INFORMATION THEORETICAL LIMIT OF COMPRESSION FORENSICS

Xiaoyu Chu[*1], *Yan Chen*[*2], *Matthew C. Stamm*[*#3] *and K. J. Ray Liu*[*4]

[*]Dept. of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA
[1]cxygrace@umd.edu, [2]yan@umd.edu, [4]kjrliu@umd.edu

[#] Dept. of Electrical and Computer Engineering, Drexel University, Philadelphia, PA, USA
[3]mstamm@coe.drexel.edu

ABSTRACT

Multimedia forensics concerns on extracting forensic information from suspicious multimedia contents. This information was embedded into the content inadvertently whenever an operation happened. Investigators may estimate the possible operations by obtaining features from the multimedia content and applying detection algorithms based on the statistics. While most existing works focus on improving detection performance and finding what more we can do, understanding the fundamental limit on the forensic information that we can obtain from the extracted features is also important. It enables us to understand the limit of forensicability. In this paper, we explore the fundamental limit of forensicability by introducing an information theoretical framework for multimedia forensics. We use mutual information as the measure of forensic information conveyed by features to investigators. To show the analytical process, we take the case of multiple JPEG compression forensics as an example. We claim that, under typical circumstances, the maximum number of compressions that we can detect by examine DCT coefficients is up to 4, in an expected sense. In addition, we also find the patterns of compression quality factors that contain the most and least forensic information.

Index Terms— Digital Forensics, JPEG Compression, Information Theory, Fundamental Limit.

1. INTRODUCTION

Due to the easy access and editing of multimedia signals, verifying the authenticity of multimedia content becomes important. In order to achieve this goal, many forensic techniques have been developed to trace processing histories of multimedia signals [1]. For example, an image's capture device can be determined through multiple evidences [2–5]. Many editing operations can also be detected, such as contrast enhancement [6], resampling [7], median filtering [8], and compressions [9–11]. Among these processing histories, image compression history is of particularly forensic important because: 1) the estimated quantization table can help investigators to identify the capture device; 2) multiple compressions reveal information of possible manipulations, since recompression happens whenever an image is re-saved after editing.

In order to increase the forensicability of investigators, researchers have been endeavoring to improve the current detection performance [12] and explore solutions to identify more complicated operations [13]. Taking compression history as an example, many forensic schemes have been proposed to achieve better performance of double compression detection [10, 11, 14–20]. Researchers also attempt to detect three or more times of compressions using first digit features, where it encountered some difficulty when the number of compressions needed to be detected reaches four [21]. Additionally, many new evidences of manipulations have been found to help investigators make more reliable decisions [22, 23].

As we continuously challenge and improve forensicability, a question would naturally arise - Is there a fundamental limit on forensicability that we can never break? In other words, while we are trying so hard to find "what we can do", should we understand and acknowledge "what we can not do"? Answering this question would be both important for investigators and forgers. On one hand, investigators would know their limit and how far current technology can go. On the other hand, forgers could do manipulations to the extent that beyonds this limit without worrying the exposure of their traces.

In this paper, we introduce an information theoretical model to explore the fundamental limit of forensicability. We understand forensics as the procedure of extracting forensic information from statistics of multimedia content, like features, where the information was inadvertently embedded to the content whenever an operation modifies the statistics. As long as the content has limited capacity of the statistics, there will be a limit on the amount of information that the content can contain. We find this limit by examining the extent of mutual information between features and possible operations, which also implies the extent of fundamental forensic information that investigators can obtain from extracting these features regardless of explicit detectors. To demonstrate the effectiveness of our framework, we use it to examine the case of multiple compression detection and answer the question of how many compressions at most that we can detect.

2. INFORMATION THEORETICAL FRAMEWORK

Let us first review the process that a multimedia signal may go through in a typical forensic analysis system. As it is shown in Fig. 1, an unaltered multimedia signal may experience some operations, like editing or acquisition process, to become our inspection signal. Then, investigators try to estimate what operations this signal has gone through by extracting various features from the multimedia content. Based on these features, explicit detection algorithms will be applied to detect possible operations.

By exploring the fundamental limit of forensicability, we are answering "how much forensic information about the operations we can, at most, obtain from the extracted features regardless of the detectors?" In other words, we are concerning the relationship be-

This work is supported in part by the NSF grant CCF1320803

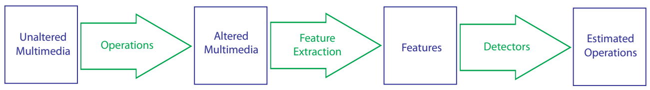

Fig. 1. Traditional process that a multimedia signal may go through when considering forensics.

