Anti-forensic Approach to Remove Stego Content from Images and Videos

P. P. Amritha[1,*], M. Sethumadhavan[1], R. Krishnan[1]
and Saibal Kumar Pal[2]

[1]*TIFAC-CORE in Cyber Security, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India*
[2]*Scientific Analysis Group, DRDO, Delhi, India*
E-mail: pp_amritha@cb.amrita.edu; m_sethu@cb.amrita.edu; drrkdrrk@gmail.com; skpal@hqr.drdo.in
[*]*Corresponding Author*

Received 07 April 2018; Accepted 21 May 2018;
Publication 02 April 2019

Abstract

Covert transmission of information hidden in different media to either a general or targeted audience constitutes steganography. However, this technique can be misused to transmit undesirable information. Traditionally the removal of such content necessitated the knowledge of the steganographic algorithm used. However, we address the scenario where such stego is removed using generic image processing operations along with an anti forensic method without assuming any knowledge of the steganographic algorithm used. The application of generic image processing operations also causes degradation of cover image, which can also be restored using this anti forensic method. Our procedure has been tested on a variety of steganographic algorithms including HUGO-BD, WOW, Synch and J-UNIWARD. By applying universal steganalysis we found that all images which have been subjected to our procedure have become stego free. However, a direct evaluation of the stego content assuming knowledge of the stego content and its location showed that 80 percentage of the stego is removed without significantly impacting the visual image quality. Video stream containing isolated static images have been addressed in this paper. The peak signal-to-noise ratio and structural

Journal of Cyber Security and Mobility, Vol. 8_3, 295–320.
doi: 10.13052/jcsm2245-1439.831

similarity metric values of cleaned images and videos are found to be in the range 30dB–40dB and 0.81–0.99 respectively.

Keywords: Steganography, Steganalysis, Image processing, Variational deconvolution, Markov features.

1 Introduction

Digital steganography is the process of securely embedding secret information in media like text, image, video and audio. It can also be used to hide the transmission of undesirable information through images via internet which is a social security problem. Detection and removal of such undesirable information from normal images is a challenging problem. Steganalysis is a type of countermeasure for steganography, which tries to detect the presence of secret messages in suspicious media and prevent further usage. Steganalysis can be blind or targeted [1, 2]. The targeted steganalysis identifies the steg content and then neutralizes it. This necessitates the knowledge of the algorithm used for stego generation and hence its use is limited to specific scenarios. The blind stego removal attempts to sanitize the content without assuming any knowledge of the algorithm. Today many terrorist organizations and hostile governments have been reported to use steganography in images in public repositories for recruiting and communicating malicious content [3]. These are not specifically addressed to a person but are more in the nature of casting a fishing net. Similarly drug dealers and financial crooks use such catch all tools.

We present an approach to remove the hidden data embedded in an image and video using suitable image processing operations along with an anti forensic method called variational deconvolution [4] for image quality enhancement. We rely on the basic idea that steganographic algorithms cannot protect the embedded information against technical modifications like noise that may occur during transmission. Our method is to apply image processing operation (can be called as showering techniques) [5] to remove (or sanitize) the stego content from the images and videos when it comes into or goes out of the system, followed by anti forensic method to destroy the stego content intensely and also to enhance the quality of showered images and videos. Then we apply qualitative approaches like RS steganalysis [6] and universal steganalysis [7] to check the performance of the system and quantitative measures like Bit Error Rate (BER) and Markov features to know how much percentage of stego is being removed. Finally, Peak Signal-to-Noise Ratio (PSNR) and Structured Similarity Index Measure (SSIM) are used for quality assessment of sanitized images and videos. Video which contains hidden static

frames can be removed by our techniques and can be evaluated using PSNR. When the proposed approach in this paper is deployed in a system, it effectively acts as Persistent Stego Incident Response System (PSIRS). [See Section 3].

While one or two attempts [8, 9] for removal of stego using image processing operations have been reported in literature, a comphrehensive set of methods for handling different types of images (textured and non-textured) and many different steganographic algorithms like HUGO-BD, WOW, Synch and J-UNIWARD have not been addressed. In addition our approach is to remove steg content without impacting the original content. The removal of steg content is also validated using universal steganalysis. The restoration of the original image and video after removal of steg content has also been attempted using deconvolution and we have been able to achieve 30–40 dB peak signal-to-noise ratio and structural similarity metric value between 0.81–0.99, which is usually taken to mean that perceptual visual quality is not impaired.

2 Related Work

In this section, we summarize some of the methods of active steganalysis and point out the difference with our technique for destroying hidden data.

Fabien A. P. Petitcolas, et al. [10] introduced attacks that enable the hidden information to be removed or otherwise rendered unusable. StirMark [11] attack introduces a practically unnoticeable quality loss in the image if it is applied only once. In [12] Fisk, G., et al. introduced the concept of Minimal Requisite Fidelity as a measure of the degree of signal fidelity that is both acceptable to users and destructive to covert communications utilizing TCP/IP suite. The concept of steganographic sanitization was introduced in [13]. The method successfully sanitizes images from 26 different steganographic methods using different levels of scrambling. After performing different levels of scrambling most of the images were sanitized with only level one distortion. Our work enhances by restoring the quality of the image and quantitative approach was introduced to estimate the percentage of stego content removed after sanitization and is applicable to all types of image formats.

Sieffert, M., et al. [14] introduced a framework that sniffs all HTTP traffic, reconstruct the image that are transmitted through the packets and test each image against all known steganalysis algorithms. This system does not destroy stego content, but detect hidden information and block the communication if stego is present. Fawzi Al-Naima, et al. [15] proposed a steganographic firewall which destroys the embedded information in images, and not to inspect their existence and blocks the suspected medium (cover) as the normal firewall does. Firewall acts as a filter, that lets the clean files pass through it,

but it destroys the information partially or completely, that might possibly be embedded inside it. Compared to this, our techniques try to restore the input image to the original cover as close as possible. Francia implemented the concept of steganography obliterator [16], proactively cleansing the cover media from possible steganographic data embeded using LSB technique. Smith, C. B. and Agaian, S. S [17] examined the removal of hidden content initially by using the concept of bit deletion in spatial domain and JPEG re-encoding in the transform domain. Both methods maintained the quality of image. Then image denoising methods were also introduced to remove the hidden data from images. Smith, C. B and Agaian, S. S [8, 18] considered stego removal as a denoising process. They estimated the percentage of stego removed using BER. The recovered message was less than half the length of the bits embedded. But denoising degraded the image quality. An architecture called stego scrubbing for removing the stego was discussed [19] by P. A. Lafferty, et al. The authors proposed a system which will be acting like a guard or firewall. Researching solely in the domain of images, Lafferty conducted experiments [9] using different collection of images and steganographic algorithms in both domains. Our work enhances this work by applying image enhancement method on showered image and used steganalysis algorithm to detect the stego removal. Chandramouli [20] is the first to significantly mention about mathematical framework for active steganalysis. Active steganalysis, extraction of a hidden message with little or no prior information, is formulated as a blind system identification problem within this framework. BER and PSNR are used as performance measures. Spread Spectrum steganography and Watermarking are easy to break by using this framework.

Nutzinger, M. [21] focused on preventing steganographic usage in digital audio data. Their system uses natural modification like noise addition or shifiting of sampling values and simulates packet loss and malicious modification like variable time delay and frequency shifting for steganography prevention. These techniques were able to maintain an acceptable audio quality. Another method which will effectively remove the steganographic information in spatial and frequency domain for image, video and audio was introduced by Sharp, A., et al. [22–24]. Authors proposed an attack using discrete spring transform which will only distort the numerical value of the carrier media while keeping visual quality in a high level. Blasco, Jorge, et al. [25] suggest a framework to forbid the steganography usage through HTTP. Different sanitizers that eliminate hidden content from any kind of information transmitted through HTTP were proposed.

