

Towards Higher Detection Accuracy in Blind Steganalysis of JPEG Images

Mehdi Zohourian, Morteza Heidari, Shahrokh Ghaemmaghami, and Iman Gholampour
Institute of Communication Acoustics, Ruhr-Universität Bochum, Bochum, Germany
Electrical Engineering Department and Electronics Research Institute,
Sharif University of Technology, Tehran, Iran

mehdi.zohourian@rub.de, m_heidari@alum.sharif.edu, ghaemmag@sharif.edu, imangh@sharif.edu

Abstract—A new steganalysis system for JPG-based image data hiding is proposed in this paper. We use features extracted from both wavelet and DCT domains that are refined later in the sense of utmost discrimination between the clear and stego images in the classification system. Statistical properties of the SVD of wavelet sub-bands are combined with the extended DCT-Markov features, and the features that are most sensitive to the data embedding are chosen through a SVM-RFE based selection algorithm. Experimental results show significant improvement over baseline methods, especially for steganalysis of Perturbed Quantization (PQ), which is known to be one of most secure JPG-based steganography schemes, with 90.5% average detection accuracy at low embedding rates..

Keyword; image steganalysis, steganography, wavelet, Singular Value Decomposition, DCT.

I. INTRODUCTION

Blind image steganalysis is a technique used for detecting the existence of the data hidden in an image, where no information about the stenographic algorithm is available or usable. Principally, stenographic methods, on one hand, attempt to make minimal changes to the statistical characteristics and the perceptual contents of the cover image. This is while the steganalysis approaches, on the other hand, aim at detection of even minor alterations happen to the cover signals, due to the data embedding process. Steganalysis methods are based on extraction and examination of features sensitive to the hiding algorithms. This is the essential game-like battle between steganography and steganalysis.

Early steganalysis approaches, such as chi-square attack [1], and Ker's method [2], employed the general shape of the image histograms and some first order statistics of the image as features for steganalysis. Later, higher-order statistics of the cover image attracted attention of steganalyzers, e.g. Lyu and Farid [3] used higher order statistical moments of wavelet coefficients as the feature set or WAM steganalyzer [4] whose features were chosen to be higher order absolute moments of stego-signal estimated in wavelet domain. Some recent steganalysis methods are concerned of correlation in spatial

domain, between pixels, as well as the correlation between coefficients of the image in the transform domain, often in the DCT (Discrete Cosine Transform) and wavelet domains. For instance, Gul [5] used singular value decomposition (SVD) of the suspicious image as the feature set and reached a good result for steganalysis in spatial domain. Pevny and Fridrich proposed a steganalyzer for JPEG images in [6], which merged the extended DCT and Markov features that improved the detection scores in steganalysis of some selected steganography methods. This method, however, failed to detect some well-known data hiding approaches like PQ (Perturbed Quantization) steganography at lower embedding rates.

Furthermore, another method for blind image steganalysis is presented in [7] which uses statistical moments of the SVD components that are extracted from the DCT coefficients. For the DCT calculation, this paper uses a method different from conventional DCT calculation on image signals, which makes the steganalysis method more sensitive to data hiding.

In this paper, a universal steganalyzer for JPG images is proposed that uses the features extracted from both wavelet and DCT domains. Statistical features extracted from Wavelet Singular Value Decomposition (WSVD) of image are combined with 274-D DCT features selected and employed in [6]. This combined feature vector is ranked using Support Vector Machine Recursive Feature Elimination (SVM-RFE). Then a 160-D feature set of most effective features is fed to a SVM classifier to discriminate between clean and stego images.

The paper is organized as follows. In the next section, we introduce the WSVD image steganalysis method. Section III reviews the merged extended DCT and Markov features given in [6], explains the relation between WSVD features and those in [6], and extracts optimal features based on the SVM-RFE. Experimental results are presented and discussed in Section IV and a conclusion is given in section V.

