

# IMAGE FORENSICS USING GENERALISED BENFORD'S LAW FOR ACCURATE DETECTION OF UNKNOWN JPEG COMPRESSION IN WATERMARKED IMAGES

*Xi Zhao<sup>1</sup>, Anthony TS Ho<sup>1</sup>, Yun Q Shi<sup>2</sup>*

<sup>1</sup>Dept. of Computing, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, Surrey, GU2 7XH, UK

<sup>2</sup>Dept. of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102, USA

## ABSTRACT

In the past few years, semi-fragile watermarking has become increasingly important as it can be used to verify the content of images and to 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) applied to the watermarked images. In this paper, we 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, to accurately detect the unknown QF of the images. The results obtained show an overall average QF correct detection rate of approximately 99% when 5% of the pixels are subjected to image content tampering, as well as compression using different QFs (ranging from 95 to 65). Consequently, our proposed image forensics method can adaptively adjust the threshold for images based on the estimated QF, therefore, improving the accuracy rates in authenticating and localising the tampered regions for semi-fragile watermarking.

**Index Terms**—Semi-fragile Watermarking, Generalised Benford's Law, DCT, JPEG Compression, Image Authentication

## 1. 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 [1-7]. 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 [2] [4]. 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  $\tau$ , this block is classified as a tampered block, and vice versa. This is represented in equation (1):

$$cr(w, w') < \tau, \max(\tau) - \tau = TM \quad (1)$$

where  $\max(\tau)$  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 [3] [4]. An example is illustrated in Figure 1, where the watermarks  $w$  are embedded into each side of threshold  $\tau$  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  $\tau$  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  $w' = 0$  [4].

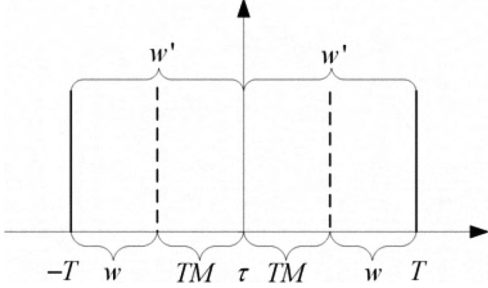


Figure 1 illustrates the pre-determined threshold during the watermark embedding process.

The third approach uses a threshold for comparison with the result of applying the Tamper Assessment Function (TAF) during the authentication of images [7]. The extracted watermarks  $w'$  and their corresponding original watermarks  $w$  are calculated by using TAF, as in equation (2):

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

where  $N_w$  is the length of the watermark. The TAF value is compared with a threshold  $\tau$ , where  $0 \leq \tau \leq 1$ . 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  $\tau$  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 JPEG coefficients of the compressed watermarked image [5-7]. Alternatively, Error Correction Coding (ECC) has been used for improving watermark detection and authentication rates [3].

However, the relationship between QF and threshold has not been discussed in the literature. If the QF could be estimated, then appropriate thresholds could be adapted for each test image, before initialising the watermark extraction and authentication process. The use of Benford's Law has already been applied to image forensics of JPEG compressed images [8]. In this paper, we analyse the relationship between QF and threshold, and propose a framework that further explores generalised Benford's Law as an image forensics technique, in an effort to accurately detect the unknown JPEG compression in semi-fragile watermarking images.

The rest of this paper is organised as follows. Section 2 demonstrates a simple semi-fragile watermarking scheme to explain the relationship between threshold, QF, missed detection rate and false alarm rate when authenticating test images. Section 3 describes the background of Benford's Law, generalised Benford's Law and their relationship with the watermarked image, JPEG compressed watermarked image. Section 4 describes the proposed image forensics method and experimental results are presented in Section 5. Finally, Section 6 presents the conclusion and future work.

## 2. THRESHOLD IN SEMI-FRAGILE WATERMARKING

In this section, the feasibility of our proposed method is investigated in detail. By analysing the first approach previously reviewed in [2] [4], a simple semi-fragile watermarking algorithm based on discrete cosine transform (DCT) and the importance of threshold is also described.

### 2.1 Watermark embedding process

As shown in Figure 2, the original image is divided into non-overlapping sub-blocks of  $8 \times 8$  pixels and DCT is applied to each block.

