ADVANCES IN CYBER-PHYSICAL SYSTEMS

Vol. 6, Num. 2, 2021

DETECTION OF STEGO IMAGES WITH ADAPTIVELY EMBEDDED DATA BY COMPONENT ANALYSIS METHODS

Dmytro Progonov

Igor Sikorsky Kyiv Polytechnic Institute, 37, Prosp. Peremohy, Kyiv, 03056, Ukraine.
Authors' e-mail*: d.progonov@kpi.ua*

Submitted on 21.09.2021

***Abstract*: Ensuring the effective protection of personal and corporate sensitive data is topical task today. The special interest is taken at sensitive data leakage prevention during files transmission in communication systems. In most cases, these leakages are conducted by usage of advance adaptive steganographic methods. These methods are aimed at minimizing distortions of cover files, such as digital images, during data hiding that negatively impact on detection accuracy of formed stego images. For overcoming this shortcoming, it was proposed to pre-process (calibrate) analyzed images for increasing stego-to-cover ratio. The modern paradigm of image calibration is based on usage of enormous set of high-pass filters. However, selection of filter(s) that maximizes the probability of stego images detection is non-trivial task, especially in case of limited a prior knowledge about embedding methods. For solving this task, we proposed to use component analysis methods for image calibration, namely principal components analysis. Results of comparative analysis of novel maxSRMd2 cover rich model and proposed solution showed that principal component analysis allows increasing detection accuracy up to 1.5% even in the most difficult cases (low cover image payload and absence of cover-stego images pairs in training set).**

***Index Terms*: adaptive embedding methods, communication system security, digital images, steganalysis**

I. INTRODUCTION

Reliable protection of public and private confidential data is topical task today [1]. Particular attention is paid to the counteraction of covert data transmission, which is able to overcome existing intrusion detection systems (IDS). In this case, unauthorized transmission of sensitive data is provided by negligible alteration of cover files, for example digital images (DI), in order to embed messages (stego data) [2, 3].

The state-of-the-art methods for revealing of formed stego images are based on statistical analysis of DI with usage of rich models (e.g. SRM-based models [6]), convolutional neural networks, such as SR-Net [5] and Zhu-Net [7], deep autoencoders, for example ASSAF network [8]. Feature of these methods is pre-processing (calibration) of analyzed image for increasing stego-to-cover ratio. The calibration is aimed at suppression of cover's content or alterations caused by message hiding [9].

Modern calibration methods are based on suppression of CI content by applying of various high pass filters. Such filters are determined manually for maximization of detection accuracy for predefined set of embedding methods [10]. The task of selection optimal filters those are applicable for wide range of embedding methods is currently unsolved in general case, while near-optimal filters are proposed for specific steganographic methods only.

To overcome this limitation, we proposed to use component analysis methods. Such methods allow effectively un-mixing image's components under some common assumptions about theirs statistical features, such as variance, mutual information etc. The paper is devoted to investigation of image calibration performance in case of usage principal component analysis to extract and suppress image's components connected with message hiding.

II. RELATED WORKS

Message embedding into innocuous files and further transmission of altered (stego) files in communication network became one of widespread method for creating hidden communication channels between attackers in last years [1]. Negligible alterations of covers during data hiding make it hard to detect of formed stego files with usage of modern IDS.

Modern methods of cover files, such as DI, steganalysis can be divided into three groups [3]: signature, statistical and structural methods. The feature of first group is usage of preliminary known alterations (signatures) of cover images (CI) that are specific for steganographic methods. This makes signature-based stegdetector (SD) one of the most effective detection methods. Nevertheless, considerable limitation of such SD is necessity to known in advance of signatures that is inappropriate for detection of previously unknown embedding methods.

Methods of statistical and structural analysis share the same idea that stego images can be detected by abnormal changes of CI parameters. Still, these methods differ in approaches to estimate of image parameters – with usage methods either of statistical, or structural analyses.

Despite wide range of proposed steganalysis method, they performance considerably decrease in case of processing of stego images formed according to novel adaptive embedding methods (AEM). Feature of such methods is minimization of CI distortions caused by message hiding that considerably decrease SD performance.

For overcoming this limitation, it was proposed to pre-process (calibrate) of analyzed DI for improving stego/cover

ratio and simplify further detection [9]. This approach was developed in SRM rich models that are based on utilization of enormous set of high pass filters [6]. SRM models allow considerably improving detection accuracy at the cost of computational complexity and adaptability to new embedding methods. Therefore, further step in development of effective calibration method was done by usage of convolutional neural networks [5, 7]. Such networks allow learning appropriate high pass filtering during training stage that simplifies image calibration procedure. Nevertheless, high computational complexity of re-tuning of pre-trained network for detection of stego images formed according to new method limits their usage in real applications.

The task of image calibration can be reformulated as a search of effective method for image denoising under limited or even absent information about features of introduced noises (alterations) during message hiding. For solving this task we proposed to utilize component analysis methods that provide fast and flexible solutions for signal decomposition under limited information about theirs statistical features. The work is devoted to performance analysis of well-known Principal Component Analysis (PCA) for the task of calibration stego images formed according to novel AEM.

III. ADAPTIVE EMBEDDING METHODS FOR DIGITAL IMAGES

The state-of-the-art paradigm of DI steganography is based on minimization of CI alteration during message hiding [11]. This leads to considerable decreasing of stego images unmasking features (e.g. changes of statistical features) that decrease performance of modern stegdetectors.

Mentioned breakthrough of novel steganography methods is achieved by representation of message hiding procedure as the optimization problem with constraints [12]:

$$D(\mathbf{X},\mathbf{Y})=\sum_{i,j}\rho_{i,j}(\mathbf{X},\mathbf{Y})\xrightarrow{|\mathbf{M}|=\text{const}}\min, \quad (1)$$

where $\mathbf{X},\mathbf{Y}\in\Im=\left\{0,1,\ldots,2^k-1\right\}^{M\times N}$ are cover and stego images of size MxN pixels correspondingly; $k\in\mathrm{N}$ is color bit-depth; $D(\cdot,\cdot)$ is empirical function for estimation of CI distortion during stego data hiding; $\rho_{i,j}(\cdot,\cdot)$ is empirical function for estimation cover image's statistical features alteration by changes of $(i,j)^{\text{th}}$ pixel; $\mathbf{M}$ is binary representation of m-bits stego data.

In the general case, the function $\rho(\cdot)$ in (1) allows estimating changes of CI statistical parameters caused by a single pixel alteration as well as non-linear dependencies between these changes by embedding series of bits [12]. The former alteration can be performed using common statistical models of DI [4, 6]. The latter one highly depends on mutual influence of altered pixels that requires utilization of computationally intensive methods for such dependency estimation. In most cases, mentioned dependencies may be estimated only for small (short) message (up to 100 bits) [12]. Therefore, majority of modern embedding methods includes "simplified" functions $\rho(\cdot)$ that provide tractable estimation on single pixels alterations only.

The selection of CI pixels to be used for stego bits hiding is usually made by analysis of statistical parameters of current pixels neighborhood (clique) [12]. This allows providing low cover image alteration during message hiding by preserving tractable complexity of the embedding algorithm.

