Increase in level of detection of secret embedding into compressed JPEG-images based on selection of informative features

O Shumskaya¹²

¹ Laboratory of Big Data Technologies of Sociocyberphysical Systems, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, 39 14-th Line, VI, St. Petersburg, 199178, Russia
² E-mail: shumskaya.oo@gmail.com

Abstract. The research of informational content of features in the spatial and the frequency domains at steganalysis of the compressed JPEG-images is presented. Modern steganalysis methods, based on which the feature set has been created, are considered. As a method of informative features selection has been applied the greedy algorithm with an exception which has allowed reducing number of features in the set from 43 to 26. The experiments have shown that steganalysis on the basis of informative features allows increasing the accuracy of correct classification of initial images to 94%, of images with attachments by the PM1 and the Jsteg methods by 25 and 18 percent respectively. The issue of steganalysis of images with artificially created distortions is mentioned. The experiments showed that steganalysis based on informative features allow classifying correctly such images with the accuracy 96%.

1. Introduction
It is difficult to present today any field of human activity where digital images are not used: photos, schemes, schedules, charts, drawings, etc. Every day users post into the Internet millions of images of various characters, save the pleasant pictures, without suspecting about their possible contents. Active flows of digital images in the Internet allow people imperceptibly exchanging the secret information by use of digital steganography. The authors in [1] explain steganography as the science about the ways of information transfer, storage, which provide concealment of existence of this information in some signal; it provides various methods of concealment of data in digital images.

The criteria of efficiency of steganographic embedding are capacity, obscurity (lack of visible distortions of digital objects – stego-containers) and stability before steganalysis – the science about ways of identification of the hidden messages’ existence facts in digital objects. Steganalysis in most cases consists in search of characteristics (features) of a digital object, which change during steganographic embedding.

Identification of the unmasking features can be made by means of the analysis of images arrays (stego-images and images which aren't containing secret information). However there are tens of thousands of various features of digital images in the spatial and the frequency domains. Many of features are linearly dependent, and some do not bear information about distortions, which were made during steganographic embedding. As features in a set are considered in interaction, it is important to reveal a set of informative features.
Steganography methods are divided to embedding into spatial domain and into frequency one. However, steganalysis methods can be based on features in frequency domain, in spatial one and there are methods based on a mixed feature set.

2. Steganalitical Features of Digital Images

Steganalysis methods are divided into methods based on features in the frequency domain and methods based on features in the spatial domain of the image. Methods with combined sets, which include features both from the spatial domain of the image and from the frequency domain, appear more often. In each steganalysis method, a vector with which a qualifier works is formed of the received values of features. Before beginning classification, it is necessary to provide training: the qualifier defines what intervals of features’ values belong to the initial images and to the stego-images. Comparing the data obtained during tutoring with the calculated vector of features of the image, the qualifier defines contents of the image.

Let us consider some features used in the modern steganalysis methods.

In the work [2] the author considers embedding to the quantized DCT-coefficients area of the image. The feature set includes features in the frequency domain, based on ratios between the energy collected in separate frequency coefficients of the DCT-spectrum. Author explains the choice of such feature set with the fact that these values concentrate in themselves the maximum information on the image’s internal contents.

Let $F_1 = E(f_{[0]}^{1\ldots n})^{-1} E(f_0)$, $F_2 = E(f_{[0]}^{1\ldots n})^{-1} \sum_{[n]} E(f_n)$, $F_3 = E_{[\parallel]}^{-1} E_{[\parallel]}^{-1}$, $F_4 = E_{[\parallel]}^{-1} E_{[\parallel]}^{-1}$, $F_5 = E_{[\parallel]}^{-1} E_{[\parallel]}^{-1}$, where $E(f_0)$ – mean value of zero AC-coefficients frequencies at blocks; $E(f_{[0]}^{1\ldots n})$ – mean value of those AC-coefficients, absolute value of which is equal to 1; $E_{[\parallel]}$ – energy of those AC-coefficients, absolute value of which is $>1$.

