Signal Processing 125 (2016) 315–328

Contents lists available at ScienceDirect

Signal Processing

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

Reversible data hiding by adaptive group modification on histogram of prediction errors

Reza Moradi Rad [a], KokSheik Wong [b,*], Jing-Ming Guo [c]

[a] School of Engineering Science, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada V5A 1S6
[b] Faculty of Computer Science & Information Technology, University of Malaya, Kuala Lumpur, Malaysia
[c] Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 21 August 2015
Received in revised form
27 January 2016
Accepted 1 February 2016
Available online 13 February 2016

Keywords:
Adaptive group modification
AGM
Histogram shifting
Reversible
Data hiding
Prediction error

ABSTRACT

In this work, the conventional histogram shifting (HS) based reversible data hiding (RDH) methods are first analyzed and discussed. Then, a novel HS based RDH method is put forward by using the proposed Adaptive Group Modification (AGM) on the histogram of prediction errors. Specifically, in the proposed AGM method, multiple bins are vacated based on their magnitudes and frequencies of occurrences by employing an adaptive strategy. The design goals are to maximize hiding elements while minimizing shifting and modification elements to maintain image high quality by giving priority to the histogram bins utilized for hiding. Furthermore, instead of hiding only one bit at a time, the payload is decomposed into segments and each segment is hidden by modifying a triplet of prediction errors to suppress distortion. Experimental results show that the proposed AGM technique outperforms the current state-of-the-art HS based RDH methods. As a representative result, the proposed method achieves an improvement of 4.30 dB in terms of PSNR when 105,000 bits are hidden into the test Lenna image.

1. Introduction

Data hiding is the art and science of exploiting a content (e.g., image, audio, and video) to accommodate payload. The hidden information can be used for various purposes, including identity authentication, content protection, and covert communication [1–7]. For many data hiding methods, the distortion introduced to the host content due to hiding the external/extra information (hereinafter referred to as payload) is permanent and hence the original content cannot be completely recovered from the processed content [1]. However, some applications deal with sensitive and crucial information, including medical image, military image, picture of crime scene and digital archive of rare artwork, where they cannot tolerate any distortion. As such, reversible data hiding (RDH) is proposed to enable perfect recovery of the original content directly from its processed counterpart.

In the last decade, numerous RDH techniques are proposed for image. In particular, the techniques in the spatial domain can be coarsely divided into three frameworks, namely, compress-and-append, expansion based (EB), and histogram shifting based (HS). The earliest RDH methods are based on the compress-and-append framework [8–10], which exploits the losslessly compressible property of the host image to vacate room for data hiding. On the other hand, the EB framework creates some features that are representable by small magnitudes using a decorrelation function [11]. The first method of this kind was proposed by Tian [12], where the difference (in terms of value) between two adjacent pixels is expanded (i.e., multiplied by 2). This forces all the least significant bits (LSBs) of the differences to be zero, and hence the LSBs of these

* Corresponding author. Tel.: +603 7967 6417; fax: +603 7957 9249.
E-mail addresses: rmoradir@sfu.com (R.M. Rad), koksheik@um.edu.my (K. Wong), jmguo@seed.net.tw (J.-M. Guo).

http://dx.doi.org/10.1016/j.sigpro.2016.02.001
0165-1684/

differences can be utilized to host payload. Another representative method following this framework is proposed by Thodi et al. [13], which is based on prediction error expansion (PEE). However, the severe distortion caused by EB RDH methods motivated researchers to find alternative solutions to minimize the distortion [14]. To this end, Ni et al. [15] pioneered the HS based RDH method in 2006. After generating the histogram of pixel intensity values, they shift the bins between the zero and peak points to create empty bins for data hiding. Although HS achieves better visual quality than EB, its hiding capacity is usually lower than that of EB, especially when it is compared to PEE-based method [14]. Furthermore, the performance of HS based technique is highly dependent on the occurrence of the most frequent pixel value and the local characteristics of the host image, viz., zero and peak points.

