Contents lists available at ScienceDirect

Journal of Information Security and Applications

journal homepage: www.elsevier.com/locate/jisa

A review of image steganalysis techniques for digital forensics

Konstantinos Karampidis[a,*], Ergina Kavallieratou[a], Giorgos Papadourakis[b]

[a] *Department of Information & Communication Systems Engineering, University of the Aegean, 83200 Karlovasi, Samos, Greece*
[b] *Department of Informatics Engineering, Technological Educational Institute of Crete, 71410 Heraklion, Crete, Greece*

ARTICLE INFO

Article history:

Keywords:
Image steganalysis
Steganography
Digital forensics
Universal steganalysis
Deep learning

ABSTRACT

Steganalysis and steganography are the two different sides of the same coin. Steganography tries to hide messages in plain sight while steganalysis tries to detect their existence or even more to retrieve the embedded data. Both steganography and steganalysis received a great deal of attention, especially from law enforcement. While cryptography in many countries is being outlawed or limited, cyber criminals or even terrorists are extensively using steganography to avoid being arrested with encrypted incriminating material in their possession. Therefore, understanding the ways that messages can be embedded in a digital medium –in most cases in digital images-, and knowledge of state of the art methods to detect hidden information, is essential in exposing criminal activity. Digital image steganography is growing in use and application. Many powerful and robust methods of steganography and steganalysis have been presented in the literature over the last few years. In this literature review, we will discuss and present various steganalysis techniques – from earlier ones to state of the art- used for detection of hidden data embedded in digital images using various steganography techniques.

1. Introduction

Steganography is the art of covered or hidden messaging. It is far different from Cryptography which is the art of making something inevitable to understand, unless the cryptography key is known. Steganography hides a message in a medium -which is in plain sight-, but no one understands hidden message's existence unless he is aware of it. It is an ancient technique and the etymology of the word comes from ancient Greek words: steganos (cover) + grapho (write). This technique of hiding messages is very common now days since cryptography in many countries is forbidden or limited by law [1]. In many cases, if a suspect maintains cryptographic content and refuses to reveal the cryptography key, the authorities automatically consider him as guilty.

Steganalysis is the opposite procedure of steganography. Primarily, we try to detect the existence of steganographic content in a digital device and secondly discover the hidden message. From this point of view, steganalysis can be classified into two major categories i.e. passive or active. Passive steganalysis tries to classify a cover medium as stego and identify the steganographic embedding algorithm, while active steganalysis additionally tries to estimate the embedded message length and ideally extract it from the cover medium.

Digital forensics is a relative new field in Computer Science and focuses on the acquisition, preservation and analysis of digital evidence. As Palmer said, digital forensics are "the use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitation or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations."[2].

Both steganography and cryptography intend to hide information and often both are used together. Though cryptographic messages are easily detectable while they are meaningless, steganography messages appear to be normal at first sight. Based on knowledge of the actual message, the availability of the original cover file and the steganography tool, the following types of technical steganalysis can be distinguished [3]:

- Stego only attack - only the stego object is available for analysis.
- Known cover attack - the cover and the stego object are both available for analysis.
- Known message attack - the message is known and can be compared with the stego object.
- Chosen stego attack - the stego object and the stego tool (algorithm) are available for analysis.

[*] Corresponding author.
E-mail addresses: karampidis@aegean.gr (K. Karampidis), kavallieratou@aegean.gr (E. Kavallieratou), papadour@cs.teicrete.gr (G. Papadourakis).

https://doi.org/10.1016/j.jisa.2018.04.005
2214-2126/© 2018 Elsevier Ltd. All rights reserved.

- Chosen message attack - the steganalyst generates stego-media from some steganography tool or algorithm from a known message. The goal in this attack is to determine corresponding patterns in the stego-media that may point to the use of specific steganography tools or algorithms.
- Known stego attack - the steganography tool (algorithm) is known and both the original and stego-object are available.

Cover medium can be an image file, an audio file, a video file, a network packet or even a text file. It is obvious that as more elements are known to a digital forensics examiner, the more effective steganalysis will be. Furthermore, steganalysis becomes more complex when moving from detection only, to detecting and deciphering the embedded message i.e. moving from passive to active steganalysis. As steganography becomes more widely available and data either on digital devices or internet increases, the detection of steganographic content by digital forensics examiners becomes highly important. Theoretically, this concerns any type of digital objects, but practically -in most cases- audiovisual files are more frequently met. This literature review will deal with image steganography and analyze state of the art methods of steganalysis.

More than one hundred methods extended to any type of image steganalysis were recorded and presented. Two major approaches were adopted by scientists. The first one refers to extraction of statistical features from stego and clean images. These statistical features are compared then, in order to discriminate clean from stego images.

The second general approach is by employing machine learning techniques. Thus, features are extracted from images (both clean and stego), a classifier is trained, and finally unseen images are presented to the model for evaluation. Typical paradigms of the utilized classifiers are mostly Support Vector Machines (SVM) and artificial neural networks. In both approaches an interesting subject discussed widely in each paper – and a critical step for achieving best results- is feature extraction and selection. Many techniques were used for this, such as statistics (mean, kurtosis, skewness, histogram analysis etc.), covariance matrix, similarity measures between pixels etc. Apart from the two prementioned approaches, modern methods employ deep learning techniques such as convolutional neural networks or deep autoencoders, where feature extraction and selection is made in an almost automatic way.

The performance and the quantitative analysis of the techniques discussed in the following sections has also been given, by using metrics such as the detection rate, the error rate and ROC curves in specific embedding rates. In appendix we also provide tables (Table 1 to Table 5) for each steganalysis category. These tables besides basic information (i.e. author, date, method in brief) also indicate the evaluation metric, dataset and number of images used, in order to make the comparison between methods from the same steganalysis category more distinct.

In [4,5], authors propose a different taxonomy of steganalysis i.e. specific and statistical while in [6] and [7]. Paper in [8] only refers to steganalysis methods applicable to jpeg images, while in [9] authors only refer to methods for universal (blind) detection for image steganography. Our review provides a detailed reference from earlier steganalysis methods to state of the art, refers to all steganalysis categories and not only to specific ones (such as jpeg, universal etc.) and it is up to date including current trends like deep learning techniques.

All presented papers were retrieved by Google Scholar. Primarily, the search term “image AND steganalysis” was given and 5590 results were retrieved. In order to reduce the number of given papers, we searched again under the following constraints: a) the search term should be part of paper’s title, so that the presented papers by Google Scholar should be more relevant to our subject b) only papers from year 2000 since today are acceptable c) patents were not included d) books were not included. Search queries used were: i) allintitle: "steganalysis", which resulted to 2080 results ii) allintitle: "image steganalysis", which resulted to 344 results iii) allintitle: "lsb", which resulted to 2150 results iv) allintitle: "lsb matching" which resulted to 159 results v) allintitle: "universal steganalysis" which resulted to 75 results. Combining all the above search results from the different given search terms and by reading the abstract of each paper to determine if the paper was relevant to our research, we ended up having more than 100 papers which are presented in the following sections.

The rest of the paper is organized as follows. In Section 2 the taxonomy of steganalytic techniques is presented. Section 3 examines visual steganalysis, while Section 4 presents signature steganalysis techniques. Statistical steganalytic techniques are discussed in Section 5. Spread Spectrum Steganalysis is discussed in Section 6, while in Section 7 the Transform Domain Steganalysis techniques are presented. Finally in Section 8, Universal (blind) Steganalysis methods are examined, and Section 9 presents the conclusions derived from this review.

2. Taxonomy of steganalysis techniques

The simplest method of steganography is by embedding a message after the end of file (EOF) or by embedding hidden information into exif header. Both methods are simple and fast, but they are vulnerable to steganalysts. Even by looking the file with a hex editor, the message -if unencrypted- can be revealed. This simple technique is effective for people with little or none steganalysis knowledge, but it is very easy for digital forensic examiners to detect and retrieve the hidden information from the cover medium. Consequently, new steganography techniques were developed and new steganalytic approaches were proposed. Depending on the attack method a forensic examiner uses, six major categories are introduced:

- visual steganalysis
- signature or specific steganalysis
- statistical steganalysis
- spread spectrum steganalysis
- transform domain steganalysis
- universal or blind steganalysis

3. Visual steganalysis

Visual attacks are the simplest form of steganalysis. A visual attack is the examination of the suspicious image with the naked eye to identify any noticeable discrepancies. This turns to be very difficult, since the alterations made to an image when a message is embedded, do not result in quality degradation. Most steganographic algorithms create stego objects that are similar to their cover medium. However, when unaltered parts of a stego image are removed, it is possible to observe signs of manipulation. Hence, if a steganalyst can identify those features of the image that characterize it as stego, a visual attack may reveal the existence of a hidden message. The most common form of a visual attack concerns Least Significant Bit (LSB) steganography. The image is converted to its binary form and the bits in the LSB plane are retrieved. In an image usually, there are as many even values as there are odd, typically saying that there are approximately as many 1’s as there are 0’s in its LSB plane. When text is converted to binary however, there are often more 0’s than 1’s. This indicates a visual inconsistency and helps the forensic examiner to classify the image as stego. However, this steganalytic technique is successful only when a poor steganographic algorithm was used to produce the stego image. Typical software paradigms following that embedding

Table 1
Synoptic presentation of LSB steganalysis methods.

Authors - Ref	Year	Database	# Of images	Method	Accuracy – Detection rate – Error rate
Westfeld and Pfitzmann [16]	2000	No database used	5	Chi-squared detects of POVs	Various tests depending on steganography tool (Steganos, S-Tools, Jsteg, EZStego) and size of embedding message
Westfeld [17]	2003	No database used	7	Chi-squared detects of POVs	Various tests for 10 versions per true color image with different steganographic message sizes
Fridrich et al. [19]	2000	Color images, 350 × 250 pixels, stored as JPEGs	300	Raw Quick Pairs method. Statistical analysis of the image colors in the RGB cube	Various tests showing threshold and error probability for several different test message sizes and different secret message sizes.
Fridrich et al. [20]	2001	No database used	3	RS steganalysis	Various tests and results depending on initial bias, steganographic tool (Steganos, S-Tools, Hide4PGP) and image used (its size).
Avcibas et al. [21]	2001	Images were obtained from (1)	1800	Similarity measures between 7th and 8th plane	Various tests and results depending on embedding percentage (1−15%) and steganographic scheme (Outguess−, Outguess+, F5, LSB, LSB±).
Lyu et al. [22]	2003	Images were obtained from (1)	1800	Higher order Statistics – SVM Classifier	Various tests and results depending on embedding message length (32 × 32–256 × 256) and steganographic scheme (JSteg, Outguess−, Outguess+, EZStego,LSB) and classification method (Fisher Linear discriminant analysis or SVM).
Dumitrescu et al. [23–25]	2003	No known database used	29	Finite state machine	Various tests and results depending on embedding message length
Roue et al. [26]	2004	Kodak image database (2)	108	Marginal and joint probabilistic distributions of the image	Accuracy 70%
Lu et al. [27]	2004	No database used	4	Finite state machine with a new least square estimation	Various tests and results depending on embedding message length
Avcibas et al. [28]	2005	Images obtained from (3)	22	Analysis of Variance (ANOVA) - multivariate regression	Performance varies from 75% to 100% depending the watermarking algorithm used.
Dumitrescu et al. [29]	2005	Same as used in [23–25] plus ten colored high-resolution (2310 × 1814) uncompressed scanned images	39	High-order statistics of the samples	Various tests and results depending on embedding message length
Li Zhi et al. [30]	2003	No database used	4	Gradient Energy-Flipping Rate Detection (GEFR)	Various tests and results depending on embedding rate.
Zhang and Ping [31]	2003	USC-SIP1 Image database (4) CBIR Image Database (5)	5	Translation coefficients between difference image histograms	Various tests and results depending on embedding ratio.
Fridrich et al. [32].	2003	Color GIF images obtained using four different digital cameras, originally stored as high-quality JPEG images and later resampled to 800 × 600 Pixels.	180	Pairs Analysis	Various tests and results
Ker [34]	2004	2200 simulated uncompressed images, all 512 × 512. 5000 JPEG images, all 900 × 600. 10,000 JPEG images, sizes varying between 890 × 560 and 1050 × 691. 7500 JPEG images of very variable quality	24,700	Improved RS & Pair Analysis	Various tests and results depending on embedding message length
Celik et al. [35]	2004	Kodak Photo CD Images (2)	108	Feature set based on rate-distortion characteristics of images. Bayesian classifier preceded by a Karhunen-Loeve transform	Various tests and results depending on embedding rate (0–1.0bpp)
Benton and Chu [36]	2005	No database used	1000	RS method for feature extraction. Decision trees and neural networks for classification.	Various tests and results depending on embedding rate and classification method (decision tree & neural network).
Fridrich et al. [37]	2004	Images acquired from a digital camera, downsampled from original 2272 × 1704 resolution to 800 × 600 and converted to grayscale.	60	Estimation of hidden message via weighted stego image	Various tests and results.

