A Universal Image Forensic Strategy Based on Steganalytic Model

Xiaoqing Qiu
School of Information Science and Technology
Sun Yat-sen University
Guangzhou, P.R. China
qiuxq3@mail2.sysu.edu.cn

Haodong Li
School of Information Science and Technology
Sun Yat-sen University
Guangzhou, P.R. China
lihaod@mail2.sysu.edu.cn

Weiqi Luo*
School of Software
Sun Yat-sen University
Guangzhou, P.R. China
weiqi.luo@yahoo.com

Jiwu Huang
College of Information Engineering
Shenzhen University
Shenzhen, P.R. China
jwhuang@szu.edu.cn

ABSTRACT

Image forensics have made great progress during the past decade. However, almost all existing forensic methods can be regarded as the specific way, since they mainly focus on detecting one type of image processing operations. When the type of operations changes, the performances of the forensic methods usually degrade significantly. In this paper, we propose a universal forensics strategy based on steganalytic model. By analyzing the similarity between steganography and image processing operation, we find that almost all image operations have to modify many image pixels without considering some inherent properties within the original image, which is similar to what in steganography. Therefore, it is reasonable to model various image processing operations as steganography and it is promising to detect them with the help of some effective universal steganalytic features. In our experiments, we evaluate several advanced steganalytic features on six kinds of typical image processing operations. The experimental results show that all evaluated steganalyzers perform well while some steganalytic methods such as the spatial rich model (SRM) [4] and LBP [19] based methods even outperform the specific forensic methods significantly. What is more, they can further identify the type of various image processing operations, which is impossible to achieve using the existing forensic methods.

Categories and Subject Descriptors

I.4 [**Image Processing and computer vision**]

*Corresponding author

IH&MMSec'14, June 11–13, 2014, Salzburg, Austria.

http://dx.doi.org/10.1145/2600918.2600941.

Keywords

Universal Forensics; Tampering Detection; Steganalysis

1. INTRODUCTION

Nowadays, with the powerful and user-friendly digital image editing software such as PhotoShop and GIMP, it becomes easy to modify digital images without leaving any perceptible artifacts. Once such tools are abused for those forgers, it would lead to some potential serious moral, ethical and legal consequences. Therefore, digital image forensics [21] have become an important issue and have attracted increasingly attention.

Up to now, many forensic methods have been proposed for different forensic scenes, such as exposing splicing images [18, 5], identifying image compression history [3, 13], detecting some image processing operations [14, 20, 23], and so on. However, most exiting methods just consider only one type of image processing operations, and these methods are difficult to be extended for other image operations. For example, a set of features proposed for identifying JPEG compression may not be suitable for identifying image splicing, re-sampling, and/or blurring operations. Moreover, most forensic methods usually assume that the questionable image has been processed by a specific image processing operation or not, and they aim to make a binary decision. In most forensic cases, however, such an assumption is not very reasonable since no prior information is available for a given image. If the suspicious image may be performed with several possible operations, all specific forensic methods become poor or even useless. Therefore, a universal forensic strategy is needed in this case. To our best knowledge, none related literatures have been reported previously.

In this paper, we try to find a universal strategy for detecting various types of image processing operations. We begin with studying the common artifacts introduced by different types of image processing operations, and found that any processing would inevitably modify many image pixels and destroy some inherent statistics within an original image, which is quite analogous to image steganography. Based on

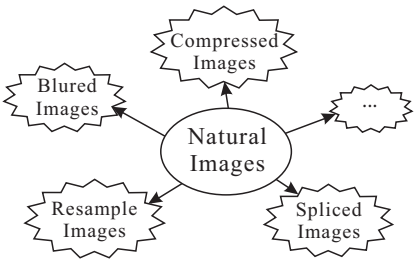


Figure 1: Identification of image processing operations based on natural image model