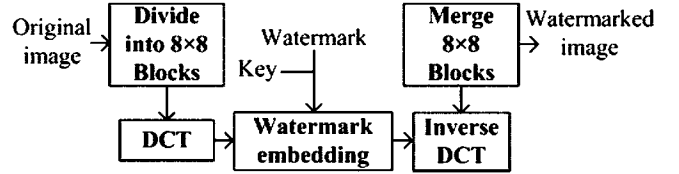


Figure 2 illustrates the watermark embedding process

The watermark embedding process is achieved by modifying the random selected mid-frequency (shaded blocks in Figure 3) of the DCT coefficients in each block as follows:

$$coef' = \begin{cases} coef, & (coef \geq T \wedge w = 1) \vee (w \leq -T \wedge w = -1) \\ \alpha, & (coef < T \wedge w = 1) \\ -\alpha, & (coef > T \wedge w = -1) \end{cases} \quad (3)$$

where  $coef$  is the original DCT coefficient,  $coef'$  is the modified DCT coefficient.  $w$  is the watermark bits generated via a pseudo-random sequence (1 and -1) using a secret key.  $T > 0$  determines the perceptual quality of the watermarked image and  $\alpha \in [T/2, T]$  is a constant. The inverse DCT is then applied to each block to obtain the watermarked image. Figure 3 illustrates examples of  $8 \times 8$  DCT block with different watermark sequences and embedding locations for each block.

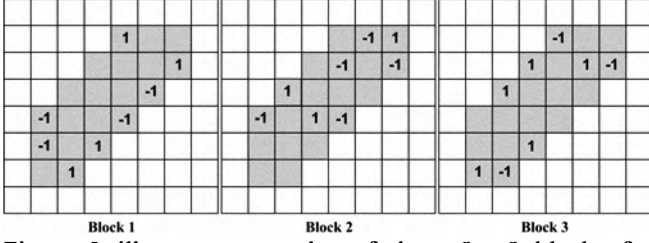


Figure 3 illustrates examples of three  $8 \times 8$  blocks for watermark embedding

## 2.2 Watermark detection and authentication process

In Figure 4, the test image is first divided into non-overlapping sub-blocks of  $8 \times 8$  pixels, and DCT is then applied to each block.

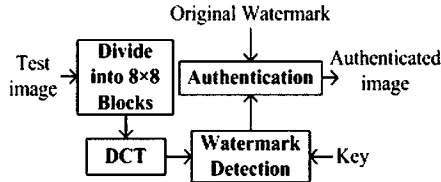


Figure 4. An illustration of the watermark detection and authentication processes.

The watermark detection algorithm shown in equation (4) is then applied.

$$w' = \begin{cases} 1, & coef' \geq 0 \\ -1, & coef' < 0 \end{cases} \quad (4)$$

where  $w'$  is the extracted watermark bits and  $coef'$  is the DCT coefficient of the test image. The extracted watermark bits from each block are compared with its corresponding original watermark  $w$  bits to obtain the correlation coefficient  $cr$  as shown in equation (5):

$$cr(w, w') = \frac{\sum (w' - \bar{w}') (w - \bar{w})}{\sqrt{\sum (w' - \bar{w}')^2 \sum (w - \bar{w})^2}} \quad (5)$$

The correlation coefficient of each block is then compared with a pre-determined threshold  $-1 \leq \tau \leq 1$  as below:

$$Block = \begin{cases} un-tampered, & cr(w, w') \geq \tau \\ tampered, & cr(w, w') < \tau \end{cases} \quad (6)$$

## 2.3 The importance of threshold

The magnitude of threshold affects the false alarm rate ( $P_F$ ) is the percentage of un-tampered blocks detected as tampered and the missed detection rate ( $P_{MDR}$ ) is the percentage of tampered blocks detected as un-tampered.

Figure 5 shows that the missed detection rate decreases if the threshold is in close proximity to 1. This

also leads to an increase in the false alarm rate. However, if the threshold is set to be of a close proximity to -1, then the missed detection rate increases and the false alarm rate will decrease. This results in a dilemma in determining a suitable threshold. For the proposed semi-fragile watermarking scheme, the threshold is set as 0.5, which provides a good trade-off between  $P_F$  and  $P_{MDR}$ .

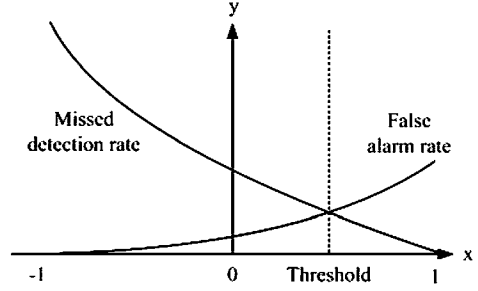


Fig. 5. the relationship among threshold,  $P_F$  and  $P_{MDR}$ .

