Multimedia Tools and Applications (2019) 78:23309–23328
https://doi.org/10.1007/s11042-019-7654-9

Steganalysis aided by fragile detection of image manipulations

Ping Wang[1] · Fenlin Liu[1] · Chunfang Yang[1] · Xiangyang Luo[1]

Received: 21 August 2018 / Revised: 8 April 2019 / Accepted: 15 April 2019 /
Published online: 2 May 2019

Abstract
Steganalysis is usually considered as a two-class classification problem of differentiating between covers and stegos. However, in the real world, the cover image may have undergone various operations, which causes two problems that some processed covers tend to be judged as stegos by the steganalyzer and the stegos processed before information embedding may be easily missed, resulting in the high false alarm rate and the high missed detection rate of steganalysis respectively. To address the former problem, this paper proposed a steganalysis framework based on the combination of the image forensics and the steganalysis tools to reduce the false alarms. First, the fragile detection of image manipulations which is not robust to steganography is applied to separate the normally processed images from the investigated images. Then remaining images are fed to the trained classifier for stegnalysis. The experimental results on gamma transformed images validate the effectiveness of the proposed steganalysis framework that the false alarm rates of steganalysis can be reduced when the investigated image dataset contains normally processed images.

Keywords Fragile detection · Gamma transformation · Image manipulation · Steganalysis

1 Introduction

Steganography is a technique for covert communication, which embeds secret messages into ordinary digital media without drawing suspicion, while steganalysis, the counterpart of steganography, mainly aims to judge whether the unknown digital media carrier secret messages. In the decades, researchers have proposed many effective steganalysis algorithms, where the steganalysis framework based on the feature vector and the classifier has become the mainstream of steganalysis. The prevailing steganalysis algorithms analyze the effects

This work was supported in part by the National Natural Science Foundation of China (No. 61772549, 61872448, U1736214, 61602508, and 61601517), and the National Key R&D Program of China(No. 2016YFB0801303 and 2016QY01W0105).

✉ Fenlin Liu
liufenlin@vip.sina.com

[1] Zhengzhou Science and Technology Institute, Zhengzhou, 450001, China

of steganography on image statistics to construct steganalysis features, such as SPAM [27] and SRM series [9–11, 32]. The features of labeled images are used to train the classifiers, such as the support vector machine SVM [7] and the ensemble classifier [19], which are applied to steganalysis in practice.

However, a common issue to most steganalysis tools is that the investigated images are supposed to be from a single type of images directly, that is, the covers are of the same type and the stegos are generated directly by embedding secret messages into the covers. Moreover, the performances of these steganalysis tools are usually evaluated on the image databases with only a few image types. While in the real world, there are thousands of image types due to the rapid development of image processing technology. It is very easy to deal with digital images for a certain application [12]. Images on the web pages, received by communication tools and in the social network, may have undergone various image manipulations. As all these images can be taken as the covers for steganography, it is key to deal with the heterogeneous images for the application of steganalysis tools in practice. In fact, the statistical distributions of covers may be quite different from that of natural images. Although most existing steganalysis methods achieve excellent results in the certain experimental settings, they may suffer the problem of cover source mismatch when the training set does not contain the images of the same type as the testing images. While it is fallacious to try to train the classifier on a large heterogeneous data set of mixed sources [17], the detection results are not reliable in the real world due to the complicated case where the image may have undergone various image manipulations before information embedding.

There are two possibilities that image manipulations affect the detection performances of steganalysis tools in the real world. On the one hand, image manipulations may change the statistical distributions of cover images, which makes it hard to distinguish a stego from the normally processed images. The steganalysis tools may judge the processed covers as stegos, resulting in a high false alarm rate. While in the real world, the stegos are relatively hardly observed, and high false alarm rates could make the steganalysis system collapse due to a large number of misjudged covers. Therefore, real-world steganalysis should be required to have very low false alarm rates. On the other hand, steganography on the processed images may make the stego statistics similar to the cover images statistics, resulting in a high missed detection rate of steganalysis. Besides. to avoid the problem of cover source mismatch, the training set should cover the types of testing images. However, enlarging the training set and increasing the types of training data may reduce detection accuracies. Besides, due to the huge number of image types, the training set could hardly contain all image types. Hence, most of the steganalysis tools are hardly directly applicable in the real world.

To improve the reliability of steganalysis in the real world, some methods are proposed to deal with heterogeneous images. In [13], He et al. selected the characteristic function moments of the image and its wavelet subbands as features to classify the natural, the stego and the sharpened images. The method aims to reduce false alarms by differentiating stego images and processed images. Considering that a cover may have undergone image manipulations before information embedding, another scheme [1, 15, 21] is proposed for steganalysis on images with different types separately. In [1], Barni et al. used the forensics tools to aid the steganalysis of heterogeneous images. They firstly differentiated camera images and computer-generated images, and then used the steganalyzer explicitly trained to work with images belonging to the correct class. Li et al. [21] applied the image pre-classification to cluster the image into different classes. For each cluster, the steganalyzer is trained separately. However, the image type in each cluster is unknown, while with the knowledge of the image type, a higher accuracy of steganalysis may be achieved. In [15], different tools were selected for steganalysis based on the knowledge of whether the

investigated bitmap images have undergone the JPEG compression or not. Furthermore, with the knowledge of quality factors, a much more reliable detection accuracy could be achieved.

Fragile detection of image manipulation is one kind of image forensics technology which could identify the last applied image operation, while failing to detect the targeted operation if it is followed by another operation, such as steganography. In this paper, we proposed a steganalysis framework based on the combination of the image forensics and the steganalysis tools to attenuate the problem that normally processed images are judged as stegos. Firstly, the normally processed images are separated from the investigated images by fragile detection of image manipulations. Then the steganalysis is conducted on the unlabeled images. The identified normally processed images are judged as covers, which reduces the false alarm rates of steganalysis. Different from existing methods based on the image multi-classification, any effective fragile detection of image manipulations can be applied to the proposed steganalysis framework, ensuring the extensibility of the framework. As image gamma transformation is one of the most frequently-used operations for image contrast changing, it is considered to be a practical tool as an assistant of steganography [31]. The image manipulation gamma transformation, and two steganographic schemes, LSB matching and S-UNIWARD [14], are conducted in the experiments, which validates the effectiveness of the proposed for improving the reliability of steganalysis in the real world. The experimental results show that the steganalysis false alarm rates of the proposed framework is smaller than that of the steganalysis without image forensics where the classifiers are trained on both original covers and heterogeneous images respectively.

2 Steganalysis error probability

Generally, the performance of a steganalyzer is assessed by the average detection error of the covers and stegos. Let C and S note the covers and stegos respectively, N_C and N_S present their numbers, and N stand for the total image number, i.e., $N = N_C + N_S$. Then the steganalysis error probability is

$$P_E = \frac{N_C}{N} P_{FA} + \frac{N_S}{N} P_{MD} \tag{1}$$

where P_{FA} and P_{MD} are the false alarm rate and missed detection rate, which are defined as

$$\begin{cases} P_{FA} = \frac{N_C^S}{N_C} \\ P_{MD} = \frac{N_S^C}{N_S} \end{cases} \tag{2}$$

where N_C^S and N_S^C are the numbers of the misjudged covers and stegos respectively.

At present, most of the prevailing steganalysis schemes use machine learning method where steganalytic features are fed to the classifier. Figure 1 gives the steganalysis framework where the classifier is trained by the features of covers and stegos, and then the trained classifier is used to detect the unknown images. For the sake of clarity and comparison, the steganalysis framework is called as the *traditional* steganalysis framework in this paper. The construction of steganalysis features is motivated to capture the changes on the image by steganography and the steganalysis error comes from two parts, namely, the false alarms that natural images are judged as stegos and the missed detections that stegos are judged as covers.

