Signal Processing 94 (2014) 118–127

Contents lists available at ScienceDirect

Signal Processing

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

Reversibility improved data hiding in encrypted images ☆

Weiming Zhang *, Kede Ma, Nenghai Yu

School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China

ARTICLE INFO

Article history:
Received 16 November 2012
Received in revised form
20 June 2013
Accepted 22 June 2013
Available online 29 June 2013

Keywords:
Reversible data hiding
Image encryption
Privacy protection
Histogram shift

ABSTRACT

A novel reversible data hiding technique in encrypted images is presented in this paper. Instead of embedding data in encrypted images directly, some pixels are estimated before encryption so that additional data can be embedded in the estimating errors. A benchmark encryption algorithm (e.g. AES) is applied to the rest pixels of the image and a special encryption scheme is designed to encrypt the estimating errors. Without the encryption key, one cannot get access to the original image. However, provided with the data hiding key only, he can embed in or extract from the encrypted image additional data without knowledge about the original image. Moreover, the data extraction and image recovery are free of errors for all images. Experiments demonstrate the feasibility and efficiency of the proposed method, especially in aspect of embedding rate versus Peak Signal-to-Noise Ratio (PSNR).

1. Introduction

Reversible data hiding (RDH) has the capability to erase the distortion introduced by embedding step after cover restoration. It is an important property that can be applied to many scenarios, such as medical imagery, military imagery and law forensics. For this reason, RDH becomes a hot research topic and is extensively studied over the years.

Until now, many RDH techniques have been proposed based on three fundamental strategies: lossless compression-appending scheme [1,2], difference expansion (DE) [3,4] and histogram shift (HS) [5]. Some recent arts combined the three strategies to residuals of the image such as prediction errors (PE) [6–9] to achieve better performance. Almost all state-of-the-art RDH algorithms consist of two steps. The first step generates a host sequence with small entropy, i.e., the host has a sharp histogram which usually can be realized by using PE combined with the sorting technique [10] or pixel selection [11]. The second step reversibly embeds the message in the host sequence by modifying its histogram with methods like HS and DE. Optimal coding methods for modifying the histogram were proposed by Zhang et al. [12] and Lin and Chung [13]. On the other hand, some robust RDH methods have also been proposed [14–17].

Nowadays with the increasing demand of privacy protection, the ability to embed information in encrypted data will be useful in cloud computing [18]. In [19], the content owner encrypts the signs of host discrete cosine transform (DCT) coefficients. Different fingerprints are generated at the receiver side by decrypting only a subset of the coefficients with different decryption keys. In [20], during H.264/AVC compression, the intra-prediction mode, motion vector differences and DCT coefficients' signs are encrypted, while watermarking process proceeds on the DCT coefficients' amplitudes adaptively. In [21], a commutative watermarking and encryption system is presented based on a layered scheme and a key dependent transform domain. However, the data embedding is not reversible with the above-mentioned techniques [19–21]. As can be seen, there are some promising applications if RDH can be

☆ This work was supported in part by the Natural Science Foundation of China under Grants 61170234 and 60803155, by the Strategic and Piloted Project of CAS under Grant XDA06030601, and by the Funding of Science and Technology on Information Assurance Laboratory under Grant KJ-13-02.

* Corresponding author. Tel.: +86 551 3600683.
E-mail addresses: weimingzhang@yahoo.cn, zwmshu@gmail.com (W. Zhang), k29ma@uwaterloo.ca (K. Ma), ynh@ustc.edu.cn (N. Yu).

0165-1684/$ - see front matter Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2013.06.023

adopted in encrypted images. For instance, by such technique, we can embed notations into an encrypted medical image. With the notation, a server can manage the image or verify its integrity without knowing the image content, and thus the privacy of the patient is protected. On the other hand, a doctor, having the key, can decrypt and restore the image losslessly.

