Active Steganalysis of Transform Domain Steganography Based on Sparse Component Analysis

Hamed Modaghegh *
Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
hamed.modaghegh@stu-mail.um.ac.ir
Seyed Alireza Seyedin
Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
seyedin@um.ac.ir

Received: 20/Aug/2014 Revised: 19/Jan/2015 Accepted: 27/Jan/2015

Abstract

This paper presents a new *active* steganalysis method to break the *transform* domain steganography. Most of steganalysis techniques focus on detecting the presence or absence of a secret message in a *cover* (*passive* steganalysis), but in some cases we need to extract or estimate a hidden message (*active* steganalysis). Despite the importance of estimating the message, little research has been conducted in this area. In this study, a new *active* steganalysis method based on Spars Component Analysis (SCA) technique is presented. Here, the sparsity property of the *cover* image and hidden message has been used to extract the hidden message from a stego image. In our method, the *transform* domain steganography is formulated mathematically as a linear combination of sparse sources. Thus, the *active* steganalysis can be presented as a SCA problem. The feasibility of the SCA problem solving is confirmed by Linear Programming methods. Then, a fast algorithm is proposed to decrease the computational cost of steganalysis and still maintains the accuracy. The accuracy of the proposed method has been confirmed in different experiments on a variety of *transform* domain steganography methods. According to these experiments, our method not only reduces the error rate, but also decreases the computational cost compared to the previous *active* steganalysis methods in the literature.

Keywords: Sparse Component Analysis (SCA); Active Steganalysis; Blind Source Separation (BSS); Transform Domain steganography.

1. Introduction

After the seminal study of Johnson and Jajodia [1], [2], steganalysis has attracted growing attention [3]–[7]. Different types of steganalysis techniques (STs), mostly *passive*, have been proposed [4]–[7]. While Steganography deals with hiding information by embedding a message in another object (*cover*) such as an image, steganalysis focuses on revealing those hidden messages from the *cover*. Steganalysis has gained prominence in the international security since the detection of hidden messages can lead to the prevention of catastrophic events, such as terrorist attacks.

Current STs focus on detecting the presence of a hidden message in the *cover* (*passive* manner). An in-depth review of *passive* STs has been presented by Nissar et al. [5]. They attempted to classify various approaches. They have categorized STs into signature and statistical techniques. Their categorization is either based on the signature of the applied technique or the image statistics which is used to detect the presence of hidden messages. Furthermore, in their classification, each category is sub-divided into specific and universal approaches.

Specific steganalysis targets a particular steganographic technique [4], [7], [8]. These methods analyze the embedding operation and concentrate on some image features or statistics. As a result, it may fail if any other steganography method is used or simply a change occurs in the steganography algorithm. Consequently, universal STs [9]–[11], were introduced to overcome the deficiency of specific STs. These methods could detect embedded messages using any type of steganographic technique even in the absence of prior knowledge of embedding technique. Most of them train a classifier with *cover* and stego images in the detection procedure.

Following the detection procedure, sometimes it is necessary to extract and determine the content of the hidden message (*active* steganalysis). In fact, by revealing the hidden messages, *active* steganalysis complements the *passive* one. Most STs deal with *passive* techniques and little attention has been paid to *active* methods [12]–[16]. In this scope, some researchers focus on *active* STs based on the blind sources separation (BSS) [3], [15]–[19]. This study focuses on this class of *active* STs and discusses advantages and disadvantages of these methods.

It is worth mentioning that most BSS-based *active* STs take advantage of the independency property of the image without using sparsity property of the hidden message to achieve better results. Moreover, all *active* STs, which use only one stego image to extract the hidden message, [15]–[17], [19] increase the computational cost. They need at least two observed signals and use a denoising algorithm to generate them. This algorithm usually

* Corresponding Author

increases the computational complexity and making ST as time-consuming algorithm.

In this paper, a new *active* ST with only one stego image is introduced which does not require a denoising algorithm. Eliminating the denoising algorithm makes this ST more efficient than the previous BSS-based STs. Our method uses sparsity property of sources to separated *cover* and hidden message. It is enough to *cover* and hidden message be sparse in two different dictionaries. Based on sparsity property, an optimization problem is proposed and the feasibility of solving it with linear programming methods is examined. Then, a fast algorithm based on fast *transforms* is proposed to solve the problem and extract the hidden message.

To this end, the paper has been organized as follows: in section II, a brief overview of the current BSS-based *active* STs is presented and their advantages and disadvantages are discussed. Section III explains the sparse component analysis and source separation problem briefly. In the next section, the details of our *active* steganalysis method are presented. Finally, a discussion of the experimental evaluation is made in section VI and conclusions are drawn in section VII.

2. A Brief History on Active STs

Chandramouli [3] developed the first *active* ST based on the BSS model to challenge the linear steganography. His proposed method was based on the BSS model with a hypothesis that the *cover* image and hidden message were independent. However, his proposed method needed at least two stego images with the same message, *cover* and key but different embedding strength factor. However, these conditions are not practical since steganalyst can usually access one stego image only.

Fan et al. [15] tried to apply a method to realize *active* steganalysis when there was only one stego image copy. Their method was based on Independent Component Analysis (ICA) [20] and Hidden Morkov Tree (HMT) model[21]. The former is a popular BSS technique and the latter is mainly applied to denoise an image in the *transform* domain. They adopted HMT model to obtain the second copy of stego image and then the optimized ICA was applied to achieve the *active* steganalysis.

Another study with the view of active steganalysis as BSS problem was presented in [16]. It solely used a single copy of stego image. The maximum a posteriori (MAP) estimator was adopted to provide an estimate of the *cover* image. Two *active* steganalysis schemes was introduced in this method; the first scheme was similar to [15] which considered the estimated version as another stego image. In the second scheme, besides the original stego image, two other stego images were generated from the estimated image. All the three images provide an input to the ICA algorithm. These schemes were applied to extract messages from the least significant bit (LSB) steganography in spatial, discrete cosine *transform* (DCT), and discrete wavelet *transform* (DWT) domains. The results indicated that the second scheme has a better performance than the first one. The method proposed in [17] was similar to that of [16]. However, HMT model was applied in [17] while the MAP estimator was used in [16] to gain an estimate of the *cover* image.

