Classification of PQ Stego-Images and F5 Stego-Images using SVM

Rita Chhikara¹, Deepika Bansal²

¹Department of Computer Science and Engineering, ITM University Gurgaon, Haryana, India
²Department of Computer Science and Engineering, ITM University Gurgaon, Haryana, India

Abstract— PQ and F5 are two typical steganography methods of JPEG images and have been used widely. To compare the robustness of the above two algorithms, to withstand steganalytic attack a classification algorithm based on sensitive features is presented. SVM is used as a classifier using sensitive features extracted from two domains DCT and DWT. A comparison between the two types of features is also shown. Experimental results show that the PQ method can withstand steganalytic attack by a margin of 30% accuracy as compared to F5, hence PQ is more reliable for Steganography.

Keywords— Steganography, PQ, F5, SVM, Discrete Cosine Transformation, Discrete Wavelet transformation.

I. INTRODUCTION

Steganography is the art of hiding information inside the images, audio or video. The main aim of steganography is to hide the existence of the secret information hidden in any cover file used. While using steganography, we can prevent our secret message from being damaged or accessed by any attacker. The three policies of computer security are confidentiality, integrity & availability. Steganography is used to maintain confidentiality. Throughout the history, various steganography techniques were being used, for example wax covered tablets, hidden tattoos, invisible inks, microfilms, microdots, null ciphers, etc [1].

The most widely used image steganographic techniques are substitution technique in (i) Spatial Domain: Spatial domain technique embeds secret bits directly in the cover file. The commonly used spatial domain techniques is Least significant bit insertion (LSB). In LSB, the secret bits are inserted in the least significant bits of cover image. LSB is of 2 types [2]: LSB Replacement & LSB Matching. (ii) Transform domain[3] like Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT) and Fast Fourier Transform (FFT) (iii) Spread spectrum technique[4].

Fridrich proposed a feature set mainly derived from marginal and joint statistics of DCT coefficients and employed the concept of calibration[5]. Another JPEG feature set was proposed by Shi et al., which is based on Markov models of DCT plane and called “Markov features”[6]. Result indicated that these two feature sets could achieve good performance. However these researchs on single feature have some drawbacks such as biased detection for different steganographic algorithms. To overcome the limitation of single feature method, Tomas Pevny et al. rebuilt the “DCT features” and the “Markov features”, the resulting feature sets are called “DCT extended features” and “Markov reduced features” in this paper, and merged them to produce a 274-dimensional feature vector, which is named “Merged features”[7]. The features were further extended by Farid et al[8] to wavelet domain and indicated that steganography algorithms are more sensitive to high order statistics of subband coefficients of wavelet decomposition.

In [9] author has classified Jsteg and F5 using SVM and wavelet features, however they have not taken DCT domain features into consideration. In our proposed work we have shown comparison between features in both domain for classifying PQ and F5 steganography algorithm.

The paper is divided as follows, section II explains the two steganography algorithms, section III gives proposed method Feature extraction is highlighted in section IV and SVM (Support Vector Machine) is explained in section V. Experimental Analysis is given in section VI followed by conclusion in section VII.

II. STEGANOGRAPHY ALGORITHMS

A. F5

F5 withstands visual and statistical attacks offering a large steganographic capacity. F5 implements matrix encoding to improve the efficiency of embedding. F5 employs permutative straddling to uniformly spread out the changes over the whole steganogram. It shuffles all coefficients using a permutation first. Then, F5 embeds into the permuted sequence[10].

The algorithm F5 has the following coarse structure:

1. Start JPEG compression. Stop after the quantisation of coefficients.
2. Initialise a cryptographically strong random number generator with the key derived from the password.
3. Instantiate a permutation (two parameters: random generator and number of coefficients).

ISSN :2277-6370
4. Determine the parameter \( k \) from the capacity of the carrier medium, and the length of the secret message.
5. Calculate the code word length \( n = 2^k - 1 \).
6. Embed the secret message with \((1, n, k)\) matrix encoding.
   (a) Fill a buffer with \( n \) nonzero coefficients.
   (b) Hash this buffer (generate a hash value with \( k \) bits).
   (c) Add the next \( k \) bits of the message to the hash value (bit by bit, xor).
   (d) If the sum is 0, the buffer is left unchanged. Otherwise the sum is the buffer’s index 1 \( \ldots \) \( n \), the absolute value of its element has to be decremented.
   (e) Test for shrinkage, i.e. whether we produced a zero. If so, adjust the buffer (eliminate the 0 by reading one more nonzero coefficient, i.e. repeat step 6a beginning from the same coefficient). If no shrinkage occurred, advance to new \( c \) coefficients behind the actual buffer. If there is still message data continue with step 6a.
7. Continue JPEG compression (Huffman coding etc.).

