Color images steganalysis using rgb channel geometric transformation measures

Hasan Abdulrahman, Marc Chaumont, Philippe Montesinos, Baptiste Magnier

SECURITY AND COMMUNICATION NETWORKS
Security Comm. Networks 2015; **00**:1–12
DOI: 10.1002/sec

RESEARCH ARTICLE

Color Images Steganalysis Using RGB Channel Geometric Transformation Measures

Hasan ABDULRAHMAN [2,4], Marc CHAUMONT [1,2,3], Philippe MONTESINOS [4] and Baptiste MAGNIER [4]

[1] Nîmes University, Place Gabriel Péri, 30000 Nîmes Cedex 1, France.
[2] Montpellier University, UMR5506-LIRMM, 34095 Montpellier Cedex 5, France.
[3] CNRS, UMR5506-LIRMM, 34392 Montpellier Cedex 5, France.
[4] Ecole des Mines d'Alès, LGI2P, Parc Scientifique G.Besse, 30035 Nîmes Cedex 1, France.

ABSTRACT

In recent years, information security has received a great deal of attention. To give an example, steganography techniques are used to communicate in a secret and invisible way. Digital color images have became a good medium for digital steganography due to their easy manipulation as carriers via Internet, e-mails, or used on websites. The main goal of steganalysis is to detect the presence of hidden messages in a digital media. The proposed method is a further extension of the authors previous work: steganalysis based on color feature correlation and machine learning classification. Fusing features with those obtained from Color-Rich Models allows increasing the detectability of hidden messages in the color images. Our new proposition uses two types of features, computed between color image channels. The first type of feature reflects local Euclidean transformations and the second one reflects mirror transformations. These geometric measures are obtained by the sine and cosine of gradient angles between all the color channels. Features are extracted from co-occurrence correlation matrices of measures. We demonstrate the efficiency of the proposed framework on three steganography algorithms designed to hide messages in images represented in the spatial domain: S-UNIWARD, WOW, and Synch-HILL. For each algorithm, we applied a range of different payload sizes. The efficiency of the proposed method is demonstrated by the comparison with the previous authors work and the Spatial Color Rich Model and CFA-Aware features for steganalysis.

KEYWORDS
Steganalysis; Color Spatial Rich Model; CFA-aware steganalysis; channel correlation; ensemble classifier; steganography.

Received . . .

1. INTRODUCTION

Steganalysis, the art of detecting hidden information, has received a great deal of attention in recent years. There are many researchers working on solutions ensuring the detection of hidden messages inside digital media. As a result, there are many techniques and methods that are currently used in the field of steganography and steganalysis [1].

Modern information security techniques demonstrate that cryptography alone is not enough to ensure the safe communication of a hidden message. Indeed, it is simple to corrupt, sabotage or delete a file containing secret/encrypted message, as they may be tracked. In addition, the presence of encrypted information itself is valuable information. Additionally, when any person finds and sees an encrypted message, this makes possible its decryption. For these reasons, it is common to work with steganography, encrypting the messages, and then hiding them in a digital medium. By the way, steganography is not intended to replace cryptography but supplement it to make the detectability of the secret messages more and more difficult [2].

More specifically, steganography is the art of hiding the presence of a communication, by embedding messages within a media such as audio, image or video files, in a way that is hard to detect. The steganographer objective is thus to hide the fact that there are information, hidden in a media [3].

Image steganography techniques based on the modification are predominantly classified into the spatial and frequency domains [4]. In the spatial domain, pixel values are

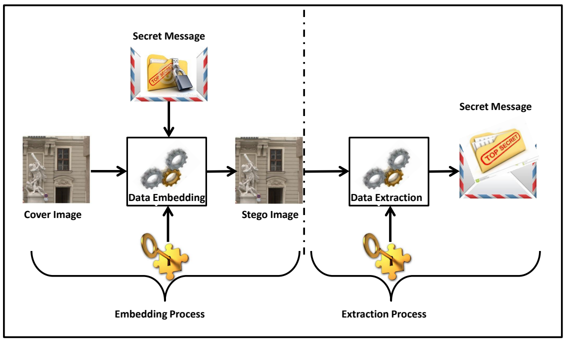


Figure 1. Basic Steganography Model.

used directly to embed the message bits. In the frequency domains [5], the frequential coefficients are used to embed the message. Each domain has several different algorithms. Generally, steganography is made of two parts, messages are embedded inside the digital media in the first part (the embedding) and they are extracted in the second part (the extraction) [6], as illustrated in Fig.1.

Although the embedded messages inside the digital medium involves some slight changes in this medium, these changes modify slight coefficient values of the image [18]. These changes are difficult to identify by a common user. On the other hand, steganalysis research aims to develop some methods, theories and applications that are effectively able to detect these minor modifications in order to detect hidden messages in this medium. Although, the real-world uses significantly more color images than grayscale images, there is a lot of research in steganalysis of grayscale images compared to color images [7].

In this article, we describe further extensions of the recent method described by Addulrahman *et al.* [8]. We propose new features to enhance the Color Rich Model [9], which is formed by co-occurrence matrices of residuals taken across the color channels.

The rest of this paper is organized as follows. Section 2 is dedicated to steganalysis methods for digital color images. Section 2.1 describes Color Spatial Model steganalysis, and Section 2.2 describes Color Filter Array aware features for steganalysis [22]. We present a detailed description of our proposed method in Section 3 by recalling the color channel correlation and the mirror transformations. The ensemble classifier used in this work is explained in Section 4. Experimental results and comparisons are given in Section 5. Finally, Section 6 gives some conclusions and perspectives.

2. RELATED WORK

In recent years, there have been a few techniques involving color steganalysis methods. In this regard, the earliest work was reported by Fridrich *et al.* [7]. The authors have developed an influential approach for color steganalysis to detect stego images which are created by embedding a message inside the pixels randomly (using the Last Significant Bit (LSB) steganography method). They have found the relative number of close colors pairs between the original image and the stego-image. Let us note $(R1, G1, B1)$ and $(R2, G2, B2)$ respectively the three color channels specifying the red, green and blue components of two color images. They show that if two colored pixels $(R1, G1, B1)$ and $(R2, G2, B2)$ for these two images are close, then the condition $(R1 - R2)^2 + (G1 - G2)^2 + (B1 - B2)^2 \leq 3$ must be satisfied. Thus, after the embedding process, the number of unique colors will be increased in stego images more than the number of unique colors in the cover image.

Fridrich *et al.* [10] have introduced a reliable steganalysis algorithm to detect LSB embedding in randomly non-sequential scattered pixels in both 24-bit color and grayscale images. In this method, the message length is derived by searching the lossless capacity in the LSB. Additionally, Westfeld and Pfitzman [11] have applied a robust statistical attack method based on statistical analysis of Pairs of Values ($PoVs$) that are exchanged during message embedding. This method detects very reliable stego images with hidden messages which are embedded in sequential pixels using EZ Stego, S-tools, J-Steg, and Steganos methods. A set of $PoVs$ is used to detect the presence of secret messages in digital images. However, this method is not efficient with embedded messages in random pixels.

