 CMC, vol.59, no.1, pp.315-329, 2019

Color Image Steganalysis Based on Residuals of Channel Differences

Yuhan Kang[1], Fenlin Liu[1], Chunfang Yang[1,*], Xiangyang Luo[1] and Tingting Zhang[2]

Abstract: This study proposes a color image steganalysis algorithm that extracts high-dimensional rich model features from the residuals of channel differences. First, the advantages of features extracted from channel differences are analyzed, and it shown that features extracted in this manner should be able to detect color stego images more effectively. A steganalysis feature extraction method based on channel differences is then proposed, and used to improve two types of typical color image steganalysis features. The improved features are combined with existing color image steganalysis features, and the ensemble classifiers are trained to detect color stego images. The experimental results indicate that, for WOW and S-UNIWARD steganography, the improved features clearly decreased the average test errors of the existing features, and the average test errors of the proposed algorithm is smaller than those of the existing color image steganalysis algorithms. Specifically, when the payload is smaller than 0.2 bpc, the average test error decreases achieve 4% and 3%.

Keywords: Color channel, channel difference, color image, steganalysis, steganography.

1 Introduction

Steganography is the science and art of concealing secret messages in unsuspected digital media, and the generated digital media are referred to as stego objects. Steganalysis is used to detect the stego objects and extract the secret messages. Among steganographic algorithms [Xiang, Li, Hao et al. (2018); Zhang, Qin, Zhang et al. (2018); Yao, Zhang and Yu (2016); Tang, Chen, Zhang et al. (2015)] with text, image, and video as cover, the image steganographic algorithms attract widespread attention from steganalysts. Steganalysts have designed numerous effective steganalysis algorithms [Yang, Liu, Luo et al. (2013); Pevný, Bas and Fridrich (2010)] for a number of classic image steganographic algorithms such as least significant bit steganography (LSB), F5 [Westfeld (2001)], and MB [Sallee (2005)], where many steganalysis algorithms could even locate or extract the secret messages in certain cases [Yang, Luo, Lu et al. (2018)]. Currently, steganalysts have also proposed a number of high-dimensional steganalysis

[1] Zhengzhou Science and Technology Institute, Zhengzhou, 450001, China.

[2] Wirtschaftswissenschaftliche Fakultät, Heinrich-Heine-Universität, Universitätsstraße 1, Haus 19, Zimmer 218, Düsseldorf 40225, Germany.

[*] Corresponding Author: Chunfang Yang. Email: chunfangyang@126.com.

features, such as the rich model [Fridrich and Kodovský (2012)], the projection spatial rich model (PSRM) [Holub and Fridrich (2013b)], maxSRM [Denemark, Sedighi, Holub et al. (2014)], and phase-aware projection features (PHARM) [Holub and Fridrich (2015)], for the new content-adaptive image steganographic algorithms, such as highly undetectable stego (HUGO) [Pevný, Filler and Bas (2010)], wavelet obtained weights (WOW) [Holub and Fridrich (2012)], universal wavelet universal wavelet relative distortion (UNIWARD) steganography [Holub and Fridrich (2013a)], and studied how to select the most effective features from them [Ma, Luo, Li et al. (2018)].

Steganalysts currently focus on steganalysis for steganographic algorithms with grayscale images as covers [Song, Liu, Yang et al. (2015); Zhang, Liu, Yang et al. (2017)]. However, color images are more widely used in our life and daily work. And for color image steganography, some steganalysis algorithms have also been proposed. Here, they are classified into various types, including steganalysis algorithms based on changes of color number [Fridrich, Du and Long (2000); Su, Han, Huang et al. (2011)], steganalysis algorithms based on inter-channel texture consistency [Abdulrahman, Chaumont, Montesinos et al. (2016b)], steganalysis algorithms based on co-occurrence matrices across channels [Goljan, Fridrich and Cogranne (2014); Goljan and Fridrich (2015); Liao, Chen and Yin (2016)], steganalysis algorithms based on inter-channel prediction errors [Lyu and Farid (2004); Liu, Sung, Xu et al. (2006); Li, Zhang and Yu (2014)], and steganalysis algorithms based on combinations of different channel features [Abdulrahman, Chaumont, Montesinos et al. (2016a)].

The steganalysis algorithms based on changes of color number primarily used the characteristic that steganography will increase the number of colors or similar color pairs to detect color stego images. For example, Fridrich et al. [Fridrich, Du and Long (2000)] extracted the ratio of similar color pairs in the color pairs appearing as features, and Su et al. [Su, Han, Huang et al. (2011)] embedded fixed ratios of random information in the given image, and then extracted the increased numbers of different colors and similar color pairs as features to detect the color stego image of LSB steganography. The steganalysis algorithms based on inter-channel texture consistency primarily extracted the statistical features that could reflect the strong consistency between the texture of different channels to detect color stego image. For example, Abdulrahman et al. [Abdulrahman, Chaumont, Montesinos et al. (2016b)] used the cosine and sine of the angle between the gradients of different channels to depict the texture direction consistency of different channels, extract their co-occurrence matrices, and combined them with SCRMQ1 (spatio-color rich model with quantization step $q=1$) [Goljan, Fridrich and Cogranne (2014)] to improve the detection accuracy of color stego images. The steganalysis algorithms based on the co-occurrence matrices across channels primarily captured the correlation between different channels by extracting the co-occurrence matrices across the residuals of three channels to detect the color stego image. For example, Goljan et al. [Goljan, Fridrich and Cogranne (2014)] extracted the co-occurrence matrices between the residuals of three channels and the rich model features of each channel, and then merged them into the color image steganography detection features, SCRMQ1. Goljan et al. [Goljan and Fridrich (2015)] divided the image pixels into blocks according to the color filter array characteristics from the imaging principle of