All these *active* steganalysis methods (ICA-based *active* steganalysis) use ICA technique to separate message from image. This technique is an inherently high computational cost technique. Ambalavanan and Chandramouli, on the other hand, introduced another *active* steganalysis based on a different BSS technique [22] in order to reduce the computational cost and to improve the performance of message extraction [18]. However, their efforts were not successful.

Modaghegh et al. [19] introduced an *active* ST that reduced the computational cost and error rate . Their method was basically a combination of the blind source separation technique and MAP estimator. Additionally, they presented a new geometrical BSS method based on the minimum range of mixed sources which reduced the computational cost of their *active* ST. Their experiments showed that their *active* ST not only reduced the error rate, but also decreased the computational cost compared to the previous *active* STs. Nevertheless, all of these methods involve a denoising algorithm which increase the computational cost required to generate an estimated version of the *cover*.

3. Preliminary: Sparse Component Analysis and Blind Source Separation

A brief overview of the sparse component analysis and blind source separation problems and the possible solutions [23] are given in this section. The goal of the source separation is to retrieve an unknown source signals S from observed signals X where X and S are row vectors. The observed signals are often assumed to be a linear instantaneous mixture of source signals. Thus, it can be written as $X = A \times S$, where A denotes the unknown mixture matrix. Since A and S are unknown, some assumptions are needed to solve this problem. Basically, it is necessary to have prior knowledge of the source properties such as its independency, sparsity, bounded/unbounded states and so on. The methods based on the property of source independency are called independent component analysis [20]. On the other hand, the sparse component analyses (SCAs) are based on the property of the source sparsity [24], meaning that each source is seldom *active* and mostly (nearly) zero. Let us consider a sparsity model for the probability distribution function of the sources as follows,

$$P_{s_k}(s_k) = p_{s_k}\delta(s_k) + (1 - p_{s_k}) f_{s_k}(s_k) \quad (1)$$

Where p_{sk} is the sparsity factor of the source and f_{sk} denotes the distribution of s_k when the corresponding source is *active*.

In the general sparse representation framework, a signal vector $X \in R^N$ is modeled in the dictionary Φ as the linear combination of T elementary waveforms φ_i (atoms of dictionary):

$$X = \Phi\boldsymbol{\alpha} = \sum_{i=1}^{T} \alpha_i \varphi_i \quad (2)$$

In the case of overcomplete representation, the number of waveforms or atoms $\phi_{i\,(1\le i\le T)}$ that constitute columns of the dictionary Φ is greater than the space dimension in which X lies: $T>N$, or even $T>>N$ for highly redundant dictionaries. The separation problem of a signal or image in Φ is concerned with recovering the coefficient vector α in Eq.(2). However, as there are N equations and T unknowns, the problem has no unique solution. The solution to the underdetermined system of linear equations $X = \Phi\alpha$ can be achieved by reducing the space of candidate solutions. In the SCA problem, sparsity imposes constraints on the solutions, meaning that among all solutions of $X = \Phi\alpha$, the sparsest one is preferred (with the least number of nonzero entries α_i). In other words, the sparse decomposition problem entails solving the following minimization problem:

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^T} \|\boldsymbol{\alpha}\|_0 \quad s.t. \quad X = \Phi\boldsymbol{\alpha} \quad (3)$$

As can be seen, Eq.(3) is a *combinatorial* optimization problem that requires enumerating all collections of atoms in Φ to find the smallest set that synthesizes X. This made authors turn to approximations or relaxations of Eq. (3). Donoho and Huo [25] proposed a method to relax the non-convex ℓ_0 sparsity measure by substituting the problem in Eq.(3) with the convex problem:

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^T} \|\boldsymbol{\alpha}\|_1 \quad s.t. \quad X = \Phi\boldsymbol{\alpha}. \quad (4)$$

This problem is called *Basis Pursuit* (BP) [26]. Unlike Eq.(3), Eq.(4) is a computationally tractable convex optimization problem that can be solved efficiently by linear programming methods [27]. However, BP is not able to find a general solution for Eq.(2). Under appropriate conditions on Φ and X, nonetheless, BP can offer a general optimal solution of Eq.(2).

On the other hand, some authors have attempted to provide a more effective solution for problem Eq.(2) [28], [29]. The morphological diversity concept introduces a new data modeling framework that allows having both a sparse representation and a fast algorithm that exploits the structure of the dictionary. Morphological diversity assumes that the signal X can be modeled as the sum of K components s_k that are morphologically different:

$$X = \sum_{k=1}^{K} s_k \quad (5)$$

Where s_k is a morphological component. Each s_k is sparse in a given dictionary Φ_k, which is associated with implicit fast analysis/synthesis *transforms* such as wavelet and DCT.

4. The Proposed Active Steganalysis Method

In this section, we introduce our *active* steganalysis method that can be used for breaking the *transform* domain steganography techniques [30]–[34]. In the first subsection, *transform* domain steganography is formulated mathematically as a linear combination of sparse sources. Then, *active* steganalysis is formulated as a new SCA problem. In the next subsection, the feasibility of solving the new SCA problem is discussed. Finally, with the goal of reducing computational cost, a fast algorithm for solving the new SCA problem is proposed. At this point, we are ready to illustrate the details of our proposed method.

4.1 SCA as Active Steganalysis

Common steganography techniques can be modeled as an additive embedding, i.e. the sum of image features and hidden message. *Transform* domain steganography methods use *transform* coefficients as image features to embed hidden message [34]. Thus, these methods can be formulated as follows:

$$Y_x = Y_c + \beta W \quad (6)$$

Where Y_x denotes the *transform* coefficients of stego image, β is the embedding coefficient and W and Y_c are respectively the hidden message and *transform* coefficients of *cover* image. To obtain the stego image, we need to apply inverse *transform* to Y_x, so we have:

$$X = C + \text{invTrans}(\beta W) = C + S_{\beta w} \quad (7)$$

To perform *active* steganalysis, we need to extract hidden message W from stego image X. In this paper, we formulate *active* steganalysis as a source separation problem. From a BSS viewpoint, the *cover* image and inverse *transform* of the message are sources mixed to form the stego image as an observation. In this case, since the number of observations is less than the number of sources, the problem is underdetermined, and therefore there is an infinite number of solutions. As a result, in order to select one solution among all available solutions it is necessary to impose certain additional constraints based on the prior knowledge. In previous BSS-based *active* STs, the independency of image and hidden message was used as an additional constraint [3], [15]–[19]. However, in this paper, we show that both sources in the stego image are sparse in their dictionaries and use sources sparsity as an additional constraint to solve the BSS problem.

