MULTICLASS DETECTOR FOR MODERN STEGANOGRAPHIC METHODS

Dmytro Progonov

Abstract: *Creation of advanced steganalysis methods for reliably detection of hidden messages in widespread multimedia files, such as digital images, is topical task today. One of the key requirements to such methods is ability to reveal the stego files even in case of limited or absent information relating to used embedding methods. For solving this task there was proposed the multiclass stegdetector, based on applying the powerful methods of digital image structural analysis. Obtained earlier results confirmed the high efficiency of proposed stegdetector by message hiding in cover image's transformation domain. There is conducted analysis of stegdetector performance in case of message hiding according to advanced adaptive steganogaphic methods, such as HILL, MiPOD and Synch algorithms. It is shown that usage the "extended" cover image model, includes not only statistical, but also correlation and fractal features, gives opportunity to improve the detection accuracy of stegdetector in most difficult cases of image steganalysis.*

Keywords: *digital images steganalysis, adaptive embedding methods, multiclass stegdetector.*

ITHEA Keywords: *K.6.5 Management of computing and information systems. Security and Protection; I.4.10 Image processing and computer vision. Image representation.*

Introduction

Protection of private as well as state-owned sensitive information is urgent problem today. Considerable quantity of freely available malware, ransomware and operation system's backdoors packets allow any users of Internet to create the personal toolbox for attacking not only private computers, but also the information infrastructures systems of governmental agencies as well as private corporations. Distinctive feature of such attacks is wide usage of complicated methods for creation the hidden communication [Cisco, 2015; Cisco, 2016; FireEye, 2015]. These channels are integrated into information flows in telecommunication systems, like email, social networks, file sharing networks, which complicates the issue of theirs detection and counteraction by state security analytics agencies.

It is worth noting that in most cases information relating to data embedding process is limited or even absent. Therefore, applying of known signature or statistical steganalysis methods does allow providing the high accuracy of stego files detection. That is why development of new steganalysis approaches,

which allow detecting the hidden messages in case of limitation or absence the advance information regard used steganography technique, are required to be developed.

Related work

For revealing the hidden communication channels there are proposed considerable numbers of targeted steganalysis methods, based on usage the signature database and statistical models of cover files, such as digital images [Fridrich, 2009; Cox et al, 2008; Böhme, 2010]. Advantage of these methods is high accuracy of hidden messages (stego files), but only when embedding method is a priory known. For improvement the performance of signature and statistical stegdetectors in case of limited information relating to used steganalysis technique, there was proposed to use the rich cover model [Fridrich and Kodovsky, 2012a], obtained by merging of several statistical models in spatial as well as JPEG domains. Nevertheless practical usage of proposed stegdetectors is limited due to ample quantity of cover's model parameters which should be computed, for instance 34,671 parameters for SRM [Fridrich and Kodovsky, 2012a] and 35,263 features for J+SRM [Fridrich and Kodovský, 2012b] models.

Alternative approach to stego image detection is based on usage the simplified or approximate cover models [Avcibas et al, 2003; Farid, 2001]. Obtained universal (blind) stegdetectors give opportunity to overcome mentioned drawbacks of targeted steganalysis methods and reveal the hidden messages when there is no information about embedding process. But usage of approximate cover model makes unfeasible elicitation of slight changes of parameters the sophisticated cover models, which are widely used in modern embedding algorithm. It leads to deterioration of stegdetectors performance, especially in case of usage the adaptive steganographic techniques, such as HUGO algorithm, MiPOD algorithm and UNIWARD family of embedding methods.

For overcome mentioned drawbacks of well-known steganalysis methods, there was proposed to use the powerful methods of structural analysis for revelation the slight changes the cover image fine structure, caused by message hiding [Progonov and Kushch, 2014a; Progonov and Kushch, 2014b; Progonov and Kushch, 2015a; Progonov and Kushch, 2015b]. Based on developed methods of structural steganalysis it was proposed the multiclass stegdetector (MCS), which gives opportunity not only reveal the stego images, but also determinate the class of steganographic methods used for theirs creation. Results of comparative analysis the performance of MCS in case of stegodata hiding in transformation domains [Progonov, 2016] confirmed the high efficiency of proposed approach. Therefore it is of interest further examination of multiclass stegdetector performance by stego image formation according to advance embedding methods.

Task and challenges

Our purpose is investigation the performance of proposed multiclass stegdetector in case of usage the modern adaptive methods for data embedding in digital images.

