Information Sciences 179 (2009) 3332–3348

Contents lists available at ScienceDirect

Information Sciences

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

A path optional lossless data hiding scheme based on VQ joint neighboring coding

Jun-Xiang Wang [a], Zhe-Ming Lu [a,b,*]

[a] School of Information Science and Technology, Sun Yat-Sen University, XinGangXi Road 135, Haizhu District, Guangzhou 510275, PR China
[b] Institute of Astronautic Electronic Engineering, School of Aeronautics and Astronautics, Zhejiang University, Hangzhou 310027, PR China

ARTICLE INFO

Article history:
Received 26 November 2008
Received in revised form 7 May 2009
Accepted 30 May 2009

Keywords:
Information hiding
Reversible data hiding
Lossless steganography
Joint neighboring coding

ABSTRACT

Data hiding is an important technique for covert communication that embeds secret data into a cover image with minimal perceptible degradation. Lossless data hiding is a special type of data hiding technique that guarantees not only the secret data but the cover media can be reconstructed without any distortion. In this paper, a novel path optional lossless data hiding scheme based on the joint neighboring coding (JNC) of the vector quantization (VQ) index table is proposed. The proposed scheme generates a VQ index table based on the cover image first. Next, according to an initial key and secret data content, different adjacent indices may be chosen to perform joint neighboring coding for each index and hide secret data. Finally, an appropriate output codestream is generated based on the minimal length principle. Our main contributions lie in three aspects: (1) the method combines the novel path_based shift method with the traditional JNC method to improve the capacity and stego image quality simultaneously. (2) The secret data extraction and cover image recovering processes are separated and both can be publicized to users. (3) The method is path optional to meet various users' requirements. To testify the superiority of the proposed method, we compare it with the side match VQ (SMVQ)-based and modified fast correlation VQ (MFCVQ)-based algorithms. According to the experimental results, the proposed scheme outperforms the SMVQ-based and MFCVQ-based algorithms in four aspects, i.e., hiding capacity, stego image quality, transmission efficiency and security level.

1. Introduction

With the rapid development of computer sciences, the Internet appears and allows more and more users to communicate and exchange information at any time and any location. Therefore, a large amount of digital information, such as digital images, video, audio and so on, is transmitted over the Internet every day. Nowadays, the Internet is an open channel, and thus security problems such as interception, modification and others have become more and more serious. To solve these problems, researchers have proposed various approaches such as encryption, steganography and data hiding techniques. The encryption transforms important data into an unrecognizable form and only the authorized user can recover them to their original form by using the secret key shared between senders and receivers. However, the meaningless form maybe arouses much more attentions of unauthorized users to decrypt the stego messages and even attack them. To avoid unnecessary attentions, steganography is proposed to hide secret information into a cover medium. Similar to steganography, various data hiding techniques that facilitate the transmission of secret information in the cover media without

* Corresponding author. Address: Room 261, Teaching Building 11, Yuquan Campus, Zhejiang University, Hangzhou 310027, PR China. Tel./fax: +86 571 86971612.
E-mail address: zhemingl@yahoo.com (Z.-M. Lu).

0020-0255/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2009.05.021

arousing attentions of malicious attackers are proposed. For traditional data hiding techniques, the cover media may be only slightly modified and these changes do not arouse suspicion of potential attackers. A drawback of these methods is that the cover media will be injured and cannot be recovered after secret data extraction. However, in several special fields, such as military, medical and legal fields, it is crucial to recover the original cover media without any distortion after the secret data are extracted. To solve this problem, researchers have further proposed various reversible data hiding algorithms. In a reversible data hiding scheme, the secret data are embedded in a reversible way so that authorized users can not only extract the hidden data but also recover the original cover media without any distortion.

Data hiding technique has become a highlighted research topic in recent years. Data hiding algorithms are applied in the spatial domain, transform domains and other compressed domains. In the spatial domain, Lee and Chen [17] modified the least significant bit (LSB) of each pixel in the cover image to embed the secret data and then Chang et al. [5] employed a dynamic programming strategy to find the optimal LSB substitution for data hiding. In 2006, by combining both run-length encoding and modular arithmetic, Chang et al. [11] proposed two efficient data hiding methods to embed bitmap files and general gray scale files into gray scale images, respectively. In transform domains, Chang et al. [4] employed the mid-frequency coefficients of the DCT transformed cover image to embed secret data. In the same way, Iwata et al. [16] utilized the boundaries between zero and non-zero DCT coefficients to hide secret data.