The hidden message is sparse because imperceptibility is a fundamental requirement of steganographic methods. This means that embedding a hidden message should not significantly change the *cover* so the hidden messages remain indiscernible to the human eye. Because of this feature, a hidden message needs to have short length and low amplitude. In other words, most of elements of a hidden message vector W are zero and the embedding coefficient β is also small. In this regard, the message source can be viewed as a sparse source in *transform* domain (Φ_1 dictionary).

On the other hand, images and practical signals are not, in general, strictly sparse. Though they may be compressible or weakly sparse in *transform* domains (Φ_2 dictionary) such as wavelet or DCT. This means that most α_is in Eq.(2) are near zero, and one can neglect all but perhaps a small fraction of the coefficients without significant loss. Thus, the image source can be viewed as a sparse source in Φ_2 dictionary.

Since the two sources are sparse, the original image can be separated from the hidden message if the stego image is searched smartly for sparse component, meaning that *active* steganalysis can be viewed as a SCA problem. As stated in Section 3, we can formulate SCA problem as an optimization problem Eq.(3). As our SCA problem is underdetermined and there are an infinite number of solutions, we need to use our prior knowledge to find the accurate answer (*cover* and hidden message), including the knowledge of α_{1i} amplitude in dictionary Φ_1 which is corresponding to the hidden message. As α_{1i} is small due to imperceptibility feature of the steganography method, Eq.(3) can be written as follows:

$$\min_{\boldsymbol{\alpha_i}\in\mathbb{R}^T}\left(\|\boldsymbol{\alpha_1}\|_0+\|\boldsymbol{\alpha_2}\|_0\right)\quad s.t.\quad \begin{matrix} x=\Phi_1\boldsymbol{\alpha_1}+\Phi_2\boldsymbol{\alpha_2} \\ |\alpha_{1i}|<\alpha_{th}\end{matrix} \tag{8}$$

$$and \qquad \Phi=\bigcup_{k=1}^{2}\Phi_k$$

Where Φ_2 and α_2 are dictionary and coefficient vector of the *cover* respectively. This optimization problem has an extra constraint $|\alpha_{1i}| < \alpha_{th}$ which helps separate hidden message from the *cover*.

4.2 Feasibility Problem

The extra constraint distinguishes this problem from normal SCA problems. Since it is a complex non-convex problem, we can simplify this problem by converting l_0 norm to l_1 norm, formulating this problem like *Basis Pursuit* as follows:

$$\min_{\boldsymbol{\alpha_i}\in\mathbb{R}^T}\left(\|\boldsymbol{\alpha_1}\|_1+\|\boldsymbol{\alpha_2}\|_1\right)\quad s.t.\quad \begin{matrix} X=\Phi_1\boldsymbol{\alpha_1}+\Phi_2\boldsymbol{\alpha_2} \\ |\alpha_{1i}|<\alpha_{th}\end{matrix} \tag{9}$$

This extra constraint also distinguishes this problem from BP. However, this is a linear programming problem which can be solved by LP methods. To do so, we need to convert Eq.(9) to a canonical form of LP problem as stated below:

$$\min B^T\boldsymbol{\alpha}\quad s.t.\quad D\boldsymbol{\alpha}\le b\quad and\quad \boldsymbol{\alpha}\ge 0 \tag{10}$$

Where B and b are vectors of (known) coefficients and D is a (known) matrix of coefficients. The inequalities $D\alpha \le b$ and $\alpha \ge 0$ are the constraints which specify a convex polytope over which the objective function is to be optimized. Clearly, Eq.(9) can be converted into the canonical form Eq.(10) by adding extra unknowns [27]. Then, this LP problem can be solved using simplex or interior point method[27].

However, Eq.(9) does not provide a solution for Eq.(8) in general. But under appropriate conditions on Φ and X, BP can provide the globally optimal solution for Eq.(8). Thus, practical algorithms can solve problems that seem computationally intractable on the surface. Many studies have focused on sufficient (and sometimes necessary) conditions under which the problem BP recovers the sparsest solution of an underdetermined system of linear equations. For instance, sufficient conditions based on the mutual coherence of Φ were introduced by several authors (see, for example, [25], [35]–[37]). The mutual coherence μ_Φ of Φ is defined as:

$$\mu_\phi=\max_{i\ne j}\left|\left\langle\varphi_i,\varphi_j\right\rangle\right| \tag{11}$$

This quantity can be viewed as a worst-case measure of the resemblance between all pairs of atoms. Donoho and Huo [25] showed that for dictionaries with small μ_Φ, the solution of BP is unique and this unique solution is a point of equivalence of Eq.(3) and Eq.(4). It can be shown that these conditions are available for our *active* steganalysis problem Eq.(9). Thus, we can easily solve Eq.(9) instead of non-convex Eq.(8) and obtain the hidden message. In the case Eq.(9), if the mutual coherence between atoms of two dictionaries Φ_1 and Φ_2 is low, this condition is met. Since in steganography, Φ_1 and Φ_2 correspond to *transform* domains, the *cover* and hidden message need to be sparse in two different *transforms* to acquire a small μ_Φ. Under this condition, the *active* steganalysis problem is feasible, and it can be solved with LP methods.

4.3 Proposing Fast Algorithm

In many cases, BP-like synthesis algorithms are computationally expensive. In this subsection, an alternative to these approaches has been proposed. We have chosen an approximation to our true minimization task to find a simplified optimization problem which is computationally effective. Our proposed method is similar to Morphological Component Analysis (MCA) method, which can be seen as a kind of *Basis Pursuit* method [26] called MCA steganalysis (MCAS). The algorithm is based on the Block-Coordinate-Relaxation method[38], with some changes made by the properties of the *active* steganalysis such as the hidden message amplitude constraint imposed on the reconstructed components.