Figure 6 illustrates the overall relationship between threshold,  $P_F$  and  $P_{MDR}$  for the proposed semi-fragile watermarking scheme. The watermarked image 'Lena' has been tampered with a rectangular block and JPEG compressed at QF=75. Figure 6 (a) shows the pre-determined threshold  $\tau = 0.5$  used for authentication. The authenticated image shows that the proposed semi-fragile watermarking scheme can localise the tampered region with reasonable accuracy, but with some false detection errors.

In Figures 6 (b) and 6 (c), the lower and upper thresholds  $\tau = 0.3$  and  $\tau = 0.7$  were used for comparison, respectively. Figure 6 (b) shows that the false alarm rate has decreased whilst the missed detection rate has increased in the authenticated image. Figure 6 (c) shows the image has a lower missed detection rate but with a higher false alarm rate. From this comparison,  $\tau = 0.5$  was chosen for JPEG compression at QF=75. However, if QF =95, then  $\tau = 0.5$  may not be adequate as shown in Figure 7 (a). The missed detection rate is higher than Figure 7 (b) with  $\tau = 0.9$ . Therefore, it would be advantageous to be able to estimate the QF of JPEG compression, so that an adaptive threshold can be applied for increasing the authentication accuracy. In this paper, we propose the use of generalised Benford's Law to estimate the QF, and this will be explained in the next section.



Fig. 6. Different thresholds for QF=75

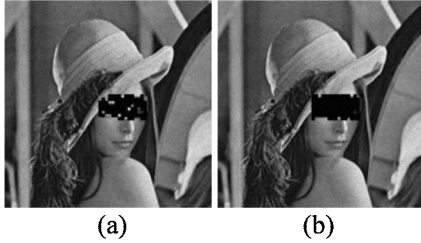


Fig. 7. Different thresholds for QF=95

### 3. BENFORD'S LAW FOR SEMI-FRAGILE WATERMARKING

#### 3.1 Background of Benford's Law

Benford's Law was introduced by Frank Benford in 1938 [9] and then was developed by Hill [10] for analysis of the probability distribution of the first digit (1-9) of the number from natural data in statistics. Benford's Law has also been applied to accounting forensics [11] [12]. Since the DCT coefficients of a digital image obey Benford's Law, it has recently attracted a significant amount of research interests in image processing and image forensics [8] [13] [14]. The basic principle of Benford's Law is given as follows:

$$p(x) = \log_{10} \left( 1 + \frac{1}{x} \right), \quad x = 1, 2, \dots, 9 \quad (7)$$

where  $x$  is the first digit of the number and  $p(x)$  is the probability distribution of  $x$ .

In contrast to digital image watermarking which is an "active" approach by embedding bits into an image for authentication, image forensics is essentially a "passive" approach of analysing the image statistically to determine whether it has been tampered with. Fu *et al.* [8] proposed a generalised Benford's Law, used for estimating the QF of the JPEG compressed image, as shown in equation (8).

$$p(x) = N \log_{10} \left( 1 + \frac{1}{s + x^q} \right), \quad x = 1, 2, \dots, 9 \quad (8)$$

where  $N$  is a normalisation, and  $s$  and  $q$  are model parameters [8]. Their research indicated that the probability distribution of the first digit of the JPEG coefficients obey generalised Benford's Law after the quantisation. Moreover, the probability distributions were not following the generalised Benford's Law if the image had been compressed twice with different quality factors. Thus, by utilizing this property, the QF of the image can be estimated. In this paper, we propose to use generalised Benford's Law for detecting unknown JPEG compression QF to improve the authentication process, during the semi-fragile watermarking authentication process.

#### 3.2 Benford's Law, Generalised Benford's Law vs. Watermarked images

The feasibility of generalised Benford's Law for use in semi-fragile watermarking was first investigated. In our experiment, we selected 1338 uncompressed grayscale images from the Uncompressed Image Database (UCID) [15] for analysis to ensure that there was no compression performed on the images previously. Throughout this section we adhere to the same terminology as used in [8], where "Block-DCT coefficients" refers to the  $8 \times 8$  block-DCT coefficients before the quantisation, and "JPEG coefficients" refers to the  $8 \times 8$  block-DCT coefficients after the quantisation.

