THE COMPARISON OF CLASSIFIERS IN IMAGE STEGANALYSIS

Martin BRODA, Vladimír HAJDUK, Dušan LEVICKÝ
Department of Electronics and Multimedia Communications, Faculty of Electrical Engineering and Informatics,
Technical University of Košice, Letná 9, 042 00 Košice, Slovak Republic, tel. +421 55602 2861,
e-mail: {martin.broda, vladimir.hajduk, dusan.levicky}@tuke.sk

ABSTRACT

In this paper, proposed steganalytic method utilized for the detection of secret message is based on extraction of statistical features from cover and stego images in JPEG file format together with calibration technique. The steganalyzer concept uses Support Vector Machines (SVM) classification or Bayes classifier for training a model that is later used by the same steganalyzer in order to identify between a clean (cover) and stego image. The aim of the paper was to compare detection accuracy (ACR) of the trained models for two types of classifiers: Support Vector Machines and Bayes classifier. In this paper, five models created between cover and stego images (images obtained by nsF5, Model Based 1, Model Based 2, Modulo Histogram Fitting with Dead Zone and Pertubed Quantization steganographic method) was tested.

Keywords: *steganalysis, image, classifier, Bayes, SVM, secret message*

1. INTRODUCTION

Steganography is the art of hiding secret information in unsuspicious data (cover data). More accurately, it deals with establishment subliminal channels and transporting confidential messages through it. While steganography was related with transfer of physical objects in the past, nowadays, is focused on transfer data in the digital form such as digital images, videos, audios and texts [1]. In the article still images were utilized.

The most popular method in the image steganography is LSB (Least Significant Bit). Secret message is embedded to least significant bits of either coded words. This substitution is performed in spatial or transformed domain. Steganographic methods utilized in the work are based on embedding information in DCT (Discrete Cosine Transformation) domain.

Steganalysis aims to detect the presence of hidden message inside apparently-innocent covers [2]. It is performed by in advance-trained model obtained in training phase of steganalytic process. Training phase requires high computational complexity than embedding process of steganography.

The method of mentioned training is machine learning. Machine learning is a science discipline that belongs to artificial intelligence. It is inspired by human learning system and gives this ability of self-learning to machines. Machine learning is utilized to solve two main problems: classification and sequential problems. The former deals with making a decision to classify some problem to one of certain classes. If these classes are presented in training, it is supervised learning. In addition, in sequential problems learner knows start and finish position only and seeks road to achieve that. In this case we discuss unsupervised learning [3].

If steganalytic technique is adapted to steganographic method and its characteristics then this technique can achieve higher efficiency in the process of detection. Such a system of steganalysis is called targeted steganalysis. On the other hand, there is a blind steganalysis. It has no information about used steganographic method. Blind steganalysis usually extracts more statistical features in spatial and transform domain for detection more than one steganographic tool. Even, it is appropriate to detect new not well-known algorithms, too. Both targeted and blind steganalysis extract features in training and testing phase of the process as well. Approach of steganalytic analysis together with extraction of the 274 statistical features was used in this paper.

The main part of steganalytic system is classifier. Classifier works in both training and testing phases of the steganalytic system. Classifier is able to put a testing object to the appropriate class using pre-calculated model in a training process. This work was aimed to comparison of efficiency of two well-known classifiers, SVM and Bayes classifier in specific tool of image steganalysis.

The paper is organized as follows. In Section 2, image steganalysis is described, including block diagram of testing and training phase. Descriptions of both tested classifiers are in the same section as well. In Section 3, experimental results are shown and the paper is concluded in Section 4.

2. IMAGE STEGANALYSIS

The steganalysis is scientific discipline and its primary function is detection of secret message in multimedia or detection of subliminal communication that is defined between two participants. If process of steganalysis is able to reveal secret communication, steganographic system is defined as broken and purpose of steganography is defeated. Steganalytic method is defined as successful, when stego image can be differentiated from cover image with higher probability as random guessing. Steganalysis can be supplemented by activity of extraction secret message's intelligence what requires a set of techniques for further analysis and increase of computational demands [5].

The main idea of steganalysis in static images is detection changes in statistic properties of cover image after embedding a secret message. Therefore, the calculation of those statistical features is very important in design of steganalytic method. The features distinguish the differ-

ence between stego and cover image and represent the input for classifier block as is illustrated in Figure 1.