(*continued on next page*)

Table 1 (*continued*)

Authors - Ref	Year	Database	# Of images	Method	Accuracy – Detection rate – Error rate
Ker et al. [38]	2008	1600 raw digital camera images. 3000 images from NRCS website (6). 1040 images supplied by researchers at Binghamton University	5640	Improved new weighted stego estimators	Various tests and results depending on image set and statistical measures (IRQ, Mean Error, Mean absolute error)
Chen et al. [39]	2006	USC-SIP1 Image database (4) CBIR Image Database (5)	520	Features based on 7th and 8th bit plane randomness tests. SVM for classification.	Various tests and results depending on embedding rate (0.01–0.15bpp) and steganographic scheme used (F5, LSB, LSB ±)
Bhattacharyya et al. [40]	2011	No database used	20	Auto-regressive model and SVM classifier	Various tests and results depending on embedding rate (0.01–1.0bpp)
H.B.Kekre et al. [41]	2011	No database used. BMP images (30 color and 30 grayscale) of size 128 × 128	60	Feature vectors derived from gray level co-occurrence matrix (GLCM). Euclidean distance as metric for classification.	Various tests showing detection accuracy per feature per embedding length
Fillatre [42]	2012	BOSSBase v0.92 (7)	9074	Adaptive statistical test based on the likelihood ratio.	Various statistical tests concerning BOSSBase images and comparison with other methods.
Fridrich et al. [44]	2013	BOSSBase v0.92 (7)	9074	Machine learning based detector utilizing co-occurrences of neighboring noise residuals as features.	Various tests and results concerning average detection error for different versions of the rich model, dependence on the change rate for two selected quality factors etc.

technique are Camouflage and JpegX [10,11], both early steganographic software that nowadays are outdated and least used due to their ease of detection [12]. A poor algorithm will embed the message bits directly after converting from ASCII to binary, and this will lead to the increase in 0's. This attack is usually related to palette images for LSB embedding in indices to the palette. Nevertheless, this technique has very poor results when trying to distinguish noisy images from stego images. In the less likely case that a forensic examiner detects the cover images in a digital device, the stego images are compared with the respective original cover images and differences are observed. Another indication of the existence of hidden messages is by trying to detect blank spaces in the possible stego images. That is, because some stego algorithms crop and pad the image in order to fit it into a fixed size. Moreover, differences in file size between cover image and stego images, increase or decrease of unique colors in stego images can also be used as indicators for the detection of hidden messages. Fig. 1 shows a clean (unaltered) image, while Fig. 2 is the same image with an embedded text file.

Fig. 3 shows the LSB plane of Fig. 1, while Fig. 4 shows the LSB plane of Fig. 2.

These examples show that sequential LSB embedding is easily detectable. For this reason, new steganographic software was developed, which embeds data to the carrier file in a randomized way. Fig. 5 shows the LSB plane of Fig. 1 and Fig. 6 shows the LSB plane, when randomized LSB embedding was performed. When Figs. 3 and 4 with Figs. 5 and 6 are compared respectively, we conclude that randomized LSB embedding is very strong to visual attacks.

4. Signature steganalysis

Another steganalytic technique is to observe any repetitive patterns (signatures) of a steganography software. These techniques search for signature patterns to determine the presence of a hidden message. For example, the string CDN is always added in the end of file when a message is embedded in an image with Hiderman steganography software as shown on Fig. 7.

Masker [13] – another steganography software – uses the last 77 bytes of a stego file for its signature. Jpegx [11] before embedding the hidden message at the end of jpeg's file marker, adds the sequence 5B 3B 31 53 00. There are many steganalytic software tools which scan files and identify signatures from various embedding algorithms. Therefore, it is rather easy for a forensic examiner to discover steganographic content if the stego image was produced with a tool which embeds its signature in the stego file. A method for identifying steganographic content in JPEG images regardless the tool's signature was proposed by Fridrich [14,15]. The image is divided into 8 × 8 blocks and the quantization matrix is extracted by analyzing the values of Discrete Cosine Transform (DCT) coefficients in all 8 × 8 blocks. The quantization table is then compared with standard JPEG quantization table for compatibility. If there are any incompatible blocks the image is characterized as stego.

5. Statistical steganalysis

Statistical steganalysis concerns those techniques developed by analyzing the embedding procedure and determining certain statistics that get modified as a result of the embedding process. Therefore, an in depth understanding of embedding process is needed in order to achieve maximum steganalytic accuracy. In spatial domain, the steganographic algorithm is applied directly on the pixels of the image. One of the earliest techniques is the so called Least Significant Bit Substitution (LSB) technique. Two different LSB approaches were introduced i.e. LSB replacement and LSB matching.

5.1. LSB replacement

In LSB Replacement, the cover image bytes have their least significant bits replaced by the secret data. There are two different embedding schemes in Least Significant Bit Substitution algorithms i.e. sequential and randomized. Sequential embedding denotes that the algorithm starts at the first pixel of the cover image and embeds the bits of the message data in order until whole message is embedded. Randomized embedding, disperses the positions of the values that will be modified to contain the bits of the embedded data.

Westfeld and Pfitzmann [16] proposed the first statistical steganalysis technique. The technique identifies Pairs of Values (POVs)

Table 2
Synoptic presentation of LSB matching steganalysis methods.

Authors- Ref	Year	Database	# Of images	Method	Accuracy – Detection rate – Error rate
Zou et al. [47]	2006	2812 images from Vision Research Lab (8). 1096 sample images included in the CorelDRAW Version 10.0 software CD#3 (9)	3908	2-D Markov chain of thresholded prediction-error image along with horizontal, vertical and diagonal directions serve as features. SVM with linear and nonlinear kernel as classifier.	52.28% for 0.01 bpp embedding rate – 97.75% for 0.1 bpp
Malekmohamadi et al. [48]	2009	Grayscale images taken from USC-SIPI (4)	200	Gabor filter coefficients and statistics of the gray level co-occurrence matrix of images as features. SVM as classifier	94.50% average detection rate for clean and stego images. Embedding rate 0.141 bpp.
Pevny et al. [49]	2010	9200 raw images from digital camera. BOWS2 (10) (10,700 grayscale images). NRCS (6) (1576 raw scans of film converted to grayscale with fixed size 2100 × 1500) - JPEG85 (9200 images from camera compressed by JPEG with quality factor 85). JOINT (images from all four databases above, 30,800 images approx.)	30,800 at topmost	Local dependences between differences of neighboring cover elements are modeled as a Markov chain, whose empirical probability transition matrix is taken as a feature vector. SVM as classifier	0.08–0.057 error rate when Embedding rate is 0.25bpp. 0.02 – 0.026 error rate when Embedding rate is 0.50bpp.
Zhang et al. [50]	2010	NRCS image database (6)	3185 color images in TIFF format converted to grayscale.	Statistical modeling of pixel difference distributions.	Embedding rate 50–100%: 68.48–98.27% True Positive respectively
Fridrich et al. [51]	2011	BOSSBase v0.92 (7)	9074 grayscale images	High-dimensional (33,963) feature vector, along with the use of ensemble classifiers obtained by fusing decisions of simple detectors implemented using the Fisher linear discriminant.	Embedding rates from 0.1–0.5bpp. Error rate from 21.0% to 7.3% respectively.
Gul et al. [52]	2011	BOSSBase v 0.92 (7) BOSSRank image set (7)	10,074 images	Features extracted by applying a function to the image constructing the k variate probability density function (PDF) estimates, and downsampling it by a suitable downsampling algorithm. Linear nad SVM classifier.	Accuracy 85% when using a SVM.
Fridrich et al. [53]	2012	BOSSBase v0.92 (7)	9074 grayscale images	Rich image models combined with ensemble classifiers	Payload from 0.0.5–0.40bpp. Error estimates on Mean Absolute Deviation from 0.065–0.0035 respectively.
Pevny et al. [54]	2012	Raw images from digital camera converted to grayscale and to JPEG (quality factor 80)	9200	Feature vector extracted from the investigated object and the embedding change rate. Support vector regression was utilized then.	Various tests and results depending on embedding rate and comparison to prior art
Cogranne et al. [55]	2013	BOSSBase v0.92 (7) BOWS Database (10)	9074 raw images 10,000 images	Generalized likelihood ratio test.	Various tests and results
Holub et al. [59]	2013	BOSSBase v1.01 (7)	10,000 raw images	Projection of neighboring residual samples onto a set of random vectors. Histogram of the projections was used as feature vector.	Various test along with the detection error for different embedding rate and three different content adaptive steganographic algorithms in the spatial domain
Xia et al. [60]	2014	NRCS (6) 3161 images each of them was split to four other in order and converted to grayscale BOSSBase v0.92 (7) - 9074 raw images	12,644	Co-occurrence matrix was used to model the differences with the small absolute value to extract features. SVM as classifier.	Various test concerning the detection of HUGO evaluated by “detection reliability” p ($p = 2$A-1, where A is the area below the ROC curve)
Xia et al. [61]	2016	NRCS (6) 3161 images each of them was split to four other in order and converted to grayscale BOSSBase v0.92 (7) - 9074 raw images	12,644	Calculation of the center of mass (COM) of the characteristic function of difference histogram (DHCF). SVM as classifier.	Various test on different embedding rate (0.10–1.0bpp) to two different datasets, with minimized classification error as metric.

(*continued on next page*)

Table 2 (*continued*)

Authors- Ref	Year	Database	# Of images	Method	Accuracy – Detection rate – Error rate
Goljan et al. [63]	2015	BOSSBase v1.01 (7)	10,000 raw images	Method discussed in [41] along with additional features extracted by three-dimensional co-occurrences of residuals computed from all three-color channels.	Various tests for different embedding rates (0.05–0.5bpp) with average detection error as metric, on variations of BossBase dataset and its grayscale versions.
Chen et al. [64]	2016	BOSSBase (7) – 10,000 images NRCS (6) – 3161 images	13,161	Calculation of the difference histogram characteristic function (DHCF) and the moment of DHCFs (DHCFM) and used them as discriminative features. Features were calibrated by decreasing the influence of image content on them and a SVM classifier was trained	Various test for embedding rate 0.25bpp. Results in papers figures
Lerch-Hostalot et al. [65]	2016	BOSSBase (7)	10,000	Unsupervised steganalysis method combined with artificial training sets and supervised classification.	Various tests for different embedding rates (0.1bpp, 0.2bpp, 0.25bpp, 0.4bpp) for three different steganographic algorithms. Comparison results with other methods
Sandoval et al. [66]	2017	BOWS (10) – 10,000 images UCID (11) – 1338 images	11,338	12 relevant features based on the probability density function (PDF) of difference of adjacent pixels and the co-occurrence matrix of the image. SVM as classifier	Various tests for different embedding rates (100%, 75%, 50%, 25%). 87.2% average detection accuracy.

Table 3
Synoptic presentation of Spread Spectrum Image Steganography (SSIS) steganalysis methods.

Authors- Ref	Year	Database	# Of images	Method	Accuracy – Detection rate – Error rate
Harmsen et al. [69]	2003	Images from Kodak PhotoCD PCD0992 (12)	24	Histogram Characteristic Function (HFC) – Center of Mass (COM). Mahalanobis distance as metric	95% accuracy at embedding rate of 1bpp
Chandramouli et al. [70]	2003	2D DCT coefficients of Lena image	1	First technique deploys regression. Second technique exploits higher order statistics	45% (approx.) estimation of message bits for the first technique. 70% (approx.) estimation of message bits for the second technique.
Wang et al. [71]	2003	Lena, Jet and Baboon	3	Histograms of pixel differences. Kolmogorov–Smirnov (KS) binary hypothesis test for classification.	Authors don't provide experimental results on large number of images.
Rongrong et al. [72]	2006	No reference by the authors.	300	Calculation of the scatter (variance) difference in both cover and it's "possible" stego image. Difference between the two scatters classifies the image.	Accuracy over 90%.
Sullivan et al. [73]	2005	1. digital camera images, partitioned into smaller sub-images 2. scanned photographs 3. scanned, downsampled, and cropped photographs 4. images from the Corel volume Scenic Sites. Images were converted losslessly to PNG and color images were converted to grayscale.	No reference by the authors	Markov random chain for modeling the correlation between pixels. SVm for classification.	95% accuracy
Li et al. [74]	2013	Variations of Baboon image	1	Multicarrier iterative generalized least-squares core algorithm	Authors compare their method with other ones. No experiments on a large database.

Table 4
Synoptic presentation of transform domain steganalysis methods.