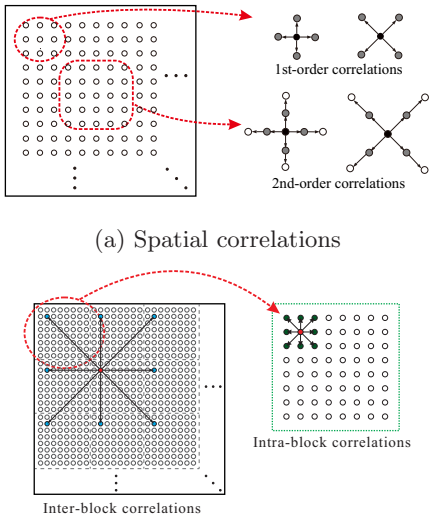
this analysis, we propose a forensic strategy via borrowing some powerful features from universal steganalysis. Unlike the typical forensic methods, the proposed strategy is universal in two senses. First of all, without changing the features, it can be widely used to identify those images modified by different kinds of operations from original ones. Secondly, the proposed strategy can also be used to further identify the types of the operations. In the experiments, we tested six typical kinds of image processing on 1050 natural images downloaded from the First IFS-TC Image Forensics Challenge [1]. The experimental results have shown the superiority of the proposed strategy compared with those specific forensic methods.

The rest of this paper is organized as follows. Section 2 analyzes the universal image forensics from the view of steganalysis; Section 3 describes the proposed universal strategy based on steganalytic model; Section 4 shows the experimental results and discussions. Finally, the concluding remarks and future works will be given in Section 5.

2. UNIVERSAL IMAGE FORENSICS FROM THE VIEW OF STEGANALYSIS

To develop a universal forensic method for detecting various image processing operations, such as lossy compression, region blurring, re-sampling, splicing, and so on, we should concentrate on the common artifacts left by various operations rather than some specific artifacts introduced by a certain operation as it did in most previous forensic methods. One of the options is to model the inherent statistical properties of original natural images, the common objects that various operations would perform on. As illustrated in Fig. 1, if such an ideal model is available, it is possible to identify various image processing operations since different operations usually modify the model in different manners and/or strengths in the corresponding feature space.

As we know, there are many inherent statistical properties within natural images, and some properties may be useful in our forensic scene. For instance, as illustrated in Fig. 2, the adjacent pixel values in spatial domain as well as their corresponding DCT frequency coefficients are highly correlated. However, any image processing operation would inevitably modify some pixels values within an image. As a result, the correlations among those modified pixels and their neighbors would be changed or even broken. If such correlated properties can be well studied and modeled, it is expected that those modifications in an image can be detected effectively. Therefore, our key issue is how to model the inherent correlations within natural images.



(a) Spatial correlations

(b) Frequency correlations

Figure 2: Illustration of the correlations in spatial and frequency domain within a natural image

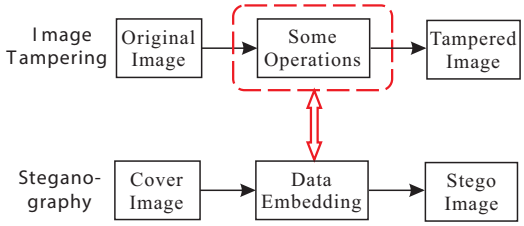


Figure 3: Illustration of the relationship between image tampering and steganography

Fortunately, some powerful statistical features can be borrowed from another research field - steganalysis, which aims to expose those stego images with hidden messages embedded by steganography. As illustrated in Fig. 3, if the operation of data embedding in steganography is regarded as a specific type of image tampering [1], then the tasks of image forensics and steganalysis become exactly the same, namely, differentiating those natural images (*i.e.* cover images) from the tampered ones (*i.e.* stego images). Please note that most typical image processing operations would change the inherent properties (such as the correlations and so on) more severely comparing with the data embedding operation. The reasons are list as follows.

