Steganalytic Features for JPEG Compression-Based Perturbed Quantization

Gökhan Gül, Ahmet Emir Dirik, and İsmail Avcıbaş, *Member, IEEE*

***Abstract*—Perturbed quantization (PQ) data hiding is almost undetectable with the current steganalysis methods. We briefly describe PQ and propose singular value decomposition (SVD)-based features for the steganalysis of JPEG-based PQ data hiding in images. We show that JPEG-based PQ data hiding distorts linear dependencies of rows/columns of pixel values, and proposed features can be exploited within a simple classifier for the steganalysis of PQ. The proposed steganalyzer detects PQ embedding on relatively smooth stego images with 70% detection accuracy on average for different embedding rates.**

***Index Terms*—Singular value decomposition (SVD), steganalysis, steganography.**

I. Introduction

STEGANOGRAPHY hides information in a manner that the existence of the message is unknown. The goal of steganography is to communicate as many bits as possible without creating any detectable artifacts in the cover-object. If any suspicion about the secret communication is raised, then the goal is defeated. Steganalysis is the art of detecting the presence of covert communication between sender and receiver. A steganographic scheme is considered secure if no existing steganalysis method distinguishes cover and stego-images with a success better than random guessing. The embedding process on an object, while being perceptually transparent, leaves statistical artifacts that can be used to distinguish stego and cover-objects. The argument that data hiding methods leave telltale effects is common to all steganalysis methods [1]–[4]. Recently proposed perturbed quantization (PQ) steganography [5] is a quite successful data hiding approach for which current steganalysis methods fail to work [6]. In other words, PQ does not leave any traces in the form that the current steganalysis methods can catch. However, linear dependency between image rows and/or columns in the spatial domain is affected by PQ embedding due to random modifications on discrete cosine transform (DCT) coefficients' parities during data hiding. In this letter, the change in linear dependency is analyzed by singular value decomposition (SVD), and several features are derived from SVD. By a statistical hypothesis test, we justify the effectiveness of the features and then use these features to build a classifier to differentiate cover and stego-images embedded with the JPEG-based PQ steganography method. The rest of this letter is organized as follows: In Section II, we briefly describe PQ, SVD, and the features derived from SVD. We show the effectiveness of the features in discriminating cover and stego-images via analysis of variance (ANOVA) and scatter diagrams. Data set and experimental results are given in Section III. Conclusions are drawn in Section IV.

Manuscript received April 10, 2006; revised July 19, 2006. This work was supported in part by TÜBİTAK under Project 104E056. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Fernando Perez-Gonzalez.

G. Gül and A. E. Dirik and are with the Electronics Engineering Department, Uludag University, 16059 Bursa, Turkey (e-mail: gokhan@uludag.edu.tr; emirdirik@yahoo.com).

I. Avcıbaş is with the Electrical-Electronics Engineering Department, Baskent University, 06530 Ankara, Turkey (e-mail: avcibas@baskent.edu.tr).

Digital Object Identifier 10.1109/LSP.2006.884010

II. PQ and SVD-Based Features

In PQ steganography, the cover-object is applied an information-reducing operation that involves quantization such as lossy compression, resizing, or A/D conversion before data embedding. The quantization is perturbed according to a random key for data embedding, therefore called "*perturbed quantization*." PQ steganography, which uses JPEG compression for information reducing operation, is different from their DCT-based counterparts. Since message bits are encoded by changing DCT parities after quantization, the cover image can be thought of just as a recompressed input image. To achieve high embedding rates, recompression is realized by doubling the input quantization table with the assumption that recompression of cover JPEG images does not draw any suspicion because of its wide usage in digital photography [5]. Since the original cover image is recompressed via embedding operation, its compressed version should be considered as "cover" instead of original image. While we answer "if any message bits are hidden" for any other steganographic method, we answer "if quantization steps are perturbed" for PQ to make steganalysis possible. As a result of permuting of slight perturbations and altering the quantization steps of nonzero DCT coefficients, statistical properties of an image change in different regions in such a way that these changes balance each other as a whole. The permutation of DCT coefficients makes it impossible to predict which sections of an image are affected by data embedding. Unlike statistical properties, the inherent linear dependency cannot be preserved because any little change made on rows in any region yields a linear dependency modification, regardless of being negative or positive. Since PQ steganography changes DCT coefficients randomly, the linear dependencies between columns and rows of a cover image, especially in smooth regions, decrease according to embedding noise in the DCT domain. The decrease of linear dependencies between rows and columns can be detected by checking singular values resulting from SVD over a given image.