Research showed that increase in values $E(f_0), \sum_{[n]} E(f_n), E_{[\parallel]}^{-1}$, and reduction of values $E(f_{[0]}^{1\ldots n})$, $E_{[\parallel]}^{-1}$ is characteristic of additive embedding of information into the quantized DCT-coefficients of the JPEG-image.

Embedding into digital images is carried out sequentially in each block of the image. The interblock correlation is characteristic of adjacent blocks. Changes, made to image’s blocks during embedding can lead to disruption of bonds in between:

$$F_4 = \left(\sum_{i=1}^{\frac{N}{2}} (x_i - \bar{x})^2 \sum_{j=1}^{\frac{N}{2}} (y_j - \bar{y})^2\right)^{1/2} \sum_{i=1}^{\frac{N}{2}} (x_i - \bar{x}) (y_j - \bar{y}),$$

(2)

where $\bar{x}, \bar{y}$ – mean values of adjacent blocks’ AC-coefficients; $x_i, y_i$ – $i$-th AC-coefficient of adjacent blocks.

In papers [3-4] is considered the set, which combine features, both in the frequency domain, and in the spatial domain as a steganalysis basis:

- general histogram of DCT-coefficients;
- individual histograms of the first five AC-coefficients, for everyone separately;
- interblock dependences diversely (features in the spatial domain);
- dual histograms of AC-coefficients which value is in the range $[-5, 5]$, for everyone the separate histogram which represents the matrix reflecting on what place, how many times totally in all blocks the coefficient met a certain value:

$$f_5, \ldots, f_{15} = \left\| \sum_{k=1}^{N} \delta(d, d_k(i, j)) \right\|_{L_1}^{1} \left\| \sum_{k=1}^{N} \delta(d, d_k(i, j)) \right\|_{L_1},$$

(3)
where \( d \) – the fixed value of coefficient, \( d \in [-5, 5] \); \( B \) – blocks number in the image; \( i, j \) – coordinates of coefficient’s position in the block; \( \delta(u, v) = \begin{cases} 1, & u = v \\ 0, & \text{else} \end{cases} \); \( L_1 \) norm – maximum of the elements’ amounts at columns.

The authors calculate each characteristic twice: for the researched image (\( J_1 \)) and for the image which is received by truncating of the researched image on top and at the left by 4 pixels (\( J_2 \)). The authors claim that when truncating the image at the left and on top, the division of the image into blocks moves, coefficients of the discrete cosine transform are exempted from influence of last quantization and contain only statistical data of the image, which are just important in steganalysis case.

Thus, the value of the functionality will be finite feature:

\[
F_{5, \ldots, 15} = \| f_{5, \ldots, 15}(J_1) - f_{5, \ldots, 15}(J_2) \|_i,
\]

(4)

| Feature       | Formula |
|---------------|---------|
| spatial energy| \( F_{16} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{i,j}^2 \), where \( N \) – brightness gradation of stego-image; \( P \) – connectivity matrix. |
| entropy       | \( F_{17} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{i,j} \log P_{i,j} \) |
| uniformity    | \( F_{18} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \left(1 + |i + j| \right)^{-1} P_{i,j} \) |
| contrast      | \( F_{19} = \sum_{n=0}^{N-1} n^2 \left[ \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{i,j} \right] \cdot |i - j| = n \) |
| i-average     | \( F_{20} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} iP_{i,j} \) |
| j-average     | \( F_{21} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} jP_{i,j} \) |
| i-dispersion  | \( F_{22} = \sum_{i=1}^{N} \sum_{j=1}^{N} (i - F_{20})^2 P_{i,j} \) |
| j-dispersion  | \( F_{23} = \sum_{i=1}^{N} \sum_{j=1}^{N} (j - F_{21})^2 P_{i,j} \) |
| covariance    | \( F_{24} = \sum_{i=1}^{N} \sum_{j=1}^{N} (i - F_{20})(j - F_{21}) P_{i,j} \) |
| correlation   | \( F_{25} = \sum_{i=1}^{N} \sum_{j=1}^{N} F_{22}^{-1} (i - F_{20})(j - F_{21}) P_{i,j} \) |

For the solution of the steganalysis task it is possible to use not only the features created especially for its carrying out but also features representing some characteristics calculated for digital images. In particular, it can be the textural features applied in different tasks of image identification. 10 features given in [5] were included in the researched feature set as \( F_{16}, \ldots, F_{25} \).