Advanced methods for data embedding in digital images

For message hiding in digital images, there was proposed significant number of steganographic methods. Such methods can be divided into four groups [Fridrich, 2009; Cox et al, 2008; Böhme, 2010]:

1. Model preserving methods – are designed to preserve the simplified model of the cover source. The examples of such methods are MBS1 and MBS2 algorithms.
2. Mimicking natural image processing methods – the goal of such methods is to masquerade the embedding as some natural process of images, such as noise superposition during image acquisition. In this group of steganographic methods can be included the stochastic modulation method.
3. Steganalysis-aware methods – use known steganalysis attacks as guidance for the design the embedding process. As examples it should be mentioned (±1) algorithm, F5 algorithm and HUGO algorithms.
4. Minimal-impact (adaptive) methods – are based on minimizing the total cost (impact) of data hiding during formation of stego image. The total cost is measured as sum of embedding changes at each cover image element during hiding the separate stegobit. The most well-known adaptive methods are WOW method, UNIWARD family of steganographic algorithms, Synch algorithm.

The stego scheme, based on model-preserving principle, will be undetectable as long as the chosen model completely describes the cover images. Due to lack of accurate models for real images, there are used simplified model of image, for instance based on preserving its first-order statistics or histogram [Fridrich, 2009]. Applying by stego analytic more precise model of cover image source, for example, including the high-order statistics, allows reliably detecting the stego images, formed according to such schemes.

By usage the stego methods from the second group, even if the effect of embedding were in-distinguishable from some natural processing, the obtained stego images should stay compatible with the distribution of cover images. Distinction between the cover image dataset used by steganographer and steganalytics can be used by the latter for reliably detection the formed stego images.

The most secure stego schemes for message hiding today are related to groups of steganalysis-aware and minimal-impact methods. Such methods are typically realized in two steps: firstly, compute the cost of changing each cover image's pixel with usage of predefined distortion function. Secondly, secret

message is embedding while minimizing the sum of cost of all changed pixels. Such approach gives opportunity to create high robust embedding methods, which are most challenging to steganalysis. Well-known examples of such methods are WOW [Holub and Fridrich, 2012], UNIWARD method's family [Holub et al, 2014], HILL [Li et al, 2014] and MiPOD [Sedighi et al, 2016] embedding algorithms. Let us consider such algorithm in more details.

Peculiarity of first adaptive embedding methods was usage of heuristic-defined function $\rho(\mathbf{x})$ for estimation the cover image $\mathbf{x}$ distortion due to message hiding. Applying of simplified image model, which does not capture the interpixels dependences, allows represent $\rho(\mathbf{x})$ as superposition of local disturbances ρ_{ij} of cover image's characteristics due to stegodata embedding. One of the well-known examples of such distortion function was proposed in the WOW embedding algorithm [Holub and Fridrich, 2012]:

$$\rho_{ij} = \sum_{l=1}^{L} \frac{1}{\left| \sum\limits_{(m,n)\in M\times N} \left|\mathbf{R}_{mn}^{(l)}\right| \cdot \left|\mathbf{R}_{mn}^{(l)} - \mathbf{R}_{[ij]mn}^{(l)}\right| \right|},$$

where $\mathbf{R}^{(l)} = \mathbf{x} * \mathbf{K}^{(l)}$ – lth residuals, obtained by convolution of cover image $\mathbf{x}$ and lth direction filter $\mathbf{K}^{(l)}$; $\mathbf{R}_{[ij]}^{(l)} = \mathbf{x}_{[ij]} * \mathbf{K}^{(l)}$ – lth residuals, calculated for cover image after hiding separate stegobit by altering the pixel brightness at position $\mathbf{x}_{[ij]}$; $\aleph_L = \left\{\mathbf{K}^{(1)}, \mathbf{K}^{(2)}, \ldots, \mathbf{K}^{(L)}\right\}$ – bank of directional filters; M, N – size of cover image $\mathbf{x}$. For additional decreasing the number of disturbed pixels stegodata is preprocessed with usage of syndrome-trellis codes. WOW algorithm forces the distortion to be high where the content is predictable in at least one direction (smooth areas and clean edges) and low where the content is unpredictable in every direction (as in textures).

Modification of WOW's distortion function was proposed in HILL algorithm [Li et al, 2014]:

$$\boldsymbol{\rho} = \frac{1}{\left|\mathbf{x} * \mathbf{H}\right| * \mathbf{L}_1} * \mathbf{L}_2,$$

where $\mathbf{H}$ – high-pass filter (Ker–Böhme kernel); $\mathbf{L}_1, \mathbf{L}_2$ – correspondingly, low-pass (averaging) filter of support 3×3 and 15×15 pixels. Low-pass filtering of the costs $\boldsymbol{\rho}$ allows improving empirical security

due to increasing the entropy of embedding changes in highly textured regions and, therefore, reducing the distortion for the same payload.

Further development of WOW's distortion function is universal wavelet relative distortion (UNIWARD) [Holub et al, 2014]:

$$\rho(\mathbf{x},\mathbf{y}) = \sum_{k=1}^{3}\sum_{u=1}^{n_1}\sum_{v=1}^{n_2}\frac{\left|W_{uv}^{(k)}(\mathbf{x}) - W_{uv}^{(k)}(\mathbf{y})\right|}{\sigma + \left|W_{uv}^{(k)}(\mathbf{x})\right|},$$

where $\mathbf{x},\mathbf{y}$ – correspondingly cover and stego images; $W_{uv}^{(k)}(\mathbf{x})$ – uvth wavelet coefficient in the kth subband of the first level the two-dimensional discrete wavelet transformation the cover image; $\sigma(\sigma > 0)$ – constant stabilizing the numerical calculations. Usage of proposed distortion function allows create the state-of-art uniform approach to cover image parameters disturbances regardless of the message embedding domain [Holub et al, 2014].