Authors- Ref	Year	Database	# Of images	Method	Accuracy – Detection rate – Error rate
Liu et al. [75]	2003	No reference by the authors.	125	Extract features through DFT, DCT, DWT transform. Neural network as classifier.	Average accuracy 80.2%
Liu et al. [76]	2004	Part of USC-SIPI database (4) and images acquired from digital camera and the internet.	3056	Spectrum analysis and energy differences score differences in the histograms of clean and stego images. A threshold determines whether the image was stego or clean.	Successful detection rate of 99%
Liu et al. [78]	2004	First image set includes images as Lena, Peppers etc., digital photography taken by digital camera. Second image set is from corel image database (9)	183	Statistical analysis of the texture of the image. Neural network as classifier.	Successful detection rate of 84%
Sullivan et al. [79]	2004	Digital orthophoto quarter-quadrangle aerial images, Corel PhotoCD (9) images, and images taken with a Canon digital camera.	3000	Histogram as an empirical probability mass factor (PMF) for feature extraction. Supervised learning for classification.	Various tests on each image dataset. Error rate varies from 0.001 to 0.083 when depending on images quality factor.
Shi et al. [80]	2006	No reference by the authors.	7560	Second order statistics along with threshold utilization for dimensionality reduction. SVM as classifier.	Various tests depending on steganographic algorithm.
Kodovsky et al. [82]	2010	JPEG images acquired by 22 different digital cameras at full resolution in a raw format and then converted to grayscale.	6500	First method of estimation of change rate using the maximum likelihood principle. Second method based on minimizing a penalty function on cover images while increasing it on stego images.	Various tests and results on different estimators using median absolute error, median bias and interquartile range (IQR) as performance measures.
Liu et al. [83]	2010	Images from (13)	17,051	Extended the method discussed in [60] proposing a new approach for feature extraction. SVM as classifier	Various tests and results
Sheikhan et al. [84]	2010	UCID (11) Images were converted from TIFF to JPEG.	1338	Contourlet coefficients and co-occurrence metrics of sub-band images for features extraction. SVM as classifier	Average accuracy 96.29%
Kodovsky et al. [85]	2012	Images taken from camera	6500	Ensemble classifier	Various tests and results. Median (MED) testing error over ten different splits of the CAMERA database into a training and testing set, as well as the median absolute deviation (MAD) values.
Cho et al. [89]	2013	UCID (11) INRIA Holidays dataset (14)	2829	Decomposed image blocks	Various tests and results depending on method and classifier.
Lakshmi et al. [90]	2014	No reference by the authors.	20	Authors exploited a 3-Level DWT and calculated the energy value for both training and testing dataset. SVM as classifier	Accuracy 90%
Holub et al. [91]	2015	BOSSbase v1.01 (7)	10,000	Features derived first-order statistics of quantized noise residuals obtained from the decompressed JPEG image using 64 kernels of the DCT. Tests on selected state-of-the-art JPEG steganographic schemes.	Various tests and results depending on quality factor and steganographic scheme.

exchanged during message embedding. POVs can be pixel values, quantized DCT coefficients, or palette indices that differ in the LSB. Westfeld and Pfitzmann claimed that the frequencies of each of the two-pixel values in each POV tend to lie far from the mean of the POV. The Chi-squared attack detects these near-equal POVs in images and consequently embedded information. The Chi-squared method reliably detects sequentially embedded messages but has low success when embedding is randomized. A more generalized approach of chi square attack was used to detect messages that are randomly scattered in an image [17,18].

Fridrich et al. [19] proposed a method for detecting LSB embedding in 24-bit color images, the so called Raw Quick Pair (RQP) method. RQP analyzes close pairs of colors created by LSB embedding. Close color pairs indicate that two colors differ only at LSB. The process of embedding messages into images increases the number of close color pairs. Therefore, by counting the number of close color pairs we can characterize an image as stego or not. Authors showed that even for secret message capacities of 0.1 – 0.3 bits per pixel, it is possible to achieve a high degree of detection reliability. The drawback of this method is that it can be applied only to color images.

For this reason, Fridrich et al. proposed a new scheme for detection of LSB embedding in color and grayscale images, the so-called RS steganalysis [20]. This technique divides the image into groups and measures noise in every group. Afterwards, flipping of the LSBs of a fixed set of pixels within each group (by using a mask i.e. the

Table 5
Synoptic presentation of universal or blind steganalysis methods.

Authors- Ref	Year	Database	#Of images	Method	Accuracy – Detection rate – Error rate
Farid [92]	2002	Images were obtained from (1)	1800	Wavelet-like decomposition to build higher order statistical models of natural images. Fisher linear discriminant analysis for discrimination of images	Accuracy varies from 1.3% (LSB – message length 32×32) to 94% (Jsteg – message length 256×256).
Lyu et al. [93]	2004	Natural images downloaded from www.freefoto.com	40,000	Extended their work in [11] by applying their method to color images. SVM as classifier.	Various results depending on image (grayscale or color), embedded message length (from 10×10 to 80×80) and steganographic algorithm.
Lafferty et al. [95]	2004	No reference by the authors	2000	Local binary pattern texture operator as feature extractor. Artificial neural network for classification.	Various results depending on embedded message length (60 bytes to 100 bytes) and steganographic algorithm.
Xuan et al. [96]	2005	CorelDraw image database (9)	1096	Feature vector formed from the first three moments of characteristic function of wavelet sub-bands with the 3-level Haar wavelet decomposition. Bayes classifier.	Various results depending on embedded message length (10×10 to 80×80) and steganographic algorithm.
Shi et al. [97]	2005	CorelDraw image database (9)	1096	Features derived from the statistical moments of characteristic functions of the prediction-error image, the test image, and their wavelet sub-bands. Artificial neural network as classifier.	Detection rate 99.5%
Lie et al. [98]	2005	Variations of 132 images such as Lena, Baboon, Barbara etc.	2088	Features extracted from spatial and DCT domains. Nonlinear neural classifier.	Detection rates approx. 90%.
Farid et al. [99]	2006	Natural images downloaded from www.freefoto.com	40,000	Extended their work in [69] by including phase statistics in addition to first and higher order magnitude statistics. SVM as classifier.	Various results depending on quality factor (70–90, jpeg images) and steganographic algorithm used.
Chen et al. [100]	2006	CorelDraw version 11 CD#4 (9)	1349	Features extracted from projection histogram of Empirical Matrix and from prediction-error image. SVM as classifier.	Detection rate 98.1%
Sun et al. [101]	2008	Grayscale images in raw format downloaded from the website of vision research lab, University of California	600	Features extracted from co-occurrence matrices of thresholded differential images. SVM as classifier.	Various tests and results depending on payload (0.1–0.3bpp) and steganographic method used (LSB, ± 1). Combined Detection rate 72.2%.
Zhao et al. [102]	2011	UCID (11)	1388	Features from generalized difference images and color correlogram.	Detection rates from 61.85% to 100% depending on steganographic scheme and payload.
Zong et al. [103]	2012	NRCS (6) plus some common standard images.	2056	Method based on the correlation of inter- and intra-wavelet sub-bands in the wavelet domain and feature extraction from the co-occurrence matrix. SVM as classifier	Various tests and detection rates concerning feature combination, embedding method and image size.
Ghanbari et al. [104]	2012	USC-SIPI (4) BSDS (15) Images from internet	800	Features extracted from the GLCM of the original image and stego image. MLP as classifier.	Accuracy 80%
Zhang et al. [105]	2013	Images were obtained from (1)	5000	Method based on sparse representation. Sparse Representation Classification (SRC) algorithm and SVM for classification	Various results depending on embedding rate (25–100%), steganographic scheme and classification method (SRC–SVM)
Devi et al. [106]	2013	No reference made by authors	5931	Method based on minimizing image-to image variations.	Various results depending on embedding rate (0–100%).
Verma [107]	2014	Gray scale BMP images of size 256×256	60	Features extracted by Gray Level Co-Occurrence Matrix. MPL with Pre-processed Vectors Diagonal Back Propagation Algorithm (PVDBPA), was used as classifier.	Various results depending on version of algorithm used.
Lu et al. [108]	2014	BOSSBase v1.01 (7)	5000	Feature selection method based on the Fisher criterion.	Various results depending on embedding ratio (bpp) and embedding method.
Tang et al. [109]	2016	BOSSBase v1.02 (7)	10,000	Feature selection method based on the Fisher criterion, in which the separability of single-dimension and multiple dimension features, combined with measurement of the Euclidean distance, is analyzed.	Various results depending on embedding ratio (bpp) and embedding method.
Qian et al. [110]	2015	BOSSBase v1.01 (7) ImageNet (16) - 100,000 randomly selected images)	120,000	Deep Learning with convolutional neural networks (CNN)	Various tests depending on image database, embedding ratio (0.3–0.5bpp) and embedding method. Error rate as metric.

(*continued on next page*)

Table 5 (*continued*)

Authors- Ref	Year	Database	#Of images	Method	Accuracy – Detection rate – Error rate
Desai et al. [111]	2016	CorelDraw (9) BSDS500 (17)	1400	A reduced dimensional merged feature set for universal image steganalysis using Fisher Criterion and ANOVA techniques was used. SVM with RBF kernel as classifier	97% detection accuracy in various steganographic methods.
Couchot et al. [114]	2016	BosBase v1.01 (7) Raise database (18)	18,156	Deep Learning with convolutional neural networks (CNN)	Various tests depending on embedding method and theory different versions. Accuracy as metric.
Sajedi [115]	2016	Washington University image database (19) 3959 images were taken with six cameras with different resolutions	4959	Feature extraction via fuzzy if–then rules. SVM as classifier.	Embedding rate 0.05–0.4bpp. Average accuracy on different steganographic methods from 79–91%
Rostami et al. [116]	2016	BOSSBase (7)	10,000	Feature selection method based on based on optimization process of Particle Swarm Optimization (PSO) and AUC as fitness function. SVM as classifier.	Embedding rate 0.4bpp, detection accuracy 82.62%
Wu et al. [117]	2017	BOSSBase (7)	10,000	Deep residual network (DRN).	Average error rate 6.48%
Ye et al. [118]	2017	BOSSBase (7) BOWS (10)	20,000	Deep Learning with convolutional neural networks (CNN)	Various embedding rates (0.05–0.5bpp). Low detection error on various steganographic algorithms.
Nouri et al. [119]	2017	UCID (11)	506	Alteration of singular value curve was used to construct the steganalysis feature vector.	Embedding rates of 0.05, 0.1, 0.2 and 0.4 bpp. Various results on different steganographic algorithms. Comparison with other relevant feature extraction methods.

Fig. 1. Clean image of Lena.

Fig. 2. Stego image of Lena.

pattern of pixels to flip) is performed and every group is classified as regular or singular depending on whether the pixel noise within the group is increased or decreased. The classification is repeated for a dual type of flipping. RS steganalysis proved to be more reliable than Chi-square method.

Avcibas et al. [21] used image quality metrics -selected based on the analysis of variance (ANOVA) technique- as feature sets, to distinguish between cover-images and stego-images. The classifier between cover and stego-images was built using multivariate regression on the selected quality metrics and was trained based on an estimate of the original image. The embedded message sizes were 1/10, 1/40 and 1/100 of the cover image size depending on steganographic scheme used. The detection rate varied from 65% to 80%.

Lyu et al. [22] used higher-order statistics to capture certain properties from natural images. These properties were used as features to train a SVM. Several experiments were conducted depending on the varied embedding rate and the steganographic algorithm. The obtained classification accuracy reached a maximum of 94%.

Dumitrescu et al. based on Fridrich's work, presented a generalized case of methods given in [23–25]. They used a finite state machine whose states were selected multisets of sample pairs called trace multisets. This finite state machine helped them to formulate a quadratic function that estimates the length of embedded information with high precision.

Roue et al. [26] proposed an improvement in this method, by using marginal and joint probabilistic distributions of the image.

Lu et al. [27] also proposed a variation of method presented in [23]. They combined the statistical measures developed in [23] and a new least square estimation. The proposed method in comparison to SPA, showed less false alarm rate (13.79% when SPA false rate is 5%). Moreover, the estimating precision is approximately 9%

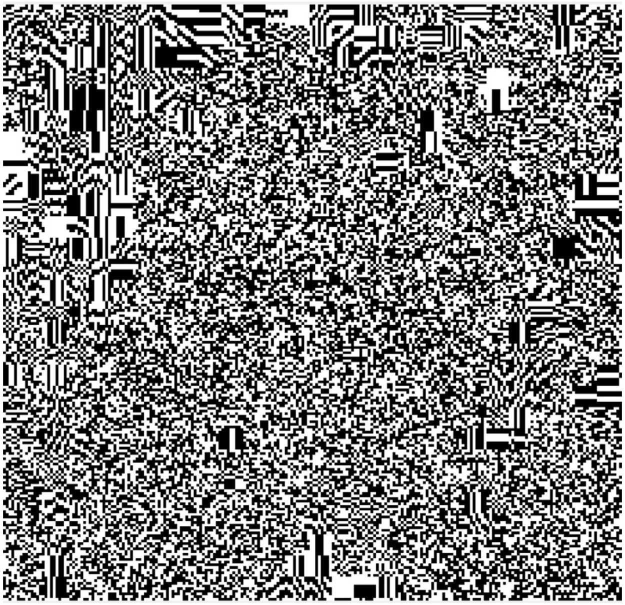


Fig. 3. LSB of clean image.

