Signal Processing 100 (2014) 169–185

Contents lists available at ScienceDirect

Signal Processing

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

A local adaptive model of natural images for almost optimal detection of hidden data ☆

Rémi Cogranne [a,*], Cathel Zitzmann [b], Florent Retraint [a], Igor V. Nikiforov [a], Philippe Cornu [a], Lionel Fillatre [c]

[a] ICD - LM2S - UMR STMR CNRS - Troyes University of Technology, 10004 Troyes Cedex, France
[b] ICD - LM2S - UMR STMR CNRS - EPF École d'ingénieurs, 10000 Troyes, France
[c] I3S - University of Nice Sophia-Antipolis - UMR7271 - UNS CNRS, 06900 Sophia Antipolis, France

ARTICLE INFO

Article history:
Received 30 July 2013
Received in revised form
23 January 2014
Accepted 28 January 2014
Available online 5 February 2014

Keywords:
Image parametric model
Statistical hypothesis test
Steganalysis
Local non-linear model
Nuisance parameters
Digital forensics

ABSTRACT

This paper proposes a novel methodology to detect data hidden in the least significant bits of a natural image. The goal is twofold: first, the methodology aims at proposing a test specifically designed for natural images, to this end an original model of images is presented, and, second, the statistical properties of the designed test, probability of false alarm and power function, should be predictable.

The problem of hidden data detection is set in the framework of hypothesis testing theory. When inspected image parameters are known, the Likelihood Ratio Test (LRT) is designed and its statistical performance is analytically established. In practice, unknown image parameters have to be estimated. The proposed model of natural images is used to estimate unknown parameters accurately and to design a Generalized Likelihood Ratio Test (GLRT). Finally, the statistical properties of the proposed GLRT are analytically established which permits us, first, to guarantee a prescribed false-alarm probability and, second, to show that the GLRT is almost as powerful as the optimal LRT. Numerical results on natural image databases and comparison with prior art steganalyzers show the relevance of theoretical findings.

1. Introduction

Information hiding concerns the transmission of a secret message buried in a host digital medium. It has received increasing interest in the last decades driven by the large number of ensuing applications such as watermark-based authentication and fingerprint tracing. Unfortunately, malicious uses of data hiding have also emerged; the "prisoners problem" [43] exemplifies such a use of steganography. Alice and Bob, two prisoners, communicate by imperceptibly embedding a secret message $\boldsymbol{M}$ into a cover-object $\boldsymbol{C}$ to obtain an innocuous looking stego-object $\boldsymbol{S}$, which is then sent through a public channel. Wendy, the warden, examines all their communications in order to detect whether the inspected object, denoted $\mathbf{Z}$ as it can be a cover-object $\boldsymbol{C}$ or a stego-object $\boldsymbol{S}$, contains a secret message $\boldsymbol{M}$ or not. With many tools available in the public domain, steganography is within reach of anyone. It is thus crucial for security forces to detect

☆ This work was supported by National Agency for Research (ANR), Program (Project ANR-07-SECU-004), by the Prevention of and Fight against Crime Programme of the European Union European Commission – Directorate-General Home Affairs (2centre.eu project) and by Troyes University of Technology (UTT) strategic program COLUMBO.

* Corresponding author.
E-mail address: remi.cogranne@utt.fr (R. Cogranne).
URL: http://www.lm2s.utt.fr/fr/membres/cogrannr.html (R. Cogranne).

http://dx.doi.org/10.1016/j.sigpro.2014.01.027
0165-1684

efficiently data hidden within a (possibly very large) set of media.

1.1. State of the art

Among the wide range of steganographic schemes proposed in the past decade, the vast majority insert the (binary) message $\boldsymbol{M}$ in the Least Significant Bits (LSB) plane of the cover medium [4,13]. More precisely, two embedding functions have been widely studied: LSB replacement and LSB matching (or ± 1 embedding). The LSB replacement method consists in substituting cover medium LSB by bits of secret message. The LSB matching scheme has been proposed as an improvement on LSB replacement; when the hidden bit to be inserted does not match the LSB of cover medium sample, it is proposed to randomly increment or decrement cover sample value.

