Signal Processing: *Image Communication* 29 (2014) 385–399

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

A novel steganalysis framework of heterogeneous images based on GMM clustering☆

Xiaodan Hou [a,*], Tao Zhang [a], Gang Xiong [b], Zhibo Lu [a], Kai Xie [a]

[a] *Zhengzhou Information Science and Technology Institute, No. 837, P.O. Box 1001, Zhengzhou 450002, Henan, China*
[b] *National Digital Switching System Engineering & Technological R&D Center, Zhengzhou 450002, Henan, China*

ARTICLE INFO

Article history:
Received 4 May 2013
Received in revised form
4 December 2013
Accepted 16 January 2014
Available online 30 January 2014

Keywords:
Steganalysis
Steganography
Clustering
Gaussian mixture model
Texture complexity

ABSTRACT

The current steganalysis frameworks involve a large number of techniques for feature extraction and classification. However, one of their common defects is treating all images as equal, thus ignoring the variability of statistical properties of different images, which motivates us to propose a novel steganalysis framework based on Gaussian mixture model (GMM) clustering in the study, targeting at heterogeneous images with different texture complexity. There are two main improvements compared to the current steganalysis frameworks. First, in the training stage, the GMM clustering algorithm is exploited to classify the training samples into limited categories automatically, and then design corresponding steganalyzers for each category; second, in the testing stage, the posterior probability of testing samples belonging to each category is calculated, and the samples are submitted to the steganalyzers corresponding to the maximum posterior probability for test. Extensive experimental results aiming at least significant bit matching (LSBM) steganography and two adaptive steganography algorithms show that the proposed framework outperforms the steganalysis system that is directly trained on a mixed dataset, and also indicate that our framework exhibits better detection performance compared to the representative framework for using image contents in most circumstances and similar detection performance in few cases.

1. Introduction

Since the early 1990s, modern information hiding technologies have become a research hotspot in the information security field. During the past decade, the game between steganography, as an important technique for information hiding, and its backward steganalysis is growing. The purpose of digital steganography is to transform the hidden secret information in a digital cover media, such as digital audio, image, video, etc., without causing suspicion of a third part, thereby achieving covert communication. On the contrary, steganalysis exploits the difference between cover media and stego media, to detect the existence of hidden information. There are primarily two types of steganalysis methods, namely, targeted steganalysis method and general blind detection. Targeted steganalysis is designed for one or for one category of steganography algorithm. The knowledge of specific details of the steganography algorithm concerned is required. However, a blind detection method does not need the prior information about the steganography algorithm. It usually uses a pattern classification method based on machine learning. Hence, the general blind detection has become a mainstream steganalysis method. In this study, we consider digital images as covers and investigate the steganalysis of images.

General blind detection normally consists of the training stage and the testing stage of the classifiers. The core is how to extract a classification feature set with strong discernibility. Consequently, the difference between all methods mainly lies in the classification features extracted. Typical classification features include image quality measure feature [1],[2],

☆ Projects 61272490 and 60903221 supported by NSFC.

* Corresponding author. Tel.: +86 371 81630726; fax: +86 371 81630662.
E-mail address: hxd2305@163.com (X. Hou).

0923-5965/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.image.2014.01.006

statistical moment feature of probability density function (PDF) characterized by wavelet domain and high-order [3]'[4]'[5], PDF statistical moment feature of local linear transform (LLT) coefficients [6], histogram characteristic function (CF) center of mass feature [7], CF moment feature of wavelet coefficient histogram [8]'[9]'[10], empirical matrix (or co-occurrence matrix) feature [11]'[12]'[13] and multi-domain integrated feature [14]'[15]'[16].