In our algorithm, we assumed that a dictionary can be built by amalgamating two subdictionaries (Φ_1, Φ_2) such that for each k, the representation of s_k in Φ_k is sparse and not sparse – or at least not as sparse – in other Φ_l , $l \ne k$. In other words, the subdictionaries Φ_k must be mutually incoherent. Thus, the dictionary Φ_k plays the role of a discriminant between the hidden message and *cover*, preferring the component s_k over the other part. This is a key observation for the success of the separation algorithm. If this condition is not satisfied and the hidden message and *cover* sources are sparse in the same dictionary, other *active* steganalysis methods can be used [19].

Since most *transform* domain steganography methods [30]–[34] use fast *transforms*, in our algorithm, the matrix Φ_k and its transpose Φ_k^T corresponding to each *transform*

are never explicitly constructed in the memory. Instead, they are implemented as a fast implicit analysis and synthesis *transforms* taking a signal vector X and returning $\Phi_k^T X = T_k(X)$ (analysis side), or taking a coefficient vector α_k and returning $\Phi_k \alpha_k = \Phi_k^{-1}(X)$ (synthesis side). In the case of a simple orthogonal basis, the inverse of the analysis *transform* is trivially $T_k^{-1} = \Phi_k$. The use of this transformation instead of the dictionary is what makes our method computationally efficient.

One of the important ingredients of our algorithm is the coordinate relaxation. If all component coefficients α_l but the kth are fixed, then a solution can be achieved through thresholding the coefficients of the marginal residuals $r_k = X - \sum_{l \neq k} \Phi_l \alpha_l$ in Φ_k. The other components are relieved of these marginal residuals r_k and are likely to contain mainly the salient information of s_k. This intuition dictates a coordinate relaxation algorithm that cycles through the components at each iteration and applies a thresholding to the marginal residuals. MCAS algorithm is summarized as follows:

Algorithm 1:

1. Initialize iteration number $j=1$, N_{iter} the number of iterations, threshold δ_0, and step size $\lambda = \delta_0 / N_{iter}$.
2. Calculate the residual $r_1=X-s_2$, $r_2=X-s_1$.
3. Calculate the *transform* T_k of r_k and obtain $\alpha_k = T_k(r_k)$ for $k=1,2$.
4. Apply hard threshold to the coefficient α_k with the δ_j threshold and obtain $\hat{\alpha}_k$ for $k=1,2$.
5. Reconstruct s_k by $s_k=T^1(\hat{\alpha}_k)$ for $k=1,2$.
6. Apply the constraint correction if $l_0(\hat{\alpha}_1) \leq N_{msg}/K_1$, $s_1=0$.
7. Update the threshold by $\delta_j=\delta_{j-1}-\lambda$.
8. If $l_0(\hat{\alpha}_1) \leq N_{msg}$, update the $j=j+1$ and return to Step 2, else , finish.

Unlike BP, the MCAS is stage wise and exploits the fact that the dictionary is structured (union of *transforms*), and the atoms enter the solution by groups, rather than individually. As such, MCAS is a salient-to-fine process in which the most salient content of each morphological component is iteratively computed at each iteration. These estimates are then progressively refined as the threshold δ evolves toward δ_{min}. In the above algorithm, we use hard thresholding instead of soft thresholding due to our formulation of the l_0 sparsity penalty term and the fact that hard thresholding provides better results [39].

Besides fast *transforms* and coordinate relaxation, another important ingredient of MCAS is iterative thresholding with varying threshold. The way the threshold is decreased along the iterations of the MCAS algorithm is significant in terms of separation quality. There are two types of decreasing threshold strategy: prefixed decreasing threshold strategy and adaptive strategy. Given the information we have about the amplitude of α_1, the prefixed decreasing threshold is selected. In our study, the threshold is decreased linearly so the threshold δ sequence is as follows:

$$\delta_j = \delta_0 - j \times (\delta_0 - \delta_{min}) / N_{iter} \quad (12)$$

The first threshold δ_0 can be set automatically to a sufficiently large value which is greater than $\hat{\beta}$ (the estimate of the message embedding coefficient), e.g. $\delta_0 = K_1 \hat{\beta}$ where K_1 can be a number between 1.5 and 3. For an exact representation of the data with the morphological components, δ_{min} must be set to zero. In our algorithm, considering the information we have about the message embedding rate ($_{Nmsg}$) the algorithm can be repeated until the number of non-zero element of $\hat{\alpha}_1$, i.e. $l_0(\hat{\alpha}_1)$, become equal to N_{msg}. Since some *passive* steganalysis methods [40]–[45] can estimate the message embedding rate, it is practical to use this prior information in our algorithm.

The key part of our algorithm is step 6. In regular SCA method [28], [29], [46] all s_k are updated in every iteration. In our algorithm, however, the hidden message source has low and almost equal amplitude. Thus, we use this prior information to update the message source in case the number of extracted hidden message is greater than N_{msg}/K_2 (where K_2 is a number between 3 and 5). Otherwise, the extracted data does not belong to the hidden message source and s_1 is not updated. Under ideal circumstances where mutual coherence is zero, this step would be unnecessary. In practical conditions, however, the mutual coherence between Φ_1 and Φ_2 is not zero and the image source s_2 also produces some large elements in α_2. Under these conditions, some elements of α_1 acquire large amplitudes, not corresponding to hidden message s_1 and should be discarded. Fortunately, since the number of these large amplitude samples in hidden message coefficient vector α_1 is low, it can be discarded in step 6.

Another challenge facing our *active* steganalysis method is the message dictionary selection, i.e. the detection of the dictionary used in steganography algorithm. To do so, first a proposition based on the central limit theorem is presented and then it is used to detect the embedding dictionary.

Proposition 1: suppose that Φ_1 and Φ'_1 are uncorrelated dictionaries which correspond to orthogonal *transforms*. Furthermore, α_1 and α'_1 are defined as follows:

$$s_1 = \Phi_1 \boldsymbol{\alpha}_1 = \sum_{i=1}^{T} \alpha_{1i} \varphi_{1i}$$
$$s_1 = \Phi'_1 \boldsymbol{\alpha}'_1 = \sum_{i=1}^{T} \alpha'_{1i} \varphi'_{1i} \quad (13)$$

Now if s_1 is sparse in Φ_1 with sparsity factor $p_{s1}=1-L/T$ and $|\alpha_{1i}|_{1<i<L<T} = \beta$, then the probability distribution of $\acute{\alpha}_{1i}$ is normal and $\acute{p}_{s1} < p_{s1}$.