a camera, and then computed the co-occurrence matrices across residuals in different channels from each block and merged them as the final feature set for steganalysis. Liao et al. [Liao, Chen and Yin (2016)] obtained the regions with complex texture in each channel and the regions with complex texture in any channel, and then calculated the co-occurrence matrices from residuals of these two types of regions in each channel and combined them as steganalysis features, that improved the detection accuracy of new adaptive steganography, such as WOW and S-UNIWARD. The steganalysis algorithms based on inter-channel prediction errors considered the correlation between channels when calculating the prediction errors of the image elements (such as pixel or wavelet coefficients) or their features, and then combined the features of prediction errors with other features to detect the color stego image. For example, Lyu et al. [Lyu and Farid (2004)] utilized the correlation between horizontal, vertical, and diagonal wavelet sub-band coefficients of different scales and different color channels to calculate logarithmic prediction errors, extracted their statistic features such as mean, variance, skewness, and kurtosis, and then used one-class support vector machines to realize pure blind detection of color image steganography. Liu et al. [Liu, Sung, Xu et al. (2006)] measured the correlation coefficients between the LSB planes of different color channels and the correlation coefficients between the prediction errors of each channel, and then combined them with the features reflecting the correlation in each channel to improve the color stego image detection performance of LSB matching. Li et al. [Li, Zhang and Yu (2014)] calculated the prediction errors of other channels from Y-channel by differences, and extracted the Markov features, Pevný features (PEV), co-occurrence matrix features, and their calibration features of the prediction error, and then combined them with the statistical features in the Y-channel to improve the detection performance of the color JPEG image steganography. The steganalysis algorithms based on the combination of different channel features extracted the features from three channels of the color image and then combined them to obtain the steganalysis feature. For example, Abdulrahman et al. [Abdulrahman, Chaumont, Montesinos et al. (2016a)] used steerable Gaussian filters to construct gradient magnitudes and derivatives of each channel, and then calculated the co-occurrence matrices from them as features SGF (steerable Gaussian filters) which is combined with the spatio rich model with quantization step $q=1$ (SRMQ1) features and color rich model with quantization step $q=1$ (CRMQ1) to train the steganalyzer.

Compared with simply applying the steganography detection algorithm of grayscale images in three channels and fusing the results, the above algorithms improve the detection accuracy of color image steganography. Specifically, Abdulrahman et al. [Abdulrahman, Chaumont, Montesinos et al. (2016a)] combines three color steganalysis features with great performance, and obtains a better result. However, two of them, SRMQ1 and SGF, fail to consider a number of types of relationships between different color channels. As the pixels in different color channels exhibit strong correlative dependence and interplay, it is very likely that extracting these two features from the differences of different channels will further improve the accuracy of color image steganalysis.

In view of this, this study proposes a steganalysis algorithm for color images based on residuals of channel differences. The advantages of the channel difference feature are

first analyzed from the view of the variance change rate. A steganalysis feature extraction method, based on residuals of channel differences, is then proposed. Then the proposed feature extraction method is used to improve two high-dimensional rich model features--the spatio rich model with quantization step $q=1$ (SRMQ1) and steerable Gaussian filters (SGF). Combining the improved features with the 22563-dimension steganalysis feature reported in Abdulrahman et al. [Abdulrahman, Chaumont, Montesinos et al. (2016a)], a color image steganalysis algorithm based on residuals of channel differences is proposed. Finally, the experimental results indicate that, when secret messages are embedded into the R, G, or B color channels by new adaptive steganography such as WOW and S-UNIWARD, the improved features significantly enhance the steganalysis performance compared to the original features, and the proposed color image steganalysis algorithm has significantly smaller detection error rates than existing algorithms. Specifically, at smaller embedding rates, the maximum decreasing amplitude of detection errors can attain 3%.

2 Advantage of channel difference features

Let X and Y be random variables representing pixel values at the same position in channel I and channel II, respectively, and N_X and N_Y be random variables representing steganographic signals at the same positions in channel I and channel II, respectively. The means of X, Y, N_X and N_Y are denoted as μ_X, μ_Y, μ_{N_X} and μ_{N_Y}, respectively, and the variances are denoted as σ_X^2, σ_Y^2, $\sigma_{N_X}^2$ and $\sigma_{N_Y}^2$. It assumed that the correlation coefficient between X and Y is r.

Proposition 2.1. When random message bits are embedded into color images by additive noise, if the variances of the steganographic signal in the same positions in channel I and channel II are equal, that is:

$$\sigma_{N_X}^2=\sigma_{N_Y}^2, \tag{1}$$

and the correlation coefficients between pixel values in the same positions in channel I and channel II and their variances satisfy the following relationships:

$$r>\left|\frac{\sigma_X^2-\sigma_Y^2}{2\sigma_X\sigma_Y}\right|, \tag{2}$$

then the variance change rate of the difference between the pixels in the same position in channel I and channel II after steganography, $\Delta D(X-Y)$, is greater than the variance change rate of the pixel at the corresponding position of either channel I or channel II, $\Delta D(X)$ or $\Delta D(Y)$.

