenario with the highest forensicability, i.e., the case having the best detection performance, we still cannot obtain perfect detection. Furthermore, the distance of the best performance to perfect detection will be larger with the increase of M. Therefore, when M increases, it will be much harder to detect all hypotheses, which verifies our theory.

C. Estimation Error Probability Lower Bound

According to theorem 1, forensicability determines the lower bound of error probabilities. In this section, we perform an experiment to examine the effectiveness of the lower bound by comparing the theoretical lower bound of all possible error probabilities with the experimental error probability obtained from a specific estimator. We perform this comparison on two examples of $\mathcal{Q}_{20}$, which are constructed by randomly selecting quantization step sizes from the candidate set in (26).

The experimental error probabilities of a specific estimator are obtained as follows. For each $M \in [2, 20]$, $\mathcal{Q}_M$ is obtained

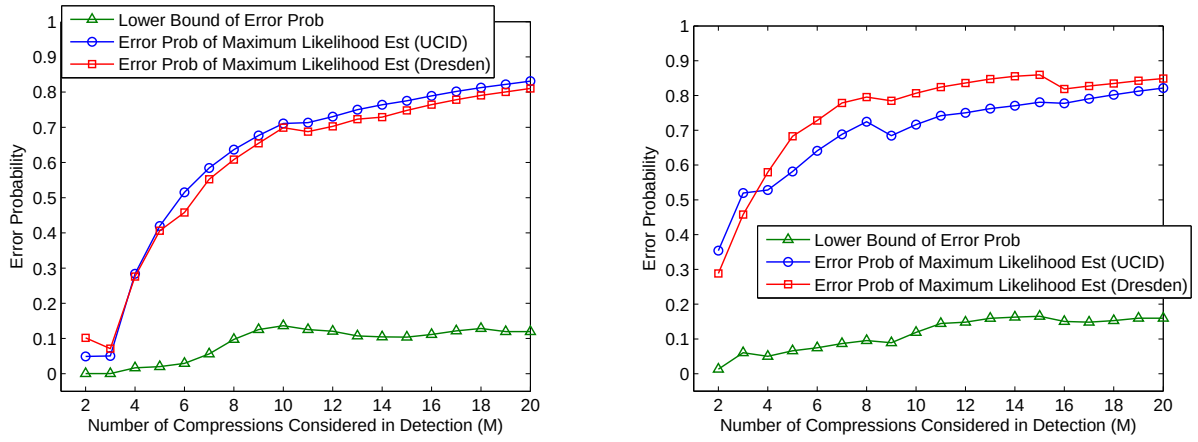


Fig. 10. Experimental error probabilities of a maximum likelihood estimator on two different image databases, compares with the theoretical lower bound of error probabilities for all estimators, when two randomly selected $\mathcal{Q}_{20}$'s are taken as examples: (a) $\mathcal{Q}_{20} = \{..., 8, 11, 13, 6, 5\}$ and (b) $\mathcal{Q}_{20} = \{..., 11, 9, 7, 8, 13\}$. Experimental error probabilities are obtained for both UCID images (plotted in blue circles) and images from Dresden database (plotted in red squares).

as the last M quantization step sizes in $\mathcal{Q}_{20}$. The 1338 uncompressed images from the UCID database are first used to construct a test database. Specifically, for each M, we JPEG compress each of the 1338 images M times using quality factors, whose quantization step sizes in the $(2, 3)$ subband are $\{q_{20-M+1}, \ldots, q_{20}\}$. The resulting 1338 images compose the data set of M times compressed images. Then, normalized DCT coefficient histograms in the $(2, 3)$ subband are extracted for analysis. Their theoretical distributions are also calculated based on $\mathcal{Q}_{20}$ and the estimated λ's from their uncompressed versions.

Given the assumption of uniform priors and the proposed conditional distribution of a normalized histogram given the theoretical distribution in (17) and (19), we use the maximum likelihood estimator to estimate the number of compressions for each M. This estimator is optimal for minimizing the mean square error of the estimation. Specifically, when M hypotheses of X are considered in the system, let m be the actual number of compressions that an image has gone through. Its normalized DCT coefficient histogram is denoted as $\underline{y}_m$. Then the maximum likelihood estimator for m is

$$\hat{m} = \underset{1 \le m^* \le M}{\arg\max} \ \mathbb{P}(\underline{Y}_{m^*} = \underline{y}_m), \tag{27}$$

where the distribution of $\underline{Y}_{m^*}$ is given in (17) and (19). The error probabilities of this estimator are plotted in Fig. 10 using blue circles.