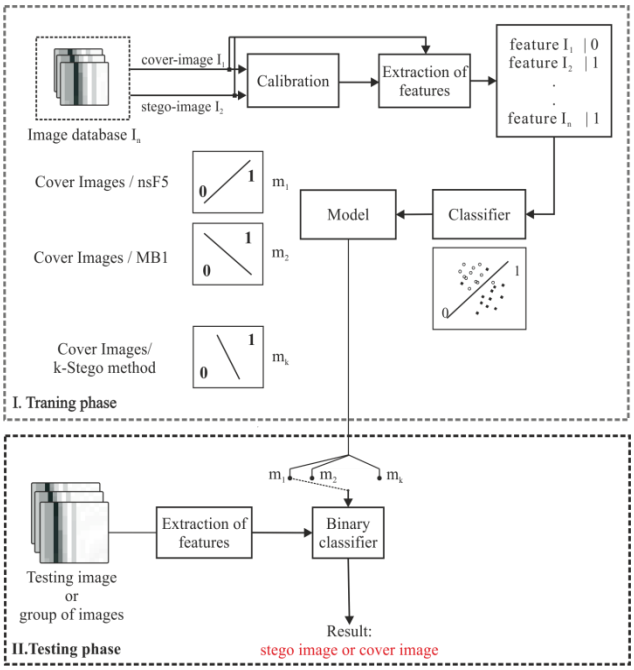

Fig. 1 Block diagram of proposed image steganalytic method

The image database consists of several thousand of images that were taken by different types of cameras using different camera's settings and resolutions. Stego images are created by embedding a secret message with several steganographic methods (e.g. nsF5 [6], MB [7] and others used in JPEG files). In next step, statistical features are extracted from stego or cover images, whereby we obtain two sets of statistical parameters that are separated according to identifier.

Proposed steganalytic method in this article includes 274 statistical features (reasons for the selection of these statistical parameters and more details are stated in article [8] :

- Global histogram from all *64 x* n_B (total blocks of image) DCT coefficients and local histograms in mode $(i, j) \in \{(1,2),(2,1),(3,1),(2,2),(1,3)\}$. The central part <-5,5> of this histogram was selected due to maximum energy situated on this interval. *(66 statistical features)*
- Dual histogram *(99 statistical features).*
- Functions of intra blocking dependencies of DCT coefficients – Variation. *(1 statistical feature)*
- Integral measures of intra blocking dependence. *(2 statistical features)*
- Functions from co-occurrence matrix *C* of neighboring DCT coefficients. *(25 statistical features)*
- Parameters of Markov model *(81 statistical features).*

Next block in steganalytic scheme is classifier, where input of classifier is set of statistical features calculated in previous step. Result of classification process is trained model between cover images and stego images that were obtained by specific steganographic method. This paper was focused on the comparison of two classifiers: SVM and Bayes classifier. Details of these classifiers are stated in chapter 2.1.

2.1. Classifiers

2.1.1. Support Vector Machines

Support Vector Machines (SVM), purposed by Vapnik [9], is method of machine learning which is used to classify linear separated or non-separated problems. Based on input data, SVM computes parameters of the separated hyperplane to classify data to appropriate class. Problem of the training model is to find this optimal border by witch cover and stego characteristic features are divided. (Figure 2)

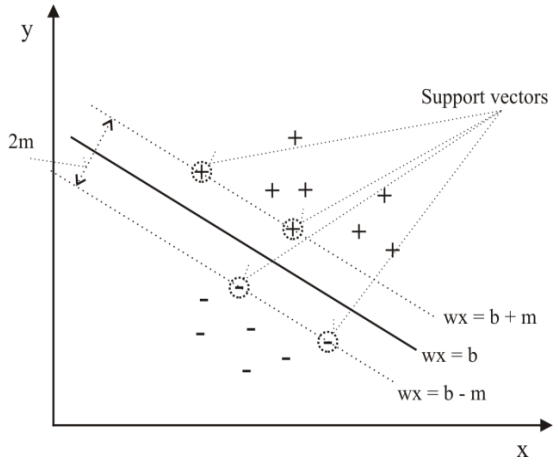

Fig. 2 Linear separated problem classified by SVM

Optimal separated hyperplane is defined as [10]:

$$wx = b \tag{1}$$

where: x – input vector, w – vector of weighting coefficients, b – offset. Hyperplane is situated in the middle of range $2m$, given by support vectors.

