Hybrid Dictionary Learning for JPEG Steganalysis

Zhihao Xu, Yanqing Guo, Jun Guo, and Xiangwei Kong
School of Information and Communication Engineering
Dalian University of Technology, Dalian, China, 116023
E-mail: {xuzhihao, guojun}@mail.dlut.edu.cn; {guoyq, kongxw}@dlut.edu.cn

***Abstract*—In order to distinguish cover images and stego images, JPEG steganalysis technology has growing ties with machine learning in recent years. As an important research field in machine learning, dictionary learning (DL) has been successfully applied to various tasks, but its application in steganalysis is insufficient. In this paper, we propose a hybrid dictionary learning framework for JPEG steganalysis based on the fact that features of stego images have a close connection with image content but not helpful for steganalysis. We learn class-specific dictionaries for both cover and stego images to obtain their different particularities simultaneously. Besides, we learn a shared dictionary which can represent the common content of both sides. In such a way, class-specific dictionaries are used for classification while the shared dictionary contributes to reconstructing data. In addition, the proposed method also learns a synthesis dictionary for representation, and an analysis dictionary to achieve good classification. Compared with previous methods, our method dose not need l_0-norm or l_1-norm constraint and employs linear projection for obtaining the discriminative codes to make our method more efficient. So the hybrid dictionary is a phrase with double meaning, it is not only the combination of the class-specific dictionary and the shared dictionary, but also the incorporation of the synthesis dictionary and the analysis dictionary. The experimental results indicate that our hybrid dictionary learning based steganalysis method could achieve good performance.**

I. Introduction

The purpose of steganography is hiding information into a cover object, then getting a stego object for transmitting in an insecurity channel [9]. Steganalysis is revealing the presence of secret messages embedded in objects. Both fields have developed rapidly in the past few decades [2], [4], [20], [21]. Historically, two disparate kinds of ways are used in steganalysis. One is built on employing a statistical model to design the detector. On the other hand, the detector can be constructed by machine learning when the cover objects are well represented by vectors. Statistical signal processing approaches are completely different from machine-learning steganalysis. The former is only focusing on distinguishing the two classes instead of figuring out the model of the cover and stego objects. Therefore, the detector can be obtained by using finite number of training examples. Generally speaking, in the same station, feature-based steganalysis usually performs better than analytically derived detectors.

There are two parts in machine-learning steganalysis: the classification algorithm and the features [5], [6], [10]. Classification tools are very important, and the support vector machine (SVM) with Gaussian kernel is usually used when the feature's dimension is small. But with growing feature spaces and training sets, it becomes computationally unfeasible for SVM to search for hyperparameters. Hence, simpler classifiers such as ensemble classifier (EC) are needed [7]. EC could not only keep the classification accuracy, but also reduce the calculation complexity greatly and increase classification speed. It is comprised by a series of linear base learners which are trained by random training sets. The ensemble classifiers make decisions according to all the base learners. Although there are many differences between the laboratory and the real world, we could get some inspirations from the lab. Many researchers have proposed feature extraction methods for image steganalysis, but there exists lots of similarities between the stego and cover images' context which affect the classification.

In order to improve the ability of distinguishing the cover and stego images, we propose a hybrid dictionary learning scheme for steganalysis. We can regard the binary classification of steganalysis as image classification with two categories [12], [14]. However, the major difference is the degree of differences. In steganography, stego images are obtained by slightly amending DCT coefficients or other numerical factors of the cover images. If we employ the traditional coding methods for the steganalysis features [11], [13], [15], the learned dictionary may have good representation ability but poor discriminability. In other words, most dictionary atoms represent the common parts, few atoms are used to encode the differences, which would result in a low classification accuracy since cover and stego features are encoded by identical dictionary atoms.

To address this issue, we learn a shared dictionary for both cover and stego image features and a class-specific dictionary for each of them. The shared dictionary could encode both two kinds of features and contribute to the representing power of dictionary, and the class-specific dictionary is used to encode the differences of cover and stego images [1], [3], [8]. But, the sparse solving process of dictionary learning requires a lot of computation. In order to improve the efficiency, we are motivated by a new dictionary learning model which dose not need l_0-norm or l_1-norm constraint, and save much more time than previous methods [19]. With the proposed hybrid dictionary learning method, the overall dictionary could lead to better image representation and achieve a satisfying steganalysis performance.