An overwriting approach was introduced by Siddeeq Y. Ameen and Muthana Al-Badrany [26], where random data written again and again over

stego images to remove the stego content. Filters based denoising approach and wavelet thresholding were also incorporated to remove the stego content. PSNR calculated after destruction showed the cover image quality has been enhanced with denoising techniques. No details were provided about exactly what percentage of stego was removed. Our work differs from this by using additional filtering operations and present estimation of stego content removed by quantitative approach.

3 Persistent Stego Incident Response System: Design and Implementation

Stego incidences which are sanitized (cleaned) by our procedure will not be reported but removed. The removal of stego content is verified by universal steganalysis algorithm. In case the universal steganalysis algorithm identifies the stego content which has not been removed by our algorithm then that case will be reported. Hence, we call our system as Persistent Stego Incident Response System. Figure 1 shows the outline of proposed system. In this work, we extend our initial work [5] by applying the proposed techniques on latest stego algorithms in spatial and transform domains. We also incorporated universal steganalysis and calculated second order Markov features to validate efficiency of our system and claim that it is impossible to retrieve the stego content. This system can be opted as a universal approach for removing the stego content by suppressing the possible carrier of steganographic information and making the carrier available for further use. This system has four components. First one is the classification part, which will classify the images into textured and non-textured. Second is the showering techniques which will suppress stego content by using image filtering operations. Third component is for restoring the degraded image and fourth for obtaining quantitative and qualitative measures for assessing the performance of the system. If input is video, all the frames will be extracted and then individual frames are given to the showering module for further processing.

3.1 Objective of this Work

This paper tries to evaluate showering techniques on newer algorithms like HUGO-BD (Bounding Distortion) [27], WOW [28], Synch [29], J-UNIWARD [30] and TPVD [37]. Our earlier work [5] addresses this problem, when steganographic algorithms like PVD [31], LSB matching [32], OUTGUESS [33] and Wavelet based [17] were used in images. Removal of static hidden frames from videos is also tried here.

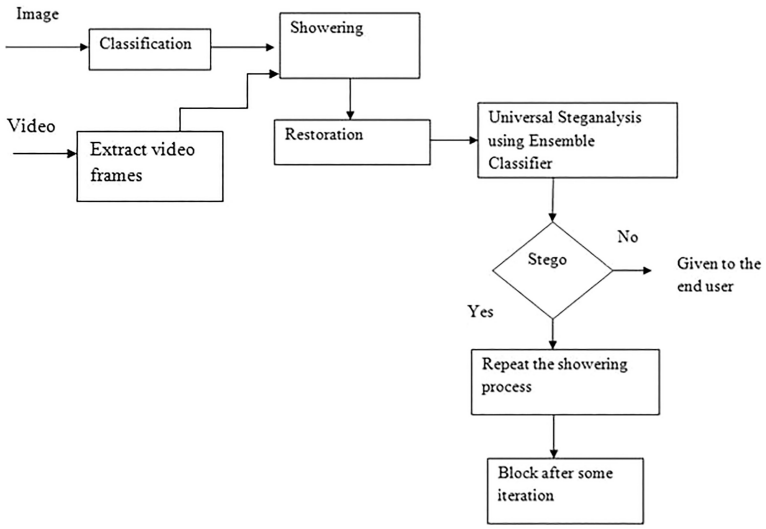

Figure 1 Proposed showering mechanism.

3.2 Procedure

1. Classification: This is to avoid delay in prioritizing the level of showering, since performance of showering operation depends on the type of images. We used Gray-Level Co-occurrence Matrix (GLCM), as proposed by Haralick [34] to automatically classify the input into textured and non-textured images. Although 28 features can be derived from GLCM, usually only following five important features are considered. Correlation, Homogeneity, Dissimilarity, Energy and Entropy. We have considered two of them namely; Correlation and Homogeneity which allowed us to classify textured and non-textured images.
2. Showering: With multiple types of steganographic methods available, each modifying images in different way, and embedding the data in different places, it should be expected that different embedding methods will require different showering methods. To evaluate the effectiveness of the showering techniques, images with known stego content were used and the impact of these techniques on the removal of the said stego content was estimated. Showering can be achieved either using radio-metric operations or geometric operations or a combination of the two. The radiometric operations include Median, Mean, Gaussian, Wiener filters of different window sizes, Wiener restoration, and Wavelet thresholding [26]. The geometric operations include rotation through various angles.

- Median filter: We took different window size of 3×3, 5×5 and 7×7 and calculated median on sliding window basis.
- Gaussian filter: We took different window size of 3×3, 5×5 and 7×7 and each with standard deviation of 0.5, 2, 3.5, 6.5, 8, 9.8.
- Mean filter: This type of operation takes the average pixel value of all pixels in an 3×3, 5×5 and 7×7 pixel neighborhood, compute the average value, and replace the pixel value of original image to that computed average value on a sliding window basis.
- Wiener filter: We have taken filter with different window size of 3×3, 5×5 and 7×7. Removing noise from each pixel is based on statistics estimated from a local neighborhood.
- Wiener restoration: By using this filter we can reverse the affect of denoising filter. Image quality lost during filtering can be re-gained without restoring the secret. A small amount of stego is again removed due to this process.
- Wavelet thresholding: We have used hard threshold function for thresholding. In this the wavelet coefficients below a given value are settled to zero.
- Rotation: We rotate the image *n* degrees, and again by *-n* degrees. Rotated the orginal image through various angles of 1, 2, 3, 4, 5, 10, 15, and 20 degrees.

Apart from the filters mentioned above, the five combinational filters used here are summarized in Table 1. So a total of 27 different filters were applied on stego images to remove the stego content. Showering can be done in frequency domain as well.

3. Restoration: Some of the filters used for showering do not guarantee the quality of image after filtering. So we integrate this component to enhance the quality of the showered image using variational deconvolution method [4]. This method approximates the showered image to original cover by applying deconvolution. Even though Wei Fan deconvolution techniques seem to work for all types of filtering, use of particular kernel for different filtering is likely to work better. Suitable kernel has to be

Table 1 The five different combination filters

Combination filter 1	Gaussian, Median, rotate degree 5
Combination filter 2	Gaussian, Wiener, rotate degree 5, Median
Combination filter 3	Gaussian, Median, Wiener, rotate degree 5, Wiener restore
Combination filter 4	rotate degree 5, Gaussian, Median
Combination filter 5	Median, rotate degree 5, Gaussian

selected for different filtering. By employing variational deconvolution method, we could maintain the PSNR value above 30dB. Quality of the restored image is verified by taking the PSNR and SSIM values.

4. Steganalysis: This component assesses the performance of the system by providing qualitative measures using RS Steganalysis, Chi-square attack [35], difference image histogram [36] and universal steganalysis [7]. BER and Markov features were used to estimate the percentage of stego removed in both domains. If stego content is detected even after going through several levels of showering process, the system block the usage of such images. Novelty of this paper stands on successful application of our approach on newer algorithms like HUGO-BD, WOW, Synch and J-UNIWARD which are considered to be superior to PVD, LSB matching, OUTGUESS and Wavelet based steganography. These algorithms increases the security of additive steganographic schemes for digital images represented in the spatial and transform domain. Ensemble classifier is used to verify the efficiency of showering. This indicates that our showering method is effective.

3.3 Evaluation of the System

In this section we assess the efficacy and efficiency of the overall system. In terms of efficacy, we evaluate if the PSIRS is able to remove stego content from images sent through emails and from system resources so that hidden data cannot be recovered. For establishing this we use a rich model based universal steganalysis [7]. For testing efficiency, we have measured the stego removed in terms of BER and Markov features.