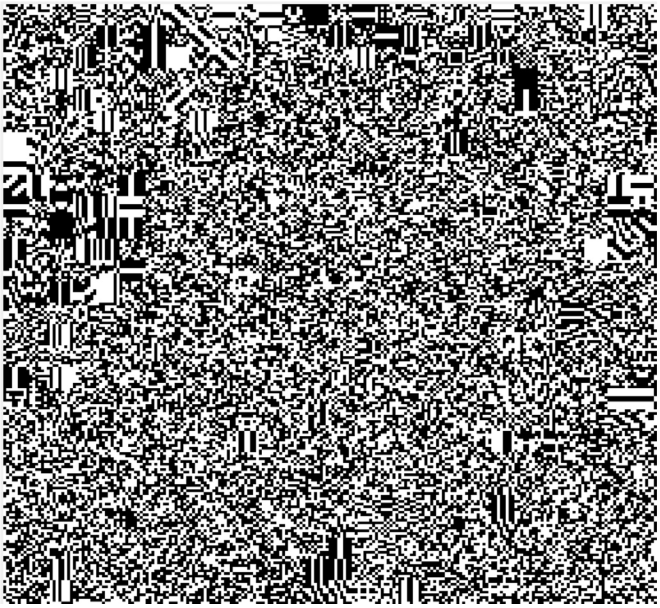


Fig. 4. LSB of stego image.

higher than that of SPA method (88%) if the embedding ratio is lower than 10%.

Avcibas et al. [28] proposed a method which searches the 7th and 8th bit planes of an image and calculates several similarity measures. Their approach was based on the fact that correlation between the specific planes of the image and the binary texture characteristics within these bit planes, are different in a stego and a cover image. Several features were calculated, and these features were utilized to train a SVM to classify images as clean or stego. The classifier was trained with all embedding percentages from 1% to 15% and the detection rate varied from 48.80% to 92.17% depending also on the steganographic scheme used.

Dumitrescu et al. [29] also proposed another method that exploits high-order statistics of the samples in order to derive a detection equation. They estimated the hidden message's length by measuring signature statistical quantity. This method proved to be effective on both color and grayscale images.

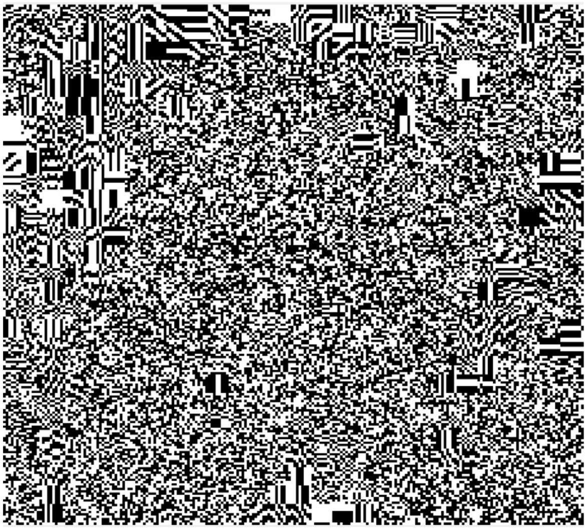


Fig. 5. LSB of clean image.

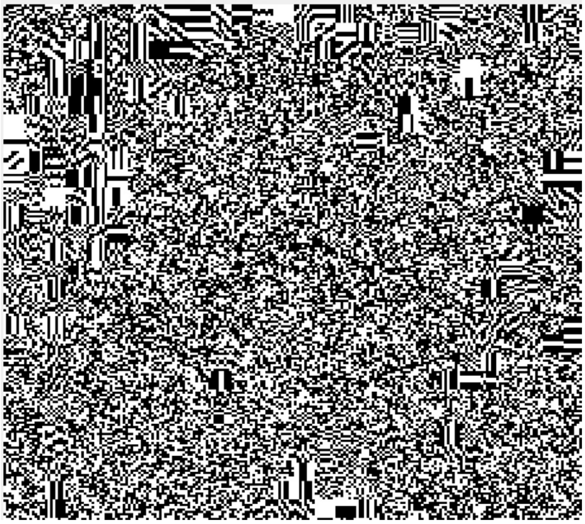
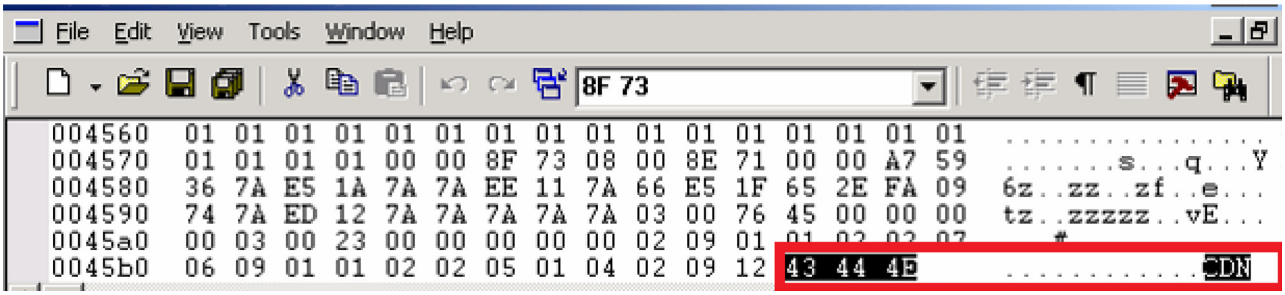


Fig. 6. LSB randomized.

Li Zhi et al. [30] proposed Gradient Energy-Flipping Rate Detection (GEFR). GEFR calculates the gradient energy both of the cover and the stego image. Then the Gradient Energy curve is utilized to estimate the message length. When embedding rate is more than 0.05 bits per pixel, the technique reliably detects the presence of the secret message.

Zhang and Ping [31] proposed another technique for grayscale images. The technique is based on the difference image histogram. Translation coefficients between difference image histograms were utilized as the measure to indicate the weak correlation between the least significant bit (LSB) plane and the remaining bit planes. This measure was then used to construct a classifier in order to discriminate the stego-image from the carrier-image. Embedding rates varied from 0% to 100% in 10% increments, while the detection rate reached an average of 96.03% at topmost. The proposed algorithm works well both for sequential or random LSB replacement and shows better performance and computation speed than RS analysis.

A method for 8-bit GIF images known as Pairs Analysis was proposed by Fridrich et al. [32]. The technique uses patterns formed



```
004560  01 01 01 01 01 01 01 01 01 01 01 01 01 01 01 01  ................
004570  01 01 01 01 00 00 8F 73 08 00 8E 71 00 00 A7 59  .......s...q...Y
004580  36 7A E5 1A 7A 7A EE 11 7A 66 E5 1F 65 2E FA 09  6z..zz..zf..e...
004590  74 7A ED 12 7A 7A 7A 7A 7A 03 00 76 45 00 00 00  tz..zzzzz..vE...
0045a0  00 03 00 23 00 00 00 00 00 02 09 01 01 02 02 07     #
0045b0  06 09 01 01 02 02 05 01 04 02 09 12 43 44 4E     .............CDN
```

Fig. 7. LSB plane after randomized embedding.

by pairs of colors (color cuts) to estimate the length of the secret message. The structure of the color cuts is measured using an entropy-like quantity R -which is in fact a quadratic function of the secret message length- and based on R, they estimate the unknown message length from the stego image. This technique outperforms the Chi-square attack [21] and for BMP and palette images it produces more reliable results than RS steganalysis [20]. Nevertheless, for grayscale images, Pairs Analysis is slightly worse than RS steganalysis.

Ker in [33] evaluated both Pairs Analysis and RS steganalysis techniques and proposed improvements in both of them for grayscale images [34].

Another method was presented by Celik et al. [35]. Based on the observation that hidden message embedding increases image's entropy and various hiding method introduce small imperceptible distortions, they formed a feature set based on rate-distortion characteristics of images. This feature set was utilized to train a Bayesian classifier preceded by a Karhunen–Loeve transform and eventually classify images as clean images or stego. Embedding rate was 0.1, 0.2, 0.4, 0.6, 0.8, and 1.0 bits per pixel (bpp) and 27% of the cover images were mislabeled as stego-images, while the miss rate decreased with increasing embedding rate.

Benton and Chu [36] used decision trees and neural networks in order to discriminate clean from stego images. In order to extract features, they used the RS method with a slightly different approach than the original RS, as the goal of their approach was to decide whether the image contained hidden data and not in estimating the embedding probability.

Fridrich et al. [37] introduced the concept of a weighted stego image and then formulated the problem of determining the unknown message length as a simple optimization problem. The accuracy of this method in detection of hidden information and estimation of embedding ratio is relatively high.

Ker et al. [38] revisited weighted stego image steganalysis for estimating LSB replacement payload sizes in digital images. They suggested new weighted stego estimators by upgrading the method's three components i.e. cover pixel prediction, least-squares weighting, and bias correction. These new methods compared to other structural detectors, managed to improve accuracy while not being complex.

Chen et al. [39] proposed a technique based on 7th and 8th bit plane randomness tests. A scan of the two-bit planes was performed, and two binary sequences were obtained. Afterwards, the randomness of these two sequences were tested by several randomness tests respectively. The results of the randomness tests were used as attributes to construct a classifier to distinguish between stego and cover images. The results showed the detection accuracy of method was higher than 95% to stego images with an embedding rate higher than 0.05 bits per pixel.

Bhattacharyya et al. [40] used an auto-regressive model and a SVM classifier to detect the presence of the hidden messages, along with multiple regression parameters in order to predict the length of the hidden information. Embedding rates varied from 10% to 100% with 10% increments and maximum accuracy achieved.

Kekre et al. [41] used feature extraction and distance measures to detect stego images. The extracted feature vectors were derived from gray level co-occurrence matrix (GLCM) as they noticed that there is a difference between the features of stego and non-stego images. Afterwards, they compared distance metrics like Absolute distance and Euclidean distance for classification and they concluded that Euclidean distance gives the best results. Their method works in case of both grayscale and color images.

Fillatre [42] designed an adaptive statistical test that its probability distribution is always independent of the unknown image parameters i.e. the mean level and the covariance matrix of the image. The unknown parameters are replaced by estimates based on a local linear regression model. The specific statistical test is based on the likelihood ratio. Experiments were conducted on real natural images derived from BOSSBase image set [43] and the proposed method was also compared to other state of the art. The resulted ROC curve showed the relevance of the method.

Fridrich et al. [44] proposed a machine learning based detector utilizing co-occurrences of neighboring noise residuals as features. Researchers adapted the features for detection of LSB replacement by making them aware of pixel parity. Then they introduced two key novel concepts – calibration by parity and parity-aware residuals. It was shown that, for a known cover source when a binary classifier can be built, its accuracy is better in comparison with the best structural and WS detectors in both uncompressed images and in decompressed JPEGs. This improvement is especially significant for very small change rates.

Verma et al. [45] used a Difference Image Histograms (DIH) for both suspicious and original image, then flipped LSB bits to both images, reconstructed the DIH and compared them in order to characterize the suspicious image as stego or clean.

5.2. LSB matching

LSB replacement technique proved to be very vulnerable. In order to avoid certain statistical attacks Sharp [46] introduced LSB matching steganography technique. In the LSB Matching embedding algorithm each secret data bit is compared with the least significant bit of the corresponding cover byte. If the two compared bits match, no change is made while in the case of a mismatch the cover byte is incremented or decremented at random. Let C is the cover image, C_i the ith LSB bit, M the hidden message, M_i the ith bit of M, S the resulted stego image and S_i the ith LSB of the stego image. Eq. (1) shows the embedding process for LSB matching

$$S_i = \begin{cases} C_i, if\ M_i = C_i \\ C_i - 1,\ \ if\ M_i \neq C_i\ and\ C_i \neq 0 \\ C_i + 1,\ \ if\ M_i \neq C_i\ and\ C_i = 0 \end{cases} \tag{1}$$

LSB Matching retains the characteristics of LSB replacement but it is more difficult to be detected from statistical perspective. Consequently, previous mentioned methods on LSB replacement have low detection accuracy on LSB matching.

Zou et al. [47] proposed a steganalysis system based on 2-D Markov chain of thresholded prediction-error image. A non-linear SVM was utilized as classifier and extensive experiments were conducted which showed very good results. Embeddings rates varied from 0.01bpp to 0.3bpp and the average detection rate was 52.28% to 97.75% respectively. This method also performs well as a universal stego detector.

Malekmohamadi et al. [48] proposed a method for steganalysis of grayscale images using spatial and Gabor features. They used spatial relationships between pixels of clean and stego images for feature selection. Those features were utilized to train a SVM classifier. Gabor filter coefficients were also used to form their input vectors for training an agent. First and higher order statistics from the whole image and its DCT transform have been employed. The trained model was then applied to unseen altered and clean images. The results showed a high correct detection rate i.e. 93% for altered images and 96% for clean images while the embedding rate for the algorithm was 14.1%.

Pevny et al. [49] proposed a novel approach to steganalysis of LSB matching by introducing a new feature set, the so called SPAM feature set. The local dependences between differences of neighboring cover elements are modeled as a Markov chain, whose empirical probability transition matrix is taken as a feature vector. The conducted experiments showed that SPAM feature set can reliably detect algorithms hiding in the block DCT domain as well.