A. Singular Value Decomposition and Derived Features

SVD is an extremely powerful tool in linear algebra. SVD decomposes a matrix $A \in \mathbb{R}^{m \times n}$ into the product of two or-

thonormal matrices $U \in \mathrm{IR}^{m\times m}$, $V \in \mathrm{IR}^{n\times n}$, and a diagonal matrix $S \in \mathrm{IR}^{m\times n}$ as in

$$A = USV^{\mathrm{T}}. \tag{1}$$

The diagonal elements of matrix S are nonnegative and sorted in decreasing order $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(m,n)}$, where m, n are the dimensions of A. Singular value vector Sv is obtained by the diagonal entries of matrix S, and singular value vector has the entire energy of the matrix A

$$Sv = Diag(S) \tag{2}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{A}(i,j)^2 = \sum_{i=1}^{\min(m,n)} Sv(i)^2. \tag{3}$$

Proposition: For a given square matrix A, if the number of linearly dependent rows is i, the number of linearly dependent columns is j, and the number of zeros in the singular value vector of A is k, then

$$k = \begin{cases} i-1, & if\ i > j \\ j-1, & if\ i < j. \end{cases} \tag{4}$$

Proof: Due to the definition of singular value decomposition, U and V are the orthonormal matrices and of full rank. Thus, the rank of matrix A equals to the number of nonzero elements of matrix S. The rank of a matrix is the minimum number of linearly independent rows or columns plus one. Therefore, the number of zero elements of diagonal of matrix S equals to the maximum number of linearly dependent rows or columns minus one.

While linear dependency of rows and columns increases, the singular values with high indexes tend to decrease, and singular values with low indexes tend to increase, i.e., energy flows from high indexed singular values toward low indexed ones, and vice versa. Therefore, if all rows or all columns of a matrix are linearly dependent, whole energy of the matrix will be concentrated in the first index of the singular value vector Sv.

The number of linearly dependant rows is determined by checking zero values at a certain index in Sv for 50% overlapping $W\times W$ windows over a given image. For each window, Sv vector is computed and the ith indexed singular value is checked whether or not it is zero. If the ith singular value is zero, then the *linear dependency counter* that indicates the overall linear dependency in an image is incremented. Finally, the linear dependency counter is normalized by the number of sub-blocks in the image. Linear dependency counter values of 300 stego and 300 recompressed cover images are given in Fig. 1. These images are randomly selected from the image database in [6]. The linear dependency counter is computed by checking zeros at the eighth singular values in Sv vectors for 16 × 16 windows. As it is seen in Fig. 1, the number of zero singular values in stego-images is significantly less than the number of zero singular values in recompressed cover-images. This observation, which is also the pillar of the proposed scheme, suggests that PQ embedding decreases the linear dependency of the pixel rows, and suitable features measuring the extent of linear dependency can be discriminative enough to design a steganalyzer. In the following, we define some SVD-based features to measure various aspects of linear dependency.

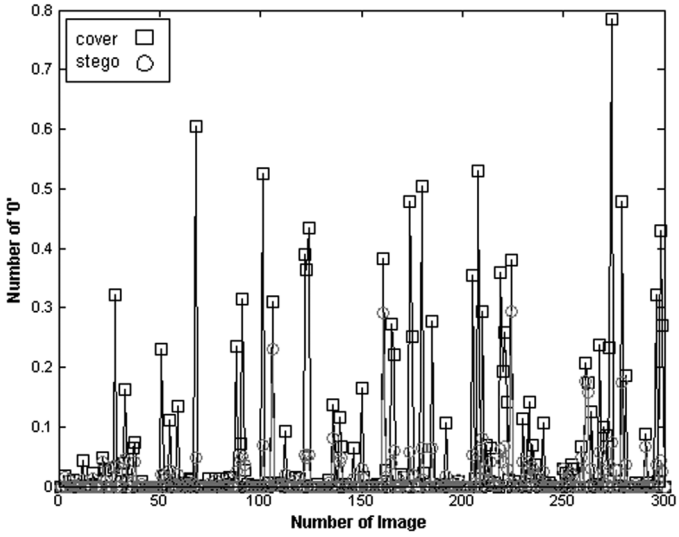

Fig. 1. Number of zero singular values in normalized singular value vectors (Sv) for stego (0.2 bpc embedding rate) and recompressed cover images.

B. SVD-Based Features

In order to obtain SVD-based features, images are first normalized to zero mean, and then SVD is applied to $W\times W$ ($W = 3, 4, \ldots 20$) blocks to obtain Sv vectors for each block. Singular values are normalized with the sum of singular values in Sv vector to reduce the effects of different energy values in different images. SVD-based features are defined as follows.

