Information Sciences 358–359 (2016) 164–190

Contents lists available at ScienceDirect

Information Sciences

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

A novel general multiple-base data embedding algorithm

Wei-Sung Chen[a,1], Yi-Kai Liao[a,1], Yun-Te Lin[a,b], Chung-Ming Wang[a,*]

[a] Department of Computer Science and Engineering, National Chung Hsing University, 145 Xingda Road, South District, 40227, Taichung, Taiwan

[b] NAR Labs, Taichung Branch, National Center for High-Performance Computing, No. 22, Keyuan Road, Central Taiwan Science Park, 40763, Taichung, Taiwan

ARTICLE INFO

Article history:
Received 9 October 2015
Revised 31 December 2015
Accepted 28 March 2016
Available online 7 April 2016

JEL classification:
C61: optimization techniques
Programming models
Dynamic analysis

Keywords:
Multiple-base
Data embedding
Payload
Image quality
Optimal base vector
Prediction

ABSTRACT

This paper presents a general multiple-base (GMB) data embedding algorithm to conceal a serial secret bit stream equivalent to an M-ary secret digit in a pixel-cluster consisting of n pixels, where M is automatically determined by the initial input (n, F) given by the end user. Through the change of two parameters, n and M, the proposed algorithm offers a multiple-purpose message embedding style to produce a high quality embedded image or provide a large embedding payload. Inspired by a single base (SB) data embedding approach, this study first introduces a multiple-base (MB) scheme which adopts an n-tuple optimal base vector (*OBV*) to conceal a secret M-ary digit with minimal pixel distortion, where M is the product of all vector components in the *OBV*. This study extends the MB scheme to develop the GMB algorithm, which supports a serial secret bit stream as a secret message. Four binary to M-ary conversion schemes are introduced, allowing the GMB algorithm to carry an extra secret bit per pixel-cluster, offering a larger payload without increasing the pixel distortion caused by data embedding. The proposed algorithm is analyzed, and mathematical expressions are derived so that prior to a real message embedding, it is possible to predict the expected payloads and the corresponding image quality. Finally, we extend the GMB algorithm to support content-adaptive data embedding. To the best of the authors' knowledge, the proposed algorithm is the first multiple-purpose data embedding technique, providing greater flexibility and offering large payloads or high image quality. Experimental results demonstrate that the proposed scheme outperforms current state-of-the-art competitors.

1. Introduction

Digital image data embedding utilizes images as a venue to host external information. Applications of image data hiding include, among others, confidential communication, security protection, authentication, secret data storing and content annotation [7]. In general, image data embedding algorithms in the spatial domain content can be classified into two pursuit disciplines which concern the image quality or the embedding capacity, the payloads. The first discipline aims to generate a resultant image that conceals secret messages with high image quality. Algorithms in the first discipline, such as the exploiting modification direction scheme [34], minimize per pixel modification imposed on the carrier to carry more secret bits. The produced image demonstrates a perceptually high quality that is visually similar to the host image.

* Corresponding author: Tel.: +886422840497x915, fax: +886422853869.
E-mail addresses: wschen213@gmail.com (W.-S. Chen), narumi1001@hotmail.com (Y.-K. Liao), lsi@cs.nchu.edu.tw (Y.-T. Lin), cmwang@cs.nchu.edu.tw (C.-M. Wang).
[1] Tel.: +886-4-22840497 ext 904, fax: +886-4-22853869.

http://dx.doi.org/10.1016/j.ins.2016.03.045
0020-0255/

On the other hand, the second discipline aims to increase the quantity of messages conveyed in the carrier. Algorithms in the second discipline, such as the optimal pixel adjustment process (OPAP) [1], the pixel-value differencing scheme [29], and the notational system method [33], embed more messages by delicately altering the pixel values. Offering a large payload, the resultant image demonstrates less perceptually high quality but is visually bearable to some extent.

While many image data embedding algorithms have been proposed to date, they were designed to perform a single-purpose, pursuing either high image quality or a large payload, but not both. An image data embedding algorithm that satisfies a multiple-purpose style provides significant benefits; it reduces the complexity of the algorithm, shortens the developing time, and simplifies the implementation intricacy. Real applications will certainly benefit from these advantages when developing multi-purpose data embedding algorithms.