Ker, in [12], enhanced techniques for the detection of LSB Matching method from grayscale images into the color images experiment. Beginning with the outlining of Harmsen method [13], by using Histogram Characteristic Function (HCF), Ker has described two new ways: the first one by calibrating the output Center Of Mass (COM) using a down sampled image, and the second way by computing the adjacency histogram instead of the usual histogram to detect an additive noise based steganography.

Thiyagarajan *et al.* [14] has developed a steganalysis method based on color model conversion. Indeed, considering color images, to detect hidden messages, they convert Red (R), Green (G) and Blue (B) channels of the images to the Hue Saturation and Intensity (HSI) color model. Stego images are then generated by implementing different color image formats, using the last significant bit steganography method. Finally, cover and stego images are recognized using a threshold value which depends on the correlation between pixel pairs in terms of color components.

Lyu *et al.* [15] have described a steganalysis algorithm that exploits the inherent statistical regularities of the original images. The statistical model consists of first and higher order color wavelet statistics of noise residuals obtained using predictors of coefficients in Quadratic Mirror Filter (QMF) decomposition of the image from all three color channels. Finally, they estimate that the addition of color statistics provides considerable improvement in overall detection accuracy.

Krichner *et al.* [16] proposed a steganalysis method to detect LSB replacement steganography in color images. Also, the authors have enhanced the Weighted Stego (WS) image steganalysis method [17] by replacing the cover predictor in WS with position specific predictors, to detect stego images produced from covers that exhibit traces of Color Filter Array (CFA) interpolation. This technique explains the local predictability of pixels, depending on their position in the (CFA) interpolation to compute the differences between cover images and stego images. The detector exploits only dependencies within a color channel due to color interpolation at cover generation.

Olguin-Garcia *et al.* [19] have developped a new approach for color image steganalysis depending on Histogram Characteristic Function Center of Mass ($HCFCoM$) detecting histogram changes in each R, G,and B channels. The stego images are created by using LSB Matching steganography method. Then, the Probability Density Function (PDF) is computed to find the adequate threshold, and different threshold values are determined with different payloads.

The most recent and efficient methods in color image steganalysis are explained in detail in the two following sections.

2.1. Color Spatial Rich Model steganalysis

As it is well known, embedding a message in an image modifies some pixel values. Indeed, this modification provides slight changes to the pixel values where the message is embedded. It is a difficult task to detect and extract the sensitive features. Many methods apply high-pass filters to the target image, and then compute high order statistics on the filtered images. Goljan *et al.* [9] have introduced efficient color image features which are an extension of the Spatial Rich Model [18], produced from two different sets of features. First of all, this method extracts the noise residual from each color channel separately. Let us note that X_{ij} is a pixel value of an 8-bit grayscale cover image. We can specify the red, green and blue channel of color images by the following formula:

$$\mathbf{R}_{ij} = \hat{X}_{ij}(\mathcal{N}_{ij}) - c \cdot X_{ij}, \tag{1}$$

where:

- $c \in \mathbb{N}$, is the residual order,
- $\mathcal{N}_{ij}$, is a local neighborhood of pixel X_{ij} at coordinates (i, j),
- $\hat{X}_{ij}(\cdot)$ is a predictor of $c \cdot X_{ij}$, $X_{ij} \notin \mathcal{N}_{ij}$, $X_{ij} \in \{0, ..., 255\}$.

Many diverse submodels built from the differences between neighboring pixels are combined in the Rich Model, all of the submodels $(\mathbf{R}_{ij}) \in \mathbb{R}^{n_1 \times n_2}$ are formed from noise residual images of size $n_1 \times n_2$ computed using high pass filters of the following form:

$$\mathbf{R}_{ij} \leftarrow tranc_T\left(round\left(\frac{R_{ij}}{q}\right)\right), \tag{2}$$

where:

- $R_{ij} = \begin{cases} tranc_T(x) & = x \quad \textbf{for } x \in [-T, T], \\ tranc_T(x) & = T \cdot sign(x) \quad \textbf{otherwise.} \end{cases}$
- q is the quantization step,
- $round$ is a function for rounding to an integer value.

The Spatio-Color Rich Model consists of two different components. On one hand, the Spatial Rich Model ($SRMQ1$) [18] with a fixed quantization $q = 1$ and truncation $T = 2$ yields a dimensionality of 12753 features. These features are computed from each R, G and B color channel separately. Finally, the three dimensionality features are added together to keep the same dimensionality as for grayscale images. On the other hand, from the same noise residuals (i.e. $SRMQ1$), the $CRMQ1$ builds a collection of 3D color co-occurrence matrices, taking three color values at the same position (across the three channels of each pixel). Thus, with fixed truncation $T = 3$ and quantization $q = 1$, $CRMQ1$ produces 5404 features per image.

2.2. *CFA*-aware features steganalysis

Digital cameras capture color images using a single sensor in conjunction with a Color Filter Array (CFA) interpolation. The CFA allows us to capture only one part of the spectrum though the sensor so that only one color is measured at each pixel (red, blue or green) and so the resulting images are called mosaic images. To construct a color image, a demosaicking algorithm is used in order to interpolate each color plane (i.e. CFA interpolations). Several patterns exist for the color filter array, with the most common being Bayer CFA [20]. During this process, the green color channel is the most important factor which determines the luminance of the color image, 50% of the pixels in the Bayer CFA structure are assigned to the green channel, while 25% are assigned to the red channel and 25% to the blue color channel [21].

Goljan *et al.* introduced in [22] the CFA-aware CRM for color image steganalysis. The features are made from two parts, the first one is the Color Rich Model $CRMQ1$ explained in section 2.1 with $T \in \{2, 3\}$. The second part is the CFA-aware feature, which consists of three combinations: RB/GG *split*, $R/B/GG$ *split* and NII/INI *split*.

Let us note, if X has a true-color image size of $n_1 \times n_2$, where n_1 and n_2 are even numbers, $(0 \leq i < n_1, 0 \leq j < n_2)$. Considering a typical Bayer mosaic, the G sub-image has twice as many pixels as the R and B sub-images. We

DOI: 10.1002/sec
Prepared using secauth.cls

must mention that, all the color images used in this method are cropped from one pixel position which is the upper left pixel corresponding to a non-interpolated blue in the Bayer CFA. The color noise residuals Z=($z_{ij}^{(R)}$, $z_{ij}^{(G)}$, $z_{ij}^{(B)}$) is computed as Eq.1, corresponding to CFA used map.

