

An Image Adaptive Steganography Algorithm Based on Sparse Representation and Entropy

Chunmei Yu^{1,a}, Jianjun Wang^{2,b}

¹ Department of Electronic Engineering, Fudan University, Shanghai 200433, China

² Department of Electronic Engineering, Fudan University, Shanghai 200433, China

^a11210720034@fudan.edu.cn, ^bwangjj@fudan.edu.cn

Keywords: sparse representation; entropy; image adaptive steganography; steganalysis

Abstract: According to the image processing theory, complex texture image regions have larger entropy values, an adaptive steganography algorithm in the sparse domain is proposed here. The algorithm first divides an image into nonoverlapping image blocks, then computes the entropy of each block and chooses the mean of all entropies as a threshold. Image blocks whose entropy is higher than the threshold are selected for sparse decomposition as complex texture image regions. The secret message is embedded into the decomposition coefficients, and then the stego image is reconstructed with modified coefficients. The experimental results prove that, given the same payload, the proposed adaptive algorithm has a better performance against the steganalyzer.

Introduction

Steganography is a technology that tries to embed secret messages to general objects (images, audio, etc.) so as not to attract outside attention. As a main adversary, steganalysis aims to detect the presence of hidden messages. Existing steganalysis algorithms often rely on the statistical characteristics of image signals, but ignore the change on the semantic level. This inspired researchers to embed secret information in the semantic content of a carrier signal to improve the security. Image sparse representation is capable of representing the image in a sparse and efficient way, so it can be applied to image steganography.

Currently only a few papers preliminarily explored the application of image sparse representation to image steganography. Cancelli embedded secret information by modifying the decomposition coefficients [2,3,4]. It is demonstrated that the same embedding algorithm applied in the sparse domain is more secure than it is in the pixel domain. However these papers just applied sparse representation to image steganography, did not take the influence of other factors, such as images content, into account. And studies have found that if better steganalysis tools are used, such as the SPAM (subtractive pixel adjacency matrix) steganalyzer [5], the possibility of detection will increase greatly.

Considering these problems, an image adaptive steganography algorithm based on sparse representation and entropy is proposed. The rest of the paper is organized as follows. In Section II, a brief introduction of steganography using image sparse representation is given. The proposed algorithm is presented in Section III. Section IV reports the results of the experiments that we carried out to validate the proposed technique. Conclusions are given in Section V.

Related Work

In the sparse representation theory, each signal is represented with a linear combination of few basic elements. The elements $g_i \in R^N$ for the decomposition of signals are called atoms, the set of atoms $D = \{g_i\}_{i=1,\dots,K}$ is known as a dictionary. If $K > N$, atoms are linear correlated, the dictionary D is redundant.

There exists two critical issues to apply sparse representation to image steganography: the

construction of the dictionary D for image sparse representation and how to get the efficient and unique decomposition coefficients of an image given the dictionary D . The following provides solutions for these two problems.

Construction of The Redundant Dictionary. The goal is to embed secret messages at a semantic level, so the required redundant dictionary should be as semantically meaningful as possible. Generally, the dictionary can be constructed through the following three methods: 1) generated manually [4]; 2) generated directly using existing generating functions, like DCT transform, Haar transform [6], and Curvelet transform [7,8]; 3) generated by learning, the dictionary is learned given training samples.

Here the K-SVD (K-means Singular Value Decomposition) algorithm is used to obtain the redundant dictionary. The objective function can be expressed as:

$$\min_{D, X} \{\|Y - DX\|_F^2\} \quad \text{s.t. } \forall i, \|x_i\|_0 \leq T_0, \quad (1)$$

Where $D \in \mathbb{R}^{N \times K}$ is the dictionary, $Y = \{y_i\}_{i=1}^n$ is the set of n training signals, $X = \{x_i\}_{i=1}^n$ is the coefficient matrix for the sparse representation of signals. T_0 is the maximum number of non-zero elements for sparse decomposition coefficients.

