Steganalysis by Recombining Spatial Rich Model

Ling Gan\textsuperscript{a} and Ning Sun\textsuperscript{b}

School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China

\textsuperscript{a}ganling@cqupt.edu.cn, \textsuperscript{b}S160201003@stu.cqupt.edu.cn

Abstract. Steganalysis is an important topic in the field of information security. Steganalysis is mainly based on the traditional machine learning method, represented by the spatial rich model, and has high detection performance. Aiming at the problems of high feature dimension, strong redundancy and long feature extraction time in the spatial rich model, this paper conducts a series of feature selection operations and explores four different model combination methods: single quantization factor combination method, single filter kernel type combination method, mixed selecting combination method, Pearson correlation coefficient combination method. In addition, a feature preprocessing means are added. The experiment proves that the combination of mixed selecting combination and Pearson correlation coefficient can be stable with the performance of steganalysis before dimension reduction, but the performance dimension ratio is higher.

1. Introduction
Steganalysis is the opposite direction of steganography, which can be seen as an attack on steganography, similar to the "spear and shield" in the fable story. The means of steganalysis are generally divided into special steganalysis and general steganalysis. Special steganalysis is mainly based on the principle and characteristics of a known steganography, such as $\chi^2$ methods [1], RS [2], SPA [3], WS [4]. They have higher detection accuracy. But the scope of use is too narrow and the application is limited. Due to the recent rapid development of steganography, the special steganalysis technology can not meet the requirements of steganalysis, making general steganalysis the focus of research in the field of steganalysis.

The main technique of general steganalysis is based on machine learning method, which mainly includes two steps: (1) Extracting the feature that can reflect the change of statistical characteristics of the cover before and after embedding information. (2) Training the classifier to complete classification.

In terms of feature extraction, the early statistical features are relatively simple, and the feature dimensions are relatively low. For example, the image histogram-based changes proposed by Westfall[1]. Harmsen perform low-pass filtering on histograms to smooth histograms [5]. Li used the histogram feature function ratio of the difference image to design the steganalysis feature on the basis of the predecessors [6]. Sullivan first regarded the change of adjacent pixels as a Markov random process, and the Markov transition probability matrix of neighboring pixels as steganalysis features [7]. In 2010, Pevný designed SPAM features using pixel-difference first-order and second-order Markov transition probability matrices [8]. The above method achieved great success at the time, but with the development of steganography, the steganography can maintain many complex features. The above
feature dimensions are relatively low, and cannot completely capture the correlation change between pixels. From the empirical point of view, it is obvious that the more features, the more reflective the image content, the more accurate the results of the steganalysis. So the scholars continue to study the features that better reflect the image content. In 2012, Fridrich carefully designed a spatial rich model SRM [9], which is a multi-model combination feature, which uses linear and nonlinear calculations to quantify the noise of images, and uses the high-order co-occurrence matrix as the feature to construct 106 sub-models, which are combined with each sub-model and have dimensions up to 34671 dimensions. On the basis of the spatial rich model, various variant versions have appeared, such as PSRM [10], tSRM [11], maxSRM [12], $\sigma$ SRM [13] and other steganalysis methods.

In terms of classifier, with the increase of feature dimension, FLD and SVM are limited by the computational complexity, which can not meet the requirements of steganalysis. It is necessary to design a classifier that matches high-dimensional features. Fridrich designed a special classifier of steganalysis, which uses FLD as a sub-classifier, and multiple sub-classifiers judge the random sampling of high-dimensional features, and finally vote the fusion of the classification results of all sub-classifiers into the final discriminant result. The ensemble classifier quickly complete classification under the premise of ensuring classification quality [14].

This paper focuse on the feature extraction in the spatial rich model steganalysis. Based on the steganalysis of spatial rich model, four different model combinations are explored, and feature preprocessing methods are added to reduce the feature dimension of SRM, save the time of feature extraction, and improve the performance dimension ratio of steganalysis.

2. Spatial rich model steganalysis method
The SRM steganalysis [9] is similar to the common pattern recognition, which adopts the structure of “feature + classifier”, but there are also differences: the training set and test set required for the study are not completely collected. The stego images are obtained by using someone steganography to embed secret information into cover images.