To examine the experimental result for different databases, we also used the Dresden Image Database [32] to obtain the experimental error probabilities, which are plotted in Fig. 10 using red squares. This database contains 1491 unprocessed images, with each has size of 2000×3008 or larger. Unlike the UCID, the Dresden Image Database has a small portion of images (31 images) that have low visual qualities due to any of the following reasons: overexposure, underexposure, out of focus, image defects, and camera shaking. The maximum likelihood estimator was applied on this database in the same way that is applied on the UCID database. We can see that the detector performs similarly on these two databases.

Then, for every M, the theoretical lower bound of error probabilities is calculated for each image, i.e., each estimated λ, using (3), then we take the mean value and plot it in Fig. 10 using green triangle.

Both examples in Fig. 10(a) and Fig. 10(b) show that the error probability of the specific estimator is higher than the theoretical lower bound, which verifies the validity of our proposed lower bound. For the example in Fig. 10(a), when $M < 4$, the experimental results are approximate to their lower bounds, which means that using maximum likelihood estimator for this case can yield the best performance. The occurrence of experimental results being very close to the theoretical lower bounds also proves the effectiveness of our proposed error probability lower bound. For the example in Fig. 10(b), the experimental result is much worse than that in Fig. 10(a), even when detecting double compressions, i.e., $M = 2$. This matches the results in forensic literatures of detecting double compressions, which shows difficulty when the detected image has a secondary compression quality factor lower than the primary one [11]. The distance between the experimental error probability of one specific estimator and the theoretical error probability lower bound of all estimators suggests the existence of better estimators or better features.

D. Maximum Number of Detectable Compressions

Given the error probability lower bound, we can determine what is the maximum number of compressions investigators can detect by using corollary 1. First, based on Theorem 1, we use the highest reachable forensicability for each M to calculate the minimum lower bound of error probabilities for all possible compression quality factors. The calculation results are shown in Table I. From this table we can see that, for double compression detection where $M = 2$, the lower bound of error probability is approximately 0 (note that it is not exactly zero, it is just smaller than the precision of Matlab

processor), which matches the result of existing techniques [11]. Furthermore, the table shows that the minimum lower bound of error probability increases dramatically with M.

TABLE I
$\min_{\mathcal{Q}_M} P_e^0$ FOR DIFFERENT M.

M	2	3	4	5	6
$\min_{\mathcal{Q}_M} P_e^0$	0	3.9×10^{-9}	5×10^{-5}	2.1×10^{-4}	0.0016

Then, to determine the point where we cannot perfectly detect any more compressions, we adopt the concept of expected perfect detection defined in definition 2 and use the conclusion in corollary 1. For example, if the forensic investigator performs experiments on a test database of size $S = 5000$, then because $\min_{\mathcal{Q}_4} P_e^0 < 1/S = 2 \times 10^{-4}$ but $\min_{\mathcal{Q}_5} P_e^0 > 2 \times 10^{-4}$, we claim that no expected perfect detection exists for $M > 4$.

Furthermore, by noticing that

$$\frac{1}{\min_{\mathcal{Q}_4} P_e^0} = 20000, \qquad \frac{1}{\min_{\mathcal{Q}_5} P_e^0} = 4762, \tag{28}$$

we have the following conclusion. For any database of size bigger than 4762 and smaller than 20000, expectedly, no perfect detection can be achieved for detecting more than 4 times of JPEG compressions. In other words, for typical sizes of database, investigators can only detect up to 4 times of JPEG compressions using DCT coefficient feature.

We note that, since we are analyzing the minimum lower bound of error probability, which is the best performance we may get from all estimators and all compression quality factors, these results only provides an upper limit of investigators' capability. In other words, "cannot perfectly detect 5 compressions" does not mean "can perfectly detect 4 compressions for sure". Our theorem tells what we cannot do rather than what we can do.