The advance adaptive embedding methods HUGO [11], MG [13] and MiPOD [14] were considered in the work. The HUGO method is based on minimizing the following empirical function $D(\cdot)$ in (1):

$$\min_{\pi}\mathrm{E}_{\pi}(D)=\sum_{\mathbf{Y}\in\Upsilon}\pi(\mathbf{Y})\cdot D(\mathbf{X},\mathbf{Y}),|\mathbf{M}|=\mathrm{H}(\pi), \quad (2)$$

$$\mathrm{H}(\pi)=-\sum_{\mathbf{Y}\in\Upsilon}\pi(\mathbf{Y})\cdot\log(\pi(\mathbf{Y})),$$

where $\mathbf{Y}$ is a stego image sampled from the set of all stego images Υ; π is probability distribution of selection of some stego image from the set Υ; $\mathrm{E}_{\pi}(D)$ is averaging operator for function $D(\cdot)$ over distribution π; $\mathrm{H}(\pi)$ is entropy function for distribution π.

Filler et al. [11] proposed to use co-occurrence matrix $\mathbf{C}_{k,l}$ for solving the optimization problem (2), namely estimation of CI distortions caused by message hiding:

$$D(\mathbf{X},\mathbf{Y})=\sum_{c\in\aleph}\sum_{(k,l)\in\Im}\omega_{k,l}\cdot\mathbf{C}^{c}_{(k,l)}(\mathbf{X},\mathbf{Y})$$

where $\aleph=\left\{\rightarrow,\leftarrow,\uparrow,\downarrow\right\}$ is set of scanning directions during co-occurrence matrix $\mathbf{C}_{k,l}$ estimation; $\omega_{k,l}>0$, $\forall(k,l)\in\Im$, is weighting coefficient; $\mathbf{C}^{c}_{(k,l)}$ is co-occurrence matrix by fixed brightness (k,l) of pixels and scanning direction c. For example, the matrix $\mathbf{C}^{\rightarrow}_{(k,l)}$ for grayscale images $\mathbf{X}$ and $\mathbf{Y}$ with size of MxN pixels by row-wise scanning (left-to-right) may be estimated according to next formula [11]:

$$\mathbf{C}^{\rightarrow}_{(k,l)}(\mathbf{X},\mathbf{Y})=\frac{1}{N(M-2)}\cdot$$

$$\cdot\sum_{i,j}\left|\left[\left(\mathbf{D}^{\rightarrow}_{i,j},\mathbf{D}^{\rightarrow}_{i,j+1}\right)(\mathbf{Y})=(k,l)\right]_{I}-\right. \quad (3)$$

$$\left.-\left[\left(\mathbf{D}^{\rightarrow}_{i,j},\mathbf{D}^{\rightarrow}_{i,j+1}\right)(\mathbf{X})=(k,l)\right]_{I}\right|,(k,l)\in\Im,$$

$$\left(\mathbf{D}^{\rightarrow}_{i,j},\mathbf{D}^{\rightarrow}_{i,j+1}\right)(\mathbf{X})=(k,l)\Leftrightarrow$$

$$\Leftrightarrow\left(\mathbf{D}^{\rightarrow}_{i,j}(\mathbf{X})=k\right)\wedge\left(\mathbf{D}^{\rightarrow}_{i,j+1}(\mathbf{X})=l\right),$$

where $\mathbf{D}^{\rightarrow}_{i,j}(\mathbf{U})$ is array of pixel brightness differences for grayscale image U on coordinate (i,j) by image row-wise processing (left-to-right). Estimation of $\mathbf{C}^{c}_{(k,l)}$ matrices for other scanning directions may be performed in similar way to (3) [11].

The feature of MG and MiPOD methods is minimization of both the CI distortion, and statistical stegdetector performance (estimated detection accuracy) during stego data embedding [13, 14]. This is achieved through the usage of Gaussian mixture models (GMM) for estimation cover image's noises parameters [14].

The cover image processing pipeline is similar for both MG [13] and MiPOD [14] methods. At the first stage, the CI is pre-processed (filtered) for suppressing the influence of cover image context using a filter F_{dn}:

$$\mathbf{r} = \mathbf{X} - F_{dn}\left(\mathbf{X}\right).$$

Then, variance σ_l^2 of pixels brightness for computed residuals r is calculated using next linear model:

$$\mathbf{r}_l = \mathbf{G}\mathbf{a}_l + \xi, l \in \left[1; M \cdot N\right]. \quad (4)$$

Sedighi et al proposed to use Maximum Likelihood for estimation of mentioned model parameters [14]:

$$\sigma_l^2 = \frac{\left\| \mathbf{P}_{\mathbf{G}}^{\perp} \mathbf{r}_l \right\|_F}{p^2 - q}, q \in \mathrm{N}, \quad (5)$$

where $\mathbf{P}_{\mathbf{G}}^{\perp}$ is projection operator for residuals $\mathbf{r}_l$ (4) on sub-space with (p^2-q) dimensionality, create from eigenvectors of matrix $\mathbf{G}$; $\left\|\cdot\right\|_F$ is Frobenius norm. Residuals $\mathbf{r}_l$ are computed within neighborhood of pxp pixels for current l^{th} pixel.

The presented simplified estimation of variance σ_l^2 (5) is used for MG method [13], while MiPOD method uses more accurate estimation:

$$\sigma_l^2 = \frac{\left\| \mathbf{r}_l - \mathbf{G}\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{r}_l \right\|_F}{p^2 - q}, q \in \mathrm{N}.$$

At the third stage, the magnitude β_l, $1 \le l \le M \cdot N$, of pixels brightness changes that minimizes the deflection coefficient ς^2 between cover and stego image distributions is estimated:

$$\varsigma^2\left(\beta_l\right) = 2\sum_{l=1}^{M \cdot N} \beta_l^2 \sigma_l^4 \xrightarrow{\sum_{l=1}^{M \cdot N} \mathrm{H}_4(\beta_l) = const} \min, \quad (6)$$

$$\mathrm{H}_4\left(z\right) = -2z\log\left(z\right) - \left(1-2z\right)\log\left(1-2z\right),$$

where $\mathrm{H}_4(z)$ is ternary entropy function. The deflection coefficient ς^2 (6) provides statistical measurement of divergence between cover and stego images distribution that reflects expected performance of statistical SD [13, 14].

The mentioned optimization task for coefficient ς^2 (6) can be solved using Lagrange multipliers method [14]. Then, optimal values of β_l and Lagrange multipliers λ_L can be calculated by numerical solving of next equations:

$$\beta_l^2 \sigma_l^4 = \frac{1}{2\lambda_L} \ln\left(\frac{1-2\beta_l}{\beta_l}\right), l \in \left[1; M \cdot N\right].$$

Estimated optimal values of β_l are used for calculation corresponding values of ρ_l function during embedding stegobit into l^{th} pixel of CI:

$$\rho_l = -\ln\left(\beta_l - 2\right), l \in \left[1; M \cdot N\right]. \quad (7)$$

At the last stage, message **M** bits are embedding to CI using trellis-code using magnitudes of pixels brightness alteration estimated with ρ_l (7).

It should be noted that the GMM used in MG and MiPOD methods allows accurately estimating of local alterations of pixels brightness during stego image formation [14]. This provides high robustness of formed stego images to known statistical steganalsis methods without involving of computationally intensive methods for image modeling, such as Random Markov Fields [15].