- In the modern steganography such as HUGO [16] and WOW [6], the embedding changes are mainly located at the textural/noisy regions that are difficult to model. However, both textural and smooth regions within a natural image would be modified a lot for most image processing operations, such as JPEG compression, blurring, and median filtering.

- Both the number and the magnitude of modified pixel values in most image processing operations are much

[1]Note that both operations would modify many pixel values, and destroy some inherent correlations in original images.

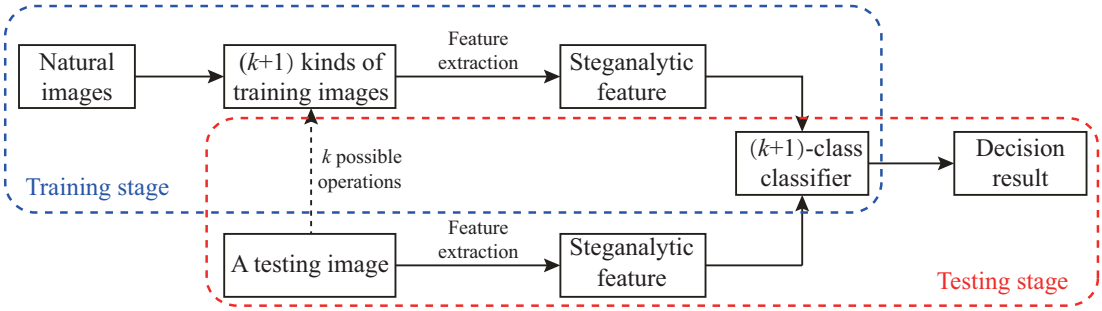


Figure 4: The diagram of the proposed strategy

larger than those in steganography. For most steganography, the modification magnitude is usually limited in ± 1 for every pixel value, and the modification rate is typically less than 9% (around 0.4bpp using the WOW embedding algorithm [6]) for the secure embedding [2]. Based on our experiments on 1050 natural images in dataset [1], however, the average modification rates for JPEG compression, gamma correction, median filtering, and Gaussian filtering are 36.55%, 48.71%, 35.84%, and 37.93%, respectively, all are more than three times over the secure limit (*i.e.* 9% modification rate).

- No obvious visual artifacts would be introduced into those stego images. However, some image contents would be changed significantly for some tampering operations, such as copy-paste tampering, image splicing and in-painting.

Based on the above analysis, it is expected that some steganalytic features which can effectively model those inherent correlation properties within natural images would be also promising in image forensics.

3. THE PROPOSED STRATEGY

In this section, we will propose a universal image forensic strategy based on steganalytic model. First of all, we should analyze what kind of steganalytic features is suitable in the proposed strategy. It is known that the steganalytic methods can be divided into two different types, *i.e.* targeted and universal methods [11]. Targeted steganalytic methods require the knowledge of the targeted steganography. It may become useless for other steganographic ones, which is similar to those existing image forensic methods being effective to detect the corresponding image operations. Unlike the targeted steganalytic methods, universal steganalytic methods try to model some inherent statistical properties within natural images, and extract the steganalytic features without any information about the targeted steganography. Therefore, those universal steganalytic features should be considered in the proposed strategy. Besides, the supervised learning is needed to train a binary or multi-class classifier, which depends on the the number of possible image processing operations under investigation.

[2]Usually, the higher the embedding rate, the better the steganalytic performances. Based on the results shown in [6], when the embedding rate is smaller than 0.4bpp, the security of WOW is still acceptable evaluated on the state-of-the-art steganalytic SRM model [4].

The proposed strategy is illustrated in Fig. 4. Assume the number of possible image processing operations is k for a given testing image, where $k \geq 1$. In the training stage, we first collect sufficient natural images, and we create the corresponding $k+1$ kinds of images (including the class of original natural images themselves) using the k possible operations respectively, and then we extract the steganalytic features from every training image. Finally, we can train a (k+1)-class classifier. In the testing stage, the steganalytic features of the testing image are extracted and then are fed to the obtained classifier to get the decision results.