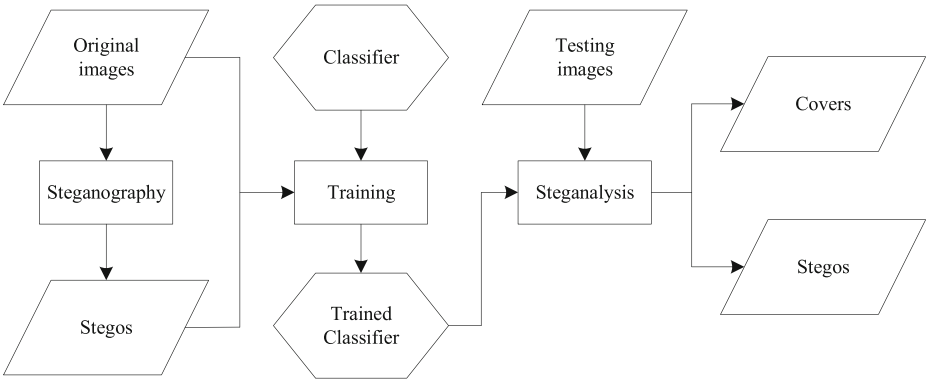


Fig. 1 Steganalysis framework in the traditional mode

While the testing image set consists of heterogeneous images, the false alarm rate can be represented as the sum of the detection errors of all image classes. Assume there are m kinds of images which have undergone various operations. Let C_0 represent the original image, C_i represent the image processed by the i-th operation, where $1 \leq i \leq m$. Then the false alarm rate could be rewritten as

$$\begin{aligned} P_{FA} &= \sum_{i=0}^{m} \frac{N_{C_i}^S}{N_C} \\ &= \sum_{i=0}^{m} \frac{N_{C_i}}{N_C} \cdot \frac{N_{C_i}^S}{N_{C_i}} \end{aligned} \tag{3}$$

For the sake of brevity and readability, we assume the ratio of the covers number and the stegos number are fixed. In this way, we could focus on the analysis of the false alarm rate. Assume that $m = 1$, namely, the covers are consists of two kinds of images: the original images and the images processed by an operation. In this case, the false alarm is

$$\tilde{P}_{FA} = \frac{N_{C_0}}{N_C} \cdot \frac{N_{C_0}^S}{N_{C_0}} + \frac{N_{C_1}}{N_C} \cdot \frac{N_{C_1}^S}{N_{C_1}} \tag{4}$$

Note that if the covers are all original images, then the false alarm rate is

$$\bar{P}_{FA} = \frac{N_{C_0}^S}{N_{C_0}} \tag{5}$$

Then

$$\tilde{P}_{FA} = \bar{P}_{FA} + \frac{N_{C_1}}{N_C} \left(\frac{N_{C_1}^S}{N_{C_1}} - \frac{N_{C_0}^S}{N_{C_0}} \right) \tag{6}$$

When $\frac{N_{C_1}^S}{N_{C_1}} > \frac{N_{C_0}^S}{N_{C_0}}$, $\tilde{P}_{FA} > \bar{P}_{FA}$, which means that if the detection error probability of C_1 is larger than that of the natural images, then the average false alarm rate of steganalysis on heterogeneous images would increase. Moreover, the more images C_1 are in the heterogeneous images, the larger is the false alarm rate. The similar conclusion could be made in the

cases where $m > 1$. Specifically, when $\frac{N_{C_i}^S}{N_{C_i}} = 1$, $1 \le i \le m$, the false alarm rate reaches the maximum that

$$\begin{aligned} P_{FA}^{\max} &= \sum_{i=0}^{m} \frac{N_{C_i}}{N_C} \cdot \frac{N_{C_i}^S}{N_{C_i}} = \frac{N_{C_0}}{N_C} \cdot \frac{N_{C_0}^S}{N_{C_0}} + \sum_{i=1}^{m} \frac{N_{C_i}}{N_C} \cdot \frac{N_{C_i}^S}{N_{C_i}} \\ &= \frac{N_{C_0}}{N_C} \cdot \frac{N_{C_0}^S}{N_{C_0}} + \sum_{i=1}^{m} \frac{N_{C_i}}{N_C} = \frac{N_{C_0}}{N_C} \cdot \frac{N_{C_0}^S}{N_{C_0}} + 1 - \frac{N_{C_0}}{N_C} \\ &= 1 - \frac{N_{C_0}}{N_C} \left(1 - \frac{N_{C_0}^S}{N_{C_0}} \right) \end{aligned} \tag{7}$$

Steganography could be regarded as a special image operation that adds a few noises to images, and the image visual effect is hardly altered. While the normal image operations, such as contrast enhancement, usually change the visual effect for certain applications, they may introduce a lot of noises. When the normally processed images are fed to a steganalysis classifier which is trained to discriminate original images and stegos, they may be judged as stegos with high probabilities. As there is a great number of processed images in the real world, the false alarm rates of traditional steganalysis schemes would be very high in practice applications.

3 Proposed steganalysis framework

An intuitive idea to remove the effects of image operations on the steganalysis performance in the real world is to separate the processed images from the investigated images before steganalysis. In this paper, we take advantage of image forensics.

Image manipulation identification is an image forensics technology which could judge whether an image has undergone the specific operation. At present, the research of image manipulation identification technology has made a lot of achievements, including the detection of median filtering [18, 26, 42], contrast changing [5, 29], blurring [4, 22, 44], rescaling [3] and JPEG compression [2, 16, 38]. Most of the existing image manipulation detection tools are capable of detecting the last operation applied to the investigated image even if the image has undergone various operations. However, if the targeted operation is followed by another operation, some forensics tool may fail. As the counterpart of robust detection, the detection of the targeted operation is called as fragile detection to the post-operation which could make the forensics tool fail.

While the fragility of an image manipulation detection may be taken as a weakness in image forensics [33, 34], it could be exploited to assist the steganalysis in the real world. Consider the case where the investigated images include original images, normally processed images and stegos. Note that the stego may have undergone image operations before information embedding. If an image is a stego, then the fragile detection of image manipulations would not judge the stego as the normally processed image due to its fragility to steganography. Moreover, if an image is a normally processed image, then it does not carry any secret message. Therefore, the false alarm rate of steganalysis could be reduced by the fragile detection of image manipulations applied before the steganalyzer. Figure 2 gives the proposed steganalysis framework based on the fragile detection of image manipulations. Firstly, the fragile forensics tools are applied to separate the normally processed images from the investigated images, and then the remaining images are steganalyzed.

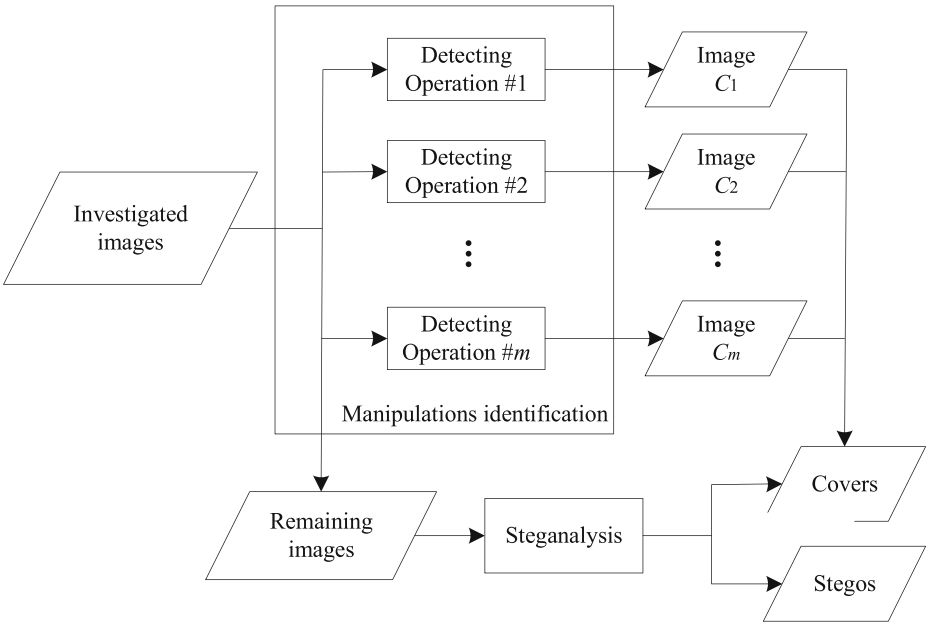


Fig. 2 Steganalysis framework based on the combination of the image forensics and the steganalysis

It is worth noting that the proposed framework is different from the one proposed in [13] which uses the characteristic function moments for multiclassification. In this paper, the specific images are separated by the corresponding fragile forensics tools before steganalysis. Any fragile forensics tools for image manipulations detection could be applied to the proposed framework. Thus, it possesses the feature of extensibility.