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

tween operations and features. This motivates us to abstract all the processes happened between operations and features as a channel, as shown in Fig. 2. Due to the content-dependent characteristic of multimedia signal, the channel is random instead of deterministic where the randomness comes from the content-dependent noise. In such a case, each operation may generate different features for different multimedia signal, i.e., a distribution of features can be obtained for each operation. When these feature distributions overlap too much, we may not be able to distinguish the corresponding operations. In this paper, we propose to quantify the forensic information that the feature can tell about the operations by measuring the mutual information between them. Note that comparing with the related work in [24], whose definitions of distinguishability between operation chains are constrained to simple hypothesis model, our measure of mutual information is also suitable for more general forensic models.

To demonstrate the effectiveness of this framework, we use multiple JPEG compression forensics as an example. When an image is compressed into a JPEG format, discrete cosine transform (DCT) will be applied on each 8 by 8 block of the image, where DCT coefficients are obtained for all subbands. Then, these coefficients will be quantized and lossless coded to form the JPEG file. Decompression goes the reverse direction. In these processes, only the quantization is lossy, due to which the JPEG compression fingerprints are introduced [9]. If the image is re-saved with different quantization tables, new patterns of the histogram of DCT coefficients may be introduced due to the double quantizations, which can be used as the fingerprints of double JPEG compressions [10].

Since the histogram of DCT coefficients is a common-used feature in the literature for JPEG compression forensic, in this paper, we will evaluate the maximal number of JPEG compressions that can be distinguished by using this feature. The system model for this specific problem is shown in Fig. 3, where $X \in \{1, 2, ..., M\}$ represents the number of JPEG compressions that is applied on the unaltered image and $\underline{Y} = [y_{-N}, ..., y_0, ..., y_N]$ is the vector form of the normalized DCT histogram that we observe.

We take one subband as an example to illustrate the relationship between X and $\underline{Y}$. We model the distribution of DCT coefficients of an uncompressed image, D_0, as a Laplace distribution [25],

$$\mathbb{P}(D_0 = d) = \frac{\lambda}{2} e^{-\lambda |d|}, \tag{1}$$

with λ being the Laplacian parameter. Let $\mathcal{Q}_M = \{q_1, q_2, ..., q_M\}$, where the order of elements matters, be the set of possible quantization step sizes that may be used for the subband in the compressions. In multiple compression detection problem, we are given a JPEG image and try to identify the number of compressions that has been applied before the last one. Therefore, we keep the last compression the same for all hypothesis. That is, if $X = m, 1 \leq m \leq M$, then the DCT coefficient D_0 should have been quantized by step

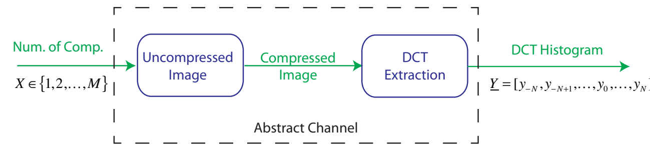

Fig. 3. Model for multiple compression detection forensics.

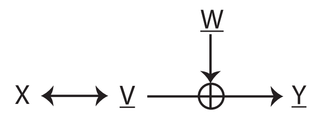

Fig. 4. Abstract channel inner structure for the model in Fig. 3.

sizes $\{q_{M-m+1}, q_{M-m+2} ..., q_M\}$ in order, and the m-times quantized coefficient D_m is

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

With (1) and (2), we can theoretically derive the distribution of D_m, which only has nonzero values at integer multiples of q_M, as shown below,

$$\begin{aligned} \underline{v}_m(\lambda, \underline{q}_m) &= [v_{m,-N}, v_{m,-N+1}, ..., v_{m,N}] \\ &= [\mathbb{P}(D_m = -Nq_M), ..., \mathbb{P}(D_m = Nq_M)], \end{aligned} \tag{3}$$

where $\underline{q}_m = [q_{M-m+1}, q_{M-m+2} ..., q_M]$, and $\underline{v}_m(\lambda, \underline{q}_m)$ stands for the theoretical distribution of DCT coefficients if m times of compressions have been done. However, due to the model mismatch and/or the rounding and truncation in the compression and decompression, noise will be introduced and thus we may not observe the theoretical distribution $\underline{v}_m(\lambda, \underline{q}_m)$. Let us denote the DCT histogram that we actually observe as