The solution of equation (1) is an iterative process. First, we find the coefficient matrix X for signals Y assuming D is fixed. Then, the dictionary is updated according to the coefficients. The update of the dictionary is performed by column. Take the update of the k th atom and its corresponding coefficient x_T^k (the k th row of X) for example:

1) Find the corresponding indices of all the sample signals that use the atom g_k in $\{y_i\}$ and form a set, defined as $w_k = \{i \mid 1 \leq i \leq n, x_T^k(i) \neq 0\}$.

2) Compute the overall representation error matrix, E_k , by

$$E_k = Y - \sum_{j \neq k} g_j x_T^j. \quad (2)$$

3) Define Ω_k as a matrix of size $N \times |w_k|$, with ones on the $(w(i), i)$ th entries and zeros elsewhere. Use $E_k^R = E_k \Omega_k$ to find error columns that correspond to samples that use the atom g_k .

4) Apply SVD decomposition $E_k^R = U \Delta V^T$. The updated dictionary column is

$$g_k = U(:, 1), \quad (3)$$

the coefficient vector $x_R^k = x_T^k \Omega_k$ is

$$x_R^k = V(:, 1) \Delta(1, 1) \quad (4)$$

Repeat until convergence (stopping rule). Note that, the atoms of the learned dictionary is real-valued. This will result in a quantization error when the stego image is reconstructed, so the atoms should be further binarized with only the 0 and 1 values are allowed.

Stable Sparse Decomposition Algorithm for Image Stegonagraphy. The commonly used MP (matching pursuit) algorithm [7,8] is a greedy algorithm that selects the dictionary atoms sequentially. It can be described as two steps:

$$[c_k, g_{\gamma_k}] = S(R^{k-1}, D), \quad (5)$$

$$R^k = U(R^{k-1}, c_k, g_{\gamma_k}), \quad (6)$$

where $S(\cdot)$ is a selection function and $U(\cdot)$ is an update function. R^k is the residual signal, c_k is the specific weight of the k th atom, γ_k is the index of the k th atom of the decomposition.

However it's not suitable to directly apply the MP algorithm to image steganography, because the insertion of the message may cause the instability of decomposition path and decomposition coefficients. To solve the problem, Cancelli defined the selection rule as follows:

$$c_k^* = \max\{c \geq 0 : R^k = R^{k-1} - c g_{\gamma_k} \geq 0\}, \quad (7)$$

$$R_{\gamma_k}^k = R^{k-1} - c_k^* g_{\gamma_k}, \quad (8)$$

$$\gamma_k^* = \arg \min_{\gamma_k \in \{1, 2, \dots, D\}} \sum_{i,j} \|R_{\gamma_k}^k(i, j)\|^2, \quad (9)$$

The equation (7) guarantees that every pixel value of the residual image is non-negative. Using the selection rule described above, the instability of decomposition coefficients can be avoided. In other words, given the same decomposition path, the receiver can obtain the same modified version of the coefficients, the algorithm is stable.

Proposed Image Adaptive Steganography Algorithm

As mentioned previously, the principle of steganography in the sparse field is to embed secret information in the semantic content of the image. From the aspect of image composition, complex texture regions are often contain many different structural elements, while the constitution of smooth areas are more single. Besides, in the background of fixed luminance, the human visual system is more sensitive to the grayscale change in smooth regions and rarely notices the change in complex texture regions. These two points inspire us to select the complex texture regions of an image and embed information in the decomposition coefficients of these blocks only.

The three principal approaches used in image processing to describe the texture of a region are statistical, structural, and spectral. In statistical approaches, entropy is an efficient measure, defined as

$$E_k = - \sum_{i=0}^{255} P_i \cdot \log_2 P_i. \quad (10)$$

where P_i is the probability of grayscale value i . Entropy is a measure of variability and is 0 for a constant image. As its magnitude increases, the image texture is often more complex. Therefore, we can calculate the entropy to select complex texture image blocks.

Secret Information Embedding Process. The detail is presented in Figure 3-1.