Upper-mentioned case is for linear separated problems. If the problem is not linearly separated, input vector is transformed to space with more dimensions. It is achieved using a kernel function (see Figure 3) [11].

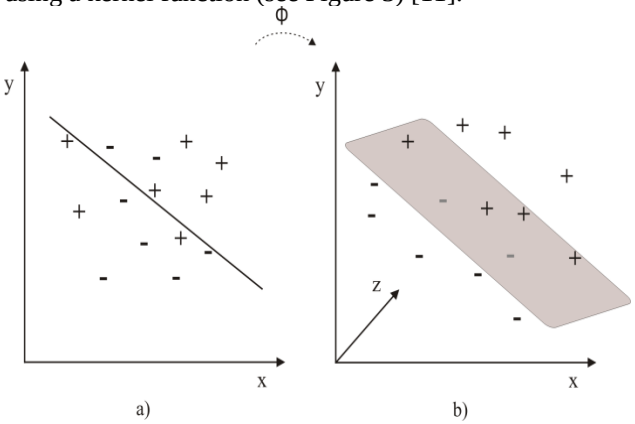

Fig. 3 Linear non-separated problem (a), Transformation to multidimensional space (b)

Now, the classifier searches for optimal separated plane in multidimensional space. Separated hyperplane in multidimensional space is defined:

$$w\Phi(x) = b \tag{2}$$

where: Φ(x) – transformation of vector *x* to multidimensional space by the kernel function.

2.1.2. Naive Bayes Classifier

In machine learning, naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. Naive Bayes has been studied extensively since the 1950s [12].

Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers.

The idea behind a Bayesian classifier is that, if an agent knows the class, it can predict the values of the other features. If it does not know the class, Bayes' rule can be used to predict the class given the feature values. In a Bayesian classifier, the learning agent builds a probabilistic model of the features and uses that model to predict the classification of a new example.

An advantage of naive Bayes is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix [12].

The classifier works as follows. Given a set *D* of n dimensional vectors $x\ (x_1, x_2, x_3, ..., x_n)$, and m classes : $C_1, C_2, ..., C_m$ the Naive Bayes classifier predicts *x* belongs to class C_m if:

$$P(C_i|x) > P(C_j|x) \tag{3}$$

for all *j* between *1* and *m*. The above conditional probability can be expressed using the Bayes theorem:

$$P(C_i|x) = \frac{P(x|C_i)P(C_i)}{P(x)} \tag{4}$$

As *P(x)* is constant, eq. 4 reduces to maximizing:

$$P(x|C_i)P(C_i) \tag{5}$$

While this approach is computationally expensive for large *n*, to ease the burden "class conditional independence" is assumed resulting in:

$$P(x|C_i) = \prod_{k=1}^{n} P(x_k|C_i) \tag{6}$$

3. EXPERIMENTAL RESULTS

Our database contained 18 000 real images taken by different camera types (Nikon D3200, Nikon D3100, Nikon D3000, Nikon D60, Olympus FE-115, Olympus X-715, Panasonic Lumix DMC-FZ5, Samsung S730, Samsung Galaxy ACE, Sony Ericsson C702 and Sony Ericsson W580). This image set included images with various quality and resolution, whereby pictures were taken in different light conditions and various scenes. Created database was divided into two categories: training and testing part.

The selection was implemented on the basis of specific cameras in order to preserve the maximum of diversity. The 2000 images were selected from the group of training images and the 200 pictures were chosen from testing database. Image spatial resolutions were modified because of higher diversity. The image database included these types of resolutions: 320×240 (QVGA), 480×320 (HVGA), 640×480 (VGA), 800×600 (SVGA), 1024×768 (XGA), 1600×1200 (UXGA) and 1920×1080 (HD 1080). In training process, the secret message was embedded using five steganographic tools (nsF5 [6], MB1 [7], MB2 [7], MHF-DZ [13] and PQ [14]).

The image database was divided into 8 parts with the 250 images so that the every part included all image resolutions, all types of camera, etc. Consequently, the secret message with variable length was inserted into images in specific groups (8 different sizes of secret message). Variable size of the secret message was applied because of increase in the sensitivity of the trained model. The size of secret message was expressed using parameter *Paylod* [%] (payload 100 % explains maximal size of the secret message for specific steganographic tool). This process was repeated for every tested steganographic method.