Figure 8 illustrates the comparison between the probability distribution of Benford's Law, mean distribution of 1<sup>st</sup> digit of block-DCT coefficients of 1338 images and the watermarked images. The average PSNR between the original images and watermarked images was approximately 35.71dB, which is considered to be of acceptable image quality. Figure 8 shows that the distribution of the 1<sup>st</sup> digits of the block-DCT coefficients for the uncompressed images obeys Benford's Law closely. This was also observed by Fu *et al.* in their analysis [8]. In terms of the watermarked images, the mean distribution also follows Benford's Law. The mean standard deviations of the 1338 uncompressed images and their watermarked images are considerably small, as shown in Table I. The average  $\chi^2$  divergence [8] for watermarked images is also small at 0.0115. This indicates a good fitting between Benford's Law and watermarked images. The  $\chi^2$  divergence is shown in equation (9).

$$\chi^2 = \sum_{i=1}^9 \frac{(p_i' - p_i)^2}{p_i} \quad (9)$$

where  $p_i'$  is the actual 1<sup>st</sup> digit probability of the DCT coefficients of the watermarked images and  $p_i$  is the 1<sup>st</sup> digit probability from Benford's Law in equation (7). Hence, the results indicated that the probability distribution 1<sup>st</sup> digits of the block-DCT coefficients of the watermarked images follow Benford's Law. Figure 9 (a) illustrates an example of  $8 \times 8$  DCT coefficients. The 1<sup>st</sup> digits of the AC coefficients are then extracted as shown in Figure 9 (b).

Figures 10-12 illustrate the comparisons between the probability distribution of Benford's Law, generalised Benford's Law and the mean distributions of the 1<sup>st</sup> digits of block JPEG coefficients of the watermarked images compressed at QF=100, 75, 50, respectively. Table II summarises the mean standard deviations obtained for the 1338 original and watermarked images, JPEG compressed at the three QF rates are considerably small. Furthermore, as shown in TABLE III, the  $\chi^2$  divergences are also calculated by using equation (9), where  $p_i'$  is the actual 1<sup>st</sup> digit probability of the JPEG coefficients of the compressed

watermarked images,  $p_i$  is the 1<sup>st</sup> digit probability from generalised Benford's Law in equation (8) and  $N$ ,  $s$  and  $q$  are model parameters gained from [8]. These results also indicate the good fitting between generalized Benford's Law and watermarked images compressed with different QFs, respectively.

The results indicated that the probability distributions of the 1<sup>st</sup> digits of JPEG coefficients of the watermarked images, in Figures 10-12, obey generalised Benford's Law model proposed by Fu *et al.* [8], in equation (8). Hence, we could employ their model to estimate the unknown QF of test images to adjust the threshold for authentication. The improved authentication process is described in next section.

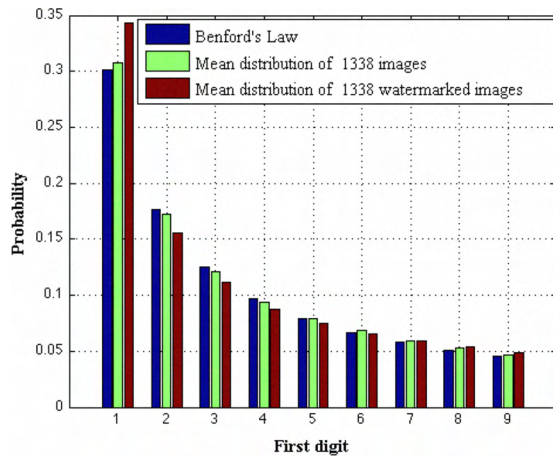


Fig. 8. 1<sup>st</sup> digit of block-DCT coefficients

TABLE I

Mean standard deviations of 1338 images

1 <sup>st</sup> digit	Original images	Watermarked images
1	0.0139	0.0145
2	0.0084	0.0078
3	0.0067	0.0068
4	0.0050	0.0049
5	0.0037	0.0030
6	0.0032	0.0023
7	0.0028	0.0021
8	0.0028	0.0023
9	0.0022	0.0021

1.3e+3	4.7	3.2	-0.19	0.25	-0.5	-4.5	5.6
7.9	-0.7	0.6	-4.9	1.9	2.9	-3.7	3.3
-5.0	-0.2	-1.6	1.7	-0.6	-0.4	1.8	-2.2
2.3	1.1	1.7	0.9	-0.7	-1.3	0.2	1.1
-1.0	-1.2	-0.3	-1.4	1.7	1.1	-1.4	-0.6
1.2	0.4	-1.8	-0.1	-2.0	-0.7	1.6	0.7
-1.7	0.2	3.1	1.6	1.6	-2.2	-1.2	-0.9
1.3	-0.4	-2.4	-1.6	-0.8	1.9	0.5	0.6