- 1) Divide the cover image into $w \times w$ nonoverlapping blocks, calculate entropy value E_k of each block.
- 2) Set the average block entropy value \bar{E} as threshold ET . The blocks whose entropy value satisfies $E_k > ET$ are classified as complex texture regions.
- 3) Decompose the selected complex texture image blocks with the binarized dictionary complete and non-negative constrained matching pursuit algorithm, get the decomposition path and decomposition coefficient matrix.
- 4) Use the embedding algorithm to hide secret information in the non-zero decomposition coefficients. Besides the first decomposition coefficient usually shares most of the image energy, so it should be kept unchanged during embedding.
- 5) Reconstruct the stego image with the modified decomposition coefficients and the redundant dictionary.

Secret Information Extraction Process. For grayscale images, in addition to the redundant dictionary, the cover image is also required to extract the secret information.

- 1) Divide the cover image into $w \times w$ blocks, select the complex texture blocks and get the decomposition path of these blocks.
- 2) Decompose the stego blocks which corresponds to the complex texture cover blocks with the decomposition paths computed in step 1) to obtain the decomposition coefficients.
- 3) Use the corresponding extraction algorithm to obtain the secret information embedded in the decomposition coefficients.

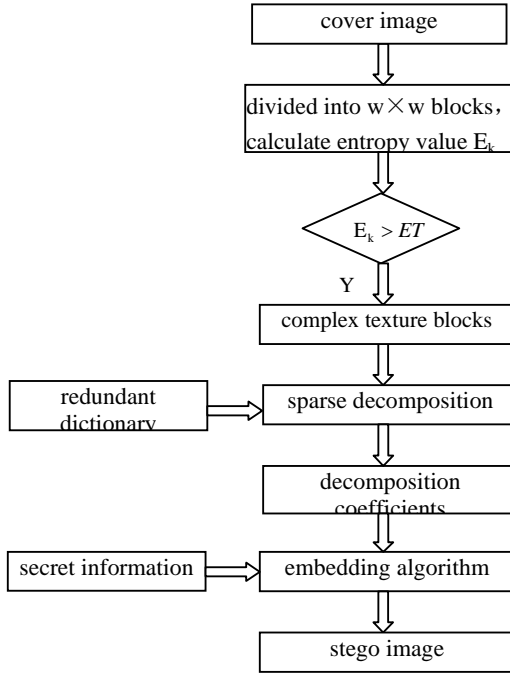


Figure 3-1. Embedding process

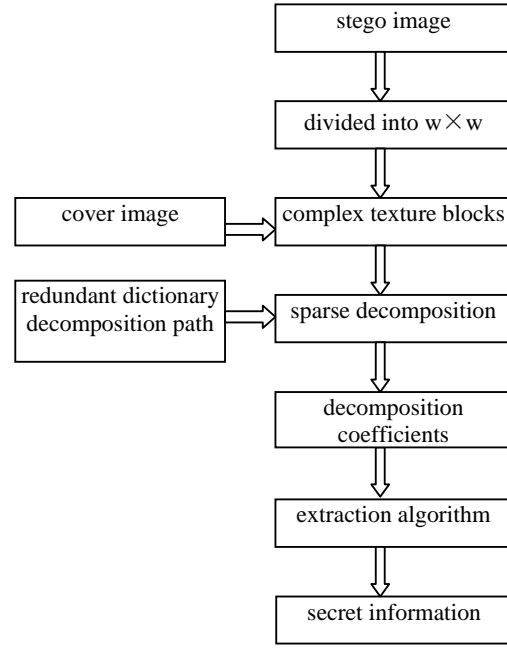


Figure 3-2. Extraction process

Experimental Results and Analysis

In this section, experimental results are reported to evaluate the security of the adaptive steganography.

Experimental Set. UCID (Uncompressed Color Image Database) 0 contains 1338 uncompressed color images of 512×384 size. The images show a variety of scenarios, such as natural scenery, indoor and outdoor man-made objects. In the experiments, all the images are all converted to 8-bit grayscale ones and then cropped into 384×384 size.

The image is divided into 4×4 nonoverlapping blocks. The training signals are taken from standard images and the initial dictionary is randomly generated. After training, we binarize the atoms using OTSU method and remove the identical atoms. At last, a binary dictionary containing 305 different atoms is constructed.