2.1. Obtain residual image and co-occurrence matrix
To reflect changes in pixel correlation, a variety of linear and nonlinear filter kernels are used to generate a rich variety of residual images. The filter kernel can be roughly divided into six classes: 1st, 2nd, 3rd, EDGE3$\times$3, EDGE5$\times$5 and SQUARE. The name of the filter kernel consists of four parts: filter mode, number of filters, symmetry factor, and symbiotic matrix scan direction. There are two filtering modes: the filter residual value is determined by a filter called “spam”, otherwise it is called “minmax”. The residual value of the "minmax" type is obtained by taking the maximum and minimum values of multiple filters. The number of filters is that the filter kernel consists of several filters. The symmetry factor is that the filter kernel can have several different forms by rotation or mirror symmetry. There are three kinds of symbiotic matrix scanning directions: “h” stands for horizontal scanning, “v” stands for vertical scanning, if there is no “h” or “v”, it means horizontal and vertical symmetry.

Table 1 shows the specific classification of the six classes of filter kernels in the spatial rich model. The residual value obtained by filtering has a large amplitude range. The larger the residual value of the image element, the smaller the correlation between the pixel and the surrounding pixels, and the smaller the effect of the embedding. Since the co-occurrence matrix generated by the residual image will be characterized as a feature, the dimension of the co-occurrence matrix cannot be too high, and the range of the residual value needs to be controlled, so quantization and truncation operations are taken:
Table 1. Six kinds of filter kernels for spatial rich model.

| 1st       | 2nd       | 3rd       | EDGE $3 \times 3$ | EDGE $5 \times 5$ | SQUARE |
|-----------|-----------|-----------|-------------------|-------------------|--------|
| minmax22h | minmax22h | minmax22h | minmax22h         | minmax22h         | spam11 |
| minmax24  | minmax24v | minmax24v | minmax24v         | minmax24v         | minmax22h |
| minmax24  | minmax24v | minmax24  | minmax24          | minmax24          | minmax24 |
| minmax34  | minmax34  | minmax34  | minmax34          | minmax34          | minmax41 |
| minmax34h | minmax34h | minmax34h | spam14h, v        | spam14h, v        | minmax34v |
| minmax41  | —         | minmax41  | —                 | —                 | —      |
| minmax48h | —         | minmax48h | —                 | —                 | —      |
| minmax48v | —         | minmax48v | —                 | —                 | —      |
| minmax54  | —         | minmax54  | —                 | —                 | —      |
| spam14h, v| —         | spam14h, v| —                 | —                 | —      |

$$R_{i,j} \leftarrow \text{trunc}_c \left( \text{round} \left( \frac{R_{i,j}}{q} \right) \right)$$

$$\text{trunc}_c(x) = \begin{cases} 
T & x > T \\
T - x & T \geq x \geq -T \\
-x & x < -T 
\end{cases}$$

$$q \in \{c, 1.5c, 2c \} \quad c > 1$$

In summary, quantization factor $q$ and truncation threshold $T$ are very important for steganalysis. If the truncation threshold is too high, it can introduce a large number of statistical features that are not statistically significant. Because the changes to pixel values usually does not exceed 2, so the threshold $T = 2$. In addition to the above two parameters, the order $d$ of the co-occurrence matrix is also an important parameter. The order of the co-occurrence matrix is too large, which makes the feature too sparse. The order of the co-occurrence matrix is too small, which makes the diversity of features not reflected, so $d = 4$ is suitable.

2.2. Ensemble classifier design

Steganalysis is essentially a two-category problem. The use of classifiers to detect images is an important step in general-purpose image steganalysis. J Fridrich proposed an ensemble classifier for steganalysis that integrates multiple Fisher Linear Discriminators (FLDs). In order to improve the discriminating accuracy of the classifier, it is necessary to increase the diversity of the FLD. Therefore, each FLD randomly samples the samples in the training set as training samples, and randomly extracts the feature subsets of the training sample images for training. In the classification stage, after all the basic classifiers respectively predict the results, the results are subjected to unweighted voting fusion. The number of votes is the final prediction result, and the number of votes is the same, and a result is randomly predicted. As shown in Fig. 1.
3. Recombine spatial rich model

3.1. Feature similarity measure
The Pearson correlation coefficient $\rho_{xy}$ is a more complex method of judging similarity than the Euclidean distance. It is the most commonly used linear correlation coefficient to reflect the degree of linear correlation between two variables. The correlation coefficient ranges from -1 to 1. The larger the absolute value, the stronger the correlation.

$$
\rho_{xy} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mathbb{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}
$$

(4)

3.2. Feature preprocessing
The spatial rich model steganalysis is based on the statistical generation of co-occurrence matrices in the residual image as features, and the extracted features are up to 34671 dimensions. In this paper, all the image extraction features of the BOSSbase 1.01 database are obtained as a feature matrix of 10000×34671, as shown in Fig. 2.

An in-depth observation found that all images have a value of zero in some dimensions. There is no contribution to such a full 0 dimension in the subsequent classifier training and testing process. To solve this problem, this paper proposes to add the feature preprocessing process after feature extraction, delete the all 0 dimension, and further improve the performance dimension ratio of the steganalysis.