$$\underline{Y}_m = [Y_{m,-N}, ..., Y_{m,N}] = [h(-Nq_M), ...h(Nq_M)], \tag{4}$$

where $h(\cdot)$ represents the normalized histogram at a certain location. By assuming that the noise $\underline{W}$ is an additive random variable, we have

$$\underline{Y}_m = \underline{v}_m(\lambda, \underline{q}_m) + \underline{W}. \tag{5}$$

In summary, given λ and $\mathcal{Q}_M$, for each X, there is a theoretical DCT coefficient distribution $\underline{V} \in \{\underline{v}_1(\lambda, \underline{q}_1), \underline{v}_2(\lambda, \underline{q}_2), ..., \underline{v}_M(\lambda, \underline{q}_M)\}$. However, due to the model mismatch and/or rounding and truncation effect, additive noise is introduced and thus the observed histogram $\underline{Y}$ is a noisy version of $\underline{V}$. The relationships among these random variables are shown in Fig. 4.

3. CHANNEL NOISE MODELING AND ESTIMATION

In order to model the channel noise $\underline{W}$, we use the real data statistics to analyze how the observed normalized histogram differs from the theoretical distribution. Without loss of generality, we examine the case where the image is compressed once, i.e., $X = 1$. We use the 1338 uncompressed images from UCID database [26], and JPEG compress them with quality factor 80. In order to derive the theoretical DCT distributions for each single compressed image, we first estimate the Laplacian parameter λ using DCT coefficients statistics from its uncompressed version; then, $\underline{v}_1(\lambda, q)$ is calculated according to (3). The observed normalized DCT histograms, denoted as $\underline{y}_1$, are directly obtained from the single compressed images. By subtracting $\underline{v}_1(\lambda, q)$ from $\underline{y}_1$, we obtain the

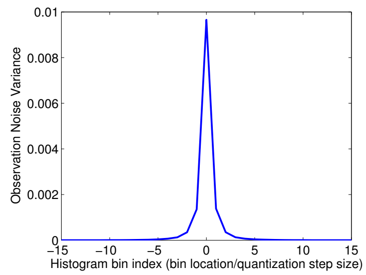

Fig. 5. Channel noise variance for each DCT histogram bin.

channel noise, denoted as $\underline{w} = [w_{-N}, ..., w_N]$ for all images. From histograms of w_n, we observe that Gaussian distributions can approximately describe the channel noise, i.e., $\underline{W} \sim \mathcal{N}(0, \Sigma)$. By assuming independence among different histogram bins, we have $W_n \sim \mathcal{N}(0, \sigma_n), -N \leq n \leq N$.

To further model the variance σ_n, we plot the variance of W_n for $-14 \leq n \leq 14$ in Fig. 5. We can see that variance are not even for different histogram bins. Specifically, we observe higher variance of W_n when $n \to 0$. Since the theoretical distribution of DCT coefficients for a single compressed image $\underline{v}_1(\lambda, q)$ is quantized Laplacian, which has similar shape with the variances of the observation noise, we model σ_n as an increasing function of $v_{1,n}$. Specifically, we model $\sigma_n = cv_{1,n}^{2\alpha}$, where c and α are positive constants. Such a modeling is reasonable because that observation noise tends to be larger for higher magnitude histogram bins.

By generalizing from this specific case of single compression, our model of the channel noise for m times compressions is $\underline{W} \sim \mathcal{N}\big(0, diag\big(c\underline{v}_m^{2\alpha}(\lambda, \underline{q}_m)\big)\big)$. Thus, the conditional probability of the channel output $\underline{Y}$ given input X can be described as

$$\mathbb{P}(\underline{Y}|X) = \mathbb{P}(\underline{Y}|\underline{V}) \sim \mathcal{N}(\underline{V}, diag(c\underline{V}^{2\alpha})). \quad (6)$$