- Bit Error Rate: This metric is designed to capture bit by bit error in the extracted secret message with the original secret image.
- First order Markov feature: This metric is to capture correlation between adjacent two pixels at a time. Let the secret image be, $I_1 = [x_i]$, $1 \leq i \leq mn$ and the secret image extracted from restored image is $I_2 = [y_i]$ of size *mn*

$$C = \sum_i [\partial(x_i, y_i) \&\& \ \partial(x_{i+1}, y_{i+\,1})] \ , \quad where \quad 1 \leq i \leq mn - 1$$

- Second order Markov feature this metric is designed to capture correlation between adjacent three pixels at a time by comparing the extracted secret message and the original secret image.

$$C = \sum_i [\partial(x_i, y_i) \&\& \ \partial(x_{i+1}, y_{i+\,1}) \&\& \ \partial(x_{i+2}, y_{i+2})],$$
$$where \ 1 \leq i \leq mn - 2$$

where && is the logical operator AND and *C* gives the count of equal number of adjacent bits in the image.
We define delta function as

$$\partial(x,y) = \left\{ \begin{array}{l} 1\ ,\ if\ x = y \\ 0\ ,\ \ otherwise \end{array} \right\}$$

Total number of secret bits minus *C* value will give the count of removed bits. Second order Markov feature is found to be more accurate than BER and first order Markov feature. Table 2 to 5 summarizes the average stego removed using three metrics.

Table 2 Stego removed (on average) from images stegoed with HUGO

Showering Process	Type of the Image	BER	First Order Markov Feature (Percentage)	Second Order Markov Feature (Percentage)	PSNR (dB)
Median	Textured	0.48	74	86	30.2
(3 × 3) filter	Non-Textured	0.47	71	84	35.1
Gaussian	Textured	0.49	74	86	33.1
(3 × 3) filter	Non-Textured	0.48	70	82	40.2
Rotation	Textured	0.49	73	86	20.2
	Non-Textured	0.48	71	84	22.3
Wiener filter	Textured	0.49	74	87	28.0
	Non-Textured	0.49	75	90	35.2
Combination	Textured	0.49	73	85	20.4
filters	Non-Textured	0.48	70	82	25.4

Table 3 Stego removed (on average) from images stegoed with WOW

Showering Process	Type of the Image	BER	First Order Markov Feature (Percentage)	Second Order Markov Feature (Percentage)	PSNR (dB)
Median	Textured	0.49	77	90	30.5
(3 × 3) filter	Non-Textured	0.51	76	89	29.9
Gaussian	Textured	0.52	77	89	36.1
(3 × 3) filter	Non-Textured	0.49	74	87	37.0
Rotation	Textured	0.48	75	88	20.4
	Non-Textured	0.47	73	86	23.7
Wiener filter	Textured	0.49	73	87	30.3
	Non-Textured	0.49	77	89	35.5
Combination	Textured	0.51	75	87	22.1
filters	Non-Textured	0.50	74	88	25.0

Table 4 Stego removed (on average) from images stegoed with Sync

Showering Process	Type of the Image	BER	First Order Markov Feature (Percentage)	Second Order Markov Feature (Percentage)	PSNR (dB)
Median	Textured	0.50	75	87	32.8
(3 × 3) filter	Non-Textured	0.51	76	89	29.9
Gaussian	Textured	0.48	73	85	39.4
(3 × 3) filter	Non-Textured	0.49	74	87	37.0
Rotation	Textured	0.48	72	85	25.8
	Non-Textured	0.47	73	86	23.7
Wiener filter	Textured	0.51	75	87	32.5
	Non-Textured	0.49	77	89	35.5
Combination filters	Textured	0.50	75	87	25.0
	Non-Textured	0.50	73	85	25.0

Table 5 Stego removed (on average) from images stegoed with J-UNIWARD

Showering Process	Type of the Image	BER	First Order Markov Feature (Percentage)	Second Order Markov Feature (Percentage)	PSNR (dB)
Median	Textured	0.50	75	90	28.1
(3 × 3) filter	Non-Textured	0.47	73	90	32.3
Gaussian	Textured	0.48	74	86	35.2
(3 × 3) filter	Non-Textured	0.34	72	80	38.1
Rotation	Textured	0.49	74	90	20.1
	Non-Textured	0.49	72	81	21.3
Wiener filter	Textured	0.49	73	80	25.2
	Non-Textured	0.49	75	90	35.1
Combination filters	Textured	0.51	73	81	20.4
	Non-Textured	0.50	74	90	25.5

4 Experimental Results

4.1 Stego Removal from Images

To conduct experiments, we used 225 textured and non-textured images of size 256 x 256 taken from the BOSSbase database ver.1.01 [41]. Initially we trained the Support Vector Machine for classifying textured and non-textured images and we considered only two features, Homogeneity and Correlation to decide a new image is textured or not. We implemented the system for stego removal against four latest steganographic algorithms, HUGO-BD (Bounding Distortion), WOW and Synch in spatial domain and J-UNIWARD

in transform domain. All 225 images were subjected to stego embedding at various payloads (5%, 30%, 70%, and 100% of image size) using these four algorithms. The final showered (cleaned) images were again subjected to universal steganalysis using SRMQ1 (Spatial domain Rich Model with the fixed Quantization) features [38] for spatial domain and CC300 features [39] for transform domain to check the performance.

Percentage of stego removed was calculated using BER, first and second order Markov features by comparing the original and extracted secret from the cleaned image. Performance of showering techniques in terms of stego removed is explained below by considering only the first order Markov feature. Evaluation by other two metrics is summarised in Table 2,3,4 and 5. Results in Table 2 show that Median filter and Gaussian filter remove 86 percentage of the stego while maintaining the PSNR above 30 dB for images stegoed with HUGO. Likewise, Tables 3 to 5 show the corresponding stego removal for WOW, Sync and J-UINWARD.

Figures 2, 3 and 4 show the percentage of stego removed calculated using first order Markov feature for textured and non-textured images. Figures 5, 6 and 7 show the corresponding PSNR and SSIM values. It is evident from the Figure 2 that 70 percentage of the secret is removed after applying Median filter, average filter, Gaussian filter of size 3, rotation of degree 1, 5, 10 and 20, Wiener filter and by all combinations filters from textured images. But, while considering the quality of the image, Median filter and Gaussian filter of size 3 is finally chosen to remove stego content from textured images.

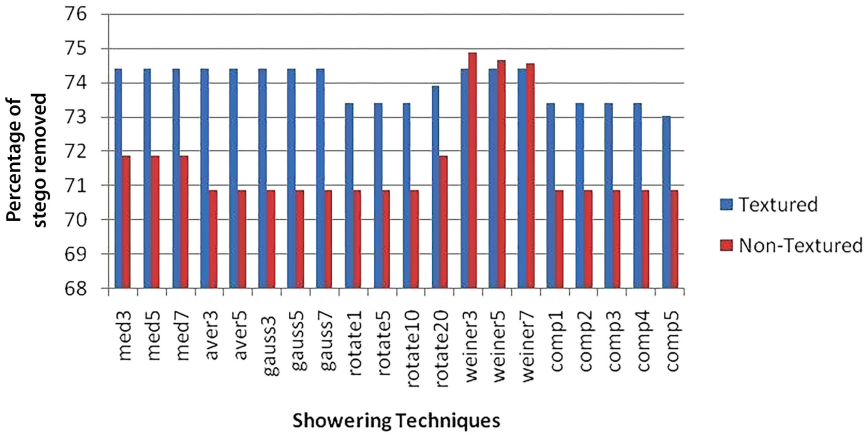

Figure 2 Percentage of stego removed from images stegoed with HUGO-BD.