4 Steganalysis aidded by gamma transformation detection

In this section, we are considering the case where gamma transformation is involved. Firstly, the application of gamma transformation in the steganography [31] is reviewed. Then we construct a new feature for gamma transformation detection. Finally, a steganalysis scheme aided by gamma transformation detection is presented.

4.1 Application of gamma transformation in steganography

Image gamma transformation is an operation for image contrast changing, whose basics form is $s = r^{\gamma}$, where $r \in [0, 1]$ is the input and $\gamma > 0$ is the only decisive parameter which controls the direction and intensity of the transformation. Generally, due to the limited storage space, the pixel values of the digital image need to be truncated. In this paper, we consider the 8-bit images, which is widely used as covers in steganography. Thus, the form of image gamma transformation is

$$y = \text{round}\left(255 \times \left(\frac{x}{255}\right)^{\gamma}\right) \tag{8}$$

where $x, y \in \{n|n \in [0, 255] \cap \mathbf{Z}\}$ represent image pixel values before and after gamma transformation respectively, and round$(\cdot)$ is the rounding operation.

As widely used in image processing, gamma transformation is considered to be a practical tool as an assistant of steganography. In [31] (in Chinese with English abstract), Sun et al. analyzed the deviation of an image statistical feature and pointed out that the gamma transformed images tend to be judged as stegos in steganalysis. Based on the conclusion, they proposed a steganography scheme that embeds information into gamma transformed images. Firstly, the image is gamma transformed with parameter $\gamma = 1 + \Delta$, where Δ is the disturbance factor of the gamma transformation parameter. Then the secret message is embedded into the gamma transformed images. In this way, the normal gamma transformed images will be judged as the stegos, resulting in a high false alarm rate.

4.2 Fragile detection of gamma transformation

If the gamma transformed images without information embedding could be separated from the investigated images, then the steganalysis performance would be improved. The idea can be achieved by fragile detection of gamma transformation as it is effective for the detection of the last applied gamma transformation, while it will fail if the gamma transformed images have undergone steganography at last.

For gamma transformation detection, Stamm and Liu [30] exploited the high-frequency coefficients of image histogram characteristics function. They pointed out that after gamma transformation, the high-frequency coefficients will be enlarged. Therefore, they constructed a feature by averaging the high-frequency coefficients to detect gamma transformation. Cao et al. [6] proposed a gamma transformation detection method based on the number of image histogram gaps. They observed that after gamma transformation, many gaps emerge that their values are zero and their adjacent histogram bins are nonzero. In our prior works [35, 36], we analyzed effects of gamma transformation on the image histogram and pointed out that gamma transformation introduces zero-value histogram bins whose locations are closely related to the gamma transformation parameter. Based on the conclusion, we proposed a manipulation detector and a parameter estimator for image gamma transformation.

The methods above have more or less taken advantage of the zero-value histogram bins introduced by gamma transformation. While after steganography, the zero-value histogram bins may be filled, therefore these methods are fragile to steganography. Based on this feature, the zero-value histogram bins could be exploited to separate the gamma transformed images without information embedding from stegos, and it need not consider whether the stegos have undergone gamma transformation.

More specifically, we assume that the steganography applies $\pm K$ operation to pixels to embed the information, such as EA [23], HUGO [28], S-UNIWARD [14], and HILL [20] etc. Then the relationship between the image histograms before and after steganography could represented by the following equation

$$h_s(n) = \sum_k \alpha_{n-k,k} h_c(k) \tag{9}$$

where $h_s(n)$ and $h_c(n)$ are the histogram bins of the stego and the cover at n respectively, $\alpha_{i,j} \geq 0$ and $\sum_i \alpha_{i,j} = 1$. As to the steganography LSB matching with the payload ρ, $\alpha_{-1,j} = \alpha_{1,j} = 0.25\rho$ and $\alpha_{0,j} = 1 - 0.5\rho$. Therefore, the zero-value histogram bin

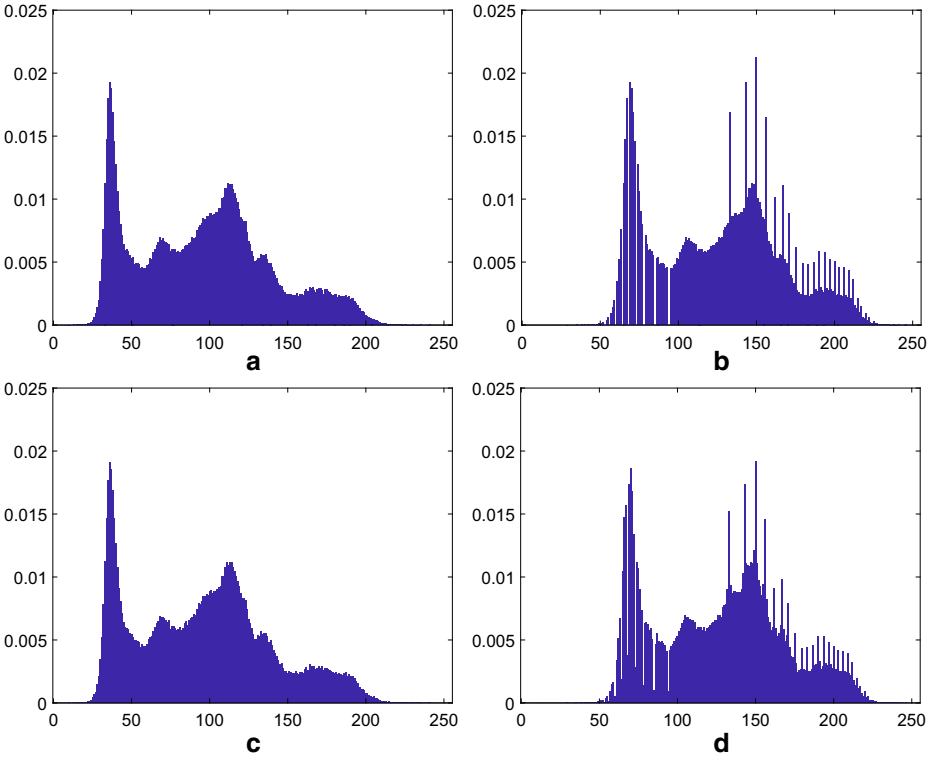


Fig. 3 Histograms of image Lena. **a** original; **b** gamma transformed with $\gamma = 0.67$; **c** LSB matching on (**a**) with $\rho = 0.4$; **d** LSB matching on (**b**) with $\rho = 0.4$

will be filled due to the shares of its adjacent nonzero histogram bins. Figure 3 shows the histograms of the typical image *Lena* with size 512×512 under gamma transformation and LSB matching. The parameter of gamma transformation is $\gamma = 0.67$ and the payload of LSB matching is $\rho = 0.4$. Figure 3a and b are the histograms of the original image and the gamma transformed image, and Fig. 3c and d are the histograms of the stegos which are generated by embedding information into the original image and the gamma transformed image respectively. It is showed that only the histogram of gamma transformed image in Fig. 3b owns many zero-value histogram bins.