Zhang et al. [50] proposed a LSB matching steganalytic method based on statistical modeling of pixel difference distributions. The method examines the number of non-zero difference values from stego-images and the number of the zero-difference value. Afterwards, the estimation of the relative error between the estimated and actual values of the number of the zero-difference value is used as the classification feature.

Fridrich et al. [51] attacked a content-adaptive steganographic algorithm (HUGO) and identified features capable of detecting payload embedded using such schemes. Afterwards they utilized ensemble classifiers obtained by fusing decisions of base learners trained on random subspaces of the feature space. The best performance achieved on BOSSRank test set [43] was 80.3% and the embedding rate was 0.4bpp.

Gul et al. [52] attacked HUGO as well. First, they extracted features by applying a function to the image constructing the *k* variate probability density function (PDF) estimates, and downsampling it by a suitable downsampling algorithm. Images from BOSSBase were used as training set while BOSSRank was the test set, with an embedding rate at 0.4bpp. Feature selection improved very slightly the detection accuracy i.e. 0.3% in average. The best detection rate attained was 85% when 957 features were selected and a SVM was utilized as classifier.

Fridrich et al. [53] used rich image models combined with ensemble classifiers in order to automatize steganalysis for a wide spectrum of steganographic schemes. They assembled a rich model of the noise component by considering various qualitatively diverse relationships between pixels. Then, ensemble classifiers were used to assemble the model and the final steganalyzer.

In [54] authors used a 275-dimensional feature vector to discriminate stego from clean images. This feature vector was consisted of 193 features (calculated from DCT coefficients) and 81 calibrated Markov features, while the 275th feature improved the accuracy of the steganalyzer, helping it to adjust to different values of features on images of different size. Then, by using regression they learned the relation between the feature's location and the change rate. This method is applicable for both LSB replacement and matching steganography.

Cogranne et al. [55] presented a test for LSB Matching detection. Authors introduced a test based on the likelihood ratio, which maximizes the detection power regardless the embedding rate is. Afterwards, they calculated the statistical properties of this test and finally they presented a generalized likelihood ratio test by replacing the unknown medium parameters by their estimation. The proposed test was performed on BOSSBase and BOWS [56] image sets, both publicly available. Authors also compared their proposed method with other state of the art such those described in [57,58] and the resulted ROC curves showed that the proposed method performs well.

In [59] Holub et al. proposed an alternative statistical representation. The authors projected neighboring residual samples onto a set of random vectors and took the first-order statistic (histogram) of the projections as the features. To evaluate the performance of their method authors attacked three steganographic schemes on two different test sets, with an embedding rate from 0.1bpp to 0.4bpp. Authors also contrasted the results against several state-of-the-art domain specific features sets.

Xia et al. [60] showed that LSB matching smoothes the histogram of multi-order differences. Based on this observation, they used the co-occurrence matrix to model the differences with the small absolute value to extract features. Support vector machine classifiers were trained with these features, to distinguish stego images from the original ones. Experiments were carried out on three test sets, the embedding rate varied from 0.1bpp to 1.0bpp and comparison to other state of the art methods has also been made.

Xia et al. [61] proved that after embedding a message with LSB Matching, the histogram of the differences between pixel gray values is smoothed by the stego bits even if there is a large distance between the pixels. Also, the center of mass of the characteristic function of difference histogram (DHCF COM) decreases after messages are embedded. Thus, the DHCF COMs were calculated and used as features and a SVM was trained to detect the existence of hidden messages. Feature sets from adjacent and non-adjacent pixels were made, namely DHCF COMs#1 and DHCF COMs#2. BOSSBase and NRCS [62] were the two image sets utilized as test sets. Moreover, the proposed method was compared with the methods described in [58] and [49]. Features extracted from nonadjacent pixels do not depend on image correlation. This may be the reason that the combination of SPAM and features in DHCF COMs#2 can get a better detection result.

In [63] an extension of the spatial rich model [53] for steganalysis of color images was proposed. The additional features used, were extracted by three-dimensional co-occurrences of residuals computed from all three-color channels. These features can capture dependencies across color channels. Experiments were conducted on three image databases - different color versions of BOSSBase v1.01 - with an embedding rate 0.1 bpp for LSB Matching and 0.4 bpp for WOW. These experiments showed that the proposed feature set (18,157 features) proved to be extremely powerful for detection of LSB Matching steganography in images. The average detection error for one payload is 0.0297–0.1790 (LSB Matching for the three test sets), while for different payloads (0.05–0.5 bpc) is also small as shown in paper's Figs. 2 and 3.

Chen et al. [64] proposed a method that calculates the differences among pixel pairs and proved that the histogram of difference values is smoothed by stego noises. They calculated the difference histogram characteristic function (DHCF) and the moment of DHCFs (DHCFM) and used them as discriminative features. Features were calibrated by decreasing the influence of image content on them and a SVM classifier based on the calibrated features, was trained. BOSSBase and NRCS were used as training and test sets and the embedding rate was 0.25bpp. Experimental results demonstrate that the DHCFMs calculated with nonadjacent pixels were helpful to detect stego messages hidden by LSB matching.

Lerch-Hostalot et al. [65] provided an unsupervised steganalysis method that combined artificial training sets and supervised classification. This method assumes that the embedding algorithm and the approximate bit rate used by the steganographer are known. BOSSBase image set was used to produce stego images with various embedding rates (0.10bpp, 0.20bpp, 0.25bpp and 0.40bpp). The model has been tested on three steganographic methods and the extensive comparative experiments done, showed that the proposed method achieves better classification accuracy than that obtained of traditional supervised steganalysis (Rich Models, Ensemble Classifiers etc.)

In [66] Sandoval et al. chose the 12 most relevant features based on the probability density function (PDF) of difference of adjacent pixels and the co-occurrence matrix of the image. This feature vector trained a SVM to distinguish stegoimages from the natural images. To evaluate the proposed steganalysis scheme, they used two image data sets, BOWS and UCID [67] under four different embedding rates or payloads, i.e. 100%, 75%, 50% and 25%. Experimental results showed that the proposed scheme provides better performance - 87.2% in average- than previously proposed methods.

6. Spread spectrum steganalysis

Spread Spectrum Image Steganography (SSIS) was first described by Marvel et al. [68]. SSIS embeds the hidden information within noise, which is then added to the digital image. This noise if kept at low levels, is not distinguishable to the human eye.

Harmsen et al [69] presented a spread-spectrum steganalysis method for color images, using Histogram Characteristic Function (HFC) -which is the Fourier Transform of image histogram- and exploiting the properties of center of mass of HCF where center of mass is the first order moment. Two different experiments were conducted. The first detected images when the embedding method was known, and the detection rate was 94.68%. The second one detected images when the embedding method was unknown. The detection rate in this case was 95.89%. In both cases the embedding rate of 1 bpp.

Chandramouli et al. [70] proposed two other steganalysis schemes for spread spectrum steganography. The first scheme does not exploit higher order statistics. It uses regression techniques to estimate cover image from stego image. Afterwards in order to obtain the estimate of the secret message, the estimated value is subtracted from the stego image. The second exploits higher order statistics. Experiments showed that in comparison to the first proposed scheme, exploiting higher order statistics improved performance of steganalysis.

Wang et al. [71] proposed a technique for block DCT based steganography. Authors noticed that pairs of neighboring pixels within an 8 × 8 block have different statistics from those across two 8 × 8 blocks. Two histograms of pixel differences were computed for which a Kolmogorov–Smirnov (KS) binary hypothesis test discriminates stego from clean images.

Another method is given by Rongrong et al. [72]. This method is based on block based scatter (variance) difference detection. Primary, after applying a spatial filter, the cover image is restored. Afterwards, the spread spectrum process is performed on the test image several times and the scatter of low frequency coefficients in each DCT block is estimated. The same process is applied over the estimated cover image and its own scatter is estimated as well. Finally, the difference between the two scatters determines if there is a spread spectrum message.

Sullivan et al. [73] proposed a steganalysis method suitable for grayscale images. They modeled the correlation between pixels in an image, by utilizing a Markov random chain. Afterwards a SVM was trained with both clean images and images embedded with spread spectrum steganography.

Li et al. [74] developed a low complexity multicarrier iterative generalized least-squares core algorithm to extract unknown messages, hidden in image hosts via spread-spectrum embedding.

7. Transform domain steganalysis

As more attacks on various steganographic schemes were presented by steganalysts, there was the need of finding steganographic methods more robust to attacks such as compression, filtering etc. Various transform domains techniques were utilized such as Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT) and Fast Fourier Transform (FFT) in order to hide information in transform coefficients of the cover images.

Liu et al. [75] transformed digital images (both clean and stego) in DFT, DCT, DWT transform domains. Each image was divided into 8 × 8 sub-block and DCT was performed in each sub-block. In DFT and DWT data hiding process, selected metrics were based on magnitude (two statistics, image and its sub block). Then the three levels of DWT were taken under consideration in each training image, and the mean value, variance, skewness and kurtosis of each part of every level was calculated. This procedure produced 36 features for each image. Added with DCT's 4 statistics based image metrics, a 40-d feature vector was created. An artificial neural network was trained then and the average detection rate was 80.2%.

Another approach was given in [76]. They proposed a method specific for wavelet domain quantization modulation technique [77]. It has been observed that histogram shape of cover image is smoother than stego image. Spectrum analysis and energy differences was used to score for differences in the histograms of clean and stego image and a threshold was used to determine whether the image was stego or clean.

Liu et al. [78] proposed another technique based on statistical analysis of the texture of the image. Once more, they used a neural network as a classifier for clean and stego images but their approach for extracting features was different from [75]. Specifically, wavelet coefficients in each sub-band of a three-level wavelet transform were modeled as a Generalized Gaussian distribution (GGD) with two parameters α and β. Consequently, nine pairs of those parameters i.e. eighteen image features were utilized as inputs of a neural network. Authors hided a 64 × 64 binary bitmap in images from two test sets and the average correct detection rate reached up 84%.

Sullivan et al. [79] proposed a steganalysis method specific to Quantization Index Modulation (QIM) data hiding. They tested QIM that embeds in 8 × 8 blockwise DCT coefficients of an image. They used the histogram as an empirical probability mass factor (PMF) for acquiring a 300-dimensional feature vector. A supervised learning procedure was employed later to train the classifier. Three different image sets were used as training and test sets. Experiments were conducted in a supervised learning approach and showed that even when the quality factor was unknown on a mixed (from all three datasets) set of images, the detection error was low i.e. 0.01–0.083.

Shi et al. [80] presented a new steganalysis scheme to effectively detect the advanced JPEG steganography. They worked on JPEG 2-D arrays formed from the magnitudes of JPEG quantized block DCT coefficients. Difference JPEG 2-D arrays along horizontal, vertical and diagonal directions were then used to enhance changes caused by JPEG steganography and Markov process was applied to modeling these difference JPEG 2-D arrays so as to utilize the second order statistics for steganalysis. Furthermore, a thresholding technique was developed to reduce the dimensionality of feature vectors, in order to make the computational complexity of the proposed scheme wieldy.

Westfeld [81] proposed a methodology to apply higher order steganalytic attacks from the spatial domain to the transform domain. More specifically, 72 methods derived from the spatial domain were examined. There was also examined the proposed method's detection power and precision compared to prior methods and finally designated the way properties like image size and JPEG quality effect the ranking of the proposed attacks.

Kodovsky et al. [82] introduced two approaches for detecting hidden data using LSB embedding in quantized DCT coefficients of a JPEG file. At first, a change rate estimation using the maximum likelihood principle was introduced. Due to this model's high complexity, another method was proposed also, based on minimizing a penalty function on cover images while increasing it on stego images.

Liu et al. [83] expanded the work in [80] and proposed a new scheme. The features of the joint density of the differential neighboring in the DCT and the DWT domains and the errors of the polynomial fitting on the histogram of the DCT coefficients constitute the original ExPanded Features (EPF). Features were also extracted from the calibrated version i.e. the so-called reference EPF features. The difference between the original and the reference EPF features was calculated then, and finally the original EPF features and the difference were merged to form the feature vector for classification. Feature selections techniques were applied and a SVM was used to detect stego-images.

Sheikhan et al. [84] extracted statistical features of Contourlet coefficients and co-occurrence metrics of sub-band images. In order to decrease extracted features, ANOVA method was performed, and the selected features were utilized to train a nonlinear SVM for classifying images as stego or clean.

Kodovsky et al. [85] proposed the use of an ensemble classifier instead of a SVM due to the fact that ensemble classifiers are computationally less complex compared to SVMs. The lower training complexity makes possible to work with high-dimensional cover models and train on larger training sets. A 7. 850- dimensional Cartesian- calibrated feature set (*CF**) was used to train the ensemble classifier to detect nsF5 (an improved version of F5 [86]) steganographic algorithm. When unknown test images were presented to the model, the obtained testing error was 0.1702. Authors also tested their method with stego images produced from other steganographic schemes such as YASS [87] and MBS [88] with payload from 0.01 to 0.05 bpac. The performance of the ensemble classifier using *CF** features was also compared to the state-of-the-art classifiers. The reported median (MED) testing error as well as the median absolute deviation (MAD) values showed that the proposed method performs well.