To estimate c and α, we use the statistics of $\underline{V}_1$ and $\underline{Y}_1$ as we obtained at the beginning of this section. It is observed that each image may result in a different estimated λ from others. Therefore, the instances of $\underline{V}_1$ are different due to their dependences on λ. Let $\underline{v}_{\lambda_i}$ denote the theoretical distribution of single compressed DCT coefficients for image i, with $v_{\lambda_i,n}$ representing its nth element. Let $Y_{\lambda_i,n}$ denote the random variable of the observed normalized histogram bin magnitude at the nth location. Thus, according to our model, $Y_{\lambda_i,n} \sim \mathcal{N}(v_{\lambda_i,n}, cv_{\lambda_i,n}^{2\alpha})$. Then, the likelihood probability of observing the nth normalized histogram bin from image i as $y_{\lambda_i,n}$ is

$$\mathbb{P}(Y_{\lambda_i,n} = y_{\lambda_i,n}) = \frac{1}{\sqrt{2\pi cv_{\lambda_i,n}^{2\alpha}}} exp\Big\{-\frac{(y_{\lambda_i,n} - v_{\lambda_i,n})^2}{2cv_{\lambda_i,n}^{2\alpha}}\Big\}. \quad (7)$$

By maximizing the likelihood probabilities for all histogram bins from all images, we obtain the estimates of c and α as,

$$(\hat{c}, \hat{\alpha}) = \underset{c>0,\alpha>0}{\arg\max} \log \sum_{i=1}^{K} \sum_{n=-N}^{N} \mathbb{P}(Y_{\lambda_i,n} = y_{\lambda_i,n}), \quad (8)$$

According to Karush-Kuhn-Tucker conditions,

$$\sum_{i=1}^{K} \sum_{n=-N}^{N} (y_{\lambda_i,n} - v_{\lambda_i,n})^2 \ln v_{\lambda_i,n} (\frac{1}{v_{\lambda_i,n}})^{2\hat{\alpha}} = \hat{c} \sum_{i=1}^{K} \sum_{n=-N}^{N} v_{\lambda_i,n},$$

$$\sum_{i=1}^{K} \sum_{n=-N}^{N} \frac{(y_{\lambda_i,n} - v_{\lambda_i,n})^2}{v_{\lambda_i,n}^{2\hat{\alpha}}} = \hat{c}K(2N+1). \quad (9)$$

By solving (9), we can derive the optimal c and α. Specifically for the UCID database, we have $c = 0.0494$ and $\alpha = 0.744$.

4. MUTUAL INFORMATION AND EXPECTED PERFECT DETECTION

By definition, the mutual information $I(\underline{V}; \underline{Y})$, which equals to $I(X; \underline{Y})$, describes the amount of information that $\underline{Y}$ can tell about $\underline{V}$, and thus X. Therefore, we use $I(\underline{V}; \underline{Y})$ to measure the forensic information that the DCT histogram contains about the number of compressions. In forensic analysis, we usually assume that each possible operation happens with equal probability. Therefore, we use a uniform prior, i.e., $\mathbb{P}(X = m) = \frac{1}{M}$, to calculate the mutual information, as shown below,

$$I(\underline{V}; \underline{Y}) = \log_2 M - \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}\Big[\log_2 \sum_{j=1}^{M} exp\Big(\Phi_j^m(\underline{V})\Big)\Big], \quad (10)$$

where

$$\Phi_j^m(\underline{V}) = \sum_{n=-N}^{N} \Big[\alpha \ln \frac{v_{m,n}}{v_{j,n}} - \frac{(y_n - v_{j,n})^2}{2cv_{j,n}^{2\alpha}} + \frac{(y_n - v_{m,n})^2}{2cv_{m,n}^{2\alpha}}\Big], \quad (11)$$

with $[v_{m,-N}, ..., v_{m,N}]$ representing the theoretical distribution of DCT coefficients if it is compressed m times, and $[y_{-N}, ..., y_N]$ is the normalized DCT histogram that we observe.

The second term in (10) is the conditional entropy $H(\underline{V}|\underline{Y}) = H(X|\underline{Y})$, which describes the amount of confusion on X given the knowledge of $\underline{Y}$. Thus, the smaller $H(\underline{V}|\underline{Y})$ is, i.e., the closer $I(\underline{V}; \underline{Y})$ is to $\log_2 M$, then, the more forensic information that we can extract from DCT histograms towards detecting M times of compressions. Consequently, the better detection performance we can get regarding distinguishing among $1, 2, ..., M$ times of compressions. To further describe the relationship between mutual information and detection performance, we have the following theorem to show that the conditional entropy provides a lower bound on the error probability:

Theorem 1 *Consider any estimator $\hat{X}$ such that $X \to \underline{Y} \to \hat{X}$. If it is better than a random decision, where the decision is made by randomly pick one from the set of X with uniform probability, let $P_e = \mathbb{P}(X \neq \hat{X})$, then we have*

$$P_e^0 \leq P_e \leq \frac{M-1}{M}, \quad (12)$$

where P_e^0 is unique and satisfies

$$H(P_e^0) + P_e^0 \log_2(M-1) = H(X|\underline{Y}). \quad (13)$$

This theorem can be proved by following the derivation of Fano's inequality in [27].