The Embedding Capacity. For image steganography, the embedding capacity is expressed in bit per pixel (bpp). Given the above redundant dictionary, the average maximum payload allowed by the MPSteg algorithm0 is 0.7437bpp, while the adaptive algorithm is 0.5106bpp. The capacity declines slightly because only complex texture image blocks are selected for embedding in the adaptive algorithm.

Security Analysis. Existing steganalyzers can be classified into two categories: targeted and blind. Targeted steganalyzers normally consider special statistical features that are related to the data embedding procedure, while blind steganalyzers generally extract features without a priori knowledge about the embedding procedure.

The SPAM method 0 is an effective blind steganalyzer, it has a good detection performance against the steganography in the spatial domain and the transform domain. The algorithm use first-order and second-order Markov chains to model the differences between adjacent pixels, and then subsets of sample transition probability matrices are used as features for detection. In our experiments, the second-order SPAM features with 686 dimensions are used.

The cover and stego images are used to build a training and a test set, both containing

50% cover and 50% stego images. The training set is built randomly with 20% of the images, and the remaining 80% forms the test set. The Fisher linear discriminant(FLD) is used for classification. To avoid any dependence of the results upon the specific training and test sets, the experiments are repeated 20 times, each time with a different training and test set. The final receiver operating characteristic(ROC) curve is produced by vertically averaging the obtained 20 curves.

The horizontal coordination of the ROC curve indicates the false alarm rate, while the vertical one indicates the detection rate. If the ROC curve is far above the diagonal line, it is generally considered that the classifier performs better. As an overall measure of classification accuracy, the area under curve(AUC) value is also given in the figures. In the experiments, stego images are created with 0.3bpp and 0.5bpp payloads.

We can apply any method that is usually applied in the pixel domain to embed a message. Here the LSB matching algorithm is first adopted. Comparison is carried out between the following method—the conventional ± 1 algorithm, ± 1 applied in the MP domain(MPSteg), ± 1 applied in the adaptive MP domain(the proposed method).

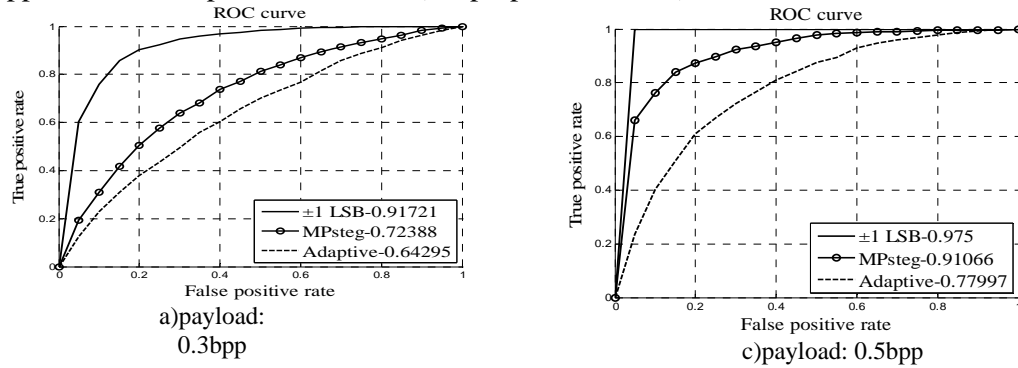


Figure 4-2 Performance comparison: MPSteg vs. Adaptive

The results shown in figure 4-2 indicate that the adaptive method is much less detectable, then the MPSteg method, the conventional ± 1 algorithm has the worst performance. The gap is more explicit with increased payloads.

Second the section-wise EMD algorithm is adopted. Comparison between the MPSteg method and the proposed adaptive method is presented in figure 4-3.