(a)

4	3	1	2	5	4	5
7	7	6	4	1	2	3
5	2	1	1	6	4	1
2	1	1	9	7	1	2
1	1	3	1	1	1	1
1	4	1	1	2	7	1
1	2	3	1	1	2	1
1	4	2	1	8	1	5

(b)

Fig. 9. 1<sup>st</sup> digit of 8 × 8 Block-DCT coefficients

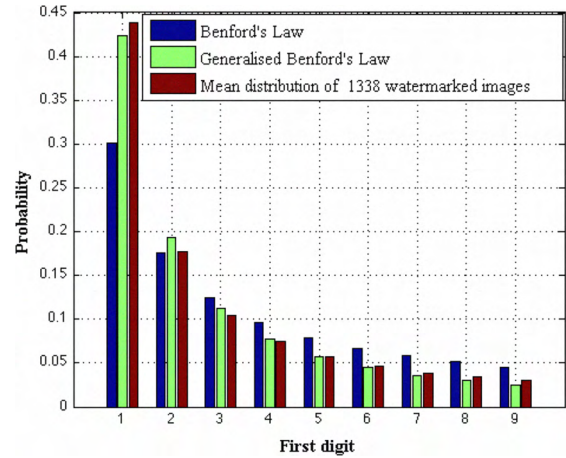


Fig. 10. 1<sup>st</sup> digit of JPEG coefficients (QF=100)

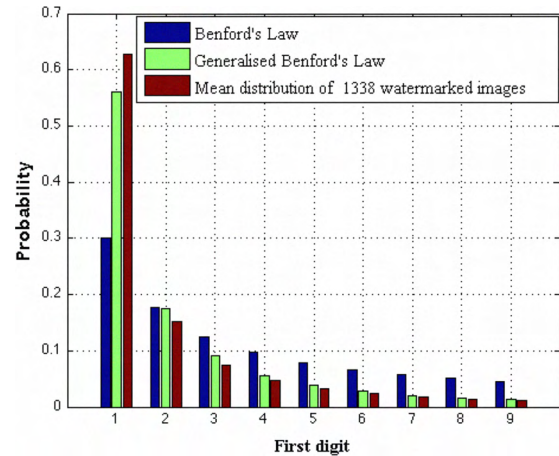


Fig. 11. 1<sup>st</sup> digit of JPEG coefficients (QF=75)

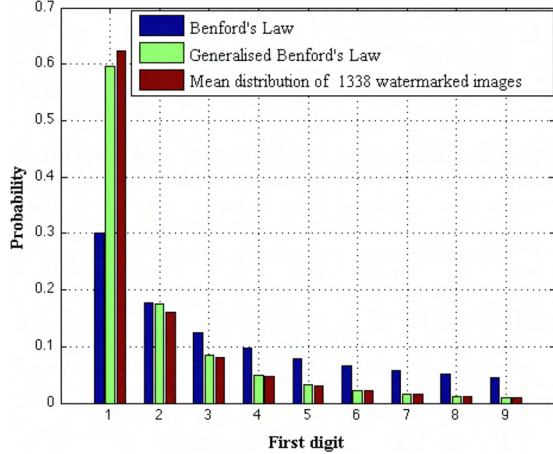


Fig. 12. 1<sup>st</sup> digit of JPEG coefficients (QF=50)

TABLE II  
Mean standard deviations of 1338 JPEG compressed images

1 <sup>st</sup> digit	Original images			Watermarked images		
	QF100	QF75	QF50	QF100	QF75	QF50
1	0.0828	0.0327	0.0399	0.0664	0.0514	0.0509
2	0.0165	0.0067	0.0089	0.0122	0.0132	0.0149
3	0.0169	0.0066	0.0088	0.0143	0.0111	0.0112
4	0.0163	0.0058	0.0072	0.014	0.0082	0.0084
5	0.0142	0.0049	0.0059	0.0121	0.0064	0.0065
6	0.0123	0.0043	0.0048	0.0102	0.0052	0.0051
7	0.0107	0.0037	0.0039	0.0087	0.0042	0.0041
8	0.0094	0.0032	0.0033	0.0075	0.0035	0.0034
9	0.0084	0.0027	0.0027	0.0065	0.003	0.0028

TABLE III  
Average  $\chi^2$  of 1338 compressed watermarked images

QF	Model Parameters			$\chi^2$
	N	q	s	
100	1.456	1.47	0.0372	0.0257
70	1.412	1.732	-0.337	0.0292
50	1.579	1.882	-0.2725	0.0166

#### 4. THE IMPROVED AUTHENTICATION METHOD

In this section, we explain the improved authentication process which uses the generalised Benford Law model. In Figure 13, the test image is divided into non-overlapping blocks of  $8 \times 8$  pixels and DCT is then applied to each block. The watermark detection process then extracts the watermark bits using a secret key.

