Chapter 4
Image Forensics Using Generalised Benford's Law for Improving Image Authentication Detection Rates in Semi-Fragile Watermarking

Xi Zhao
University of Surrey, UK

Anthony T. S. Ho
University of Surrey, UK

Yun Q. Shi
New Jersey Institute of Technology, USA

ABSTRACT

In the past few years, semi-fragile watermarking has become increasingly important to verify the content of images and localise the tampered areas, while tolerating some non-malicious manipulations. In the literature, the majority of semi-fragile algorithms have applied a predetermined threshold to tolerate errors caused by JPEG compression. However, this predetermined threshold is typically fixed and cannot be easily adapted to different amounts of errors caused by unknown JPEG compression at different quality factors (QFs). In this paper, the authors analyse the relationship between QF and threshold, and propose the use of generalised Benford's Law as an image forensics technique for semi-fragile watermarking. The results show an overall average QF correct detection rate of approximately 99%, when 5%, 20% and 30% of the pixels are subjected to image content tampering and compression using different QFs (ranging from 95 to 65). In addition, the authors applied different image enhancement techniques to these test images. The proposed image forensics method can adaptively adjust the threshold for images based on the estimated QF, improving accuracy rates in authenticating and localising the tampered regions for semi-fragile watermarking.

DOI: 10.4018/978-1-4666-1758-2.ch004

INTRODUCTION

Nowadays, the popularity and affordability of advanced digital image editing tools, allow users to manipulate images relatively easily and professionally. Consequently, the proof of authenticity of digital images has become increasingly challenging and difficult. Moreover, image authentication and forensics techniques have recently attracted much attention and interest from the Police, particularly in law enforcement applications such as crime scene investigation and traffic enforcement applications.

Semi-fragile watermarking has been used to authenticate and localise malicious tampering of image content, while permitting some non-malicious or unintentional manipulations. These manipulations can include some mild signal processing operations such as those caused by transmission and storage of JPEG images. In the literature, a significant amount of research has been focused on the design of semi-fragile algorithms that could tolerate JPEG compression and other common non-malicious manipulations (Lin & Chang, 2000; Lin et al., 2000; Zou et al., 2006; Zhu et al., 2007a; Zhu et al., 2007b; Yu et al., 2000; Kundur & Hatzinakos, 1999). However, watermarked images could be compressed by unknown JPEG QFs. As a result, in order to authenticate the images, these algorithms have to set a pre-determined threshold that could allow them to tolerate different QF values when extracting the watermarks.

The art of determining the threshold values for semi-fragile watermarking schemes has been extensively documented by several researchers. In this paper, we review three common approaches. The first approach uses a threshold for authenticating each block of the image (Lin et al., 2000; Zhu et al., 2007a). In this scheme, if a block of correlation coefficients cr (between the extracted watermark w' and its corresponding original watermark w) is smaller than threshold τ, this block is classified as a tampered block, and vice versa. This is represented in Equation (1):

$$cr\left(w, w'\right) < \tau, \;\; \max\left(\tau\right) - \tau = TM \tag{1}$$

where $\max\left(\tau\right)$ is the maximum threshold value with $w = w'$, and TM is the JPEG compression tolerance margin. We discuss this approach in more detail in the next section. The second approach uses a threshold, which has been pre-determined during the watermark embedding process (Zou et al., 2006; Zhu et al., 2007a). An example is illustrated in Figure 1, where the watermarks w are embedded into each side of threshold τ according to the watermark value (e.g., 0 or 1), by shifting or substituting the corresponding coefficient. The value of T and $-T$ controls the perceptual quality of the watermarked image. Threshold τ is determined empirically to detect the watermark while extracting the watermarks w'. TM is the JPEG compression tolerance margin. If $w' > \tau$ then $w' = 1$, otherwise $w' = 0$ (Zhu et al., 2007a).

The third approach uses a threshold for comparison with the result of applying the Tamper Assessment Function (TAF) during the authentication of images (Kundur & Hatzinakos, 1999). The extracted watermarks w' and their corresponding original watermarks w are calculated by using TAF, as in Equation (2):

$$TAF\left(w, w'\right) = \frac{1}{N_w} \sum_{i=1}^{N_w} w\left(i\right) \oplus w'\left(i\right) \tag{2}$$

where N_w is the length of the watermark. The TAF value is compared with a threshold τ, where $0 \le \tau \le 1$. If $TAF(w, w') > \tau$, then the watermarked image is considered as a tampered image, otherwise it is not. The tolerance margin can also be denoted as $TM = 1 - \tau$. The thresholds τ mentioned previously are pre-determined which will result in some fixed tolerance margins. A significant amount of research has been dedicated to improving the watermark embedding algorithms by analysing the characteristics of