

Image Forensic of Glare Feature for Improving Image Retrieval Using Benford's Law

Ghulam Qadir*, Xi Zhao, Anthony TS Ho and Matthew Casey

University of Surrey

Guildford, UK, GU2 7XH

*Email: G.Qadir@surrey.ac.uk

Abstract— Whilst it is sometimes essential that a scene is well lit before image capture, too much light can cause exposure or glare-based problems. Typically, glare is introduced to images when the camera is pointed towards the light source, and results in a visible distortion in the image. In this paper, we analyse and identify images that contain the ‘glare’ property using the empirical Benford’s Law. The experiment is performed on 1338 images, and extracts discrete wavelet High High (HH), High Low (HL) and Low High (LH) sub bands as raw data. The significant digit from each coefficient of all sub bands is then calculated. We then analyse the probability of occurrence of large digits against smaller digits to detect anomalies. All images containing these anomalies are further analysed for the identification of additional salient features. This analysis is performed in accordance with the Benford’s Law plot and the help of probability intensity histogram and divergence. Our results indicate that 142 images have irregular Benford’s Law curves. For most images, the irregularity occurs at the 5th digit. After visual examination, we have found the unbalanced light and high level of brightness in these images. To measure the intensity of light in an image, we compute the probability histogram of gray levels. These results also correlate with the irregular peak identified from the Benford’s Law curves. In addition, the divergence is then computed, which shows the deviation between the actual Benford’s Law curve and the Benford’s Law graph of an image. Our proposed technique is novel and has a potential to be an image forensic tool for quick image analysis.

I. INTRODUCTION

The field of digital image forensics is striving hard to restore the lost trust in digital content. Images, which are becoming a part of today’s life, are growing vulnerable to digital forgery [1]. The readily available image manipulation software, such as Photoshop and GIMP are primarily responsible for this. Application of state of the art digital forensic methods for examination and authentication of digital images is not possible when images are available in such huge quantity. In order to trace or extract images of certain features or characteristics, an automated system is useful which will save precious time lost while scanning each image manually. There are systems already in place doing image retrieval with the help of feature extraction algorithms such as edge detection, etc.

A statistical law called, Benford’s Law [2]–[5], has been used previously in accounting forensics to detect fraudulent data by Nigirini [5]. Similarly, another study in the field of psychology noticed a peak at the digit 5 when people are asked to choose from a tampered set of data [6]. In recent years it has attracted the attention of image processing experts.

In 2005, Acebo *et al* [7], showed how light intensity in an image can be used to determine if an image is genuine or computer generated. Unlike [7], we have applied DWT before calculating 1st digit probabilities and analysed images in the frequency domain, which is better to separate edge details from low frequencies. Similarly, Fu *et al* [8] have applied Benford’s Law on DCT coefficients in order to detect unknown JPEG compression. In our previous work [9], we have analysed DWT coefficients using Benford’s Law and audited the processing history applied to JPEG2000 images. We have noticed a sharp peak at digit 5 in Benford’s Law curve for some images, which became the basis for this paper.

Therefore, we propose a novel use of the Benford’s Law to identify unbalanced lighting in an image with the help of DWT. In this paper, we will analyse the irregularity in the Benford’s Law curve that appears for images with a certain feature. This feature is ‘glare’, when appeared in or near, a field of view induces unbalanced light that makes an image comparatively brighter in various parts. In effect, the image loses its visual quality. To determine if an image possesses unbalanced lighting, we will apply a single level of DWT to an image and compute its 1st digit probabilities of DWT coefficients. Images with irregular Benford’s Law curves are then identified and analysed further. The divergence, adopted from Fu *et al* [8] and Acebo *et al* [7], which shows how much the Benford’s Law curve of an image deviates from the actual Benford’s Law, is also calculated. Furthermore, the intensity histogram, used for measuring the strength of gray levels in various sections of an image, will compare with its 1st digit probabilities.

The rest of the paper is organised as follows: Section II, will briefly describe Benford’s Law, Discrete Wavelet Transform and intensity distribution for gray level images. Section III illustrates our proposed glare image detection method, followed by our experimental results and analysis in Section IV. This is followed by conclusion and future work in Section V.

II. BACKGROUND

A. Benford’s Law

Benford’s Law, also known as the 1st digits law and significant digits law, was introduced by Frank Benford in 1938 [3]. Then, Hill [4] expressed Benford’s Law as a logarithmic distribution, for analysis of the probability distribution of the 1st digit (1 – 9) of the number from natural data in

statistics. The distribution for Benford's Law can be expressed by Equation 1.