IV. METHODS OF DIGITAL IMAGE COMPONENT ANALYSIS

A numerous number of tasks in the digital image processing domain are based on usage of data transformation methods, such as Fourier transform, wavelet transform to name a few. These transformations simplify analysis and extraction of signal's components that present interest for data enhancement or compression tasks, for example noises and distortions.

The selection of an optimal transformation for specific problem (e.g. signal filtering, compression) is non-trivial task that depends on statistical features of test signals packet. The common approach for solving this task is usage of Fourier transform mathematical apparatus for selection an appropriate basis function [16], like wavelets, shearlets, bandlets etc. Despite effectiveness of such approach in image processing tasks, selection of an optimal basis function that minimizes image's restoration error remains open task today. The task is particularly solved for some practical cases, such as suppression of Gaussian noise, anisotropic noise etc.

Message hiding into cover image's noise decreases performance of well-known spectral transformation due to limited a prior information about features of embedding for selection of an appropriate basis functions. Recently, special interest is taken at component analysis methods for overcoming mentioned limitation [17]. The feature of component analysis is ability to signals decomposition by the criterion of their statistical characteristics. This makes component analysis methods attractive candidates for image steganalysis tasks.

One of well-known example of component analysis methods is principal components analysis (PCA). The PCA provides fast decomposition of multidimensional signals that is based on theirs energy (variance of elements values). The method is widely used in digital image processing domain for effective suppression of noises and distortions under limited aprior information about theirs statistical features [18].

Principle component analysis is based on signal decomposition into orthogonal components by the criterion of energy (variance of elements values) of these components:

$$\sum_{i=1}^{K} \left\| \mathbf{x}_i - \mathrm{L}_K \right\|_2 \to \min, \quad (8)$$

where $\mathbf{x}_i \in \mathrm{R}^N$, $1 \le i \le K$, is current vector from train set; $\mathrm{L}_K \subset \mathrm{R}^n$ is the best linear approximation of train vectors distribution; $\left\|\cdot\right\|_2$ is Euclidean norm. Let us note that any K-dimensional linear manifold in may be represented by a set of linear forms of orthonormal vectors $\left\{\mathbf{a}_0, \cdots, \mathbf{a}_K\right\} \subset \mathrm{R}^n$:

$$\mathrm{L}_K = \left\{\mathbf{a}_0 + \gamma_1\mathbf{a}_1 + \cdots + \gamma_K\mathbf{a}_K \mid \gamma_i \in \mathrm{R}, i \in \left[1; K\right]\right\}.$$

Then, the decomposition (8) may be presented in the form:

$$\left\|\mathbf{x}_i - \mathrm{L}_K\right\|_2 = \left\|\mathbf{x}_i - \mathbf{a}_0 - \sum_{j=1}^{K} \mathbf{a}_j \left\langle \mathbf{a}_j, \mathbf{x}_i - \mathbf{a}_j \right\rangle\right\|_2,$$

where $\langle\cdot,\cdot\rangle$ is dot product. Solution of current approximation task can be represented as set of nested linear manifolds $\mathrm{L}_0 \subset \ldots \subset \mathrm{L}_{K-1}$. These manifolds are fully described by set of orthonormal vectors (vectors of principal components) $\{\mathbf{a}_1, \cdots, \mathbf{a}_K\} \subset \mathrm{R}^n$ and vector $\mathbf{a}_0 \subset \mathrm{R}^n$. Each of a vector can be obtained by solving of approximation task for L_i manifold with using of generalized least square:

$$\mathbf{a}_i = \arg\min_{\mathbf{a}_i \in \mathrm{R}^n} \left(\sum_{j=1}^{K} \left\|\mathbf{x}_i - \mathrm{L}_j\right\|_2 \right).$$

This approximation task may be reduced to the equal task of diagonalization of covariance matrix $\mathbf{C}=(c_{ij})$, $1 \le i,j \le K$:

$$c_{ij} = \frac{1}{K-1} \sum_{l=1}^{K} \left(x_{li} - \overline{\mathbf{X}}_i\right) \cdot \left(x_{lj} - \overline{\mathbf{X}}_j\right),$$

where $\overline{\mathbf{X}}_k$ is the mean of vector $\mathbf{X}_k$ elements. By this representation, the PCA is related to spectral decomposition of covariance matrix $\mathbf{C}$, namely to represent of train vectors space as sum of orthogonal eigenspaces $\mathbf{C}_i$. Then, matrix $\mathbf{C}$ can be represented as linear form of orthogonal projection operators on these eigenspaces with weights $\lambda_i \ge 0, \forall i$.

Let us present a train set of row-vectors as $\mathbf{X}=\{\mathbf{x}_1, \ldots, \mathbf{x}_K\}^{\mathrm{T}}$, $\mathbf{x}_i \in \mathrm{R}^n$, then covariance matrix may be written as $\mathbf{C}=1/(K-1)\cdot\mathbf{X}^T\mathbf{X}$ and its spectral decomposition corresponds to singular decomposition of data matrix $\mathbf{X}$:

$$\mathbf{X} = \sum_{l=1}^{p} \sigma_l \mathbf{b}_l^T \mathbf{a}_l^T,$$

where σ_l is singular value of matrix $\mathbf{X}$; $\mathbf{a}_l \in \mathrm{R}^n$, $\mathbf{b}_l \in \mathrm{R}^m$ are right and left singular vectors correspondingly. The right singular vectors relate to principal components vectors and they are eigenvectors of covariance matrix $\mathbf{C}$ that corresponds to positive eigenvalues $\lambda_l = 1/(m-1)\cdot\sigma_l^2$.

The PCA is widely used in digital image denoising tasks, so far as it does not require a priori data on the statistical features of the image. In this case, image denoising is performed by removing of components with the smallest singular values that corresponds to noises. This makes possible to use thresholding to isolate and suppress noise, whose statistical characteristics differ from image components. This ability makes PCA an attractive candidate for image calibration task in digital image steganalysis.

V. EXPERIMENTS

The performance analysis of PCA usage for cover and stego images calibration was done on a subset of 10,000 grayscale images with size of 512x512 pixels that are randomly sampled from a standard data package ALASKA [19]. Stego images were formed according to advance adaptive methods HUGO, MG and MiPOD. The cover image payload was varied in the following range – 3%, 5%, 10%, 20%, 30%, 40% and 50%.

The stegdetector was constructed using second-order SPAM model [20] (686 features) and Random Forest classifier [21]. The SD was tested using standard cross-validation procedure by minimizing of total error P_e [21]:

$$P_e = \frac{1}{2}\left(P_{FA} + P_{MD}\left(P_{FA}\right)\right),$$

where P_{FA} and P_{MD} are probabilities of false alarm (type I error) and missed detection (type II error) correspondingly. Validation of SD was performed 10 times by pseudo random splitting of image dataset into train (50%) and tests (50%) samples.

