DO BIOMETRIC IMAGES FOLLOW BENFORD'S LAW?

Aamo Iorliam[1], Anthony TS Ho[1], Norman Poh[1] and Yun Q Shi[2]
[1]University of Surrey, Guildford, UK, GU2 7XH
{a.iorliam, a.ho, n.poh}@surrey.ac.uk
[2]New Jersey Institute of Technology, Newark,
NJ 07102, USA; shi@njit.edu

ABSTRACT

Tampering of biometric samples is becoming an important security concern. Tampering can take place at the sensor level (spoofing), and through the backend, e.g., replacing the template with another sample. One example of backend attack is manipulating the original biometric image, i.e. contaminating the template. We study one particular aspect of tampering: image manipulation. In the forensics literature, Benford's law has been reported to be very effective in detecting tampering of natural images. In this paper, our motivation is to examine whether biometric images will follow the Benford's law and whether or not they can be used to detect potential malicious tampering of biometric images. We find that, the biometric samples do indeed follow the Benford's law; and the method can detect tampering effectively, with Equal Error Rate (EER) of 0.55% for single compressed face images, 2.7% for single compressed fingerprint images, 4.3% for double compressed face images and 3.7% for double compressed fingerprint images.

Index Terms**—**Benford's law, forensic biometrics

1. INTRODUCTION

1.1 Motivation

Digital tampering is so rampant recently because of easy access to digital processing tools such as Photoshop [18]. The most common manipulations that can likely be applied to biometric data (e.g. face image) include: copy-paste manipulation which has an inserted region that is uncompressed, but the composite biometric face image is saved in a Joint Photographic Expert Group (JPEG) format; copy-paste manipulation which has an inserted region that is compressed and the composite biometric face image is saved in JPEG format; and inpainting manipulation of JPEG biometric face image [10]. Another great concern of biometrics is related to biometric sensors which include replay-attack and print- attack, which are all spoofing approaches to fool biometric sensors, thus resulting in sensor tampering [19]. We are however, concerned with biometric image manipulation in this paper. Biometric modalities consists of face, fingerprints, iris, retina, teeth, hand geometry, and skin when considering biological measurements, whereas when considering behavior information, the data consists of voice, gait, keyboard stroke, signature or other written scripts [14]. However most of these modalities are digitally captured and can be exposed to digital tampering.

Digital forensic techniques have been used to solve challenges in biometrics and vice-versa. Yan and Osadciw [1], combined biometrics and forensics for face recognition, whereas Meuwly and Veldhuis in [14] stated how the field of biometrics and forensic science could contribute and benefit from each other and also showed the need for new techniques in the area of forensic biometrics.

In the forensics literature, Benford's law has been reported to be very effective in detecting tampering of natural images. Benford's law of "anomalous digits" was coined by Frank L. Benford in 1938 [2], which is also described as the first digit law, considers the frequency of appearance of the most significant digit (MSD), for a broad range of natural and artificial data [3]. The Benford's law as described by Hill [4] can be expressed in the form of a logarithmic distribution, when considering the probability distribution of the first digit from 1 to 9 for a range of natural data. Naturally generated data are supposed to obey this law whereas tampered or randomly guessed data are supposed to disobey this law [5].

Since the first paper on Benford's law, there have been several developments concerning this law such as its application to financial data [6]. Considering image processing, Jolion [7], stated that this law works well on the magnitude of the gradient of an image and also for the Laplacian pyramid code. Acebo and Sbert [8] used the Benford's law on synthetic images; however in 2007 Gonzalez *et al.* [3] argued that their approach did not follow this law for many real images. Even though images in the pixel domain did not follow the Benford's law, they did observe that images when transformed to Discrete Cosine Transform (DCT) followed this law. They also showed that the generalized Benford's law could detect hidden data in a natural image [3]. Fu *et al.* [5] used this law on DCT coefficients with the aim of detecting unknown JPEG compression. Qadir *et al.* [9] analyzed the Discrete Wave Transform (DWT) coefficients using Benford's law and audited the processing history applied to JPEG2000 images where they observed a sharp peak at the digit five for some images when looking at the Benford's law curve. They also proposed the use of the law to identify unbalanced lighting in an image with the help of DWT [9]. Li *et al.* [10] used the statistical features of the first digits of individual alternate current and support vector machine to detect and locate the tampered region. By Alternate Current (AC), we mean the 63 values achieved excluding the Direct Current (DC) value as described in section 2.2. Li and Kong recently used the