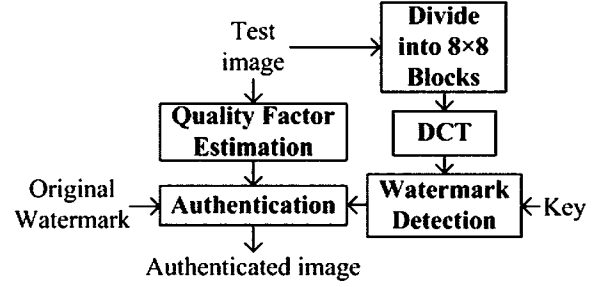


Fig. 13. Improved authentication process

The same test image is also used for detecting the QF by the quality factor estimation process. This process works by firstly classifying the test image as compressed or uncompressed by adapting from [8]. If the test image has been compressed, the test image is then recompressed with the largest QF, from QF=100 to QF=50, in decreasing steps of 5. We decrease in steps of 5 as this gives us the most frequently used quality factors for JPEG compressed images (i.e. 95%, 90%, 85% etc.). For each compressed test image, the probability distribution of the 1<sup>st</sup> digits of JPEG coefficients is obtained. Each set of values are then analysed by employing the generalized Benford's Law equation and using the best curve-fitting to plot the data. In order to obtain the goodness of fit, we calculate the sum of squares due to error (SSE) of the recompressed images. We can detect the QF of the test image by iteratively calculating the SSE for all QFs (starting at QF=100, and decreasing in steps of 5), and as soon as  $SSE \leq 10^{-6}$ , we have reached the estimated QF for the test image. As per the pseudocode below, the threshold  $10^{-6}$  has been set to allow us to detect the QF of the test image. This threshold value was reported in [8], and has been verified by the results in our experiment.

```

If  $SSE \leq 10^{-6}$ 
    Then QF has been detected.
    Break,
  
```

End

Figure 14 illustrates the results of estimating the QF for a test image that has previously been compressed with QF=70. Three curves have been drawn in order to fit the three probability distribution data sets: generalized Benford's Law for QF=70, the test image recompressed with QF=70, and separately recompressed at QF=90. The distribution of QF=90 shows the worst fit and is considerably fluctuated, while the distribution of QF=70 is a generally decreasing curve, which also follows the trend of generalized Benford Law. These results indicate that if the test image has been double compressed without the same quality factor, the probability distribution would not obey the generalised Benford's Law.

Once the QF is estimated, the threshold  $\tau$  can be adapted according to different estimated QFs, based on the following conditions:

$$\tau = \begin{cases} 0.9 & QF \geq 90 \\ 0.7 & 90 < QF < 75 \\ 0.5 & QF \leq 75 \end{cases} \quad (10)$$

Finally, the correlation coefficient between original watermarks and extracted watermarks for each block is compared using the attuned threshold  $\tau$  to authenticate, in order to determine whether any blocks have been tampered with. This is similar to the authentication process as described in Section 2.

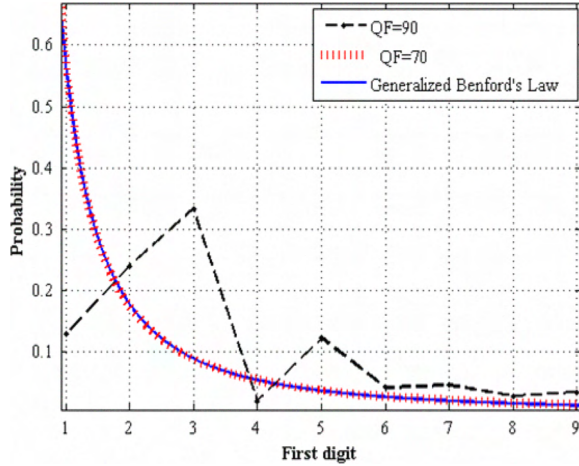


Fig. 14. Estimating the QF of a watermarked image

## 5. EXPERIMENTAL RESULTS

The watermarked images are generated by our proposed semi-fragile watermarking algorithm (as discussed in Section 2) using the 1338 test images from UCID [15]. In order to achieve a fair comparison, different embedding parameters are randomised for each image such as the watermarks location, watermark string and watermark bits. For our analysis, four types of test images with and without attacks are considered as shown in Figure 15.

