Signal Processing 119 (2016) 102–114

Contents lists available at ScienceDirect

Signal Processing

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

A modular framework for color image watermarking

Marco Botta [a], Davide Cavagnino [a,*], Victor Pomponiu [b]

[a] Dipartimento di Informatica, Università degli Studi di Torino, Corso Svizzera 185, 10149 Torino, Italy
[b] Information Systems Technology and Design, Singapore University of Technology and Design, Singapore 138682, Singapore

ARTICLE INFO

Article history:
Received 24 April 2015
Received in revised form
19 June 2015
Accepted 24 July 2015
Available online 5 August 2015

Keywords:
Color image
Compressed image
Information hiding
Fragile watermarking
Genetic Algorithms
Karhunen–Loève transform

ABSTRACT

We present an algorithm for fragile watermarking of color, or multi-channel, images either in uncompressed format, in lossless compressed format, or in compressed format with locally compressed units (like JPEG). The watermark is embedded into the Karhunen–Loève transform (KLT) coefficients of the host image, and the security of the method is based on the secrecy of the KLT basis computed from a corresponding secret key image. The watermark bits may be embedded with various methods, in particular the use of syndrome coding has proven flexible and effective. Genetic Algorithms (GAs) are used to find the optimum pixel modification leading to the watermark embedding. The resulting watermarked images have shown to have both a high objective quality (in terms of PSNR and SSIM) and a high subjective quality (tested by a group of observers). At the same time, the watermark extraction is very sensitive to small (± 1 intensity level for single pixel) modifications, ensuring image authentication with very high probability.

1. Introduction

Different types of data, like images, videos and sounds are evermore widely distributed thanks to computing power and network capacity. At the same time new malware may cancel the security and integrity of these media. For example, unauthorized copy, forgery and tampering are possible attacks to the digital objects.

To contrast the possible risks many techniques have been developed in the field of computer security: for example, digital signatures guarantee integrity and authenticity; message authentication codes may also be used when non-repudiation is not required.

Digital watermarking defines a broad family of methods having in common the characteristic of embedding a signal into a digital object. Depending on the requirements of the application at hand, this signal may be designed to be altered at the minimum modification of the object or, conversely, to resist modifications aimed at its removal.

It is common to give a high level description of watermarking techniques required by an application by defining two phases: the embedding phase and the verification phase, separated by a transmission, as shown in Fig. 1; in this context, transmission possibly means sending data over a communication channel and/or storing data on a media. In the embedding phase a function is used to modify the host signal S (or some of its features, like linear transform coefficients) with the objective of hiding a watermark W: typically, to improve the security of the scheme a secret key K is used to control how W is stored in S. The output of the embedder is a watermarked signal S^w. During the transmission, S^w may be subject to modifications η due, for example, to noise, compression and/or filtering. At the receiving end, the application for which the method is developed may require one or both of two possible results from the watermark verification phase (note that these results may be obtained sequentially,

* Corresponding author.
E-mail addresses: marco.botta@unito.it (M. Botta), davide.cavagnino@unito.it (D. Cavagnino), victor.pomponiu@ieee.org, victor.pomponiu@gmail.com (V. Pomponiu).

http://dx.doi.org/10.1016/j.sigpro.2015.07.018
0165-1684/

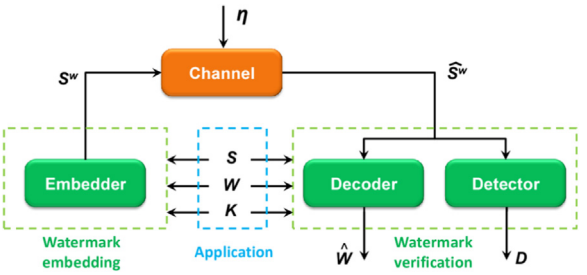

Fig. 1. The general representation of the watermarking from a communication point of view: s denotes the host signal, W is the watermark signal, K is the secret key, S^w is the watermarked signal, η is the generic modification, $\hat{S}^w$ is the possibly perturbed signal, $\hat{W}$ is the extracted watermark and D is the detector decision.

independently or at the same time according to the specific watermarking algorithm): (1) an estimation $\hat{W}$ of the watermark by a decoder, and/or (2) a Boolean decision D whether or not W is correctly present in the possibly modified signal $\hat{S}^w$. The verification phase always requires the key K and the signal $\hat{S}^w$, but some methods may also need the watermark W and/or the original signal S.