The corresponding author is Yanqing Guo. This work is supported by the National Natural Science Foundation of China (Grant No. 61402079).

II. Hybrid Dictionary Learning

A. Basics of Dictionary Learning

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N] \in \mathbb{R}^{l \times N}$ be the original data set. The basic dictionary learning (DL) scheme is to obtain an overcomplete dictionary $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_K] \in \mathbb{R}^{l \times K}$ which can provide an effective representation for each sample $\mathbf{x}_i \in \mathbb{R}^l$. Let $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_N] \in \mathbb{R}^{K \times N}$ denote $\mathbf{X}$'s corresponding codes over the learned dictionary $\mathbf{D}$. Most existing DL methods aim to solve:

$$\min_{\mathbf{D},\mathbf{A}} \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 + \tau_1 \|\mathbf{A}\|_p + \tau_2 \phi(\mathbf{X}, \mathbf{D}, \mathbf{A}, \mathbf{M}). \quad (1)$$

where τ_1 and τ_2 are regularization parameters, $\mathbf{M}$ is the label matrix of $\mathbf{X}$. The term $\|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2$ is the reconstruction error, and $\|\mathbf{A}\|_p$ is an l_p-norm constraint on $\mathbf{A}$. $\phi(\mathbf{X}, \mathbf{D}, \mathbf{A}, \mathbf{M})$ denotes functions which can promote the discriminability of $\mathbf{D}$ or $\mathbf{A}$.

At present, the most methods of DL uses are to learn a dictionary for all classes, some obtain a classifier for sparse coding in the meantime. But, they use l_0-norm or l_1-norm to constraint the coding coefficients in order to get sparse codes. That will lead to excessive calculation and cost too much time. In this paper, we employ dictionary pair learning (DPL) model to acquire a synthesis dictionary and an analysis dictionary simultaneously. No l_0-norm or l_1-norm bound term is needed in DPL, the model employs matrix multiplication to acquire codes which spends less time and achieves a better classification accuracy.

B. Dictionary Pair Learning Model

The traditional dictionary learning model employs dictionary $\mathbf{D}$ to represent the data, and requires l_0-norm or l_1-norm to proceed sparse coding. If we could get an analysis dictionary, denoted by $\mathbf{\Omega} = [\omega_1; \omega_2; \cdots; \omega_K] \in \mathbb{R}^{K \times l}$, where ω_i denotes the ith row of the $\mathbf{\Omega}$, in some way, the sparse codes $\mathbf{A}$ are obtained by $\mathbf{\Omega X}$, this process could simplify calculation a lot. DPL based on the idea learns an analysis dictionary $\mathbf{\Omega}$ and a synthesis dictionary $\mathbf{D}$ at the same time, the DPL model is expressed as follow:

$$\min_{\mathbf{D},\mathbf{\Omega}} \|\mathbf{X} - \mathbf{D}\mathbf{\Omega}\mathbf{X}\|_F^2 + \tau \phi(\mathbf{X}, \mathbf{D}, \mathbf{\Omega}, \mathbf{Y}). \quad (2)$$

where $\mathbf{D}$ and $\mathbf{\Omega}$ are a couple of dictionaries, $\mathbf{D}$ is used to reconstruct original signals, and $\mathbf{\Omega}$ is multiplied with $\mathbf{X}$ to obtain the codes. $\phi(\mathbf{X}, \mathbf{D}, \mathbf{\Omega}, \mathbf{Y})$ plays an important role in DPL model's discrimination power, and this term has to be considered in classification. We learn a synthesis dictionary $\mathbf{D}_c$ and an analytic dictionary $\mathbf{\Omega}_c$ corresponding to class c. Recently, studies on sparse coding indicate the signals from class c can be represented well by the corresponding dictionary when the signals from different classes are disparity to each others. The signals from class i $(i \neq c)$ are projected to be zero, because of $\mathbf{\Omega}$'s structured quality, i.e.,