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

Where x is the 1st digit of the number and $p(x)$ is the probability distribution of x . A typical probability distribution of Benford's Law is shown in Figure 1. Any peaks or irregularities in the curve mean that the data has been tampered with or is unnatural [5], [6], [7].

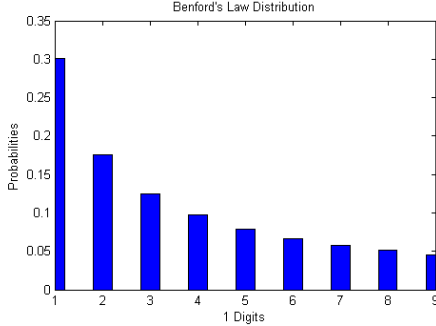


Fig. 1. Probability Distribution of Benford's Law

B. Discrete Wavelet Transform

As shown in Figure 2, Discrete Wavelet Transform (DWT) decomposes an image into a power of single resolution levels by applying a collection of low-pass and high-pass filters onto the image in vertical and horizontal directions. This resolution level consists of four sub-bands, which are Low Low (LL), Low High (LH), High Low (HL) and High High (HH). The LL sub-band contains most energy and therefore can be decomposed further [11]. The rest of the sub-bands contain edge details.

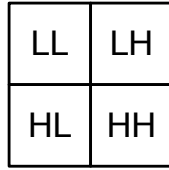


Fig. 2. DWT Decomposition

C. Intensity Distribution

Digital images (and particularly gray scale images) can be described in finite terms and discrete quantities, which are pixel and amplitude [12]. A pixel, which contains a gray level, is arranged into two dimensional orders to form a digital image. The intensities are spread randomly over the entire image. In order to calculate the strength of various intensity levels or gray levels, a histogram is computed. As Gonzalez

and Woods [12] discussed, the probability of occurrence of intensity level r_k in an image is approximated by Equation 2.

$$p_r(r_k) = \frac{n_k}{MN}, k = 0, 1, 2, \dots, L - 1 \quad (2)$$

Where M and N are the total number of pixel in an image, n_k is the number of pixels that have intensity r_k , and L is the possible intensity levels in the image (256 for 8 bit image).

III. PROPOSED METHOD

In this section, we explain the proposed glare image detection process which uses the Benford's Law model. As shown in Figure 3, the test image is first decomposed into LL, LH, HL and HH sub-bands through single level DWT (2, 3 and 4 levels of DWT can also be applied, but could decrease the detection rates). In our experiment, we used the wavelet with Daubechies 9/7 filter, which is standard for lossy JPEG2000 compression. Secondly, the DWT coefficients of LH, HL and HH sub-bands are extracted (LL sub-band is not used as it contains highest amount of energies and is more sensitive to low frequencies [11]), and then the probability distribution of the 1st digits of these DWT coefficients is obtained by using the Benford's Law model, as shown in Equation 1. Finally, in order to detect glare, these 1st digit probabilities undergo an image retrieval process, as per the pseudocode below.

```

if  $p(x) < p(x + 1)$  then
  This 1st digit's probability does not obey Benford's Law
  This image could have glare
end if

```

In addition, with the purpose of further investigate the relationship between the 1st digit probability and intensity distribution, the intensities of the image are calculated using Equation 2, which estimate the strength of gray levels present in an image in the spatial domain. Moreover, the divergence is also calculated, which measures how much deviation of the 1st digit probability graph from the actual Benford's Law curve. The divergence adopted from Acebo *et al* [7] and Fu *et al* [8], is shown in Equation 3.

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

Where p'_i is the actual 1st digit probability of the DWT coefficients and p_i is the Benford's Law from Equation 1.

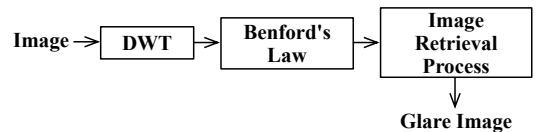


Fig. 3. Proposed glare image detection process

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Our proposed glare image detection method has detected 142 (128, 123 and 121 if 2, 3 and 4 levels of DWT applied, respectively) out of 1338 images from UCID [10], which had irregular Benford's Law curves. Figures 4(b) and 6(b) show two irregular Benford's Law curves of images 'Statue 1' (Figure 4(a)) and 'Street' (Figure 6(a)), which compared with the actual Benford's Law curve. By comparing with the regular Benford's Law curve of image 'Statue 2' (Figure 5(a)) visually, the unbalance light caused by glare can also be found from Figures 4(a) and 6(a), which possesses extra amount of brightness in some parts. Moreover, the intensity distribution histograms of a normal image 'Statue 2', in Figure 5(c), is somewhat smoother than the glare images 'Statue 1' and 'Street', in Figures 4(c) and 6(c)).