The SPAM model used in SD is based on the analysis of the degree of mutual correlation for the adjacent pixels brightness [20] using Markov chains. Estimation of the brightness correlation was performed by scanning of analyzed image in row and column-wise fashions. Let us consider ad example of the correlation estimation for the case of row-wise processing (left-to-right scanning) of test grayscale image $\mathbf{U}$. The differences of adjusted pixels brightness in this case can be calculated as [20]:

$$\mathbf{D}^{\rightarrow}_{i,j} = \mathbf{U}_{i,j} - \mathbf{U}_{i,j+1}.$$

Then, considered differences are used for parameters estimation of first and second order Markov chains:

$$\mathbf{M}^{\rightarrow}_{u,v} = \Pr\left(\mathbf{D}^{\rightarrow}_{i,j+1} = u \mid \mathbf{D}^{\rightarrow}_{i,j} = v\right),$$

$$\mathbf{M}^{\rightarrow}_{u,v,w} = \Pr\left(\mathbf{D}^{\rightarrow}_{i,j+2} = u \mid \mathbf{D}^{\rightarrow}_{i,j+1} = v, \mathbf{D}^{\rightarrow}_{i,j} = w\right),$$

where $(-T) \le u,v,w \le T$ are values of differences between adjacent pixels brightness (states of Markov chain); $T \in \mathrm{N}$ is threshold. If probabilities $\Pr\left(\mathbf{D}^{\rightarrow}_{i,j} = v\right)$ or $\Pr\left(\mathbf{D}^{\rightarrow}_{i,j+1} = v, \mathbf{D}^{\rightarrow}_{i,j} = w\right)$ equal to zero, corresponding values of $\mathbf{M}^{\rightarrow}_{u,v}$ and $\mathbf{M}^{\rightarrow}_{u,v,w}$ are set zeros as well. Calculation of Markov chains parameters for other scanning directions may be performed in the same way [20].

Finally, parameters of SPAM models are calculated by averaging of estimated parameters for Markov chains [20]:

$$\mathbf{F}_{1\ldots k} = \frac{1}{4}\left(\mathbf{M}^{\rightarrow} + \mathbf{M}^{\leftarrow} + \mathbf{M}^{\uparrow} + \mathbf{M}^{\downarrow}\right),$$

$$\mathbf{F}_{(k+1)\ldots(2k)} = \frac{1}{4}\left(\mathbf{M}^{\square} + \mathbf{M}^{\square} + \mathbf{M}^{\square} + \mathbf{M}^{\square}\right).$$

This transformation is used the standard assumption that image's statistical features is robust to D affine transformation (rotation, flipping). The total number of parameters of first order SPAM model is equal to $k=(2T+1)^2$, while for the second order model is $k=(2T+1)^3$. During performance analysis, we used second order SPAM model for estimation statistical parameters of DI due to its high accuracy by preserving a tractable computational complexity.

It should be noted that image calibration during steganalysis leads to extending of features, which can be used for stego image detection. We may cluster these features into next groups [9, 22]:

1. Features of uncalibrated image – correspond to the case of using SPAM-features for the original (uncalibrated) image:
$$\mathbf{F}_{nc} = F_{SPAM}\left(\mathbf{U}\right).$$
2. Features of the calibrated image – correspond to the case of calculation of DI features after applying of PCA:
$$\mathbf{F}_{calib} = F_{SPAM}\left(C_{PCA}\left(\mathbf{U}\right)\right).$$
3. Linearly transformed features of the calibrated image – correspond to the difference between the features of calibrated and original images:
$$\mathbf{F}_{DF} = \mathbf{F}_{calib} - \mathbf{F}_{nc}.$$
4. Cartesian product of the features for calibrated and original images:
$$\mathbf{F}_{CC} = \left\{\mathbf{F}_{calib}; \mathbf{F}_{nc}\right\}.$$

The practical application of the $\mathbf{F}_{nc}$ features is limited due to insufficient accuracy of known statistical models to detection of weak (negligible) differences between cover and stego images [2]. Consequently, $\mathbf{F}_{CC}$ features are widely used instead of $\mathbf{F}_{nc}$ in majority of SD. Also, usage of linearly transformed features $\mathbf{F}_{DF}$ is taken special interest due to promising results for widespread AEM [22, 23].

Along with type of features used for SD training, stegdetectors performance significantly depends on fraction F_α of pairs of cover-stego images features utilized by training stage [24]:
$$F_\alpha = \frac{\left|\left\{\left(\mathbf{X},\mathbf{Y}\right):\left(\mathbf{X}_i,\mathbf{Y}_{\mathbf{X}_i}\right), i \in S_{train}\right\}\right|}{\left|S_{train}\right|}\cdot 100\%,$$
where S_{train} is set of digital images used during training of stegdetector; $\mathbf{Y}_{\mathbf{X}_i}$ is stego images formed from cover $\mathbf{X}_i$. The F_α parameter varies from 0% (absent of cover-stego images pairs in training set) to 100% (training set consists only from cover-stego images pairs). The former case corresponds to the real situation when steganalytics do not have access to stego encoder and may use only captured stego images. The latter one relates to the widely considered situation when steganalytics have full access to stego encoder for stego images generation, but they are limited in knowledge about features of embedding process. Both cases were considered during our analysis for more accurate estimation of image calibration procedure performance by PCA.

Performance analysis of image calibration procedure with PCA was done by variation of both types of features, and F_α parameters used during SD training procedure. At the first stage, the case of comparative analysis of state-of-the-art maxSRMd2 [6] and solely SPAM (without image calibration) models was considered. The feature of maxSRMd2 is utilization of extensive set of high-pass filters during image calibration stage for effectively suppression of CI content. This leads to enormous set of 12,753 features for such model, which complicates fast reconfiguration of SD for new embedding methods detection.

The dependencies of P_e error on the cover image payload by usage of maxSRMd2 and SPAM models and proposed approach for the adaptive steganographic methods HUGO, MG and MiPOD by F_α=100% are shown in Fig. 1.

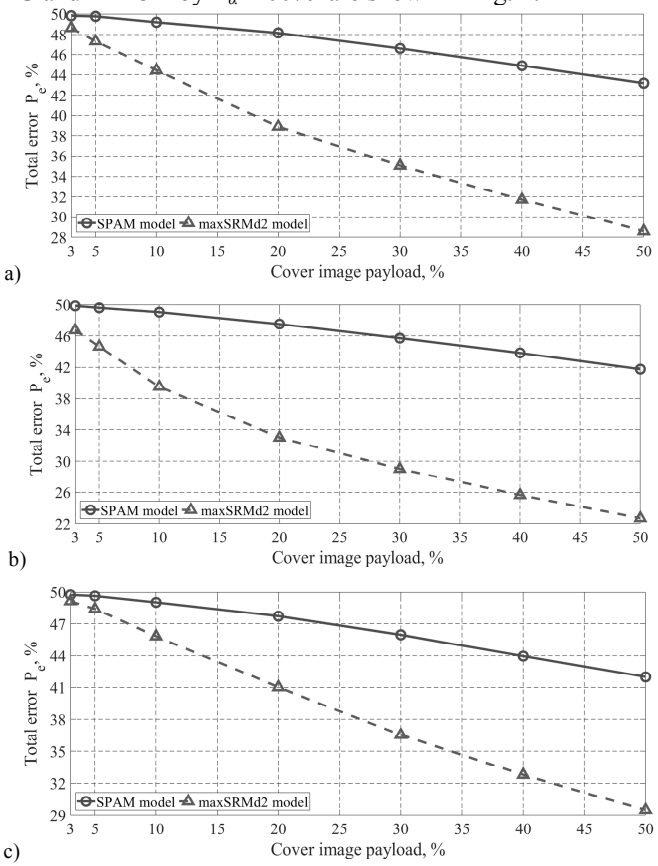

Fig. 1. Dependencies of P_e *error on the cover image payload by usage of maxSRMd2 and SPAM models and proposed approach for the adaptive steganographic methods HUGO (a), MG (b) and MiPOD (c) by* F_α*=100%.*