II. WAVELET SINGULAR VALUE DECOMPOSITION (WSVD) STEGANALYSIS

Multi-resolution based image representation is one of thriving approaches to image steganalysis used so far. It has been shown earlier that, due to the energy compactness and decorrelation properties of the wavelet transform, most of the changes to the image statistics made by the data embedding could be caught in the wavelet domain [3]. However, despite such decorrelation properties, the wavelet transform retains some correlation between sub-bands in different scales [3]. This is something undesirable in steganalysis, as detecting the alterations made to the statistics of the image by the hidden message is significantly affected by the correlation between the selected features. Accordingly, it is believed that using a method to reduce the aforementioned remaining correlation between the wavelet coefficients can improve the detection rate in steganalysis, particularly at low embedding rates. Methods proposed earlier, based on log prediction error [3], Markov transition probability matrix [8], and co-occurrence matrix [9], have already been used for this purpose. These methods, however, considerably increase the dimension of the feature sets and hence the time needed to process the features.

In fact, the change happens to the statistical structure of the image, following the embedding process, arises from the general dissimilarities between the statistics of the image and that of the embedded message. In most cases of image steganography, the message is pseudo-randomized using a pseudo-random generator prior to the embedding. The cover image is rather composed of visually perceptible objects, hence far from random. This statistical/perceptual difference is the main basis of discrimination between innocent and stego signals in a steganalysis system. Here, we use Singular Value Decomposition (SVD), which is a matrix factorization tool based on eigenanalysis, to examine the correlation between the wavelet coefficients. In particular, SVD is employed to determine rank of the matrix of wavelet coefficients, which is expected to have an ascending change when a random message is embedded to the image.

To analyze the embedding process, it is assumed that the hidden message is an additive, zero-mean, white Gaussian iid, z , embedded in the cover signal, s , to build the stego signal, x . This is shown as:

$$x = s + z \quad (1)$$

Where z is independent of s .

By applying the wavelet transform to both sides of (1), we get:

$$X = S + Z \quad (2)$$

where X , S , Z are the matrices of the wavelet transform of x , s and z , respectively. Assuming r to be the rank of matrix S , we have:

$$\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_r \geq \alpha_{r+1} = \alpha_{r+2} = \dots = \alpha_n = 0 \quad (3)$$

Where α_i ($1 \leq i \leq n$) are SVD values.

According to the iid property of matrix Z with σ^2 variance, we can write:

$$E\left(\frac{ZZ^t}{n}\right) = \sigma^2 I_m \quad (4)$$

where I_m is m -dimensional identity matrix.

Therefore, it is easy to show that:

$$E\left(\frac{XX^t}{n}\right) = E\left(\frac{SS^t}{n}\right) + \sigma^2 I_m \quad (5)$$

The SVD decomposition of $m \times n$ matrix S is

$$S = U \Xi V^T \quad (6)$$

Where U and V are two column-orthogonal matrixes, and Ξ is a diagonal matrix with elements α_i ($1 \leq i \leq n$) as mentioned in (3). Therefore, according to equation (5) we have:

$$SS^t = U \Xi^2 V^T \quad (7)$$

So, for stego signal X , it can be shown that:

$$X \approx U \left(\Xi^2 + n\sigma^2 I_m \right)^{\frac{1}{2}} V^T \quad (8)$$

It is clear that the singular values $\alpha_{r+1}, \alpha_{r+2}, \dots, \alpha_n$ will become nonzero. Therefore, the rank of matrix will be increased. This shows that the SVD components of a matrix (image) is quite sensitive to data embedding that can serve as a sign of the hidden contents in steganalysis. In addition, by using the SVD of the wavelet coefficients, the number of resulting singular values is less than the number of wavelet coefficients, leading to reducing the number of calculations.