As a special branch of data hiding, the reversible data hiding techniques are similarly designed and applied in above-mentioned domains. The first reversible embedding method proposed by Barton [2] is employed in spatially uncompressed images. In this scheme, several bits of the original cover image are compressed by an embedding operation, and then both the compressed and secret data are hidden in the cover image. Later, Honsinger et al. [15] proposed a reversible embedding method based on the addition modulo 256, which flips pixels to implement the reversible embedding process. The scheme can avoid both overflow and underflow, but would cause annoying salt and pepper noises when pixel values are near the upper or lower bound. In 2003, Tian [21] proposed a different method based on modifying the difference between a pair of pixel values, and the expandable pairs are used to circumvent overflow and underflow. Later, Tian [22] further proposed a reversible data hiding technique based on the difference expansion (DE) technique to increase the capacity and improve the image quality. The drawback of this scheme lacks the ability of capacity control because of the unforeseen but necessary location map. In 2004, Alattar [1] proposed a generalized difference expansion method that hides several bits of secret data in the difference expansion of vectors consisting of adjacent pixels. The compressed bits of the original cover image are not required for realizing reversibility. In 2008, Lin et al. [18] proposed a multilevel reversible data hiding scheme based on the difference image histogram modification which used the peak point to hide messages. The scheme utilized a multilevel hiding strategy to achieve large hiding capacity and keep the distortion comparatively low. In 2009, Tsai et al. [23] proposed a reversible image hiding scheme based on histogram shifting for medical images. To enlarge the embedding capacity, the similarity of neighboring pixels was explored. In addition, the prediction technique and the residual histogram of the predicted errors of the cover image were used to hide the secret data.

In order to save the bandwidth of networks and the storage space for the growing sizes of various multimedia files, many well accepted image compression algorithms, such as VQ and JPEG, have been proposed and employed in the reversible data hiding techniques. In 2001, to authenticate digital images in the JPEG domain, Fridrich et al. [14] presented an invertible data hiding scheme which employs an order-2 function to modify the quantization table and losslessly embeds one bit of secret data per DCT coefficient. Later, Xuan et al. [25] proposed to hide data into the high-frequency wavelet coefficients using the histogram modification method, and employed a preprocessing step to prevent pixels from over-flowing and under-flowing. In 2006, Bausys et al. [3] presented a semi-blind reversible pixel-wise image authentication framework in the frequency domain. In 2007, Chang et al. [10] presented a lossless steganography scheme for hiding secret data in each block of quantized DCT coefficients in JPEG images. As one of the most commonly studied image compression techniques, VQ has been widely applied in data hiding techniques because of its simplicity and cost-effective implementation. In 2005 and 2006, Chang et al. [12,13,8] proposed several reversible data hiding algorithms based on the SMVQ or modified SMVQ. As the SMVQ-based algorithm utilizes the subcodebook generated from the main codebook to encode image blocks, the bit rate is relatively low. However, the quality of the stego images relies heavily on the side match distortions. Almost at the same time, Yang et al. [26] proposed MFCVQ-based reversible watermarking that adopts four adjacent blocks to encode the current block and utilizes certain flag bits to denote the chosen block. However, the method sacrifices the image quality to make reversibility feasible, and its hiding capacity is low. In 2007, Chang et al. [9] proposed a reversible scheme for VQ-compressed images based on a declustering strategy and the local spatial characteristics of the cover images. The method can not only embed different numbers of secret bits according to various declustering strategies but also achieve a large embedding capacity. However, this scheme sacrifices the quality of the stego images. In 2009, Chang et al. [7] presented a novel joint neighboring coding (JNC) scheme in the VQ index domain for reversible data hiding. In this paper, the JNC method was first introduced and applied in the data hiding. In 2009, Chang et al. [6] proposed a locally adaptive coding method for reversible information hiding in the VQ compressed domain. In the same year, based on the algorithm proposed in the literature [7], Wang et al. [24] proposed a lossless data hiding scheme based on improved VQ index joint neighboring coding. The method achieves high quality of the stego images and high hiding capacity. However, the security level of the proposed method is not high enough because the secret data are directly exposed in the output codestream. To enhance the security, in 2009, Lu et al. [20] proposed an improved reversible data hiding scheme based on VQ index residual value coding. The literature selected various neighboring encoded indices to encode the current index according to the secret data content. compared with reference [24], the method improves the security of secret data, but the capacity greatly decreases. In this paper, a novel path