In this work, an efficient HS based RDH method using histogram of prediction error is proposed to simultaneously improve hiding capacity and image quality. By adaptively shifting histogram of prediction errors, only selected bins are vacated to host the payload. Our design goal is to minimize the number of shiftings for smaller prediction errors, which are of significantly higher occurrences, and vice versa. The proposed method also aims to avoid utilizing bins of higher frequencies except when it is necessary, such as handling payload of larger size. The proposed method is able to hide a larger number of bits while causing less visual distortion when compared to the conventional HS based RDH methods. This work makes the following contributions: (1) performing detailed analysis on the conventional HS based RDH methods and summarizing their development trends; (2) prioritizing prediction error bins by adaptive shifting and hiding to suppress distortion; and (3) modifying prediction errors in group to embed multiple bits (variable length, up to 5 bits) at a time for improving hiding efficiency.

The rest of this paper is structured as follows: Section 2 surveys the conventional HS based RDH methods and details their development trends. The proposed method is presented in Section 3. An analysis of the proposed HS based RDH method is performed in Section 4. Experimental results are presented in Section 5 and Section 6 concludes this paper.

2. Related work and observed trends

Ni et al. [15] pioneered the idea of HS to achieve RDH. Their method can be decomposed into three main steps: (1) identify the peak and zero bins in the histogram of image pixel values; (2) make room for data hiding by shifting the bins between the zero and peak points; and (3) hide information by modifying the pixels assuming the peak value. High output image quality is generally achieved in HS based RDH method because any modified pixel value differs from its original counterpart by at most unity, resulting in a mean square error (MSE) of at most unity for each modified or shifted pixel. Nonetheless, the performance of HS method is highly dependent on the local characteristics of the host image, viz., the number of occurrences of the most frequent pixel value. HS is also of high computational complexity due to the unavoidable pre-processing, such as performing $x \leftarrow x+1$ for all pixel values $x > \tau$ for some threshold τ (usually the peak or zero point).

On the other hand, Hong et al. [11] proposed a HS based RDH method by exploiting prediction error. In particular, the Median Edge Detection (MED) predictor in JPEG-LS [16,17] is utilized to predict the pixel intensity value. Furthermore, bins associated with the two peak error values (e.g., 0 and -1) are exploited to hide the payload. Li et al. [18] then proposed a reversible watermarking method based on adaptive prediction error expansion. Two strategies, namely, adaptive hiding and pixel selection, are employed in their work. Although Li et al. [18] achieved high hiding capacity, quality of the processed image is degraded due to the overutilization of smooth areas for hiding a large payload. In a recent work, Kuo et al. proposed a diamond encoding method exploiting the modification direction to improve hiding capacity [19]. This method is able to hide $(2k^2+2k+1)-ary$ data by modifying a pair of pixels in the host image.

Over the years, many HS based RDH methods are proposed to enhance the performance of the basic technique [15]. The main effort has been channeled to identify new feature and exploit it for data hiding by modifying the histogram of this newly identified feature [11], as well as improving the histogram manipulation mechanism. In the following subsections, these efforts for improvement are categorized and detailed in two general parts of HS based RDH, namely, histogram generation and histogram manipulation.

2.1. Histogram generation

2.1.1. Identifying features with sharper distribution

Identifying features with sharper distribution results in higher hiding capacity. In particular, since the (most) frequently occurring features are conventionally utilized to host the payload, these features will have higher counts in a sharper distribution, hence more information can be hidden. This idea has been explored in [3–7,18,20–31]. These techniques can be further divided into three categories. The first category employs a pixel predictor to generate prediction errors and uses them as the host feature [4–7,18,21,24,25,27–31]. The second category exploits the difference values among neighboring pixels following a specific scanning order to maximize the decorrelation process [23,32]. The third category utilizes the differences among corresponding pixels in the sub-sampled image [20,3].

2.1.2. Increasing the total number of features

Increasing the total number of features (i.e., data carriers) leads directly to higher hiding capacity. In other words, with the availability of larger number of features, more features will potentially satisfy the requirement to be exploited as the venue for data hiding. In particular, features of the same type can be simply explored in other directions/dimensions as proposed in [32–34] to generate more elements in the histogram.