978-1-4799-4370-8/14/$31.00 ©2014 IEEE

DCT coefficients to restore blood vessels patterns from JPEG compressed skin images which can further aid in forensic investigation and analysis [15]. From literature search, it appears that the use of Benford's law has not been applied to the area of biometric images/ biometric modalities.

In this paper we analyse the block-DCT coefficients of uncompressed biometric images such as face and fingerprint images and show that they closely follow the standard Benford's law, and that JPEG coefficients of these images also closely follow the generalized Benford's law as proposed by Fu *et al.* [5]. We therefore propose a novel use of the Benford's law on biometric modalities such as face images, and fingerprints to determine if such modalities are uncompressed or JPEG compressed images with the help of DCT. We also calculate the average of divergence for the fitting of the biometric face and fingerprint images to show if these data sets obey the law as used in [5].

1.2 Paper Organization

Section 2 gives a brief description of Benford's law, and Discrete Cosine Transform (DCT). Section 3 shows our experiments, data sets used and results. The potential applications of this law to biometrics are discussed in Section 4. Section 5 gives the discussion. The conclusion and future work is given in Section 6.

2. BENFORD'S LAW AND BLOCK-DCT COEFFICIENTS

2.1. Benford's Law

A typical distribution of the Benford's law can be seen in Figure 1. Therefore any data that closely follows this pattern follow the standard Benford's law.

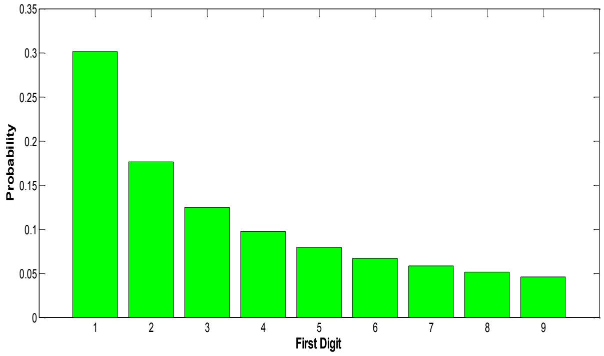


Fig. 1. Probability distribution of Benford's law

When considering the Most Significant Digit (MSD) where 0 is excluded, and the datasets satisfy the Benford's law, then the law can be expressed as Equation 1 [4].

$$p(x) = log_{10}\left(1+\frac{1}{x}\right), x= 1, 2,…,9 \qquad (1)$$

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

The generalized Benford's law which was described by Fu *et al.* [5], to closely follow a logarithmic law is defined in Equation 2.

$$p(x) = Nlog_{10}\left(1+\frac{1}{s+x^q}\right), x = 1,2,…,9 \qquad (2)$$

where N is a normalization factor which makes $p(x)$ a probability distribution. The model parameters in this case are represented by s and q which describes the distributions for different images and different compression QF's as defined in [5]. The s and q are data-dependent and they are determined by Matlab curve fitting tool box as used in [5].

2.2. Block-DCT Coefficients

The Benford's law has been studied closely for the JPEG image compression [5]. The DCT is used to transform a signal or image from the spatial domain to the frequency domain. It is however noted that, the direct conversion of a 2D spatial function *f(x,y)* into the 2D spectrum F(*u,v*) of spatial frequencies and vice-versa does not lose any information from the signal or image [11]. When considering images, the 2D DCT is used because of the 2D signals of images. According to Fu *et al.* [5], JPEG image compression is block-DCT based and has the 8 X 8 Block-DCT, Quantization, and Entropy Coding.