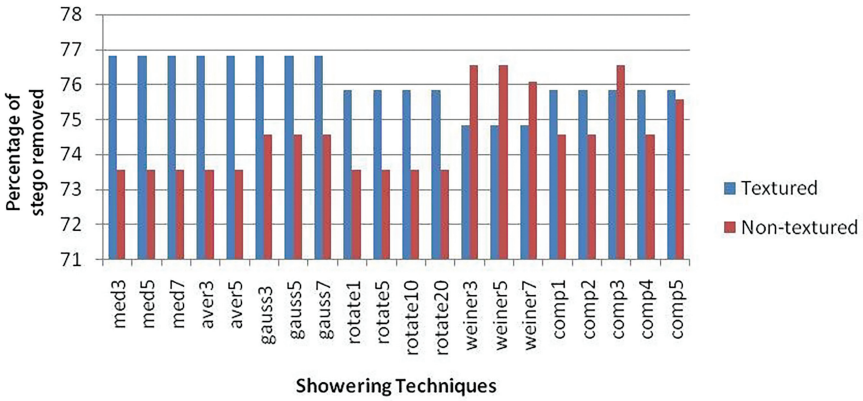

Figure 3 Percentage of stego removed from images stegoed with WOW.

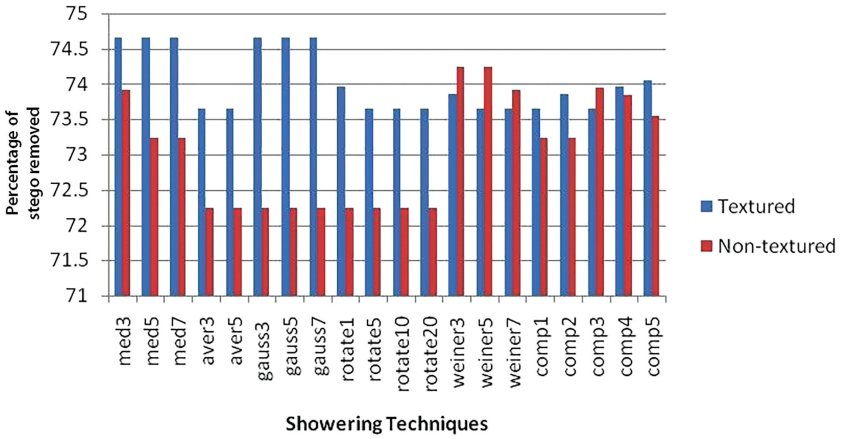

Figure 4 Percentage of stego removed from images stegoed with J-UNIWARD.

Likewise, for non-textured images, Wiener filter and Median filter of size 3 is preferred to remove stego while maintaining the PSNR above 30 dB as seen in Figure 5 (a). Rotation will definitely destroy very well but restoration is not good quality. When showering techniques are applied against WOW, results showed that 76 percentage of the secret is removed after applying Median filter and Gaussian filter of size 3 from textured images as seen in Figure 3.

Likewise, for non-textured images, Wiener filter, combination filter 3 are preferred to remove stego while maintaining the quality of the image as seen in Figure 6(a). Experiments with stego images created using Sync algorithm showed similar results as WOW for both textured and non-textured images.

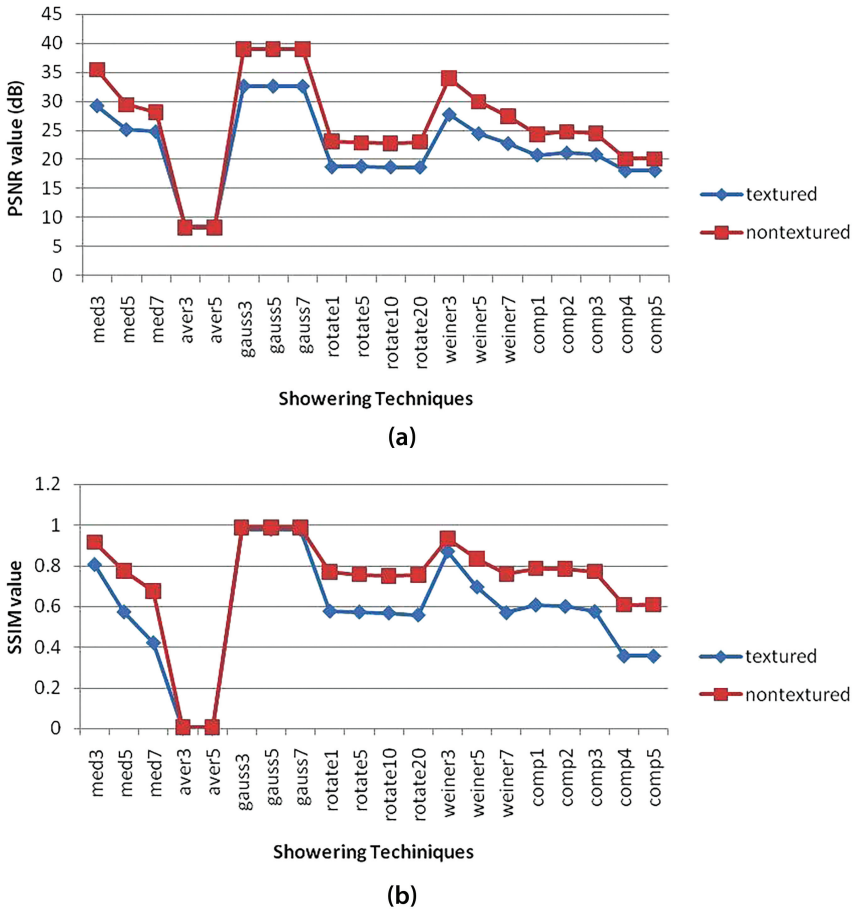

Figure 5 (a) PSNR values of restored images (HUGO-BD) (b) SSIM values of restored images (HUGO-BD).

For destroying stego content in transform domains, Gaussian and Median filters are preferred for textured images and while for non-textured images Wiener filter, median filter and Combination filters 1 and 2 are preferred (Figure 4). These filters removed 70 percentage of the stego content while maintaining the quality which can be seen in Figure 7(a). It is observed that average filter performs badly in terms of preserving quality for all the four algorithms. All combination filters used for experiment are providing good results in terms of stego removal but results showed that using combination filters 1, 2, 3 (see Table 1) stego is removed above 75% and at the same time maintains quality above 25 dB.