First of all, the following four index sets must be generated:

$$\begin{aligned} X_B &= \{(i,j)|i \text{ even}, j \text{ even}\}, \\ X_{G1} &= \{(i,j)|i \text{ odd}, j \text{ even}\}, \\ X_{G2} &= \{(i,j)|i \text{ even}, j \text{ odd}\}, \\ X_R &= \{(i,j)|i \text{ odd}, j \text{ odd}\}. \end{aligned}$$

Four 3D co-occurrence matrices are computed from residual samples due to the above index sets.

$$C_{d_1d_2d_3}^{(B)} = \sum_{(i,j)\in X_B} \left[(z_{ij}^{(R)}, z_{ij}^{(G)}, z_{ij}^{(B)}) = (d_1, d_2, d_3) \right], \tag{3}$$

$$C_{d_1d_2d_3}^{(G1)} = \sum_{(i,j)\in X_{G1}} \left[(z_{ij}^{(R)}, z_{ij}^{(G)}, z_{ij}^{(B)}) = (d_1, d_2, d_3) \right], \tag{4}$$

$$C_{d_1d_2d_3}^{(G2)} = \sum_{(i,j)\in X_{G2}} \left[(z_{ij}^{(R)}, z_{ij}^{(G)}, z_{ij}^{(B)}) = (d_1, d_2, d_3) \right], \tag{5}$$

$$C_{d_1d_2d_3}^{(R)} = \sum_{(i,j)\in X_R} \left[(z_{ij}^{(R)}, z_{ij}^{(G)}, z_{ij}^{(B)}) = (d_1, d_2, d_3) \right]. \tag{6}$$

From the above four co-occurrence matrices, three combinations of features are generated to form the total number of features with the $CRMQ1$ set:

The first combination is called $RB/GGsplit$ which generates 4146 features. $C_{d_1d_2d_3}^{(R)}$ and $C_{d_1d_2d_3}^{(B)}$ are treated and added together, the same thing is applied to $C_{d_1d_2d_3}^{(G1)}$ and $C_{d_1d_2d_3}^{(G2)}$ as in Eq.'s 5 and 6.

$$C_{d_1d_2d_3}^{(RB)} = C_{d_1d_2d_3}^{(B)} + C_{d_3d_2d_1}^{(B)} + C_{d_1d_2d_3}^{(R)} + C_{d_3d_2d_1}^{(R)}, \tag{7}$$

$$C_{d_1d_2d_3}^{(GG)} = C_{d_1d_2d_3}^{(G1)} + C_{d_3d_2d_1}^{(G1)} + C_{d_1d_2d_3}^{(G2)} + C_{d_3d_2d_1}^{(G2)}. \tag{8}$$

$R/B/GG$ *split* represents the second set and produces 10323 features. This part can be considered as an important component in this method, because it gives a considerable number of features. It can be generated from the concatenation of $C_{d_1d_2d_3}^{(R)}$, $C_{d_1d_2d_3}^{(B)}$, and $C_{d_1d_2d_3}^{(G1)} + C_{d_1d_2d_3}^{(G2)}$.

The third set corresponds to the NII/INI *split*; 'N' meaning non-interpolated and 'I' interpolated respectively, in the RGB triple. The 'NII' pixels correspond to the same set as RB but the two co-occurrence matrices are directionally symmetrized differently. This set generates 5514 features from two co-occurrence matrices:

$$C_{d_1d_2d_3}^{(NII)} = C_{d_3d_2d_1}^{(B)} + C_{d_1d_2d_3}^{(R)}, \tag{9}$$

$$C_{d_1d_2d_3}^{(INI)} = C_{d_1d_2d_3}^{(GG)}. \tag{10}$$

All these features are gathered in a one dimensional vector, while all detectors are trained as binary classifiers implemented using Kodovsky ensemble classifiers [26], as explained in the following Section 4.

3. FEATURES DESCRIPTION

Our proposition is to enrich the $SCRMQ1$ with an inter-channel correlation which is composed of three sets of features. The first set, produced by [9], gives 18157 features. The second set, produced by our first method [8], gives 3000 features. Additionally, the third set, produced by a second method, gives 3000 features; they are obtained from the new correlation of different R, G and B channel gradients, as shown in Table I.

Table I. Features description with their dimmensionalities corresponding to q and T.

Feature set	$SCRMQ1$	$\mathcal{C}_{RG}/\mathcal{C}_{RB}$	$\mathcal{S}_{RG}/\mathcal{S}_{RB}$
Dim. Symmetry	yes	yes	yes
Dimension	18157	3000	3000

The following section recalls the RGB Channel Correlations which gives an explanation to our proposition, then section 4 explains the ensemble classifiers used in this approach.

3.1. RGB Channel Correlation

In this section, we introduce an inter-channel correlation measure, and demonstrate that it can be linked to first order Euclidean invariants (see Hilbert [23] for the invariant theory). Such invariants have mainly been used for stereo-matching [24]. In this paper, we show that the information provided can enhance steganography detection. The underlying idea here, is that if one channel has been affected by steganography, the inter channel correlation will measure the local modifications.

Starting from the local correlation of red and green channels (similar to the correlation of red and blue channels):

$$Corr_{R,G}(i,j,k,l) = \sum_{(i',j')\in \mathcal{W}_{i,j}} X_{i',j'}^{(R)} \cdot X_{k+i',l+j'}^{(G)} \tag{11}$$

with:

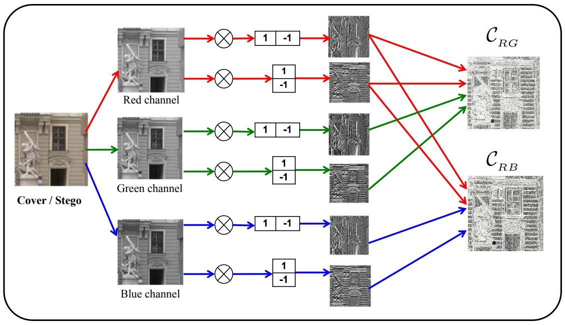


Figure 2. Features extraction: Cosine of the gradient angles [8].

- $X^{(R)}_{i',j'} \in [0, 255]$, being a pixel value at position (i', j') in the red channel,
- $X^{(G)}_{k,l} \in [0, 255]$, being a pixel value at position (k, l) in the green channel,
- $\mathcal{W}_{i,j}$, representing a small window centered in (i, j).

Considering $(k, l) = (0, 0)$ and a limited development of $X^{(R)}$ and $X^{(G)}$ around (i, j), then:

$$Corr_{R,G}(i,j,0,0) = \sum_{\substack{\mathbf{h}=(i'-i,j'-j)\\ (i',j')\in\mathcal{W}_{i,j}}} \left(X^{(R)}_{i,j} + \nabla X^{(R)}_{i,j} \cdot \mathbf{h}\right)\left(X^{(G)}_{i,j} + \nabla X^{(G)}_{i,j} \cdot \mathbf{h}\right). \tag{12}$$

Developing this equation leads to four terms. Three of which are constant or not informative, then there is only one informative term :

$$\nabla X^{(R)}_{i,j} \cdot \nabla X^{(G)}_{i,j}\,. \tag{13}$$

If only one channel has been altered locally, the gradient in this channel is modified. Consequently, the scalar product of two channel gradients reflects the change in the cosine of the difference between the two gradient angles.

Similarly, we can apply the same computation for the red and blue channel and then obtain :

$$\nabla X^{(R)}_{i,j} \cdot \nabla X^{(B)}_{i,j}\,. \tag{14}$$

As stated by Gouet *et al.* [24] (and following the Hilbert theory [23]), it is unnecessary to investigate the $\nabla X^{(G)}_{i,j} \cdot \nabla X^{(B)}_{i,j}$ term, as it is already implicitly contained in the first two expressions (Eq. 13 and 14).