In this work, we first decompose image into three levels through Haar wavelet transform, and decompose the first-scale diagonal sub-band to enhance performance of the system. Next, singular values of each sub-band are calculated. Subsequently, we extract five typical features from the resulting 16 vectors. These features include the first three statistical moments, logarithm of geometrical mean, and the condition number that is the ratio of maximum singular value to the minimum one. Consequently, there are 80 features used for this part of the steganalysis. As mentioned earlier, embedding data in an image increases rank of the matrix of image. Therefore, distributions of the singular values before and after embedding are expected to be different. This was also verified by our experiments conducted over a large number of images. Fig.1 illustrates distributions of singular values of image Lena before and after the data embedding.

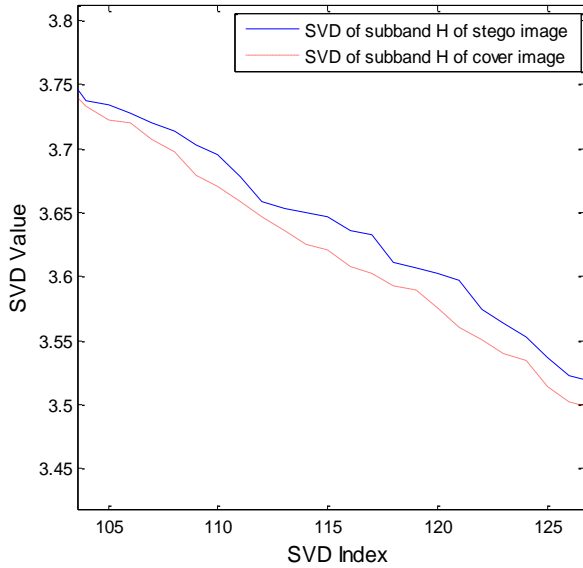


Fig.1. SVD distribution of sub-band H of clean and stego images of Lena damaged by PQ method (the vertical axis is log scaled).

As shown in Fig.1, singular values with higher indices are increased, while those with lower indices do not change much. These differences could be caught by some statistical analysis techniques using mean, variance, skewness, and/or the condition number. In order to intensify the effect of higher indices singular values compared to the lower indices ones, the logarithm of geometrical mean is used as another feature set for extracting the statistical differences of the SVD values [5].

These 80 features, called WSVD features, are combined with 274 extended DCT and Markov features, as introduced in [6]. The resulting feature set is ranked by the SVM-RFE classifier. It is shown that 160 high-ranked features, among the total 354 features, can often yield a higher detection rate in steganalysis of JPEG images with a lower computational complexity, as compared to the baseline methods.

III. COMBINE WSVD WITH EXTENDED DCT AND MARKOV FEATURES

Pevny and Fridrich [6] introduced a steganalysis method for JPEG images based on merging two groups of features; the DCT features and the calibrated version of Markov features described in [8]. Both of them result in a 274-D feature set. This 274-D feature set is one of the most competent one used for steganalysis of JPEG images, so far, to attack some well-known steganographic algorithms such as F5, Model Base (MB), Outguess and Steghide. However, they have not been found efficient for detecting more robust steganography methods like the PQ.

We combine the WSVD features with the abovementioned 274-D features to improve the detection accuracy of some well-known steganography methods that has not been detected precisely so far. In fact, there are especial properties in each

group of these features which provide a significant improvement to the detection accuracy of the steganalysis method, when they are combined together. In case of JPEG steganography, in which the DCT coefficients of the image are manipulated, still the statistical properties of the wavelet sub-bands could be altered as well. However, it has been shown that using the DCT features in the feature set, in addition to the wavelet features, makes the steganalysis system more sensitive to data embedding in the DCT domain [10]. Apparently, the question about this fusion of features is how these two feature sets should be combined.

• Feature selection

One of the most important issues in steganalysis is feature selection. There are some methods used for this purpose such as analysis of variance (ANOVA) used in [11] to select statistically significant features. However, this method neglects the interaction between features and considers the statistical significance of individual features [10]. Recently, an efficient method for feature selection based on support vector weights has been proposed in [10], which is called SVM Recursive Feature Elimination (SVM-RFE). This method is utilized in this paper in order to rank the 354 features of combined WSVD and extended DCT-Markov features. The high-ranked features are then fed into the SVM classifier to classify clean and stego images of several steganography methods like LSB and $\text{LSB} \pm 1$ embedding, Steghide [13], F5 [14], PQ (Perturbed Quantization) [15], and MB1 (Model Based) [16].