It is also noted that, for databases bigger than 20000, the maximum number of compressions can be detected may be less than 4. It implies that the number of detectable compressions depends on the test database size. It is reasonable because, as the database size goes bigger, there will be higher probability that we may meet an instance that is hard to detect and thus error may occur.

E. Quality Factor Patterns having the Highest and Lowest Forensicabilities

As Fig. 9 shows, forensicability varies significantly with $\mathcal{Q}_M$. In order to characterize this effect, we examine all combinations of quantization step sizes and their forensicabilities. From there, we find the patterns of $\mathcal{Q}_M$ which will yield the highest and lowest forensicabilities, as they are shown in Fig. 11.

We find that, if the next compression always uses a higher quality factor than the previous one, forensicabilities will be the highest, i.e., they are easiest to be detected. Denote the set of quality factors yielding the highest forensicabilities as $\mathcal{Q}^h$, then

$$\mathcal{Q}^h = \{\mathcal{Q}_M | q_m < q_{m-1}, \forall 1 < m \leq M, M \in \mathbb{Z}^+\}. \tag{29}$$

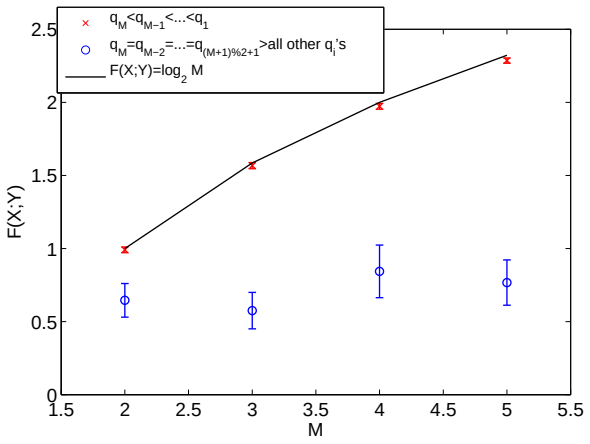


Fig. 11. Patterns of $\mathcal{Q}_M$ yielding the highest and lowest forensicabilities.

To explain this phenomenon, let us examine a DCT coefficient histogram of an image that has been compressed m times using decreasing quantization step sizes $a_1 > a_2 > \ldots > a_m$ in the concerned subband. Recall the discussion in Section III-A, the singly quantized coefficients D_1 obeys a quantized Laplace distribution with quantization step size a_1. Then, given that the next quantization step size is smaller than the current one, when re-quantizing this histogram, every bin will remain its original value but be shifted to its nearby integer multiple of a_2. Zeros may be introduced into the histogram of D_2, but all nonzero histogram bins will be the same as those in the histogram of D_1. Similar analysis applies for the following quantizations. Therefore, the normalized DCT histogram after m times of quantizations will have all of its nonzero bins being equal to those after the first quantization.

For detecting M times of compressions with quantization step sizes $\mathcal{Q}_M$, we are distinguishing the following M hypotheses on the DCT coefficient histogram:

$$\begin{cases} H_1 : 1 \text{ time of quantization by } q_M, \\ H_2 : 2 \text{ times of quantizations by } \{q_{M-1}, q_M\} \text{ in order}, \\ H_3 : 3 \text{ times of quantizations by } \{q_{M-2}, q_{M-1}, q_M\}, \\ \vdots \\ H_M : M \text{ times of quantizations by } \{q_1, q_2, \ldots, q_M\}. \end{cases} \tag{30}$$

It is easy to notice that, for different hypotheses, the first quantization step sizes are different. Thus, for case of $q_1 > q_2 > \ldots > q_M$, theoretically, the nonzero bins of the normalized histogram obtained from one hypothesis are completely different from those obtained from another hypothesis. Furthermore, there may also have cases where a location of a zero histogram bin in one hypothesis has a nonzero bin in another hypothesis. This will further enlarge the disparity of DCT histograms obtained from different hypotheses. Therefore, the complete distinguishability of theoretical distributions of DCT coefficients among different hypotheses results in the easiest detection and the highest forensicability.