$$\mathbf{\Omega}_c \mathbf{X}_i \approx 0, \forall i \neq c. \quad (3)$$

Therefore, the matrix $\mathbf{\Omega X}$ should be block diagonal. Except it, synthesis dictionary $\mathbf{D}$ is structed, so, class-specific dictionary $\mathbf{D}_c$ is excepted to be reconstruct the signals $\mathbf{X}_c$ well, then the first term of Equation (2) is:

$$\min_{\mathbf{D},\mathbf{\Omega}} \sum_{c=1}^{C=2} \|\mathbf{X}_c - \mathbf{D}_c \mathbf{\Omega}_c \mathbf{X}_c\|_F^2. \quad (4)$$

Through the above analysis, we consider Equation (3) and Equation (4) in the mean time, then the following model could be obtained:

$$\begin{aligned} \min_{\mathbf{D},\mathbf{\Omega}} \quad & \sum_{c=1}^{C=2} \|\mathbf{X}_c - \mathbf{D}_c \mathbf{\Omega}_c \mathbf{X}_c\|_F^2 + \lambda \left\|\mathbf{\Omega}_c \overline{\mathbf{X}}_c\right\|_F^2, \\ s.t. \quad & \|\mathbf{d}_i\|_2^2 \leqslant 1. \end{aligned} \quad (5)$$

where $\overline{\mathbf{X}}_c$ indicates the supplementary set of the $\mathbf{X}_c$, λ is the factor of the penalty term, $\mathbf{d}_i$ is the ith column of the synthesis dictionary $\mathbf{D}$.

C. Hybrid Dictionary Learning

The cover images and stego images are significantly similar in the visual angle, then the class-specific dictionary of them will share a number of atoms together which are important to reconstruct the signals and are useless for classification. These atoms can be replaced with each other when they are used to represent the signal, cover and stego images may be mixed in classification. To address this issue, we propose to learn a shared dictionary for all categories and a category-specific dictionary for each category in the feature coding process. Therefore, the combination of the commonality and the particularity (corresponding to the specific class) can faithfully represent the samples from each class, and the particularities are more discriminative and more compact for steganalysis. We denote the dictionary as $[\mathbf{D}_c, \mathbf{D}] = \mathbf{G}_c \in \mathbb{R}^{l \times (K_c + K)}$. In order to express commodiously, we introduce a selection multiplier $\mathbf{Q}_c = [\mathbf{q}_1^c, \mathbf{q}_2^c, \cdots, \mathbf{q}_K^c] \in \mathbb{R}^{(K_c+K) \times K}$, the ith column is:

$$\mathbf{q}_i^c = [\underbrace{0, \cdots, 0}_{K_c}, \underbrace{\overbrace{0, \cdots, 0}^{i-1}, 1, 0, \cdots, 0}_{K}]^T. \quad (6)$$

Then we have $[\mathbf{D}_c, \mathbf{D}]\mathbf{Q}_c = \mathbf{G}_c \mathbf{Q}_c = \mathbf{D}$. Therefore we obtain the objective function f:

$$\begin{aligned} \min_{\mathbf{G},\mathbf{\Omega}} \quad & \sum_{c=1}^{C=2} \|\mathbf{X}_c - \mathbf{G}_c \mathbf{\Omega}_c \mathbf{X}_c\|_F^2 + \lambda \left\|\mathbf{\Omega}_c \overline{\mathbf{X}}_c\right\|_F^2 + \\ & \alpha \sum_{j=1, j \neq c}^{J=2} \|\mathbf{G}_c \mathbf{Q}_c - \mathbf{G}_j \mathbf{Q}_j\|_F^2, \\ s.t. \quad & \|\mathbf{g}_i\|_2^2 \leqslant 1. \end{aligned} \quad (7)$$

where α is a regularization parameter, and we employ $\sum_{c=1}^{C=2} \sum_{j=1, j \neq c}^{J=2} \|\mathbf{G}_c \mathbf{Q}_c - \mathbf{G}_j \mathbf{Q}_j\|_F^2$ to make cover and stego images share the same dictionary.