Proof. When random message bits are embedded into color images by additive noise, because of the randomness of the embedded information, the stego noise is independent of the pixel value. Therefore, the variances of pixels X and Y in the same positions in channel I and channel II after steganography are as follows:

$$D(X+N_X)=\sigma_X^2+\sigma_{N_X}^2 \tag{3}$$

$$D\left(Y+N_Y\right)=\sigma_Y^2+\sigma_{N_Y}^2 \tag{4}$$

The variance change rates of pixels X and Y in the same positions in channel I and channel II are:

$$\Delta D\left(X\right)=\frac{D\left(X+N_X\right)-D\left(X\right)}{D\left(X\right)}=\frac{\sigma_{N_X}^2}{\sigma_X^2} \tag{5}$$

$$\Delta D\left(Y\right)=\frac{D\left(Y+N_Y\right)-D\left(Y\right)}{D\left(Y\right)}=\frac{\sigma_{N_Y}^2}{\sigma_Y^2} \tag{6}$$

In the cover image, if pixels at the same positions in channel I and channel II are differentiated, the variance of the difference is as follows:

$$D\left(X-Y\right)=\sigma_X^2+\sigma_Y^2-2r\sigma_X\sigma_Y \tag{7}$$

In the stego image, if pixels at the same positions in channel I and channel II are differentiated, the variance of the difference is as follows:

$$D\left(X+N_X-Y-N_Y\right)=\sigma_X^2+\sigma_Y^2-2r\sigma_X\sigma_Y+\sigma_{N_X}^2+\sigma_{N_Y}^2 \tag{8}$$

Therefore, the variance change rate of the difference between the pixels in the same positions in channel I and channel II after steganography is:

$$\Delta D\left(X-Y\right)=\frac{D\left(X+N_X-Y-N_Y\right)-D\left(X-Y\right)}{D\left(X-Y\right)}=\frac{\sigma_{N_X}^2+\sigma_{N_Y}^2}{\sigma_X^2+\sigma_Y^2-2r\sigma_X\sigma_Y} \tag{9}$$

Subtracting the variance change rate of channel I from the variance change rate of the pixel difference between channel I and channel II, the following result is obtained:

$$\Delta D\left(X-Y\right)-\Delta D\left(X\right)=\frac{\sigma_X^2\sigma_{N_Y}^2-\sigma_{N_X}^2\sigma_Y^2+2r\sigma_X\sigma_Y\sigma_{N_X}^2}{\sigma_X^2\left(\sigma_X^2+\sigma_Y^2-2r\sigma_X\sigma_Y\right)} \tag{10}$$

The denominator of (10) is always positive because the correlation coefficient between the pixels in the same positions in channel I and channel II is in the range [-1, 1], viz. $-1\le r\le 1$. Therefore, when the correlation coefficient between pixels in the same positions in channel I and channel II in the cover image satisfies the relationship shown in (11), (10) is positive. In other words, the variance change rate of the difference between pixels in the same positions in channel I and channel II is greater than the variance change rate of the pixel in the same position in channel I.

$$r>\frac{\sigma_{N_X}^2\sigma_Y^2-\sigma_{N_Y}^2\sigma_X^2}{2\sigma_X\sigma_Y\sigma_{N_X}^2}=\frac{\sigma_Y^2-\frac{\sigma_{N_Y}^2}{\sigma_{N_X}^2}\sigma_X^2}{2\sigma_X\sigma_Y} \tag{11}$$

Similarly, by subtracting the variance change rate of channel II from the variance change rate of the difference between pixels in the same positions in channel I and channel II, it also can be derived that the variance change rate of the difference between pixels in the same positions in channel I and channel II is greater than that of channel II when the correlation coefficient between pixels in the same positions in channel I and channel II in the cover image satisfies the relationship shown in (12).

$$r > \frac{\sigma_{N_Y}^2 \sigma_X^2 - \sigma_{N_X}^2 \sigma_Y^2}{2\sigma_X \sigma_Y \sigma_{N_Y}^2} = \frac{\sigma_X^2 - \frac{\sigma_{N_X}^2}{\sigma_{N_Y}^2}\sigma_Y^2}{2\sigma_X \sigma_Y} \tag{12}$$

Eqs. (11) and (12) show that if $\sigma_{N_X}^2 = \sigma_{N_Y}^2$ and the correlation coefficient r satisfies the relationship (2), the variance change rate of the difference between pixels in the same positions in channel I and channel II after the steganography is greater than the variance change rate of the pixel in the corresponding position of either channel I or channel II. Therefore, the proposition is proved.

One of the necessary conditions of proposition 1 is that the variances of the steganographic signal in the same positions in channel I and channel II satisfy (1). In [Sangwine and Horne (1998)], Sangwine et al. reported that there is always a strong correlation between the three color channels R, G, and B of natural color images and gave the correlation coefficients between color channels as $r_{\text{B-R}} \approx 0.78$, $r_{\text{R-G}} \approx 0.98$, and $r_{\text{G-B}} \approx 0.94$. This strong positive correlation makes the texture complexity of the same region in different channels significantly close.

When the same adaptive steganography algorithm is used to embed information in two channels, the similar texture complexity in the same position of two channels results in the distortion of the two corresponding pixels being significantly close after changing. As the probability of change for each pixel in adaptive steganography is determined by the distortion function and the length of the embedded information, the probability of two pixels in the same position being changed is approximately equal when the random information with equivalent lengths are embedded in the two channels. This also makes the variance of the steganographic signal at the same position approximately equal, that is,

$$\sigma_{N_X}^2 \approx \sigma_{N_Y}^2 \tag{13}$$

Therefore, one of the necessary conditions of proposition 1, (1), is reasonable.