It should be noted, that considered embedding algorithms allow minimize the distortion of cover image parameters by message hiding, but do not taking into account the statistical detectability of obtained stego images. Design of distortion functions that measure cover image distortions as well as statistical detectability of formed stego images is one of open problems in digital image steganography today [Ker et al, 2013]. For solve this problem there were proposed various approaches, based on usage only pixels, which have the smallest impact on the empirical statistical distribution of pixels groups [Pevný, 2010] or usage the distortion functions, which are optimized to minimize the empirical detectability in terms of the margin between cover and stego images represented using low-dimensional features [Filler and Fridrich, 2011]. These approaches are limited to empirical "models" that need to be learned from a database of images and, therefore, may become highly detectable should the Warden choose a different feature representation [Filler and Fridrich, 2011]. For overcome mentioned drawback there was proposed to model the cover pixels as a sequence of independent Gaussian random variables with unequal variances (multivariate Gaussian or MVG). It gives opportunity to achieve the empirical security of the embedding methods, which was subpar with respect to state-of-the-art steganographic methods [Holub and Fridrich, 2012; Holub et al, 2014]. Example of steganographic techniques, based on such approach, is MiPOD embedding method, which uses the locally-estimated multivariate Gaussian cover image model.

Message hiding in grayscale cover image $\mathbf{x}$ with size $M \times N$ (pixels) according to MiPOD method is carried out in several steps [Sedighi et al, 2016]:

1. Suppress the image content $\mathbf{x} = (x_1, x_2, \ldots, x_L)$, $L = M \cdot N$, using a denoising filter F:

$$\mathbf{r} = \mathbf{x} - F(\mathbf{x}),$$

where $\mathbf{x}$ is represented in column-wise order;

2. Measure pixels residual variance σ_l^2 using Maximum Likelihood Estimation and local parametric linear model:

$$\mathbf{r}_l = \mathbf{G}\mathbf{a}_l + \xi_l, \tag{1}$$

where $\mathbf{r}_l$ – represents the value of the residual $\mathbf{r}$ inside the $p \times p$ block surrounding the l[th] residual put into a column vector of size $p^2 \times 1$; $\mathbf{G}$ – a matrix if size $p^2 \times p$ that defines the parametric model of remaining expectation; $\mathbf{a}_l$ – a vector of $q \times 1$ of parameters; ξ_l – the signal whose variance is need to be estimated.

The pixels residual variance σ_l^2 is estimated according to further formula:

$$\sigma_l^2 = \frac{\left\| \mathbf{P}_{\mathbf{G}}^{\perp} \mathbf{r}_l \right\|^2}{p^2 - q},$$

where $\mathbf{P}_{\mathbf{G}}^{\perp} = \mathbf{I}_l - \mathbf{G}\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{G}^T$ – the orthogonal projection of the residual $\mathbf{r}_l$, estimated according to (1), onto the $p^2 - q$ dimensional subspace spanned by the left null space of $\mathbf{G}$; $\mathbf{I}_l$ – the $l \times l$ unity matrix.

3. Determine the probability of l[th] embedding change $\beta_l, I \in \{1, 2, \ldots, L\}$ that minimize the deflection coefficient ς^2 between cover and stego image distributions:

$$\varsigma^2 = 2\sum_{l-1}^{L} \beta_l^2 \sigma_l^{-4}, \tag{2}$$

under payload constrain

$$R = \sum_{l=1}^{L} H(\beta_l),$$

where $H(z) = -2z\log z - (1-2z)\log(1-2z)$ – is ternary entropy function; R – cover image payload in nats.

Minimization of (2) can be achieved by using the method of Lagrange multipliers. The change rate β_l and the Lagrange multiplier λ can be determined by numerically solving of further $(I+1)$ equations:

$$\beta_l \sigma_l^{-4} = \frac{1}{2\lambda} \ln\left(\frac{1-2\beta_l}{\beta_l}\right), l \in \{1,2,\ldots,L\},$$

$$R = \sum_{l=1}^{L} H(\beta_l).$$

4. Convert the change rate β_l to cost ρ_l:

$$\rho_l = \ln(1/\beta_l - 2); \quad (3)$$

5. Embed the desired payload R using syndrome-trellis codes (STCs) with pixel costs determined according to (3).

Applying the locally-estimated multivariate Gaussian cover model in MiPOD algorithm gives opportunity to derive a closed-form expression for the performance of the detector but, at the same time, complex enough to capture the non-stationary character of natural images [Sedighi et al, 2016].