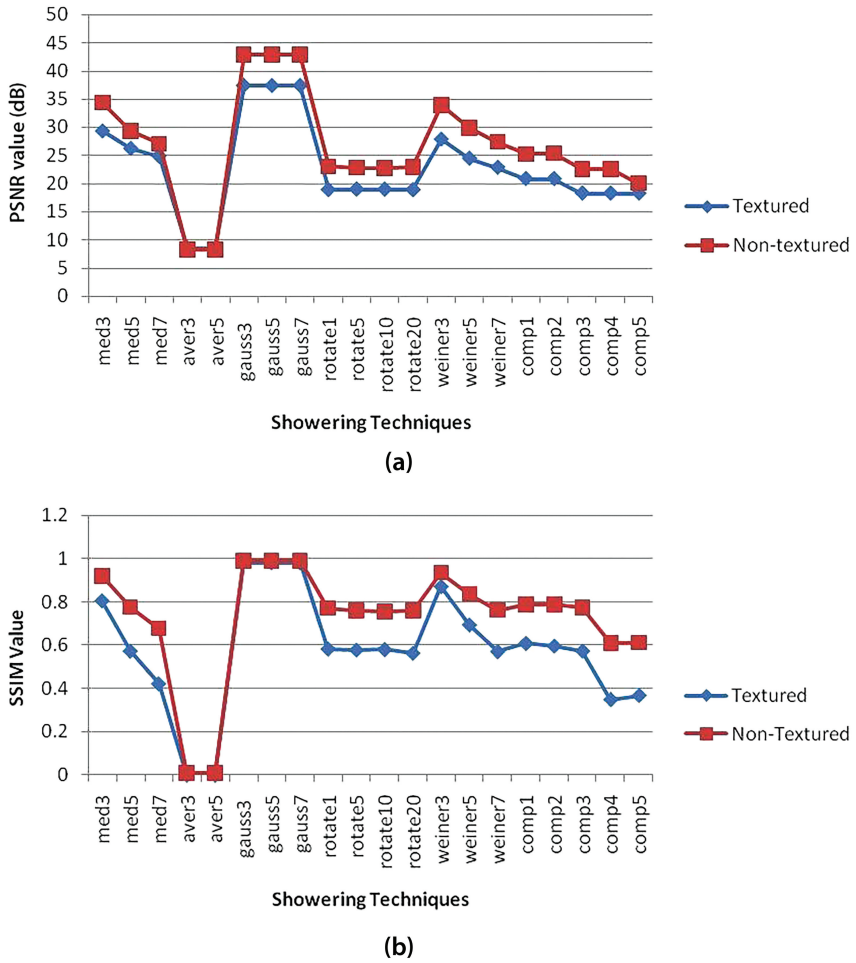

Figure 6 (a) PSNR values of restored images (WOW) (b) SSIM values of restored images (WOW).

Table 6 Universal Steganalysis against HUGO-BD, WOW and Sync

Input to the Classifier	Accuracy Percentage	False Positive Rate (FPR)
Showered image	100	0

We used ensemble classifier for detecting presence of remanent stego content on showered (cleaned) image. Table 6 shows the detection accuracy of the classifier and false positive rate (i.e rate in which cover wrongly classifying as stego) when we trained the classifier with SRMQ1 features of the cover and

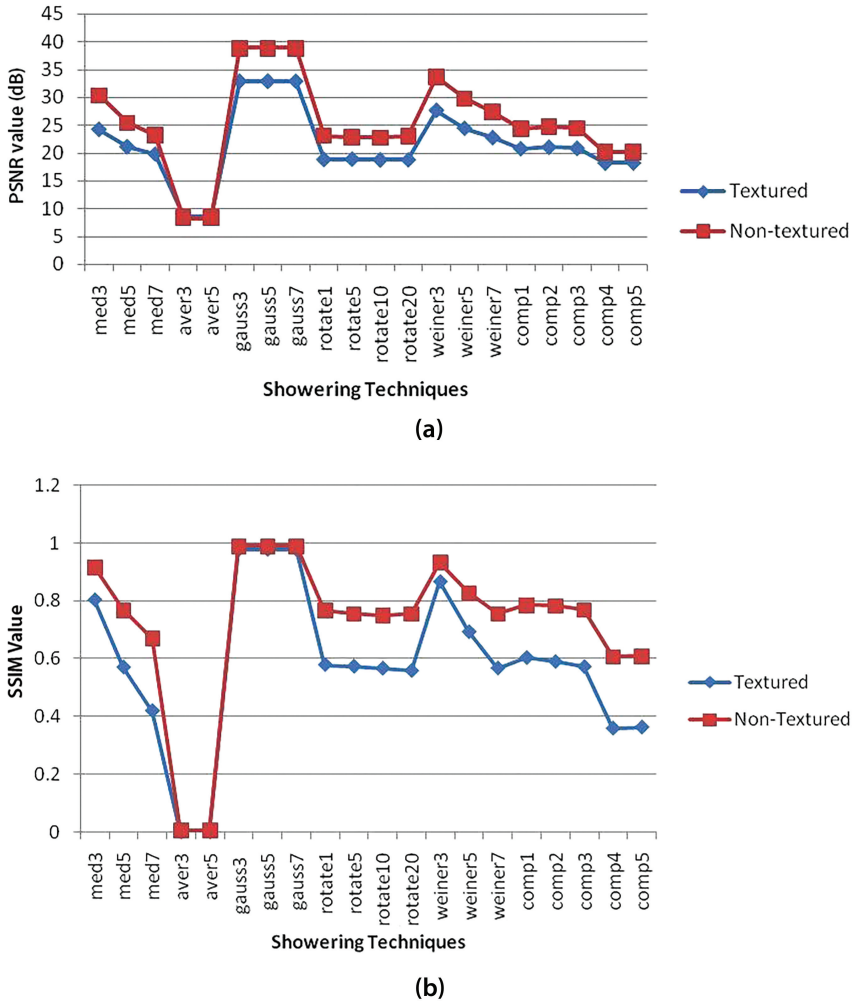

Figure 7 (a) PSNR values of restored images (J-UNIWARD) (b) SSIM values of restored images (J-UNIWARD).

stego images embedded using HUGO, WOW and Sync. The classifier gave the accuracy of detection as 100 percentage which means that all showered images were correctly classified as cover since stego content was removed from the image, which justifies that our showering algorithm is efficient. Table 7 shows the detection accuracy of the classifier and false alarm rate when trained the classifier with CC-C300 features of the cover and stego images embedded using J-UNIWARD. Classifier gave 95 percentage accuracy and at most 10

Table 7 Universal Steganalysis against J-UNIWARD

Input to the Classifier	Accuracy Percentage	False Positive Rate (FPR)
Showered image using Median filter	0.937	0.071
Showered image using Gaussian filter	0.924	0.080
Showered image using Wiener filter	0.955	0.053
Showered image using 5 degree rotation filter	0.915	0.089
Showered image using Five combination filter	0.962	0.036

out of 112 showered images were only wrongly classified as stego which shows that our showering algorithm is efficient. Accuracy and False Positive Rate are calculated as follows;

Accuracy = (TP+TN)* 100/ (P+N) FPR = FP/(TN+FP). where P = TP+FN, N = FP+TN, TP denote true positive, TN true negative, FP false positive, FN false negative and FPR false positive rate.

Since the showered images are almost like cover images we have tried to evaluate them using steganalysis based on all the different steganographic algorithms. i.e even if the stego involved in test image was WOW it was tested using universal steganalysis which was trained on HUGO-BD Synch and J-UNIWARD and all of them indicated that the showered image is devoid of any stego content. This implies that our showering procedure is able to remove stego content without the knowledge of the stego algorithm. We summarize the results as follows:

1. Median is not a smoothing filter hence it can work equally well in textured and non-textured images
2. Homogeneity and Correlation will determine the texture property as given in Table 8. When homogeneity of a textured image increases from 0.60 to 0.85, showering effect by Wiener filter increases, while that of Gaussian filter decreases. Similarly when correlation of a textured image increases from 0.45 to 0.70 showering effect by Wiener filter increases while that of Gaussian filter decreases. These are shown in Figures 8 and 9. These observations are found to be true in the case of non-textured images (Figures 10 and 11) also. This validates our claim that Gaussian filter is working better in textured images and Wiener filter for non-textured images.