After embedding, the final database consisted of the 2000 images (for every steganographic tool) for the feature extraction. Consequently, these statistical parameters represented the input for classifier. A part of features' extraction is calibration technique that performs cropping of picture by 4 pixels in each direction. The calibrated image has very similar statistical features to cover image. The calibration was executed on image database in order to acquire difference statistics of DCT coefficients what means a feature vector. In training phase, there were created steganalytic models for binary classification, e.g. *model cover – nsF5 stego images, cover – MB1 stego images etc.* for every tested steganographic method.

In testing phase, there was realized experiment for the verification of detection accuracy of created models for specific steganographic methods using L-SVM classifier or Bayes classifier. L-SVM classifier was used in configuration with linear kernel function and Naive Bayes classifier was tested with the normal Gaussian distribution.

The steganalyzer performance is highly susceptible to embedded data rate. The tested steganographic methods possess with non-equal embedding capacity what did not allow us to show comparable results of final detection accuracy for all values of the secret messages.

Table 1 shows Accuracy (*ACR*) and True Positive Rate (*TPR*) of the trained model for different algorithms, payloads and classifiers.

Table 1 Accuracy (ACR) and True Positive Rate (TPR) of trained model for different algorithms, payloads and classifiers

Testing algorithm	Payload	L-SVM		Bayes	
		TPR	ACR	TPR	ACR
nsF5	25%	0,63	**0,74**	0,49	**0,61**
	50%	0,94	**0,89**	0,91	**0,82**
	75%	0,98	**0,91**	0,91	**0,82**
	100%	1	**0,92**	0,9	**0,81**
MHF-DZ	25%	0,51	**0,61**	0,44	**0,56**
	50%	0,54	**0,62**	0,46	**0,57**
	75%	0,6	**0,65**	0,48	**0,58**
	100%	0,72	**0,71**	0,54	**0,61**
MB1	25%	0,75	**0,77**	0,65	**0,68**
	50%	0,9	**0,84**	0,76	**0,74**
	75%	0,96	**0,87**	0,84	**0,78**
	100%	1	**0,89**	0,88	**0,8**
MB2	25%	0,84	**0,81**	0,67	**0,7**
	50%	0,92	**0,85**	0,83	**0,78**
	75%	0,98	**0,88**	0,89	**0,81**
	100%	1	**0,89**	0,92	**0,82**
PQ	25%	0,95	**0,97**	0,97	**0,93**
	50%	0,94	**0,97**	0,94	**0,91**
	75%	0,91	**0,95**	0,92	**0,9**
	100%	0,91	**0,95**	0,92	**0,9**

Table 1 and Figure 4 show that the better accuracy of detection was achieved using SVM classifier for all types of tested steganographic tools. On the other hand, Bayes classifier had advantage in a smaller computational complexity and smaller time required for training of model. For example, Bayes classifier was able to perform training of *model Cover - MB2* with the 2000 images in less than 30 seconds. On other hand, SVM classifier achieved training time: 10 minutes in the same case. Specific test was executed using Intel Core i5 processor with the clock frequency 2,5 GHz.

The best results of *ACR* in testing process were attained for the model trained between cover and PQ stego images. On the other hand, model based on steganographic method MHF-DZ achieved the smallest level of detection accuracy.

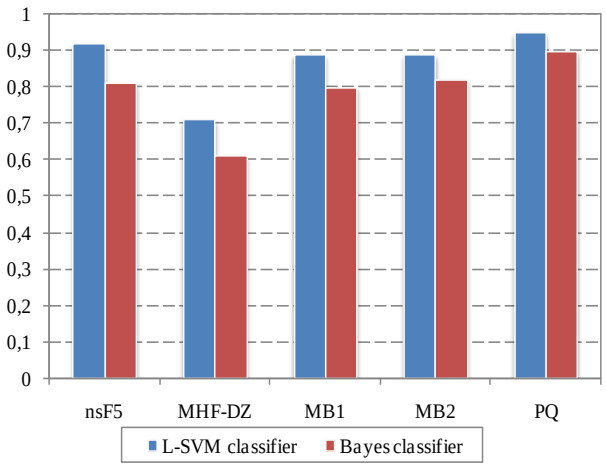

Fig. 4 Comparison of accuracy for specific steganalytic model using L-SVM or Bayes classifier