Proof:

Since Φ_1 and Φ'_1 are orthogonal transform, we have:

$$\begin{bmatrix} \varphi_{11} \\ \dots \\ \varphi_{1T} \end{bmatrix} = \begin{bmatrix} e_{11} & \dots & e_{1T} \\ \dots & & \\ e_{T1} & & e_{TT} \end{bmatrix} \times \begin{bmatrix} \varphi'_{11} \\ \dots \\ \varphi'_{1T} \end{bmatrix} \quad (14)$$

Thus, each atom φ_{1i} can be written as $\varphi_{1i} = e_{i1}\varphi'_{11} + \dots + e_{iT}\varphi'_{1T}$.Now, substituting φ_{1i} in Eq.(13) :

$$s_1 = \Phi_1 \boldsymbol{\alpha}_1 = \sum_{i=1}^{T} \alpha_{1i} \varphi_{1i} = \sum_{i=1}^{L} \pm\beta\varphi_{1i} = \beta \sum_{i=1}^{L} (\pm e_{i1}\varphi'_{11} \pm ... \pm e_{iT}\varphi'_{1T}) = \sum_{i=1}^{T} \alpha'_{1i}\varphi'_{1i} \quad (15)$$

Thus:

$$\alpha'_{1i} = \beta(\pm e_{i1} \pm ... \pm e_{iT}) \quad (16)$$

In the case of orthogonal *transforms*, if we suppose that e_{ij} is an independent random variable with identical distribution (i.i.d), then $\acute{\alpha}_{1i}$ will be the sum of a large number of i.i.d random variables. Therefore, according to the *central limit theorem* in the probability theory, $\acute{\alpha}_{1i}$ will have approximately normal distribution. Since $\acute{\alpha}_{1i}$ is a normally distributed random variable, most of $\acute{\alpha}_{1i}$ s are non-zero and therefore $\acute{p}_{s1}$ will be less than *ps1*.

According to this proposition, if signal s_1 is sparse in dictionary Φ_1 then it would be less sparse in another dictionary Φ'_1. Thus, if a message is embedded in the *transform* domain Φ_1 and the hidden message is searched in another dictionary Φ'_1, the extracted message will be less sparse. This property is used to detect the embedding dictionary. To do so, we need to run algorithm 1 until the condition stated in step 6 is met. Then, the corresponding threshold δ is saved for each dictionary. This process is executed for all *transforms* that might be used in steganography and the dictionary with the greatest threshold is selected. The results of running algorithm for some *transforms* will be presented in the next section.

5. Experimental Results

At this point, the simulation results of our *active* steganalysis method are presented so that we can evaluate its performance. The simulation was done in MATLAB environment. To evaluate the steganalysis method, we used Berkeley Segmentation Data Set and Benchmarks 500 (BSDS500) [47], a dataset consisting of 500 natural color images. These color images, span a range of indoor and outdoor scenes, are in JPEG compressed format with a quality of 75%. We converted these images into grayscale images using only the central 256×256 region of each image.

Quality factor 75 has been selected because the default quality setting on most digital cameras and image editors is 75, which provides a good tradeoff between the file size and the perceived quality. With this quality factor, the sparsity factor of DCT coefficient (p_{s2}) for our dataset is about 0.8. The sparsity factor of the embedded message (ps1) is set to 0.1 (i.e. the message is embedded in 10% of DWT coefficients) in the first simulation (Table 1), while different embedding rates are selected in the next simulation (Table 2). For example, in an image of 256×256 pixels, the number of DWT (Haar Wavelet) coefficients is 256×256. Thus, when data is embedded in 10% of coefficients, the capacity of data hiding will be 0.1 bit/pixel. In other words, the number of bits embedded in the image is 256×256×0.1=6553.6 with a data hiding capacity of 6553.6 bit/image. Binary message bits (± 1) are randomly embedded in the DWT coefficients with β=4. As an example, the original, extracted and stego images with 0.1 bit/pixel embedded message rate have been shown in Fig. 1.

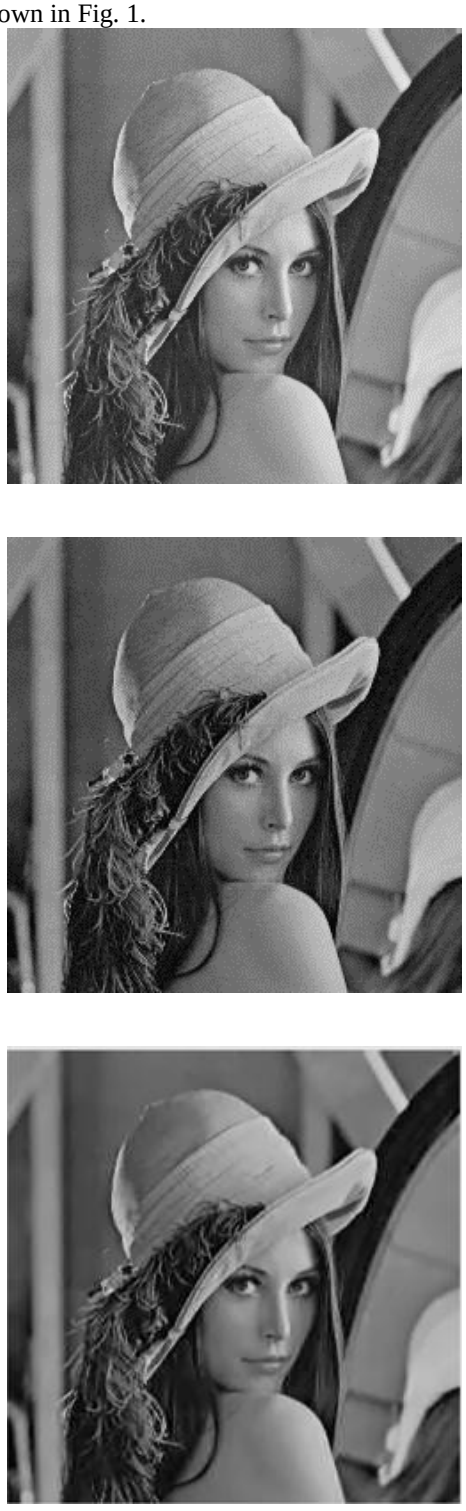

Fig. 1. Original Lena image (a) stego image (b) extracted image (c).