To achieve the first digits' probability distribution based on the DCT, the partitioning of an original uncompressed image into a non-overlapped 8 X 8 pixel blocks is first performed. This process brings about the block-DCT coefficients. A 2D DCT is applied to each block in order to convert it to a frequency space. These results into 64 values, the value at the upper-left corner is referred to as the DC Coefficient and the other 63 values are referred to as the AC coefficients. The quantization table is therefore applied to each block of the DCT coefficients [11]. After this process, the JPEG coefficients are produced. We are therefore interested in studying the probability distribution of the first digits of the block-DCT coefficients of biometric colored face images and biometric gray level fingerprint images to see whether they follow the standard Benford's law or not. We will then analyse the probability distribution of first digits of the JPEG coefficients to determine whether they also follow the generalized Benford's law or not.

3. EXPERIMENTS

The goals of the first set of experiments are to investigate the block-DCT and JPEG coefficients for both face and fingerprint images to determine whether they follow the standard Benford's law and generalized Benford's law. The second sets of the experiments are to show the performance of the study.

The divergence in [5] used here to show how data samples departs from the Benford's law is expressed in Equation 3.

$$x^2 = \sum_{i=1}^{9} \frac{(p'_i - p_i)^2}{p_i}, i = 1,2,…9 \qquad (3)$$

where p'_i is the actual first digit probability of the block-DCT coefficients of the biometric face and fingerprint images and p_i is the Benford's law as given in Equation 1.

3.1 Data Sets

We used two biometric databases for our experiment. For the face modality, we used CASIA-FACEV5 which consists of 2500 color images of 500 subjects. The face images are 16 bit color BMP files with image resolution of 640 X 480 [12]. For our experiments we use frontal cropped single face image for 80 subjects, making a total of 80 face images. We chose only the images of 80 subjects because of uniformity with our fingerprint datasets which are also 80 in number. It is however not clearly stated, whether or not this database is uncompressed; which will be further analysed in this paper. The FVC2000 has four different databases (DB1, DB2, DB3, and DB4) [13]. 80 gray fingerprint images from DB1 have been chosen for the experiment. For the face images, the luminance component is used and the probability distribution of the first digit of the block-DCT is performed on each of the face image and fingerprint image which is learnt from [5].

3.2 Results for Standard and Generalised Benford's Law

In the above experiments, we observe that the probability distributions of the first digits of the block-DCT coefficients for uncompressed face and fingerprint images closely follow the standard Benford's law as shown in Figures 2(a) and 2(b), respectively.

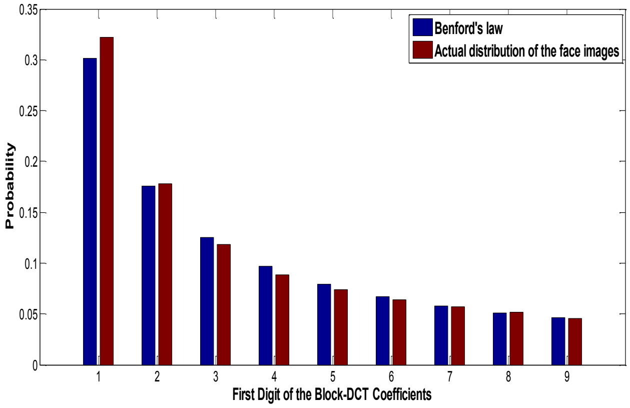


Fig.2. (a) Block-DCT coefficients of face images closely follow Benford's law

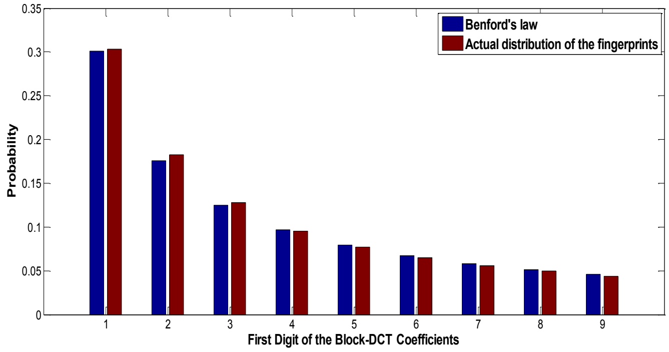


Fig.2. (b) Block-DCT coefficients of fingerprint images closely follow Benford's law

We obtained a divergence of 0.000917 and 0.000064 for the face images and fingerprint images, respectively. As can be seen from the results, the smaller the divergence the better the fitting and this clearly shows that uncompressed biometric data closely followed the standard Benford's law.