However, the natural images and stegos may also have zero-value histogram bins or unsmooth histogram envelopes. Therefore, the existing gamma transformation detection will judge these images as the gamma transformed, which led to the missed detection of steganalysis. Considering the issue, we constructed a new feature based on the zero-value histogram bins and their adjacent histogram bins values. It is observed that in the histograms of natural images and stegos, the values of histogram bins adjacent to the zero-value histogram bins are usually small, as shown in Fig. 3a, c and d, while in the histogram of the gamma transformed images without information embedding they are relatively large, Fig. 3b. Besides, the numbers of zero-value histogram bins of the gamma transformed images without information embedding are usually larger than that of the natural images and stegos. Based on the observation, we multiply the two histogram bins adjacent to the

zero-value histogram bin, and take the sum of all the products as the feature to detect gamma transformation, namely,

$$F = \sum_{x \in \Phi} h(x-1) \cdot h(x+1) \tag{10}$$

where $\Phi = \{x | h(x) = 0, 0 < x < 255\}$ is zero-value histogram bin locations set. In the construction of feature F, we consider both the number of zero-value histogram bins and the values of histogram bins adjacent to them. By this way, the value of F is small for the natural image and the stego, while it is large for the gamma transformed image. Therefore, using F to detect gamma transformation, we can avoid judging the stegos as gamma transformed images, which ensures the low missed detection rate of steganalysis when the forensics tool is used to reduce the false alarm rate.

4.3 Steganalysis aided by gamma transformation detection

Figure 4 show the flow chart of steganalysis aided by gamma transformation detection. Given an unknown image, the feature F is extracted firstly to detect gamma transformation. Based on the result, it is decided whether or not to further apply the steganalysis tool. In this paper, we use the steganalysis feature SRM [11] and the ensemble classifier [19] for steganalysis. The detailed steps are as follows.

#1 Calculate the image histogram $h(x)$.
#2 Find the zero-value histogram bin locations Φ.
#3 Calculate the image feature F according to (10).
#4 Detect gamma transformation according to the predefined threshold η using the following rule

$$\delta = \begin{cases} \text{image is gamma transformed,} & F > \eta \\ \text{image is not gamma transformed,} & F \leq \eta \end{cases} \tag{11}$$

If gamma transformation is presented, then judge the image as the cover; otherwise, take the next step.
#5 Extract the steganalysis feature SRM.
#6 Feed the SRM to the trained ensemble classifier for steganalysis.

5 Experimental results

We use 10,000 original images in BossBase-1.01[1] to validate the effectiveness of the proposed framework. The original images are with fixed size 512×512 coming from rescaled and cropped natural images of various sizes. For experiments, They are firstly gamma transformed with parameter $\gamma = 1 + \Delta$, where $\Delta \in \{\pm 0.1, \pm 0.2\}$. Meanwhile, the original image is considered as the gamma transformed with $\gamma = 1$, namely, $\Delta = 0$. Then, information is embedded into each image by LSB matching and S-UNIWARD with payload $\rho \in \{0.1, 0.2, 0.3, 0.4\}$. Hence, there are 450, 000 images in total in the constructed image dataset. At last, the proposed framework is applied for steganalysis.

At first, the performance of the feature F for gamma transformation detection is tested. Then the detected gamma transformed images are separated from the image dataset, and the

[1] http://agents.fel.cvut.cz/stegodata/

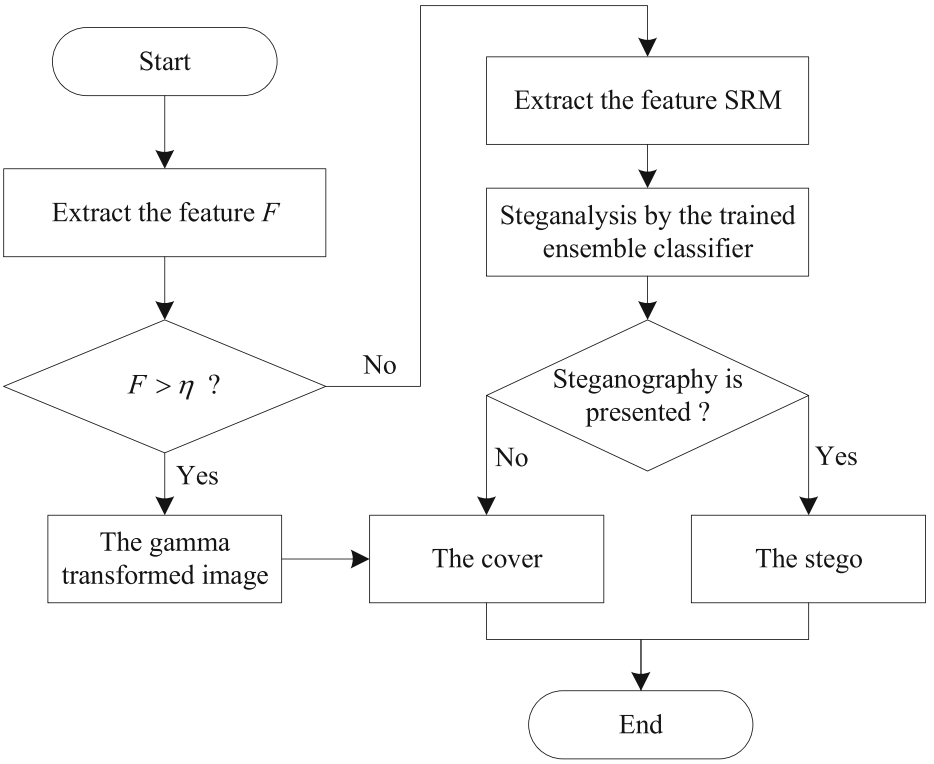


Fig. 4 Flow chart of steganalysis aided by gamma transformation detection

remaining images are fed to the trained classifier for steganalysis. Images in each class are divided into two parts. One is for classifier training and threshold setting, and the other is for performance testing, including gamma transformation detection and steganalysis. It is worth mentioning that the missed detections of steganalysis should include the false positives of gamma transformation detection that the stegos are judged as gamma transformed. Meanwhile, the false alarm rate of steganalysis is the ratio of the number of misjudged covers to the number of all covers, instead of to the number of covers fed to the ensemble classifier.

5.1 Gamma transformation detection

We randomly selected 5,000 images to test the performance of gamma transformation detection. For contrast experiments, the methods proposed in [30] and [6] are used, which are called as STM and CGM in the rest of this paper. Figure 5 presents the ROC curves of gamma transformation detection results. Here, the gamma transformed image is considered as positive. The results show that the proposed method outperforms the other two methods. Besides, the proposed method achieves a high detection accuracy of gamma transformation at a very low false positive rate. For example, at the false positive rate of 0.02, the detection accuracies are 99.86%, 99.77%, 94.7%, and 94.97% when $\Delta = -0.2, -0.1, 0.1$ and 0.2 respectively. This is significantly important to the followed steganalysis, because the low false positive rate of gamma transformation detection means few stegos are separated out in the gamma transformation detection.