For SVM-RFE step, the goal is to maximize the sensitivity of the steganalysis method to all mentioned steganography schemes, so we use those different steganographic methods to create stego images class. Additionally, we keep the embedding rate within the target range of low rates to let the SVM-RFE algorithm prioritize the most sensitive features at the selected low rates.

To improve the time needed to classify cover and stego images, we disregard features with higher indices, which may not give especial results in the detection accuracy rates. Doing so experimentally on different algorithms, we found features with indices more than 160 do not impact the detection accuracy specifically. These 160 features comprised 27% of WSVD features in addition to 73% of extended DCT-Markov features which most of them are calibrated Markov features. These features are selected in order to feed to the SVM classifier with the RBF kernel. The RBF kernel is selected because of its localized and finite responses over the entire range of the real x-axis.

IV. EXPERIMENTS AND DISCUSSIONS

We use 2000 images of different kinds taken from CorelDraw image database [12]. All images are converted to gray level images of the size 512×512 and JPEG compressed by quality factor 80. 1000 images of this dataset

are used as clean images and the other 1000 images are used to generate stego images employing the chosen steganographic methods. To assess the proposed method, six typical steganography schemes, including LSB and $\text{LSB} \pm 1$ embedding, Steghide [13], F5 [14], PQ (Perturbed Quantization) [15], and MB1 (Model Based) [16] methods with three embedding rates are used.

Results of the experiments are listed in Table I. For an overall comparison, results are compared to methods using Markov features [8] and 274-D features of Pevny and Fridrich [6], and blind steganalysis method that is presented in [7]. To have a fair comparison, all methods are simulated on the database under same condition. Results for each method are an average of ten different iterations to make them more reliable. For each iteration, images for training and test steps are chosen randomly. 1300 images are used in training step. Steganalysis systems are tested on the other 700 images of the dataset. As shown in table I, significant improvement is achieved using the proposed method, especially for steganalysis of the PQ algorithm, as compared to results using the methods given in [6][7] [8].

TABLE I. COMPARISON OF DETECTION RATES (%) USING THE PROPOSED AND THE TWO REFERENCE STEGANALYSIS METHODS.

Data Hiding Method	Embed Rate (%)	Markov based [8] (%)	Method in [7] (%)	F-274 [6] (%)	Proposed Method (%)
LSB	10	99	99.3	99.1	99.5
	30	99.2	99.3	99.5	99.6
	70	99.5	99.5	99.6	99.8
$\text{LSB} \pm 1$	10	98.5	98.3	98.8	99.4
	30	98.9	99.1	99.1	99.4
	70	99.3	99.4	99.5	99.6
STEGHI DE	10	87.5	89.9	91.3	98.75
	30	89.8	92.7	94.2	99.3
	70	91.0	95.3	97.0	99.4
F5	10	87.5	89.6	95.1	98
	30	90.8	95.5	96.8	98.2
	70	99.0	99.2	99.5	99.5
PQ	10	61.4	68.6	78.0	85.0
	30	68.0	73.7	81.0	89.5
	70	72.0	84.5	87.1	97.0
MB1	10	99	92.9	99.1	99.5
	30	99.2	94.5	99.5	99.5
	70	99.6	99.6	99.6	99.6

To examine these results precisely, the ROC curves of the proposed method and its counterparts: SVBS [11] and Empirical Matrix [17], which are the specific steganalyzers of the PQ steganography, also the 274-D features based method

of Pevny and Fridrich [6], and SVD-DCT method given in [7] are computed and presented in Fig. 2, for low rate PQ steganography. To draw ROC curve generalized eigenvector from training step is used on the results of test step. As indicated in this figure, the proposed method yields higher detection accuracy at different embedding rates, as compared to its competitors.