In image 'Statue 1', due to its glare affect, the top boundary of the image can not be found visually, which also correlates with its 1st digit probabilities, which do not obey Benford's Law. Especially, its irregular peak at digit 5 in Figure 4(b), which is nearly five times higher. Similarly, the top right boundary of image 'Street' is also difficult to find visually and the amount of its unbalanced light is much less than the image 'Statue 1'. Furthermore, in Figure 6(b), the irregular peak at digit 5 from its 1st digit probabilities is only twice as high. In addition, the intensity distribution histogram in Figure 6(c), is smoother than image 'Statue 1' in Figure 4(c). Therefore, we conclude that the amount of unbalanced light in an image could affect the magnitude of the digit 5 in the probabilities. In contrast, the distribution of gray levels in the intensity histogram in Figure 5(c) for image 'Statue 2', is spread across a wider range of gray levels smoothly. In Figure 5(b), the image obeys Benford's Law and there appears to be no significant brightness in the image. The divergence between these 1st digit probabilities and Benford's Law is shown in Table I.

TABLE I
DIVERGENCE FROM BENFORD'S LAW CURVE

Image	Divergence
Statue 1	1.7590
Statue 2	0.0024
Street	0.0919

From Table I, we observe that the divergence of image 'Statue 2' is showing the best fit to Benford's Law, at 0.0024. In comparison, the worst fit at 1.759 is image 'Statue 1' and followed by image 'Street' at 0.0919. Furthermore, by deducing common artefacts that are present in both sets of images that follow or do not follow Benford's Law, the only feature left is glare (extra brightness). Hence, we can conclude that the glare feature in an image results in an irregularity in its 1st digit probabilities.

In addition, we also simulated an attack by appending artificial glare into the images. The attack is implemented by using "lens flare filter" function with the brightness increased to 135% in Adobe Photoshop CS5 image editing software.

From this experiment, the results showed that both regular and irregular 1st digit probabilities are not influenced by the artificial glare, and therefore our method could also identify whether there is natural or artificial glare present in an image.

V. CONCLUSION AND FUTURE WORK

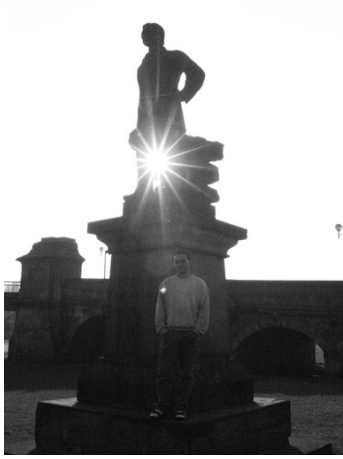
In this paper, we presented a method of extracting images with glare (such as unbalanced lighting) out of the bulk of images in DWT domain using Benford's Law. Any irregularities in an image could be detected as peaks via the Benford's Law curves. The peaks were mainly located at digit 5. We found via experiments that 142 images have this irregularity from 1338 images, and the amount of glare feature (unbalanced light) in an image could affect the magnitude of digit 5 in 1st digit probabilities. In order to further analyse the irregular Benford's Law curve, the intensity distribution histogram of test images was also calculated for comparison. We found a good correlation between the 1st digit probabilities and intensity distribution of its gray level. The divergence was also calculated between the 1st digit probability curve and Benford's Law, such as image 'Statue 2', at 0.0024. Our method could also identify whether there is natural or artificial glare present in an image. In future work, more tests will be performed on images under different natural light conditions. Further research with Fourier Transform instead of DWT for comparative analysis will also be considered.

ACKNOWLEDGMENT

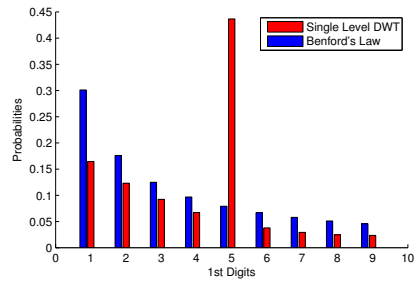
Ghulam Qadir would like to acknowledge the financial support of COMSATS Institute of Information Technology, Islamabad, Pakistan for the sponsorship of his PhD scholarship.