Normalizing these expressions, we obtain the cosine of rotation angles, between channel gradients:

$$\mathcal{C}_{RG} = \frac{\nabla X^{(R)}_{i,j} \cdot \nabla X^{(G)}_{i,j}}{|\nabla X^{(R)}_{i,j}|\,|\nabla X^{(G)}_{i,j}|}, \tag{15}$$

$$\mathcal{C}_{RB} = \frac{\nabla X^{(R)}_{i,j} \cdot \nabla X^{(B)}_{i,j}}{|\nabla X^{(R)}_{i,j}|\,|\nabla X^{(B)}_{i,j}|}. \tag{16}$$

Fig. 2 illustrates our preprocessing steps [8] to obtain the cosine of rotation angles, between channel gradient. Note that gradients derivatives of each channel are estimated by a convolution with a [-1; 1] mask (horizontal and vertical).

3.2. Mirror transformations

In the preceding section, we have seen that the inter-channel correlation is linked with the scalar product of gradients (i.e. Euclidean invariants). This means that if we are able to measure the absolute value of a rotation angle between two channel gradients, we still need the direction of the rotation, which is linked this time to Mirror transformations (as illustrated in Fig. 3).

Our proposition is to add two new features sets based on the determinants of channel gradients. Similar to that applied in the recent work of Abdulrahman *et al.* [8], the features are directly linked to the correlation in order to obtain new features of Sine of the gradients angle. Finally, as illustrated in Fig. 4, we normalize these determinants by gradient norms to obtain the sine of the rotations:

$$\mathcal{S}_{RG} = \frac{\nabla X^{(R)}_{i,j}[0] \cdot \nabla X^{(G)}_{i,j}[1] - \nabla X^{(R)}_{i,j}[1] \cdot \nabla X^{(G)}_{i,j}[0]}{|\nabla X^{(R)}_{i,j}|\,|\nabla X^{(G)}_{i,j}|}, \tag{17}$$

$$\mathcal{S}_{RB} = \frac{\nabla X^{(R)}_{i,j}[0] \cdot \nabla X^{(B)}_{i,j}[1] - \nabla X^{(R)}_{i,j}[1] \cdot \nabla X^{(B)}_{i,j}[0]}{|\nabla X^{(R)}_{i,j}|\,|\nabla X^{(B)}_{i,j}|}, \tag{18}$$

with $\nabla X[0]$ (resp. $\nabla X[1]$) the first (resp. second) component of the vector ∇X i.e. corresponding to the horizontal and the vertical derivatives (see Fig. 4).

3.3. Complete feature set

Our features, are computed from $\mathcal{C}_{RG}$, $\mathcal{C}_{RB}$, $\mathcal{S}_{RG}$ and $\mathcal{S}_{RB}$ correlations by computing the co-occurrence matrices as in the Rich Model [18]. We used different values of the quantization $q \in \{0.1, 0.3, 0.5, 0.7, 0.9, 1\}$ with fixed truncation T=1. The reason for using these different values of quantization q is that $\mathcal{G}_{RG}$, $\mathcal{G}_{RB}$, $\mathcal{S}_{RG}$ and $\mathcal{S}_{RB}$ belong

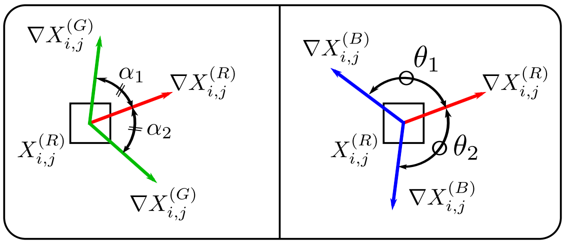


Figure 3. Rotation angle between two channel gradients $\cos(\alpha_1) = \cos(\alpha_2)$ but $\sin(\alpha_1) = -\sin(\alpha_2)$
$\cos(\theta_1) = \cos(\theta_2)$ but $\sin(\theta_1) = -\sin(\theta_2)$.
Sine is essential to determine the direction of the rotation.

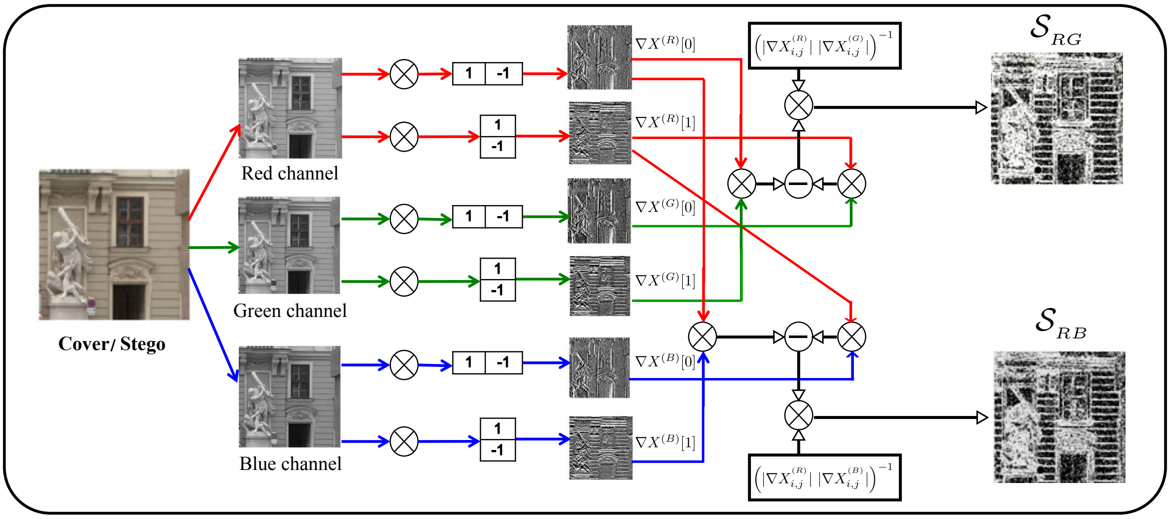


Figure 4. Features of extraction: Sine of the gradients angles extracting information from the direction of the local rotation.

to $[-1, 1]$. Moreover, the use of these values gives more accurate features and avoids the generation of too many zero values caused by the truncation step in the co-occurrence vector. For each quantization, we obtain 12 submodels from methods 1 [8] and 12 submodels from the new proposed method 2 [*]. The submodels from the Color Rich Models [9] give 18157 features, those of the method 1 [8] give 3000 features, and those of our proposed method 2 give 3000 features. Accordingly, the final feature vector collects a final set of 24157 features.

4. THE ENSEMBLE CLASSIFIERS

An ensemble of classifiers [25] is a set of classifiers whose individual decisions are combined and organized into weighted or unweighted votes to classify the data sets (in this work, features represent these data sets, as detailed in the previous sub-section).

Modern steganalysis methods for digital images are based on feature extraction. These methods need machine learning techniques to detect if the media contains hidden messages or not. In our work, we choose ensemble classifiers [26] because of their efficient classification performance for large scale learning.

Kodovsky *et al.* [26] proposed ensemble classifiers[†] which is a machine learning tool for steganalysis, consisting of many classifier L independently trained (B_l) designed to keep complexity to a minimum and make the overall process simple.

Each base learner is trained on randomly selected subspaces d_{sub}-dimensionals of the original feature space, from the entire full d-dimension feature space. The authors use Ficher Linear Discriminants (FLD) as base learners and the final decision is made by aggregating the decision of individual base learners. Let d be a full dimensional feature space, N^{trn} and N^{tst} the number of training and testing samples from each class. First, the classifiers construct a number L of FLD base learners (B_l) with $l \in \{1, ..., L\}$. Each one performs its learning on a subspace of d_{sub} dimension, where $d_{sub} << d$. From the i^{th} image, a feature vector, $f_i \in \mathbb{R}^d$, is extracted, and then mapped, such as $\mathbb{R}^d \rightarrow \{0, 1\}$, where $'0'$ stands for cover and $'1'$ for stego.