Usage of maxSRMd2 model allows considerably improving detection accuracy in comparison with proposed image calibration method (Fig. 1) – the reduction of P_e error varies from 9.5% for low ($\Delta_p \le 10\%$) to 19% for high ($\Delta_p \ge 50\%$) cover image payload. Revealed detection accuracy improvement significantly depends on used embedding method – the biggest improvement is achieved for MG (Fig. 1b, up to 9.5%) and HUGO (Fig. 1a, up to 4.8%) methods, while much smaller for MiPOD method (Fig. 1c, up to 3.2%). Thus, the obtained results confirm the general trend in the steganalysis domain – usage of rich models with extensive set of image filters allows significantly increases SD performance.

For comparison, the dependencies of P_e error on the cover image payload by usage of maxSRMd2 and SPAM models and proposed approach for the adaptive steganographic methods HUGO, MG and MiPOD by F_α=0% are shown in Fig. 2.

The absence of cover-stego pairs in the training set (F_α=0%, Fig. 2) leads to a significant reduction in the efficiency of both approaches:

- For the SPAM model, the reduction is 6.1% for the HUGO method and 5.8% for other methods;

- For the maxSRMd2 model, the reduction is 15.1% for the MiPOD method and 17.5% for other methods.

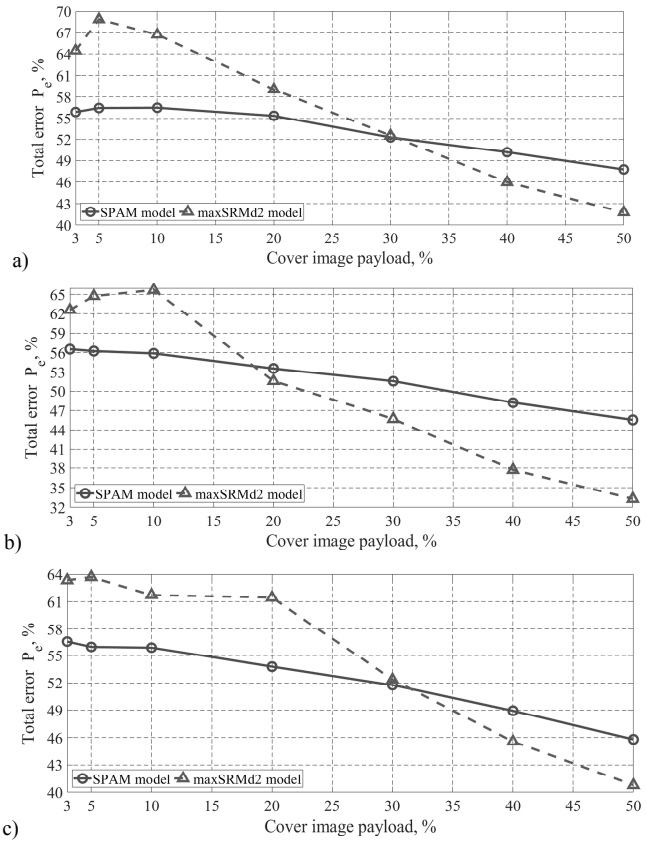

Fig. 2. Dependencies of P_e *error on the cover image payload by usage of maxSRMd2 and SPAM models and proposed approach for the adaptive steganographic methods HUGO (a), MG (b) and MiPOD (c) by* F_α*=0%.*

It should be noted that usage of the maxSRMd2 model in this case leads to a significant deterioration in the detection accuracy in comparison with much simpler SPAM model (Fig. 2). The biggest decrease of detection accuracy was revealed for low cover image payload ($\Delta_p \leq 10\%$, up to 12.5% reducing). Both models perform almost the same for the medium cover image payload ($10\% < \Delta_p < 50\%$).

Rich model maxSRMd2 allows reducing P_e error in comparison with SPAM model only for high cover image payload ($\Delta_p \geq 50\%$). Nevertheless, obtained "gain" of detection accuracy in this case (Fig. 2) is much smaller than for the previous one (Fig. 1) – up to 12.3% for the MG method, and up to 5.6% for other methods.

The obtained results are consistent with the previously obtained estimates for adaptive steganographic methods HUGO and S-UNIWARD [24] and it confirms the limited effectiveness of rich models by working under limited a prior information about used embedding methods (namely impossibility to obtain stego image for arbitrary cover one). Therefore, we consider only this situation (F_α=0%) during investigation as the closest one to the real cases of image steganalysis in the wild.

At the second stage, the case of image calibration with proposed PCA method was considered. The image denoising was performed by suppression a fraction Δ_C of image's components with the smallest singular values and, correspondingly, energies. Then, SPAM model was applied for feature extraction from denoised images.

The relative detection accuracy indicator P_Δ was used for estimating difference of P_e error for initial (processing of non-calibrated images) and considered (image calibration with PCA) cases:

$$P_\Delta = P_e^{SPAM} - P_e^{PCA}.$$

Positive values of the P_Δ index correspond to the case, when applying of proposed approach (PCA-based image denoising) allows improving stegdetector accuracy. The negative ones coincide with the case of decreasing SD performance by usage of PCA-based denoising in comparison with initial (non-calibrated) case.

Including of features for calibrated images allows improving detection accuracy [9]. As an example, we may mention maxSRMd2 model [6] where Cartesian calibrated features $\mathbf{F}_{CC}$ are used. Therefore, it takes an interest to estimate the gain in detection accuracy by utilization of linearly transformed $\mathbf{F}_{DF}$ and Cartesian calibrated features $\mathbf{F}_{CC}$ in comparison with calibrated $\mathbf{F}_{calib}$ ones.

The dependencies of P_Δ index on the CI payload and fraction ΔC of used image's components during PCA by usage of SPAM models and adaptive embedding methods HUGO for calibrated $\mathbf{F}_{calib}$, linearly transformed $\mathbf{F}_{DF}$ and Cartesian calibrated $\mathbf{F}_{CC}$ features by F_α=0% are shown in Fig. 3.

Applying of PCA for image calibration allows reducing the value of P_Δ index on 1.5% in the most difficult case – the low ($\Delta_p \leq 10\%$, Fig. 3) cover image payload. Mentioned decreasing was revealed for $\mathbf{F}_{calib}$ (Fig.3a) and $\mathbf{F}_{DF}$ (Fig.3b) features for all considered range of cover image payload values. At the same time, usage of Cartesian calibrated features $\mathbf{F}_{CC}$ (Fig.3c) does not allow improving detection accuracy. This can be explained by limited ability to reveal weak changes of cover image parameters during message hiding by analysis of doubled feature vector (features of initial and calibrated images).