Mentioned additive distortion functions use simple assumption that cost of not making a change is always zero. It does not take into account the influence of surrounding pixels on analyzed pixel's brightness value, which leads to underestimate the cover image distortion by message hiding. Therefore it was proposed to use the non-additive distortion functions for improving the empirical security of embedding schemes [Denemark and Fridrich, 2015].

In the work there was also investigated the case of usage the Synch scheme [Denemark and Fridrich, 2015] for improving the MiPOD embedding algorithm. The main steps of stegodata $\mathbf{m}$ embedding in in grayscale cover image $\mathbf{x}$ with size $M \times N$ (pixels) according to Synch-MiPOD algorithm are:

1. Divide message into two equal size parts:

$$\mathbf{m} = \mathbf{m}_1 \cup \mathbf{m}_2;$$

2. Compute the cost $\rho_{ij}, i \in \{1,2,\ldots,N\}, j \in \{1,2,\ldots,M\}$ from the cover image $\mathbf{x}$ according to formula (3);

3. Set stego image $\mathbf{y}$ is equal to cover image $\mathbf{x}$;

4. For each stego image pixel compute the cost of its modification in range $\Delta \in \{-1;0;+1\}$:

$$\rho_{ij}^{(+1)} = D_A\left(\mathbf{y}, x_{ij} + 1\mathbf{y}_{\sim ij}\right), \tag{4}$$

$$\rho_{ij}^{(0)} = D_A\left(\mathbf{y}, x_{ij}\mathbf{y}_{\sim ij}\right), \tag{5}$$

$$\rho_{ij}^{(-1)} = D_A\left(\mathbf{y}, x_{ij} - 1\mathbf{y}_{\sim ij}\right), \tag{6}$$

where

$$D_A\left(\mathbf{x},\mathbf{y}\right) = \sum_{x_{ij}\neq y_{ij}} D\left(\mathbf{x}, \mathrm{y}_{ij}\ \mathbf{x}_{\sim ij}\right)$$

is additive approximation of the distortion function

$$D\left(\mathbf{x},\mathbf{y}\right) = \sum_{((i,j),(k,l))\in\wp} S_{\wp}\left(x_{ij} - y_{ij}, x_{kl} - y_{kl}\right),$$

$$S_{\wp}\left(a,b\right) = \begin{cases} 0 & when \quad a = b, \\ A_{\wp} & when \quad |a| + |b| = 1, \\ \nu A_{\wp} & when \quad (a \neq b) \wedge \left(|a| + |b| = 2\right), \end{cases}$$

$\wp$ – index set of all two-pixels cliques formed by two vertically and horizontally adjacent pixels; $A_{\wp} = \left(\rho_{ij} + \rho_{kl}\right)/2$ – average clique cost; $\nu\,(\nu \geq 0)$ – parameter controlling the strength of penalizing desynchronized changes; $y_{ij}\mathbf{x}_{\sim ij}$ – shorthand for $\mathbf{x}$ in which only the (i,j) pixel x_{ij} was changed to y_{ij};

5. Embed i[th] element of message $\mathbf{m}_q$ into cover image, by taking into consideration the computed costs (4)-(6), with usage of STCs;

6. Repeat steps #4-5 q times ($q = 2$);

7. Repeat step #6 k times ($k \in \{1,2,\ldots,K\}$).

Embedding with different costs of all three possibilities $\{-1;0;+1\}$ requires the use of the so-called multi-layer STCs [Filler et al, 2011]. It should be mentioned that the costs A_{ij} are computed only once before the embedding starts and are kept the same throughout the embedding, i.e., they are not recomputed after every k sweep. Finally, the recipient reads the secret message using the same STCs applied to each sublattice and concatenating both parts.

Structural steganalysis of digital images

The most common approach to revealing the stego image is based on analysis the alteration of cover image's statistical characteristics, such as first-order statistics, second-order statistics and so on [Fridrich, 2009; Böhme, 2010]. There was proposed considerable number of powerful statistical steganalysis methods, based on applying the rich models if cover image in spatial (SPAM, SRM models) as well as JPEG (CC-PEV, CC-JRM models) domains. Despite of high accuracy the stego image detection, there is significant limitation of practical usage of mentioned methods, connected with great number of model's parameters, for instance 22,510 parameters for CC-JRM [Fridrich and Kodovský, 2012b] model, 34,671 parameters for SRM model [Fridrich and Kodovsky, 2012a]. It leads to sizeable increasing the stegdetector tuning and image processing times, which is inappropriate for real-time detection systems.

For overcome mentioned drawback of statistical steganalysis methods, there was proposed to use the powerful methods of digital image structural analysis, such as variogram analysis [Progonov and Kushch, 2014b], multifractal detrended fluctuation analysis [Progonov and Kushch, 2014a; Progonov and Kushch, 2015a] and multifractal analysis [Progonov and Kushch, 2015b].