We also observe that the JPEG coefficients of the biometric face and fingerprint images closely followed the generalized Benford's law based on a Quality Factor (QF) of 100, N=1.456, q=1.47, and s=0.0372 as shown in Figures 3 (a), and in 3(b), respectively. We used QF=100 because this QF produces the best image quality when compared with other QF's. The plots also showed the standard Benford's law for comparison. We also showed how the JPEG coefficients of the biometric face and fingerprint images follow the generalized Benford's law by using the divergence formula in Equation 3, but this time the p'_i is the actual first digit probability of the JPEG coefficients of the biometric face or fingerprint images and p_i is the logarithmic law (generalized Benford's law) as used in [5]. The divergence of 0.0031 and 0.0014 were obtained for QF=100 for both the biometric face and fingerprint images respectively, which showed a good fitting when using the model parameters as used in [5]. Other QF's were also found to give a good fitting.

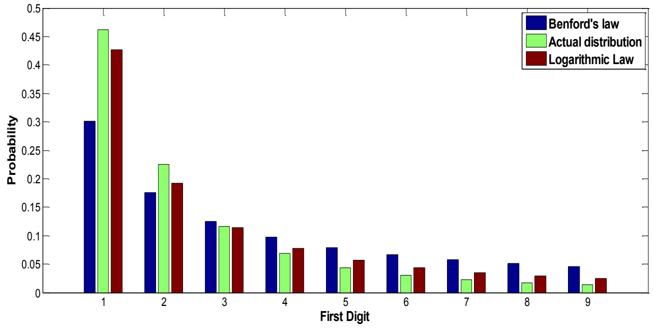


Fig.3. (a) JPEG coefficients of face images closely follow generalized Benford's law at QF=100

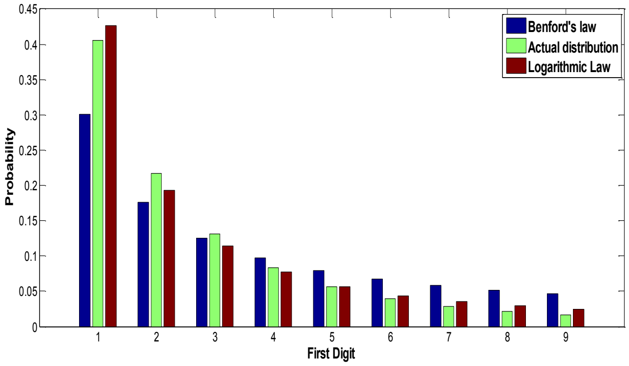


Fig.3. (b) JPEG coefficients of fingerprint images closely follow generalized Benford's law at QF=100

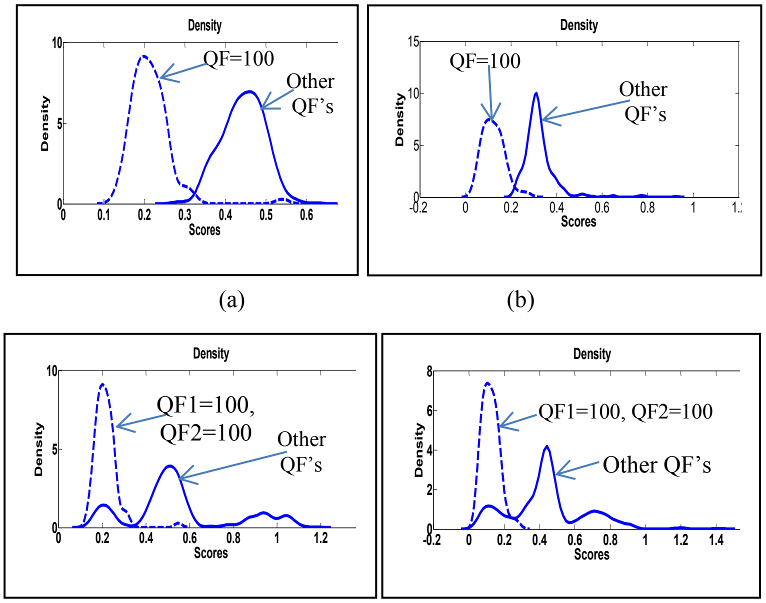
Other divergence values for uncompressed face and fingerprint images are given in Table 1 for different QF from 100 to 50 in step of -10.