Note that many methods have been proposed to improve the embedding security, see [33,38,50] and the references therein. Because those methods rely on LSB replacement or LSB matching embedding functions, efficient and reliable detection of these two basic schemes remains at the heart of steganalysis. In the present paper it is assumed that the embedding algorithm is unknown but relies on the LSB replacement method. This method is simple, easy to implement and is used in about 70% of available steganographic software on the Internet [22]. In the literature, many different algorithms have been proposed to detect LSB replacement steganography, see [4,13] and the references therein. These methods can be roughly divided into the four following categories [4]:

1. *Hypothesis-theory-based detectors* include only a few works [12,14,17,48,51]. Although the statistical properties of those tests can be analytically established, they lack an accurate cover image model which results in overall poor detection performance.
2. *Structural detectors* aim at detecting specific modifications due to the parity structure of the LSB replacement using local pixels' correlation. While these detectors are efficient [15,20,28,32] they rely on empirical pixel correlation models and do not exploit statistical methods. Hence, their performance remains analytically unknown.
3. Similarly, *Weighted Stego-image detectors* (WS) achieve an overall good performance [19,29]. However, they rely on a local autoregressive image model, which is obviously simplistic, and their statistical properties have not yet been analytically established.
4. Lastly, *blind detectors*, which rely on machine learning, have recently been widely developed [21,30,37,45] and achieve very accurate classification. As in all applications of supervised learning, a difficult problem is choosing an appropriate features set. Moreover, the issue of establishing classification error probabilities remains open in the framework of statistical learning [39]. Besides, blind detectors are very sensitive to the cover source mismatch [2] problem; hence, one needs images from the same camera to correctly train the classifier and such data might be impossible to access.

1.2. Contributions and organization of this paper

The performance of a steganalyzer is usually only evaluated using simulation on a large database but is seldom formally established. In practice, this is a major drawback because in an operational context, the most important and difficult challenge is to guarantee a prescribed false-alarm probability without which the inspection of a large set of images is very difficult. This is the main goal of this paper and it can only be achieved using an accurate image model for an accurate assessment of the proposed detector properties. In fact, the more accurate the image is, the less modeling error will impact the ensuing test and the more reliable the detector will be. However, in the field of steganalysis the use accurate models reflecting the statistical properties of natural images have not yet been proposed. The few prior works that apply hypothesis testing theory for establishing the statistical performance of detectors use rather simplistic image models. More precisely, the first works that cast the problem of steganalysis within the framework of hypothesis theory and established the statistical properties of the proposed test are based on the simplistic assumption that all the pixels share the same expectation and the same variance [14]. Our previous works [6,7,17,48] proposed to use a model in which pixels have different expectations, but this expectation is modeled with a simplistic piecewise-polynomial model. With the same model some prior works [10,12,52] study analytically the impact of quantization of detector performances. However, a piecewise-polynomial model is obviously simplistic and, hence, even though for most images the detectors perform well, for a significant proportion of images the obtained results do not match with the established ones.

It should be highlighted that dealing with such complex objects as images is difficult because (1) the statistical properties of pixels change within an image and (2) the unknown expectation of pixels acts as a nuisance parameter which prevents the construction of reliable detectors and (3) the hidden information is embedded in pixels LSB, hence, the modifications it is aimed at detecting are very small.

To address these difficulties, a novel methodology is proposed in the present paper. This methodology essentially relies on hypothesis testing theory and uses an accurate and original model adapted to natural images that allows, in practice, a precise assessment of the proposed detector properties, see Fig. 1. The main contributions of the proposed methodology are the following:

1. A local adaptive model of expectation of pixels, that is adapted according to the local content of each area of an image, is proposed to take into account the specificity of natural images, such as textures, edges and non-stationary blur (which usually changes over the whole