III. Optimization

The objective function is hard to solve in mathematics, so we bring a variable matrix $\mathbf{A}$ in the function.

$$\begin{aligned}\min_{\mathbf{G},\mathbf{\Omega}} \quad & \sum_{c=1}^{C=2} \|\mathbf{X}_c - \mathbf{G}_c\mathbf{A}_c\|_F^2 + \tau \|\mathbf{\Omega}_c\mathbf{X}_c - \mathbf{A}_c\|_F^2 + \\ & \lambda \left\|\mathbf{\Omega}_c\overline{\mathbf{X}}_c\right\|_F^2 + \alpha \sum_{j=1,j\neq c}^{J=2} \|\mathbf{G}_c\mathbf{Q}_c - \mathbf{G}_j\mathbf{Q}_j\|_F^2, \\ s.t. \quad & \|\mathbf{g}_i\|_2^2 \leqslant 1.\end{aligned} \tag{8}$$

where τ is the coefficient of penalty term. The function can be solved, because each of the terms is constraint by Frobenius nom. We use random numbers which mean value is zero to initialize the synthesis dictionary $\mathbf{G}$ and the analysis dictionary $\mathbf{\Omega}$, and normalization processing is carried out. Then, we iteratively update $\mathbf{A}$, $\mathbf{\Omega}$ and $\mathbf{G}$.

1) To update $\mathbf{A}$ with fixed $\mathbf{G}$ and $\mathbf{\Omega}$, we require to solve:

$$\min_{\mathbf{A}} \sum_{c=1}^{C=2} \|\mathbf{X}_c - \mathbf{G}_c\mathbf{A}_c\|_F^2 + \tau \|\mathbf{\Omega}_c\mathbf{X}_c - \mathbf{A}_c\|_F^2. \tag{9}$$

Equation (9) can derive to be a least squares problem, we get the (10) by calculating the derivate.

$$\mathbf{A}_c^* = \left(\mathbf{G}_c^T\mathbf{G}_c + \tau\mathbf{I}\right)^{-1} \left(\tau\mathbf{\Omega}_c\mathbf{X}_c + \mathbf{G}_c^T\mathbf{X}_c\right). \tag{10}$$

2) To update $\mathbf{\Omega}$ with fixed $\mathbf{A}$ and $\mathbf{G}$, we require to solve:

$$\min_{\mathbf{\Omega}} \sum_{c=1}^{C=2} \tau \|\mathbf{\Omega}_c\mathbf{X}_c - \mathbf{A}_c\|_F^2 + \lambda \left\|\mathbf{\Omega}_c\overline{\mathbf{X}}_c\right\|_F^2. \tag{11}$$

We can also get the solutions by calculating the derivate.

$$\mathbf{\Omega}_c^* = \tau\mathbf{A}_c\mathbf{X}_c^T \left(\tau\mathbf{X}_c\mathbf{X}_c^T + \lambda\overline{\mathbf{X}}_c\overline{\mathbf{X}}_c^T + \tau\mathbf{I}\right)^{-1}. \tag{12}$$

3) To update $\mathbf{G}$ with fixed $\mathbf{A}$ and $\mathbf{\Omega}$, we require to solve:

$$\begin{aligned}\min_{\mathbf{G}} \quad & \sum_{c=1}^{C=2} \|\mathbf{X}_c - \mathbf{G}_c\mathbf{A}_c\|_F^2 + \alpha \sum_{j=1,j\neq c}^{J=2} \|\mathbf{G}_c\mathbf{Q}_c - \mathbf{G}_j\mathbf{Q}_j\|_F^2, \\ s.t. \quad & \|\mathbf{g}_i\|_2^2 \leqslant 1.\end{aligned} \tag{13}$$

We update class-specific dictionaries respectively by the ADMM algorithm as a result of they are affect to each other.