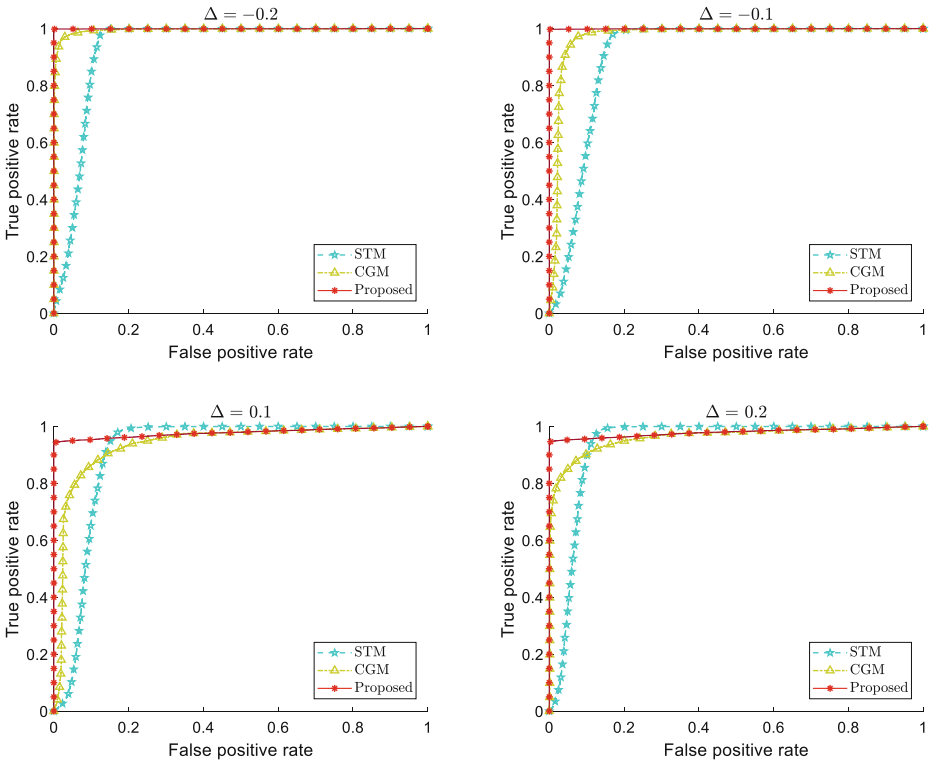


Fig. 5 ROC curves of gamma transformation detection results

In practice, gamma transformation detection according to the value of feature F needs a predefined threshold η. Figure 6 shows the distributions of the feature F of the original images, the gamma transformed images without information embedding and the stegos generated by LSB matching, where the payload $\rho = 0$ represents the images which have not undergone steganography. There are 5,000 images in each class, and the region where the values are larger than 6×10^{-7} is compressed for a clear comparison between the gamma transformed images without information embedding and the other kinds of images. It is showed in Fig. 6 that the F values of original images and stegos are all smaller than 6×10^{-7}, while most of the gamma transformed images without information embedding are over than 6×10^{-7}. It indicates that by the thresholding method based on the F value, the gamma transformed images without information embedding could be separated from the original images and stegos. Considering the missed detection rate of steganalysis, the value of η should be large so that few stegos will be judged as the gamma transformed. While a too large threshold will lead to a high false alarm rate of steganalysis that many gamma transformed images will be missed detected and further undergo steganalysis. According to the experimental results, the threshold is set as $\eta = 10^{-6}$.

Table 1 gives the error probability of gamma transformation detection for each kind of images using the feature F when $\eta = 10^{-6}$. The results show that images that have not undergone gamma transformation ($\Delta = 0$) are all correctly classified, whether the steganography is applied or not. With respect to LSB matching, all stegos are judged as the not gamma transformed, while there are a few stegos generated by S-UNIWARD that

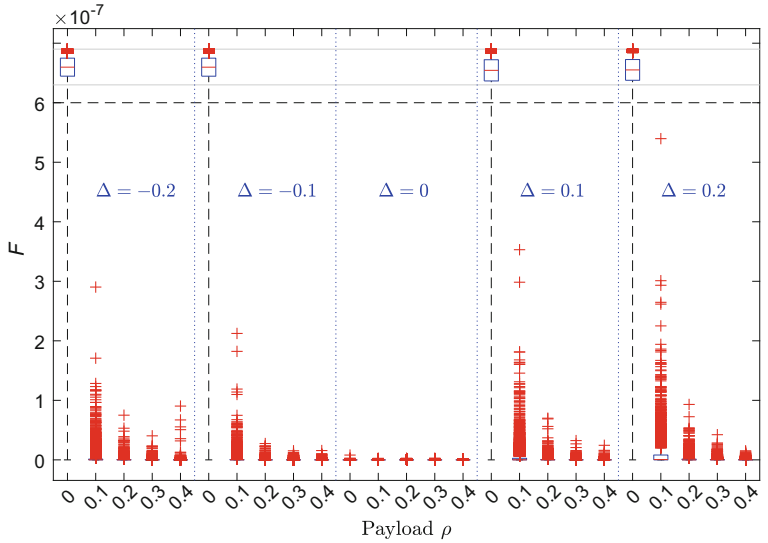


Fig. 6 Plots of F values of heterogeneous images. The region where the values are larger than 6×10^{-7} is compressed. The F values of original images and stegos are all smaller than 6×10^{-7} while most of the gamma transformed images are over than 6×10^{-7}

are judged as the gamma transformed, which are going to be judged as covers directly without steganalysis. Therefore, judging the stegos as the gamma transformed may enlarge the missed detection rate of steganalysis. Besides, there are some false detections of gamma transformed images. As the zero-value histogram bins introduced by gamma transformation only distribute in one side of histogram [35, 36], there may be no zero-value histogram bins in the histograms except for that in the two histogram sides. Therefore, these images will be further steganalyzed.

5.2 Steganalysis

After gamma transformation detection, the remaining images are going to be steganalyzed. In the proposed mode, the classifier trained on the original images and the corresponding stegos is used, and only the images which are not judged as not gamma transformed are tested. Steganalysis in the traditional mode is compared in two ways due to the types of the

Table 1 Error probability of gamma transformation detection for each kind of images

Δ	$\rho = 0$	LSB Matching				S-UNIWARD			
		$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$	$\rho = 0.4$	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$	$\rho = 0.4$
−0.2	0.0108	0	0	0	0	0.082	0.0195	0.007	0.0019
−0.1	0.014	0	0	0	0	0.0652	0.0175	0.006	0.0019
0	0	0	0	0	0	0	0	0	0
0.1	0.1825	0	0	0	0	0.0638	0.0226	0.0088	0.0030
0.2	0.1595	0	0	0	0	0.0838	0.0302	0.0147	0.0054

dataset for classifier training. Namely, the classifier is trained on the original images and the corresponding stegos, or on the heterogeneous images. The former and the latter trained ensemble classifiers are referred to as S-EC and M-EC respectively here. All investigated images in the testing sets are steganalyzed in the traditional mode. For the experiments of steganalysis in the traditional mode using M-EC, one fifth of each type of images are randomly selected to form the heterogeneous images set for classifier training and testing. We test 10 times for each image class and take the average of the 10 results for the performance evaluation.