The results of extracting message from 500 stego images using the proposed steganalysis method, ICA-

based *active* steganalysis methods [15][16][17] and Modaghegh et al.'s method [19] are shown in Table 1. In this table, for ease of comparison and preparation of similar conditions for all methods, ICA-based *active* steganalysis methods are simulated by "Bayesian Least Squares-Gaussian Scale Mixture" (BG) denoising algorithm [48], which has been employed in Modaghegh et al.'s steganalysis method [19]. This algorithm is one of the most effective denoising algorithms for removing homogeneous additive noise from natural images.

It is important to note that the results of ICA-based *active* steganalysis are for steganography methods with the message embedded in block-DCT coefficient [15][16][17] [19] not the DWT coefficient. Nonetheless, because there is no similar study in *active* steganalysis, these methods have been compared in general.

Table 1. A comparison between the proposed *active* steganalysis (message in DWT coefficient) and the ICA-based *active* steganalysis methods (message in DCT Coefficient) for the message extracted from 500 stego images.

***Active* steganalysis method**	**Mean of true extracted message bits (bit/image)**	**Mean of false extracted message bits (bit/image)**	**Mean error rate**
Our proposed *active* steganalysis	5521.27	1021.34	% 15.64
Modaghegh et al.'s *active* steganalysis [19]	5121.96	1551.96	% 23.29
ICA-based *active* steganalysis	4785.52	1638.48	% 25.51

In Table 1, false extracted message bits are embedded message bits that their sign of detected samples is not equal to the embedded ones. Moreover, the error rate is defined as follows:

$$\text{Error rate} = \frac{\text{False extracted bits}}{\text{All embedded message bits}} \qquad (17)$$

As noted earlier, in above simulations, it is supposed that the message embedding rate is known. This, however, should not be considered as a restrictive assumption because there are *passive* steganalysis methods [40], [45], [49] which can precisely estimate the message embedding rate.

As shown in Table 1, our proposed method has lower error rate and higher true detected bits compared to that of ICA-based *active* steganalysis methods.

Table 2 also shows the comparative error rate of steganalysis methods for different message embedding rates. The results confirm that our method has almost similar performance in all low embedding rates, but when the embedding rate increases, the error rate is increased gradually. It is not surprising since the embedding message was supposed to be sparse, i.e. steganography has low embedding rate.

Table 2. Mean error rate for messages extracted from 100 stego images. Message embedding rate varies from 0.02 to 0.25 bits/pixel.

Message embedding rate (bit/pixel)	**0.02**	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**
Mean error rate of the proposed *active* steganalysis	11.33%	12.19%	15.64%	19.80%	25.73%	30.76%
Mean error rate of Modaghegh et al.'s *active* steganalysis [19]	26.65%	24.01%	22.89%	22.29%	21.98%	21.95%
Mean error rate of the ICA-based *active* steganalysis	23.03%	24.29%	24.67%	25.35%	25.64%	26.01%

We have also drawn the comparative figure for different embedding rates (Fig. 2).

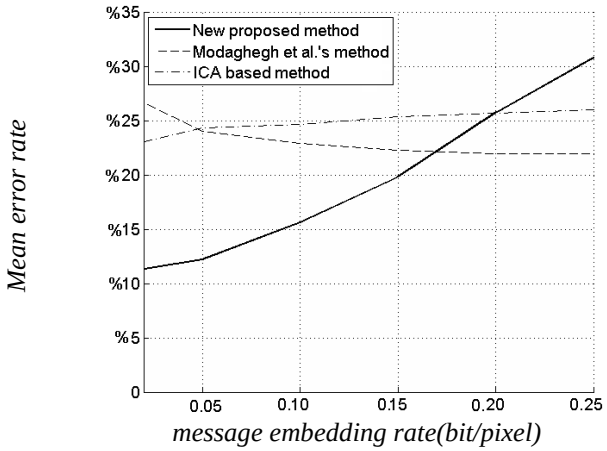

Fig. 2. Mean error rate for messages extracted from 100 stego images.

Additionally, our steganalysis method is simulated for different wavelet *transform* such as Symlet, Haar and Coiflet in Table 3. The condition of message embedding and steganalysis is similar to Table 1. The mean error rate, as shown in Table 3, is about 15% for different wavelet *transforms*.

Table 3. Mean error rate for messages extracted from 500 stego images. The message is embedded with different *transforms*.

Transform (Dictionary)	**Mean error rate**
Coiflet Wavelet	15.89%
Reverse Biorthogonal Wavelet	15.34%
Haar Wavelet	15.64%
Biorthogonal Wavelet	15.33%
Symlet Wavelet	15.40%

As mentioned earlier, our active ST does not need to know the *transform* domain in which the message is embedded, as it is determined by the authors. To do so, we embed the message in the DWT domain with Coiflet Wavelet, trying to extract it by another Wavelet. Then, the error rate and threshold of these Wavelets are compared (Table 4). As can be seen, the threshold of Coiflet Wavelet is greater than the other wavelets, suggesting that the message is probably embedded in this *transform*. Additionally, for two wavelets with great threshold, it can be concluded that two dictionaries are mutually coherent, so either of them can be used for extraction. For example, the results of data extracted by Meyer Wavelet correspond to Coiflet Wavelet as they are mutually coherent.

Table 4. Mean error rate and threshold for message extraction from 500 stego images. Message is embedded in Coiflet Wavelet and extracted with different dictionary.