TABLE 1

Quality Factor	Divergence (face images)	Divergence (fingerprint images)
100	0.052	0.0231
90	0.1313	0.0994
80	0.1625	0.1143
70	0.1663	0.1167
60	0.1638	0.1109
50	0.1701	0.1168

3.3 Results for Performance Evaluation

In order to measure the performance of this experiment, the kernel density estimation technique was used to first estimate the probability density function (pdf) from the data representing similarity scores acquired from the data sets used for this experiment [16]. This was performed for both the single compression and double compression, bearing in mind that QF=100 indicate an image with best quality in forensics. In view of this, a single compression of 100 and double compression of QF_1=100 and QF_2=100 are both considered images with the best quality. A single compression of QF_1=50 to 100 in step 10 was performed for all the data sets, and a double compression of QF_1=55, 65, 75, 85, 100 and QF_2=70, 80, 90, 95, 100 was also performed on all the data sets as shown in Figures 4(a), 4(b), 4(c), and 4(d), respectively.



(a) (b)

(c) (d)

Fig.4. The pdf of the divergence of: (a) singly compressed face images of QF=100 versus the pdf of other QF's from 90 to 50 in step -10; (b) singly compressed fingerprint images of QF=100 versus the pdf of other QF's from 90 to 50 in step -10; (c) doubly compressed face images of QF_1=100, QF_2=100 versus other QF's including QF_1=55, 65, 75, 85 and QF_2=70, 80, 90, 95; (d) doubly compressed fingerprint images of QF_1=100, QF_2=100 versus other QF's including QF_1=55, 65, 75, 85 and QF_2=70, 80, 90, 95.

To detect the performance, the detection error tradeoff (DET) was used to show how images with a single compression QF=100 and the rest of the compression were detected, and also how images with a double compression QF_1=100, QF_2=100 were detected from the other images with different quality factors [17]. In order to assess how well divergence as defined in (3) can be used to discriminate between uncompressed images and compressed images with different quality factors, we use the DET. This is because the plot can handle the different proportion between the total numbers of compressed images versus the uncompressed ones. If classification error is used, the error estimate will be dominated by the class having the larger number of samples. Using a DET curve effectively eliminates this potential source of bias in error estimation. It is particularly useful for binary classification problem such as our case.

A DET curve is a plot of False Rejection Rate (FRR) versus False Acceptance Rate (FAR) in the inverse of the standard normal deviates scales [17]. Using the compressed images as the reference, which is the target class to be detected, their respective definitions are:

$$\mathrm{FAR}(\Delta)=\frac{\#\ of\ falsely\ accepted\ \ uncompressed\ images\ at\ \Delta}{Total\ \#\ of\ uncompressed\ biometric\ images}$$

$$\mathrm{FRR}(\Delta) = \frac{\#\ of\ falsely\ rejected\ \ compressed\ images\ at\ \Delta}{Total\ \#\ of\ compressed\ biometric\ images}$$