REFERENCES

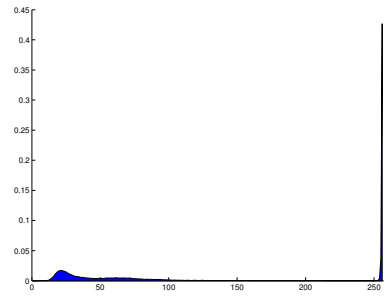
- [1] Farid, H., *A Survey of Image Forgery Detection*, IEEE Signal Processing Magazine, Papers 26(2), pp. 16-25, 2009.
- [2] Newcomb, S., *Note on the frequency of use of the different digits in Natural Numbers*, American Journal of Mathematics, Papers 4(1), pp. 39-40, 1881.
- [3] Benford, F., *The law of anomalous numbers*, Proc. Amer.Phil.Soc.78, pp. 551-572, 1938.
- [4] Hill, T. P., *A statistical derivation of the significant-digit law*, Statistical Science, Papers (10), pp. 354-363, 1996.
- [5] Nigrini, M.J., *I've got your number*, Journal of Accountancy, May, 1999.
- [6] Burns, B. D., *Sensitivity to Statistical Regularities: People (Largely) Follow Benford's Law*, in: N.Taatgen and H.Van Rijn (eds.), Proc. Thirty-First Annual Conference of the Cognitive Science Society, Cognitive Science Society, Austin, TX, pp. 28722877, 2009.
- [7] Acebo, E. and Sbert, M., *Benford's law for natural and synthetic images*, Computational Aesthetics in Graphics, Visualizations and Imaging, 2005.
- [8] Fu, D., Shi, Y. Q. and Su, Q., *A generalized Benford's Law for JPEG coefficients and its applications in image forensics*, Proc. SPIE 6506, 1L1-1L11, 2007.
- [9] Qadir, G., Zhao, X., and Ho, A. T. S., *Estimating JPEG2000 compression for image forensics using Benford's Law*, Proc. SPIE 7723, 77230J, 2010.
- [10] Schaefer, G. and Stich, M., *UCID - An Uncompressed Colour Image Database*, Proc. SPIE, Storage and Retrieval Methods and Applications for Multimedia 2004, pp. 472-480, San Jose, USA.
- [11] Farid, H., and Lyu, Siwei., *Higher-order wavelet Statistics and their Application to Digital Forensics*, Proc. IEEE Workshop on Statistical Analysis in Computer Vision, 2003.
- [12] Gonzalez, R. C., and Woods, R. E., *Digital Image Processing*, Pearson Prentice Hall, 2008.



(a)



(b)

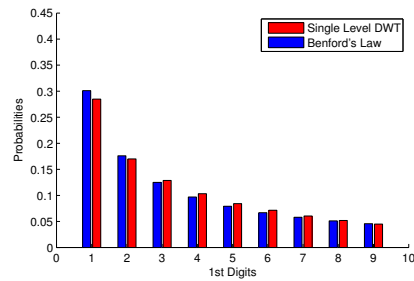


(c)

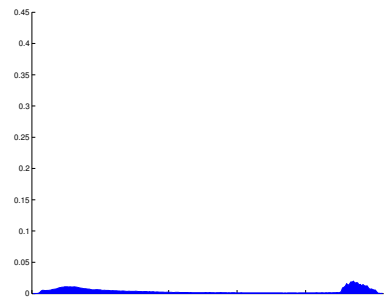
Fig. 4. (a) Statue 1, (b) 1st digit probabilities & Benford's Law, (c) Probability Histogram



(a)



(b)

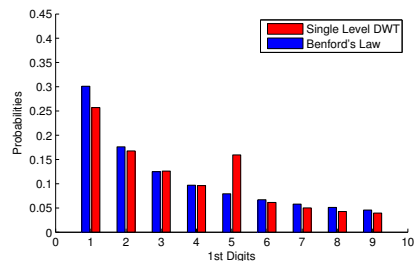


(c)

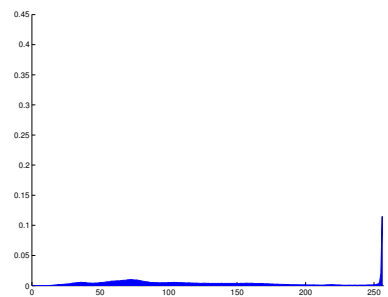
Fig. 5. (a) Statue 2, (b) 1st digit probabilities & Benford's Law, (c) Probability Histogram



(a)



(b)



(c)

Fig. 6. (a) Street, (b) 1st digit probabilities & Benford's Law, (c) Probability Histogram