International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 10 Issue 2, February-2021, Impact Factor: 7.957

Coupling of New Model Feature Extraction and Support Vector Machine for Blind Image Steganalysis

Rafidison Maminiaina Alphonse[1], Ramafiarisona Hajasoa Malalatiana[2], Pr Randriamitantsoa Paul August[3]

[1,2,3]Telecommunication Automatic Signal Image Research Laboratory/Doctoral School in Science and Technology of Engineering and Innovation/ University of Antananarivo, Antananarivo 101, Madagascar

ABSTRACT

Information security is very important in different domain especially in information technology. A confidential information supposed to not be transferred out of group may be shared by a member to external person. A solution should be in place to detect such malicious action. This paper presents a method to detect the presence of hidden information in image commonly known as steganalysis. The approach here is based on Support Vector Machine (SVM) and classified as passive algorithm. It means that we are not interested to find the steganography method and hidden information, we destroy directly the steg support image. Surely, we have our own technical image feature extraction by introducing Intersecting Cortical Model (ICM) neural network, Discrete Cosine Transform (DCT), Mojette transform, Discrete Wavelet Transform (DWT) to ensure a good performance result.

Keywords: Feature extraction, Intersecting Cortical Model, Mojette Transform, steganalysis, Support Vector Machine.

1. INTRODUCTION

Steganalysis is an art to detect the presence of hidden information done with steganography method. This discipline is a dual of steganography. For clear understanding, we consider three persons: first a prisoner, second a guardian and third partner in crime. The first person tries to send a secret message to third person passing through second one, he applied steganography algorithm. The guardian on his turn, he prepares a steganalysis method in the aim to analyze the transited information. This guardian can be passive, active, malicious. If he just observes the traffic, he is called “passive” but if he tries to modify the medium to delete steganography process, he is “active”. He can try to understand the algorithm and extract the hidden message, so in this case, he is a “malicious” guardian.

In this paper, we are content with detecting the presence of secret message in medium image. To attend the objective, we follow the following plan: firstly, we will present an overview of steganalysis followed directly by our related work and testing results. Our steganalysis algorithm has two parts: feature extraction and classification. Certainly, we have a discussion and conclusion before closing this present article.

2. STEGANALYSIS

Steganalysis is to identify suspected data, determine hidden data, and recover the hidden data. Steganalysis can be divided into four categories: visual, structural, statistical, and learning steganalysis. Visual steganalysis is to investigate visual artifacts in the stego-images, where try to catch visual difference by analyzing stego-images. Structural steganalysis looks into suspected signs in the media format representation since the format is often changed when the secret message is embedded. RS analysis and pair analysis are included in the structural steganalysis. Statistical steganalysis utilizes statistical models to detect steganography techniques. Statistical steganalysis can be divided into specific statistical and universal statistical steganalysis. Learning steganalysis also called blind steganalysis is one of universal statistical steganalysis since cover images and stego-images are used as training datasets.

Other classification of steganalysis can be divided into sixcategories as shown in Fig.1. It is depending on what kind of attacks a forensic examiner uses.

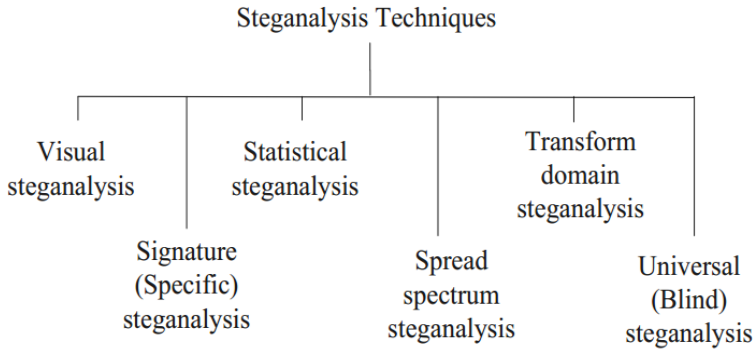

Fig. 1: Steganalysis classification

In Figure 1, universal or blind steganalysis techniques are based on detecting the secret messages without regard to steganography techniques. Comparing with other steganalysis techniques, universal steganalysis technique is very difficult to find extraction features [1][2].