The application domain defines the characteristics the watermark must satisfy. Watermarks may be *robust* or *fragile*. A robust watermark is designed to be detectable even if the digital object containing it is (maliciously) modified. Conversely, a fragile watermark is designed to be altered at the minimal modification of the object. Typical applications are copyright protection for robust watermarking and content authentication for fragile watermarking.

Watermarking algorithms that need the host object (i.e. the original un-watermarked object) when verifying the presence of the watermark are called *non-blind* (or *informed*) whilst methods that only use the watermarked object are called *blind*.

Digital objects may be represented in different domains: for example, a sound may be represented as samples in the time domain, or in the frequency domain obtained through Fourier analysis. As a direct consequence of this, a watermark may be embedded considering one (or more) of the possible representations of the object. In the case of images, watermarks are typically inserted in the *spatial domain*, i.e. the pixels of the image, or in a *frequency domain*, like the Fourier Transform domain or the Discrete Cosine Transform domain. Also, other domains have been exploited for images, like the *fractal* and the Singular Value Decomposition (SVD) domains.

A further characteristic of watermarking algorithms is the ability to restore the host image from the watermarked image: this is called reversibility and the algorithm possessing it is called *reversible*. This feature is typically required by algorithms applied in the medical field.

The present paper presents a non-reversible algorithm for the fragile blind watermarking of color images, both for lossless compressed images and for lossy compressed (at block level) images: in particular we demonstrate the application on the JPEG compressed format. Due to the modular structure of the proposed algorithm, and its composition of basic units, we call it Multichannel Image Modular Integrity Checker (MIMIC). Its main characteristics are

- *detection* of image modifications: the watermark is highly sensitive to even small changes (one intensity level in one color channel) of the image;
- *localization*: the modification is identified at block level (i.e. a group of contiguous pixels);
- *invisibility*: the watermark signal is imperceptible for general applications of the image, in particular it is invisible to humans, due to the high PSNR (>55 dB) of the resulting images. Thus the proposed algorithm is tailored to many application scenarios; but due to the non-reversibility, the host image cannot be restored, so this algorithm cannot be used when the original image is required, like in all medical applications;
- *security*: due to the methodology and the secret key, the watermark is *secure* against various possible attacks; in particular, we will show that it is able to detect transplantation, birthday and cut-and-paste attacks.

The algorithm introduces the following new features with respect to previously published fragile watermarking algorithms:

- it extends our previous algorithm for gray-scale images to color images;
- it may use syndrome coding for embedding the watermark bits, in order to increase the objective quality and to ease the embedding effort;
- it proposes a modular architecture that increases flexibility in improving the functionalities and allows for application customization of the performance;
- it improves the objective quality of the watermarked images w.r.t. other fragile watermarking algorithms.

The following sections will present a number of scientific works related to watermarking that we consider pertinent to our algorithm (Section 2) and the set of mathematical tools we applied in our algorithm (Section 3). The core of our work is the algorithm developed that we detail in Section 4 and whose performance is shown in Section 5 where many experimental results over a large set of (publicly available) images are presented. Our conclusions are drawn in the Section 6 where we also discuss the obtained results. Supplementary Appendix contains details about the mathematical tools we used.

2. Related works

In general, watermarking schemes devised for gray-scale images can be extended and adapted to color images although several approaches have used the color information as an independent feature within the watermarking process in order to achieve an imperceptible watermark.

Depending on how the host image is perturbed during the embedding phase, state-of-the-art methods which are used for color image watermarking can be categorized into