Another necessary condition of proposition 1 is that the correlation coefficients between the pixels in the same positions in channel I and channel II and their variances satisfy (2). As reported in Sangwine et al. [Sangwine and Horne (1998)], the strong positive correlation between the two channels results in the left-hand side of (2), the correlation coefficient between two channels, being greater than zero and close to one. In addition, the similar texture complexity of the same position in the two channels makes the variances of the pixel values in the same position approximately equal, resulting in the right-hand side of (2) being approximately zero. Therefore, the correlation coefficients between pixels in the same positions in different color channels and their variances should be able to satisfy (2) in the bulk of color images. The proportion of images whose two channels satisfy (2) will be calculated from the 10,000 color images from BOSSbase database[3].

Let channel I and channel II represent channel R and channel G, respectively, and the value of the right-hand side of (2) is subtracted from the value of the left-hand side for each image, that is:

[3] http://agents.fel.cvut.cz/stegodata/RAWs/

$$res(r_{RG},\sigma_R,\sigma_G)=r_{RG}-\left|\frac{\sigma_R^2-\sigma_G^2}{2\sigma_R\sigma_G}\right| \tag{14}$$

In a color image, if the correlation coefficient between channel R and channel G and their variance results in (14) being greater than zero, it can be concluded that channel R and channel G of this image satisfy (2). From the statistical results shown in Fig. 2(a), there are 9992 images whose channel R and channel G satisfy (2) in the 10,000 color BOSSbase images. Similarly, there are 9887 images whose channel R and channel B satisfy (2), and 9927 images whose channel G and channel B satisfy (2), as shown in Figs. 2(b) and 2(c). It can be seen that the ratio of images whose two color channels satisfy (2) is greater than 98%, which indicates that the second necessary condition of proposition 1 is reasonable.

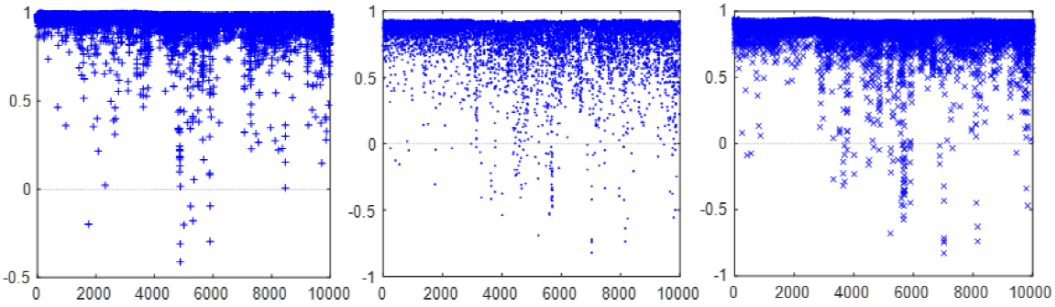

Figure 1: Value of function "res" (14) with correlation coefficient between different color channels and their variances as parameters (x-label: Number of pictures and y-label: Value of "res"): (a) Channel R and channel G, (b) Channel R and channel B, and (c) Channel G and channel B

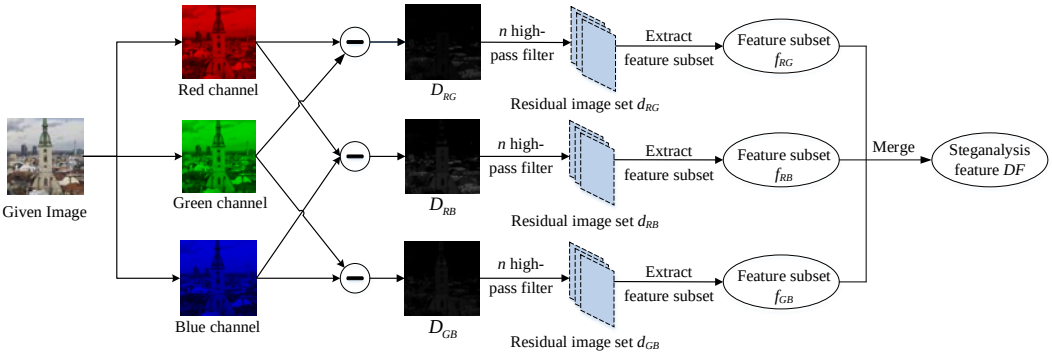

Figure 2: Procedure of steganalysis feature extraction based on residuals of channel difference

The above analysis of the conditions of proposition 1 indicates that, for most of the color images, the variance change rate of the difference between pixels in the same positions in different channels after steganography should be greater than that of the corresponding

pixel in any of the channels. When the means of the steganographic signal, N_X and N_Y, are zero, the mean change rate of the pixel value after steganography is zero and the mean change rate of the difference between pixels in the same positions in different channels is also zero. Therefore, proposition 1 indicates that steganography has a greater effect on the distribution of pixel differences between two channels than on any single channel. In addition, the features extracted from the pixel difference between two channels should be able to detect the color stego image more effectively.