The compression quality factors resulting in the lowest forensicabilities, as it is shown in Fig. 11, are those which use same quality factors periodically. More specifically, denote

the set of quality factors yielding the lowest forensicabilities as $\mathcal{Q}^l$. We have found that

$$\mathcal{Q}^l = \{\mathcal{Q}_M | q_M = q_{M-2} = \ldots = q_{(M+1)\%2+1} > \text{all other } q_i's, M \in \mathbb{Z}^+\}, \quad (31)$$

where $\%$ is a remainder operator. The reason can be explained by the following theorem.

Theorem 2: Given a quantized DCT coefficient D_{m-2} with the last quantization step size as q_{m-2}. We further quantize it two more times using quantization step sizes q_{m-1} then q_m. The obtained coefficient is denoted as D_m. If the quantization step sizes satisfy $q_m = q_{m-2} > q_{m-1}$, then the DCT coefficient remains the same after these two more compressions, i.e., $D_m \equiv D_{m-2}$.

Proof: Take any possible value of $D_{m-2} = l_{m-2}q_{m-2}$, where $l_{m-2} \in \mathbb{Z}$, after the two quantizations, we obtain

$$D_m = round\bigg(round\Big(\frac{l_{m-2}q_{m-2}}{q_{m-1}}\Big)\frac{q_{m-1}}{q_m}\bigg)q_m. \quad (32)$$

Given that $\forall A \in \mathbb{R}, A - 1/2 < round(A) \leq A + 1/2$, we have

$$\begin{aligned}
\frac{D_m}{q_m} &> round\bigg(\Big(\frac{l_{m-2}q_{m-2}}{q_{m-1}} - \frac{1}{2}\Big)\frac{q_{m-1}}{q_m}\bigg) && (33)\\
&= round\Big(l_{m-2} - \frac{1}{2}\frac{q_{m-1}}{q_m}\Big) && (34)\\
&> round\Big(l_{m-2} - \frac{1}{2}\Big) && (35)\\
&> l_{m-2} - 1, && (36)
\end{aligned}$$

where (34) and (35) are obtained from the condition $q_{m-2} = q_m > q_{m-1}$. Since $\frac{D_m}{q_m}$ is an integer, we obtain $\frac{D_m}{q_m} \geq l_{m-2}$. Similarly, we can prove that $\frac{D_m}{q_m} \leq l_{m-2}$. Thus,

$$D_m = l_{m-2}q_m = D_{m-2}. \quad (37)$$

■

Given the above theorem, the M hypotheses in (30) can be reduced to only singly quantized hypothesis and double quantized hypothesis. Specifically, all odd numbered hypotheses will be identical to each other. While all even numbered hypotheses will be simplified to 2 times of quantization with different primary quantization step sizes. Furthermore, for the simplified double quantization hypotheses, the second quantization step size is larger than the first one, which is harder for estimation compared to its opposite case. Therefore, such a pattern of compression quality factors is the hardest to be detected, and thus has the lowest forensicability. Moreover, since the estimation performance will always be similar to a double compression detection regardless of how many compressions investigators really want to detect, forensicability almost remains the same as M increases.

F. Optimal Strategies for Forgers and Investigators

The fundamental measurement of forensicability can also be used to obtain the optimal strategies for both investigators and forgers. In this multiple compression detection system, investigators try to detect the number of compressions forgers have done on an image. Thus, investigators can choose examined subbands to maximize forensicability, while forgers have the right of choosing compression quality factors to minimize forensicability. Given that forensicability is a function of both subband parameter λ and compression quality factors $\mathcal{Q}_M$, we model the optimal strategies for forensic investigators and anti-forensic forgers in this multiple compression detection system as

$$\delta_F = \arg\max_{(i,j)} \mathbb{E}_{\mathcal{Q}_M}\Big[F_{\lambda_{(i,j)},\mathcal{Q}_M}(X;\underline{Y})\Big], \quad (38)$$

$$\delta_{AF} = \arg\min_{\mathcal{Q}_M} \mathbb{E}_{\lambda_{(i,j)}}\Big[F_{\lambda_{(i,j)},\mathcal{Q}_M}(X;\underline{Y})\Big], \quad (39)$$

respectively, where $(i,j), i,j \in [1,8]$, denotes the subband index.