Table 8 Ranges of values for GLCM features

Features	Textured	Non-Textured
Homogeneity	0.60–0.85	0.85–0.95
Correlation	0.45–0.70	0.70–0.95

Table 9 Stego removal (on average) from videos stegoed in I-P-B frames

Showering Process	PSNR (dB)	SSIM	Second Order Markov Feature (Percentage)
Median filter (3 × 3)	35.1	0.90	89
Gaussian filter (3 × 3)	33.2	0.85	90
Wiener filter	30.5	0.82	93
Rotation	23.0	0.65	95
Combination filters	26.4	0.79	85

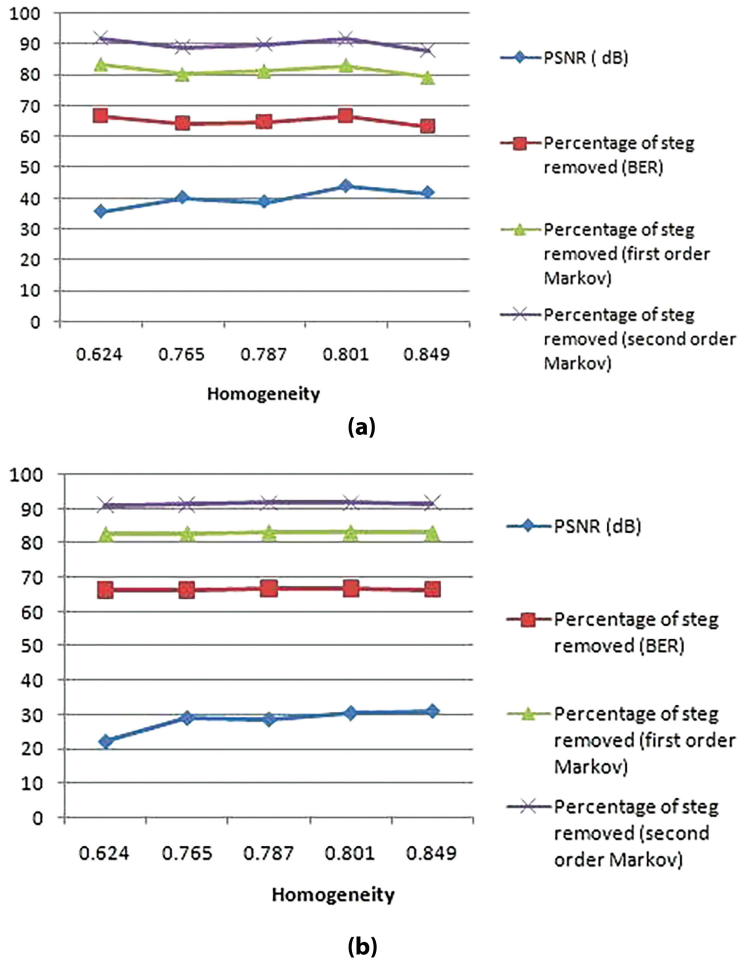

Figure 8 Effect of (a) Gaussian and (b) Wiener filter in removing stego from textured images by considering homogeneity property.

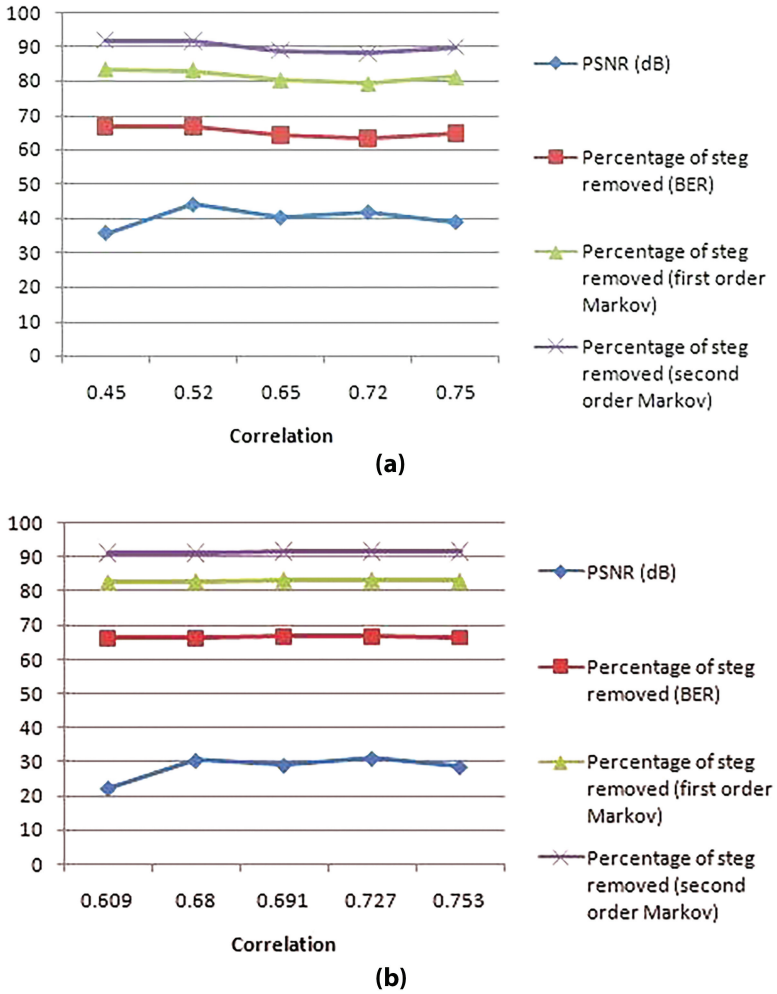

Figure 9 Effect of (a) Gaussian and (b) Wiener filter in removing stego from textured images by considering correlation property.

We analysed the effect of Gaussian filter (with varying standard deviations) on removing stego content from images and its effect on image quality. Results showed that when the standard deviation increases the removal of stego content is more but quality of image is not preserved. Gaussian filter with standard deviation 0.5 removes 90 percentage of the stego content and preserves quality above 30 dB as shown in Figure 12.

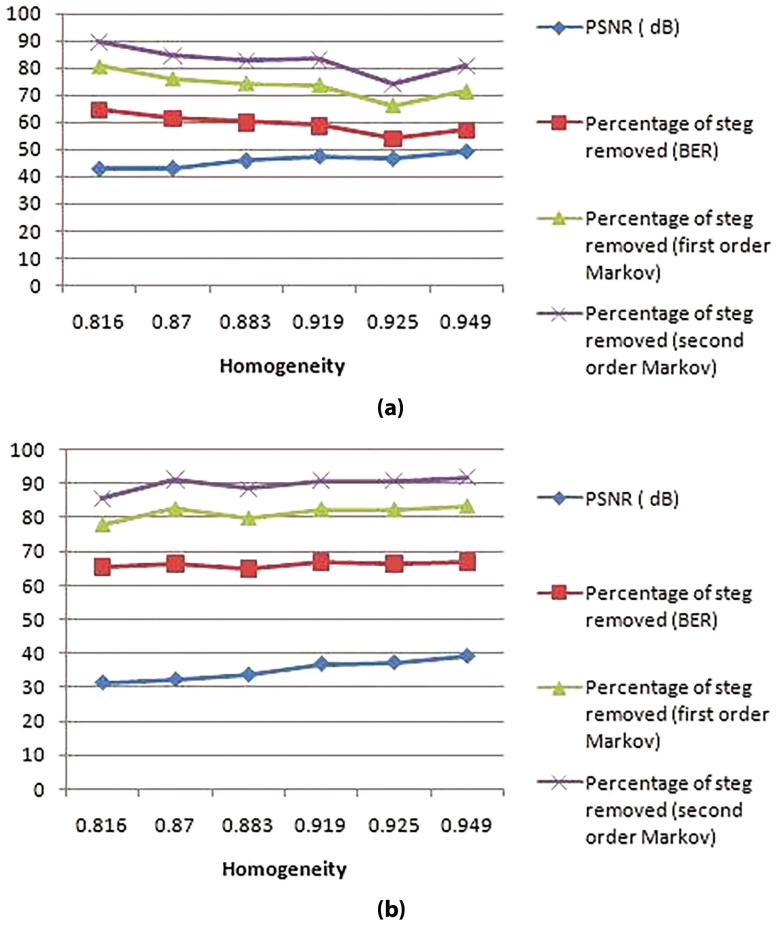

Figure 10 Effect of (a) Gaussian and (b) Wiener filter in removing stego from non-textured images by considering homogeneity property.