3 Color image steganalysis features based on residuals of channel differences

In existing color image steganalysis algorithms, some features are independently extracted from the residuals of each color channel, and then combined. According to the analysis in Section 2, if the residuals are computed from the differences between different color channels, the steganalysis features extracted from them should have better detection performance. Therefore, this section proposes the following steganalysis feature extraction method based on the residuals of channel differences. As shown in Fig. 3, the primary steps of the method are as follows:

1) Calculate the channel differences between any two of the three colors channels R, G, or B, in a color spatial image, as follows:

$$\begin{cases} D_{RG}(i,j) = X_R(i,j) - X_G(i,j) \\ D_{RB}(i,j) = X_R(i,j) - X_B(i,j) \\ D_{GB}(i,j) = X_G(i,j) - X_B(i,j) \end{cases} \tag{15}$$

where D_{RG}, D_{RB}, and D_{GB} are the differences between channels red and green, red and blue, and green and blue, respectively;

2) The n high-pass filters are used to filter the three channel differences D_{RG}, D_{RB}, and D_{GB}, and then the residual image sets $\mathbf{d}_{RG}$, $\mathbf{d}_{RB}$, and $\mathbf{d}_{GB}$ of three channel differences are generated as follows:

$$\begin{cases} \mathbf{d}_{RG} = \{d_{RG,1}, d_{RG,2}, \ldots, d_{RG,n}\} \\ \mathbf{d}_{RB} = \{d_{RB,1}, d_{RB,2}, \ldots, d_{RB,n}\} \\ \mathbf{d}_{GB} = \{d_{GB,1}, d_{GB,2}, \ldots, d_{GB,n}\} \end{cases} \tag{16}$$

3) Each residual image in the residual image sets $\mathbf{d}_{RG}$, $\mathbf{d}_{RB}$, and $\mathbf{d}_{GB}$ is seen as a feature extraction source, from which statistical features such as the co-occurrence matrices or Markov transition probability matrices are extracted as the feature subsets $\mathbf{f}_{RG}$, $\mathbf{f}_{RB}$, and $\mathbf{f}_{GB}$:

$$\begin{cases} \mathbf{f}_{RG} = \{f_{RG,1}, f_{RG,2}, \ldots, f_{RG,n}\} \\ \mathbf{f}_{RB} = \{f_{RB,1}, f_{RB,2}, \ldots, f_{RB,n}\} \\ \mathbf{f}_{GB} = \{f_{GB,1}, f_{GB,2}, \ldots, f_{GB,n}\} \end{cases} \tag{17}$$

4) Merge the features in feature subsets $\mathbf{f}_{RG}$, $\mathbf{f}_{RB}$, and $\mathbf{f}_{GB}$ to obtain the steganalysis feature based on the residual of channel differences.

4 Color image steganalysis algorithm based on residuals of channel differences

In existing color steganalysis features, SRMQ1 feature has become one of the mainstream steganalysis features because it considers a number of types of relationships between neighboring samples of noise residuals obtained by linear and non-linear filters and achieves excellent performance. In recent years, SGF feature is proposed as a new feature to significantly improve the steganalysis performance. This feature takes into account the correlation between adjacent pixels inside each channel. The image content components are eliminated by a number of filters of different orientations, which helps to capture the steganographic signals with higher signal-to-noise ratio, and results in a more reliable detection of stego images. However, these two features fail to consider the correlation between channels. This section tries to apply the feature extraction method proposed in Section 3 to SRMQ1 and SGF features, which are called DSRMQ1 and DSGF. And a color image steganalysis algorithm is proposed by combining the DSRMQ1 and DSGF features with the CRMQ1, SRMQ1, and SGF features. As supervised learning techniques typically outperform unsupervised learning techniques in both the accuracy and efficiency [Xiang, Zhao, Li et al. (2018)], supervised learning techniques are more widely used in steganalysis. And the ensemble classifier [Kodovský, Fridrich and Holub (2012)] has become a popular learning tool used in steganalysis. Therefore, it will also be used in the proposed color image steganalysis algorithm. The algorithm comprises steganalyzer training and stego image detection. The detailed procedures are as follows:

Algorithm 1: Training color steganalyzer.

Input: Color image training set, including cover training images and corresponding stego training images.

Output: Trained steganalyzer.

1) Steganalysis features extraction. A 39722-dimensional steganalysis feature is extracted from each training image as follows.

I. CRMQ1, SRMQ1, and SGF features extraction. The CRMQ1, SRMQ1, and SGF features with dimensions 22563 are extracted by the method in Abdulrahman et al. [Abdulrahman, Chaumont, Montesinos et al. (2016a)];

II. Channel differences computation. Compute the differences between two pixel values in the same position in any two color channels to get the channel differences;

III. DSRMQ1 feature extraction. The method presented in Section 3 is used to extract the 12753-dimensional DSRMQ1 feature from the channel differences D_{RG}, D_{RB}, and D_{GB};

IV. DSGF feature extraction. The method presented in Section 3 is used to extract the 4406-dimensional DSGF features from the channel differences D_{RG}, D_{RB}, and D_{GB};

V. Features merging. Merge the features extracted in asteps I, III, and IV to generate the 39722-dimensional color image steganalysis feature.

2) Ensemble classifier training. The group of label and corresponding steganalysis feature of each training image are taken as a training sample to train the ensemble classifier, that will be used as the steganalyzer.

Algorithm 2: Detecting color stego image.

Input: The given color image and the steganalyzer obtained by Algorithm 1.