The proposed method has also been evaluated for steganalysis of images that have undergone the data hiding at low embedding rates, which is known as a challenging issue in steganalysis. The ROC curves for different steganography methods at 10% embedding rate are depicted in Fig. 3, which indicates considerable superiority of the proposed method over the other illustrated methods for detecting the PQ steganography at low embedding rates. The reason is that most of JPG steganalysis methods as well as those included in [6][8][17] consider only the statistical properties of DCT coefficients as features for classification, whereas, not only does the dependencies of DCT parities of the image are changed through PQ algorithm, the statistical properties of wavelet sub-bands both in low and high frequencies will be influenced as well. Therefore, combining these two groups of features may detect the statistical changes made to DCT coefficients by PQ algorithm more precisely rather than using just one group. The other reason that is worth considering is that in the proposed method we use SVM-RFE to select those features that are sensitive enough to distortion caused by different data hiding methods. This algorithm ranks elements of the feature vector in a string. Those features that are at the end of this string do not change under data embedding specially at low rates, so they have a negative influence on accuracy of the classifier. Furthermore, the SVM-RFE algorithm significantly reduces the feature set dimension (from 354-D to 160-D in our case), which helps to boost the classification speed.

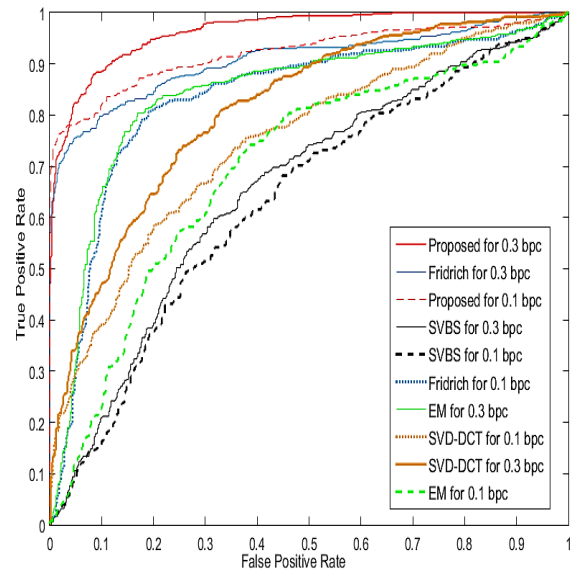


Fig. 2. ROC curves for steganalysis of PQ steganography.

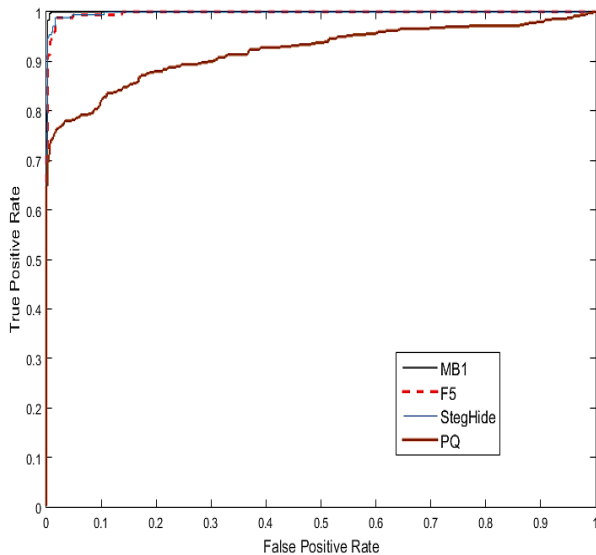


Fig.3. ROC curves of detecting steganography methods at 10% embedding rate using the proposed steganalyzer.

The idea of combining wavelet features with DCT ones has also been proposed in [10]; however, this method considers only features extracted from low frequency wavelet sub-bands of image, while it is obvious that both low and high frequency sub-bands of image are influenced through embedding algorithms.