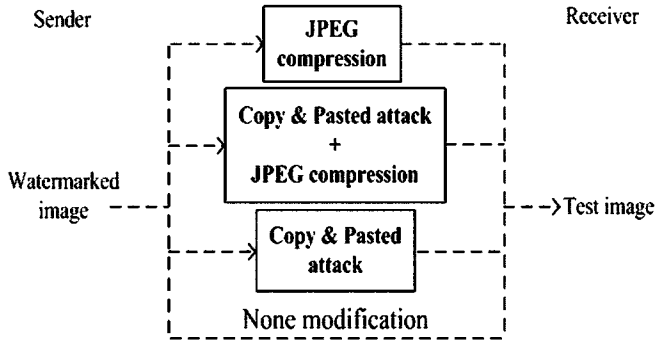


Fig. 15. Four types of test images with and without attacks

Table IV summaries the results obtained for test images that have been JPEG compressed only. To evaluate

the accuracy of the quality factor estimation process, each test image has been blind compressed from  $QF=100$  to  $QF=50$  in decreasing steps of 5. For each compression, the quality factor estimation process was used to determine the QF. The mean estimated QFs for all 1338 test images and each correctly identified detection accuracy rate  $P_{de}$  for each JPEG compression quality factor are shown in Table IV, based on equation (11).

$$P_{de} = \frac{\hat{\partial}}{\beta} \times 100\% \quad (11)$$

where  $\hat{\partial}$  is the number of correctly detected QF and  $\beta$  is the number of images tested. The mean estimated QF results indicate the QFs can be estimated with high accuracy. The only exceptions for lower correct detection rates,  $P_{de}$ , were obtained for  $QF=50$ ,  $QF=60$ , and  $QF=100$ . In the case of  $QF=50$ ,  $P_{de}$  was very low at approximately 18.2%, meaning that the process was probably detecting QFs close to  $QF=55$ . For  $QF=60$ , and  $QF=100$ , the detection rates were slightly better at 38.6% and 65.7%, respectively.

For comparison, both the mean estimated QF value and correct detection rate were used for each result to estimate the actual QF for the images. The QFs were then grouped into three different ranges:  $QF \geq 90$ ,  $90 > QF > 75$  and  $QF \leq 75$ . The grouping into three QF ranges did not have an overall effect on the authentication process. Results obtained for  $P_{de2}$  also showed the correct detection accuracy rates in these QF ranges were on average at 99%.

Table V summaries the results obtained for test images that have been attacked via copy & paste and then JPEG compressed. Each watermarked image has been tampered randomly in different regions by applying a copy & paste attack to 5% of the watermarked image (9830 pixels in 384512 pixels image), and also compressed with different QF values. The results showed that the quality factor estimation process was highly accurate even under these attacks.

From Table V, the lowest correct detection rates were obtained for  $QF=50$ ,  $QF=60$ , and  $QF=100$ . Two other experiments were performed with the test image subjected to only the copy & paste attack and with the test image without any modification. The detected QFs achieved for both experiments were approximately 99, and fit well in the upper range of  $QF \geq 90$ . Similarly, the results of  $P_{de2}$  also showed the correct detection rates in the three ranges were highly accurate with an overall average of 99%. As such, the threshold can be adapted into the three QF ranges according to the estimated QF of each test image as described in Section 4.

TABLE IV  
JPEG compression only

Actual QF	Mean Estimated QF	$P_{de}$	$\tau$	$P_{de2}$
100	98.16	65.7%		
95	94.87	97.3%	0.9	98.8%
90	90.06	98.2%		
85	84.20	91.4%	0.7	99.1%
80	79.77	97.5%		
75	75.35	97.0%		
70	69.77	98.8%		
65	64.42	93.7%	0.5	99.4%
60	62.42	38.6%		
55	55.15	94.1%		
50	54.25	18.2%		

TABLE V  
Copy & paste (5%) + JPEG compression

Actual QF	Mean Estimated QF	$P_{de}$	$\tau$	$P_{de2}$
100	98.60	72%		
95	95.00	100%	0.9	99.1%
90	90.14	98.6%		
85	84.83	97.9%	0.7	99.3%
80	79.95	99.6%		
75	75.22	99.1%		
70	69.87	99.5%		
65	64.46	98.7%	0.5	99.2%
60	61.54	63.9%		
55	54.93	96.6%		
50	53.32	20.4%		

## 6. SUMMARY

In this paper, we presented the relationship between QF and threshold, and proposed a framework incorporating the generalised Benford's Law as an image forensics technique to accurately detect unknown JPEG compression levels in semi-fragile watermarked images. We reviewed three typical methods of employing predetermined thresholds in semi-fragile watermarking algorithms and the limitations of using predetermined thresholds were also highlighted.