Cho et al. [89] discriminated a stego image from its cover image based on steganalysis of decomposed image blocks. They decomposed the image into blocks, classified those blocks into multiple classes and found a classifier for each class. Afterwards by integrating results of all image blocks via decision fusion, the whole image was classified as stego or clean. During their research they observed that the performance of block-based image steganalysis is less sensitive to the decision fusion methods but more sensitive to the classifier choice.

In [90] Lakshmi et al. exploited a 3-Level DWT and calculated the energy value for both training and testing dataset. The extracted features trained a SVM which was utilized for classification. Stego images were created by hiding multiple images in the cover medium and the accuracy obtained was 90%.

Holub et al. [91] introduced a novel feature set for steganalysis of JPEG images. The features are first-order statistics of quantized noise residuals obtained from the decompressed JPEG image using 64 kernels of the DCT (the so-called undecimated DCT). The proposed steganalysis feature set has low computational complexity, lower dimensionality in comparison with other rich models, and a competitive performance with respect to previously proposed JPEG domain steganalysis features.

8. Universal or blind steganalysis

Universal steganalysis tries to detect the embedded messages regardless the steganographic technique applied to cover image. The main difficulty is to find relevant features which are characteristic for stego images. Afterwards machine learning techniques are used to build a detection model from the experimental data. When the method identifies stego images regardless the steganographic method the hidden message was embedded in cover medium, we can call it a universal (blind) method, while when the method attacks specific steganographic algorithms we may call it a semi-blind one.

The first attempt to build a universal steganalyzer was made from Avcibas [21]. This method was based clearly on statistical as already stated earlier.

Farid [92] used a wavelet-like decomposition to build higher order statistical models of natural images. A Fisher linear discriminant analysis was then used to discriminate between untouched and altered images. Accuracy varied from 1.3% to 94% depending on steganographic algorithm and message length (32×32 to 256×256).

Lyu et al. [22] also used wavelet-like decomposition to build higher-order statistical models of natural images. A SVM was used afterwards to discriminate clean and stego images. An extension was proposed in [93] from the same researchers in order to apply their model to color images. A one class SVM (OC-SVM) was used to simplify the training process of the classifier.

Harmsen et al. [69] considered hidden information as additive noise. Therefore, they introduced a blind detection scheme that used only statistics from unaltered images. By calculating the Mahalanobis distance from a test Center of Mass (COM) to the training distribution, a threshold was used to identify steganographic images.

Trivedi et al. [94] presented a steganalysis method for sequential steganography. Abrupt changes in statistics due to sequential steganography were exploited to estimate the message location and length. These abrupt changes were used as a feature that distinguishes sequential steganography embedding from other types of embedding. Sequential probability ratio test was employed as a mathematical tool, and as a result cumulative sum (CUSUM) test statistics was derived for detecting steganography.

Lafferty et al. [95] proposed a method which utilizes the local binary pattern texture operator to examine the pixel texture patterns within neighborhoods across the color planes. An artificial neural network was used as classifier. For the training set the embedding rate was 0.0049 bpp, while for the test image sets the embedding rate was 0.0082 bpp. Accuracy depending on steganograhic algorithm used, varied from 86.5% to 88.6%.

A semi blind steganalysis technique based on multiple features formed by statistical moments of wavelet characteristic functions was proposed by Xuan et al. [96]. A 39-d feature vector was formed from the first three moments of characteristic function of wavelet sub-bands with the 3-level Haar wavelet decomposition. A Bayes classifier was used for classification. This method is also effective for spread spectrum hiding methods.

Shi et al. [97] proposed a blind steganalysis system, in which the statistical moments of characteristic functions of the prediction-error image, the test image, and their wavelet subbands were selected as features. An artificial neural network was utilized as the classifier and the model's average accuracy reached 98.7%. Authors also compared the classifier by deploying their method with a Bayes classifier and proved that the artificial neural network had better classification results.

Lie et al. [98] used two image features in order to build a blind model. Their technique is based on the analysis of two properties in the spatial and DCT domains. A non-linear neural classifier based on these two extracted features was used to achieve classification. A database composed of 2088 plain and stego images (generated by using six different embedding schemes) was used for evaluation. The proposed model managed to give 90% positive detection rate regardless the embedding technique. The embedding rate varied from 0.01 to 2.66 bpp depending on the steganographic scheme.

Farid and Lyu again extended [99] their model to include phase statistics in addition to first and higher order magnitude statistics, extracted from multi-scale, multi-orientation image decompositions. Experiments were conducted on a large collection of images, concerning eight different steganographic embedding algorithms and results showed that this method is reliable.

Chen et al. [100] used the projection histogram of EM to extract features composed of two parts i.e. the moments of projection histogram (pH) and the moments of the characteristic function of pH. Features were extracted also from prediction-error image in order to enhance performance. A SVM was utilized as classifier. The proposed model was tested on six (6) different steganographic schemes and the average detection rate was 98.1%.

Sun et al. in [101] introduced a universal steganalysis method based on co-occurrence matrix of differential image. They calculated the forward difference in three directions (horizontal, vertical and diagonal), towards adjacent pixels to obtain three-directional differential images for a natural image. Then they set a threshold and removed the redundant information. The co-occurrence matrices of thresholded differential images was used as features for steganalysis. A SVM with RBF kernel was used as a classifier to discriminate stego images and cover images. This method is effective in steganographic schemes applied in spatial domain.

Zhao et al. [102] proposed a steganalysis algorithm for palette-Based images. More specifically they focused on cover images of GIF format, transformed from natural images. They extracted features from generalized difference images and color correlogram. A two-class classification scheme was used to differentiate cover images and stego images, with high accuracy when the embedding rate was no less than 20%.

Zong et al. [103] proposed a blind JPEG steganalysis method based on the correlation of inter- and intra-wavelet sub-bands in the wavelet domain and feature extraction from the co-occurrence matrix. A two-order wavelet decomposition was performed, the joint probability density of each sub-band's difference coefficients (with adjoining coefficients in the horizontal, vertical, and diagonal directions) is calculated and the entropy and energy were extracted from the joint probability density matrix as features. Then, the image was decomposed into three sub-bands, and the Probability Density Function (PDF) was extracted from each sub-band's wavelet coefficient. Finally, these three kind features were combined to detect the image. A SVM was utilized as a classifier.

Ghanbari et al. [104] proposed a new algorithm for steganalysis using GLCM matrix properties. They used a combined method of steganography based on both location and conversion to hide the information in the original image and called it image-steg1 image. Then, they hided the information in image-steg1 again and called it image-steg2. Using GLCM matrix properties, they discovered some different features in the GLCM of the original image and stego images. These features were extracted and used for training a multilayer perceptron (MLP) neural network.

Zhang et al. proposed [105] a universal steganalysis method for jpeg images based on sparse representation. Sparse representation, is to convey the main body of information with as little information as possible, thus simplifying the solving process of information processing. This method has high detection accuracy and overcomes the "over-fitting" problem of traditional classifiers.

Devi et al. [106] presented four different steganalysis techniques applicable when binary images (black and white) are used as a cover image. Their method improved steganalysis techniques by minimizing image-to image variations. Image-to-image variation is defined as the difference between the underlying statistic of one image and that of another. They estimated the cover image from the stego image, then they computed the difference between the two to minimize the image-to image variation and finally they extracted the feature set from this difference. Their method can detect the stego object and estimate the length of the embedded message.

Verma [107] used a multilayer perceptron with backpropagation algorithm as a model for image classification. Moreover, he used Pre-processed Vectors Diagonal Back Propagation Algorithm (PVDBPA) to perform the operations which can detect the presence of hidden message. Furthermore, BMP steganalysis using Gray Level Co-Occurrence Matrix has been examined by using feature vectors and analyzed them through Euclidean distance which was taken as a measure.

In [108] Lu et al. proposed a steganalytic feature selection method based on the Fisher criterion, in which the separability of single-dimension and multiple dimension features, combined with measurement of the Euclidean distance, is analyzed. The proposed method has been used to analyze the features (in spatial and frequency domain) and select feature components to reduce the dimensionality. Experimental results showed that the proposed method can effectively reduce the feature dimension and also improve the steganalytic efficiency.

Tang et al. [109] proposed an adaptive steganalytic scheme based on embedding probabilities of pixels. Six different embedding rates (0.05–0.5 bpp) to images from BOSSBase image set, were tested. Experiments evaluated on four typical adaptive steganographic methods, have shown the effectiveness of the proposed scheme, especially for low embedding rates, for example, lower than 0.20 bpp.

Qian et al. [110] were the first to introduce convolutional neural networks (CNN) in order to detect the existence of steganographic content. The proposed model can capture the complex dependencies that are useful for steganalysis. Compared with other existing methods, this model can automatically learn feature representations with several convolutional layers. The feature extraction and classification steps are unified under a single architecture, which means the guidance of classification can be used during the feature extraction step. To evaluate the effectiveness of the developed model for steganalysis, authors conducted experiments on three spatial domain steganographic algorithms on various payloads (0.3–0.5 bpp). Results compared to other state-of the-art steganalysis methods were slightly worse.

Desai et al. [111] developed a reduced dimensional merged feature set for universal image steganalysis using Fisher Criterion and ANOVA techniques. Features were extracted from wavelet subbands and binary similarity patterns extracted from DCT of an image were merged to make a combined feature set. Fisher criterion and ANOVA test were then applied to evaluate the combined feature vector score and then only those features were selected which were found sensitive in both feature selection methods. The reduced 15-dimensional feature vector was utilized to train a SVM classifier with RBF kernel. The proposed algorithm was tested against various steganography methods at different embedding rates. Stego images were generated using state of the art steganographic algorithms and two standard image databases: CorelDraw [112] and BSDS500 [113]. A 10-fold cross validation process was performed. The proposed algorithm achieved overall 97% detection accuracy against various steganography methods.

Couchot et al. [114] proposed an architecture which embeds less convolutions, with much larger filters in the final convolutional layer. This approach is more general; therefore, it is able to deal with larger images and lower payloads. For a payload of 0.4 bpp the proposed CNN can detect stego images with an accuracy higher than 98%, whatever the steganographic algorithm chosen among three state-of-the-art, while it falls at most to 73.30% for the payload of 0.1.

Sajedi [115] proposed a method to discover special patterns that a steganography algorithm embeds in an image, the so-called Steganography Pattern Discovery (SPD). An evolutionary method was utilized to extract the signature of stego images against clean images via fuzzy if–then rules. Then, a SVM classifier was employed to detect stego images with high accuracy. Embedding rate was 0.05–0.4 bpp and the average accuracy on different steganographic methods varied from 79% to 91%.

Rostami et al. [116] proposed a feature selection method based on based on optimization process of Particle Swarm Optimization (PSO). In order to improve the performance of the method, the proposed PSO is used along with the measure of Area Under the receiver operating characteristics Curve (AUC) as the fitness function. Experimental results of the proposed method on BOSSBase image set showed that even that PSO method leads to a higher feature vector, it outperforms other state of the art feature selection approaches as the classification accuracy is higher. Moreover, the embedding rate in the dataset was 0.4bpp and the classification accuracy reached 82.62% when a SVM was utilized as classifier.

Wu et al. [117] proposed a very deep CNN model, the deep residual network (DRN). DRN model usually has a large number of network layers, which proves to be effective to capture the complex statistics of digital images. Furthermore, DRN's residual learning (ResL) method actively strengthens the signal coming from secret messages, which is extremely beneficial for the discrimination between cover images and stego images. Experiments on BOSSBase dataset (embedding rate 0.4bpp) showed that the DRN model achieves low detection error rates – 6.48% in average - for the state of the art steganographic algorithms, and outperforms the classical rich model method and several recently proposed CNN based methods.

Ye et al. [118] proposed a CNN based steganalyzer. The proposed CNN had different structure compared to the ones designed for computer vision tasks. Rather than a random strategy, the weights in first layer of the CNN are initialized with the basic high-pass filter set used in calculation of residual maps in Spatial Rich Model (SRM). Furthermore, a new activation function called truncated linear unit (TLU) was adopted in the model. Finally, the performance of the CNN based steganalyzer was boosted by incorporating the knowledge of selection channel. This approach proved capable of detecting several state-of-the-art steganographic schemes in spatial domain for a wide variety of payloads (0.05–0.5 bpp) with high accuracy.

Finally, Nouri et al. [119] proposed a scheme in which the alteration of singular value curve was used to construct the steganalysis feature vector. Two spatial and JPEG based feature vectors were extracted in the proposed statistical exploitation. Experimental results on images from two datasets, embedded with relative payloads of 0.05, 0.1, 0.2 and 0.4 bpp showed the acceptable performance of the proposed feature vectors for both universal and JPEG based steganalysis methods.