Output: The label of the given image. If the given image is a stego image, the label is set as +1, otherwise, the label is set as -1.

1) Steganalysis feature extraction. For each color image to be detected, the same procedure and parameters of step 1) in algorithm 1 are used to extract the 39722-dimensional steganalysis feature of the given image.

2) Classifying the given image. Take the 39722-dimensional steganalysis feature of the given image as input, and use the steganalyzer trained by Algorithm 1 to distinguish whether the given image is a stego image.

5 Experimental results and analysis

10,000 raw images downloaded from BOSSbase[4] were scaled to generate 10000 color cover images in "tiff" format with sizes of 512×512 pixels. The two typical adaptive steganography algorithms, WOW [Holub and Fridrich (2012)] and S-UNIWARD [Holub and Fridrich (2013a)], were used to embed pseudo-random information in the R, G, and B color channels of cover images with payloads of 0.05, 0.1, 0.2, 0.3, and 0.4 bits per channel pixel (bpc), and then generated $2\times5=10$ groups of color stego images. Then, the performance of the proposed DSRMQ1 and DSGF features and color image steganalysis algorithm were tested with these images.

In each experimental unit, 5000 images were randomly selected as training cover images, and the corresponding 5000 stego images were used as training stego images. The remaining 5000 cover images and 5000 stego images were used as test cover and stego images, respectively. And the average testing errors under equal priors [Fridrich and Kodovský (2012)] were calculated to evaluate the performance. For each payload of a steganographic algorithm, 10 experimental units were performed. The median of the average testing errors under equal priors of 10 experimental units was taken to measure detection performance. The smaller the value, the better the steganalysis performance.

5.1 Performance of steganalysis features based on residuals of channel differences

The steganalysis performance of SRMQ1 and SGF features in Abdulrahman et al. [Abdulrahman, Chaumont, Montesinos et al. (2016a)], the DSRMQ1 and DSGF features proposed in Section 4, and the combined features DSRMQ1+DSGF and DSRMQ1+DSGF+SRMQ1+SGF were tested with above experimental settings in this subsection. Tab. 1 presents the average testing errors under equal priors of the above feature types for WOW and S-UNIWARD. It can be seen that, for both the steganographic algorithms, the performance of DSRMQ1 and DSGF features are significantly better than those of SRMQ1 and SGF features. The average test errors of DSRMQ1 features for WOW and S-UNIWARD are smaller than those of the original SRMQ1 feature by the decreasing amplitudes reaching 12.19% and 13.43% respectively. The average test errors of DSGF feature for WOW and S-UNIWARD steganography are smaller than those of original SGF

[4] http://agents.fel.cvut.cz/stegodata/RAWs/

features, with decreasing amplitudes reaching 16.30% and 17.17%, respectively. If the DSRMQ1 and DSGF features are combined, the combined features exhibit better performance. Compared with the average test errors of original SRMQ1+SGF features, those of the DSRMQ1+DSGF features for WOW and S-UNIWARD are smaller, with decreasing amplitudes reaching 12.82% and 14.51%, respectively. And the average test errors of the combined DSRMQ1+DSGF+SRMQ1+SGF features for WOW and S-UNIWARD are much smaller than that, with decreasing amplitudes reaching 14.45% and 15.60%, respectively.

Above experimental results indicate that, compared to the features extracted from residuals of three channels, the features extracted from the residuals of channel differences exhibit better steganalysis performance.

Table 1: Average test errors of features before and after improving

Algorithm	Payload (bpc)	SRMQ1	DSRMQ1	SGF	DSGF	SRMQ1 +SGF	DSRMQ1 +DSGF	DSRMQ1+DSGF SRMQ1+SGF
WOW	0.05	0.4261	0.3628	0.4921	0.4811	0.4320	0.3577	0.3526
	0.1	0.3754	0.2722	0.4751	0.4401	0.3794	0.2762	0.2634
	0.2	0.3054	0.1835	0.4476	0.3690	0.3117	0.1838	0.1672
	0.3	0.2505	0.1334	0.4326	0.3041	0.2598	0.1316	0.1191
	0.4	0.2155	0.1022	0.4136	0.2506	0.2181	0.1013	0.0877
S-UNIWARD	0.05	0.4404	0.3504	0.4947	0.4730	0.4478	0.3532	0.3496
	0.1	0.3926	0.2718	0.4816	0.4421	0.3971	0.2657	0.2606
	0.2	0.3094	0.1783	0.4583	0.3575	0.3167	0.1716	0.1607
	0.3	0.2613	0.1270	0.4362	0.2967	0.2636	0.1269	0.1104
	0.4	0.2197	0.0974	0.4081	0.2364	0.2202	0.0942	0.0776

5.2 Performance of steganalysis algorithm based on residuals of channel differences

The performance of the steganalysis algorithms in Abdulrahman et al. [Abdulrahman, Chaumont, Montesinos et al. (2016a)] and [Goljan, Fridrich and Cogranne (2014)], and the proposed steganalysis algorithm are tested with above experimental settings in this subsection. The average test errors of the three steganalysis algorithms for WOW and S-UNIWARD with different payloads are presented in Tab. 2. Fig. 3 shows the receiver operating characteristic (ROC) curves of three steganalysis algorithms for WOW with payloads of 0.05, 0.1, and 0.3 bpc, and it is similar to that for S-UNIWARD. It can be seen that the proposed steganalysis algorithm outperforms the steganalysis algorithms in Abdulrahman et al. [Abdulrahman, Chaumont, Montesinos et al. (2016a)] and [Goljan, Fridrich and Cogranne (2014)] under different payloads for WOW. Specifically, when the payload is small, the advantage of the proposed steganalysis algorithm is more significant. The true positive rate of the proposed algorithm is clearly greater than those of the algorithms in Abdulrahman et al. [Abdulrahman, Chaumont, Montesinos et al. (2016a)] and [Goljan, Fridrich and Cogranne (2014)] at different false positive rates, and the maximum decreasing amplitudes of the average test errors for WOW and S-UNIWARD are between 4% and 5%. Even when the payload is greater than or equal to