$$\begin{cases} \mathbf{G}^{(r+1)} = \arg\min\limits_{\mathbf{G}} \sum\limits_{c=1}^{C=2} \|\mathbf{X}_c - \mathbf{G}_c\mathbf{A}_c\|_F^2 + \alpha \sum\limits_{j=1,j\neq c}^{J=2} \\ \|\mathbf{G}_c\mathbf{Q}_c - \mathbf{G}_j\mathbf{Q}_j\|_F^2 + \rho \left\|\mathbf{G}_c - \mathbf{S}_c^{(r)} + \mathbf{T}_c^{(r)}\right\|_F^2, \\ \mathbf{S}^{(r+1)} = \arg\min\limits_{\mathbf{S}} \sum\limits_{c=1}^{C=2} \rho \left\|\mathbf{G}_c^{(r+1)} - \mathbf{S}_c^{(r)} + \mathbf{T}_c^{(r)}\right\|_F^2 \\ s.t. \|\mathbf{s}_i\|_2^2 \leqslant 1, \\ \mathbf{T}^{(r+1)} = \mathbf{G}_c^{(r+1)} - \mathbf{S}_c^{(r+1)} + \mathbf{T}_c^{(r)}. \end{cases} \tag{14}$$

In every step of the iteration, we could obtain the closed-form result about $\mathbf{A}$ or $\mathbf{\Omega}$, and the convergence speed of ADMM is fast, so our method will simplify the calculation a lot. The training time of the method is less than most dictionary learning methods.

IV. Classification

Analytic dictionary $\mathbf{\Omega}_c$ projects the signals which are not from class c into the null space, and produces larger coefficients for the signals from class c. In the meantime, the synthesis dictionary $\mathbf{G}_c$ is used to represent the signals from class c, so the residual error $\|\mathbf{X}_c - \mathbf{G}_c\mathbf{\Omega}_c\mathbf{X}_c\|_F^2$ would be tiny. In addition to this, $\mathbf{\Omega}_c\mathbf{X}_i$ will be small, $\mathbf{G}_c$ is not used to reconstruct $\mathbf{X}_i$, hence the residual error $\|\mathbf{X}_i - \mathbf{G}_c\mathbf{\Omega}_c\mathbf{X}_i\|_F^2$ is bigger than $\|\mathbf{X}_c - \mathbf{G}_c\mathbf{\Omega}_c\mathbf{X}_c\|_F^2$. When we do the test, if the query signal y belongs to class c, the generated vector coefficients by $\mathbf{\Omega}_c$ are larger than by $\mathbf{\Omega}_i$, therefore the residual error $\|y - \mathbf{G}_c\mathbf{\Omega}_c y\|_2^2$ would be smaller than $\|y - \mathbf{G}_i\mathbf{\Omega}_i y\|_2^2$. We calculate each categories' reconstruction residual to classify the query signal y. The classifier can be expressed as follow:

$$label(y) = \arg\min_i \|y - \mathbf{G}_i\mathbf{\Omega}_i y\|_2. \tag{15}$$

V. Experimental Studies

A. Experimental data

In the experiment, we employ 10,000 images from reference data set called Break Our Steganography System ver. 1.01 [16]. These images are taken by eight digital cameras and transformed into 8-bit grayscale later, then, they are resized and carved to 512×512 pixels images, at last, the images are JPEG compressed by quality factor 85.

In order to verify the effectiveness of our method, we experiment with four modern steganographic algorithms including: nsF5, Outgess, UNIWARD, MB1. Firstly, we use these steganographic algorithms to hide information into images of the database by ten embedding rates (0.1, 0.2 , ... , 1.0 bpac), and we extract the PF-274 [17] features of the cover and stego images for steganalysis. Next, we randomly choose 3,500 cover images and corresponding stego images for training, and, the rest are used to test. At last, we adopt error rate P_E acquired by receiver operating characteristic (ROC) curve to assess our method. The computational formula of P_E is:

$$P_E = \min_{P_{FA}} \frac{P_{FA} + P_{MD}(P_{FA})}{2}. \tag{16}$$

where P_{FA} is called false positive rate and P_{MD} is called missing report rate. In the meantime, we use SVM-based steganalysis method [18] for comparing. The experiments are repeated for ten times and average values are calculated. The experimental results are shown in table **I** and **II**.

