Chapter 16
Reversible Information Hiding and Its Application to Image Authentication

Masaaki Fujiyoshi
Tokyo Metropolitan University, Japan

Hitoshi Kiya
Tokyo Metropolitan University, Japan

ABSTRACT

This chapter addresses a new class of Reversible Information Hiding (RIH) and its application to verifying the integrity of images. The method of RIH distorts an image once to hide information in the image itself, and it not only extracts embedded information but also recovers the original image from the distorted image. The well-known class of RIH is based on the expansion of prediction error in which a location map, which indicates the pixel block positions of a certain block category, is required to recover the original image. In contrast, the method described in this chapter is free from having to memorize any parameters including location maps. This feature suits the applications of image authentication in which the integrity of extracted information guarantees that of a suspected image. If image-dependent parameters such as location maps are required, the suspected image should first be identified from all possible images. The method described in this chapter reduces such costly processes.

INTRODUCTION

Information Hiding (IH) technology has been diligently studied to not only solve security-related problems, particularly to protect the intellectual property rights of digital content and covert communication, but also non security-oriented issues, such as the monitoring of broadcasts and multiplexing of captions (Cox, Miller, Bloom, Fridrich, & Kalker, 2008; Wu & Liu, 2003). IH techniques are used to embed information referred to as a payload into a target signal that is called the original or host signal. They then generate a slightly distorted signal carrying the payload by exploiting the redundancy of the original signal in the human perceptual system, and this distorted

DOI: 10.4018/978-1-4666-2919-6.ch016

signal is referred to as a stego signal. Many IH techniques extract hidden information from a stego signal, but the stego signal is left as it is.

As the original image needs to be accurately restored in military, medical, and heritage imagery applications as well as the hidden payload to be extracted, Reversible IH (RIH) methods that restore the original image from a stego image have been proposed. Of several RIH implementations (Caldelli, Filippini, & Becarelli, 2010), this chapter focuses on prediction error expansion-based RIH (PEE-RIH) (Conotter, Boato, Carli, & Egiazarian, 2010; Thodi & Rodríguez, 2007; Yang, Chung, Yu, & Liao, 2010), which is one major class in RIH because of its capabilities of accepting large payloads or serving large capacities. The method of PEE-RIH hides a portion of the payload into a pixel block of the original image by expanding or rounding prediction error, which is the difference between the target pixel value and its corresponding prediction in the block.

The choice between expansion and rounding for a pixel block is based on its corresponding prediction error; expansion is for expandable blocks and rounding is for changeable blocks. Consequently, the recovery of the original image with PEE-RIH should distinguish between the two types of blocks, but expandable blocks become changeable through the IH process. To overcome this problem, a location map, which indicates the block positions for a certain block group, is used in PEE-RIH (Kamstra & Heijmans, 2005; Thodi & Rodríguez, 2007). This location map is image-dependent, and should be memorized in a database or transmitted along with a stego image. This fact narrows its applications and decreases its practicality.

This chapter describes a new method of PEE-RIH that does not require parameters to be memorized including location maps. It can classify pixel blocks as expandable and unexpandable, even in stego images, by utilizing a threshold parameter that is introduced based on block statistics. Part of the information is hidden in each expandable block in an error expansion-based manner. These strategies free the method from location maps. In addition, the method described in this chapter is completely free from having to memorize parameters by adapting the capacity to the payload size. This feature of the method suits image authentication applications (Mahdian & Saic, 2010; Rey & Dugelay, 2002) in which the integrity of extracted information guarantees the integrity of the suspected stego image.

Digital signature technology is applied to images by considering them to be media-unaware data (Schneier, 1994) in basic image authentication. This approach should be used to transmit or store the signature along with the image itself. IH-based image authentication hides a predefined pattern in the image with a fragile IH technique in which image editing of the stego image corrupts the hidden payload (Rey & Dugelay, 2002). Image tampering is exposed by comparing the pattern and the payload extracted from the suspected image. This framework results in distorted images even if the images are genuine when an irreversible IH technique is used. Consequently, RIH-based image authentication is needed, which is free from having to transmit or store signatures, and can deliver undistorted images when the images are genuine.

BACKGROUND

This section briefly presents some implementations of RIH (Caldelli, Filippini, & Becarelli, 2010), and it then describes the most fundamental method of PEE-RIH (Thodi & Rodríguez, 2007) to clarify the problem on which this chapter is focused. The fundamental frameworks for image authentication (Mahdian & Saic, 2010) including those that are RIH-based are also mentioned to emphasize the advantages of the new method in its application to image authentication.

First, four major classes of RIH are presented here, viz., compression-based, histogram modification-based, difference expansion-based, and