3. PROPOSED UNIVERSAL STEGANALYSIS SCHEME

A. Preprocessing steps

An image will pass through different module and decompose to RGB channel before feature extraction (Fig.2). Among these modules, we have:

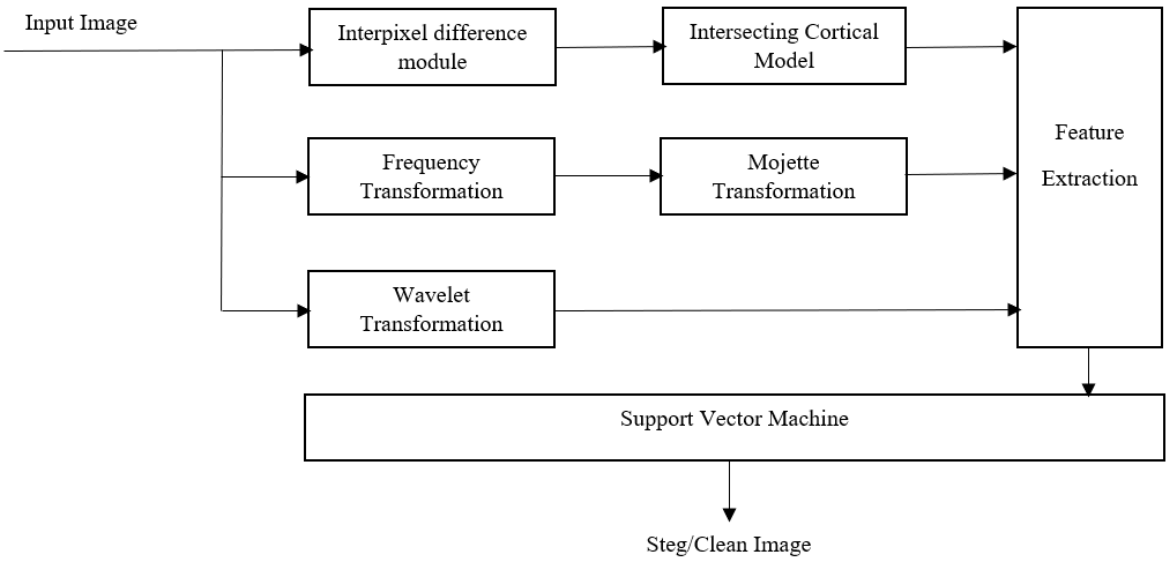

Fig. 2: Block shematic of proposed method

1) Interpixel difference module

As we know some of steganography methods changes some pixel value and to detect this change, the idea is we should calculate the difference between gray level of neighbor pixel [4].
For these calculations, column, row and diagonal are required as shown below:

$$I_{dc}(x,y) = I(x,y) - I(x,y-1) \quad (1)$$

where $I(x,y)$ is the input image with dimension $M \times N$, x is representing the row and $y$column.$I_{dc}(x,y)$ denotes the column difference. y value is starting with 2 to N and $I_{dc}(x,1) = I(x,1)$.

$$I_{dr}(x,y) = I(x,y) - I(x-1,y) \quad (2)$$

where $I_{dr}(x,y)$ is row difference and $I_{dr}(1,y) = I(1,y)$. x value is between 2 and M.

Diagonal difference has two parts: $I_{dd1}$and I_{dd2}.

$$I_{dd1}(x,y) = I(x+1,y+1) - I(x,y) \quad (3)$$

with $1 \le x \le M-1$ and $1 \le y \le N-1$.

$$I_{dd2}(x,y) = I(x+1,y-1) - I(x,y) \quad (4)$$

with $1 \le x \le M-1$ and $2 \le y \le N$. Note that $I_{dd1}$and $I_{dd2}$are initiatedsame value as$I$before applying (3) and (4).