9. Discussion and conclusion

In this review we provided a detailed report of various methods proposed for steganalysis applicable to digital images. Coming to a conclusion, about which method is more effective – in any domain- is not an easy task. There are many parameters that a digital forensic examiner must know in advance, in order to give a safe answer before deciding which method to employ. These parameters include the existence or not of the cover image, the prior knowledge of the embedded data, findings of steganography software in a suspect's computer etc. However, if we assume that in the majority of the cases only the stego object is known we can say that statistical steganalysis techniques - in any domain- are more robust and more effective than signature steganalysis. This is met for both specific and universal statistical steganalysis. In specific statistical steganalysis the proposed methods focus to the embedding procedure and attempt to find image features or statistics changed by the embedding algorithm. Thus, this steganalytic approach has excellent accuracy only when performed on the specific steganographic algorithm, but even a small change in the embedding algorithm usually results to low steganalytic accuracy. For this reason, universal statistical steganalysis is used. These methods can detect hidden messages regardless the steganographic technique that were embedded to the digital image. Typically, classification is performed based on extracted features that are dependent to a widespread diversity of embedding procedures. These methods provide less accurate results than specific statistical steganalysis methods, but they can detect unseen steganographic content. Moreover, they are more flexible than the specific ones and slight changes to classification schemes may lead to the detection of more embedding algorithms as well. In Table 6 we summarize the papers, regarding the dataset the authors used during their research. The names of the datasets are mainly abbreviations. Their full names along with their download links can be found in Appendix.

Table 6
Number of papers that datasets were used.

Image dataset	Number of papers found	Publicly available found
BOSSBase (all versions)	22	Yes
Corel	8	Yes
NRCS	7	Yes
UCID	5	Yes
USC	5	Yes
BOWS – BOWS2	4	Yes
Philip Greenspun	4	Yes
BSDS	2	Yes
CBIR	2	Yes
Kodak	2	Yes
Other	Rest of papers	Some

Observing Table 6 we can reach some useful conclusions:

- The majority of the authors choose to use publicly available datasets as benchmark. This makes the comparison between similar methods easier and the reader can determine the value that the proposed method contributes in steganalysis.
- BOSSBase [43] is by far the most utilized dataset.

Considering all this, we can tell that the most effective methods dealing with LSB embedding are the methods given in [20,24]–[26], [28,30,32,34,35]. Those methods concerning LSB matching are [51,52,54]. Methods discussed in [69,73,74] are suitable for attacking SSIS and methods in [75,80,85] are capable to detect transform domain steganography. Finally, for universal statistical steganalysis the most promising and effective techniques are [69,96,98,108,110,114] [115], [117,118]. The criterion to reach to this conclusion of most effective steganalysis techniques is the relative study conducted by the researchers in their papers.

The ultimate – yet unreachable - goal for a steganalyst, is to employ a steganalysis technique that could detect any type of steganographic embedding algorithm with low computational needs and excellent accuracy. We strongly believe that universal steganalysis combined with deep learning techniques will boost research and will provide digital forensic examiners new software tools to uncover seen of the unseen.

Appendix

Dataset links

1	Philip Greenspun	http://philip.greenspun.com/
2	KODAK	ftp://ftp.kodak.com/www/images/pcd/
3	Noname	http://www.petitcolas.net/fabien/watermarking/benchmark/image_database.html
4	USC-SIP1 Image database	http://sipi.usc.edu/services/database/Database.html
5	CBIR Image Database	http://www.cs.washington.edu/research/imagedatabase/groundtruth/
6	NRCS	http://photogallery.nrcs.usda.gov/
7	BOSSBase	http://agents.fel.cvut.cz/boss/index.php?mode=VIEW&tmpl=materials
8	Noname	http://vision.ece.ucsb.edu/∼sullivak/Research_imgs/
9	Corel	http://www.corel.com
10	BOWS 2	http://bows2.ec-lille.fr/
11	UCID	http://jasoncantarella.com/downloads/ http://vision.doc.ntu.ac.uk/
12	Kodak photo cd	http://sqez.home.att.net/thumbs/Thumbnails.html
13	Noname	http://www.cs.nmt.edu/∼IA/steganalysis.html
14	INRIA	http://lear.inrialpes.fr/∼jegou/data.php
15	BSDS	http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/fg
16	ImageNet	http://www.image-net.org/
17	BSDS500	https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/
18	Raise	http://mmlab.science.unitn.it/RAISE/
19	Washington University image database	http://imagedatabase.cs.washington.edu/

References

[1] Koops BJ, Crypto law survey, 2013. [Online]. Available: http://www.cryptolaw.org/cls2.htm#busigov, Version 27.0, February 2013, [Accessed: 07 December 2017].

[2] Palmer G, A road map for digital forensic research, 2001.

[3] Katzenbeisser S, Petitcolas FAP. Information hiding techniques for steganography and digital watermarking. Artech House; 2000.

[4] Nissar A, Mir AH. Classification of steganalysis techniques: a study. Digital Signal Processing 2010;20(6):1758–70.

[5] Chanu YJ, Singh KM, Tuithung T. Image steganography and steganalysis: a survey. International Journal of Computing Applications 2012;52(2):975–8887.

[6] Li B, He J, Huang J, Qing Shi Y. A survey on image steganography and steganalysis. Journal of Information Hiding and Multimedia Signal Processing 2011;2(2):142–72.

[7] Chandramouli R, Subbalakshmi KP. Current trends in steganalysis: a critical survey. In: ICARCV 2004 8th control, automation, robotics and vision conference, 2; 2004. p. 964–7.

[8] Pal P, Dubey S. Various JPEG image steganography techniques: a review. International Journal of Scientific & Engineering Research 2016;7(2):417–21.

[9] Luo XY, Wang DS, Wang P, Liu FL. A review on blind detection for image steganography. In: Signal processing, 88. Elsevier; 2008. p. 2138–57. 01-Sep.

[10] Breaking a steganography software: camouflage. [Online]. Available: http://www.guillermito2.net/stegano/camouflage/index.html. [Accessed: 07 December 2017].

[11] Breaking a steganography software: JpegX. [Online]. Available: http://www.guillermito2.net/stegano/jpegx/index.html. [Accessed: 07 December 2017].

[12] Analyzing steganography softwares.[Online]. Available: http://www.guillermito2.net/stegano/. [Accessed: 28 November 2017].

[13] Hide files and folders - Masker 7.5. [Online]. Available: http://www.softpuls.com/masker/. [Accessed: 28 November 2017].

[14] Fridrich J, Goljan M. “Practical steganalysis of digital images: state of the art”. In: Proc. SPIE, 4675; Apr. 2002. p. 1–13.

[15] Fridrich J, Goljan M, Du R. Steganalysis based on JPEG compatibility. In: Proc. SPIE, 4518; Nov. 2001. p. 275–80.

[16] Westfeld A, Pfitzmann A. Attacks on steganographic systems. Berlin Heidelberg: Springer; 2000. p. 61–76.

[17] Westfeld A. Detecting low embedding rates. Berlin Heidelberg: Springer; 2003. p. 324–39.

[18] Chandramouli R, Kharrazi M, Memon N. Image steganography and steganalysis: concepts and practice. Berlin Heidelberg: Springer; 2004. p. 35–49.

[19] Fridrich J, Long M. Steganalysis of LSB encoding in color images. In: IEEE international conference on multimedia and Expo. ICME2000. Proceedings of the latest advances in the fast changing world of multimedia (Cat. No.00TH8532), Vol. 3; 2000. p. 1279–82.

[20] Fridrich J, Goljan M, Du R. Reliable detection of LSB steganography in color and grayscale images. In: Proceedings of the 2001 workshop on multimedia and security new challenges - (MM&Sec '01); 2001. p. 27.

[21] Avcibas I, Memon ND, Sankur B. “Steganalysis of watermarking techniques using image quality metrics”. In: Proc. SPIE - International Society for Optics and Photonics, 4314; Aug. 2001. p. 523–31.

[22] Lyu S, Farid H. Detecting hidden messages using higher-order statistics and support vector machines. Information Hiding 2003;2578:340–54.

[23] Dumitrescu S, Wu X, Wang Z. Detection of LSB steganography via sample pair analysis. IEEE Transactions on Signal Processing 2003;51(7):1995–2007.

[24] Dumitrescu S, Wu X, Memon N, On steganalysis of random LSB embedding in continuous-tone images, In: Proceedings of the international conference on image processing, vol. 1, pp. 641–44.

[25] Dumitrescu S, Wu X, Steganalysis of LSB embedding in multimedia signals, In: Proceedings of the IEEE international conference on multimedia and expo, pp. 581–84.

[26] Roue B, Bas P, Chassery J-M. Improving LSB steganalysis using marginal and joint probabilistic distributions. In: Proceedings of the 2004 multimedia and security workshop on multimedia and security - MM&Sec ’04; 2004. p. 75.

[27] Lu P, Luo X, Tang Q, Shen L. An improved sample pairs method for detection of LSB embedding. Berlin Heidelberg: Springer; 2004. p. 116–27.

[28] Avcıbaş İ, Kharrazi M, Memon N, Sankur B. Image steganalysis with binary similarity measures. EURASIP Journal on Advances in Signal Processing 2005;2005(17):2749–57.

[29] Dumitrescu S, Wu X. A new framework of LSB steganalysis of digital media. IEEE Transactions on Signal Processing 2005;53(10):3936–47.

[30] Zhi L, Fen SA, Xian YY. A LSB steganography detection algorithm. In: *14th IEEE proceedings on personal, indoor and mobile radio* communications (PIMRC 2003); 2003. p. 2780–3.

[31] Zhang T, Ping X. Reliable detection of LSB steganography based on the difference image histogram. In: Proceedings of the IEEE international conference on acoustics, speech, and signal processing, (ICASSP ’03), Vol. 3; 2003 III-545-8.

[32] Fridrich J, Goljan M, Soukal D. “Higher-order statistical steganalysis of palette images”. In: Proc. SPIE - International Society for Optics and Photonics, 5020; Jun. 2003. p. 178–90.

[33] Ker AD. “Quantitative evaluation of pairs and RS steganalysis”. In: Proc. SPIE - International Society for Optics and Photonics; Jun. 2004. p. 83–97.

[34] Ker AD. Improved detection of LSB steganography in grayscale images. Berlin Heidelberg: Springer; 2004. p. 97–115.

[35] Celik MU, Sharma G, Tekalp AM. “Universal image steganalysis using rate-distortion curves”. In: Proc. SPIE - International Society for Optics and Photonics; Jun. 2004. p. 467–76.

[36] Benton R, Chu H. Soft computing approach to steganalysis of LSB embedding in digital images. In: ITRE 2005 3rd international conference on information technology: research and education; 2005. p. 105–9.

[37] Fridrich J, Goljan M. “On estimation of secret message length in LSB steganography in spatial domain”. In: Proc. SPIE, 5306; Jun. 2004. p. 23–34.

[38] Ker AD, Rainer B. “Revisiting weighted stego-image steganalysis”. In: Proc. SPIE, 6819; Mar. 2008. p. 1–17.

[39] Chen X-d, Sun F, Sun W. Detect LSB steganography with bit plane randomness tests. In: 2006 6th world congress on intelligent control and automation; 2006. p. 10306–9.

[40] Bhattacharyya S, Sanyal G. Steganalysis of LSB image steganography using multiple regression and auto regressive (AR) model. International Journal of Computer Technology and Applications 2011;2(4):1069–77.

[41] Kekre HB, Athawale AA, Patki SA. Steganalysis of LSB embedded images using gray level co- occurrence matrix images. International Journal of Image Processing (IJIP) 2011;5(1):36–45.

[42] Fillatre L. Adaptive steganalysis of least significant bit replacement in grayscale natural images. IEEE Transactions on Signal Processing 2012;60(2):556–69.

[43] BOSS web page. [Online]. Available: http://agents.fel.cvut.cz/boss/index.php?mode=VIEW&tmpl=materials. [Accessed: 28-Nov- 2017].

[44] Fridrich J, Kodovský J, Steganalysis of LSB replacement using parity-aware features, Lecture notes in computer science (including subseries on lecture notes in artificial intelligence lecture notes in bioinformatics (LNCS)), vol. 7692 , pp. 31–45, 2013.

[45] Verma S, Sood S, Ranade SK. Relevance of steganalysis using DIH on LSB stegnography. International Journal of Advanced Research in Computer Science and Software Engineering 2014;4(2):835–8.

[46] Sharp T. An implementation of key-based digital signal steganography. Fourth Information Hiding Workshop 2001;2137(9):13–26.

[47] Zou D, Shi YQ, Su W, Xuan G. Steganalysis based on Markov model of thresholded prediction-error image. In: 2006 IEEE international conference on multimedia and expo, ICME 2006 – Proceedings, 2006; 2006. p. 1365–8.

[48] Malekmohamadi H, Ghaemmaghami S. Steganalysis of LSB based image steganography using spatial and frequency domain features. In: IEEE international conference on multimedia and expo, (ICME 2009); 2009. p. 1744–7.