The method provided in [6] is based on the Benford’s law: the probability of appearance of digit on the first place in number is higher, than less is this digit. Based on outputs of the paper devoted to the
research of justice of the Benford’s law concerning DCT-coefficients of JPEG-images before and after quantization, authors offered a special case of the Benford’s law as quantized DCT-coefficients do not submit strictly to the Benford's law.

Features based on the Benford’s law for each of nine digits (5) and features integrating an essence of the Benford’s law and the idea of image shift by 4 pixels (6) from [3-4] were included into the feature set.

\[ F_{26, \ldots, 34} = N \log_{10} \left( 1 + \left( s + x^q \right)^{-1} \right), \]

where \( x = 1, \ldots, 9; N, s, q \) – parameters depending on quality of JPEG-compression.

\[ F_{35, \ldots, 43} = \| F_{26, \ldots, 34} (J_1) - F_{26, \ldots, 34} (J_2) \|_q. \]

Generally, digital objects steganalysis is considered as the task of two-class classification when for each analyzable object one of two outcomes is selected: it is an initial object or an object contains some hidden data.

3. Selection of Informative Features

Nowadays there is a wide choice of algorithms of informative features selection, the cornerstone of which is statistical dependences (between set elements, between elements of set and output values), comparative experiments and difficult calculations.

For research of the created earlier feature set the most suitable is the greedy algorithm with an exception as this method fast, effective according to [7-8], considers all elements of the set in total, but not as separate components.

Greedy algorithm of search is used often as it is fast and yields good result in many tasks. The group of algorithms has received such name because if one of features has been chosen in subset (or excluded), then further it remains in set (in case of the greedy inclusion) or will be forever removed (in case of the greedy exception) [9].

For formation of learning and test selections standard images for the steganography (Lena, Peppers, Baboon, Splash, Airplane, Tiffany, Girl, etc.) and images from the base UCID (Uncompressed Color Image Database) and from the base USC-SIPI ID (University of Southern California Signal and Image Processing Institute Image Database) 256x256 pixels were taken. Training (660) and test (440) selections contain images:

- without attachment;
- with attachment by classical steganographic methods: F5, PM1, Jsteg.

As in research the greedy algorithm with an exception is applied, the full feature set with 43 elements described earlier is initially considered.

In the Tab. 2 results of experiments are provided. As the feature set is quite volume, the table contains five greatest probabilities of detection on each step.

Experiments have shown that the set of 43 features in the spatial and the frequency domains can be reduced by 17 elements. It allows increasing the general accuracy of classification by 19%.

It should be noted that 7 remote features on the basis of the Benford’s law were removed quickly enough (remained only for values 3 and 4). Features based on combination of the Benford’s law with the idea of the image shift were steadier and was removed only 1 feature (for value 1). It allows making conclusion on bigger informational content and a bigger contribution of the modified version of features.

In addition during the several first iterations the big accuracy of detection was shown by sets without structural features in the spatial domain (correlation, entropy, contrast, covariance). However after removal of these 4 features in the Tab. 2 presence of other elements from this group has not been noticed almost.
4. Computing Experiments

In case of steganalysis image’s features in the spatial and the frequency domains are analyzed. However, distortions can be made in digital objects not only at steganographic embedding. Now there is a set of applications for editing images: imposing of ready filters and creation of various effects, including noise. When processing the image in an editor distortions are also made into the main features of the image. At steganalysis it is important not to classify artificially distorted image as the image with an attachment [10].