3.3. Different model combinations
There are some very similar submodels in SRM, so it is not necessary to use all the submodels, and 106 submodels can be combined in many ways. From the way of generating the symbiotic matrix feature in Section II, it has a total of 78 filter kernels. The statistical residual image is quantized by three quantization factors before generating the co-occurrence matrix. It can be combined according to the above classification.
1, 1, L, 1, corresponding to the six classes of filter kernels. There are 39 basic filtering cores, and the basic filter kernels corresponding to each class are 11, 6, 11, 5, 5, and 1, respectively. Defined as the i\textsuperscript{th} submodel under the influence of the quantization factor $q$, and the detection error $E_i$ of the i\textsuperscript{th} submodel quantized is used.

(1) Single combination of quantization factors. We can obtain 3 combined model $U_{11}, M_i^{q1}$, $U_{11.5}, M_i^{q1.5}$, $U_{2}, M_i^{q2}$ by fixing quantization factor $q = 1c, 1.5c, 2c$. The specific experimental comparison can be found in the following experimental part.

(2) Single filter core type combination. We can select all the filter kernels in a certain class to combine a new model, and generating a feature matrix with multiple quantization factors, $U_{i}, k = 1,2,L,6$.

(3) Mixed selecting combination. Compared with the submodel differences generated by different filter kernels, the submodel differences generated by the same filter kernel using different quantization factors are small. Therefore, for the basic filter kernel, only the submodel $M_i^q$ generated by the quantization factor that brings the smallest detection error $E_i^q$ can be selected, that is $q = \arg\min_q E_i^q$. Finally, 39 sub-models with the same number of basic filter kernels will be obtained, $U_{11}, M_i^q$.

(4) Pearson correlation coefficient combination. It is a heuristic combination. We can obtain a matrix of $106 \times 106$, which reflects the correlation coefficients of 106 submodels with each other. For each submodel $M_i^q$, find another submodel $M_j^q$ that is most relevant to it. If the correlation coefficient of the two submodels is greater than a threshold $T$, delete the submodel $M_j^q$. The algorithm flow is shown in Fig. 3.

![Figure 3. Feature selection flow chart.](image-url)
The above algorithm flow mainly accomplishes the task of removing similar features, only one of the similar features remains. The combined model after selecting, used in this paper, a total of 27 submodels, a total of 8892 dimensions.

4. Experiments and comparison to prior art

4.1. Setups
The experimental software and hardware configuration is as follows: CPU: Intel (R) core i5-4590 3.3GHz, memory: 8GB, Matlab R2016b. All experiments are conducted on BossBase1.01. WOW and HILL are selected to embed the secret information of 0.1bpp and 0.4bpp, and generate corresponding stego images databases. This paper measures detection performance by detecting the results of WOW and HILL. In the process of steganalysis, 5000 cover images and corresponding stego images are randomly selected to form a training set, and the remaining images are used as test sets.

4.2. Comparison to Prior Methods
For the combination model 1 and model 2 mentioned in Section Ⅲ, the experimental results are shown in Fig. 4.

It can be seen from Fig. 4 that the quantization factor has a certain impact on the performance of the steganalysis. For the WOW, whether the large or small capacity is hidden, the combined model with large quantization factor has low detection error. While for the HILL, when the small capacity is hidden, the size of the quantization factor does not play a key role. In the case of capacity, the combined model with a large value of the quantization factor has a low detection error.

As can be seen from Fig. 4, most of the high-order combination models are better than the first-order model steganalysis because the higher-order models take into account the correlation of more neighborhood pixels. The steganalysis performance of a single filter kernel has a large gap with the steganalysis performance of multiple filter kernel combinations of SRM, indicating that steganalysis requires the use of multiple filter kernels, increasing the diversity of filter kernels and facilitating detection.

It can be seen from the experiment of a single quantization factor and a single filtering kernel type that the size of the quantization factor and the diversity of the filtering kernel have a great influence on the performance of the steganalysis. Therefore, this paper proposes the third model combination method, mixed selecting combination method, the experiment is as follows:

The HILL hide the secret information of large capacity 0.4bpp, and use a single submodel to perform steganalysis one by one. All the results are shown in Fig. 5. The linear and nonlinear filtering models are separated by dashed lines, and all solid points are connected by straight lines to facilitate observation and interpretation of the results. Lines of different colors represent different quantization factors. Take the first red dot in the Fig. 5 as an example: It shows that after using the 22h filter kernel...
nonlinear filter to obtain the residual image, using ensemble classifier with the 4th order co-occurrence matrix to obtain detection error values.