V. CONCLUSION

In this paper, a new image steganalysis method for JPG images has been proposed that uses two classes of features. Statistical properties of the SVD of wavelet sub-bands has been combined with the extended DCT- Markov features that are fed to SVM-RFE classifier to rank and select the most sensitive features. Statistical analysis as well as experimental results show the superiority of the proposed scheme compared to other state-of-the-art steganalysis methods. Specifically, results show that for PQ data hiding algorithm which seems to be more resistant against steganalysis methods than others, about 8% improvement achieved in detection accuracy.

REFERENCES

- [1] A. Westfeld and A. Pfitzmann, "Attacks on Steganographic Systems," Lecture Notes in Computer Science, vol.1768, Springer-Verlag, Berlin, 2000, pp. 61-75
- [2] Andrew D. Ker, "Steganalysis of LSB Matching in Grayscale Images." IEEE Signal Processing Letters, vol. 12(6), pp. 441-444, 2005.
- [3] S. Lyu and H. Farid, "Steganalysis using higher-order image statistics," IEEE Trans Inf Forensics and Sec., vol. 1, no. 1, pp. 111-119, 2006.
- [4] M. Goljan, J. Fridrich, and T. Holotyak, "New blind steganalysis and its implications," Proc. SPIE Security, Steganography, and Watermarking of Multimedia Contents VIII, vol. 6072, pp. 607 201-1, 2006.
- [5] G. Gul and F. Kurugollu, "SVD-based universal spatial domain image steganalysis," Information Forensics and Security, IEEE Transactions on, vol. 5, no. 2, pp. 349-353, 2010.
- [6] T. Pevny and J. Fridrich, "Merging Markov and DCT features for multi-class JPEG steganalysis". In E. J.Delp and P. W. Wong, editors, Proceedings SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX, volume 6505, pages 3 1-3 14, San Jose, CA, January 29-February 1, 2007.
- [7] M. Heidari, and S. Gaemmaghami, "Universal image steganalysis using singular values of DCT coefficients". In *Information Security and Cryptology (ISCISC)*, 2013 10th International ISC Conference. pp. 1-5. IEEE, August, 2013.
- [8] Y. Q. Shi, C. Chen, and W. Chen. "A Markov process based approach to effective attacking JPEG steganography" In Proceedings of the 8-th Information Hiding Workshop, 2006.
- [9] H. Zong, F. Liu, and X. Luo. "A wavelet-based blind JPEG image steganalysis using co-occurrence matrix". 11th International Conference on Advanced Communication Technology, 3:1933-1936, 2009.
- [10] LIU, Q., SUNG, A. H., QIAO, M., CHEN, Z., AND RIBEIRO, B. 2010a. "An improved approach to steganalysis of JPEG images". Inf. Sci. 180, 9, 1643-1655.
- [11] G. Gul, A. E. Dirik, and I. Avcibas, "Steganalytic features for JPEG compression based perturbed quantization," IEEE Signal Process. Lett., vol. 14, no. 3, pp. 205-208, Mar. 2007.
- [12] Corel Draw Software, www.corel.com.
- [13] D. Artz, "Digital steganography: hiding data within data," *internet computing, IEEE*, vol. 5, pp. 75-80, 2001.
- [14] A. Westfeld, "F5 a steganographic algorithm: High capacity despite better steganalysis," 4th International Workshop on Information Hiding, Pittsburgh, PA, USA, 2001.
- [15] J. Fridrich, M. Goljan, and D. Soukal, "Perturbed quantization steganography with wet paper codes", in Proc. ACM Multimedia Workshop, Germany, 2004.
- [16] P. Sallee, "Model-based steganography", in Proc. Int. Workshop on Digital Watermarking", Seoul, Korea, 2003.
- [17] M. Abolghasemi, H. Aghaeinia, K. Faez, "Detection of Perturbed Quantization (PQ) Steganography Based on Empirical Matrix", The ISC Int'l Journal of Information Security, July 2010, Volume 2, Number 2 (pp. 119-128).