4. EXPERIMENTAL RESULTS

In the experiments, we use the image database from the first IEEE IFS-TC image forensics challenge [1], which consists of 1050 authentic and 1050 tampered images in PNG format. The original images are captured by different digital cameras with various scenes, their sizes are varying from 480×640 to 4288×4752. The tampered images are obtained via different manipulation techniques such as copy-paste, image in-painting and splicing using some image editing software such as GIMP, Adobe Photoshop *etc.* Some tampered image examples in the database are shown in Fig. 5. We have highlighted the tampered regions of these images. Please note that those tampered images are very realistic without any obvious artifacts.

We use the ensemble classifier for classification just as it did in [10]. 50% of the authentic images and 50% of the tampered images are randomly selected to train the classifier, and the detection accuracy is obtained by evaluating the remaining images in the testing stage. Please note that we repeat the training and testing 10 times and show the average results in the following tables in subsection 4.1-4.3.

4.1 Detection of Image Splicing

In this subsection, we try to evaluate the performances of six typical universal steganalytic methods (including three kinds of steganalytic features for spatial domain, *i.e.* SRM [4], LBP [19], and SPAM [15]; three kinds of steganalytic features for JPEG domain, *i.e.* CF* [10], CC-CHEN [2, 9], and CC-PEV [17, 9]) and two specific forensic methods for splicing detection (*i.e.* He [5], Shi [18]).

After splicing operation, the resulting images usually are stored in JPEG format. To evaluate the robustness against JPEG compression, all the images are JPEG compressed with a quality factor randomly selected from 75 to 95 with a step of 5. The detection accuracies for those splicing images

Figure 5: Some tampered images in the dataset [1]

Table 1: Average detection accuracies(%) for those splicing image before and after JPEG compression with different quality factors (QF). The best result is marked with an asterisk "*" in each case.

Feature Set	Spatial Steganalysis			JPEG Steganalysis			Specific Forensics	
	SRM[4]	LBP[19]	SPAM[15]	CF*[10]	CC-Chen[2, 9]	CC-PEV[17, 9]	He[5]	Shi[18]
QF=75	**88.37***	86.94	81.05	86.12	82.74	78.70	86.83	73.54
QF=80	**89.35***	88.47	82.60	87.23	83.96	80.69	86.82	74.25
QF=85	**91.22***	90.57	83.92	90.19	84.96	81.67	87.98	77.05
QF=90	**92.92***	92.89	86.50	90.98	86.58	82.74	90.47	79.73
QF=95	**94.76***	94.36	88.52	92.16	90.50	86.94	92.84	85.32
Without Compression	**97.70***	95.31	91.35	91.93	89.91	88.79	93.34	90.45

before and after JPEG compression are given in Table 1. From Table 1, three important properties can be observed.

- The detection accuracies of all the methods would increase with increasing the quality factors.
- Though the image splicing operation is performed in the spatial domain, the JPEG steganalytic methods can also achieve satisfactory results since the spatial modifications would probably destroy the inherent correlations among the adjacent DCT coefficients just as illustrated in Section 2.
- Overall, both spatial and JPEG steganalytic methods can detect image splicing, and their detection performances are similar or even much better than the two specific forensic methods, *i.e.* He [5] and Shi [18], especially when the quality factor is high. Among the eight methods, the spatial steganalytic SRM [4] performs the best in all cases.

4.2 Detection of Image Processing Operations

In this subsection, we try to determine whether or not a questionable image has been previously performed with a given image processing operation, including gaussian blurring, gamma correction, lossy JPEG compression, median filtering, and re-sampling.