Since we have just discussed the effect of compression quality factors on forensicability, let us obtain the optimal strategy for forgers (39). From the discussion in last subsection, we notice that the patterns of compression quality factors yielding the highest and lowest forensicabilities do not depend on the subband parameter λ. Instead, the results are merely dependent on how the DCT coefficients are quantized. Thus, regardless of which subband or subbands investigators will choose, $\mathcal{Q}^l$ will always yield the lowest forensicability. Thus, we obtain the optimal strategy for forgers is

$$\delta_{AF} = \mathcal{Q}^l. \quad (40)$$

We note that, when $M = 2$, we have $\delta_{AF} = \mathcal{Q}^l = \{\mathcal{Q}_2 | q_1 < q_2\}$, which is opposite to the pattern of $\mathcal{Q}^h$. This result matches our early work on the concealability-rate-distortion tradeoff of compression anti-forensics, where we found that forgers would prefer to use a lower secondary quality factor instead of a higher one in their second compression [33].

To obtain the optimal strategy for investigators, we take $\lambda_{(i,j)}$ as the mean value of all estimated λ's from the $(i,j)^{th}$ subband coefficients of 1338 uncompressed images in the UCID database. We examine the cases of detecting 2, 3, 4 and 5 times of compressions, i.e., we take $M \in [2,5]$. For each M, $\mathcal{Q}_M$ for the $(2,3)$ subband is still constructed by randomly selecting quantization step sizes from the candidate set $\{5,6,7,8,9,11,13\}$ in (26). Given that the compression quality factors corresponding to these quantization step sizes are $\{82,78,75,70,67,60,55\}$, $\mathcal{Q}_M$ for other subbands can also be determined from their corresponding quantization tables. Then, for each of the 63 alternating current (AC) DCT subbands, forensicabilities are calculated for all $\mathcal{Q}_M$'s, whose number of possibilities can reach $(7\times 6^4 =)9072$ when $M = 5$. We assume that investigators do not know the priori of the compression quality factors used by forgers. Thus, for each subband, $\mathbb{E}_{\mathcal{Q}_M}\big[F_{\lambda_{(i,j)},\mathcal{Q}_M}(X;\underline{Y})\big]$ is calculated as the mean value of forensicabilities with respect to different $\mathcal{Q}_M$'s.

By comparing the expected value of forensicabilities for all 63 subbands, we order them in descending order and take the top 9 subbands to show in Fig. 12. Our results show that, the top 9 subbands yielding the highest forensicabilities remain the same when detecting different numbers of compressions, though their orders are slightly different. Thus, if investigators take the best 9 subbands for detection, their the optimal

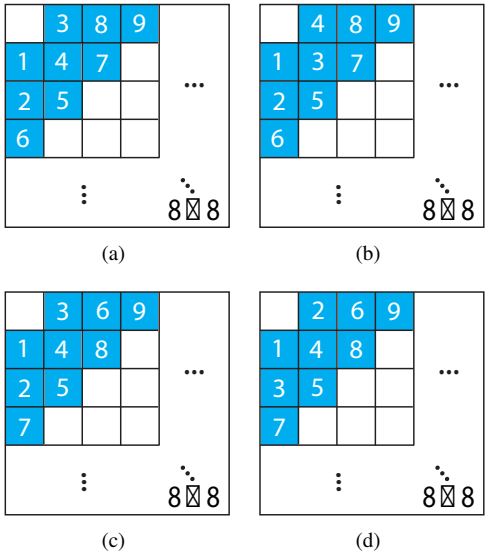


Fig. 12. The best 9 DCT subbands (shown as blue cells) for detection, which yield the highest forensicabilities for (a) $M = 2$, (b) $M = 3$, (c) $M = 4$ and (d) $M = 5$. Numbers 1 through 9 represent the order of these subbands regarding their forensicabilities from the highest to the lowest.

strategy, which is denoted as $\delta_F^{(9)}$, will be

$$\delta_F^{(9)} = \{(2,1),(1,2),(3,1),(1,3),(2,2), (4,1),(1,4),(3,2),(2,3)\}. \quad (41)$$

It matches the set of subbands that many successful double compression forensic techniques have used in their algorithms [11]. This result gives theoretical support of why we use those subbands for detecting double compressions. It also suggests that we should continue to use these subbands to detect 3, 4 or 5 times of compressions. Furthermore, the ranks on these subbands tell us which subband contains more forensic information and which one will give us the most trustful result.