B. PQ (Perturbed Quantization)

In Perturbed Quantization[11], the sender hides data while processing the cover object with an information reducing operation that involves quantization, such as lossy compression, down sampling or A/D conversion. The unquantized values of the processed cover object are considered as side information to confine the embedding changes to those unquantized elements whose values are close to the middle of quantization intervals. This choice of the selection channel calls for wet paper codes as they enable communication with non shared selection channel. Perturbed Quantization aims to achieve high efficiency, with minimal distortion, rather than a large capacity. Each coefficient in the DCT block is assigned a scalar value that corresponds to how much impact it would make to the carrier image, and then a steganographer can set a selection rule to filter out the “well behaved” coefficients.

III. PROPOSED METHOD

In this section, we give the classification framework to distinguish the PQ stego-images and F5 stego-images (see Figure 1), which is based on 274 features[13] extracted from DCT (Discrete Cosine Transform) domain and 48 features extracted from subbands coefficients of DWT (Discrete Wavelet Transform) domain. These features are then classified using SVM (Support Vector Machine).

The main steps of this framework can be described as follows:

(A) Collect image samples to construct the training image set.
(B) Divide the sample into training and test image samples
(C) Extract the multi-domains features of training images and test images respectively.

(D) Training the SVM classifier. Use the features of image whose class has been known to train the classifier, and get a classifier with good performance.
(E) Image test. Use the classifier from (D) to detect test images, and obtain accuracy for each set of test images from PQ and F5 resp.
(F) Find the accuracy of PQ and F5 test images.

IV. FEATURE EXTRACTION

A. DCT domain features

A large number of experiments indicate that steganography algorithms like jsteg, PQ, F5 change the statistical distribution of DCT coefficients at different frequency of the image, although F5 avoids the “point-pair” phenomena. It is certain that the correlation within DCT coefficients will be disturbed when image is embedded with message. Based on this theory, set up a statistical model for DCT coefficients was given by Pevny and Fridrich [13]. The various features extracted for our work from DCT domain are as given below

a. Global Histogram

The first functional is the histogram \( H \) of all \( 64 \times n_b \) luminance DCT coefficients

\[
H = (HL, \ldots, HR),
\]

where \( L = \min_{i,j,k} d_{ij}(k) \), \( R = \max_{i,j,k} d_{ij}(k) \).

b. AC Histograms

The next 5 functionals are the histograms

\[
h^{(i)} = (h_{1}^{(i)}, \ldots, h_{9}^{(i)}),
\]

of coefficients of 5 individual DCT modes \( (i,j) \in \{(1,2),(2,1),(3,1),(2,2),(1,3)\} = L \).

c. Dual Histograms
The next 11 functionals are dual histograms represented with 8×8 matrices \( g_{ij} \), \( i, j = 1, \ldots, 8 \), \( d = -5, \ldots, 5 \)
\[
g_{ij} = \sum \delta(d, d_i(k)),
\]
where \( \delta(x, y) = 1 \) if \( x = y \) and 0 otherwise.

\( d \). Variation

The next 6 functionals capture inter-block dependency among DCT coefficients. The first functional is the variation \( V \)
\[
V = \sum_{i,j=1}^{8} |d_i(I(k)) - d_i(I(k+1))| + |I_i + I_{i+1}|
\]
where \( I_I \) and \( I_c \) denote the vectors of block indices 1, \ldots, \( n_B \) while scanning the image by rows and by columns, respectively.

\( e \). Blockiness

Two next two blockiness functionals are scalars calculated from the decompressed JPEG image representing an integral measure of inter-block dependency over all DCT modes over the whole image:
\[
B_r = \sum_{i,j=1}^{8} |d_i(I(k)) - d_i(I(k+1))| + |I_i + I_{i+1}|
\]

\( f \). Co-occurrence matrix

The remaining three functionals are calculated from the co-occurrence matrix of neighboring DCT coefficients
\[
N_{00} = C_{0,0}(J_1) - C_{0,0}(J_2)
N_{01} = C_{0,1}(J_1) - C_{0,1}(J_2) + C_{1,0}(J_1) - C_{1,0}(J_2) + C_{1,0}(J_1) - C_{1,0}(J_2) - C_{0,1}(J_1) - C_{0,1}(J_2)
\]
\[
N_{11} = C_{1,1}(J_1) - C_{1,1}(J_2) + C_{1,1}(J_1) - C_{1,1}(J_2) + C_{1,1}(J_1) - C_{1,1}(J_2),
\]
where
\[
C_{st} = \sum_{i,j=1}^{8} |d_i(I(k))| \delta(t, d_i(I(k+1))) + |I_i + I_{i+1}|
\]

\( g \). Markov Features

The DCT coefficients in \( F(u, v) \) are arranged in the same way as pixels in the image by replacing each \( 8 \times 8 \) block of pixels with the corresponding block of DCT coefficients. Next, four difference arrays are calculated along four directions: horizontal, vertical, diagonal, and minor diagonal (further denoted as \( F_h(u, v), F_v(u, v), F_d(u, v), \) and \( F_m(u, v) \) respectively)
\[
F_h(u, v) = F(u, v) - F(u + 1, v),
F_v(u, v) = F(u, v) - F(u, v + 1),
F_d(u, v) = F(u, v) - F(u + 1, v + 1),
F_m(u, v) = F(u + 1, v) - F(u + 1, v + 1).
\]