0.2 bpc, the true positive rate of the steganalysis algorithm proposed in this study is still greater than those of the algorithms in Abdulrahman et al. [Abdulrahman, Chaumont, Montesinos et al. (2016a)] and [Goljan, Fridrich and Cogranne (2014)], and the average test error is decreased by approximately 2%. The outstanding performance of the proposed steganalysis algorithm can be attributed to the addition of the features extracted from the residuals of channel differences. In addition, with increased payload, the steganalysis algorithms in Abdulrahman et al. [Abdulrahman, Chaumont, Montesinos et al. (2016a)] and [Goljan, Fridrich and Cogranne (2014)] achieve significant success rates, resulting in the advantages of the proposed algorithm being less significant than those with low payloads.

Table 2: Average test errors of different steganalysis algorithms.

Algorithm	Payload(bpc)	0.05	0.1	0.2	0.3	0.4
WOW	SCRMQ1(CRMQ1 + SRMQ1) [Goljan, Fridrich and Cogranne (2014)]	0.3743	0.2686	0.1562	0.1010	0.0693
	SCRMQ1+SGF [Abdulrahman, Chaumont, Montesinos et al. (2016a)]	0.3872	0.2759	0.1606	0.1036	0.0701
	SCRMQ1+SGF+**DSRMQ1+DSGF**	**0.3344**	**0.2429**	**0.1441**	**0.0867**	**0.0597**
S-UNIWARD	SCRMQ1(CRMQ1 + SRMQ1) [Goljan, Fridrich and Cogranne (2014)]	0.3698	0.2689	0.1560	0.0955	0.0636
	SCRMQ1+SGF [Abdulrahman, Chaumont, Montesinos et al. (2016a)]	0.3786	0.2666	0.1559	0.0968	0.0642
	SCRMQ1+SGF+**DSRMQ1+DSGF**	**0.3334**	**0.2372**	**0.1344**	**0.0835**	**0.0546**

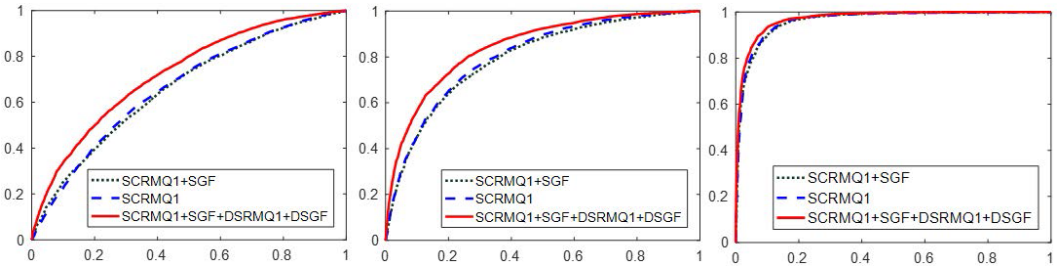

Figure 3: ROC curves of different steganalysis algorithms for WOW (x-label: False positive rate and y-label: True positive rate): (a) payload 0.05 bpc, (b) payload 0.1 bpc, and (c) payload 0.3 bpc

6 Conclusions

In existing color image steganalysis algorithms, some features do not consider the correlation between different color channels, and the performance of them can be further improved. This study points out that the influence of steganography to the distribution of channel difference is greater than that of a single channel, and features extracted from channel differences will detect the color stego image with greater success rates. Based on this, two types of typical steganalysis features of color image, SRMQ1 and SGF, are improved. A color image steganalysis algorithm is proposed by combining the improved features with the existing color image steganalysis features. Experimental results indicate that, for both WOW and S-UNIWARD, the improved

steganalysis features and algorithm significantly decrease the average test errors of the existing steganalysis features and algorithms.

In future studies, we will attempt to improve other high-dimensional steganalysis features, and design more effective steganalysis algorithms for color JPEG images.

Acknowledgement: This work was supported by the National Natural Science Foundation of China (Nos. 61772549, 61872448, U1736214, 61602508, 61601517, U1804263).

References

Abdulrahman, H.; Chaumont, M.; Montesinos, P.; Magnier, B. (2016): Color images steganalysis using RGB channel geometric transformation measures. *Security and Communication Networks*, vol. 9, no. 15, pp. 2945-2956.

Denemark, T.; Sedighi, V.; Holub, V.; Cogranne, R.; Fridrich, J. (2014): Selection-channel-aware rich model for steganalysis of digital images. *Proceedings of IEEE International Workshop on Information Forensics and Security*, pp. 48-53.

Fridrich, J.; Du, R.; Long, M. (2000): Steganalysis of LSB encoding in color images. *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 1279-1282.

Fridrich, J.; Kodovský, J. (2012): Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 868-882.