For each of 1050 authentic images in the database [1], we create five different images with a random parameter selected in Table 2. Besides the two good steganalytic methods for image splicing detection (*i.e.* SRM [4] and LBP [19]) as shown in Table 1, other five forensic methods including AR[3] [7], CE [20], JPA [12], and PPI [14] have been included in the experiment for comparative studies. For each type of image processing operations, therefore, we obtain six different classifiers using the above mentioned methods, and then we use the resulting classifiers to conduct the testing. The average detection accuracies are shown in Table 3.

From Table 3, it is observed that SRM [4] and LBP [19] usually perform the best or nearly the best in most cases (except for detecting gamma correction with the LBP [19]), which means that both the steganalytic features can be regarded as universal features for detecting different image processing operations. For those specific forensic methods, although their detection performances for the corresponding operations are good (see the underlined values in Table 3), their performances are rather poor for other operations. For instance, the method AR [7] can effectively detect the Gaussian blurring and median filtering with both accuracies larger than 97.5%, while it fails to detect gamma correction and the corresponding accuracy drops to 53.11% which is close to the random guessing.

4.3 Identification of Image Processing Operations

In this subsection, we try to identify the type of several possible operations previously used for a given questionable image. Six typical operations including image splicing in subsection 4.1 and five operations in subsection 4.2 are considered in this experiment. The test images are created similarly as described in subsection 4.1 and subsection 4.2. For each of the 1050 authentic images, therefore, we obtain six different types of tampered images.

Ensemble classifier is used for this multi-class classification via pairwise coupling method [8], which means that each pairwise comparison obtains a binary classifier to make a prediction, and then all the predictions are combined to make a final decision based on the majority voting. In our experiments, the two advanced steganalytic features *i.e.* SRM [4] and LBP [19] have been evaluated with the proposed strategy, and the confusion matrices are shown in Table 4 and Table 5, respectively. It is observed that both methods can effectively identify the type of image processing operations for a given image, especially the SRM [4]. On average,

[3]The method AR [7] can be used for detecting both Gaussian blurring and median filtering operations effectively.

Table 2: Parameters of different types of image processing operations

Type of Image Processing Operations	Parameters
Gaussian Blurring	hsize: 3×3, 5×5, 7×7, 9×9; *sigma* : 1.0, 2.0
Gamma Correction	γ: 0.5, 0.6, 0.7, 0.8, 0.9, 1.2, 1.4, 1.6, 1.8, 2.0
JPEG Compression	QF: 75, 76, 77, ..., 95
Median Filtering	R: 3×3, 5×5, 7×7, 9×9
Re-sampling	Up sampling: 1, 3, 5, 10, 20, 30, ..., 90 (%) Down sampling: 1, 3, 5, 10, 15, 20, 25, 30, 35, 40, 45 (%) Rotation angle: 1, 3, 5, 10, 15, 20, 25, 30, 35, 40, 45 (degrees)

Table 3: Average detection accuracies(%) for identifying original images and those images after a given type of image processing operations. The best result is marked with an asterisk "*" in each case, and the underlined values denote the accuracies using the specific methods detecting the corresponding type of operations.

Feature Set	Gaussian Blurring	Gamma Correction	JPEG Compression	Median Filtering	Re-sampling
AR [7]	**98.17**	53.11	65.26	**97.86**	77.56
CE [20]	69.41	**96.31***	60.41	82.57	54.55
JPA [12]	89.80	50.82	**99.18**	82.69	65.10
PPI [14]	52.79	50.28	82.91	52.49	**86.39**
SRM [4]	**99.98***	96.09	99.55	99.75	**98.90***
LBP [19]	99.90	83.64	**99.87***	**99.81***	97.91

Table 6: Average accuracies (%) along the diagonal direction in the corresponding confusion matrix for the specific forensic methods.

Feature Set	AR[7]	CE[20]	JPA[12]	PPI[14]	He[5]
Accuracy	53.60	37.91	38.74	20.68	90.83

the detection accuracies along the diagonal direction in the two confusion matrices are 96.89% and 92.42%, respectively. For the comparison purpose, we also show the average results for the other five specific forensic methods in Table 6. From Table 6, it is observed that the detection performances of most specific methods are rather poor. Please note that the result using the method [5] is still satisfactory, since it tries to detect image splicing based on the Markov features from adjacent DCT and DWT frequency coefficients. Thus, this method can also be regarded as a universal steganalytic method.