[49] Pevny T, Bas P, Fridrich J. Steganalysis by subtractive pixel adjacency matrix. Distribution 2010;5(2):215–24.

[50] Zhang T, Li W, Zhang Y, Zheng E, Ping X. Steganalysis of LSB matching based

on statistical modeling of pixel difference distributions. Information Sciences (Ny) 2010;180(23):4685–94.
[51] Fridrich J, Kodovský J, Holub V, Goljan M, Steganalysis of content-adaptive steganography in spatial domain, Lecture notes in computer science (including subseries on lecture notes in artificial intelligence lecture notes in bioinformatics (LNCS)), vol. 6958, pp. 102–17, 2011.
[52] Gul G, Kurugollu F, A new methodology in steganalysis: breaking highly undetectable steganograpy (HUGO), Lecture notes in computer science (including subseries on lecture notes in artificial intelligence lecture notes in bioinformatics (LNCS)), vol. 6958, pp. 71–84, 2011.
[53] Fridrich J, Kodovsky J. Rich models for steganalysis of digital images. IEEE Transactions on Information Forensics and Security 2012;7(3):868–82.
[54] Pevny T, Fridrich J, Ker AD. From blind to quantitative steganalysis. IEEE Transactions on Information Forensics and Security 2012;7(2):445–54.
[55] Cogranne R, Retraint F. An asymptotically uniformly most powerful test for LSB matching detection. IEEE Transactions on Information Forensics and Security 2013;8(3):464–76.
[56] BOWS-2 Web page. [Online]. Available: http://bows2.ec-lille.fr/. [Accessed: 28-Nov- 2017].
[57] Ker AD. Steganalysis of LSB matching in grayscale images. IEEE Signal Processing Letters 2005;12(6):441–4.
[58] Zhang J, Cox IJ, Doerr G. Steganalysis for LSB Matching in Images with High--frequency Noise. In: 2007 IEEE 9th Workshop on Multimedia Signal Processing; 2007. p. 385–8.
[59] Holub V, Fridrich J. Random projections of residuals for digital image steganalysis. IEEE Transactions on Information Forensics and Security 2013;8(12):1996–2006.
[60] Xia Z, Wang X, Sun X, Wang B. Steganalysis of least significant bit matching using multi-order differences. Secur Commun Netw 2014;7(8):1283–91.
[61] Xia Z, Wang X, Sun X, Liu Q, Xiong N. Steganalysis of LSB matching using differences between nonadjacent pixels. Multimed Tools and Applications 2016;75(4):1947–62.
[62] NRCS photo gallery home. [Online]. Available: https://photogallery.sc.egov.usda.gov/res/sites/photogallery/. [Accessed: 28 November 2017].
[63] Goljan M, Fridrich J, Cogranne R. Rich model for Steganalysis of color images. In: 2014 IEEE international workshop on information forensics and security (WIFS 2014); 2015. p. 185–90.
[64] Chen X, Gao G, Liu D, Zhihua X. Steganalysis of LSB matching using characteristic function moment of pixel differences. China Communications 2016;13(7):66–73.
[65] Lerch-Hostalot D, Megías D. Unsupervised steganalysis based on artificial training sets. Engineering Applications of Artifficial Intelligence 2016;50:45–59.
[66] Juarez-Sandoval O, Cedillo-Hernandez M, Sanchez-Perez G, Toscano-Medina K, Perez-Meana H, Nakano-Miyatake M. Compact image steganalysis for LSB-matching steganography. In: Proceedings - 2017 5th international workshop on biometrics and forensics, (IWBF 2017); 2017. p. 1–6.
[67] UCIDv2.0. [Online]. Available: http://jasoncantarella.com/downloads/. [Accessed: 30-Nov- 2017].
[68] Marvel LM, Boncelet CG, Methodology of spread-spectrum image steganography, 1998.
[69] Harmsen JJ, Pearlman WA. Steganalysis of additive-noise modelable information hiding. In: Security and watermarking of multimedia contents; 2003. p. 131–42.
[70] Chandramouli R, Subbalakshmi KP. Active steganalysis of spread spectrum image steganography. In: *Proceedings of the 2003 international symposium on c*ircuits and systems,(ISCAS '03), 3; 2003 III-830-III-833.
[71] Wang Y, Moulin P. Steganalysis of block-DCT image steganography. In: IEEE workshop on statistical signal processing; 2003. p. 339–42.
[72] Ji R, Yao H, Liu S, Wang L, Sun J. A new steganalysis method for adaptive spread spectrum steganography. In: 2006 international conference on intelligent information hiding and multimedia; 2006. p. 365–8.
[73] Sullivan K, Madhow U, Chandrasekaran S, Manjunath BS. Steganalysis of spread spectrum data hiding exploiting cover memory. In: Proc. SPIE 5681, 17th Annual Symposium on Electronic Imaging Science and Technology; Mar. 2005. p. 38–46.
[74] Li M, Kulhandjian MK, Pados DA, Batalama SN, Medley MJ. Extracting spread--spectrum hidden data from digital media. IEEE Transactions on Information Forensics and Security 2013;8(7):1201–10.
[75] Shaohui L, Hongxun Y, Wen G. Neural network based steganalysis in still images. Proceedings of the 2003 international conference on multimedia and expo.(ICME '03); 2003. (Cat. No.03TH8698), p. I-509..
[76] Liu S, Yao H, Gao W. Steganalysis of data hiding techniques in wavelet domain. In: Proceedings of the International conference on information technology: coding and computing, (ITCC 2004), Vol. 1; 2004. p. 751–4.
[77] Chen B, Wornell GW. Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. IEEE Transactions on Information Theory 2001;47(4):1423–43.
[78] Liu S, Yao H, Gao W, Steganalysis based on wavelet texture analysis and neural network, In: *Fifth world congress on intelligent control and automation (IEEE Cat. No.04EX788)*, pp. 4066–69.
[79] Sullivan K, Bi Z, Madhow U, Chandrosekaran S, Manjunath BS. Steganalysis of quantization index modulation data hiding. In: International conference on image processing, 2; 2004. p. 1165–8. *ICIP '04*.
[80] Shi YQ, Chen C, Chen W. A markov process based approach to effective attacking JPEG steganography. In: Information hiding. Berlin, Heidelberg: Springer Berlin Heidelberg; 2006. p. 249–64.
[81] Westfeld A. Generic adoption of spatial steganalysis to transformed domain. In: Information hiding. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 161–77.
[82] Kodovský J, Fridrich J. Quantitative steganalysis of LSB embedding in JPEG domain. In: Proceedings of the 12th ACM workshop on multimedia and security - (MM&Sec '10); 2010. p. 187.
[83] Liu Q, Sung AH, Qiao M, Chen Z, Ribeiro B. An improved approach to steganalysis of JPEG images. Information Sciences (Ny) 2010;180(9):1643–55.
[84] Sheikhan M, Moin MS, Pezhmanpour M. Blind image steganalysis via joint co-occurrence matrix and statistical moments of contourlet transform. In: 10th international conference on intelligent systems design and applications; 2010. p. 368–72.
[85] Kodovsky J, Fridrich J, Holub V. Ensemble classifiers for steganalysis of digital media. IEEE Transactions on Information Forensics and Security 2012;7(2):432–44.
[86] Westfeld A. F5 — a steganographic algorithm high capacity despite better steganalysis. In: 4th international workshop on information hiding, 2137; 2001. p. 289–302.
[87] K. Solanki, A. Sarkar, and B.S. Manjunath, YASS: yet another steganographic scheme that resists blind steganalysis, In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics LNCS), 2007, vol. 4567, pp. 16–31.
[88] Sallee P. Model-based steganography. In: Kalker T RYM, Cox I, editors. Digital watermarking, 2939. Berlin, Heidelberg: Springer; 2004. p. 154–67.
[89] Cho S, Cha B-H, Gawecki M, Jay Kuo C-C. Block-based image steganalysis: algorithm and performance evaluation. Journal of Visual Communication and Image Representation 2013;24(7):846–56.
[90] Sree Rathna Lakshmi NVS. A novel steganalytic algorithm based on III level DWT with energy as feature. Research Journal of Applied Sciences Engineering and Technology 2014;7(19):4100–5.
[91] Holub V, Fridrich J. Low-complexity features for JPEG steganalysis using undecimated DCT. IEEE Transactions on Information Forensics and Security 2015;10(2):219–28.
[92] Farid H, Detecting hidden messages using higher-order statistical models, In: *Proceedings of the international conference on image processing*, vol. 2, p. II–905–08.
[93] Lyu S, Farid H. Steganalysis using color wavelet statistics and one-class support vector machines. In: Proc. SPIE 5306; Jun. 2004. p. 35.
[94] Trivedi S, Chandramouli R. Active steganalysis of sequential steganography. In: Proc. SPIE 5020; Jun. 2003. p. 123–30.
[95] Lafferty P, Ahmed F. Texture-based steganalysis: results for color images. In: Proc. SPIE 5561; Oct. 2004. p. 145–51.
[96] Xuan G, et al.. Steganalysis based on multiple features formed by statistical moments of wavelet characteristic functions. Berlin Heidelberg: Springer; 2005. p. 262–77.
[97] Shi YQ, et al. Image steganalysis based on moments of characteristic functions using wavelet decomposition, prediction-error image, and neural network. In: IEEE international conference on multimedia expo, (ICME 2005), Vol. 2005; 2005. p. 269–72.
[98] Lie W-N, Lin G-S. A feature-based classification technique for blind image steganalysis. IEEE Transactions on Multimedia 2005;7(6):1007–20.
[99] Lyu S, Farid H. Steganalysis using higher-order image statistics. IEEE Transactions on Information Forensics and Security 2006;1(1):111–19.
[100] Chen X, Wang Y, Tan T, Guo L. Blind image steganalysis based on statistical analysis of empirical matrix. In: 18th international conference on pattern recognition (ICPR'06); 2006. p. 1107–10.
[101] Sun Z, Hui M, Guan C. Steganalysis based on co-occurrence matrix of differential image. In: 2008 international conference on intelligent information hiding and multimedia signal processing; 2008. p. 1097–100.
[102] Zhao H, Wang H, Khan MK. Steganalysis for palette-based images using generalized difference image and color correlogram. Signal Processing. 2011;91(11):2595–605.
[103] Zong H, Liu F, Luo X. Blind image steganalysis based on wavelet coefficient correlation. Digital Investigation 2012;9(1):58–68.
[104] Ghanbari S, Ghanbari S, Keshtegary M, Ghanbari N, Branch Z, Azad I. "New steganalysis method using glcm and neural network". International Journal of Computer Applications 2012;42(7):46–50.
[105] Zhang Z, Hu D, Yang Y, Su B. A universal digital image steganalysis method based on sparse representation. In: 2013 ninth international conference on computational intelligence and security; 2013. p. 437–41.
[106] Devi M, Sharma N. Improvements of steganography parameter in binary images and JPEG images against steganalysis. International Journal of Engineering Sciences and Research Technology 2013;2(8).
[107] Verma AK. A non- blind steganalysis through neural network approach. International Journal of Multidisciplinary Consortium 2014;1(1).
[108] Lu J, Liu F, Luo X. Selection of image features for steganalysis based on the Fisher criterion. Digital Investigion 2014;11(1):57–66.
[109] Tang W, Li H, Luo W, Huang J. Adaptive steganalysis based on embedding probabilities of pixels. IEEE Transactions on Information Forensics and Security 2016;11(4):734–44.
[110] Qian Y, Dong J, Wang W, Tan T, Deep learning for steganalysis via convolutional neural networks, 2015, p. 94090J.
[111] Desai MB, Patel SV, Prajapati B. ANOVA and fisher criterion based feature se-

lection for lower dimensional universal image steganalysis. International Journal of Image Processing 2016;10(3):145–60.

[112] CorelDraw image database. [Online]. Available: http://www.corel.com/. [Accessed: 30-Nov- 2017].

[113] UC Berkeley Computer Vision Group. [Online]. Available: https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/. [Accessed: 30-Nov- 2017].

[114] Couchot JF, Couturier R, Guyeux C, Salomon M, Steganalysis via a convolutional neural network using large convolution filters, pp. 1–24, May 2016.

[115] Sajedi H. Steganalysis based on steganography pattern discovery. Journal of Information Security and Applications 2016;30:3–14.

[116] Rostami V, Khiavi AS. Particle Swarm Optimization based feature selection with novel fitness function for image steganalysis. In: Artificial intelligence and robotics,(IRANOPEN 2016); 2016. p. 109–14.

[117] Wu S, Zhong S, Liu Y. Deep residual learning for image steganalysis. In: Multimedia tools and applications. US: Springer; 2017. p. 1–17. 15-Feb-.

[118] Ye J, Ni J, Yi Y. Deep learning hierarchical representations for image steganalysis. IEEE Transactions on Information Forensics and Security 2017;12(11):2545–57.

[119] Nouri R, Mansouri A. Digital image steganalysis based on the reciprocal singular value curve. Multimedia Tools and Applications 2017;76(6):8745–56.