The experiments with the initial feature set and with the set of selected informative features were made separately for 5 types of images:

- image without attachment (initial);
- images with attachments by the steganographic methods: F5, PM1, Jsteg;
- images without attachment but distorted in the popular photo editor Prisma.

| Step | Place 1 | Place 2 | Place 3 | Place 4 | Place 5 |
|------|---------|---------|---------|---------|---------|
| 1    | F25 (76.32) | F33 (76.19) | F24 (74.60) | F35 (74.60) | F1 (73.02) |
| 2    | F17 (77.78) | F4 (76.19) | F18 (76.19) | F43 (71.43) | F19 (69.84) |
| 3    | F50 (77.78) | F27 (73.02) | F10 (71.43) | F5 (69.84) | F32 (69.84) |
| 4    | F10 (77.78) | F28 (77.78) | F35 (74.60) | F2 (68.25) | F16 (66.67) |
| 5    | F26 (79.37) | F11 (77.78) | F28 (76.19) | F33 (74.60) | F40 (74.60) |
| 6    | F24 (79.37) | F5 (74.60) | F37 (74.60) | F27 (69.84) | F35 (68.25) |
| 7    | F27 (80.95) | F14 (79.37) | F12 (79.37) | F3 (73.02) | F43 (73.02) |
| 8    | F31 (84.13) | F36 (84.13) | F15 (82.54) | F23 (82.54) | F40 (82.54) |
| 9    | F1 (84.13) | F10 (84.13) | F13 (84.13) | F16 (84.13) | F5 (82.54) |
| 10   | F15 (84.13) | F8 (82.56) | F14 (82.56) | F32 (82.54) | F38 (82.54) |
| 11   | F15 (84.13) | F36 (82.56) | F1 (82.56) | F32 (82.54) | F9 (78.77) |
| 12   | F14 (84.13) | F36 (82.56) | F1 (82.54) | F21 (80.95) | F12 (80.95) |
| 13   | F13 (84.13) | F10 (82.56) | F2 (82.54) | F37 (82.54) | F53 (76.19) |
| 14   | F14 (84.13) | F40 (84.13) | F16 (82.54) | F2 (79.37) | F43 (77.78) |
| 15   | F5 (84.13) | F1 (82.54) | F12 (82.54) | F3 (79.37) | F39 (74.60) |
| 16   | F33 (85.71) | F40 (85.71) | F1 (84.13) | F12 (84.13) | F18 (80.95) |
| 17   | F34 (87.30) | F38 (84.13) | F1 (82.54) | F1 (80.95) | F5 (79.37) |
| 18   | F42 (85.71) | F5 (84.13) | F23 (84.13) | F36 (82.54) | F12 (77.78) |

Remained features removed from the full set – five dual histograms. Paying attention for what values of coefficients there were these features: -4, -3, 3, 4, 5. It is known that after the DCT and quantization the greatest number of elements is accepted by values in the range [-1, 1], and the farther the value is from borders of this range, the less is the probability to meet it in the matrix of DCT-coefficients. Therefore, dual histograms for the above-stated values were removed from the general feature set.

The sets with an exception of the first 4 features: 3 on the basis of ratios of energies in the frequency domain and interblock correlation between blocks with proximity of DC-coefficients no more than 3, – always showed result below an average. It means that these features have an important contribution to the general set and an informative component about contents of the digital image.

As a result of experiments the following set of 26 informative features has been defined:

\[ \{ F_1 - F_5, F_9 - F_{12}, F_{16}, F_{18}, F_{20} - F_{23}, F_{26}, F_{29}, F_{36} - F_{43} \} . \]
6

Figure 1. Example of images: (а) – clear image; (b) – F5; (c) – PM1; (d) – Jsteg; (e) – Prisma.