5.2.1 Stegnalysis of LSB matching

Table 2 shows the false alarm rates (P_{FA}), the missed detection rates (P_{MD}), and the average error probabilities (P_E) of steganalysis of LSB matching, where the rows indexed by *Mixed* present the overall steganalysis results of gamma transformed images with all parameters including $\rho = 0$ under each payload. In the traditional mode using S-EC, many gamma transformed images are judges as stegos, resulting in high false alarm rates. While in the proposed mode, most of the gamma transformed images without information embedding are separated in the gamma transformation detection, which reduces the probability of covers

Table 2 Final results of steganalysis of LSB matching

Δ	Traditional mode					Proposed mode				
	Classifier	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$	$\rho = 0.4$	Classifier	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$	$\rho = 0.4$
False alarm rates P_{FA}										
−0.2	S-EC	0.4978	0.2267	0.1027	0.0658	S-EC	0.0041	0.0027	0.0014	0.0002
−0.1	S-EC	0.4979	0.2514	0.1236	0.0701	S-EC	0.0037	0.0025	0.0009	0.0004
0	S-EC	0.0859	0.051	0.0331	0.0238	S-EC	0.0862	0.0511	0.0333	0.024
0.1	S-EC	0.3227	0.1292	0.0559	0.0335	S-EC	0.0397	0.0177	0.0067	0.0026
0.2	S-EC	0.3935	0.1442	0.0526	0.0286	S-EC	0.0324	0.0123	0.0042	0.0013
Mixed	M-EC	0.087	0.0464	0.029	0.0202	S-EC	0.0332	0.0173	0.0093	0.0057
Missed detection rates P_{MD}										
−0.2	S-EC	0.0377	0.0297	0.0211	0.0128	S-EC	0.0283	0.0296	0.0197	0.013
−0.1	S-EC	0.0207	0.018	0.0159	0.0125	S-EC	0.02	0.0189	0.0146	0.012
0	S-EC	0.0707	0.0415	0.0265	0.0192	S-EC	0.0698	0.0413	0.0272	0.0185
0.1	S-EC	0.0216	0.0177	0.0159	0.0121	S-EC	0.0208	0.0181	0.0154	0.012
0.2	S-EC	0.0183	0.0158	0.0146	0.011	S-EC	0.0169	0.017	0.0138	0.0112
Mixed	M-EC	0.0709	0.0371	0.0208	0.0137	S-EC	0.0312	0.025	0.0182	0.0134
Average error probabilities P_E										
−0.2	S-EC	0.2678	0.1282	0.0619	0.0393	S-EC	0.0162	0.0162	0.0105	0.0066
−0.1	S-EC	0.2593	0.1347	0.0697	0.0413	S-EC	0.0119	0.0107	0.0078	0.0062
0	S-EC	0.0783	0.0462	0.0298	0.0215	S-EC	0.078	0.0462	0.0303	0.0213
0.1	S-EC	0.1721	0.0734	0.0359	0.0228	S-EC	0.0303	0.0179	0.0111	0.0073
0.2	S-EC	0.2059	0.08	0.0336	0.0198	S-EC	0.0246	0.0147	0.009	0.0063
Mixed	M-EC	0.0789	0.0418	0.0249	0.0169	S-EC	0.0322	0.0211	0.0137	0.0095

being judged as stegos. Therefore, the false alarm rates in the proposed mode are relatively low, which contributes to the small average detection error probabilities.

As no stegos are judged as the gamma transformed in the gamma transformation detection, the final missed detection rates of steganalysis in the traditional and proposed modes are almost equal because they have the same trained classifier and the same testing images. The only difference in the results is caused by the randomicity of the selected feature subspaces. Besides, the missed detection rates when $\Delta = 0$ are higher than that when $\Delta \neq 0$, which indicates that the stegos generated by embedding information into gamma transformed images are easier to detect than that have not undergone gamma transformation. However, due to the high false alarm rates in the traditional mode when $\Delta \neq 0$, the average error probabilities are larger than that when $\Delta = 0$.

Using M-EC in the traditional mode could reduce the false alarms, especially for the steganalysis of the types of images which are gamma transformed, while the missed detection rate of the corresponding stegos is increased. In general, the performance of M-EC is better than S-EC in the traditional mode, but less than that in the proposed mode. In fact, for the application of steganalysis in the real world, it is almost impossible to train a universal classifier which is capable of detecting the stego in the heterogeneous images. Besides, the classifier trained on heterogeneous images may capture little information about the steganography as the main difference of a cover and a stego may be introduced by the image manipulations. Integrating a large number of different types of images into the training dataset will deteriorate the steganalysis performance of the trained classifier.

Overall, when the investigated image set includes gamma transformed images, the proposed framework can improve the reliability of steganalysis of LSB matching. On one hand, the false alarm rates are reduced by separating the gamma transformed images without information embedding from the investigated images. On the other hand, embedding information into the gamma transformed images by LSB matching will make it easier to detect the stegos. In this point of view, if the proposed steganalysis framework is applied and the results of gamma transformation detection are reliable, the security of steganography on gamma transformed images is lower than that on the natural images.

5.2.2 Stegnalysis of S-UNIWARD

Table 3 gives the false alarm rates (P_{FA}), the missed detection rates (P_{MD}), and the average error probabilities (P_E) of steganalysis of S-UNIWARD. The results show that the proposed mode reduces the false alarms of steganalysis and decreases the average error probabilities. However, due to the false positives of gamma transformation detection that stegos are judged as gamma transformed images, the missed detection rates of steganalysis are higher than that in the traditional mode.

In addition, in the traditional mode there is an exceptional case that the false alarm rates when $\Delta = -0.2$ is relatively smaller than that when $\Delta = -0.1$, 0.1, and 0.2. Specially, when $\rho = 0.3$, the false alarm rate when $\Delta = -0.2$ is even smaller than that when $\Delta = 0$. Moreover, the missed detection rates of steganalysis when $\Delta = -0.2$ are higher than others, which indicates that when $\Delta = -0.2$, the SRM features of the stegos generated by S-UNIWARD following the gamma transformation with parameter $\gamma = 0.8$ are more similar to the features of original images than that of other stegos. In this case, the proposed steganslysis framework is not capable of reducing the missed detections of steganalysis. While a possible solution to this problem is to judge that whether the investigated images have undergone gamma transformation and then use the classifier trained on the gamma transformed images to steganalyze the identified images, it is beyond the scope of this paper.

Table 3 Final results of steganalysis of S-UNIWARD

Δ	Traditional mode					Proposed mode				
	Classifier	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$	$\rho = 0.4$	Classifier	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$	$\rho = 0.4$
False alarm rates P_{FA}										
−0.2	S-EC	0.4386	0.3991	0.2549	0.3172	S-EC	0.0078	0.0064	0.0058	0.005
−0.1	S-EC	0.6818	0.6746	0.4924	0.51	S-EC	0.0082	0.0073	0.0062	0.0052
0	S-EC	0.4136	0.3359	0.2714	0.2201	S-EC	0.4141	0.3311	0.2716	0.2254
0.1	S-EC	0.6836	0.5984	0.4541	0.4017	S-EC	0.0959	0.0701	0.0499	0.0396
0.2	S-EC	0.6358	0.5666	0.4113	0.3829	S-EC	0.073	0.0483	0.0344	0.0268
Mixed	M-EC	0.3926	0.3102	0.2489	0.193	S-EC	0.1198	0.0926	0.0736	0.0604
Missed detection rates P_{MD}										
−0.2	S-EC	0.5101	0.436	0.4641	0.3022	S-EC	0.5471	0.4849	0.4673	0.2862
−0.1	S-EC	0.2623	0.1656	0.1808	0.1077	S-EC	0.3058	0.1966	0.1831	0.1034
0	S-EC	0.3896	0.3078	0.2411	0.1892	S-EC	0.3884	0.3069	0.2387	0.1876
0.1	S-EC	0.1802	0.1317	0.1287	0.0909	S-EC	0.2303	0.1616	0.1362	0.0882
0.2	S-EC	0.2265	0.1563	0.1615	0.1073	S-EC	0.2901	0.2004	0.171	0.1071
Mixed	M-EC	0.4003	0.3051	0.2339	0.1731	S-EC	0.3523	0.2701	0.2393	0.1545
Average error probabilities P_E										
−0.2	S-EC	0.4743	0.4175	0.3595	0.3097	S-EC	0.2775	0.2457	0.2366	0.1456
−0.1	S-EC	0.4724	0.4201	0.3366	0.3088	S-EC	0.157	0.102	0.0947	0.0543
0	S-EC	0.4016	0.3219	0.2563	0.2047	S-EC	0.4013	0.319	0.2552	0.2065
0.1	S-EC	0.4319	0.365	0.2914	0.2463	S-EC	0.1631	0.1159	0.0931	0.0639
0.2	S-EC	0.4312	0.3614	0.2864	0.2451	S-EC	0.1816	0.1244	0.1027	0.067
Mixed	M-EC	0.3964	0.3077	0.2414	0.1831	S-EC	0.2361	0.1814	0.1565	0.1075