G. Forensicabilities for Image Outliers

Given that forensicability depends on the Laplace parameter λ of DCT coefficients, it may also vary for different types of images. While our results were obtained by choosing a representative λ value and thus can be considered as the most expected performance for natural images, there are some outliers that are much harder or much easier to be detected. For example, if an image is underexposed and most of its pixels are equal to zero, then it would be very hard to detect the number of compressions on this image.

To track the change of the Laplace parameter λ for different images, we examine natural images from both the UCID database (1338 images) and the Dresden image database (1491 images). Fig. 13 shows the histogram of λ in the $(2,3)$ subband of these 2829 images. We can see that most images have their λ values close to 0.1, which was chosen as the representative value of λ in Section IV-D.

In order to examine forensicabilities for other images, we take two extreme cases of $\lambda = 0.02$ and $\lambda = 0.7$ to obtain the bounds of performance. Table II(A) and II(B) show the minimum error probability lower bound for different numbers

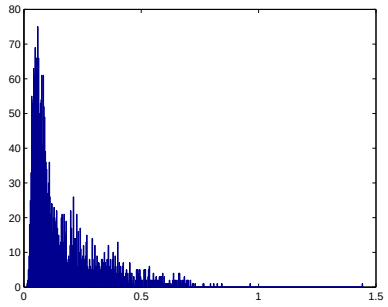


Fig. 13. Histogram of λ in subband (2,3) of images from UCID and Dresden databases.

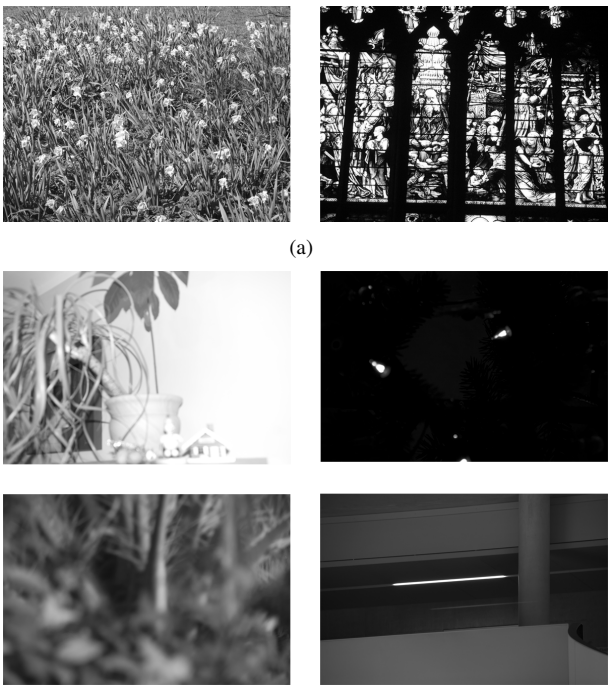


Fig. 14. Representative image outliers in UCID and Dresden databases with (a) $\lambda \cong 0.02$ and (b) $\lambda \geq 0.7$.

of compressions when $\lambda = 0.02$ and $\lambda = 0.7$, respectively. By comparing these two tables with Table I, we can see that the minimum lower bound of error probabilities $\min_{\mathcal{Q}_M} P_e^0$ increases with λ, and thus forensicability decreases with λ. This matches the results in the previous subsection where forensicability decreases for higher frequency subbands which have higher values of λ. This is because for large λ's, the DCT coefficient histograms have high kurtosis and low variances. Most bins in these histograms have small values that can be severely contaminated by noise. Only a few histogram bins have large enough values that can be used for estimation. Thus, little information can be extracted from these histograms. By following the analysis in Section IV-D we can infer that, if we have a database of size 10000, then for image outliers whose $\lambda = 0.02$, investigators can detect up to 7 times of compressions. While for image outliers whose $\lambda = 0.7$, we can only detect 2 times of compressions.

TABLE II
$\min_{\mathcal{Q}_M} P_e^0$ FOR DIFFERENT M WHEN (A) $\lambda = 0.02$ AND (B) $\lambda = 0.7$.