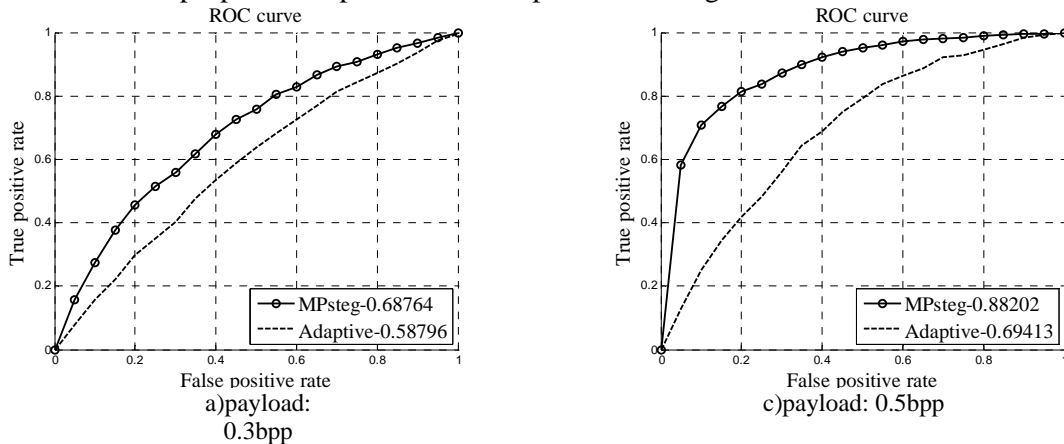


Figure 4-3 Performance comparison: MPSteg vs. Adaptive

We can see that the proposed adaptive algorithm is still less detectable. Besides, the overall performance is improved with the coding algorithm.

Summary

An image adaptive algorithm based on sparse representation and entropy is proposed. Compared to previous work, the new scheme uses the entropy to select complex texture regions, decomposes these regions and then embeds secret messages into the decomposition coefficients. Experimental results show that, though the embedding capacity has a slight decline, the proposed new scheme is more secure.

References

- [1] C. Abbas, J. Condell, C. Kevin, M. K. Paul. Digital image steganography: survey and analysis of current methods [J]. *Signal Processing*, 2010, 90(3): 727-752.
- [2] G. Cancelli, M. Barni, G. Menegaz. MPSteg: Hiding a message in the matching pursuit domain [C]. *Proc. SPIE*, 2006, 6072: 60720P_1-60720P_9.
- [3] G. Cancelli, M. Barni. MPSteg-color: A new steganographic technique for color images [C]. *Information Hiding: 9th Int. Workshop (IH 2007)*, Saint Malo, France, 2007, 4567: 1–15.
- [4] G. Cancelli, M. Barni. MPSteg-Color: data hiding through redundant basis decomposition [J]. *IEEE Trans. Information Forensics and Security*, 2009, 4(3): 346-358.
- [5] Tomáš Pevný, Patrick Bas, J. Fridrich. Steganalysis by subtractive pixel adjacency matrix [J]. *IEEE Trans. Information Forensics and Security*, 2010, 5(2): 215-224.
- [6] M. Aharon, M. Elad, A. M. Bruckstein. The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation [J]. *IEEE Trans. Signal Process.*, 2006, 54(11): 4311–4322.
- [7] S. Mallat, Z. Zhang. Matching pursuits with time frequency dictionaries [J]. *IEEE Trans. Signal Process.*, 1993, 41(12): 3397–3415.
- [8] S. Mallat, G. Davis, Z. Zhang. Adaptive time-frequency decompositions [J]. *Opt. Eng.*, 1994, vol. 33(7): 2183—2191.
- [9] G. Schaefer, M. Stich. UCID-An uncompressed colour image database [C]. *Proc. SPIE, Storage and Retrieval Methods and Applications for Multimedia*, San Jose, USA, 2004: 472–480. <http://www-staff.lboro.ac.uk/~cogs/datasets/UCID/ucid.html>.
- [10] T. Fawcett. ROC graphs: Notes and practical considerations for researchers [J]. *Machine Learning*, 2004, 31.
- [11] J. Wang, Y. Sun, H. Xu, K. Chen, H. Kim, S. Joo. An improved section-wise exploiting modification direction method [J]. *Signal Processing*, 2010, 90(11): 2954–2964.

Acknowledgement

The work on this paper was supported by Nature Science Foundation of China (Grant No. 61170207).