Since the entropy of encrypted images has been maximized and leaves no spare space for traditional RDH methods to exploit, the embedding step may not be possible by using standard RDH algorithms. Even so, some attempts have been made to accommodate additional data reversibly in encrypted images. In [22], Zhang divided the encrypted image into several blocks. By flipping 3 Least Significant Bits (LSBs) of a group of specific pixels, one bit message can be embedded into each block. Hong et al. [23] improved Zhang's method [22] by further exploiting the spatial correlation using a different estimation equation and side match technique. In order to extract data, the two methods rely on decrypted images which may be unknown for some cases. Aiming for separating data extraction from image decryption, Zhang [24] found the syndromes of a low-density parity check matrix to compress the LSBs of the encrypted image. By doing so, an extra space is created to append additional data. These techniques can only achieve low embedding capacity (achievable largest embedding rate) [22,23] or generate marked image with poor quality for high embedding capacity [24] and all of them are subject to some errors on data extraction and/or image restoration. Although the methods in [22,23] can eliminate errors by error-correcting codes, the pure embedding capacity will be further consumed. On the other hand, to locate and modify the LSBs of the encrypted image, these methods must depend on a special encryption scheme, that is, encrypting the bit planes of pixels by a stream cipher. However, most current popular image encryption techniques encrypt pixels as integers with block ciphers.

This paper proposes a novel RDH method in encrypted spatial images based on estimation technique. A large portion of pixels are utilized to estimate the rest before encryption, and then encrypted with a standard encryption algorithm. After that we encrypt the estimating errors with a special encryption scheme. By concatenating encrypted estimating errors and the large group of encrypted pixels, the ultimate version of encrypted image is formulated. The additional data can be embedded in the encrypted image by modifying the estimating errors. In general, the excellent performance can be achieved in three different prospects:

- The proposed method is completely reversible. That is, no error happens in data extraction and image recovery steps.
- The PSNR values of marked decrypted image are much higher than those previous methods [22–24] can achieve under given embedding rates.
- The extraction and decryption steps are independent, which are more natural and applicable.

The rest of this paper is organized as follows. The scheme of the proposed method is elaborated in Section 2. Abundant experimental results with the analysis of complete reversibility are presented in Section 3. We conclude our paper with a discussion in Section 4.

2. Proposed method

To enable data embedding in the encrypted image, all the three methods [22–24] try to space out room from the encrypted image directly, which follows the idea of compressing the encrypted image [25,26]. However, losslessly vacating room from the encrypted image is difficult and thus these techniques cannot achieve good image quality or realize complete reversibility. In this section, we propose a novel method to significantly improve the performance by reversing the order of encryption and vacating room. In the light of this idea, we empty out room prior to image encryption by shifting the histogram of estimating errors of some pixels and the emptied out room will be used for data hiding.

The proposed method is composed of four primary steps: vacating room and encrypting image, data hiding in the encrypted image, data extraction and image recovery. Two different schemes, extraction before decryption and decryption before extraction, are raised to cope with different applications. Fig. 1 illustrates the overview of the proposed method.

2.1. Vacating room and encrypting image

Without loss of generality, assume that the original $M \times N$ image $\mathbf{X}$ is an 8-bit gray-scale image, with pixel

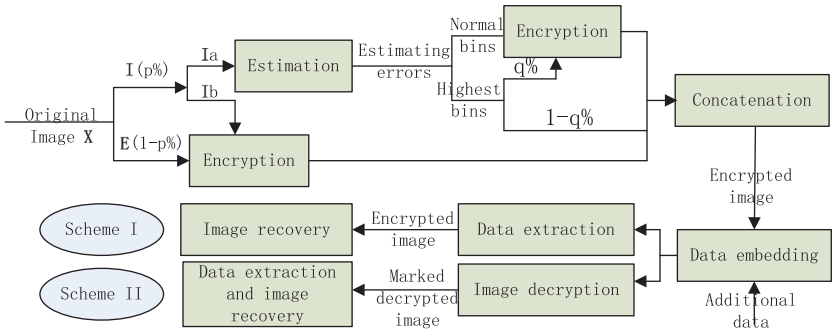

Fig. 1. The framework of proposed method.