6 Conclusions and future works

This paper addresses the problem that the normally processed images may be judged as stegos by the steganalyzer, resulting in high false alarm rates of steganalysis in the real world. A steganalysis framework based on the combination of image forensics and steganalysis tools is proposed to improve the reliability of steganalysis in the real world. Firstly, the unknown image is investigated by fragile forensics of image manipulations. If no operations are presented on the image, then it is further detected by steganalyzers. Any fragile forensics tools for image manipulations detection could be applied to the proposed steganalysis framework to improve the steganalysis performance in the real world. The validation of the proposed steganalysis framework is verified by combining the steganalysis of LSB matching and S-UNIWARD with gamma transformation detection, where a new feature is constructed as the sum of products of two histogram bins adjacent to zero-value histogram bins. The false alarm rate is reduced by separating the gamma transformed images without information embedding from the investigated images.

However, the proposed steganalysis framework introduces more factors which affect the performance of steganalysis. The steganalysis reliability now depends on both the accuracies of image manipulation identification and stego detection. For example, with a low accuracy of the image operation forensics, many normally processed images will undergo

steganalysis, which could result in a high false alarm rate of steganalysis. Meanwhile, if stegos are picked out by an image operation detector, they will be judged as covers directly without applying steganalyzers. Therefore, reliable image manipulation forensics tools are significantly important to improve the performance of steganalysis in the real world.

This paper focuses on reducing the false alarms of steganalysis in practice. While in reality, the information may be embedded into the processed images, and the stegos may trend to be judged as covers by steganalysis tools. In this case, the proposed steganalysis are not capable of reducing the missed detection rate. For the problem, further researches are needed to find the corresponding solutions.

Besides, our future research directions include but not be limited as follows. We try to extend our idea to process other type of data [24, 25, 37]. We also want to adopt multi-core CPU and many-core GPU parallel techniques [39, 45] to accelerate our algorithms in processing big image data [8, 40, 41, 43].

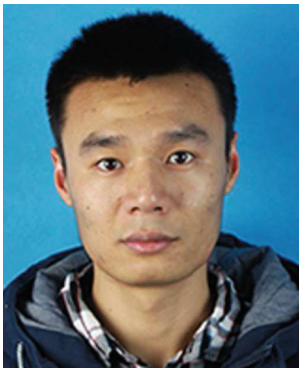
Acknowledgements The authors would like to thank the anonymous reviewers for their thorough comments and suggestions that helped to improve this paper.

References

1. Barni M, Cancelli G, Esposito A (2010) Forensics aided steganalysis of heterogeneous images. In: Proceedings of the IEEE international conference on acoustics, speech, and signal processing, ICASSP 2010. IEEE, Dallas, pp 1690–1693, https://doi.org/10.1109/ICASSP.2010.5495494
2. Bianchi T, Piva A (2012) Detection of nonaligned double jpeg compression based on integer periodicity maps. IEEE Trans Inf Forensics Secur 7(2):842–848. https://doi.org/10.1109/TIFS.2011.2170836
3. Birajdar GK, Mankar VH (2014) Blind method for rescaling detection and rescale factor estimation in digital images using periodic properties of interpolation. AEU - Int J Electron Commun 68(7):644–652. https://doi.org/10.1016/j.aeue.2014.01.013
4. Cao G, Zhao Y, Ni R (2010) Edge-based blur metric for tamper detection. J Inf Hiding Multimed Signal Process 1(1):20–27
5. Cao G, Zhao Y, Ni R (2010) Forensic estimation of gamma correction in digital images. In: Proceedings of the international conference on image processing, ICIP 2010. IEEE, Hong Kong, pp 2097–2100, https://doi.org/10.1109/ICIP.2010.5652701
6. Cao G, Zhao Y, Ni R, Li X (2014) Contrast enhancement-based forensics in digital images. IEEE Trans Inf Forensics Secur 9(3):515–525. https://doi.org/10.1109/TIFS.2014.2300937
7. Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol (TIST) 2(3):27:1–27:27. https://doi.org/10.1145/1961189.1961199
8. Chen X, He F, Yu H (2018) A matting method based on full feature coverage. In: Multimedia tools and applications. https://doi.org/10.1007/s11042-018-6690-1
9. Denemark T, Sedighi V, Holub V, Cogranne R, Fridrich J (2014) Selection-channel-aware rich model for Steganalysis of digital images. In: 2014 IEEE international workshop on information forensics and security, WIFS 2014. IEEE, Atlanta, pp 48–53. https://doi.org/10.1109/WIFS.2014.7084302
10. Denemark T, Fridrich J, Alfaro PC (2016) Improving selection-channel-aware steganalysis features. In: Alattar AM, Memon ND (eds) Proceedings of IS&T, electronic imaging, media watermarking, security, and forensics 2016. Ingenta, San Francisco, pp 1–8
11. Fridrich J, Kodovský J (2012) Rich models for steganalysis of digital images. IEEE Trans Inf Forensics Secur 7(3):868–882. https://doi.org/10.1109/TIFS.2012.2190402
12. Gonzalez RC, Woods RE (2007) Digital image processing. Prentice Hall, Upper Saddle River
13. He X, Liu F, Luo X, Yang C (2009) Classification between PS and Stego Images Based on Noise Model. In: 2009 third international conference on multimedia and ubiquitous engineering, MUE 2009. IEEE Computer Society, Qingdao, pp 31–36, https://doi.org/10.1109/MUE.2009.16
14. Holub V, Fridrich J (2013) Digital image steganography using universal distortion. In: ACM information hiding and multimedia security workshop, IH&MMSec '13. ACM, Montpellier, pp 59–68. https://doi.org/10.1145/2482513.2482514. http://doi.acm.org/10.1145/2482513.2482514