Features of Type 1: These features are the means of the number of zeros at index i in Sv vectors of blocks of size $W \times W$. Let B be the integer number of 50% overlapped $W\times W$-sized blocks in $M\times N$-sized image. Type 1 feature is defined as

$$f_W^{(1)}(i) = \frac{1}{B}\sum_{B} \delta\left(Sv(i)\right), \quad W = 3, \ldots, 20 \text{ and}$$

$$i = 1, \ldots, W, \text{ where } \delta(k) = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0. \end{cases}$$

Features of Type 2: These features are the means of the singular values at index i in Sv vectors of blocks of size $W \times W$

$$f_W^{(2)}(i) = \frac{1}{B}\sum_{B} Sv(i), \quad W = 3, \ldots, 20 \text{ and } i = 1, \ldots, W.$$

Features of Type 3: Variance of type 1 features over W and i parameters are defined as type 3 features. There are two subsets of type 3 features. For fixed i's, the first subset consists of the variances of the type 1 features over W. The second subset consists of the variances of type 1 features computed over i for fixed W's.

Features of Type 4: These features are defined the same way as type 3, but variances are computed on type 2 features rather than type 1 features.

TABLE I
ANOVA STATISTICS OF FEATURES USED IN STEGANALYSIS SCHEMES

	SVBS	WBS	FBS	BSM
1	0	1.10^{-3}	4.10^{-5}	2.10^{-6}
2	0	2.10^{-3}	3.10^{-4}	4.10^{-6}
3	0	1.10^{-2}	5.10^{-2}	2.10^{-5}
4	0	2.10^{-2}	8.10^{-2}	3.10^{-5}
5	0	5.10^{-2}	1.10^{-1}	3.10^{-4}
6	2.10^{-16}	9.10^{-2}	3.10^{-1}	3.10^{-4}
7	2.10^{-12}	1.10^{-1}	5.10^{-1}	9.10^{-4}

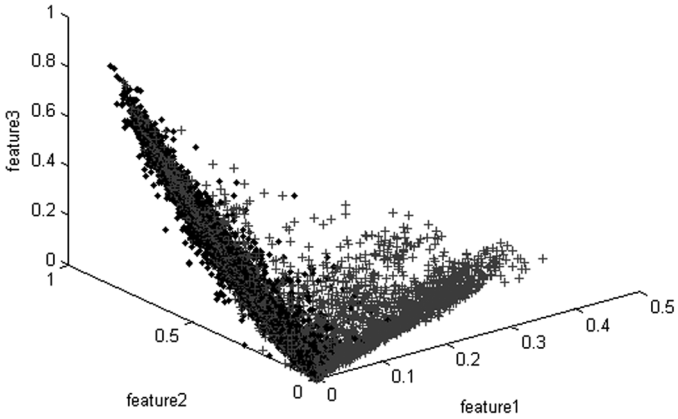

Fig. 2. The 3-D plot of type 1 and type 5 features ($f_{16}^{(5)}(8)$, $f_{17}^{(1)}(8)$, $f_{14}^{(1)}(9)$) of recompressed-cover and stego images.

Features of Type 5: These features are variants of type 1 features. Instead of counting exact zero values, singular values lying in particular intervals (e.g., $[10^{-16}, 10^{-8}]$) are counted.

C. Statistical Analysis of the Proposed Features

ANOVA is a general statistical hypothesis testing technique used when one wants to determine whether or not a number of data groups are statistically different. The *p-value* is the probability of finding in reality that there is no difference between the means. We used one-way ANOVA to test the proposed features for their effectiveness. We formed two groups, each consisting of feature vectors computed from the randomly chosen 500 cover-images and 500 stego-images (embedded with different rates). Table I gives the smallest seven p-values of the ANOVA test for the null hypothesis that the means of the groups are equal for a given feature used in the proposed SVD steganalysis (SVBS) method and other state-of-the-art steganalysis methods, feature-based steganalysis (FBS) [1], wavelet-based steganalysis (WBS) [2], and binary similarity measures (BSM)-based steganalysis [3] methods. Relatively much smaller p-values of the proposed features indicate that the feature means for two populations are significantly different and, as will be seen in the sequel, explain why it differentiates stego and cover-image sets more accurately than other steganalysis methods. In Fig. 2, we give the scatter plot of one type 5 and two type 1 features of cover and stego-images (each consisting of 4631 samples).