There are four critical issues that must be carefully considered when designing such a multi-purpose data embedding algorithm. The first issue concerns the number of pixels used for data concealment. The algorithm cannot impose any limitation for the number of pixels employed in a multi-purpose algorithm. This means that the embedding methods are not favorable if they are restricted to a single pixel [1], two pixels [5,11,16,21,23,27,30], three pixels [12,13], or four pixels [15]. Instead, a preferable algorithm is capable of employing an adjustable number of pixels because it not only provides greater flexibility but also increases the level of security, offering greater protection against malicious attackers attempting to disclose secret messages.

The second issue deals with the notational system adopted for data concealment. Intuitively, the binary system seems to be a preferable choice. However, the binary system can only be mapped into specific notational systems; for example, a binary system representing 3 bits or 4 bits can be mapped respectively into 8-ary or 16-ary notational systems. Since a single bit is the minimal unit in a binary system, it is not possible, for example, to take advantage of the 9–15 notational systems, even though they are available and may produce a less distortion for message concealment. Consequently, an algorithm considering an arbitrary M-ary notational system is a preferred choice since it is possible to take advantage of all notational systems, providing a variety of payloads to extend the coverage of real applications. Although some algorithms seem to adopt a variety of notational systems, they are dependent on the number of pixels chosen. For example, the EMD algorithm adopts n pixels leading to employing the $(2n+1)$-ary notation systems; the APPM algorithm [10] adopts the M-ary notational system ($2 \leq M \leq 64$), but it is restricted to supporting a pixel pair only; the diamond encoding method [3] adopts the L-ary notation system, where $L = 2k^2+2k+1$ ($k \geq 1$). However, it is also restricted to a pair of pixels. In the authors' opinion, a multiple-purpose algorithm should be flexible enough to encourage the independence between the number of pixels to be employed (n) and the M-ary notational system that is to be adopted.

The third issue concerns the type of secret messages. A preferable type is a serial bit stream rather than an M-ary secret digit since most digital media are representable by a serial bit stream. Some algorithms, for example the diamond encoding method and the APPM algorithm, adopt the M-ary notational system on a pixel pair and reveal a payload of $(\log_2 M)/2$ bits, which is the optimal payload. However, when M is not a power of 2 and the secret messages are represented as a serial bit stream, the payload reported is overestimated, thus producing an inaccurate payload which is usually reported in terms of bits per pixel (bpp). Surprisingly, most algorithms adopting non-binary notational system usually ignore this issue of overestimate despite the fact that a transformation loss may occur between a binary system and a non-binary notational system. In the authors' opinion, in regard to supporting messages represented by a serial bit stream, it is necessary to develop an effective conversion scheme encouraging a seamless transformation between a binary and M-ary notational system with the minimal loss of conversion.

The final issue to be discussed involves predictability. Most data embedding schemes pay little attention to forecasting their results prior to real message embedding. A multi-purpose algorithm providing predictability demonstrates manifold benefits. It can validate the correctness of experimental outcomes, disclose any possible experimental bias, reveal any distortion encountered, and facilitate the substitution of desirable host images. With an accurate predictability, it becomes feasible to aid the real experimental results allowing users to make good decisions for accomplishing their goals and objectives.

This paper presents a general multiple-base data embedding algorithm with full consideration of the above four critical issues. The proposed algorithm, referred to as GMB, adopts an adjustable number of pixels (n) which can be as small as 2 pixels to as large as 13 pixels in our experiments. In addition, the GMB algorithm makes full use of an M-ary notational system, which can be as small as 13 to as large as 67,765,824 in our experiments. Unlike previous approaches, the proposed M-ary notational system is independent of the number of pixels employed. The independence of n and M allows the proposed algorithm to demonstrate a multiple-purpose style, which can produce a resultant image with high image quality or offer a large payload for message concealment. In order to conceal a serial bit stream, this study recommends four binary to M-ary conversion schemes which effectively convert a binary message into an M-ary digit with minimal loss of conversion. In addition, a number of mathematical expressions are developed making it possible to precisely foresee the expected payload and the corresponding quality of the resultant image without conducting a real message embedding. Finally, we extend our GMB algorithm to support image content-adaptive data embedding which conceals secret messages based on the content complexity of the host image. Experimental results are compared with 12 current state-of-the data embedding works. The comparison reveals that our GMB algorithm outperforms current state-of-the-art competitors.

This paper is organized as follows. Related works are reviewed in Section 2. The single base data embedding approach (SB) is introduced in Section 3. In Sections 4 and 5, the multiple-base (MB) data embedding scheme and the proposed GMB algorithms are described, respectively. A comprehensive example is presented to illustrate the message embedding