In the learning phase, each classifier learns to map a feature vector f_i, to the correct class number:

$$\begin{aligned} FLD_l : \mathbb{R}^d &\rightarrow \{0, 1\} \\ f_i &\rightarrow FLD_l\left(f_i\right). \end{aligned}$$

Each classifier uses the training database to compute the orthogonal vector to the hyperplane separating the two classes. For a test feature, the l^{th} base learner reaches its decision by computing a projection and comparing it to a threshold. After collecting all L decisions, the final classifier selects the class which has received the most votes. Then, the decision threshold of each base learner is adjusted to minimize the total detection error under an equal prior on the training data [26]:

$$P_E = min_{P_{FA}} \frac{1}{2}\left[P_{FA} + P_{MD}\left(P_{FA}\right)\right], \qquad (19)$$

where P_{FA} represents the false alarm probability and P_{MD} the missed detection probability.

[*] For method 1 (resp. method 2) we use one symmetrized spam14h and one spam14v submodel, with 25 features each. We also use the minmax22h, minmax22v, minmax24, minmax34h, minmax34v, minmax41, minmax34, minmax48h, minmax48v, and minmax54 submodels with 45 features for each. All submodels are gathered in a one dimension vector to erect a dimensionality of $(2 \times 25 + 10 \times 45) \times 6 = 3000$ features. For more details on submodels construction, the reader is invited to look at article [18].

[†] Ensemble classifier is available at http://dde.binghamton.edu/download/ensemble.

DOI: 10.1002/sec
Prepared using secauth.cls

5. EXPERIMENTAL RESULTS

5.1. Experimental setup and protocol

All our features are calculated and formed in a one dimensional vector from 10000 color covers and 10000 color stego images for each payload of steganography methods. These features are ready to enter in the classifier. The classifiers were implemented using the ensemble classifier [26] with many FLD as a base learner. In this paper, the detection accuracy is measured by the total probability of the average of testing errors under equal priors as in Eq. 19. 5000 images from a database are randomly chosen for the training sets and 5000 for the testing sets. The ensemble classifiers apply a vote to estimate the error of detection. This process is repeated 10 times to obtain $\bar{P}_E$, the average of testing errors. $\bar{P}_E$ quantify the detectability and are collected for each method and payload to evaluate the steganalysis method. Given the decision values, ROC curves are obtained. As illustrated in Fig. 8, the area under the ROC curves is calculated as the accuracy of the ensemble classifiers.

5.1.1. Image Dataset

A raw image is a class of computer file containing untouched pixel information coming from the digital camera sensor (i.e. the pure information). These files hold a large amount of meta-information about the image generated by the camera [27].

In our work, the color image database is very carefully built depending on the CFA idea. We collected raw images from two subsets which are the most standard, and have the highest number of images captured (i.e. the Dresden Image Database's [28] 3500 full-resolution Nikon digital camera raw color images and the Break Our Steganographic System (BOSSbase[‡]), with 1000 Canon digital camera raw color images).

In order to obtain color images in Portable Pixel Map (PPM) format of size 512×512, all images take the same CFA map layout, as illustrated in Fig. 6. For this process, two steps are required. The first step consists of using a demosaicking algorithm to convert raw images into demosaicked images. The second step consists of cropping five areas from one image. Fig. 5 shows sample images produced by the cropping step.

First we used the demosaicking algorithm Patterned Pixel Grouping (PPG) from the dcraw software[§] to convert raw images into RGB images. As illustrated in Fig.6, the obtained images are such that the Bayer Pattern is always of the type $RGBG$ (red channel pixel is placed at an even position). We wrote a spatial code to start the crop from the red channel position. Indeed, from one image, this code randomly selected the red channel position and cropped five images using a size of 512×512 pixels, so that all blocks share the same CFA map layout. The final number of images is 10000 RGB color images with a size of 512×512.

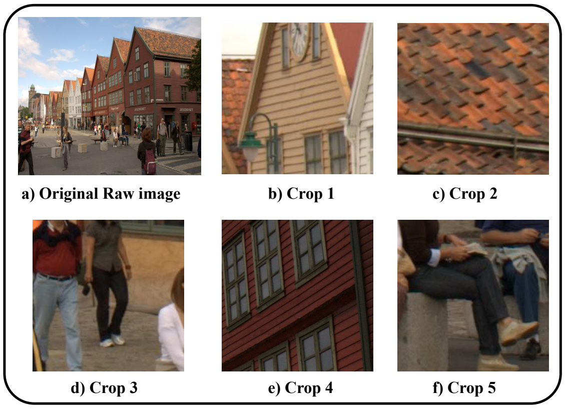


Figure 5. Sample images of our database built by random cropping from locations of red channel pixels (even position) in a Bayer pattern :

a) Original raw image 3906×2602,

b) crop 1 position x=2116, y=1928,

c) crop 2 position x=902, y=1182,

d) crop 3 position x=3080, y=436,

e) crop 4 position x=1866, y=1778,

f) crop 5 position x=650, y=1032.

5.1.2. Embedding methods

The stego images are obtained using three spatial-domain steganography algorithms. The first method is the Spatial-UNIversal WAvelet Relative Distortion (S-UNIWARD[¶]) steganography algorithm [29]. The second method is the Wavelet Obtained Weights (WOW[‖]) steganography algorithm [30]. Finally, the third method is the Synchronizing the Selection Channel (Synch-HILL[**]) steganography algorithm [31].

These algorithms are used to embed messages into color images by decomposing the R, G and B channels as three grayscale images and embedding the same proportion payload into each channel. Also, different tested payload sizes are used $\{0.01, 0.05, 0.1, 0.2, 0.3, 0.4 \text{ and } 0.5\}$ Bit Per Channel *(BPC)*.

[‡] BOSSbase can be accessed at http://www.agents.cz/boss/BOSSFinal.

[§] dcraw code is available at http://www.cybercom.net/defin/dcraw.

[¶] S-UNIWARD steganography method is available at http://dde.binghamton.edu/download/stego_algorithms/.

[‖] WOW steganography method is available at http://dde.binghamton.edu/download/stego_algorithms/.

[**] Synch-HILL steganography method is available at http://dde.binghamton.edu/download/stego_algorithms/.

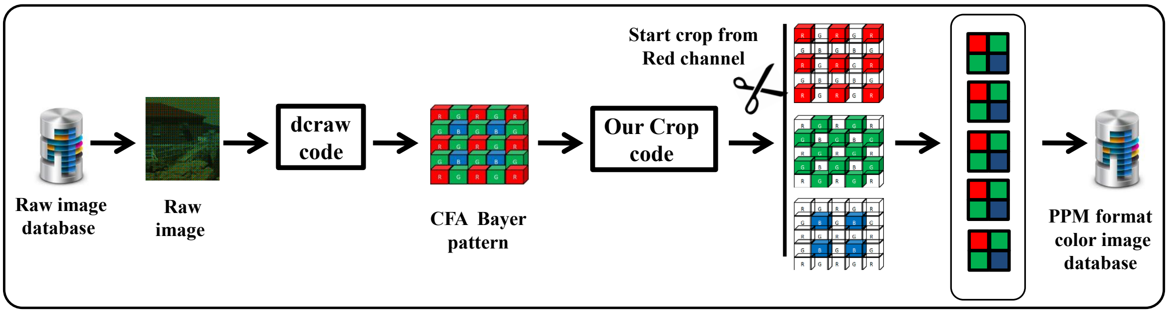


Figure 6. The preprocessing steps for building our database depending on the CFA idea.