15. Hou X, Zhang T, Xiong G, Wan B (2012) Forensics aided steganalysis of heterogeneous bitmap images with different compression history. KSII Trans Internet Inf Syst 6(8):1926–1945. https://doi.org/10.3837/tiis.2012.08.003
16. Huang F, Huang J, Shi Y (2010) Detecting double JPEG compression with the same quantization matrix. IEEE Trans Inf Forensics Secur 5(4):848–856. https://doi.org/10.1109/TIFS.2010.2072921
17. Ker AD, Bas P, Böhme R, Cogranne R, Craver S, Filler T, Fridrich J, Pevný T (2013) Moving steganography and steganalysis from the laboratory into the real world. In: ACM information hiding and multimedia security workshop, IH&MMSec '13. ACM, Montpellier, pp 45–58, https://doi.org/10.1145/2482513.2482965
18. Kirchner M, Fridrich J (2010) On detection of median filtering in digital images. In: Media forensics and security II, part of the IS&T-SPIE electronic imaging symposium, vol 7541. SPIE, San Jose, p 754110. https://doi.org/10.1117/12.839100
19. Kodovský J, Fridrich J, Holub V (2012) Ensemble classifiers for steganalysis of digital media. IEEE Trans Inf Forensics Secur 7(2):432–444
20. Li B, Wang M, Huang J, Li X (2014) A new cost function for spatial image steganography. In: 2014 IEEE international conference on image processing, ICIP 2014. IEEE, Paris, France, pp 4206–4210. https://doi.org/10.1109/ICIP.2014.7025854
21. Li W, Zhang T, Hu G, Xie K (2014) Image Pre-classification to improve accuracy of universal steganalysis. In: 2014 5th IEEE international conference on software engineering and service science, ICSESS 2014. IEEE, Beijing, pp 364–368
22. Liu G, Wang J, Lian S, Dai Y (2013) Detect image splicing with artificial blurred boundary. Math Comput Model 57(11–12):2647–2659. https://doi.org/10.1016/j.mcm.2011.06.026
23. Luo W, Huang F, Huang J (2010) Edge adaptive image steganography based on LSB matching revisited. IEEE Trans Inf Forensics Secur 5(2):201–214. https://doi.org/10.1109/TIFS.2010.2041812
24. Lv X, He F, Cai W, Cheng Y (2018) Supporting selective undo of string-wise operations for collaborative editing systems. Futur Gener Comput Syst 82:41–62. https://doi.org/10.1016/j.future.2017.11.046. http://www.sciencedirect.com/science/article/pii/S0167739X1730897X
25. Lv X, He F, Cheng Y, Wu Y (2018) A novel CRDT-based synchronization method for real-time collaborative CAD systems. Adv Eng Inf 38:381–391. https://doi.org/10.1016/j.aei.2018.08.008. http://www.sciencedirect.com/science/article/pii/S147403461730486X
26. Niu Y, Zhao Y, Ni R (2017) Robust median filtering detection based on local difference descriptor. Signal Process Image Commun 53:65–72. https://doi.org/10.1016/j.image.2017.01.008
27. Pevný T, Bas P, Fridrich J (2010) Steganalysis by subtractive pixel adjacency matrix. IEEE Trans Inf Forensics Secur 5(2):215–224. https://doi.org/10.1109/TIFS.2010.2045842
28. Pevný T, Filler T, Bas P (2010) Using high-dimensional image models to perform highly undetectable steganography. In: Information hiding - 12th international conference, IH 2010. Springer, Calgary, pp 161–177
29. Rosa AD, Fontani M, Massai M, Piva A, Barni M (2015) Second-order statistics analysis to cope with contrast enhancement counter-forensics. IEEE Signal Process Lett 22(8):1132–1136. https://doi.org/10.1109/LSP.2015.2389241
30. Stamm MC, Liu KJR (2010) Forensic detection of image manipulation using statistical intrinsic fingerprints. IEEE Trans Inf Forensics Secur 5:492–506. https://doi.org/10.1109/TIFS.2010.2053202
31. Sun X, Zhang W, Yu N, Wei Y (2017) Steganography based on parameters' disturbance of spatial image transform. J Commun 38(10):166–174. ISSN=1000–436X
32. Tang W, Li H, Luo W, Huang J (2014) Adaptive steganalysis against WOW embedding algorithm. In: ACM information hiding and multimedia security workshop, IH&MMSec '14. ACM, Salzburg, pp 91–96. https://doi.org/10.1145/2600918.2600935
33. Wan W, Wang J, Li J, Meng L, Sun J, Zhang H, Liu J (2018) Pattern complexity-based JND estimation for quantization watermarking. Pattern Recogn Lett. https://doi.org/10.1016/j.patrec.2018.08.009. http://www.sciencedirect.com/science/article/pii/S0167865518304033
34. Wan W, Wang J, Li J, Sun J, Zhang H, Liu J (2018) Hybrid JND model-guided watermarking method for screen content images. Multimed Tools Appl. https://doi.org/10.1007/s11042-018-6860-1
35. Wang P, Liu F, Yang C, Luo X (2018) Blind forensics of image gamma transformation and its application in splicing detection. J Vis Commun Image Represent 55:80–90. https://doi.org/10.1016/j.jvcir.2018.05.020
36. Wang P, Liu F, Yang C, Luo X (2018) Parameter estimation of image gamma transformation based on zero-value histogram bin locations. Signal Process Image Commun 64:33–45. https://doi.org/10.1016/j.image.2018.02.011

37. Wu Y, He F, Zhang D, Li X (2018) Service-oriented feature-based data exchange for cloud-based design and manufacturing. IEEE Trans Serv Comput 11(2):341–353. https://doi.org/10.1109/TSC.2015.2501981. doi.ieeecomputersociety.org/10.1109/TSC.2015.2501981
38. Yang J, Xie J, Zhu G, Kwong S, Shi Y (2014) An effective method for detecting double jpeg compression with the same quantization matrix. IEEE Trans Inf Forensics Secur 9(11):1933–1942. https://doi.org/10.1109/TIFS.2014.2359368
39. Yi Z, Fazhi HE, Qiu Y (2017) Dynamic strategy based parallel ant colony optimization on GPUs for TSPs. Sci China Inf Sci 60(6):68102
40. Yu H, He F, Pan Y (2018) A novel region-based active contour model via local patch similarity measure for image segmentation. Multimed Tools Appl 77(18):24097–24119. https://doi.org/10.1007/s11042-018-5697-y
41. Yu H, He F, Pan Y (2018) A novel segmentation model for medical images with intensity inhomogeneity based on adaptive perturbation. Multimed Tools Appl. https://doi.org/10.1007/s11042-018-6735-5
42. Yuan HD (2011) Blind forensics of median filtering in digital images. IEEE Trans Inf Forensics Secur 6(4):1335–1345
43. Zhang S, He F, Ren W, Yao J (2018) Joint learning of image detail and transmission map for single image dehazing. Vis Comput https://doi.org/10.1007/s00371-018-1612-9
44. Zhou L, Wang D, Guo Y, Zhang J (2007) Blur detection of digital forgery using mathematical morphology. In: Nguyen NT, Grzech A, Howlett RJ, Jain LC (eds) Agent and multi-agent systems: technologies and applications: first KES international symposium proceedings, KES-AMSTA 2007. Springer, Berlin, pp 990–998
45. Zhou Y, He F, Hou N, Qiu Y (2018) Parallel ant colony optimization on multi-core SIMD CPUs. Futur Gener Comput Syst 79:473–487. https://doi.org/10.1016/j.future.2017.09.073. http://www.sciencedirect.com/science/article/pii/S0167739X16304289



Ping Wang received his B.S. and M.S. from the Zhengzhou Science and Technology Institute in 2014 and 2017, respectively. He is currently working toward Ph.D. degree at Zhengzhou Science and Technology Institute. His research interest includes digital image forensics and steganalysis technique.

Fenlin Liu received his B.S. from Zhengzhou Science and Technology Institute in 1986, M.S. from Harbin Institute of Technology in 1992, and Ph.D. from the Northeast University in 1998. Now, he is a professor of Zhengzhou Science and Technology Institute. His research interests include digital image forensics and information hiding. He is the author or co-author of more than 90 refereed international journal and conference papers. He obtained the support of the National Natural Science Foundation of China and the Found of Innovation Scientists and Technicians Outstanding Talents of Henan Province of China.

Chunfang Yang received his B.S., M.S. and Ph. D. from the Zhengzhou Science and Technology Institute in 2005, 2008 and 2012, respectively. Now, he is with Zhengzhou Science and Technology Institute. His current research interests include image steganography and steganalysis technique.

Xiangyang Luo received his B.S., M.S. and Ph. D. from Zhengzhou Science and Technology Institute, in 2001, 2004 and 2010, respectively. He has been with Zhengzhou Science and Technology Institute since July 2004. From 2006 to 2007, he was a visiting scholar of the Department of Computer Science and Technology of Tsinghua University. From 2011, he is a postdoctoral of Institute of China Electronic System Equipment Engineering Co., Ltd. He is the author or co-author of more than 50 refereed international journal and conference papers. His research interest includes image steganography and steganalysis. He obtained the support of the National Natural Science Foundation of China and the Basic and Frontier Technology Research Program of Henan Province.