Variogram analysis is widely used for investigation the correlation characteristics of time series $\mathbf{I}(\mathbf{s})$ and based on usage the variogram function [Cressie and Wikle, 2011]:

$$2\gamma_{\mathbf{I}}(h) = 2\left(C_{\mathbf{I}}(0) - C_{\mathbf{I}}(h)\right),$$

where $C_{\mathbf{s}}(h) = \mathrm{cov}\left(\mathbf{I}(\mathbf{s}), \mathbf{I}(\mathbf{s}+h)\right)$ – covariation of values the time series adjacent elements; h – time shift (lag). In most applications further approximation of variogram $2\gamma_{\mathbf{I}}(h)$ is used [Cressie and Wikle, 2011]:

$$2\hat{\gamma}_{\mathbf{I}}(h) = \frac{1}{|N_h|}\sum_{i,j\in N_h}\left(\mathbf{I}(s_i) - \mathbf{I}(s_j)\right)^2, N_h = \left\{(i,j): s_i - s_j = h\right\},$$

where N_h – set of possible pairs of position the elements, when distance between them is equal to h. Usage of variogram approximation $2\hat{\gamma}_{\mathbf{I}}(h)$ allows estimate such correlation characteristics of time series [Cressie and Wikle, 2011]:

1. Nugget-effect – the value of correlation between adjacent elements if time series:

$$N_{\mathbf{I}} = 2\hat{\gamma}_{\mathbf{I}}(h)\big|_{h=1},$$

2. Sill – the value of maximal variance the time series element's values:

$$S_{\mathbf{I}} = 2\hat{\gamma}_{\mathbf{I}}(h)\big|_{h\to+\infty},$$

3. Range – the interval of correlation between values of adjacent elements of time series:

$$R_{\mathbf{I}} = \max\left\{h : \left(1 - \frac{2\hat{\gamma}_{\mathbf{I}}(h)}{S_{\mathbf{I}}}\right) \geq \varepsilon_R\right\}, \varepsilon_R \in \mathbb{R}_+.$$

Value of range $R_{\mathbf{I}}$ usually is determined when correlation between adjacent elements is not less than 10% [Cressie and Wikle, 2011]. Despite of high accuracy estimation of image correlation parameters by usage of variogram analysis, this approach has limited opportunity to investigate the parameters of separate image components like intrinsic noise, contours etc. It requires applying the specialized processing methods, such as multifractal detrended fluctuation analysis (MF-DFA).

MF-DFA is generalization of well-known detrended fluctuation analysis and allows not only estimate the Hurst coefficient H values, but also investigate the multifractal nature of intrinsic noise of time series [Kantelhardt et al, 2002] – spectrum of generalized Hurst exponents $h(q)$ as well as multifractal spectrum $f_h(\alpha_h)$. Variation of scaling parameter q gives opportunity to estimate the generalized Hurst exponent $h(q)$ for time series element's value fluctuation with small ($q < 0$) and ($q > 0$) large amplitude. On the other hand, discrete values of multifractal spectrum $f_h(\alpha_h)$ correspond to Hausdorff dimension of the analyzed signal subset, which exponent of Hölder condition is equal to α_h. Values of α_h are varied between $\alpha_h = \alpha_h^{\min}$, which corresponds of signal components with minimal fluctuations between adjacent pixels, to $\alpha_h = \alpha_h^{\max}$, which corresponds to most "irregular" components.

Increasing of stegdetector performance requires improving the used model of cover source. Besides the widely used statistical and correlation characteristics of cover images, it is also of interest to include the

cover-specific features, such as fractality – preserving the statistical characters on the different scales [Peitgen et al, 2014]. Multifractal analysis allows extend the opportunity of "classical" fractal analysis – gives opportunity to investigate the fractal properties of image components with usage of spectrum the generalized fractal dimensions (Renie spectrum) D_q as well as multifractal spectrum $f(\alpha)$. Spectrum D_q allows not only estimate the Hausdorff dimensions of image components with various average brightness, in particular case minimal and maximal, which are correspond to D_q^{MIN} and D_q^{MAX}, but also information (D_1) and correlation (D_2) dimensions. The former characterizes the growth rate of the Shannon entropy given by successively finer discretizations of the space, while the latter is a measure of the dimensionality of the space occupied by a set of random points.

Variogram analysis, multifractal detrended fluctuation analysis and multifractal analysis of digital images was performed according to algorithms, described in [Progonov, 2016].

Multiclass stegdetector for digital images

Joint use of mentioned methods the structural steganalysis allows not only detect the stego images, but also carry out the forensic steganalysis – ascertains the domain, where message has been hidden, estimates the payload, and determines the processing chain of cover image as well as stegodata [Progonov, 2016]. Based on these results there was developed the multiclass stegdetector, which structural scheme is shown at Fig. 1.