5.2. Results and Discussion

This section contains the experimental results of our proposed method. We illustrate these results in Table II. S-UNIWARD, WOW and Synch-HILL methods were tested with different relative payloads $\{0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$ (bpc) against three approaches: method 1 [8], the Color Rich Model [9] and the CFA-aware features steganalysis [22]. We used the same set of payload values with the same embedding methods. Our proposed second method, that uses both the sine and cosine of the gradients angle, achieved higher performance by registering 88.76%, 87.93% and 88.07% detection rates for S-UNIWARD, WOW and synch-HILL respectively (with the payload 0.5 bpc). The Color Rich Model method [9] is less efficient because it achieved respectively 86.14%, 85.27% and 85.25% detection. Also, the CFA-aware features method [22] is less efficient because it achieved respectively 87.61%, 87.04% and 87.42% detection rates. Close to the CFA-aware features method, the method of Abdulrahman *et al.* [8] is less efficient because it achieved respectively 87.54%, 86.63% and 86.77% detection rates. We noted the same trend with the rest of the payload values, as shown in Table II.

Additionally, as shown in Table II, the method of Abdulrahman *et al.* [8], that uses the cosine of the gradients angle, achieved higher performance than Color Rich Model method [9]; by registering 87.54%, 86.63% and 86.77% detection rates for S-UNIWARD, WOW and synch-HILL respectively with the payload 0.5 bpc. For the same payloads range, the Color Rich Model method [9] is less efficient because it achieved respectively 86.14%, 85.27% and 85.25% detection rates on the same test samples. Also, as shown in Table II, our proposed second method, that uses the sine and cosine of the gradients angle, achieved higher performance than CFA-aware features steganalysis method [22]; by registering 88.76%, 87.93% and 88.07% detection rates for S-UNIWARD, WOW and synch-HILL respectively with the payload 0.5 bpc. The CFA-aware features steganalysis method [22] is less efficient because it achieved respectively 87.61%, 87.04% and 87.42% detection rates on the same test samples.

Moreover, curves in Fig.7 (a) S-UNIWARD, (b) WOW and (c) synch-HILL steganography method also, illustrate the comparison between the proposed second method and the compared methods. As a result, the average testing error of the proposed second method is less than the first proposition, the Color Rich Model and CFA-aware features method. That proves the importance of the additional 3000 features proposed by the second method.

Another experiment involved embedding the entire payload in only one channel of the color image, i.e. with payload 0.2 bpc and 0.4 bpc in the green channel only. In this case, the detection rate becomes higher than the same payload distributed equally between the three color channels. Table III illustrates the comparison of detection rates between the S-UNIWARD, WOW and synch-HILL methods with payloads 0.2 bpc and 0.4 bpc embedded in one channel only and in the three channels separately. Fig. 8 (a), (b) and (c) show the ROC curves, illustrating the performance of our method 2. Finally, this experiment revealed that it is easier to detect a hidden message in only one channel than a message that is spread across all channels.

Table III. Our proposed method 2 detection rate of S-UNIWARD, WOW and Synch-HILL steganography methods at 0.2 bpc and 0.4 bpc payload embedding in the green channel compares with equal embedding in three channels.

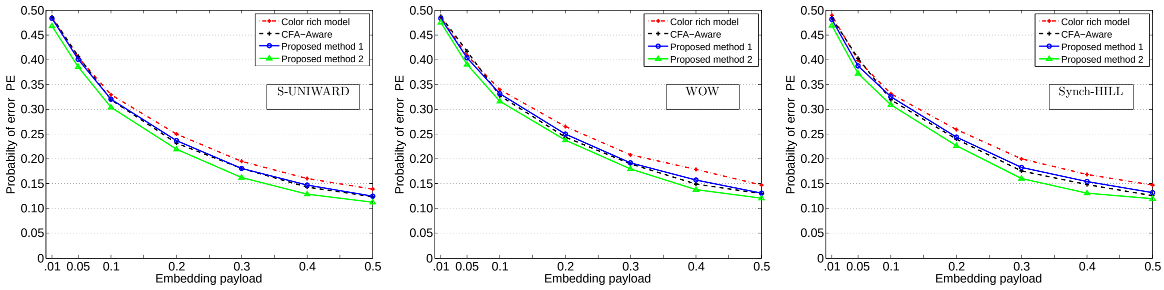
	S-UNIWARD		WOW		Synch-HILL	
Payload	G%	RGB%	G%	RGB%	G%	RGB%
0.2	90.02	78.09	88.51	76.19	89.23	77.31
0.4	96.77	87.11	94.83	86.16	94.87	86.89

6. CONCLUSION

In this paper, we have proposed new features for steganalysis of color images. Starting from the Color Rich Model proposed by Goljan *et al.* [9], we have shown that this method could be greatly enhanced by considering