(A)

M	2	3	4	5	6	7	8
$\min_{\mathcal{Q}_M} P_e^0$	0	0	1.9×10^{-9}	1.1×10^{-7}	2.2×10^{-6}	3.7×10^{-5}	5.5×10^{-4}

(B)

M	2	3
$\min_{\mathcal{Q}_M} P_e^0$	1.4×10^{-5}	0.0018

Lastly, in order to see what types of images are outliers, we select some representative images from each extreme case and show them in Fig. 14. As it is shown in Fig. 14(a), the outliers having the lowest λ's and the highest forensicabilities are highly textured images whose AC components have sufficient information to be used for forensic detection. On the other hand, the outliers having the highest λ's and the lowest forensicabilities are images having a large amount of smooth or uniform areas but few textured regions. As it is shown in Fig. 14(b), this phenomenon may be caused by overexposure, underexposure, strong blurring, or little textured image content.

V. CONCLUSION

In this paper, we proposed an information theoretical framework to explore the fundamental limit of operation forensics. In this framework, we defined forensicability in operation detection forensics as the maximum information that features contain about operations. Based on this measure, we obtained the lower bound of error probabilities for any estimators using these features. Furthermore, by introducing the concept of expected perfect detection, we were able to determine the limit of how many operations we can successfully detect. To show the effectiveness of our framework, we applied it to the case of detecting multiple JPEG compressions using DCT coefficient histogram features. By appropriate modeling of the features, we calculated forensicabilities and concluded that, under typical settings of forensic analysis where the size of the testing database is less than 20000, at most 4 times of compressions were detectable. Furthermore, based on this fundamental measurement, we found the patterns of compression quality factors holding the highest and lowest forensic information. Lastly, the optimal strategies for investigators and forgers were discussed using forensicability.

REFERENCES

[1] M. C. Stamm, M. Wu, and K. J. R. Liu, "Information forensics: An overview of the first decade," *Access, IEEE*, vol. 1, pp. 167–200, 2013.

[2] A. Swaminathan, M. Wu, and K. J. R. Liu, "Component forensics," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 38–48, Mar. 2009.

[3] X. Chu, M. C. Stamm, W. S. Lin, and K. J. R. Liu, "Forensic identification of compressively sensed images," in *Proc. IEEE ICASSP*, 2012, pp. 1837–1840.

[4] W. S. Lin, S. K. Tjoa, H. V. Zhao, and K. J. R. Liu, "Digital image source coder forensics via intrinsic fingerprints," *IEEE Trans. on Information Forensics and Security*, vol. 4, no. 3, pp. 460–475, Sep. 2009.

[5] J. Lukáš, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Trans. on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, Jun. 2006.

[6] A. Swaminathan, M. Wu, and K.J.R. Liu, "Nonintrusive component forensics of visual sensors using output images," *IEEE Trans. on Information Forensics and Security*, vol. 2, no. 1, pp. 91–106, Mar. 2007.

[7] R. Garg, A. L. Varna, and M. Wu, ""seeing" enf: natural time stamp for digital video via optical sensing and signal processing," in *Proceedings of the 19th ACM international conference on Multimedia*, New York, NY, USA, 2011, MM '11, pp. 23–32, ACM.

[8] A.C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of re-sampling," *IEEE Trans. on Signal Processing*, vol. 53, no. 2, pp. 758–767, Feb. 2005.

[9] M. Kirchner and J. Fridrich, "On detection of median filtering in digital images," *Media Forensics and Security II, Proc. of SPIE-IS&T Electronic Imaging, SPIE*, vol. 7541, 754110, 2010.

[10] M. C. Stamm and K. J. R. Liu, "Forensic detection of image manipulation using statistical intrinsic fingerprints," *Information Forensics and Security, IEEE Transactions on*, vol. 5, no. 3, pp. 492–506, 2010.

[11] T. Pevný and J. Fridrich, "Detection of double-compression in JPEG images for applications in steganography," *IEEE Trans. on Information Forensics and Security*, vol. 3, no. 2, pp. 247–258, Jun. 2008.