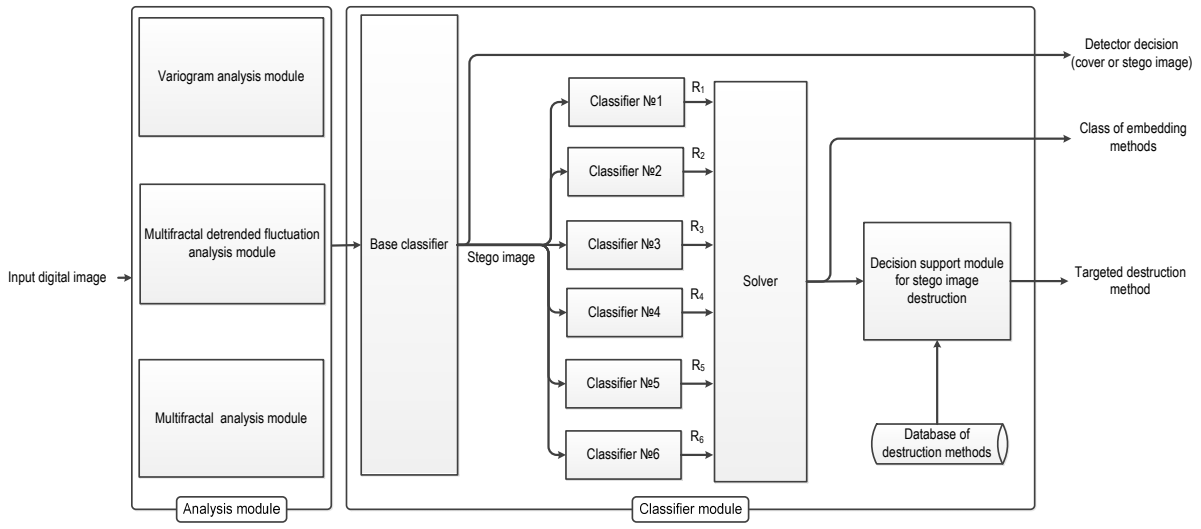

Figure 1. The flowchart of digital image processing by proposed generalized multiclass stegdetector

The stegdetector consists of two parts (Fig. 1) – analysis and classifier modules. Former module is subdivided into three modules, namely variorgam, multifractal detrended fluctuation and multifractal analyses, which are used for determination the statistical, correlation and fractal characteristics of inputted image. Obtained features are transferred to classifier module (Fig. 1). At first stage, base classifier map processing image to class of covers or stegos, depending on obtained feature values. If image is classified as containing the hidden messages (stego image, Fig. 1), additional classifier's submodules are applied for determine the class of steganographic techniques used for stego creation (Table 1).

Table 1. Classifiers for determination the class of used steganographic technique

Classifier number	Cover processing chain	Stegodata processing chain
1	One-stage, common transformation (Fourier, cosine or wavelet discrete transformations)	–
2	One-stage, uncommon transformation (for instance Singular Value Decomposition)	–
3	Two-stage, composition of common and uncommon transformations	–
4	Three-stage, composition of common and uncommon transformations	–
5	One-stage, common transformation (Fourier, cosine or wavelet discrete transformations)	Scrambling transformation
6	Two-stage, common transformation (Fourier, cosine or wavelet discrete transformations)	Scrambling transformation

Each classifier (Fig. 1) calculates the probability $P_i, i \in \{1,2,\ldots,6\}$ that analyzed image has been modified according to embedding method, belonging to corresponding class of steganographic techniques. Identification of most probable class the steganographic methods, used for creation of analyzed image, is carried out in decision support module by comparison of obtained probabilities P_i and determination of maximum probability P_i^{MAX}. According to decision of MCS there are also shown

recommendation for choosing the most effective (targeted) method for destruction the revealed stego image.

Experiments

For estimation the accuracy of stego image revealing by usage of proposed multiclass stegdetecor there were conducted the tuning and testing of MCS on test packet of 2,500 digital images from MIRFlickr-25k dataset [Huiskes and Lew, 2008]. Test packet was divided into training (1,250 images) and testing (1,250 images) subpacket in a pseudorandom manner. All images were scaled to the same size 512×512 pixels with usage of Lanczos kernel and saved in lossless JPEG format (Image Quality Factor is equal to 100%).

Payload of cover image was varied from 5% to 25% with step 5% and from 25% to 95% with step 10%.

Training of MCS was conducted with usage of image characteristics, obtained by applying the variogram analysis (39 parameters), multifractal detrended fluctuation analysis (182 parameters) and multifractal analysis (14 parameters). Estimation of mentioned features was carried out according to developed algorithms, represented in [Progonov, 2016]. Total number of used image features is equal to 235.

Testing of tuned MCS was repeated 10 times with reinitialize of training and testing subpackets. The averaged probabilities of cover and stego images attribution to steganographic technique's classes (Table 1) are shown in Table 2. For sake of convenience, the largest values of probabilities $P_i, i \in \{1,2,\ldots,6\}$ for each embedding methods are marked in thick print and underlined.