5. CONCLUDING REMARKS

In this paper, we propose a universal image forensic strategy based on steganalytic model. The main contribution of this paper is that we find that various image processing operations would inevitably modify many pixel values without considering some inherent statistical properties within natural image, such as the highly correlations among adjacent pixels, which is very similar to the data embedding operation in steganography. Based on such analysis, we build a bridge between two different research issues *i.e.* digital image forensics and steganalysis, and proposed a universal forensic strategy based on steganalytic model. The experimental results show that the proposed strategy with those universal steganalytic features usually performs well. Especially for some advanced steganalytic features such as SRM [4] and LBP [19], the proposed strategy can effectively determine whether or not a questionable image has been performed with a given type of image processing operation, their detection performances are even much better than those state-of-the-art specific forensic methods. Furthermore, the proposed strategy can further identify the type of various typical image processing operations, which has not been considered in previous forensic methods.

In the next step, more image processing operations and universal steganalytic features will be included in our experiments. Furthermore, we will extend the proposed strategy to determine the order of image processing operations [22].

6. ACKNOWLEDGMENTS

This work is supported by National Science & Technology Pillar Program (2012BAK16B06), NSFC (U1135001, 61332012, 61272191), the funding of Zhujiang Science and technology (2011J2200091), and the Guangdong NSF (S2013010012039).

7. REFERENCES

[1] Images corpus of the 1st IEEE IFS-TC image forensics challenge. Available at: `http://ifc.recod.ic.unicamp.br/fc.website/index.py?sec=5`.

[2] C. Chen and Y. Q. Shi. JPEG image steganalysis utilizing both intrablock and interblock correlations. In *Proc. IEEE Int. Symposium on Circuits and Systems*, pages 3029–3032, May 2008.

[3] C. Chen, Y. Q. Shi, and W. Su. A machine learning based scheme for double JPEG compression detection. In *Proc. 19th Int. Conf. on Pattern Recognition*, pages 1–4, Dec. 2008.

[4] J. Fridrich and J. Kodovský. Rich models for steganalysis of digital images. *IEEE Trans. Information Forensics and Security*, 7(3):868–882, 2011.

[5] Z. He, W. Lu, W. Sun, and J. Huang. Digital image splicing detection based on markov features in DCT and DWT domain. *Pattern Recognition*, 45(12):4292–4299, 2012.

[6] V. Holub and J. Fridrich. Designing steganographic distortion using directional filters. In *Proc. IEEE Int. Workshop on Information Forensics and Security*, pages 234–239, 2012.

Table 4: Confusion matrix for identifying the operation types using the SRM features [4]. Please note that the asterisk "*" here denotes that the corresponding accuracy is less than 1%.

Actual \ Predicted	Original	Gaussian Blurring	Gamma Correction	JPEG Compression	Median Filtering	Re-sampling	Splicing
Original	**96.28**	*	1.94	*	*	*	1.24
Gaussian Blurring	*	**99.54**	*	*	*	*	*
Gamma Correction	5.37	*	**93.19**	*	*	*	*
JPEG Compression	*	*	*	**99.03**	*	*	*
Median Filtering	*	*	*	*	**98.76**	*	*
Re-sampling	*	*	*	*	*	**97.49**	*
Splicing	2.69	*	*	2.57	*	*	**93.96**

Table 5: Confusion matrix for identifying the operation types using the LBP features [19]. Please note that the asterisk "*" here denotes that the corresponding accuracy is less than 1%.