From the results of table **I** and **II**, we can see our method could achieve better classification accuracy than SVM-based steganalysis method. When the steganographic algorithm is more morden and the embedding rate is lower, our method could show more advantages.

B. Experimental time

In order to compare the efficiency of our method and SVM, we conduct the experiment under the same condition: the software is MATLAB R2010a, the computer's processor is Inter(R) Core(TM) i5-3470 at 3.20 GHz, installed memory is 4.00 GB. The experimental time is exhibited in the table **III**,

TABLE I
EXPERIMENTAL RESULTS OF UNIWARD

Steganographic Algorithms	Embedding Rate	P_E by SVM	P_E by ours
UNIWARD	0.1	0.5036	0.4697
UNIWARD	0.2	0.4949	0.4378
UNIWARD	0.3	0.4130	0.4050
UNIWARD	0.4	0.3780	0.3683
UNIWARD	0.5	0.3424	0.3318
UNIWARD	0.6	0.3005	0.2946
UNIWARD	0.7	0.2531	0.2520
UNIWARD	0.8	0.2018	0.2012
UNIWARD	0.9	0.1546	0.1534
UNIWARD	1.0	0.1080	0.1079

TABLE II
EXPERIMENTAL RESULTS OF NSF5, OUTGESS AND MB1

Steganographic Algorithms	Embedding Rate	P_E by SVM	P_E by ours
nsF5	0.1	0.2694	0.2633
nsF5	0.2	0.0799	0.0778
nsF5	0.3	0.0170	0.0150
Outgess	0.1	0.0383	0.0368
Outgess	0.2	0.0069	0.0054
Outgess	0.3	0.0030	0.0019
MB1	0.1	0.0351	0.0349
MB1	0.2	0.0027	0.0016
MB1	0.3	0.0005	0.0002

we can conclude that our method needs less training time than SVM-based steganalysis method, and the testing time for each image is one-ninetieth of SVM-based steganalysis method. As a consequence, our proposed method is much more efficient in steganalysis of JPEG image.

TABLE III
THE EXPERIMENTAL TIME

Methods	Training time (s)	Testing time for each image (s)
SVM	75.16	0.00276
Ours	5.58	0.00003

VI. CONCLUSION AND FUTURE WORK

This paper has presented a novel universal JPEG image steganalysis method based on hybrid dictionary learning framework, which focuses on the fact that the stego images and the cover images are almost the same visually. We learn a class-specific dictionary for the cover and stego images respectively to encode different parts, and learn a shared dictionary to represent the common parts. Except it, our model also gets a synthesis dictionary for good representative ability and an analysis dictionary for linear projection coding. The model dose not need l_0-norm or l_1-norm constraint and makes full use of label information, the efficiency is enhanced significantly in training and testing stages. This proposed image steganalysis method can improve detection accuracy compared with the SVM-based steganalysis method.

Future researches will mainly focus on how to utilize dictionary to conduct content-mismatched steganalysis and how to make full use of the sample set to obtain a better representation. Besides, we will also investigate the combination of kernel dictionary learning methods.

REFERENCES

[1] N. Zhou, Y. Shen, J. Peng, and J. Fan.“Learning inter-related visual dictionary for object recognition,” in *Proc. IEEE Conf. Comput. Vision Pattern Recognition (CVPR)*, Providence, Rhode Island, Jun. 2012, pp. 3490-3497.

[2] V. Holub and J. Fridrich, “Low-Complexity Features for JPEG Steganalysis Using Undecimated DCT,” *IEEE Trans. Inf. Forens. Security*, vol. 10, no. 2, pp. 219-228, 2015.

[3] S. Kong and D. Wang, “A dictionary learning approach for classification: separating the particularity and the commonality,” in *Proc. European Conf. Comput. Vision (ECCV)*, Florence, Italy, Oct. 2012, pp. 186-199.

[4] D. A. Ker, P. Bas, and R. Bhme, “ Moving steganography and steganalysis from the laboratory into the real world,” in *Proc. ACM workshop on Inform. hiding and multimedia security*, 2013, pp. 45-58.