2) Intersecting cortical model

ICM is power full neural network to extract image characteristics. It is a minimal system which consists of two coupled oscillators, a small number of connections, and a nonlinear function. This system is described by the following three equations.

$$F_{x,y}[n+1] = fF_{x,y}[n] + S_{x,y} + W\{Y\}_{x,y} \quad (5)$$

$$Y_{x,y}[n+1] = \begin{cases} 1, & if\ F_{x,y}[n+1] > \theta_{xy}[n] \\ 0, & Otherwise \end{cases} \quad (6)$$

$$\theta_{x,y}[n+1] = g\theta_{x,y}[n] + hY_{x,y}[n+1] \quad (7)$$

Here the input array is S, the state of the neurons are$F$, the outputs are$Y$, and the dynamic threshold states are θ. The scalars *f* and *g* are both less than 1.0 and $g < f$ is required to ensure that the threshold eventually falls below the state and the neuron pulses. The scalar *h* is a large value the dramatically increases the threshold when the neuron fires. The connections between the neurons are described by the function *W {}* and for now these are still the 1/*r* type of connections [3].

The image resultant of I, I_{dc}, I_{dr}, $I_{dd1}$and I_{dd2} is presented as input to replace *S.F* and *W* are issued of filtering of this resultant image with mark matrix [0.707 1 0.707; 1 1 1;0.707 1 0.707]. θ is initialized as matrix of ones same size as I. ICM role here is to highlight the changed pixel value during steganography phase.

3) Frequency transformation

DCT transforms an image from spatial to frequency domain. Steganography arts use also frequency domain to ensure the robustness of the algorithm, so it is the reason that we exploit this area. DCT transform is described below:

$$I_{uv} = \alpha_u \alpha_v \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I_{xy} \cos\left[\frac{\pi.u}{2M}(2x+1)\right] \cos\left[\frac{\pi.v}{2N}(2y+1)\right] \quad (8)$$

where

$$\begin{aligned} &0 \le u \le M-1 \\ &0 \le v \le N-1 \end{aligned} \quad (9)$$

$$\alpha_u = \begin{cases} 1/\sqrt{M}, & u = 0 \\ \sqrt{2/M}, & 1 \le u \le M-1 \end{cases} \quad (10)$$

$$\alpha_v = \begin{cases} 1/\sqrt{N}, & v = 0 \\ \sqrt{2/N}, & 1 \le v \le N-1 \end{cases} \quad (11)$$

I_{xy}, $I_{uv}$are respectively input image and DCT transform matrix. $I_{u=0v=0}$ corresponds to DC coefficient and remaining value is defined as AC coefficient. In our case, we reset to 0 DC coefficient. We are interesting on AC where the hidden information is located.

4) Mojette transformation

I_{uv} with DC zero is acting the role of input of Mojette transform. We describe briefly this transformation.

Mojette transformation (MT) is deducted from discrete Radon transform. For MT, different angles are required for the calculation such as

$$\theta = arctan\left(\frac{p}{q}\right) \quad (12)$$

where (*p*,*q*) are integers restricted to PCD(*p*,*q*) = 1

$M_{p,q}$ operator help us to calculate each projection:

$$M_{p,q}f(k,l) = \sum_{k=-\infty}^{+\infty}\sum_{k=-\infty}^{+\infty} f(k,l)\Delta(b+kq-lp) \quad (13)$$

where

$$\Delta(b) = \begin{cases} 1, & if\ b = 0 \\ 0, & if\ b \neq 0 \end{cases} \quad (14)$$

and $M_{p,q}f(k,l)$is denotedas the set of I projections:

$$M_I f = \{M_{pi,qi}f; i \in \{1,2,\dots I\}\} \quad (15)$$

A simple explanation is illustrated in Figure 3 for 3x3 matrix

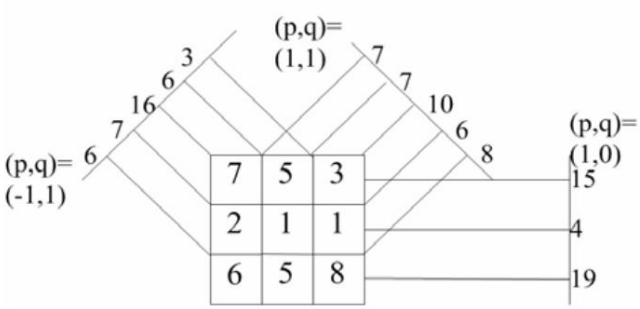

Fig. 3: Example of MT 3x3

A trap can be arrived when all projection value is equal zero, so we calculate the absolute value of each pixel. This case is called Mojette phantoms. Steganography maker took it as advantage for changing information.

We apply MT to I_{uv} with DC zero then we get 3 projections. $(p,q) = (1,0)$ doesn't have the size of other projections so we fill with 0 to have a same size. This new result will pass via MT again and we are interested on $(p,q) = (1,0)$.

5) Wavelet transformation

DWT is common operation in image processing so we will not develop here the details.

We pass to DWT level 2 parameter the input image I which returns the approximation coefficients matrix cA_1 and detail coefficients matrices cH_1, $cV_1$and cD_1 (horizontal, vertical,and diagonal, respectively).cA_1 is the original image resized. We focus with the other coefficients cH_1, $cV_1$and cD_1.

After this computing, we apply again DWT level 2 with cA_1 as input parameter then we have a new coefficients matrix cH_2, $cV_2$and cD_2. They will be exploited on feature extraction.

B. Feature extraction

Here we form one row for each input image. The method to obtain each column is illustrated through different tables below.

Table 1: Spatial domain statistics

Sr. no.	Method
	Calculate $I_s = I + I_{dc} + I_{dr} + I_{dd1} + I_{dd2}$ for each channel RGB
	Fixe the iteration number for ICM
1-3	Calculate the entropy E of $Y_{x,y}$ by considering I_s as input image.

$$E = -P_1 log_2(P_1) - P_0 log_2(P_0) \quad (16)$$

where, P_1 is the probability of presence of 1 and P_0 for 0 because ICM provides binary image.

Table 2: Frequency domain statistics

Sr. no.	Method
	Divide I into 8x8 and apply DCT
	$[I_{-1,1}, I_{1,1}, I_{1,0}]$=MT($I_{uv}$ with DC zero) for each channel RGB with formed blocks
	Form a new matrix J with $I_{-1,1}$, $I_{1,1}$, $I_{1,0}$ by filling $I_{1,0}$ with zeros.
	$[J_{-1,1}, J_{1,1}, J_{1,0}]$=MT(J)
4-6	$J_{1,0}(1,1)$
7-9	$J_{1,0}(1,2)$
10-12	$J_{1,0}(1,3)$

Table 3: Wavelet domain statistics

Sr. no.	Method
13-21	Mean of cH_1, cV_1 and cD_1 for each RGB channel
22-30	Variance of cH_1, cV_1 and cD_1 for each RGB channel
31-39	Skewness of cH_1, cV_1 and cD_1 for each RGB channel
40-48	Kurtosis of cH_1, cV_1 and cD_1 for each RGB channel
49-57	Mean of cH_2, cV_2 and cD_2 for each RGB channel
58-66	Variance of cH_2, cV_2 and cD_2 for each RGB channel
67-75	Skewness of cH_2, cV_2 and cD_2 for each RGB channel
76-84	Kurtosis of cH_2, cV_2 and cD_2 for each RGB channel

Mean:

$$M = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} I(x,y)}{M \times N} \quad (17)$$

Variance:

$$V = \frac{1}{MN-1}\sum_{x=1}^{M}\sum_{y=1}^{N}(I(x,y) - M)^2 \quad (18)$$

Skewness:

$$S = \frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N}\left[\frac{I(x,y)-M}{\sqrt{V}}\right]^3 \quad (19)$$

Kurtosis :

$$K = \left\{\frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N}\left[\frac{I(x,y)-M}{\sqrt{V}}\right]^4\right\} - 3 \quad (20)$$

4. CLASSIFICATION

Classification is one of the important issues in data mining. The goal of classification describes the relational models between classes and attributes from a predefined rule that is set up according to the goal attributes among data.
From image dataset we have two categories of images: steg and clean. Each image has 84 characteristics with one label. SVM will train the data then we test with another group of data set for classification. Radial Basis Function (RBF) is the kernel function that we use during training phase. Matlab software will take care of the compilation.