It can be observed that the current steganalysis frameworks include considerable amount of techniques for feature extraction and classification. Striking research achievements are already made in this field. But the common defect of them is to treat all images as equal, without utilizing the statistical feature differences of images. In fact, the differences in gray scale, color, shape, texture and spatial location of images will lead to diversity and complexity of image contents. Some researchers have already noticed these issues, and attained preliminary research results, such as the blind detection framework proposed by Amirkhani et al. [17] that utilized the image content. They performed tests on JPEG image steganography algorithms. The experimental results showed that the performance was greatly enhanced of the current steganalysis system. However, such approach supposes that the content category of cover image and stego image is equal. But in fact, the content category of stego image will inevitably change to some extent as a result of embedding the secret message into a cover image. Hashemipour et al. [18] proposed a blind detection framework based on multi-classification on images. The framework utilized the multi-classification method based on image features to divide the images into limited categories. Then, the steganalyzer was designed specifically for each category. This framework was tested for JPEG image steganography algorithms F5 [19] and Outguess [20] respectively, with good detection results achieved. But the approach only utilizes the non-zero alternating current (AC) coefficient of JPEG image as the basis for multi-classification, which cannot describe the great diversity of image contents. In addition, the aforementioned two frameworks only focus on the steganalysis of JPEG images, but ignore the steganalysis of spatial domain steganography.

Current literature measures image contents on the basis of texture complexity. It is shown that the flatter the image contents, the greater the differences in statistical features of cover image and stego image, and the easier for steganalysis will be. Besides, current adaptive steganography algorithms make full use of image texture complexity by embedding secret message into areas with complex texture, which increases the security of steganography. Therefore, a steganalysis framework is proposed based on Gaussian mixture model (GMM) clustering aiming at heterogeneous images constructed by images with different texture complexity. The reason for adopting the unsupervised learning method-clustering is that the supervised learning method (e.g., Fisher linear discriminant classifier, SVM classifier) acquires the knowledge of class belonging of the training sample; however, actually the class belonging of the images with different texture complexity is not obvious and exact. Moreover, because GMM nearly simulates probability distribution in any form and contains multiple models, GMM is also used in the probability distribution for clustering. In this study, GMM clustering is applied for image multi-classification and is combined with image steganalysis.

It should be noted that all blind detection methods and some of the targeted steganalysis methods are applicable to this framework. There are two main improvements compared to the current steganalysis frameworks. First, during the training stage, one preprocessing phase is added before training the steganalyzer, i.e., using GMM clustering to classify images into limited categories automatically, and then design or choose steganalysis method corresponding to each category of image to train the steganalyzer; second, in the testing stage, the posteriori probability is calculated of a given sample under test belonging to each category, and submit it to the steganalyzer corresponding to the category with the largest posteriori probability for test. Tests are conducted on the proposed framework for least significant bit matching (LSBM) steganography in spatial domain [21] and two adaptive steganography algorithms. Extensive experimental results show that the proposed framework outperforms significantly the steganalysis system that is directly trained on a mixed dataset, and also indicate that our framework exhibits better detection performance compared to the representative framework for using image contents [17] in most circumstances and similar detection performance in few cases.

The rest of the paper is organized as follows. GMM and expectation/maximization (EM) algorithm are briefly introduced in Section 2. Section 3 elaborates on the proposed framework together with image texture feature for clustering. Additionally, the efficiency of the proposed framework is proved theoretically. In Section 4, the experimental results are given with the proposed framework on LSBM steganography in spatial domain and two adaptive steganography algorithms, respectively. Comparisons are made between the proposed framework and other steganalysis frameworks. A summary on the present work and future work is given in Section 5.

2. GMM and EM algorithm

Researches on GMM started from 1894. Day [22] conducted a research on the moment estimation, least χ^2 estimation, Bayesian estimation and the maximum likelihood estimation (MLE) of GMM. It is found that the maximum likelihood estimation is superior to the other estimations. Dempster [23] invented the EM algorithm. After that, the EM algorithm has become a main algorithm of maximum likelihood estimation on GMM, and is widely applied in unsupervised learning.

2.1. GMM and MLE

A K-order probability density function of GMM is expressed as in the following equation:

$$P(\boldsymbol{x}_j|\lambda) = \sum_{i=1}^{K} \alpha_i P_i\left(\boldsymbol{x}_j; \boldsymbol{\mu}_i; \Sigma_i\right) \tag{1}$$

where $\boldsymbol{x}_j$ is a L dimensional random vector; $\alpha_i, i = 1, 2, \ldots, K$ is the weight coefficient, and satisfies $\sum_{i=1}^{K} \alpha_i = 1$. $P_i\left(\boldsymbol{x}_j; \boldsymbol{\mu}_i; \Sigma_i\right)$ is the L dimensional joint Gaussian probability