Table 3. Comparison of full set with set of informative features.

| Type of image content | Full set (43 features) | Set of informative features (25 features) |
|-----------------------|------------------------|-------------------------------------------|
| Clear                 | 90.91                  | 94.44                                     |
| F5                    | 80.65                  | 80.65                                     |
| PM1                   | 56.25                  | 81.25                                     |
| Jsteg                 | 61.70                  | 80.65                                     |
| Prisma                | 87.04                  | 96.30                                     |

The experimental results show that reduction of the feature set to 26 informative features allows increasing the volume of correct classification of images. For images with attachment by the PM1 and the Jsteg methods application of the reduced feature set show increase in accuracy of correct classification by 25 and 18.95 percent respectively. The accuracy of classification of images with artificial distortions (Prisma) increased by 9.26%, and clear images (initial) – by 3.53%. Application of the reduced feature set has not exerted impact only on images with attachment by the F5 method. Generally the accuracy of classification has increased on average by 11.34%.

5. Conclusion
Modern steganalysis methods and features in the spatial and the frequency domains of digital images, which are widely applied in steganalysis, were investigated. On the research’s basis the feature set was created.

The application of the greedy algorithm with an exception as the algorithm of informative features selection allowed reducing number of features in the set from 43 to 26.

Experiments have shown that steganalysis on the basis of informative features allows increasing the accuracy of the correct classification for images without attachment by 3.5%, with attachment by the PM1 and the Jsteg methods by 25 and 18.9% respectively.

The problem of steganalysis of images with the artificial distortions created by means of the photo editor Prisma was mentioned here. Experiments have shown that steganalysis based on informative features allows reaching the accuracy of correct classification 96%.

The conducted research shows that studying of informational content of features in the set allows achieving good results at identification of the fact of existence of confidential attachment in digital images.

References

[1] Konakhovich G F and Puzyrenko A Yu 2006 Computer steganography. Theory and Practice (Moscow: K-Press)
[2] Jia-Fa M, Xin-Xin N, Gang X, Wei-Guo Sh and Na-Na Zh 2016 A steganalysis method in the DCT domain Multimedia Tools and Applications 75 5999–6019
[3] Chen M-C 2009 Alpha-trimmed Image Estimation for JPEG Steganography Detection *Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics* (San Antonio: IEEE) pp 4581–4585

[4] Fridrich J 2004 Feature-Based Steganalysis for JPEG Images and its Implications for Future Design of Steganographic Schemes *Proceedings of the Sixth International Workshop on Information Hiding, Lecture Notes in Computer Science* (Toronto: Springer-Verlag Berlin Heidelberg) **3200** pp 67–81

[5] Andriotis P, Oikonomou G and Tryfonas T 2013 JPEG steganography detection with Benford’s Law *Digital Investigation* **9** 246–257

[6] Fu D, Shi Y Q and Su W 2007 A generalized Benford’s law for JPEG coefficients and its applications in image forensics *Proceedings of SPIE 6505, Security, Steganography, and Watermarking of Multimedia Contents IX* (SPIE) pp 1L1–1L11

[7] Berg G, Davidson I, Duan M-Y and Paul G 2003 Searching For Hidden Messages: Automatic Detection of Steganography *Proceedings of the 15th Innovative publications of Artificial Intelligence (IAAI) Conference* ed J Riedl, R Hill (Acapulco: AAAI) pp 51–56

[8] Maitra S, Paul G, Sarkar S, Lehmann M and Meier W 2013 New Results on Generalization of Roos-type Biases and Related Keystream of RC *Proceedings of the 6th International Conference on Cryptology in Africa (AFRICACRYPT)* ed M S Zediker **7918** pp 222–239

[9] Mitsel A A, Kolodnikova N V and Protasov K T 2005 Non-parametric algorithm for texture analysis of aerospace images *Bulletin of the Tomsk Polytechnic University* **308(1)** 65–70

[10] Shumskaya O O and Budkov V Yu 2018 Comparative study of classification methods in digital image analysis *Informatics, Computer Engineering and Control* **72(3)** 121-134