Transform (Dictionary)	Mean error rate	Mean of threshold	Variance of threshold
Coiflet Wavelet	15.89%	3.97	0.0062
Haar Wavelet	38.77%	3.50	0.0410
Meyer Wavelet	22.87%	3.93	0.0181
Biorthogonal Wavelet	38.77%	3.5	0.0410
Symlet Wavelet	78.30%	3.2	0.0001

Finally, we compare the computational time of ICA-based *active* steganalysis methods with our proposed steganalysis method on 2.00 GHz Pentium 4 workstation. The computational time of applying steganalysis methods to stego image with different random messages have been shown in Table 5. Here, the calculated times are greater than the ones shown in [19] because the time of common parts of two methods such as denoising algorithm and DCT calculations have not been included in [19].

Table 5. the computational time of extracting message from 500 stego images by our steganalysis and ICA-based *active* steganalysis

Steganalysis method	Computational Time (second)	
	Mean	Variance
Our proposed *active* steganalysis	2.0327	0.7120
ICA-based *active* steganalysis	8.1985	1.4202

As expected, the computational time of our proposed *active* ST is lower than that of ICA-based *active* STs since our ST, does not have the denoising algorithm and works with only one image.

6. Conclusions

In this paper, a new *active* steganalysis method based on sparsity property of signals was proposed. Our method provided satisfactory performance on stego images in which the *cover* and hidden messages were sparse in different dictionaries. We first formulated the *active* steganalysis method as an SCA problem. Then, the feasibility of solving the SCA problem was demonstrated mathematically. Since fast *transforms* are employed in most *transform* domain steganography methods, in this study a fast algorithm was presented to solve our SCA problem.

The results of experiments showed that nearly 85% of the message bits could be estimated when the sparsity factor of message was 10%. Additionally, experiments confirmed that the computational cost of our method was approximately one fourth of the previous ICA-based *active* STs.

Overall, the comparison between our proposed method and the previous *active* steganalysis schemes revealed that the use of sparsity property of signals improved the steganalysis performance (computational cost and error rate).

References

[1] N. F. Johnson and S. Jajodia, "Steganalysis of images created using current steganography software", in Information Hiding, 1998, pp. 273–289.

[2] N. F. Johnson and S. Jajodia, "Steganalysis: The investigation of hidden information", in Information Technology Conference, 1998, pp. 113–116.

[3] R. Chandramouli, "A mathematical framework for active steganalysis", Multimedia systems, vol. 9, no. 3, 2003, pp. 303–311.

[4] X. Yu, T. Tan, and Y. Wang, "Reliable detection of BPCS-steganography in natural images", in IEEE First Symposium on Multi-Agent Security and Survivability, 2004, pp. 333–336.

[5] A. Nissar and A. Mir, "Classification of steganalysis techniques: A study", Digital Signal Processing, vol. 20, no. 6, 2010, pp. 1758–1770.

[6] T. Pevny, P. Bas, and J. Fridrich, "Steganalysis by Subtractive Pixel Adjacency Matrix", IEEE Transactions on Information Forensics and Security, vol. 5, no. 2, 2010, pp. 215–224.

[7] J. Fridrich and J. Kodovskỳ, "Steganalysis of LSB replacement using parity-aware features", in Information Hiding, 2013, pp. 31–45.

[8] J. J. Harmsen and W. A. Pearlman, "Steganalysis of additive-noise modelable information hiding", in Proceedings of SPIE, 2003, vol. 5020, pp. 131–142.

[9] I. Avcibas, N. Memon, and B. Sankur, "Steganalysis using image quality metrics", Image Processing, IEEE Transactions on, vol. 12, no. 2, 2003, pp. 221–229.

[10] S. Lyu and H. Farid, "Steganalysis using higher-order image statistics", Information Forensics and Security, IEEE Transactions on , vol. 1, no. 1, 2006, pp. 111–119.

[11] Y. Yan, L. T. Li, J. B. Xue, H. G. Liu, and Q. Y. Zhang, "Study on Universal Steganalysis for BMP Images Based on Multi-Domain Features", Applied Mechanics and Materials, vol. 278, 2013, pp. 1906–1909.

[12] D.-C. Lou, C.-L. Chou, T.-C. Wei, and H.-F. Huang, "Active steganalysis for interpolation-error based reversible data hiding", Pattern Recognition Letters, vol. 34, no. 3, 2013, pp. 1032–1036.

[13] D.-C. Lou, C.-L. Chou, H.-K. Tso, and C.-C. Chiu, "Active steganalysis for histogram-shifting based reversible data hiding", Optics Communications, vol. 285, no. 10, 2012, pp. 2510–2518.

[14] L. J. T. Guang-ming, "Active Steganalysis Based on Pixels Classification in the Image", Journal of Electronics & Information Technology, vol. 8, no 34, 2012, pp.1928-1933.

[15] F. Fan, W. Jiazhen, L. Xiaoqin, and F. Huijuan, "An Active Steganalysis Method of Block-DCT Image Information Hiding", in 8th International Conference on Electronic Measurement and Instruments, 2007, pp. 2–849–2–852.

[16] B. Xu, Z. Zhang, J. Wang, and X. Liu, "Improved BSS based schemes for Active steganalysis", in Eighth ACIS