[12] M. C. Stamm, W. S. Lin, and K. J. R. Liu, "Temporal forensics and anti-forensics for motion compensated video," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 4, pp. 1315–1329, 2012.

[13] H. Farid, "Image forgery detection," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 16–25, Mar. 2009.

[14] J. F. O'Brien and H. Farid, "Exposing photo manipulation with inconsistent reflections," *ACM Transactions on Graphics*, vol. 31, no. 1, pp. 4:1–11, Jan. 2012.

[15] M. C. Stamm, X. Chu, and K. J. R. Liu, "Forensically determining the order of signal processing operations," in *Proc. IEEE WIFS*. IEEE, 2013.

[16] S. Milani, M. Tagliasacchi, and S. Tubaro, "Discriminating multiple jpeg compression using first digit features," in *Proc. IEEE ICASSP*. IEEE, 2012, pp. 2253–2256.

[17] S.-Y. Lai and R. Bohme, "Block convergence in repeated transform coding: JPEG-100 forensics, carbon dating, and tamper detection," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 3028–3032.

[18] A. Swaminathan, M. Wu, and K. J. R. Liu, "A component estimation framework for information forensics," *in Proc. IEEE 9th Workshop on Multimedia Signal Processing*, pp. 397–400, Oct. 2007.

[19] A. Swaminathan, M. Wu, and K. J. R. Liu, "A pattern classification framework for theoretical analysis of component forensics," in *in Proc. IEEE ICASSP*, March 2008, pp. 1665–1668.

[20] P. Comesaña, "Detection and information theoretic measures for quantifying the distinguishability between multimedia operator chains," in *IEEE Workshop on Information Forensics and Security*, Tenerife, Spain, 2012.

[21] Z. Fan and R. L. de Queiroz, "Identification of bitmap compression history: JPEG detection and quantizer estimation," *IEEE Trans. on Image Processing*, vol. 12, no. 2, pp. 230235, 2003.

[22] T. M. Cover and J. A. Thomas, *Elements of Information Theory, second edition*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2006.

[23] Xiaoyu Chu, Yan Chen, M.C. Stamm, and K.J.R. Liu, "Information theoretical limit of compression forensics," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, May 2014, pp. 2689–2693.

[24] T. Pevný and J. Fridrich, "Estimation of primary quantization matrix in double compressed JPEG images," in *Proc. of Digital Forensic Research Workshop*, Cleveland, Ohio, Aug. 2003.

[25] A. C. Popescu and H. Farid, "Statistical tools for digital forensics," in *6th International Workshop on Information Hiding*, Toronto, Canada, 2004.

[26] B. Mahdian and S. Saic, "Detecting double compressed JPEG images," in *3rd International Conference on Crime Detection and Prevention*, Dec. 2009, pp. 1–6.

[27] X. Feng and G. Doërr, "JPEG recompression detection," in *Proc. of SPIE, Media Forensics and Security II*, Feb. 2010, vol. 7541, pp. 0J1–0J10.

[28] F. Huang, J. Huang, and Y. Q. Shi, "Detecting double JPEG compression with the same quantization matrix," *IEEE Trans. on Information Forensics and Security*, vol. 5, no. 4, pp. 848 –856, Dec. 2010.

[29] E. Y. Lam, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans, on Image Proc.*, vol. 9, no. 10, pp. 1661–1666, Oct. 2000.

[30] G. Schaefer and M. Stich, "UCID: An uncompressed color image database," *Proc. SPIE: Storage and Retrieval Methods and Applications for Multimedia*, vol. 5307, pp. 472480, 2004.

[31] S. X. Ng and L. Hanzo, "On the MIMO channel capacity of multi-dimensional signal sets," *Vehicular Technology, IEEE Transactions on*, vol. 55, no. 2, pp. 528–536, March 2006.

[32] Thomas Gloe and Rainer Böhme, "The 'dresden image database' for benchmarking digital image forensics," in *ACM Symposium on Applied Computing*, 2010, vol. 2, pp. 1584–1590.

[33] X. Chu, M. C. Stamm, Y. Chen, and K. J. R. Liu, "Concealability-rate-distortion tradeoff in image compression anti-forensics," in *Proc. IEEE ICASSP, 2013*, May 2013, pp. 3063–3067.