[5] V. Holub and J. Fridrich, “Phase-aware projection model for steganalysis of JPEG images,” in *IS&T/SPIE Electron. Imaging. Int. Soc. for Optics and Photonics*, 2015, pp. 94090T-94090T-11.

[6] V. Holub and J. Fridrich, “Random projections of residuals for digital image steganalysis,” *IEEE Trans. Inf. Forens. Security*, vol. 8, no. 12, pp. 1996-2006, 2013.

[7] R. Cogranne, T. Denemark, and J. Fridrich, “Theoretical model of the FLD ensemble classifier based on hypothesis testing theory,” in *IEEE Nat. Conf. Parallel Computing Tech.*, 2015, pp. 167-172.

[8] S. Gao, I. W. Tsang, and Y. Ma, “ Learning category-specific dictionary and shared dictionary for fine-grained image categorization,” *IEEE Trans. Image Process.*, vol. 23, no. 2, pp.623-634, 2014.

[9] M. Li, M. Kulhandjian, D. A. Pados, S. N. Batalama, and M. J. Medley, “ Extracting spread-spectrum hidden data from digital media,” *IEEE Trans. Inf. Forens. Security*, vol. 8, no. 7, pp. 1201-1210, Jul. 2013.

[10] H. Wang and S. Wang, “Cyber warfare: Steganography vs. steganalysis,” *ACM Commun. Mag*, vol. 47, no. 10, pp. 76-82, Oct. 2004.

[11] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210-227, Feb. 2009.

[12] Z. Zhang, D. Hu, Y. Yang, and B. Su, “A universal digital image steganalysis method based on sparse representation,” in *Proc. IEEE Int. Conf. Computational Intell. and Security (CIS)*, Leshan, China, Dec. 2013, pp. 437-441.

[13] D. Zhang, M. Yang, and X. Feng, “Sparse representation or collaborative representation: Which helps face recognition?” in *Proc. IEEE Int. Conf. Comput. Vision (ICCV)*, Barcelona, Spain, Nov. 2011, pp. 471-478.

[14] J. Guo, Y. Guo, L. Li, and M. Li, “A universal JPEG image steganalysis method based on collaborative representation,” in *Proc. IEEE Int. Conf. Security Pattern Anal. Cybern. (SPAC)*, Wuhan, China, Oct. 2014, pp. 285-289.

[15] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yan, “Sparse representation for computer vision and pattern recognition,” *Proc. IEEE*, vol. 98, no. 6, pp. 1031-1044, Jun. 2010.

[16] P. Bas, T. Filler, and T. Pevny, “Break our steganographic system: The ins and outs of organizing BOSS,” in *Proc. IEEE Int. Conf. Inform. Hiding (IH)*, Prague, Czech, Jan. 2011, pp. 59-70.

[17] T. Pevny and J. Fridrich, “Merging markov and DCT features for multi-class JPEG steganalysis,” in *Proc. SPIE, Security, Steganography, and Watermarking of Multimedia Contents Conf., SPIE Electronic Imaging*, San Jose, CA, Feb. 2007, vol. 6505, pp. 03_1-03_13.

[18] C. C. Chang and C. J. Lin, “LIBSVM: A library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1-27, May 2011, [Online]. Available: http://www.csie.ntu.edu.tw/ cjlin.

[19] S. Gu, L. Zhang, W. Zuo, X. Feng,“Projective dictionary pair learning for pattern classification” in *Advances in Neural Information Processing Systems*, Canada, Dec.2014, pp. 793-801.

[20] X. Li, X. Kong, B. Wang, Y. Guo, and X. You, “Generalized transfer component analysis for mismatched JPEG steganalysis,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Melbourne, Australia, Sep. 2013, pp. 4432-4436.

[21] I. Lubenko and A. D. Ker, “Steganalysis with mismatched covers: Do simple classifiers help?” in *Proc. ACM Conf. Multimedia and Security (MM&Sec)*, New York, NY, Sep. 2012, pp. 11-18.