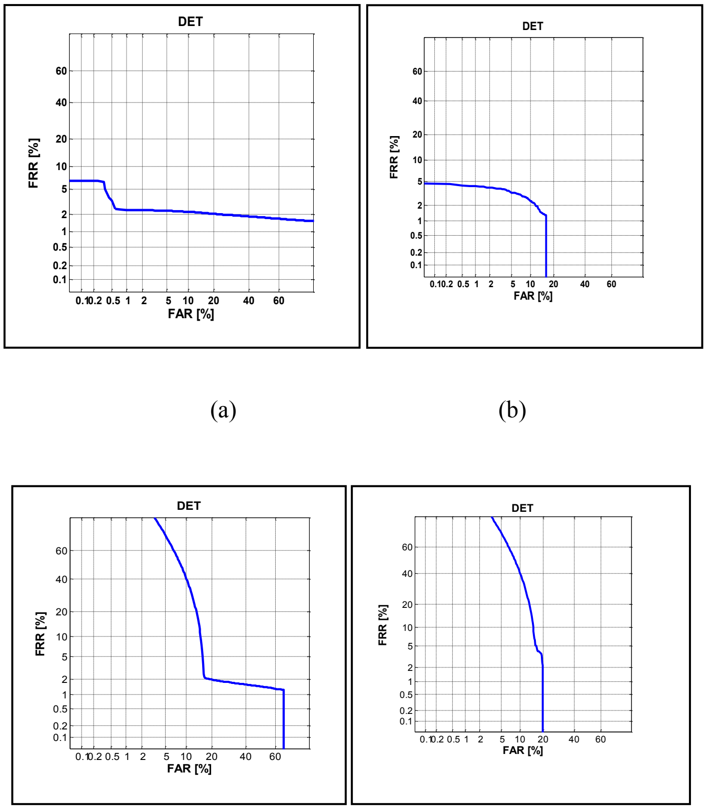
A DET curve is a plot of FRR in the y-axis versus FAR in the x-axis by varying the decision threshold such that:

$$Decision(\Delta) = \begin{cases} accept & if\ x^2 < \Delta \\ reject & otherwise \end{cases}$$

Smaller FAR and FRR values are desirable. Therefore, the DET curve of a good system should stay closer to the origin (lower left corner) of a DET curve.

Equal Error Rate (EER) is a unique operating point where FAR=FRR. This is a useful metric that summarises the entire DET curve. It is a good approximate of the Bayes error assuming that the prior probabilities of compressed versus uncompressed images are equal. Lower EER is better.

The DET curves can be seen in Figures 5(a), 5(b), 5(c), and 5(d).



(a)

(b)

(c)

(d)

Fig.5. DET curve for: (a) singly compressed face images; (b) singly compressed fingerprint images; (c) doubly compressed face images; (d) doubly compressed fingerprint images

4. APPLICATIONS TO BIOMETRIC IMAGES

For the fact that, the probability distributions of the first digits of the block-DCT coefficients of biometric face and fingerprint images which are uncompressed closely follow the standard Benford's law and the JPEG coefficients of the same data sets closely follow the generalized Benford's law considering QF's from 50 to 100 in step 10, with corresponding values of N, s and q as used in [5]. The Benford's law is very useful when no clear information is given about biometric face or fingerprint images like in the case of CASIA-FACEV5. This law can assist in determining uncompressed, and JPEG face and fingerprint images which has a potential to be very useful in the field of forensic biometrics for biometric images information.

Table 2 below shows the EER for various compressions of data sets as seen below.

Table 2

Biometric Data	EER
Single compressed face images	0.55%
Single compressed fingerprint images	2.7%
Double compressed face images	4.3%
Double compressed fingerprint images	3.7%

5. Discussion

Benford's law does not require training and the above observations shows it works well in detecting uncompressed and JPEG biometric images.

The EERs from Table 2 shows that it is easier to detect singly compressed biometric images than the doubly compressed biometric images by the Benford's law. In overall, the EERs are low, therefore showing higher separability.

6. CONCLUSION AND FUTURE WORK

This paper showed that the probability distributions of the first digits of the block-DCT coefficients of Biometric faces and fingerprint images when uncompressed closely follow the standard Benford's law. Moreover, we found that JPEG coefficients of biometric face and fingerprint images also closely followed the generalized Benford's law. Our experiments strongly supported our conjecture and show that we have succeeded in bringing a technique from forensics into biometrics, therefore contributing to the field of forensics biometrics.

In our future work, we will investigate detection of tampered biometric face and fingerprint image regions. We also wish to extend our research to other biometric modalities such as iris and speech.