TABLE II
NUMBER OF IMAGES WITH DIFFERENT EMBEDDING RATES

Embedding Rate	0.4	0.3	0.2	Mixed
Train :	641	550	540	1631
Test :	1000	1000	900	3000

D. Steganalyzer Design

Having collected the SVD-based feature vectors from both cover and stego-images, we built a linear regression classifier to differentiate them. Feature vectors were regressed to -1 and 1, respectively, if the feature vector belonged to the cover and stego-image. We expressed each decision label $g_i \in [-1, 1]$, $i = 1, \ldots, N$ as a linear combination of features, $g_i \in \beta_1 f_{1i} + \beta_2 f_{2i} + \ldots + \beta_q f_{qi}$, where $f_i = (f_{1i}, f_{2i}, \ldots, f_{qi})$ is the vector of q features computed from the ith image, and $\beta = (\beta_1, \beta_2, \ldots, \beta_q)$ are the regression coefficients. Once the regression coefficients were predicted in the training phase, they were used to test the images. For an incoming image, a feature vector f is constructed first, $g = \beta_1 f_1 + \beta_2 f_2 + \ldots + \beta_q f_q$ is evaluated next, and finally, if the evaluated value is above the threshold 0, then the decision is that the incoming image contains a message; otherwise, the decision is that it does not. The training process also employed the feature selection algorithm described in [7], both to minimize the classification error and to limit the number of features used. A 78-D feature vector, consisting of 17 type 1, 43 type 2, 4 type 3, 3 type 4, and 11 type 5 features were selected by the feature selection algorithm [7] in the training process for the maximum performance. Though features of type 1 and type 5 seem adequate to build a classifier, features of type 2 and their variants are imperative when the classifier lacks of finding zeros inside a given image.

III. EXPERIMENTAL RESULTS

A. Image Set

We used a subset of smooth, mid-quality image set from the image database used in [6], which were collected from the internet without any bias. The quality of the image subset is 80.8 with 5.1 standard deviation, and the mean of the entropy of the images is 4.0 bits with 1.6 standard deviation. These images were subjected to both recompression and PQ embedding operations as described in [5] to create recompressed cover-image and stego-image sets with different embedding rates, 0.4, 0.3, and 0.2 bits per nonzero DCT coefficient (bpc). The number of images used for training and testing is given in Table II. A mixed rate data set is formed with the cover-group and the stego-group consisting of the pool of 1641 images embedded with 0.4 bpc, 1550 images with 0.3 bpc, and 1440 images with 0.2 bpc to demonstrate that the steganalyzer works independent of the embedding rate.

B. Experiments

The image set was divided into non-overlapping training and testing subsets. Each of these subsets consisted of cover and stego-images. We first trained the linear regression classifier

TABLE III
DETECTION RATES FOR PQ STEGANALYSIS

Embedding Rate	0.4	0.3	0.2	Mixed
SVBS	%76.00	%72.85	%67.00	%70.13
WBS	%63.50	%57.45	%53.56	%58.68
FBS	%59.55	%55.25	%53.83	%56.32
BSM	%59.15	%56.45	%54.22	%56.00

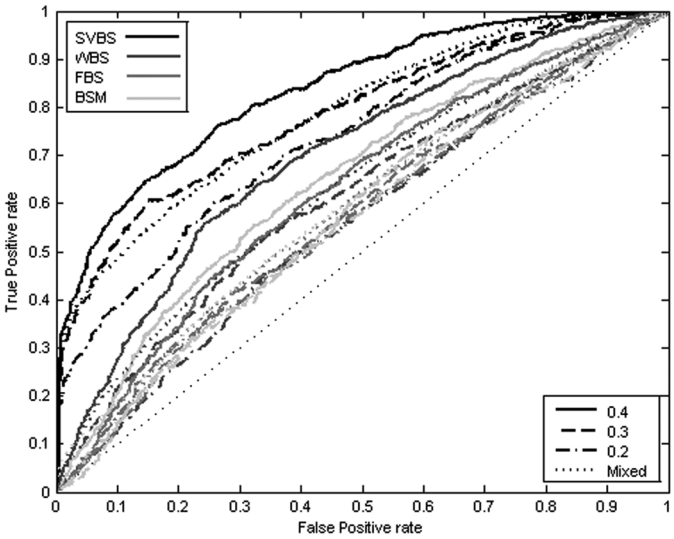

Fig. 3. ROC curves of SVBS and the other blind steganalysis schemes.