It should be mentioned that usage of proposed multiclass stegdetector allows correctly determine the cover image processing chain in case of usage the WOW and S-UNIWARD embedding methods (Table 2) – applying of common (two-dimensional discrete wavelet transformation) and specific (minimizing the distortion function value) processing methods. On the other hand, minor changes of WOW embedding scheme in HILL algorithm leads to misclassify the obtained stego images by MCS as formed according to simple embedding methods in frequency domain (class #1, please see Table 2). Obtained classification results for mentioned embedding methods remain permanent even in case of high cover image payload (more than 50%, Table 2).

In case of applying the modern adaptive steganographic schemes like MiPOD and Synch-MiPOD, multiclass stegdetector misclassify incorrectly classify obtained stego images as formed according to multistage embedding methods (Table 2), despite any cover transformations have not been applied. Misclassification of stego image in such case can be explained by disparity of used cover image model – multivariate Gaussian image model in MiPOD algorithm and union of Markov and fractal models in proposed MCS.

Table 2. Averaged probabilities of stego images attribution to considered steganographic technique's classes in case of low (10%) and high (85%) payload of cover image

Cover image payload	Embedding method	Steganographic technique's class					
		#1	#2	#3	#4	#5	#6
10%	WOW	0.39	0.17	***0.47***	0.28	0.09	0.08
	HILL	***0.42***	0.28	0.33	0.06	0.01	0.12
	S-UNIWARD	0.08	0.07	***0.58***	0.01	0.38	0.01
	MiPOD	0.02	0.15	0.34	***0.41***	0.22	0.19
	Synch-MiPOD	0.27	0.01	0.21	***0.46***	0.31	0.07
85%	WOW	0.63	0.41	***0.77***	0.22	0.11	0.18
	HILL	***0.91***	0.66	0.74	0.13	0.10	0.23
	S-UNIWARD	0.07	0.02	***0.98***	0.03	0.14	0.01
	MiPOD	0.01	0.28	***0.71***	0.68	0.07	0.11
	Synch-MiPOD	0.18	0.02	0.12	***0.89***	0.22	0.02

Conclusion

On the basis of conducted comprehensive analysis of performance the proposed multiclass stegdetector it is established that:

– It is confirmed the high efficiency of stegdetector even in case of investigation the stego images, formed according to a priory unknown embedding methods. Ability of stegdetector to determine the class of steganographic techniques, used for stego image creation, allows choose the targeted methods for hidden message destruction with minimal impact on cover image visual quality;

– Applying of adaptive embedding methods, based on usage the uncommon (multivariate Gaussian) cover model for stego image creation, allows significantly decrease the accuracy of it detection by usage of multiclass stegdetector. It is explained by usage of steganalytic "simplified" digital image model, which capture the most general features (fractality, correlation of brightness the adjacent

pixels) and has limited opportunity to represent the complicated local dependences in high-textured area of image. Overcome the revealed limitation requires creation the generalized image model for accurate capture the various features of real images, such as non-stationarity and heterogeneity.

Acknowledgement

The paper is published with partial support by the project ITHEA XXI of the ITHEA ISS (www.ithea.org) and the ADUIS (www.aduis.com.ua).

Bibliography

[Avcibas et al, 2003] Avcibas I., Memon N., Sankur B. Steganalysis using image quality metrics. IEEE Transactions on Image Processing. Volume 12, Issue 2, 2003. pp. 221–229. DOI 10.1109/TIP.2002.807363;

[Böhme, 2010] Böhme R. Advanced Statistical Steganalysis. Springer, 2010. 285 p. ISBN (eBook) 978-3-642-14313-7. ISBN (Hardcover) 978-3-642-14312-0. DOI: 10.1007/978-3-642-14313-7;

[Cisco, 2015] Cisco Systems, Inc., Annual Security Report. http://www.cisco.com/c/dam/assets/-about/ar/pdf/2015-cisco-annual-report.pdf;

[Cisco, 2016] Cisco Systems, Inc., Annual Security Report. http://www.cisco.com/c/dam/en_us/about/-annual-report/2016-annual-report-full.pdf;

[Cox et al, 2008] Cox I. J., Miller M. L., Bloom J. A., Fridrich J., Kalker T. Digital Watermarking and Steganography. Elsevier, 2008. 593 p.;

[Cressie and Wikle, 2011] Cressie N., Wikle C. Statistics for Spatio-Temporal Data. Wiley, 2011. 624 p.