Other areas of our future work will include:

1. To investigate if face images of different poses, uncontrolled conditions will follow the Benford's law.
2. Use the classification approach to detect tampering and combine the probability of tampering with the matching score.

7. ACKNOWLEDGEMENT

Aamo Iorliam would like to acknowledge the financial support of Benue State University, Makurdi, Nigeria for the sponsorship of his PhD scholarship. Dr Poh also received funding from the BEAT project (grant no. 284989) to carry out this work. The authors thank CASIA for the use of their database. Portions of the research in this paper use the CASIA-FACEV5 collected by the Chinese Academy of Sciences' Institute of Automation (CASIA).

8. REFERENCES

[1] Y. Yan, and L.A. Osadciw, "Bridging Biometrics and Forensics," EECS, Proc. SPIE 6819, Security, Forensics, Steganography, and Watermaking of Multimedia Context X, 68190Q, 2008.

[2] F. Benford, "The law of anomalous numbers," Proc. of the American Philosophical Society, vol.78, pp. 551-572, 1938.

[3] F.P. Gonzalez, G.L. Heileman, and C.T. Abdallah, "Benford's Law in Image Processing," Proc. IEEE International Conference on Image Processing, (ICIP 2007), pp. 405-408, 2007.

[4] T.P. Hill, "A statistical derivation of the significant-digit law," Statistical Science papers (10), pp. 354-363, 1996.

[5] D. Fu, Y. Q. Shi, and W. Su, "A generalized Benford's Law for JPEG coefficients and its applications in image forensics," Proc. SPIE 6506, 1L1-1L11, 2007.

[6] M. Nigrini, *A taxpayer compliance application of Benford's Law*, Journal of the American Taxation Association, vol. 1, pp.72–91, 1996.

[7] J. M. Jolion, *Images and benford's law*, Journal of Mathematical Imaging and Vision, 14(1), pp. 73–81, 2001.

[8] E. Acebo, and M. Sbert, "Benford's law for natural and synthetic images," Proc. of the First Workshop on Computational Aesthetics in Graphics, Visualization and Imaging, *2005.*

[9] G. Qadir, X. Zhao, A.T.S. Ho, and M. Casey, "Image Forensic of Glare Feature for Improving Image Retrieval Using Benford's Law," IEEE International Symposium on Circuits and Systems (ISCAS)*,* pp. 2661-2664, 2011.

[10] X. H. Li, Y.Q. Zhao, M. Liao, F.Y. Shih, and Y.Q. Shi, *Detection of Tampered region for JPEG images by using mode-based first digit features,* EURASIP Journal on Advances in Signal Processing 2012, 2012:190.

[11] N. Efford, (2000) *Digital Image Processing. A Practical Introduction Using Java^{TM},* Pearson Education, 2000.

[12] CASIA-FACEV5 (2010). *Biometric Ideal Test.* Available: http://www.idealtest.org/dbDetailForUser.do?id=9

[13] FVC2000 (2000). *Fingerprint Verification Competition Databases.* Available: http://bias.csr.unibo.it/fvc2000/databases.asp

[14] D. Meuwly, and R. Veldhuis, "Forensic biometrics: From two communities to one discipline," Proceedings of the BIOSIG 2012, International Conference of the Biometrics Special Interest Group-(BIOSIG), 2012.

[15] X. Li, and W.K. Kong, "Restoring Blood Vessel patterns for JPEG Compressed Skin Images for Forensic Analysis," IEEE International Workshop on Information Forensic and Security, 2013.

[16] E. Parzen, *On Estimation of a Probability Density Function and Mode,* The annals of Mathematical Statistics, 33(3):1065, 1962.

[17] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET Curve in Assessment of Detection Task Performance," Proc. Eurospeech '97, vol. 4, pp. 1899-1903, 1997.

[18] H. Farid, "A Survey of Image Forgery Detection," IEEE Signal Processing Magazine, Papers 26(2), pp.16-25, 2009.

[19] C.M. Roberts, "Biometric Attack Vectors and Defences, Danish Biometrics," Computers and Security, Vol.26, no.1, pp. 14-25, 2007.