Theoretically, as long as the channel is not perfectly clean, conditional entropy will not be zero, and mutual information cannot reach its maximum $\log_2 M$, i.e., perfect detection is not reachable theoretically. However, if the channel noise variance is small enough and the conditional entropy is close to zero, we can still argue the existence of perfect detections under real experiment setting as follows. Let S denote the size of the database that we use to evaluate a detector's performance, and P_e represents the theoretical error probability. We model each data in the database as an independent Bernoulli random variable, with probability P_e being the one that will cause detection error, which is denoted as "bad data". Then, it is well known that the expected time of the first occurrence of the

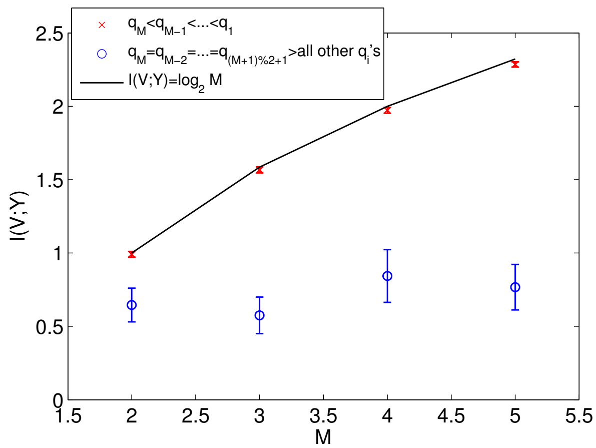

Fig. 6. The patterns of $\underline{q}$'s which yield the highest mutual information and the lowest mutual information.

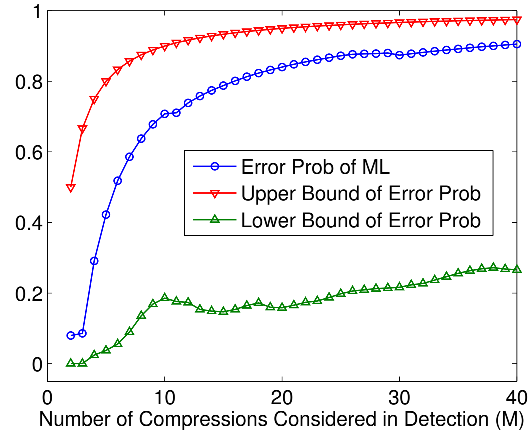

Fig. 7. Experimental error probability for one example selection of $\underline{q}$ versus the number of compressions needed to be detected, and their corresponding lower and upper bounds.

"bad data" is $1/P_e$. Therefore, if $P_e < 1/S$, expectedly, the "bad data" will not occur in the database with size of S, i.e., a perfect detection is achieved. We define the perfect detection in such sense as the *expected perfect detection*, where the expectation is taken upon all databases with the same size. The contrary of the above statement also holds, where, if $P_e > 1/S$, then the expected perfect detection cannot be achieved. In order to guarantee that any detector cannot achieve expected perfect detection, we need the lower bound of error probability, which is determined by conditional entropy, to satisfy

$$P_e^0 > 1/S. \tag{14}$$

Since the conditional entropy tends to increase with more numbers of compressions needed to be detected, i.e., larger M, this criterion can be used to determine the maximal M that we can achieve the expected perfect detection.