Actual \ Predicted	Original	Gaussian Blurring	Gamma Correction	JPEG Compression	Median Filtering	Re-sampling	Splicing
Original	**86.11**	*	11.47	*	*	1.16	1.24
Gaussian Blurring	*	**99.54**	*	*	*	*	*
Gamma Correction	20.03	*	**78.29**	*	*	*	*
JPEG Compression	*	*	*	**99.58**	*	*	*
Median Filtering	*	*	*	*	**99.26**	*	*
Re-sampling	1.01	*	*	*	*	**96.68**	*
Splicing	3.03	*	2.08	6.16	*	*	**87.46**

[7] X. Kang, M. Stamm, A. Peng, and K. Liu. Robust median filtering forensics based on the autoregressive model of median filtered residual. In *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, pages 1–9, Dec. 2012.

[8] S. Knerr, L. Personnaz, and G. Dreyfus. Single-layer learning revisited: a stepwise procedure for building and training a neural network. In *Neurocomputing*, volume 68 of *NATO ASI Series*, pages 41–50. Springer, 1990.

[9] J. Kodovský and J. Fridrich. Calibration revisited. In *Proc. 11th ACM Workshop on Multimedia and Security*, pages 63–74, 2009.

[10] J. Kodovský, J. Fridrich, and V. Holub. Ensemble classifiers for steganalysis of digital media. *IEEE Trans. Information Forensics and Security*, 7(2):432–444, 2012.

[11] B. Li, J. He, J. Huang, and Y. Q. Shi. A survey on image steganography and steganalysis. *Journal of Information Hiding and Multimedia Signal Processing*, 2(2):142–172, 2011.

[12] W. Luo, J. Huang, and G. Qiu. JPEG error analysis and its applications to digital image forensics. *IEEE Trans. Information Forensics and Security*, 5(3):480–491, 2010.

[13] W. Luo, Y. Wang, and J. Huang. Detection of quantization artifacts and its applications to transform encoder identification. *IEEE Trans. Information Forensics and Security*, 5(4):810–815, 2010.

[14] B. Mahdian and S. Saic. Blind authentication using periodic properties of interpolation. *IEEE Trans. Information Forensics and Security*, 3(3):529–538, 2008.

[15] T. Pevný, P. Bas, and J. Fridrich. Steganalysis by subtractive pixel adjacency matrix. *IEEE Trans. Information Forensics and Security*, 5(2):215–224, 2010.

[16] T. Pevný, T. Filler, and P. Bas. Using high-dimensional image models to perform highly undetectable steganography. In *Information Hiding*, volume 6387 of *Lecture Notes in Computer Science*, pages 161–177. Springer, 2010.

[17] T. Pevný and J. Fridrich. Merging markov and DCT features for multi-class JPEG steganalysis. In *Proc. SPIE, Security, Steganography, and Watermarking of Multimedia Contents IX*, volume 6505, page 650503, 2007.

[18] Y. Q. Shi, C. Chen, and W. Chen. A natural image model approach to splicing detection. In *Proc. 9th ACM Workshop on Multimedia and Security*, pages 51–62, 2007.

[19] Y. Q. Shi, P. Sutthiwan, and L. Chen. Textural features for steganalysis. In *Information Hiding*, volume 7692 of *Lecture Notes in Computer Science*, pages 63–77. Springer, 2013.

[20] M. Stamm and K. Liu. Blind forensics of contrast enhancement in digital images. In *Proc. 15th IEEE Int. Conf. on Image Processing*, pages 3112–3115, 2008.

[21] M. Stamm, M. Wu, and K. Liu. Information forensics: An overview of the first decade. *IEEE Access*, 1:167–200, 2013.

[22] M. C. Stamm, X. Chu, and K. Liu. Forensically determining the order of signal processing operations. In *Proc. IEEE Int. Workshop on Information Forensics and Security*, pages 162–167, 2013.

[23] H. Yuan. Blind forensics of median filtering in digital images. *IEEE Trans. Information Forensics and Security*, 6(4):1335–1345, 2011.