Signal Processing 88 (2008) 2138–2157

Contents lists available at ScienceDirect

Signal Processing

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

Review

A review on blind detection for image steganography

Xiang-Yang Luo [a,b,*], Dao-Shun Wang [b], Ping Wang [a], Fen-Lin Liu [a]

[a] Institute of Information Science and Technology, 450002 Zhengzhou, PR China
[b] Department of Computer Science and Technology, Tsinghua University, 100084 Beijing, PR China

ARTICLE INFO

Article history:
Received 28 August 2007
Received in revised form
24 March 2008
Accepted 26 March 2008
Available online 1 April 2008

keywords:
Steganography
Steganalysis
Blind detection
Digital image
Review

ABSTRACT

Blind steganalysis techniques detect the existence of secret messages embedded in digital media when the steganography embedding algorithm is unknown. This paper presents a survey of blind steganalysis methods for digital images. First, a principle framework is described for image blind steganalysis, which includes four parts: image pretreatment, feature extraction, classifier selection and design, and classification. We then classify the existing blind detection methods into two categories according to the development of feature extraction and classifier design. For the first category, we survey the principles of six kinds of typical feature extraction methods, describe briefly the algorithms of features extraction of these methods, and compare the performances of some typical feature extraction algorithms by employing the Bhattacharyya distance. For the second category, the development of classifier design, we make a survey on various classification algorithms used in existing blind detection methods, and detail the algorithms behind several classifiers based on multivariate regression analysis, OC-SVM, ANN, CIS and Hyper-geometric structure. Finally, some open problems in this field are discussed, and some interesting directions that may be worth researching in the future are indicated.

Contents

* Corresponding author at: Institute of Information Science and Technology, 450002 Zhengzhou, PR China. Tel.: +86 13810496126.
E-mail address: luoxy@theory.cs.tsinghua.edu.cn (X.-Y. Luo).

0165-1684/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2008.03.016

1. Introduction

Steganography is an art of hiding communication by embedding messages into an innocuous-looking cover medium such as digital image, video, audio and so on, while steganalysis focus on revealing the presence of the secret messages and extract them. Generally speaking, if an algorithm can judge whether a given image contains a secret message or not, the steganographic system is considered broken by this algorithm [1]. Hence, the first aim of steganalysis is detecting the presence of secret messages. Usually, steganalysis methods fall broadly into one of two categories: steganalysis for specific embedding or universal blind steganalysis. The former can be called as specific steganalysis, and it can reveal secret message or even estimate the embedding ratio with the knowledge of the steganographic algorithm, just like RS [2], SPA [3], DIH [4], and LSM [5] algorithms can detect the spatial LSB steganography reliably, and the algorithms of Fridrich and Pevny et al. [6–9] can determine the presence of secret message for some steganography methods that embed message in the DCT domain of the image. But the steganalysis for specific embedding is hardly practical because it is actually difficult for steganalyzers to know what steganography method was used in images. While the latter, universal blind steganalysis, can detect the secret message independent of the embedding algorithm, it is more attractive in many practical applications. Usually, it is likely that steganalysis methods that target a specific embedding method can give more accurate and reliable results than any universal blind steganalysis. Nevertheless, universal blind approaches are very important because of their flexibility and ability to be quickly adjusted to new or completely unknown steganographic methods [1].

Universal blind steganalysis is a meta-detection method in the sense that it can be adjusted, after training on original and stego images, to detect any steganographic method regardless of the embedding domain [1]. In existing literatures on non-specific steganalysis, there are also two kinds of steganalysis methods. One detects original and stego images using original images as the training set and extracting features to classify images, without the help of features of stego images. Strictly speaking, because these methods do not depend on the condition that we have known certain hiding methods used in images, we regard them as actual blind detection methods. The other kind of steganalysis method detects original and stego images by combining the original images and stego images as the training set, where the stego images are obtained using multiple steganography methods embedding messages into original images. This kind of steganalysis method assumes that it is possible to use some special hiding methods in images, but the analyzers do not know which special hiding methods are used. Hence, we can call this kind steganalysis the "half-blind" detection method. It is worth pointing out that the classifier obtained from half-blind detection methods has a certain generalizing capability. Namely, it is possible for these classifiers to detect some new and unknown steganography methods. For example, Avcibas et al. [10] made two cross-validation experiments to show the generalizing capability of the detection algorithm. In one of the cross-validation experiment, the steganalyzer trained on images embedded by Digimarc [11], and tested on images embedded by PGS [12] and Cox et al.'s spread spectrum (SS) method [13]. In another experiment, the steganalyzer trained on images embedded with Steganos [14] and S-tools [15], and tested on images embedded with Jsteg [16] for cross-validation purposes. Results showed that the classifiers are still able to classify when the tested images come from an embedding technique unknown to the steganalyzer, which indicates that the half-blind steganalysis method has a generalizing capability of capturing the general intrinsic characteristics of steganographic techniques. Hence, the kind of half-blind steganalysis is an important part of blind detection researches. In this paper, we also regard this kind of steganalysis as blind detection methods, similar to most existing references.

Since the first blind steganalysis method was presented by Avcibas et al. [17] in 2000, ever-increasing attention has been recently paid to blind steganalysis, and various techniques have been developed to detect stego images blindly in recent years. Among them, some techniques focused on features extraction of images, such as the features extraction algorithms based on image quality metrics [10,17–19], high-order probability density function (PDF) statistics moments of the decomposition subbands coefficients [20–30], center of mass (COM) of histogram characteristic functions (HCFs) [31], statistical moments of characteristic function (CF) of subband histograms [32–37], statistics of empirical or co-occurrence matrix (Co-M) [6,38–41], and merging features of spatial and DCT domains [42,43]. Some others techniques were absorbed in the selection and design of the classifier, such as the classifier based on multivariate regression analysis [10,17,19], One-Class Support Vector Machine (OC-SVM) [23,24,30,45,46], artificial neural network (ANN) [42,43,47], computational immune system (CIS) [26,27] and hyper-geometric structure [28].