It should be noted that usage of $\mathbf{F}_{DF}$ features (Fig.3b) leads to decreasing of classification accuracy by taking only fraction of image components ($\Delta_C \leq 97\%$), while suppression of the components with minimum singular values ($\Delta_C \leq 99\%$) lead to opposite effect (increasing detection accuracy). This may be explained by "spreading" of CI distortions during message hiding according to HUGO methods over several components. Consequently, removing of components with minimal singular values does not allow completely removing these distortions that leads to increasing of classification accuracy.

For comparison, the dependencies of P_Δ index on the cover image payload and fraction Δ_C of used image's components during PCA by usage of SPAM models and adaptive steganographic methods MG for calibrated $\mathbf{F}_{calib}$, linearly transformed $\mathbf{F}_{DF}$ and Cartesian calibrated $\mathbf{F}_{CC}$ features by F_α=0% are shown in Fig. 4.

The obtained results for the MG method (Fig. 4) confirm the previously obtained results for the HUGO method (Fig. 3) – usage $\mathbf{F}_{calib}$ and $\mathbf{F}_{DF}$ features allow increasing detection accuracy up to 1% in the whole range of cover image payload. On the other hand, obtained "gain" of detection accuracy by

usage of $\mathbf{F}_{calib}$ features for MG method is smaller (up to 1%, Fig. 4a) in comparison with HUGO method (up to 1.5%, Fig. 3a). Similar situation is obtained for $\mathbf{F}_{DF}$ features as well – increasing detection accuracy up to 0.5% for MG method (Fig. 4b) in contrast to 1% for HUGO method (Fig. 3b). The $\mathbf{F}_{CC}$ features allows increasing detection accuracy only for high CI payload ($\Delta_p \geq 50\%$) for both embedding methods (Fig.3-4).

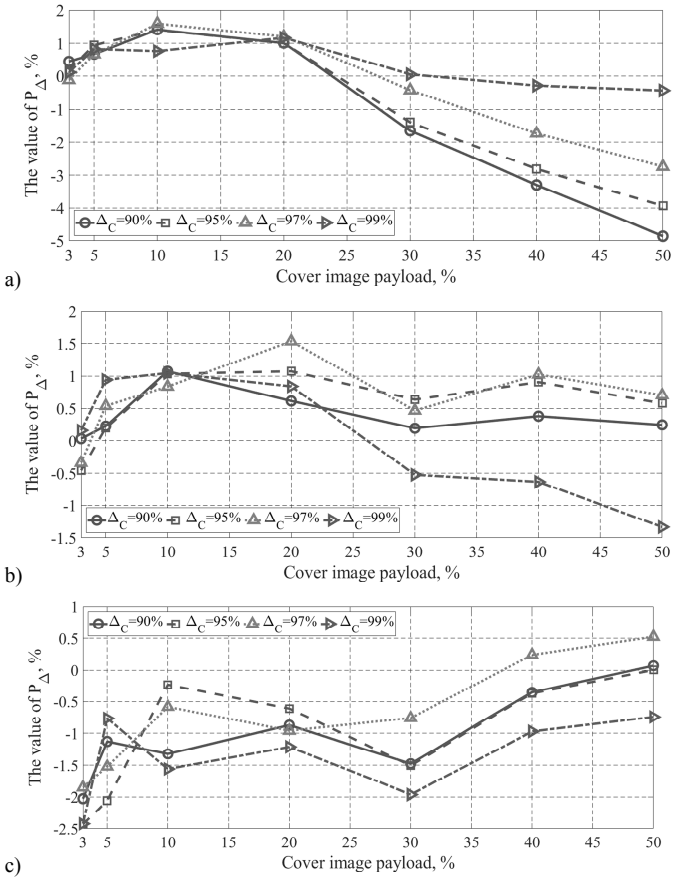

Fig. 3. Dependence of P_Δ *index on the cover image payload and fraction* Δ_C *of used image's components during PCA by usage of SPAM models and adaptive steganographic methods HUGO for calibrated* $\mathbf{F}_{\text{calib}}$ *(a), linearly transformed* $\mathbf{F}_{\text{DF}}$ *(b) and Cartesian calibrated* $\mathbf{F}_{\text{CC}}$ *(c) features by* F_α*=0%.*

Given the results, performance analysis of PCA applying for stego images formed by advanced MiPOD method is of special interest. Dependencies of P_Δ index on the CI payload and fraction Δ_C of used image's components during PCA by usage of SPAM models and MiPOD method for calibrated $\mathbf{F}_{calib}$, linearly transformed $\mathbf{F}_{DF}$ and Cartesian calibrated $\mathbf{F}_{CC}$ features by F_α=0% are shown in Fig. 5.

The results for MiPOD method (Fig. 5) correspond to the previously obtained results for HUGO (Fig. 3) and MG (Fig. 4) methods – usage of $\mathbf{F}_{calib}$ and $\mathbf{F}_{DF}$ features allow increasing detection accuracy in comparison with widely used $\mathbf{F}_{CC}$ features. Nevertheless, obtained "gain" of detection accuracy for MiPOD method (Fig. 5) is smaller than considered HUGO and MG methods – up to 1.0% for $\mathbf{F}_{calib}$ and near 0% for $\mathbf{F}_{DF}$ features. Utilization of $\mathbf{F}_{CC}$ features (Fig. 5c) does not allow improving detection accuracy at all in comparison with the case of usage the SPAM model alone.

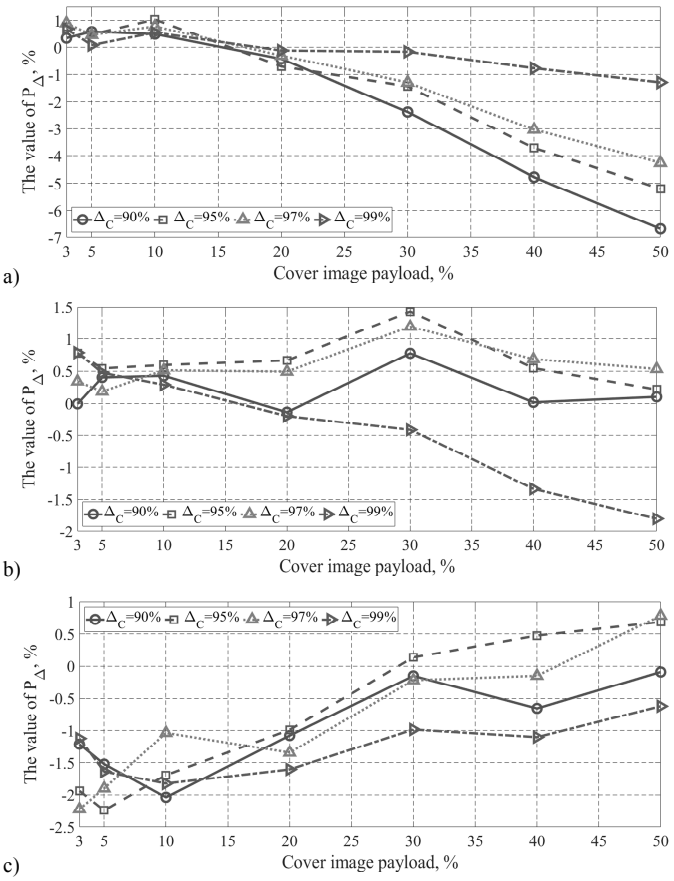

Fig. 4. Dependence of P_Δ *index on the cover image payload and fraction* Δ_C *of used image's components during PCA by usage of SPAM models and adaptive steganographic methods MG for calibrated* $\mathbf{F}_{\text{calib}}$ *(a), linearly transformed* $\mathbf{F}_{\text{DF}}$ *(b) and Cartesian calibrated* $\mathbf{F}_{\text{CC}}$ *(c) features by* F_α*=0%.*

VI. DISCUSSIONS

Obtained results of detection accuracy for stego images formed according to state-of-the-art adaptive embedding methods proved effectiveness of PCA usage for DI calibration. The comparison of detection accuracy changes by usage of considered models and embedding methods are given in Table. 1.

