Reversible Watermarking Techniques

Shao-Wei Weng¹, Yao Zhao¹, and Jeng-Shyang Pan²

¹ Institute of Information Science, Beijing Jiao Tong University, Beijing 100044, P.R. China
wswweiwei@126.com
yzhao@center.njtu.edu.cn

² National Kaohsiung University of Applied Sciences, Kaohsiung 807, Taiwan, R.O.C.
jspan@cc.kuas.edu.tw
http://bit.kuas.edu.tw/~jspan/

Summary. In this chapter, several reversible watermarking algorithms are introduced. The key idea of existing reversible watermarking techniques according to two basic classifications is generalized in this chapter. The first classification is based on Robust Spatial Additive Watermarks combined with modulo addition. The second classification is partitioned into three subsets. The first subset based on Lossless Compression Techniques contains lossless bit-plane compression methods in the spatial domain and in the Integer Discrete Wavelet Transform (IDWT) domain, the reversible RS data-embedding method, and the lossless G-LSB data-embedding method, respectively. The second subset contains the algorithm based on histogram shifting techniques. The third subset based on the value expansion techniques contains Integer Discrete Cosine Transform (IDCT) based bit-shifting method. Finally, the chapter introduces our own technique. The performance of related techniques is compared and their respective advantage and disadvantage are also analyzed in this chapter.

13.1 Introduction

Digital watermarking refers to the process of embedding some labels or signatures into digital media without introducing perceptible artifacts. It plays a vital role in the applications to copyright protection of digital media, authentication, date integrity, fingerprinting, and data hiding. However, digital watermarking usually introduces a slight but irreversible or permanent distortion to the original image. Although the distortion is generally quite small, it may not be acceptable for some applications, such as in the fields of the law enforcement, medical and military systems. It is desired to restore the original image from the watermarked image after the hidden data is retrieved due to the highly-precise requirement. The data embedding techniques, capable of reversing to the exact copy of the original image, are referred to as Reversible,
Invertible, Lossless, Distortion-Free, or Erasable Data Embedding Technique. The reversible data embedding techniques either have the desirable properties of a digital watermark, such as perceptible transparency, robustness, security, and data capacity, or have reversibility.

From the literature, most data embedding techniques cannot be completely reversed, since embedded distortion due to discarded information, quantization and integer rounding at the boundaries of the grayscale range cannot be removed. For example, by replacing the bits in the bit-planes, that is, least significant bit-plane (LSB), with watermarking bits, watermarking techniques discard all replaced bits of the bit-planes. Consequently, the bit-replacement is clearly lossy. For watermarking techniques, based on the quantization such as Vector Quantization (VQ), Quantization Index Modulation (QIM), and so on, quantization error makes that retrieved pixel values mismatch with original pixel values. Hence, there is little hope to restore the original image without distortion. For spread spectrum watermarking techniques in Discrete Cosine Transform (DCT) domain and/or Discrete Wavelet Transform (DWT) domain, round-off error and truncation error make invertible watermarking impossible. Additive, non-adaptive schemes (truncation addition) are almost lossless except for the pixels with grayscales close to 0 or 255 where truncation has occurred owing to overflow or underflow.

In all of the above mentioned embedding techniques are not suitable in some applications. Those require high-precision, such as the medical images, artworks, and so on. Some watermarking techniques have been developed in order to presented to satisfy the reversible requirement over the last a few years. All existing reversible watermarking techniques can be classified into two categories.

**Type-I algorithms** [1]-[4] are based on robust, spatial additive watermarks combined with modulo addition. These techniques add the payload by modulo addition to the host image during embedding process. At the decoder, the payload can be reconstructed from the watermarked image, and then it is subtracted to restore the original image. However, modulo additions would cause a disturbing visual artifact resembling a correlated salt-and-pepper noise into the watermarked image when pixel values close to the maximally allowed value are flipped to zero and vice versa. Type-I algorithms generally combine statistical approaches. For example, the patchwork algorithm with modulo additions is used to ensure correct watermark extraction. Hence, Type-I algorithms are robust for the data embedding and allow for extraction of hidden data even for the perturbed watermarked image. It can not ensure that the original image is precisely retrieved. The algorithm in [4] has certain degree of robustness against JPEG lossy compression. This is the only existing robust lossless data embedding algorithm for use against JPEG compression.