From these difference arrays, four transition probability matrices \( M_h, M_v, M_d, M_m \) are constructed as
\[
M_{h}(i,j) = \sum_{u=1}^{8} \sum_{v=1}^{8} \delta(F_{h}(u,v)=i,F_{h}(u+1,v+1)=j)
M_{v}(i,j) = \sum_{u=1}^{8} \sum_{v=1}^{8} \delta(F_{v}(u,v)=i,F_{v}(u+1,v+1)=j)
M_{d}(i,j) = \sum_{u=1}^{8} \sum_{v=1}^{8} \delta(F_{d}(u,v)=i,F_{d}(u+1,v)=j)
M_{m}(i,j) = \sum_{u=1}^{8} \sum_{v=1}^{8} \delta(F_{m}(u,v)=i,F_{m}(u+1,v+1)=j)
\]

\( h \). Statistical moments of wavelet subbands

Harmsen proposed an additive noise model [14] and pointed out that the effect of data hiding in spatial domain of the wavelet coefficients reflects that the peak of the histogram becomes flat and extends to the two ends. So extracting the high order statistics from images after wavelet decomposition can effectively differentiate the stego images and cover images. In this work we have firstly decomposed our image using a three level wavelet decomposition using Haar wavelet. An image and its 1-level haar wavelet decomposition is as shown

\[
\begin{array}{c|c}
\text{Functional} & \text{Dimensionality} \\
\hline
\text{Global histogram } H_t & 11 \\
\text{5 AC histograms } h_{ij}^5 & 5 \times 11 \\
\text{11 Dual histograms } g_{ij}^1 & 11 \times 9 \\
\text{Variation } V & 1 \\
\text{2 Blockiness } B_{st} & 2 \\
\text{Co-occurrence matrix } C_{st} & 25 \\
\end{array}
\]
in Fig 2. From all four components, including low pass component at each level i for i = 1,2,3. Thereafter four statistics features: mean, variance, skewness and kurtosis are extracted from them to get 48 wavelet features which are represented as DWT features in Table II.

![Wavelet Decomposition](image)

**Fig. 2 a) Image b) 1-level wavelet decomposition**

V. SVM CLASSIFIER

We employ the support vector machine (SVM) for comparative study of F5 and PQ in terms of their accuracy using the two kinds of features extracted in section IV. SVMs are able to calculate the maximum margin (separating hyper-plane) between data with and without the outcome of interest if that data are linearly separable. The margin is calculated by solving a constrained optimization problem using the Lagrangian formulation. However, such applications in real data sets are limited, and a number of adaptations have been applied to SVMs in order to improve their utility. The feature space can be modified using kernels in order to allow fitting of data that are not linearly separable. Individual kernels, such as polynomial and radial basis function transform the feature space in distinct ways, and kernel performance is dependent on characteristics of the source data. Gaussian radial-based kernels was selected in this study because of their generally good performance.

VI. EXPERIMENTAL RESULTS AND ANALYSIS

All the experiments are implemented in Matlab 2010. We have downloaded 1000 natural images from 1000pictures [16]. Those images are with different size of 1024x768, 768 · 512, 640 · 480, 512 · 512 and so on. They span an extensive range of landscape, architecture, animal, and people. For the set of 1000 natural images, we use two steganography methods, PQ and F5, to produce two sets of stego-images. In all the results presented below, for all stego-images sets and corresponding cover images sets, 600 images per set are chosen to train SVM classifier, and the remaining 400 images in each set are used to test. LibSVM[15] tool using radial basis function was employed for classification.

Table II shows the classification results of our method for features extracted from two different domains that is DCT and DWT. The accuracy of detection achieved using DCT features is less as compared to DWT features. As can further be observed SVM classifies F5 with an accuracy of almost 99.8% whereas PQ images are detected with an accuracy of 70%, hence making PQ steganography algorithm more robust to steganalytic attack using the features extracted in section IV.

<table>
<thead>
<tr>
<th>SVM CLASSIFICATION ACCURACY OF PROPOSED METHOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEATURES</td>
</tr>
<tr>
<td>F5</td>
</tr>
<tr>
<td>PQ</td>
</tr>
</tbody>
</table>

VII. CONCLUSIONS

In order to classify two typical kinds of steganography algorithms for JPEG image, PQ and F5 steganography, In this paper we gave a classification method based on sensitive features and SVM classifier. The sensitive features are extracted from both domains DCT and DWT and a comparison between the two is given in table II in experimental results. As it can be observed the DWT features give accuracy of almost 100% for F5 and 10% more for PQ, hence making them more efficient features for steganalysis. The experiment results also indicate that PQ steganography algorithm is more strong against steganalytic attack with the features extracted in section IV as compared to F5. In the next step, we shall extend our work to comparative study of other steganography algorithms using more efficient classification techniques. We shall also study the method of feature selection and optimization.

VIII. REFERENCES


F5 Tool and PQ Tool Available: http://dde.binghamton.edu/download/stego_algorithms/


LibSVM ToolBox Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm/

Image Source: www.1000pictures.com