Usage of principal components analysis allows considerably reducing the classification error P_e in comparison with novel maxSRMd2 model (Table 1). The PCA allows reducing the P_e error for the most difficult cases (low cover image payload and absence of cover-stego images pairs in training set) in comparison with advanced maxSRMd2 model.

Also, the revealed "gain" of detection accuracy by usage of PCA is preserved even by comparison with SPAM model (up to 1.5% increasing of detection accuracy) even for low CI payload.

VII. CONCLUSION

Based on the performed analysis results, limitations of well-known approach to stego image calibration with extensive set of high pass filters was revealed. The comparative analysis of standard SPAM model and novel

maxSRMd2 model showed considerable decreasing of detection accuracy in cases of real images steganalysis (absence of cover-stego images pairs in training set). Decreasing of detection accuracy for maxSRMd2 model achieved up to 12.5% for modern HUGO, MG and MiPOD methods that makes this model inappropriate for applications.

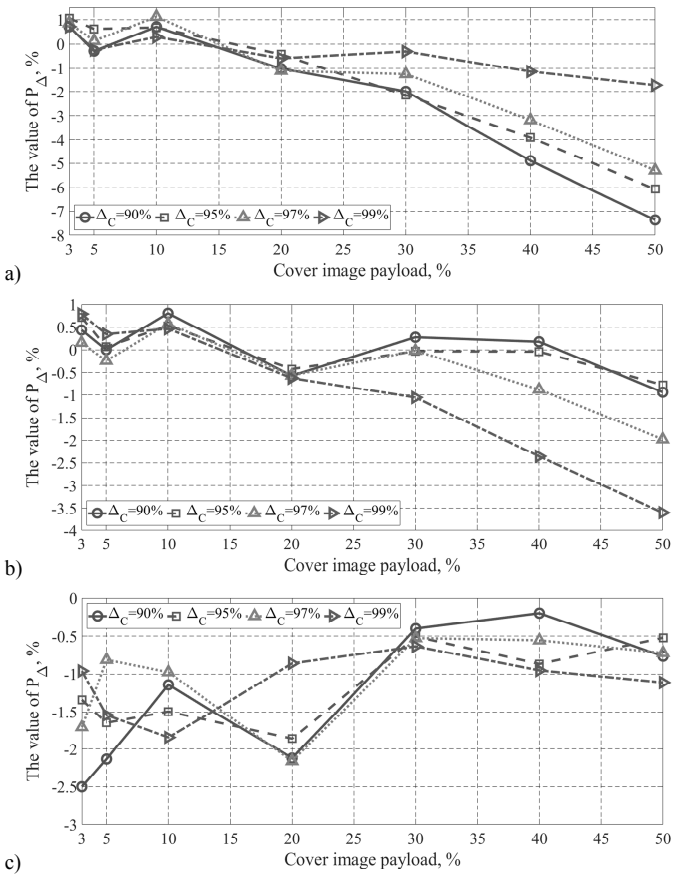

Fig. 5. Dependence of P_{Δ} *index on the cover image payload and fraction* Δ_C *of used image's components during PCA by usage of SPAM models and adaptive steganographic methods MiPOD for calibrated* $\mathbf{F}_{\text{calib}}$ *(a), linearly transformed* $\mathbf{F}_{\text{DF}}$ *(b) and Cartesian calibrated* $\mathbf{F}_{\text{CC}}$ *(c) features by* $F_{\alpha}=0\%$.

For overcoming mentioned limitations, we proposed to provide image calibration with component analysis methods, namely Principal Component Analysis. Feature of such methods is ability to signal decomposition under limited a prior information about statistical features of its components, namely noises, that makes it an attractive candidate for steganalysis.

Obtained results shown that proposed approach allows improving detection accuracy up to 1.5% in comparison with known statistical models. It should be noted that revealed "gain" of detection accuracy is preserved even in the most difficult cases (low cover image payload and absence of cover-stego images pairs in training set).

Table 1

Mean and standard deviation of the P_{Δ} indicator by stegdetector adjusting with usage of initial and calibrated (Δ_C=95%) cover and stego images by F_{α}=0% and variation of cover image payload

Stego images detection method		Cover image payload					
		Δ_P=5%		Δ_P=20%		Δ_P=50%	
		mean	std	mean	std	mean	std
HUGO embedding method							
SPAM model		0.00	0.00	0.00	0.00	0.00	0.00
maxSRMd2 model		-12.43	7.60	-3.72	4.29	5.95	1.34
PCA	$\mathbf{F}_{calib}$	0.94	0.01	1.00	0.01	-3.93	0.01
	$\mathbf{F}_{DF}$	0.20	0.00	1.07	0.01	0.58	0.01
	$\mathbf{F}_{CC}$	-2.06	0.01	-0.61	0.01	0.00	0.01
MG embedding method							
SPAM model		0.00	0.00	0.00	0.00	0.00	0.00
maxSRMd2 model		-8.52	7.53	1.97	2.73	12.26	1.42
PCA	$\mathbf{F}_{calib}$	0.48	0.01	-0.71	0.01	-5.20	0.01
	$\mathbf{F}_{DF}$	0.54	0.01	0.67	0.01	0.22	0.01
	$\mathbf{F}_{CC}$	-2.24	0.01	-0.99	0.01	0.69	0.00
MG embedding method							
SPAM model		0.00	0.00	0.00	0.00	0.00	0.00
maxSRMd2 model		-7.72	4.15	-7.68	2.66	5.02	1.95
PCA	$\mathbf{F}_{calib}$	0.62	0.01	-0.42	0.01	-6.07	0.01
	$\mathbf{F}_{DF}$	0.07	0.01	-0.42	0.01	-0.77	0.01
	$\mathbf{F}_{CC}$	-1.65	0.02	-1.86	0.02	-0.52	0.01

References

[1] V. Kopeytsev. "Steganograph in attacks on industrial enterprises". Kaspersky Inc., Tech. Rep, 2020, 6 pages. [Online] Available at:https://ics-cert.kaspersky.com/media/KASPERSKY_Steganography_in_targeted_attacks_EN.pdf (Accessed: 10 November 2021)

[2] J. Fridrich, *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge: Cambridge University Press, 2009, 437 pages, ISBN 978–0–521–19019–0, DOI: 10.1017/CBO9781139192903.

[3] G. Konachovych, D. Progonov, O. Puzyrenko. *Digital steganography processing and analysis of multimedia files*. Kyiv, 'Tsentr uchbovoi literatury' publishing, 2018, 558 pages, ISBN 978-617-673-741-4, Available at: http://pdf.lib.vntu.edu.ua/books/2019/Konahovich_2018_558.pdf (Accessed: 17 November 2021).