Characteristics of created steganalytic models can be also illustrated using ROC (Receiver Operating Characteristic) curve. The basic parameter of this curve is AUC (Area under Curve). The AUC has a value from 0 to 1 and the higher value of *AUC* explains the better detection properties of the specific model. Authors in article [15] show, both empirically and formally, that *AUC* is indeed a statistically consistent and more discriminating measure than accuracy; what means that *AUC* is a better measure than accuracy for evaluating of learning algorithms. The Figure 5 and Figure 6 illustrate ROC curves for specific models of steganographic methods with SVM and Bayes classification for maximal capacity of secret message and for every tested steganographic method.

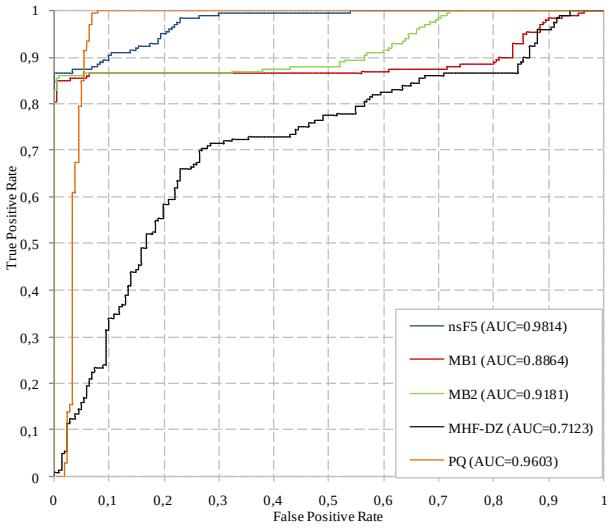

Fig. 5 ROC curves of specific steganalytic models with L-SVM classifier for Payload=100%

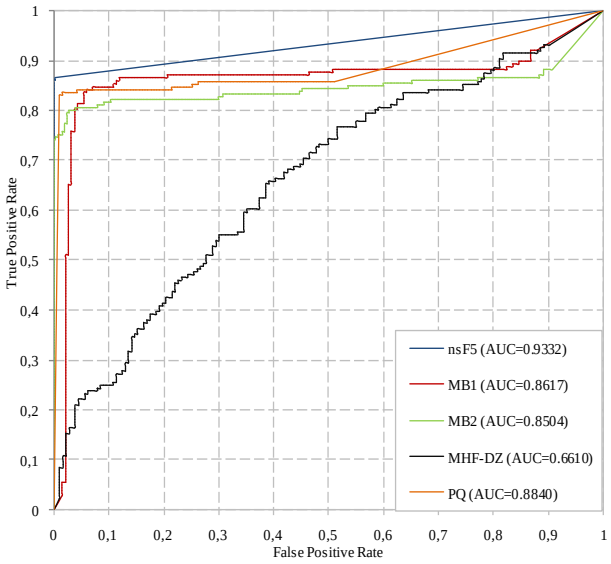

Fig. 6 ROC curves of specific steganalytic models with Bayes classifier for Payload = 100%

4. CONCLUSIONS

In this paper, universal steganalytic system with two different types of classifier was performed. The goal was to bring a comparison in accuracy of both classifiers for detection of five steganographic methods. Evaluating parameters were *ACR*, *TPR* and *AUC*. From the view of *ACR*, the support vector machines (L-SVM) achieved better performance of detection for all steganalytic mod-

els. *TPR* of the models was similar. Only for *model Cover-PQ* the value was higher in favour Bayes classifier. It was due to better detection of positive (stego) images against L-SVM. Comparison of the view of *AUC* was totally in favour L-SVM for all models, where the *model Cover-nsF5* reached the highest and *model Cover-MHF-DZ* the lowest detection for both classifiers.

ACKNOWLEDGMENTS

Paper was the result of the Project implementation: University Science Park TECHNICOM for Innovation Applications Supported by Knowledge Technology, ITMS: 26220220182, supported by the Research & Development Operational Programme funded by the ERDF [50%] and Ministry of Education of Slovak Republic VEGA Grant No. 1/0075/15 [50%].

We support research activities in Slovakia/this project is being co-financed by the European Union.