In our proposed semi-fragile watermarking method, the test image was first analysed to detect its previously unknown quality factor for JPEG compression, before proceeding with the semi-fragile authentication process. The results showed that QFs can be accurately detected for most unknown JPEG compressions. In particular, the average QF detection rate was as high as 96% for watermarked images compressed with QFs between 95-65, and 99% when the image was subjected to tampering of 5% pixels of the image and compressed with QFs between 95-65. The threshold was adapted into three specific ranges according to the estimated QF of each test image. For future

work, we plan to analyse and estimate double JPEG compression and other signal processing operations caused by transmission in semi-fragile watermarking images, as well as in robust watermarking.

## 7. REFERENCES

- [1] C.Y. Lin, and S.F. Chang, "Semi-Fragile Watermarking for Authenticating JPEG Visual Content," in *Proc. SPIE Security and Watermarking of Multimedia Contents II EI '00*, Jan. 2000.
- [2] E.T.Lin, C.I. Podilchuk, and J. Delp, "Detection of Image Alterations using semi-fragile watermarks," in *Proc. SPIE International Conference on Security and Watermarking of Multimedia Contents II*, vol. 3971, No. 14, Jan. 2000
- [3] D. Zou, Y.Q. Shi, Z. Ni, W. Su, "A Semi-Fragile Lossless Digital Watermarking Scheme Based on Integer Wavelet Transform," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 16, no. 10, pp. 1294-1300, 2006.
- [4] X.Z. Zhu, A.T.S. Ho, and P. Marziliano, "A new semi-fragile image watermarking with robust tampering restoration using irregular sampling," *Elsevier Signal Processing: Image Communication*, vol. 22, Issue 5, pp. 515-528, 2007
- [5] Y. Zhu, C.T. Li, and H.J. Zhao "Structural digital signature and semi-fragile fingerprinting for image authentication in wavelet domain," in *Proc. Third International Symposium on Information Assurance and Security*, pp. 478-483, 2007.
- [6] G.J. Yu, C.S. Lu, H.Y.M. Liao, and J.P. Sheu, "Mean quantization blind watermarking for image authentication," in *Proc. IEEE International Conference on Image Processing*, vol. 3, pp. 706-709, 2000.
- [7] D. Kundur, and D. Hatzinakos, "Digital watermarking for telltale tamper proofing and authentication," in *Proc. IEEE*, vol. 87, no. 7, pp. 1167-1180, July, 1999.
- [8] D. Fu, Y.Q. Shi, and Q. Su, "A generalized Benford's law for JPEG coefficients and its applications in image forensics," in *Proc. SPIE Security, Steganography, and Watermarking of Multimedia Contents IX*, vol. 6505, pp. 1L1-1L11, 2007.
- [9] F. Benford, "The law of anomalous numbers," in *Proc. American Philosophical Society*, vol. 78, pp. 551-572, 1938.
- [10] T.P. Hill, "The significant-Digit Phenomenon", *American Mathematical Monthly*, vol. 102, pp. 322-327, 1995.
- [11] M.J. Nigrini, "I've got your number," *Journal of Accountancy*, May, 1999.
- [12] C. Durtschi, W. Hillison, and C. Pacini, "The effective use of Benford's Law to assist in detecting fraud in accounting data," *Journal of Forensic Accounting*, vol. v, pp. 17-34, 2004.
- [13] J.M. Jolion, "Images and Benford's Law," *Journal of Mathematical Imaging and Vision*, vol. 14, pp. 73-81, 2001.
- [14] F. Perez-Gonzalez, G.L. Heileman, and C.T. Abdallah, "Benford's Law in Image Processing," in *Proc. IEEE International Conference on Image Processing*, vol. 1, pp. 405-408, 2007.
- [15] G. Schaefer, and M. Stich "UCID - An Uncompressed Colour Image Database," in *Proc. SPIE, Storage and Retrieval Methods and Applications for Multimedia*, pp. 472-480, 2004