International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007, vol. 3, pp. 815–818.
[17] L. Wenzhe, X. Bo, Z. Zhe, and R. Wenxia, "Active steganalysis with only one stego image", in Sixth International Conference on Fuzzy Systems and Knowledge Discovery, 2009, vol. 5, pp. 345–348.
[18] A. Ambalavanan and R. Chandramouli, "Blind source separation for steganalytic secret message estimation", in Proceedings of SPIE, 2007, vol. 6505, p. 650507.
[19] H. Modaghegh and S. A. Seyedin, "A new fast and efficient active steganalysis based on combined geometrical blind source separation", Multimedia Tools and Applications, pp. 1–19, 2014.
[20] C. Jutten, "Independent components analysis versus principal components analysis", in Fourth European Signal Processing Conference , 1988, pp. 643–646.
[21] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models", Signal Processing, IEEE Transactions on, vol. 46, no. 4, 1998, pp. 886–902.
[22] L. Vielva, D. Erdogmus, and J. C. Principe, "Underdetermined blind source separation using a probabilistic source sparsity model", in Proceeding of International Conference on Independent Component Analysis and Blind Source Separation (ICA), 2001, pp. 675–679.
[23] C. Jutten and M. Babaie-Zadeh, "Source separation: Principles, current advances and applications", IAR Annual Meeting Nancy France, 2006, pp. 1–10.
[24] R. Gribonval and S. Lesage, "A survey of sparse component analysis for blind source separation: principles, perspectives, and new challenges", in 14th European Symposium on Artificial Neural Networks , 2006, pp. 323–330.
[25] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition", Information Theory, IEEE Transactions on, vol. 47, no. 7, 2001, pp. 2845–2862.
[26] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit", SIAM journal on scientific computing, vol. 20, no. 1, 1998, pp. 33–61.
[27] D. G. Luenberger and Y. Ye, Linear and nonlinear programming, Springer, 2008.
[28] J.-L. Starck, Y. Moudden, J. Bobin, M. Elad, and D. L. Donoho, "Morphological component analysis", in Optics & Photonics Conference, 2005, p. 59140Q–59140Q.
[29] J. Bobin, J.-L. Starck, J. M. Fadili, Y. Moudden, and D. L. Donoho, "Morphological component analysis: An adaptive thresholding strategy", Image Processes sing IEEE Transactions On, vol. 16, no. 11, 2007, pp. 2675–2681.
[30] P. Liu, C. Chen, L. Ge, and Y. Luo, "Efficient Self-Adaptive Image Steganography Scheme Based on Iterative Blending and Integer Wavelet Transform", in Unifying Electrical Engineering and Electronics Engineering, Springer, 2014, pp. 1159–1167.
[31] X. Wang, C. Wang, H. Yang, and P. Niu, "A robust blind color image watermarking in quaternion Fourier transform domain", Journal of Systems and Software, vol. 86, no. 2, 2013, pp. 255–277.
[32] X.-Y. Wang, A.-L. Wang, H.-Y. Yang, Y. Zhang, and C.-P. Wang, "A new robust digital watermarking based on exponent moments invariants in nonsubsampled contourlet transform domain", Computers & Electrical Engineering, vol. 40, no. 3, Apr. 2014, pp. 942–955.
[33] M. Jayamohan and K. Revathy, "A Hybrid Fractal-Wavelet Digital Watermarking Technique with Localized Embedding Strength", in Wireless Networks and Computational Intelligence, Springer, 2012, pp. 584–591.
[34] Q. Cheng and T. S. Huang, "An additive approach to transform-domain information hiding and optimum detection structure", Multimedia, IEEE Transactions on, vol. 3, no. 3, 2001, pp. 273–284.
[35] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ1 minimization", in Proceedings of the National Academy of Sciences, vol. 100, no. 5, 2003, pp. 2197–2202.
[36] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases", Information Theory, IEEE Transactions on, vol. 48, no. 9, 2002, pp. 2558–2567.
[37] D. L. Donoho, "Compressed sensing", Information Theory, IEEE Transactions on, vol. 52, no. 4, 2006, pp. 1289–1306.
[38] S. Sardy, A. G. Bruce, and P. Tseng, "Block coordinate relaxation methods for nonparametric wavelet denoising", Journal of computational and graphical statistics, vol. 9, no. 2, 2000, pp. 361–379.
[39] J. Bobin, J.-L. Starck, J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation", Image Processing, IEEE Transactions on, vol. 16, no. 11, 2007, pp. 2662–2674.
[40] X. Yu, Y. Wang, and T. Tan, "On estimation of secret message length in JSteg-like steganography", in Proceedings of the 17th International Conference on Pattern Recognition, 2004, vol. 4, pp. 673–676.
[41] T. Holotyak, J. Fridrich, and D. Soukal, "Stochastic approach to secret message length estimation in$\pm$k embedding steganography", in Proceeding of SPIE, vol. 5681, 2005, pp. 673–684.
[42] M. Jiang, E. Wong, N. Memon, and X. Wu, "A simple technique for estimating message lengths for additive noise steganography", in Control, Automation, Robotics and Vision Conference, 2004, vol. 2, pp. 983–986.
[43] J. Fridrich and M. Goljan, "On estimation of secret message length in LSB steganography in spatial domain", in Electronic Imaging, 2004, pp. 23–34.
[44] J. Fridrich, M. Goljan, D. Hogea, and D. Soukal, "Quantitative steganalysis of digital images: estimating the secret message length", Multimedia Systems, vol. 9, no. 3, 2003, pp. 288–302.
[45] T. Zhang and X. Ping, "A fast and effective steganalytic technique against JSteg-like algorithms", in Proceedings of the 2003 ACM symposium on Applied computing, 2003, pp. 307–311.
[46] P. Abrial, Y. Moudden, J.-L. Starck, B. Afeyan, J. Bobin, J. Fadili, and M. K. Nguyen, "Morphological component analysis and inpainting on the sphere: Application in physics and astrophysics", Journal of Fourier Analysis and Applications, vol. 13, no. 6, 2007, pp. 729–748.
[47] "UC Berkeley Computer Vision Group - Contour Detection and Image Segmentation - Resources". [Online]. Available: http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html#bsds500. [Accessed: 31-Mar-2013].
[48] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain", Image Processing IEEE Transactions On, vol. 12, no. 11, 2003, pp. 1338–1351.
[49] J. Fridrich, M. Goljan, and D. Hogea, "Steganalysis of JPEG images: Breaking the F5 algorithm", in Information Hiding, 2003, pp. 310–323.

Hamed Modaghegh was born in Mashhad, Iran. In 2005 he received the B.Sc in Electrical Engineering from Iran University of Science and Technology, the M.Sc in Communication Engineering from Sharif University of Technology in 2007 and the Ph.D. degree in Communication from Ferdowsi University of Mashhad in 2015. His current research interests include signal processing and data hiding.

Seyed Alireza Seyedin was born in Iran. He received the B.S degree in Electronics Engineering from Isfehan University of Technology, Isfehan, Iran in 1986, and the M.E degree in Control and Guidance Engineering from Roorkee University, Roorkee, India in 1992, and the Ph.D degree from the University of New South Wales, Sydney, Australia in 1996. He has been an Associate Professor with the Department of Electrical Engineering, the University of Mashhad (Ferdowsi), Mashhad, Iran.