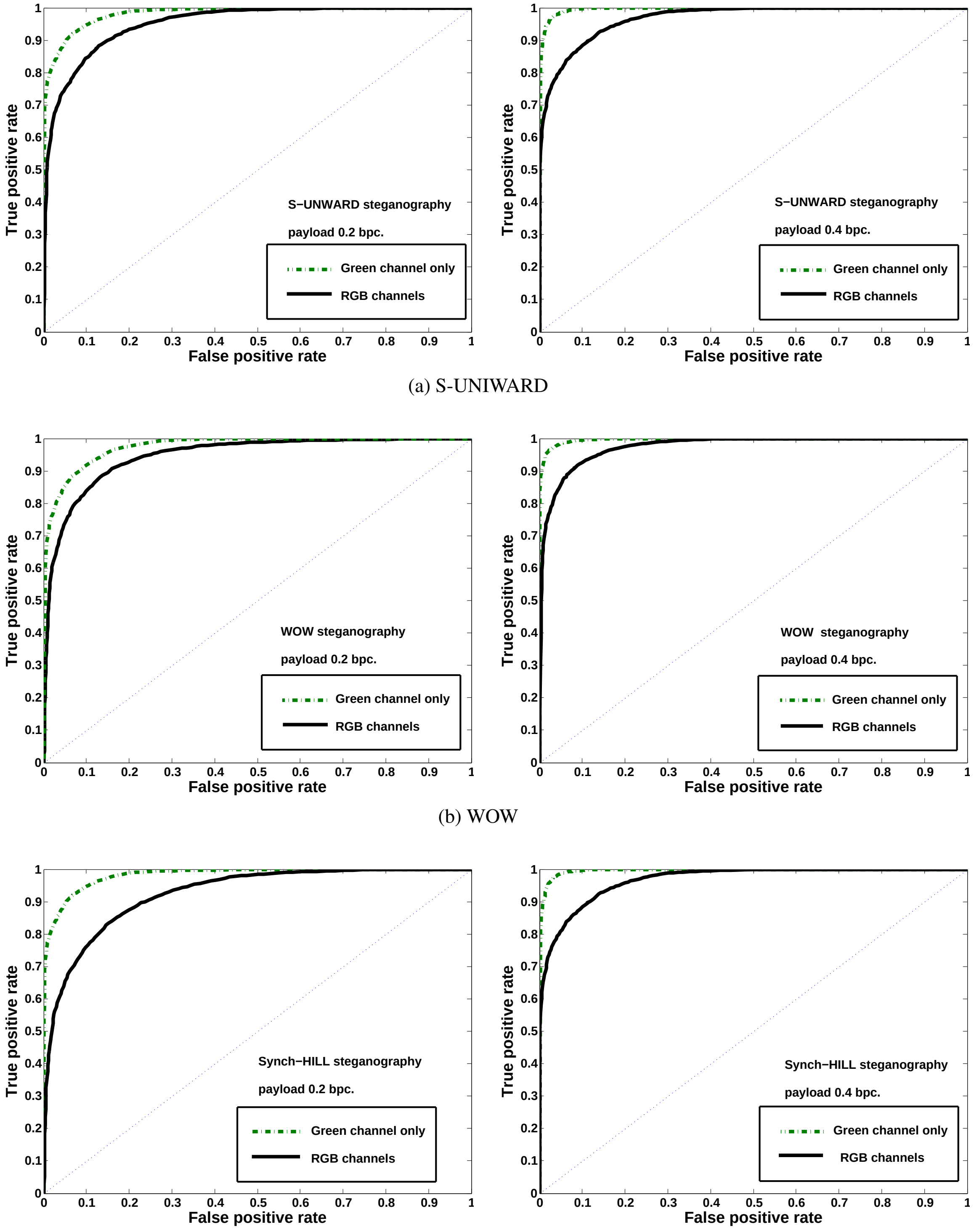
Table II. Numerical values of the average testing error $\bar{P}_E$ and the detection rate $\mathcal{P}_D\%$ for three steganography methods. For easier navigation the dark gray background column presents the first method of Abdulrahman $et\ al.$ [8] and the light gray background column presents the second proposed method.

	Color Rich		**CFA-Aware**		**Method 1**		**Method 2**	
payload	$\bar{\mathbf{P}}_\mathbf{E}$	$\mathcal{P}_\mathbf{D}\%$	$\bar{\mathbf{P}}_\mathbf{E}$	$\mathcal{P}_\mathbf{D}\%$	$\bar{\mathbf{P}}_\mathbf{E}$	$\mathcal{P}_\mathbf{D}\%$	$\bar{\mathbf{P}}_\mathbf{E}$	$\mathcal{P}_\mathbf{D}\%$
S-UNIWARD								
0.01	0.4841	51.59	0.4863	51.37	0.4830	51.70	**0.4680**	**53.20**
0.05	0.4045	59.55	0.4072	59.28	0.4010	59.90	**0.3859**	**61.41**
0.1	0.3298	67.02	0.3194	68.06	0.3203	67.97	**0.3037**	**69.63**
0.2	0.2498	75.02	0.2317	76.83	0.2370	76.30	**0.2191**	**78.09**
0.3	0.1947	80.53	0.1806	81.94	0.1808	81.92	**0.1623**	**83.77**
0.4	0.1599	84.01	0.1429	85.71	0.1470	85.30	**0.1289**	**87.11**
0.5	0.1386	86.14	0.1239	87.61	0.1246	87.54	**0.1124**	**88.76**
WOW								
0.01	0.4850	51.50	0.4875	51.25	0.4836	51.64	**0.4753**	**52.47**
0.05	0.4092	59.08	0.4174	58.26	0.4042	59.58	**0.3906**	**60.94**
0.1	0.3397	66.03	0.3275	67.25	0.3317	66.83	**0.3161**	**68.39**
0.2	0.2654	73.46	0.2440	75.60	0.2502	74.98	**0.2381**	**76.19**
0.3	0.2081	79.19	0.1895	81.05	0.1918	80.82	**0.1793**	**82.07**
0.4	0.1783	82.17	0.1487	85.13	0.1574	84.26	**0.1384**	**86.16**
0.5	0.1473	85.27	0.1296	87.04	0.1307	86.63	**0.1207**	**87.93**
Synch-HILL								
0.01	0.4893	51.07	0.4843	51.57	0.4814	51.83	**0.4687**	**53.13**
0.05	0.3991	60.09	0.4030	59.70	0.3879	61.21	**0.3720**	**62.80**
0.1	0.3311	66.89	0.3189	68.11	0.3258	67.42	**0.3086**	**69.14**
0.2	0.2595	74.05	0.2394	76.06	0.2438	75.62	**0.2269**	**77.31**
0.3	0.1997	80.03	0.1753	82.47	0.1829	81.71	**0.1607**	**83.93**
0.4	0.1684	83.16	0.1478	85.22	0.1540	84.60	**0.1311**	**86.89**
0.5	0.1475	85.25	0.1258	87.42	0.1323	86.77	**0.1193**	**88.07**



(a) S-UNIWARD (b) WOW (c) Synch-HILL

Figure 7. Avarage testing error $\bar{P}_E$ as a function of the payload for (a) S-UNIWARD,(b) WOW and (c) WOW steganography methods, comparison between the steganalysis methods (Color Rich Model, CFA-aware features steganalysis, method 1 [8] and proposed method 2).



(a) S-UNIWARD

(b) WOW

(c) Synch-HILL

Figure 8. ROC curves using our proposed method 2 feature set, for (a) S-UNIWARD, (b) WOW and (c) Synch-HILL steganography methods for payloads 0.2 bpc (up) and 0.4 bpc (down), to compare the detectability when embedding messages in only one channel with embedding messages spread in all channels.

local deformation between channels. We have proposed to add to the Color Rich Model a new set of features based on local Euclidean and mirror transformation. The euclidean transformation, proposed by Abdulrahman *et al.* [8], is estimated by a first set of features derived from correlations between the gradients of red, green and blue channels. Since these features give the cosine of angles between gradients, we still do not know the direction of the rotation between two channel gradients. Then, we have shown that by taking into account mirror transformations, we can obtain the missing information of the direction of the local rotation. According to this analysis, we add a new set of features based on the sine of local rotation angles. These two sets of features are then incorporated in the Rich Model using co-occurrence matrices in order to obtain 6000 features. The first and second set gives 3000 features each [8]. The total feature set is formed from the Color Rich Model, plus the two new sets demonstrated in this work, in order to build a vector of a total of 24157 features. We used a quantization step with a set of values that differs from the Color Rich Models. All feature vectors are fed to the Ensemble Classifier. The Ensemble Classifier is used to detect the presence of hidden messages. Eventually, multiple steganalysis comparisons have been achieved between the proposed method and the initial Color Rich Model [9] and CFA-aware features steganalysis method [22]. We have used three steganography methods (S-UNIWARD, WOW and Synch-HILL) with seven different payloads. All the experiments show that our new method outperforms the Color Rich Model and the CFA-aware feature steganalysis.

Our future work will focus on developing a new steganalysis method for digital color images using steerable filters.

6.1. Acknowledgements

The authors wish to thank the Iraqi Ministry of Higher Education and Scientific Research for funding and supporting this work.