with the features computed from the images in the training set and then classified the images in the test set (see Table II). We give the comparative detection performance of the proposed SVBS method and other steganalysis methods, FBS, WBS, and BSM, in Table III. The detection rate in Table III is defined as $1 - 0.5 \times$ (false alarm percentage + miss percentage). False alarm percentages in SVBS scheme are 0.12, 0.15, and 0.27 for 0.4, 0.3, and 0.2 embedding rates, respectively. To make a fair comparison, a linear classifer was trained for every steganalysis method on the same images and test on the same images. The receiver operator characteristic (ROC) curves are given in Fig. 3 for 0.4 bpc, 0.3 bpc, 0.2 bpc, and mixed embedding rates. The proposed scheme achieves higher detection rates than the other schemes for a relatively low entropy image set with first-order entropy less than 6 bits. The next experiment was to test the steganalysis schemes on complex images without a first-order entropy constraint. We randomly selected 1000 images from the image set in [6] and created stego-images with different embedding rates. Four hundred of these images are used for training and the remaining 600 for testing. An 8-D SVD-based feature vector was selected in the training process for SVBS. The selected features were $f_9^1(4)$, $f_{10}^1(3)$, $f_{13}^1(2)$, $f_{14}^1(2)$, $f_{16}^1(4)$, $f_{18}^1(2)$, $f_{16}^5(8)$, and $f_{20}^1(5)$. While SVBS had 69% detection accuracy for 0.4 bpc rate with this new image set, WBS had 54%, FBS had 54%, and BSM had 57% detection accuracy. For 0.3 bpc and 0.2 bpc, SVBS had 58% and 57% detection performance, respectively, while other schemes' performance ranged in 52%–50%. These results were in agreement with the results in [6]. Finally, the proposed scheme is tested on model-based steganography [8] to investigate its potential for universal steganalysis. We tested SVBS on 1800 images downloaded from the personal collection of P. Greenspun (http://philip.greenspun.com). The stego images are embedded with 0.4 rates by model-based steganography. Eight hundred images are used for training, and the remaining 1000 images are used for testing. We have obtained 78% detection performance, which indicates that the proposed features are not peculiar to PQ steganalysis.

IV. CONCLUSION

In this letter, we have proposed SVD-based features for steganalysis of JPEG-based PQ embedding. We have shown with both statistical analysis and experimental results that the PQ embedding distorts linear dependencies of rows and columns. Proposed features measuring the resulting distortion can be used in steganalysis of JPEG-based PQ. Experimental results over about 5000 images indicate the validity of the proposed scheme. Comparative results showed that the proposed scheme had better detection performance than other state-of-the-art steganalysis schemes. It is easier to detect traces of PQ embedding in smooth regions by the proposed features as smooth regions contain relatively more linearly dependent rows/columns than in complex regions. As a result, detection performance drops considerably with images having high entropies.

ACKNOWLEDGMENT

The authors would like to thank M. Kharrazi and N. Memon for providing them with the images used in experiments.

REFERENCES

[1] J. Fridrich, "Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes," in *Proc. 6th Information Hiding Workshop*, Toronto, ON, Canada, May 23–35, 2004.
[2] S. Lyu and H. Farid, "Detecting hidden messages using higher-order statistics and support vector machines," in *Information Hiding 5th International Workshop*, F. A. P. Petitcolas, Ed. New York: Springer-Verlag, 2002, vol. 2578, Lecture Notes in Computer Science, pp. 340–354.
[3] I. Avcıbaş, M. Kharrazi, N. Memon, and B. Sankur, "Image steganalysis with binary similarity measures," *EURASIP J. Appl. Signal Process.*, vol. 2005, no. 17, pp. 2749–2757, Sep. 2005.
[4] J. J. Harmsen and W. A. Pearlman, "Steganalysis of additive noise modelable information hiding," *Proc. SPIE/IS&T Electron. Imag.*, vol. 5022, Jan. 2003.
[5] J. Fridrich, M. Goljan, and D. Soukal, "Perturbed quantization steganography with wet paper codes," in *Proc. ACM Multimedia Security Workshop*, Magdeburg, Germany, Sep. 20–21, 2004, pp. 4–15.
[6] M. Kharrazi, H. T. Sencar, and N. Memon, "Benchmarking steganographic and steganalytic techniques," in *Proc. Electrical Imaging, SPIE, Security, Steganography, Watermarking Multimedia Contents VII*, San Jose, CA, Jan. 17–20, 2005.
[7] P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, pp. 1119–1125, 1994.
[8] P. Sallee, "Model-based steganography," in *Proc. Int. Workshop Digital Watermarking*, Seoul, Korea, Oct. 2003, pp. 154–167.