5. SIMULATION RESULTS

To calculate the forensic information that DCT histogram contains about the number of compressions, we use Monte Carlo simulation to obtain the mutual information $I(\underline{V};\underline{Y})$ in (10). Using subband $(2,3)$ as an example, we find that the quantization step sizes corresponding to quality factor between 50 and 100 are 1 to 14; then by excluding the trivial case where one quantization step size is an integer multiple of another, we choose the candidate quantization step sizes as $\{5,6,7,8,9,11,13\}$.To obtain the mutual information for each M, we randomly select $\mathcal{Q}_M$ from this set element-wisely and guaranteed $q_{i-1} \neq q_i$. For each selection of $\mathcal{Q}_M$, we calculate the mutual information by Monte Carlo simulation. We find that $I(\underline{V};\underline{Y})$ differs a lot with $\mathcal{Q}_M$. Then, we test on all combinations of $\mathcal{Q}_M$ and summarize the patterns of $\mathcal{Q}_M$ that can achieve the highest and lowest $I(\underline{V};\underline{Y})$. As shown in Fig. 6, those compressions that always use a smaller quantization step size conveys the maximum mutual information, which should be avoided by forgers. While those compressions that periodically use the same quantization step size with the smaller ones in the middle have the lowest mutual information, and thus will be favored by forgers.

To verify theorem 1, we randomly picked a $\mathcal{Q}_M$, and compressed the 1338 images in UCID database using quality factors corresponding to quantization step sizes $\underline{q}_m$ to obtain m times compressed images. By collecting all multiple compressed images with compression times from $m = 1$ to $m = M$, we obtain a test database of $1338 \times M$ images. Then, based on our channel model, we used maximum likelihood estimate to classify each image in this database and calculated the error probability. Fig. 7 shows these experimental probabilities for different M's, which lies between the two theoretical bounds derived in theorem 1. Additionally, from the statistics, we find that $P_e^0 > 1/S$ when $M \geq 4$. According to Theorem 1 and (14), this means that with this selection of $\mathcal{Q}_M$, up to 3 times of compressions can be detected in the sense of expected perfect detection.

Table 1. $\min_{\mathcal{Q}_M} P_e^0$ 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, we can use (14) to determine the maximum of compressions that we can detect. Since the mutual information depends on $\mathcal{Q}_M$, we use $\min_{\mathcal{Q}_M} P_e^0$ instead of P_e^0 to verify the criterion in (14). Table 1 lists the values of $\min_{\mathcal{Q}_M} P_e^0$ for different M's. If we take $S < 5000$ as a typical database size in forensics, since $5 \times 10^{-5} < 1/5000 < 2.1 \times 10^{-4}$, we are not able to obtain expected perfect detection for $M = 5$. In other words, we can claim that, no matter what detector we use and no matter what $\mathcal{Q}_M$ was used during the compressions, we can only correctly distinguish 4 times of compressions, in the sense of expected perfect detection.

6. CONCLUSION

In this paper, we proposed an information theoretical model to find the fundamental limit in forensics, where mutual information was used to quantify the amount of forensic information features convey about operations. We took the case of multiple compression detection as an example to find the maximum number of compressions that can be detected. Along the analysis, we modeled the abstract channel as a Gaussian additive channel and estimated the variance. We derived mutual information based on the channel model and used it to upper bound the performance of any detector investigator may use. Then, the expected perfect detection was defined, and based on which, we claimed that, by examining DCT coefficients, we can only detect up to 4 times of compressions under typical circumstances. In addition, we found the patterns of compression quality factors which contain the most and least forensic information.

7. 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] 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.

[3] 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.

[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] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[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. Barni and A. Costanzo, "A fuzzy approach to deal with uncertainty in image forensics," *Signal Processing: Image Communication*, vol. 27, no. 9, pp. 998–1010, 2012.

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

[14] 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.

[15] D. Fu, Y. Q. Shi, and W. Su, "A generalized Benford's law for JPEG coefficients and its applications in image forensics," in *Proc. of SPIE, Electronic Imaging, Security and Watermarking of Multimedia Contents IX*, Feb. 2007, vol. 6505, pp. 1L1–1L11.

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

[17] Y. L. Chen and C. T. Hsu, "Detecting doubly compressed images based on quantization noise model and image restoration," in *IEEE International Workshop on Multimedia Signal Processing*, Oct. 2009, pp. 1–6.

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

[19] 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.

[20] T. Bianchi and A. Piva, "Image forgery localization via block-grained analysis of jpeg artifacts," *IEEE Trans. on Information Forensics and Security*, vol. 7, no. 3, pp. 1003–1017, Jun. 2012.

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

[22] E. Kee, J. F. O'Brien, and H. Farid, "Exposing photo manipulation with inconsistent shadows," *ACM Transactions on Graphics*, vol. 32, no. 4, pp. 28:1–12, Sept. 2013.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, New York, NY, USA, 1991.