[4] J. Fridrich, J. Kodovsky. "Rich models for steganalysis of digital images", *IEEE Transactions on Information Forensics and Security*, vol. 7, iss. 3, 2012, pp. 868-882, DOI 10.1109/TIFS.2012.2190402.

[5] M. Boroumand, M. Chen, J. Fridrich. "Deep Residual Network for Steganalysis of Digital Images", *IEEE Transactions on Information Forensics and Security*, vol. 14, iss. 5, 2018, pp. 1181-1193. DOI: 10.1109/TIFS.2018.2871749.

[6] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, J. Fridrich. "Selection-Channel-Aware Rich Model for Steganalysis of Digital Images", in *IEEE Workshop on Information Forensic and Security*, Atlanta, GA, 2014, DOI 10.1109/WIFS.2014.7084302.

[7] R. Zhang, F. Zhu, J. Liu, and G. Liu, ''Efficient feature learning and multisize image steganalysis based on CNN,'' Jul. 2018, arXiv:1807.11428. [Online]. Available: http://arxiv.org/abs/1807.11428 (Accessed: 10 November 2021)

[8] A. Cohenab, A. Cohena, N. Nissim. "ASSAF: Advanced and Slim StegAnalysis Detection Framework for JPEG images based on deep convolutional denoising autoencoder and Siamese networks", *Neural Networks*, vol. 131, pp. 64-77, Nov. 2020. [Online]. DOI: 10.1016/j.neunet.2020.07.022

[9] J. Kodovskym J. Fridrich. "Calibration revisited", in *Multimedia and security: 11th ACM workshop*, Princeton, 2009, pp. 63-74, DOI: 10.1145/1597817.1597830.

[10] J. Butora, Y. Yousfi, J. Fridrich. "How to Pretrain for Steganalysis", in *ACM Workshop on Information Hiding and*

Multimedia Security, Brussels, Belgium, 2021, pp. 143-148, DOI: 10.1145/3437880.3460395.

[11] T. Filler, J. Fridrich. "Gibbs construction in steganography", *IEEE Transactions on Information Forensics Security*, vol. 5, 2010, pp. 705-720, DOI: 10.1109/TIFS.2010.2077629.

[12] T. Filler, J. Fridrich. "Design of adaptive steganographic schemes for digital images", in *Electronic Imaging, Media Watermarking, Security, and Forensics: The International Society for Optical Engineering*, San Francisco, CA, 2011, DOI: 10.1117/12.872192.

[13] V. Sedighi, J. Fridrich, R. Cogranne. "Content-adaptive pentary steganography using the multivariate generalized gaussian cover model", in *Electronic Imaging, Media Watermarking, Security, and Forensics: The International Society for Optical Engineering*, San Francisco, CA, 2015, DOI: 10.1117/12.2080272.

[14] V. Sedighi, R. Cogranne, J. Fridrich. "Content adaptive steganography by minimizing statistical detectability", *IEEE Transactions on Information Forensics Security,* vol. 11, 2015, pp. 221-234, DOI: 10.1109/TIFS.2015.2486744.

[15] Stan Z. Li. Markov Random Field Modeling in Image Analysis. In Advances in Computer Vision and Pattern Recognition, Springer, 2009, 362 pages, ISBN 978-1-84800-278-4, Available at: https://link.springer.com/book/10.1007/978-1-84800-279-1 (Accessed: 17 November 2021).

[16] S. Mallat. *A Wavelet Tour of Signal Processing. The Sparse Way*. 3rd ed. Academic Press, 2008, 832 pages, ISBN 978-0123743701, Available at: https://www.sciencedirect.com/book/9780123743701/a-wavelet-tour-of-signal-processing (Accessed: 17 November 2021).

[17] P. Comon, C. Jutten. *Handbook of Blind Source Separation*. 1st ed. Academic Press, 2010, 856 pages, ISBN 9780123747266, Available at: https://www.sciencedirect.com/book/9780123747266/handbook-of-blind-source-separation (Accessed: 17 November 2021).

[18] R. Gonzalez, R. Woods. *Digital Image Processing*. 4th ed. Pearson Press, 2017. 1192 pages, ISBN 978-0133356724, Available at: http://sdeuoc.ac.in/sites/default/files/sde_videos/Digital%20Image%20Processing%203rd%20ed.%20-%20R.%20Gonzalez%2C%20R.%20Woods-ilovepdf-compressed.pdf (Accessed: 17 November 2021).

[19] R. Cogranne, Q. Gilboulot, P. Bas. "The alaska steganalysis challenge: A first step towards steganalysis", in *Information Hiding and Multimedia Security*, Paris, 2019, ACM Press, pp. 125-137, DOI: 10.1145/3335203.3335726.

[20] T. Pevny, P. Bas, J. Fridrich. "Steganalysis by subtractive pixel adjacency matrix", *IEEE Transactions on Information Forensics Security*, vol. 5, 2010, pp. 215-224, DOI: 10.1109/TIFS.2010.2045842.

[21] J. Kodovsky, J. Fridrich. "Ensemble classifiers for steganalysis of digital media", *IEEE Transactions on Information Forensics Security,* vol. 7, 2012, p. 432-444, DOI: 10.1109/TIFS.2011.2175919.

[22] D. Progonov, V. Lucenko. "Steganalysis of adaptive embedding methods by message re-embedding into stego images", *Information Theories and Applications,* vol. 27, iss. 4, 2020, p. 3-24, Available at: http://www.foibg.com/ijita/vol27/ijita27-04-p01.pdf (Accessed: 17 November 2021).

[23] D. Progonov. "Influence of digital images preliminary noising on statistical stegdetectors performance", *Radio Electronics, Computer Science, Control,* vol. 1(56), 2021, pp. 184-193, DOI: 10.15588/1607-3274-2021-1-18.

[24] D. Progonov. "Performance of Statistical Stegdetectors in Case of Small Number of Stego Images in Training Set", in "Problems of Infocommunications Science and Technology (PIC S&T 2020)", Kharkiv, 2020, DOI:10.1109/PICST51311.2020.9467901.

Dmytro Progonov was born in Kyiv, Ukraine, in 1991. He received the B.S. and M.S. degrees in information protection systems from the Kyiv Polytechnic Institute, Kyiv, in 2011 and 2013 respectively. He received the Ph.D. degree in information security from Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, in 2016.

From 2013 to 2017, he was an Assistant with the Physics and Information Security Systems Department, Igor Sikorsky Kyiv Polytechnic Institute. Since 2017, Mr. Progonov has been an Associate Professor with the Physics and Information Security Systems Department, Igor Sikorsky Kyiv Polytechnic Institute, Kyiv. From 2021, he joined the Information Security Department, Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, as Associate Professor.

He is the author of book in the domain of digital image steganalysis, and more than 15 papers related to digital image forensics. His research interests include digital media forensics, behavior-based person authentication, machine learning and advanced signal processing. He is an Associate Editor of the journal *Information Models & Analyses*, and holds five patents.

Mr. Progonov was a recipient of the President of Ukraine Young Scientist Award in 2018.