REFERENCES

[1] LUO, X. – LIU, F. – LIAN, S. – YANG, C. – GRITZALIS, S.: "On the Typical Statistic Features for Image Blind Steganalysis." In: Selected Areas in Communications, IEEE Journal on, vol. 29, no. 7, August 2011, pp. 1404-1422.

[2] KER, A.D. – PEVNÝ, T., "The Steganographer is the Outlier: Realistic Large-Scale Steganalysis." In: Information Forensics and Security, IEEE Transactions on, vol. 9, no. 9, Sept. 2014, pp.1424-1435.

[3] MACHOVÁ, K.: Machine learning. (In Slovak), Košice: Department of Cybernetics and Artificial Intelligence, Technical University-FEI, 2002.

[5] FRIDRICH, J. – GOLJAN, M. – HOGEA, D.: "Steganalysis of JPEG images, breaking the F5 algorithm." In: Information Hiding, 5th International Workshop, Volume 2578 of Lecture Notes in Computer Science, Noordwijkerhout, the Netherlands, Springer-Verlag, New York (2002).

[6] FRIDRICH, J. – PEVNÝ, T. – KODOVSKÝ, J.: Statistically undetectable JPEG steganography: Dead ends, challenges, and opportunities. In J. Dittmann and J. Fridrich, editors, Proceedings of the 9th ACM Multimedia & Security Workshop, pages 3–14, Dallas, TX, September 20–21, 2007.

[7] SALLEE, P.: "Model-based methods for steganography and steganalysis," International Journal of Image and Graphics, p. 167-190, 2005.

[8] MAJERČÁK, D. – BÁNOCI, V. – BRODA, M. – BUGÁR, G. – LEVICKÝ, D.: "Performance Evaluation of Feature-Based Steganalysis in Steganography," 23th International Conference Radioelektronika 2013, Pardubice, April 2013, pp. 377-381.

[9] VAPNIK, V.: "The Nature of Statistical Learning Theory." New York: Springer 1995.

[10] BENNETT, K. P. – BLUE, J. A.: "A Support Vector Machine Approach to Decision Trees," Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on, vol. 3, May 1988, pp.2396, 2401 vol.3, 4-9.

[11] DE BIE, T. – CHRISTIANINI, N.: "Kernel Methods for Exploratory Pattern Analysis: a Demonstration on Text Data." Structural, Syntactic, and Statistical Pattern Reconstruction Lecture Notes in Computer Science, vol. 3138, 2004, pp. 16-29.

[12] CARUANA, R. – NICULESCU-MIZIL, A.: "An empirical comparison ofsupervised learning algorithms," Proceedings of the 23rd international conference on Machine learning, ACM, 2006, p. 168.

[13] BÁNOCI, V. – BUGÁR, G. – LEVICKÝ, D. – KLENOVIČOVÁ, Z.: "Histogram secure steanography system in JPEG file based on modulus function," Radioelektronika 2012, 22nd International Conference, Brno, pp. 263-266, 2012.

[14] FRIDRICH, J. – GOLJAN, M. – SOUKAL, D. "Perturbed quantization steganography with wet paper codes," in Proc. ACM Multimedia Security Workshop, Magdeburg, Germany, Sep. 20–21, 2004, pp. 4–15.

[15] HUANG, J. – LING, CH.: Using AUC and Accuracy in Evaluating Learning Algorithms, IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 3, March 2005, pp. 299-310.

Received January 9, 2015, accepted February 9, 2015

BIOGRAPHIES

Martin Broda was born in Prešov, Slovak republic in 1988. He received his M.Sc. degree from Faculty of Electrical engineering and Informatics, Technical University in Košice. Nowadays, he is a PhD. student at Department of Electronics and Multimedia Communications, focusing on multimedia security, image steganography, steganalysis and digital watermarking.

Vladimír Hajduk was born in Košice, Slovak Republic in 1990. He received his M.Sc. degree from Faculty of Electrical Engineering and Informatics, Technical University in Košice and now he is a PhD. student at the Department of Electronics and Multimedia Communications at the same faculty. His research interests include multimedia security, image processing, image steganography and steganalysis.

Dušan Levický was born in Slanec (Slovak Republic) in 1948. He received his M.Sc. and PhD. degrees at Technical University (TU) in Košice and now he is professor at Department of Electronics and Multimedia Communications, TU in Košice. His research interests include digital image processing, image transmission and cryptography.