4.2 Stego Removal from Videos

We used cover video of size 288 × 352 with 15 frames/sec and number of frames is 150. We created stego videos by embedding secret in the I-P-B frames using TPVD [37] algorithm and also using other algorithms mentioned in Section 3.1. Stego embedding were done on I frames similar to image embedding. For B and P frames, macro blocks were selected depending on the motion vector which gave large magnitude. Using MPEG parser we extracted I-frames, P-frames and B-frames. After extracting, all frames were subjected

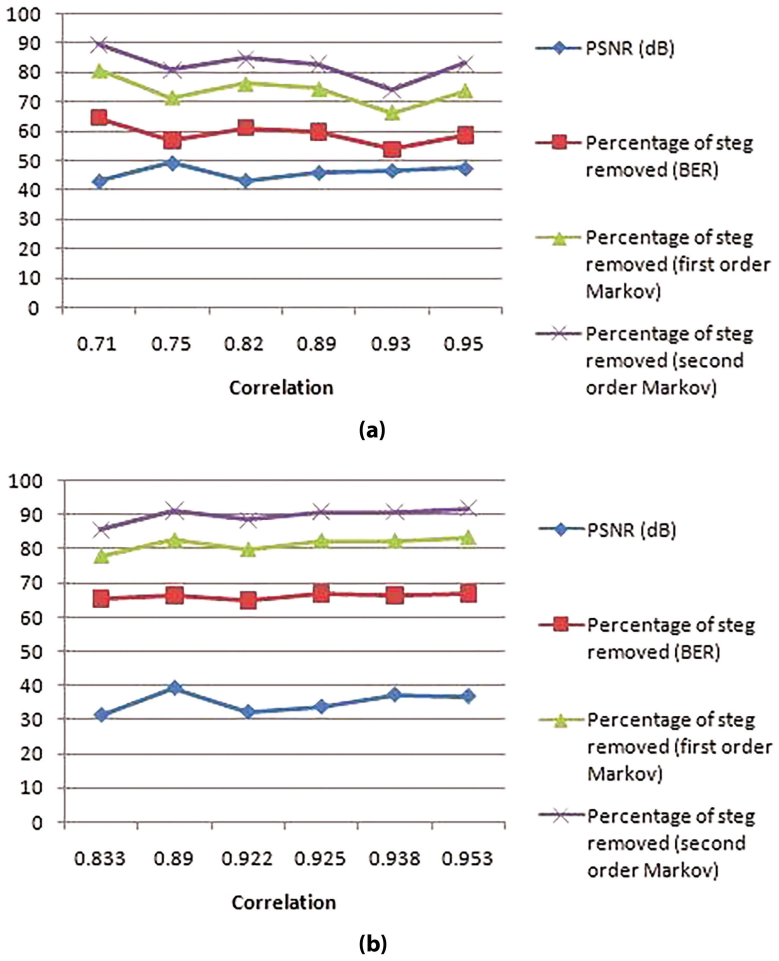

Figure 11 Effect of (a) Gaussian and (b) Wiener filter in removing stego from non-textured images by considering correlation property.

to our showering techniques along with anti forensic method. Results showed that the cleaned video preserved the perceptual video quality. Obtained PSNR, SSIM and percentage of stego removed from the video is summarised in Table 9. We concluded that Median and Gaussian filter are better than other filters in removing stego content and at the same time maintaining the video quality.

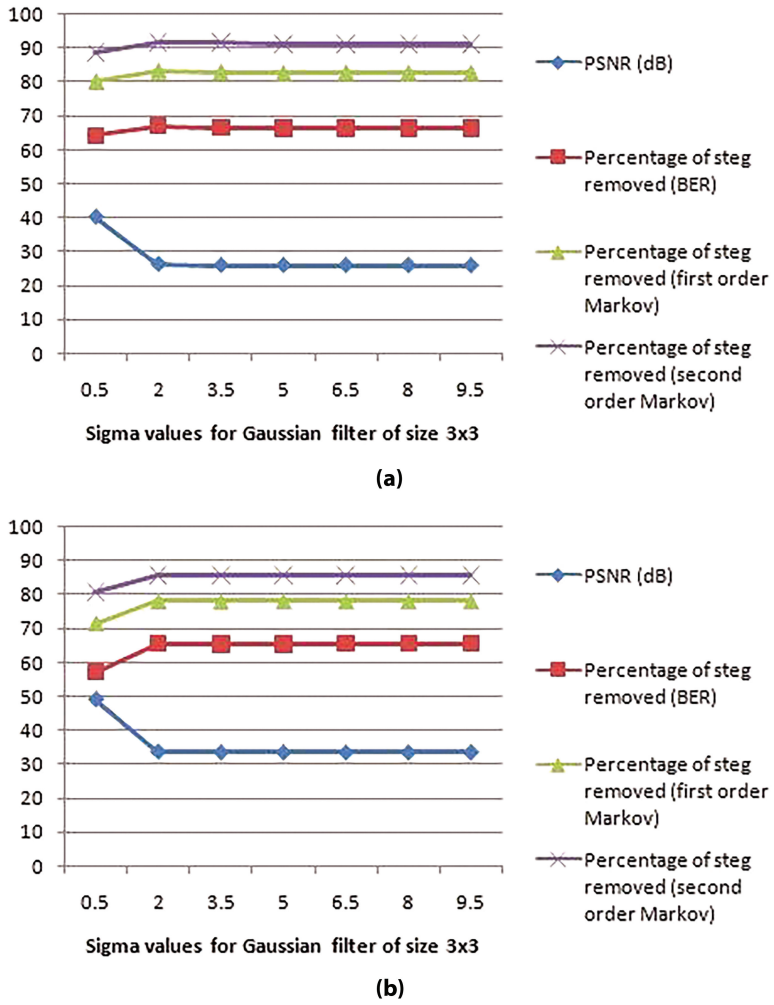

Figure 12 Effect of Gaussian filter of size 3x3 with varying standard deviation on (a) textured images (b) non-textured images.

5 Conclusion

In our earlier work we have established that our showering algorithm are effective against stego content embedded using PVD, LSB matching, OUTGUESS and Wavelet based steganography. We had shown that performance of showering depends on textured and non-textured images. In this paper we tried the showering algorithms on stego content created using

HUGO-BD, WOW, Synch and J-UNIWARD. These newer algorithms are supposed to be more difficult for steganalysis; however we were able to show that even for these algorithms, our showering methods are equally effective. To validate the statement we have a direct evaluation of stego removal as well as an evaluation using universal steganalysis. We can conclude that for textured images with different payloads embedded with spatial domain steganography, on average Gaussian and Median filter removes stego content above 80% while preserving quality. On the other hand, Wiener and combination filters removes 85% of the stego content from non-textured images. In transform domain, Median and Gaussian filters were able to destroy stego content from textured images and from non-textured images by using Median and Wiener filter while preserves the PSNR value above 30 dB. Even though other filters could completely remove stego content, they failed in preserving the PSNR to be above 30 dB. From these observations, we can conclude that Median, Gaussian and Wiener filter removes stego content to large extent without compromising quality. Second order Markov feature gave better results than BER and first order Markov feature. Showering effect on videos also showed similar results when video contains secret image or text. In future, recent works on side informed steganographic algorithms [40] and reversible data hiding methods like [42] can also be tested.

References

[1] R. Chandramouli, M. Kharrazi, N. Memon, Image steganography and steganalysis: Concepts and practice, In International Workshop on Digital Watermarking, Springer, 35–49, 2003.

[2] X.-Y. Luo, D.-S. Wang, P. Wang, F.-L. Liu, A review on blind detection for image steganography, Signal Processing, 88(9), 2138–2157, 2008.