5. EXPERIMENTAL RESULTS

A. Image dataset construction

Caltech 101 is the image dataset that we deploy for experience use. It consists of pictures of objects belonging to 101 classes, plus one background clutter class. Each image is labelled with a single object. Each class contains roughly 40 to 800 images, totaling 9146 images.

We are not interested in image classes however we separate them in 2 parts: image training data set 70% $I_{\{70\}}$ and image test data set 30% $I_{\{30\}}$. $I_{\{70\}}$ will pass via 4 steganography methods and keep a percentage of the original (clean) image $I_{\{70\}clean}$ as part of data set. Same operation will be done for $I_{\{30\}}$.

B. Test flow

Now, we have image data set described in equation (21) and (22).

$$I_{\{70\}training} = I_{\{70\}clean} + I_{\{70\}steg} \quad (21)$$

$$I_{\{30\}testing} = I_{\{30\}clean} + I_{\{30\}steg} \quad (22)$$

where $I_{\{70\}steg}$ denotes steg images from:

- Research published in International Journal of Innovative Science and Research Technology entitled "An Efficient Algorithm for Image Steganography or Visible Watermarking" in November, 2020 (Me).
- J-UNIWARD: JPEG UNIversal WAvelet Relative Distortion.
- nsF5
- UERD: Uniform Embedding Revisited Distortion

$I_{\{30\}steg}$ steg images from:

- HILL: High-pass, Low-pass, and Low-pass
- S-UNIWARD: Spatial UNIversal Wavelet Relative Distortion
- HUGO: The Highly Undetectable steGO
- WOW: Wavelet Obtained Weights

Our feature extraction algorithm collects the requested data for SVM training which utilizes $I_{\{70\}training}$ and $I_{\{30\}testing}$ dedicated for testing or classification. The system decides on his turn whether the image is steg or clean. Full details is shown in Fig.4.