Figure 5. Detection performance of 106 submodels.

It can be observed from Fig. 5 that the performance of the 39 basic models will be different under different quantization factors. Not all filter kernels have the best performance under the same quantization factor. Therefore, for each filter kernel, try to work with the best performance quantization factor and combine to generate a submodel in the combined model.

The mixed selecting combination model has only 39 submodels, a total of 12,753 dimensions, and its calculation time is short, only one-third of the original time. The combination model steganalysis method is used in this paper to compare the detection results with the SRM method as shown in Fig. 6 (left).

Figure 6. Combined model and SRM detection results.

As can be seen from Fig. 6 (left), compared with the original SRM method, although the feature dimension is reduced, the steganographic performance is not affected. The steganographic performance of the hybrid combination mode is close to that of SRM, and the error between them is negligible. To a degree, the performance of the steganalysis has reached saturation.

For the combination model 4 mentioned in Section III. The experimental results are shown in Fig. 6 (right). It can be seen from Fig. 6 that the Pearson correlation coefficient selecting combination...
method is still unaffected compared with the original SRM method, further illustrating the redundancy of the spatial rich model feature and the performance of its steganalysis has reached saturation status.

5. Conclusion
In this paper, we found that SRM feature dimension is too high, feature extraction time is long, and features are redundant. So we carried out a series of feature selection operations based on SRM, and explored the following four different model combinations: single quantization factor combination method, single filter kernel type combination method, mixed selecting combination method, and Pearson correlation coefficient combination method. In addition, feature preprocessing is added. Experiments show that the combination of mixed selecting method and Pearson correlation coefficient can reduce the dimension of the feature, and the performance of the steganalysis before reducing the dimension remains stable, with a higher performance dimension ratio.

References
[1] A Westfeld, and P Andreas. "Attacks on steganographic systems". International workshop on information hiding. Springer, Berlin, Heidelberg, 1999.
[2] J Fridrich, G Miroslav, and D Rui. "Detecting LSB steganography in color, and gray-scale images". IEEE multimedia, vol. 8, pp. 22-28, April 2001.
[3] S Dumitrescu, X Wu, and Z Wang. "Detection of LSB steganography via sample pair analysis". International Workshop on Information Hiding. Springer, Berlin, Heidelberg, 2002.
[4] J Fridrich, and G Miroslav. "On estimation of secret message length in LSB steganography in spatial domain". Security, steganography, and watermarking of multimedia contents VI. Vol. 5306. International Society for Optics and Photonics, 2004.
[5] J J Harmsen, and A P William. "Steganalysis of additive-noise modelable information hiding". Security and Watermarking of Multimedia Contents V. Vol. 5020. International Society for Optics and Photonics, 2003.
[6] X Li, T Zeng, and B Yang. "A further study on steganalysis of LSB matching by calibration". 2008 15th IEEE International Conference on Image Processing. IEEE, 2008.
[7] K Sullivan, U Madhow, S Chandrasekaran, and B.S. Manjunath. "Steganalysis for Markov cover data with applications to images". IEEE Transactions on Information Forensics and Security, vol. 1, pp. 275-287, February, 2006.
[8] T Pevny, B Patrick, and J Fridrich. "Steganalysis by subtractive pixel adjacency matrix". IEEE Transactions on Information Forensics and Security, vol. 5, pp. 215-224, February, 2010.
[9] J Fridrich, and J Kodovsky. "Rich models for steganalysis of digital images". IEEE Transactions on Information Forensics and Security, vol. 7, pp. 868-882, March, 2012.
[10] V Holub, and J Fridrich. "Random projections of residuals for digital image steganalysis". IEEE Transactions on Information Forensics and Security, vol. 8, pp. 1996-2006, December, 2013.
[11] W Tang, H Li, W Luo and J Huang. "Adaptive steganalysis against WOW embedding algorithm". Proceedings of the 2nd ACM workshop on Information hiding and multimedia security. ACM, 2014.
[12] T Denemark, V Sedighi, V Holub, R Cogranne and J Fridrich "Selection-channel-aware rich model for steganalysis of digital images". 2014 IEEE International Workshop on Information Forensics and Security (WIFS). IEEE, 2014.
[13] T Denemark, J Fridrich, and C A Pedro. "Improving selection-channel-aware steganalysis features". Electronic Imaging, vol. 2016, pp. 1-8, August, 2016.
[14] J Kodovsky, J Fridrich, and V Holub. "Ensemble classifiers for steganalysis of digital media". IEEE Transactions on Information Forensics and Security, vol. 7, pp. 432-444, February, 2012.