[3] E. Kartaltepe, J. Morales, S. Xu, R. Sandhu, Social network based botnet command and control: emerging threats and countermeasures, in: Applied Cryptography and Network Security, Springer, 11–528, 2010.

[4] W. Fan, K. Wang, F. Cayre, Z. Xiong, Median filtered image quality enhancement and anti-forensics via variational deconvolution, IEEE transactions on information forensics and security 10(5), 1076–1091, 2015.

[5] P. Amritha, M. Sethumadhavan, R. Krishnan, On the removal of steganographic content from images, Defence Science Journal, 66(6), 574, 2016.

[6] J. Fridrich, M. Goljan, Reliable detection of lsb steganography in color and grayscale images, us Patent 6, 831–991, 2004.
[7] J. Fridrich, J. Kodovsky, Rich models for steganalysis of digital images, IEEE Transactions on Information Forensics and Security, 7(3), 868–882, 2012.
[8] C. B. Smith, S. S. Agaian, On steganalysis and clean image estimation, Multimedia Forensics and Security, 212–244, 2008.
[9] P. A. Lafferty, Obfuscation and the steganographic active warden model, The Catholic University of America, 2008.
[10] F. A. Petitcolas, R. J. Anderson, M. G. Kuhn, Attacks on copyright marking systems, in:International workshop on information hiding, Springer, 218–238, 1998.
[11] J. Dittmann, M. Steinebach, A. Lang, S. Zmudzinski, Advanced audio watermarking benchmarking, in: Proceedings of SPIE, Vol. 5306, 224–235, 2004.
[12] G. Fisk, M. Fisk, C. Papadopoulos, J. Neil, Eliminating steganography in internet traffic with active wardens, in: International Workshop on Information Hiding, Springer, 18–35, 2002.
[13] A. Whitehead, Towards eliminating steganographic communication., in: PST, 2005.
[14] M. Sieffert, R. Forbes, C. Green, L. Popyack, T. Blake, Stego intrusion detection system, in: Proceedings of the fourth Digital forensic Research Workshop, 2004.
[15] F. Al-Naima, S. Y. Ameen, A. F. Al-Saad, Destroying steganography content in image files, in: The 5th International Symposium on Communication Systems, Networks and DSP, 2006.
[16] G. A. Francia III, T. S. Gomez, Steganography obliterator: an attack on the least significant bits, in: Proceedings of the 3rd annual conference on Information security curriculum development, ACM, 8591, 2006.
[17] C. B. Smith, S. S. Agaian, Denoising and the active warden, in: Systems, Man and Cybernetics, ISIC, IEEE International Conference 3317–3322, 2007.
[18] C. B. Smith, S. S. Agaian, On noise, steganography, and the active warden, Multimedia Forensics and Security, 139–162, 2008.
[19] I. S. Moskowitz, P. A. Lafferty, F. Ahmed, Stego scrubbing a new direction for image steganography, in: Information Assurance and Security Workshop, IEEE, 119–126, 2007.
[20] R. Chandramouli, A mathematical framework for active steganalysis, Multimedia systems, 9(3), 303–311, 2003.

[21] M. Nutzinger, Real-time attacks on audio steganography, Journal of Information Hiding and Multimedia Signal Processing, 3(1), 47–65, 2012.
[22] Q. Qi, A. Sharp, Y. Yang, D. Peng, H. Sharif, Steganography attack based on discrete spring transform and image geometrization, in:Wireless Communications and Mobile Computing Conference, IEEE, 554–558, 2014.
[23] A. Sharp, Q. Qi, Y. Yang, D. Peng, H. Sharif, A video steganography attack using multi dimensional discrete spring transform, in: Signal and Image Processing Applications, pp. 182–186, 2013.
[24] Q. Qi, A study on countermeasures against steganography: an active warden approach.
[25] J. Blasco, J. C. Hernandez-Castro, J. M. de Fuentes, B. Ramos, A framework for avoiding steganography usage over http, Journal of Network and Computer Applications, 35(1), 491–501, 2012.
[26] S. Y. Ameen, M. R. Al-Badrany, Optimal image steganography content destruction techniques, in: International Conference on Systems, Control, Signal Processing and Informatics, 453–457, 2013.
[27] T. Filler, J. Fridrich, Gibbs construction in steganography, IEEE Transactions on Information Forensics and Security, 5(4), 705–720, 2010.
[28] V. Holub, J. Fridrich, Designing steganographic distortion using directional filters, in: Information Forensics and Security, IEEE, 234–239, 2012.
[29] T. Denemark, J. Fridrich, Improving steganographic security by synchronizing the selection channel, in: Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security, 5–14, 2015.
[30] V. Holub, J. Fridrich, T. Denemark, Universal distortion function for steganography in an arbitrary domain, EURASIP Journal on Information Security, (1), 2014.
[31] D.-C. Wu, W.-H. Tsai, A steganographic method for images by pixelvalue differencing, Pattern Recognition Letters, 24(9), 1613–1626, 2003.
[32] A. D. Ker, Steganalysis of lsb matching in grayscale images, IEEE signal processing letters, 12(6), 441–444, (2005).
[33] N. Provos, P. Honeyman, Hide and seek: An introduction to steganography, IEEE security privacy, 99(3), 32–44, 2003.
[34] R. M. Haralick, K. Shanmugam, et al., Textural features for image classification, IEEE Transactions on systems, man, and cybernetics, (6), 610–621, 1973.

[35] C. A. Stanley, Pairs of values and the chi-squared attack, Department of Mathematics, Iowa State University.
[36] T. Zhang, X. Ping, Reliable detection of lsb steganography based on the difference image histogram, in: Acoustics, Speech, and Signal Processing, Proceedings, vol. 3, III-545, 2003.
[37] Sherly, A. P., et al. A novel approach for compressed video steganography, in Recent Trends in Network Security and Applications, vol. 89, 567–575, 2010.
[38] J. Kodovsky, J. J. Fridrich, Steganalysis of jpeg images using rich models., Media Watermarking, Security, and Forensics, 8303, 0A-1, 2012.
[39] J. Kodovsky, J. J. Fridrich, Steganalysis in high dimensions: fusing classifiers built on random subspaces, Media Forensics and Security, 78800L, 2011.
[40] T. Denemark, J. Fridrich, Steganography with Multiple JPEG Images of the Same Scene., IEEE TIFS, 12(10), 2308–2319, 2017.
[41] T. Filler, T. Pevny, P. Bas, Break our steganography system, Available at http://www, agents, cz/boss, 2010.
[42] Mon, S. Fepslin Athish, K. Suthendran, K. Arjun, and S. Arumugam, A Novel Reversible Data Hiding Method in Teleradiology to Maximize Data Capacity in Medical Images, International Conference on Theoretical Computer Science and Discrete Mathematics, Springer, 55–62, 2016.

Biographies

P. P. Amritha received her M.Tech. (Cyber Security) from Amrita Vishwa Vidyapeetham, currently pursuing her PhD at Amrita Vishwa Vidyapeetham. Her current research interests include: Steganography and code obfuscation.

M. Sethumadhavan received his PhD (Number Theory) from Calicut Regional Engineering College. Currently, he is working as a Professor in the Centre for Cyber Security, Amrita Vishwa Vidyapeetham, Coimbatore. His current research interests include: Cryptography and Boolean functions.

R. Krishnan received his PhD from IISc, Bangalore. He is currently an Adjunct Professor at Amrita Vishwa Vidyapeetham, Coimbatore. His current research interests include: Image processing and remote sensing.

Saibal Kumar Pal is presently the Director, Information Technology & Cyber Security at Defence Research & Development Organization (DRDO), New Delhi. His areas of interest are Cyber Security, Cryptography, and Computational Intelligence & High-Performance Computing. He has contributed in a number of R&D projects and international collaborations.