REFERENCES

1. Rhodes-Ousley, Mark, 2013. *The Complete Reference Information security*, (Second Edition), Copyright by The McGraw-Hill Companies, ISBN: 978-0-07-178436-8.
2. Bloisi, Domenico Daniele and Iocchi Luca, 2007. *Image Based Steganography and Cryptography*, International Conference on Computer Vision Theory and Applications (VISAPP) (1), pp. 127–134, Barcelona, Spain, March 8-11.
3. C. Hosmer, and C. Hyde, 2003. *Discovering Covert Digital Evidence*, Third Digital Forensic Research Workshop (DFRWS), pp. 1–5, Cleveland, Ohio, USA, August 6-8.
4. K. Shrikants and S. L. Nalbalwar, 2010. *Review: Steganography-Bit Plane Complexity Segmentation (BPCS) Technique*, International Journal of Engineering Science and Technology, Vol. 2, No. 9, pp. 4860–4868.
5. J. Fridrich, 2005. *Feature-based Steganalysis for JPEG Images and its Implications for Future Design of Steganographic Schemes*, 6th International Workshop on Information Hiding, LNCS, Vol. 3200, pp. 67–81, Springer-Verlag, Berlin Heidelberg.
6. J. Fridrich, 2009. *Steganography in Digital Media: Principles, Algorithms, and Applications*, Cambridge University Press, Cambridge, England.
7. J. Fridrich, and M. Long, 2000. *Steganalysis of LSB Encoding in Color Images*, IEEE International Conference on Multimedia and Expo (ICME), Vol. 3, pp. 1279–1282, New York, NY, USA.
8. H. Abdulrahman, M. Chaumont, P. Montesinos and B. Magnier, 2015. *Color Images Steganalysis Using Correlation Between RGB Channels*, 10th International Conference on Availability, Reliability and Security, (IWCC), pp. 448–454, Toulouse, France, August 24-28.
9. M. Goljan, J. Fridrich and R. Cogranne, 2015. *Rich Model for Steganalysis of Color Images*, In IEEE National Conference on Parallel Computing Technologies (PARCOMPTECH), pp. 185–190, Campus, Mathikere, Bengaluru, India, Febuary 19-20.
10. J. Fridrich, M. Goljan and R. Du, 2001. *Reliable Detection of LSB Steganography in Color and Grayscale Images*, Proceedings of ACM Workshop on Multimedia and Security, pp. 27–30, Ottawa, Canada.
11. A. Westfeld and A. Pfitzmann, 2000. *Attacks on steganographic systems*, Information Hiding, Springer, Vol. 1768, pp. 61–76.
12. A. D. Ker, 2005. *Resampling and the Detection of LSB Matching in color bitmaps*, In International Society for Optics and Photonics, Electronic Imaging pp. 1–15.
13. J. J. Harmsen and W. A. Pearlman, 2003. *Steganalysis of Additive-Noise Modelable Information Hiding*, In Electronic Imaging International Society for Optics and Photonics, In: Proc. of SPIE, pp. 131–142.
14. P. Thiyagarajan, G. Aghila, and V. P. Venkatesan, 2011. *Steganalysis using Color Model Conversion*, International Journal of Signal and Image Processing (SIPIJ), Vol. 2, No. 4, pp. 201–211.
15. S. Lyu and H. Farid, 2004. *Steganalysis using Color Wavelet Statistics and One-Class Support Vector Machines*, Proceedings (SPIE), Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents VI, San Jose, CA, Vol. 5306, pp. 35–45.
16. M. Kirchner and R. Bohme, 2014. *Steganalysis in Technicolor Boosting WS Detection of Stego Images From CFA-Interpolated Covers*, In IEEE International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP), pp. 3982–3986,

Florence, Italy, May 4–9.

17. J. Fridrich and M. Goljan, 2004. *On Estimation of Secret Message Length in LSB Steganography in Spatial Domain*, In International Society for Optics and Photonics in Electronic Imaging, in Proc. EI SPIE, Vol. 5306, pp. 23–34, Jose San, CA.
18. J. Fridrich and J. Kodovsky, 2012. *Rich Models for Steganalysis of Digital Images*, In IEEE Transactions on Information Forensics and Security, Vol. 7, No. 3, pp. 868–882.
19. H. J. Olguin-Garcia, O. U. Juarez-Sandoval, M. Nakano-Miyatake, H. Perez-Meana, 2015. *Color Image Steganalysis Method for LSB Matching*, Proceedings of the International Conference on Security and Management (SAM), the Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), pp. 309, Las Vegas, Nevada, USA, July 27-30.
20. B. E. Bayer, 1975. *Color Imaging Array*, Google Patents, US Patent 3,971,065, filed March 5, 1975, and issued July 20, 1976.
21. J. Wang, C. Zhang, and P. Hao, 2011. *New Color Filter Arrays of high Light Sensitivity and high Demosaicking Performance*, In 18th IEEE International Conference on Image Processing (ICIP), pp. 3153–3156, Brussels, Belgium, September 11-14.
22. M. Goljan and J. Fridrich, 2015. *CFA-Aware Features for Steganalysis of Color Images*, Proc. SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics, Vol. 9409, San Francisco, CA, February 8–12.
23. D. Hilbert, 1993. *Theory of Algebraic Invariants*, Cambridge University Press, Cambridge.
24. V. Gouet, P. Montesinos, and D. Pelé, 1998. *A Fast Matching Method for Color Uncalibrated Images using Differential Invariants*, Proceedings of the British Machine Vision Conference (BMVC), pp. 1–10, Southampton, September 3–7.
25. T. G. Dietterich, 2000. *Ensemble Methods in Machine Learning*, In First International Workshop on Multiple Classifier Systems (MCS), pp. 1–15, Sardinia, Italy, June 21–23.
26. J. Kodovsky, J. Fridrich, and V. Holub, April 2012. *Ensemble Classifiers for Steganalysis of Digital Media*, In IEEE Transactions on Information Forensics and Security, Vol. 7, No. 2, pp. 432–444.
27. L. Yuan and J. Sun, 2011. *High Quality Image Reconstruction from Raw and JPEG Image Pair*, In 13th IEEE International Conference on Computer Vision (ICCV), pp. 2158–2165, Barcelona, Spain, November 6–13.
28. T. Gloe and R. Böhme, 2010. *The Dresden Image Database for Benchmarking Digital Image Forensics*, Journal of Digital Forensic Practice, Vol. 3, pp. 150–159.
29. V. Holub, J. Fridrich, and T. Denemark, 2014. *Universal Distortion Function for Steganography in an Arbitrary Domain*, EURASIP Journal on Information Security, Special Issue on Revised Selected Papers of the 1st ACM Information Hiding (IH) and the ACM Multimedia and Security (MMSec) Workshop, Vol. 1, pp. 1–13.
30. V. Holub and J. Fridrich, 2012. *Designing Steganographic Distortion using Directional Filters*, In IEEE International Workshop on Information Forensics and Security (WIFS), pp. 234–239, Tenerife, Spain, December 2–5.
31. T. Denemark and J. Fridrich, 2015. *Improving Steganographic Security by Synchronizing the Selection Channel*, Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security, (ACM), pp. 5–14, Portland, Oregon, June 17–19.