[Denemark and Fridrich, 2015] Denemark T., Fridrich J. Improving Steganographic Security by Synchronizing the Selection Channel. Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security, 2015. DOI 10.1145/2756601.2756620;

[Farid, 2001] Farid H. Detecting Steganographic Messages in Digital Images. Technical Report. Dartmouth College Hanover, 2001. p.9;

[Filler and Fridrich, 2011] T. Filler, J. Fridrich. Design of adaptive steganographic schemes for digital images. Proceedings SPIE, Electronic Imaging, Media Watermarking, Security and Forensics III. Edited by A. Alattar, N. D. Memon, E. J. Delp, and J. Dittmann. 2011. pp. 1-14;

[Filler et al, 2011] T. Filler, J. Judas, J. Fridrich. Minimizing additive distortion in steganography using syndrome-trellis codes. IEEE Transactions on Information Forensics and Security. Volume 6(3), 2011. pp.920–935;

[FireEye, 2015] FireEye, Inc., HAMMERTOSS: Stealthy Tactics Define a Russian Cyber Threat Group. https://www2.fireeye.com/rs/848-DID-242/images/rpt-apt29-hammertoss.pdf;

[Fridrich, 2009] J. Fridrich Steganography in Digital Media: Principles, Algorithms, and Applications. 1st Edition. Cambridge University Press, 2009. p. 437. ISBN 978–0–521–19019–0;

[Fridrich and Kodovský, 2012a] Fridrich J., Kodovský J. Rich Models for Steganalysis of Digital Images. IEEE Transactions on Information Forensics and Security. Volume 7, Issue 3, 2012. pp. 868-882.

[Fridrich and Kodovský, 2012b] Fridrich J., Kodovský J. Steganalysis of JPEG images using rich models. Proceedings SPIE 8303, Media Watermarking, Security, and Forensics Edited by Memon Nasir D., Alattar Adnan M., Delp Edward J. doi:10.1117/12.907495;

[Holub and Fridrich, 2012] Holub V., Fridrich J. Designing Steganographic Distortion Using Directional Filters. Proceedings of IEEE Workshop on Information Forensic and Security. 2012;

[Holub et al, 2014] Holub V., Fridrich J., Denemark T. Universal distortion function for steganography in an arbitrary domain. EURASIP Journal on Information Security. 2014.

[Huiskes and Lew, 2008] Huiskes M.J., Lew M.S. The MIR Flickr Retrieval Evaluation. Proceedings of ACM International Conference on Multimedia Information Retrieval. 2008;

[Kantelhardt et al, 2002] Kantelhardt J. W., Zschiegner S. A., Koscielny-Bunde E., Bunde A., Havlin S., Stanley H. E. Multifractal detrended fluctuation analysis of nonstationary time series. Cornell University Library. Electronic Archive, 2002. https://arxiv.org/abs/physics/0202070;

[Ker et al, 2013] Ker A. D., Bas P., Böhme R., Cogranne R., Craver S., Filler T., Fridrich J., Pevný T. Moving steganography and steganalysis from the laboratory into the real world. Proceedings of the first ACM workshop on Information hiding and multimedia security (IH&MMSec '13). New York, 2013;

[Li et al, 2014] B. Li, M. Wang, J. Huang. A new cost function for spatial image steganography. Proceedings of IEEE International Conference on Image Processing (ICIP-2014);

[Peitgen et al, 2014] H.-O. Peitgen, J. Hartmut, D. Saupe. Chaos and Fractals. New Frontiers of Science. 2nd Edition. Springer, 2004. 864 p.;

[Pevný, 2010] Pevný T., TFiller T., Bas P. Using High-Dimensional Image Models to Perform Highly Undetectable Steganography. Proceedings of International Workshop on Information Hiding (IH 2010). Edited by Böhme R., Fong P.W.L., Safavi-Naini R. Springer, Berlin, Heidelberg;

[Progonov and Kushch, 2014a] Progonov D.O., Kushch S.M. Revealing of stego images with data, embedded in cover image transformation domain [In Ukrainian]. Bulletin of National Technical University of Ukraine. Series Radiotechnique. Radioapparatus Building. Vol. 57, 2014. pp. 128-142;

[Progonov and Kushch, 2014b] Progonov D.O., Kushch S.M. Variogram analysis of steganograms forme accrding to complex embedding methods. Bulletin of National Technical University of Ukraine "Lviv Polytechnic". Series Information systems and networks. Volume 806, 2014. pp.226-232;

[Progonov and Kushch, 2015a] Progonov D.O., Kushch S.M. Multifractal Detrended Fluctuation Analysis of steganograms. System research and information technologies. Volume 4, 2015. pp. 39-47;

[Progonov and Kushch, 2015b] Progonov D.O., Kushch S.M. Spectral analysis of steganograms. Radio Electronics, Computer Science, Control. Volume 2 (33), 2015. pp. 71-81;

[Progonov, 2016] Progonov D.O. Structural methods of digital image passive steganalysis. PhD Thesis. National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 2017. 293 p.;

[Sedighi et al, 2016] Sedighi V., Cogranne R., Fridrich J. Content-Adaptive Steganography by Minimizing Statistical Detectability. IEEE Transactions on Information Forensics and Security. Vol. 11, Iss. 2., 2016. pp. 221-234.

Authors' Information

Dmytro Progonov – *The Institute of Physics and Technology, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"; PhD, Associate Professor; 37, ave. Peremohy, Solomenskiy district, Kyiv, Postcode 03056, Ukraine; e-mail: progonov@gmail.com*

Major Fields of Scientific Research: digital media steganalysis, digital image forensics, machine learning, advanced signal processing