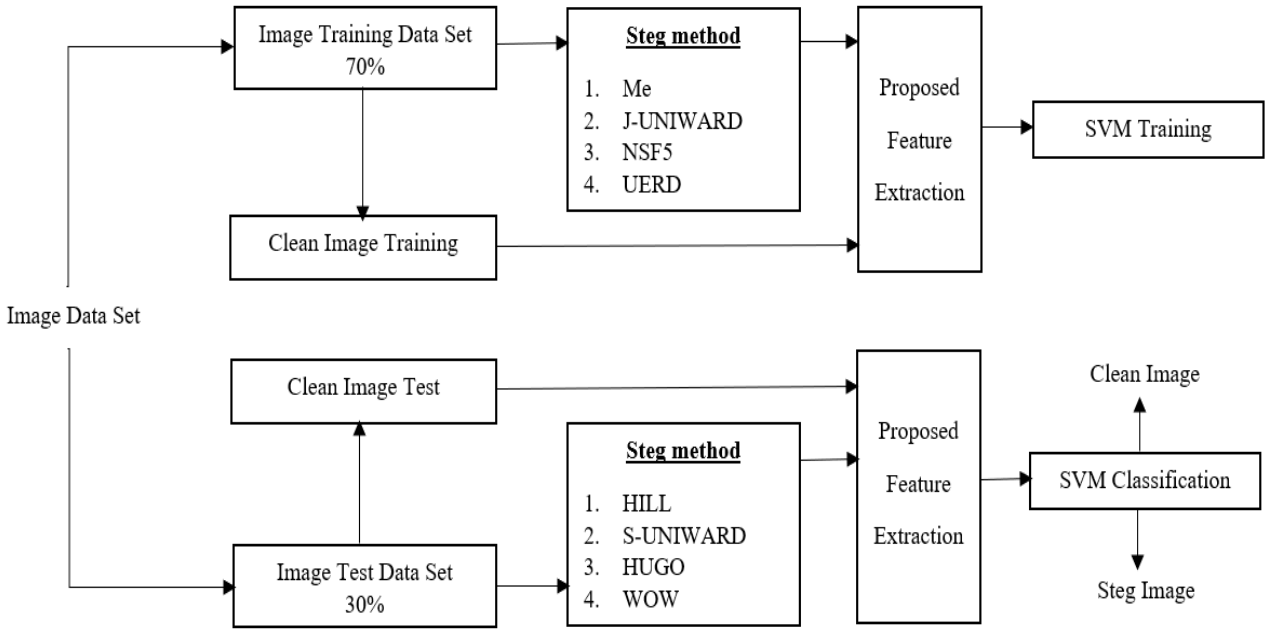

Fig. 4: Block shematic of testing flow

C. Performance results

Table 4 and 5 illustrate the data partition of image data set training and classification. Each 25% of $I_{\{70\}}$ and $I_{\{30\}}$ undergo steganography operation task and 30% present a clean image.

Table 4: Data training partition

	Training data set				
Steganography	**Me**	**J-UNIWARD**	**nsF5**	**UERD**	**Clean**
Percentage (%)	25	25	25	25	30
Partition	1595	1596	1595	1595	1914

Table 5: Data testing composition

	Testing data set				
Steganography	**HILL**	**S-UNIWARD**	**HUGO**	**WOW**	**Clean**
Percentage (%)	25	25	25	25	30
Partition	683	684	683	684	820

The success rate of this steganalysis research is around 97%. It is evident that if we increase the percentage of clean image in data set training and test, this accuracy increases.

HILL algorithm has a particular robustness and evokes a small deterioration in our performance. However, we have a considerable good result with others methods.

Table 6: Accuracy

Steganography	**Total**	**Steg**	**Clean**	**Accuracy (%)**
HILL	683	657	26	96,19
S-UNIWARD	684	667	17	97,51
HUGO	683	665	18	97,36
WOW	684	675	9	98,68
Clean	820	23	797	97,20
Summary	3554	2687	867	97,39

CONCLUSION

Steganography is a large domain and a lot of approach is already in place. Finding one way to track all existing methods is not evident nevertheless it is not a reason to not continue research in steganalysis.

This work exploits spatial, frequency domain simultaneously because each steganography searcher has his own approach. ICM is introduced to detect pixel value change in spatial domain by considering interpixel difference and DCT, MT operate in frequency domain. To ensure a good performance, DWT is required with matrix coefficient mathematical calculation. The accuracy measurement is around 97%, it can be increased by increasing ICM number iteration and extract entropy for each iteration. The disadvantage is that feature extraction process time grows but we will able to detect more hiding information in HILL steganography technical. Another option available is to drop out the second operation of MT in DCT and the image characteristics column increase too and has an impact in SVM training.

ACKNOWLEDGMENT

Authors thank Nitish Yerkiah for his support.

REFERENCES

[1] D. Lou, C. Lin, C. Liu, “Universal steganalysis scheme using support vector machines,” Optical Engineering - OPT ENG. 46. 10.1117/1.2802110, 2007.
[2] K. Jung, “A Study on Machine Learning for Steganalysis.” ICMLSC ,2019.
[3] T. Lindblad, J.M. Kinser. Image Processing Using Pulse-Coupled Neural Networks. 2nd ed., Springer Berlin Heidelberg New York, 2005, pp. 26–27.
[4] N. Pandian, “An Image Steganography Algorithm Using Huffman and Interpixel Difference Encoding,” International Journal of Computer Science & Security (IJCSS), Volume (8): Issue (6): 2014