Goljan, M.; Fridrich, J. (2015): CFA-aware features for steganalysis of color images. *Proceedings of SPIE 9409, Media Watermarking, Security, and Forensics*, pp. 1-13.

Goljan, M.; Fridrich, J.; Cogranne, R. (2014): Rich model for steganalysis of color images. *Proceedings of IEEE International Workshop on Information Forensics and Security*, pp. 185-190.

Holub, V.; Fridrich, J. (2012): Designing steganographic distortion using directional filters. *Proceedings of IEEE International Workshop on Information Forensics and Security*, pp. 234-239.

Holub, V.; Fridrich, J. (2013): Digital image steganography using universal distortion. *Proceedings of the 1st ACM Workshop on Information Hiding and Multimedia Security*, pp. 59-68.

Holub, V.; Fridrich, J. (2013): Random projections of residuals for digital image steganalysis. *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 12, pp. 1996-2006.

Holub, V.; Fridrich, J. (2015): Phase-aware projection model for steganalysis of JPEG images. *Proceedings of SPIE 9409, Media Watermarking, Security, and Forensics XVII*, pp. 1-11.

Kodovský, J.; Fridrich, J.; Holub, V. (2012): Ensemble classifiers for steganalysis of digital media. *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 432-444.

Li, F.; Zhang, X.; Yu, J. (2014): Steganalysis for color JPEG images based on ensemble proportion training. *Journal of Electronics and Information Technology*, vol. 36, no. 1, pp. 114-120.

Liao, X.; Chen, G.; Yin, J. (2016): Content-adaptive steganalysis for color images. *Security and Communication Networks*, vol. 9, no. 18, pp. 5756-5763.

Liu, Q.; Sung, A. H.; Xu, J.; Ribeiro, B. (2006): Image complexity and feature extraction for steganalysis of LSB matching steganography. *Proceedings of the 18th International Conference on Pattern Recognition*, pp. 267-270.

Lyu, S.; Farid, H. (2004): Steganalysis using color wavelet statistics and one-class support vector machines. *Proceedings of SPIE 5306, Security, Steganography, and Watermarking of Multimedia Contents VI*, pp. 35-45.

Ma, Y.; Luo, X.; Li, X.; Bao, Z.; Zhang, Y. (2018): Selection of rich model steganalysis features based on decision rough set α -positive region reduction.

https://ieeexplore.ieee.org/abstract/document/8272003.

Pevný, T.; Bas, P.; Fridrich, J. (2010): Steganalysis by subtractive pixel adjacency matrix. *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 215-224.

Pevný, T.; Filler, T.; Bas, P. (2010): Using high-dimensional image models to perform highly undetectable steganography. *Proceedings of the 12th International Workshop on Information Hiding*, pp. 161-177.

Sallee, P. (2005): Model-based methods for steganography and steganalysis. *International Journal of Image and Graphics*, vol. 5, no. 1, pp. 167-190.

Sangwine, S. J.; Horne, R. E. (1998): *The Colour Image Processing Handbook.* Springer Science & Business Media, vol. 29.

Song, X.; Liu, F.; Yang, C.; Luo, X.; Zhang, Y. (2015): Steganalysis of adaptive JPEG steganography using 2D gabor filters. *Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security*, pp. 15-23.

Su, G.; Han, L.; Huang, F.; Yang, B. (2011): A steganalysis algorithm based on statistic characteristics of the color images. *Proceedings of International Conference on Computer Science and Network Technology*, pp. 2294-2297.

Tang, S.; Chen, Q.; Zhang, W.; Huang, Y. (2015): Universal steganography model for low bit-rate speech codec. *Security and Communication Networks*, vol. 9, no. 8, pp. 747-754.

Westfeld, A. (2001): F5-a steganographic algorithm: high capacity despite better steganalysis. *Proceedings of the 4th International Workshop on Information Hiding*, pp. 289-302.

Xiang, L.; Li, Y.; Hao, W.; Yang, P.; Shen, X. (2018): Reversible natural language watermarking using synonym substitution and arithmetic coding. *Computers, Materials & Continua*, vol. 55, no. 3, pp. 541-549.

Xiang, L.; Zhao, G.; Li, Q.; Hao, W.; Li, F. (2018): TUMK-ELM: a fast unsupervised heterogeneous data learning approach. *IEEE Access*, vol. 6, pp. 35305-35315.

Yang, C.; Liu, F.; Luo, X.; Zeng, Y. (2013): Pixel group trace model-based quantitative steganalysis for multiple least-significant bits steganography. *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 216-228.

Yang, C.; Luo, X.; Lu, J.; Liu, F. (2018): Extracting hidden messages of MLSB steganography based on optimal stego subset. *Science China Information Sciences*, vol. 61, no. 11, pp. 1-3.

Yao, Y.; Zhang, W.; Yu, N. (2016): Inter-frame distortion drift analysis for reversible data hiding in encrypted H.264/AVC video bitstreams. *Signal Processing*, vol. 128, pp. 531-545.

Zhang, Y.; Liu, F.; Yang, C.; Luo, X.; Song, X. et al. (2017): Steganalysis of content-adaptive JPEG steganography based on gauss partial derivative filter bank. *Journal of Electronic Imaging*, vol. 26, no. 1, pp. 1-11.

Zhang, Y.; Qin, C.; Zhang, W.; Liu, F.; Luo, X. (2018): On the fault-tolerant performance for a class of robust image steganography. *Signal Processing*, vol. 146, pp. 99-111.