UNDERSTANDING BENFORD'S LAW AND ITS VULNERABILITY IN IMAGE FORENSICS

Jingwei Wang, Byung-Ho Cha, Seong-Ho Cho, and C.-C. Jay Kuo

Signal and Image Processing Institute and Ming Hsieh Department of Electrical Engineering
University of Southern California, Los Angeles, CA 90089-2564

ABSTRACT

In this paper, we attempt to shed light on Benford's law from the viewpoint of probability theory and point out its limitation in image forensic applications. First, we consider a generalized form of Benford's law and relate it to a random variable of a certain probability density function such as the generalized Gaussian or Laplacian function. Then, we examine the application of the generalized Benford's law to image forensic applications and point out such an application is vulnerable to the histogram manipulation attack.

Index Terms— Benford's law, generalized Benford's law, image forensics, image authentication, histogram manipulation.

1. INTRODUCTION

Digital images have been used in a wide variety of applications nowadays. Since digital images can be easily modified with editing software, their authentication has become a hot topic in image forensic analysis. The application of Benford's law to image authentication has been explored by quite a few researchers in recent years. Jolion [1] showed that the magnitude of the gradient of an image obeys this law. Acebo and Sbert [2] demonstrated that light intensities in natural images follow the Benford's law under certain constraints. Perez-Gonzales *et al.* [3] established a generalized Benford's Law by keeping two terms of the Fourier expansion of the probability density function (pdf) of the data in the modular logarithmic domain. Fu and Shi [4] showed that the first digit of block DCT coefficients follow Benford's law.

Despite the studies as mentioned above, the power and limitation of Benford's law has not yet been fully understood. In this work, we shed some light on Benford's law from the viewpoint of probability theory and relate it to a random variable of a certain pdf (*e.g.*, the generalized Gaussian or Laplacian function). Then, we examine the application of the generalized Benford's law to image forensic applications from this new viewpoint. Finally, we point out the weakness of its application with exemplary simulations. Basically, it is vulnerable to the histogram manipulation attack.

The rest of this paper is organized as follows. We first review related previous work, examine a generalization form of Benford's law and relate the law to a random variable of a certain pdf in Sec. 2. Next, we study the application of generalized Benford's law to image forensics and point out its vulnerability to the histogram manipulation attack in Sec. 3. Finally, concluding remarks and future research extensions are described in Sec. 4.

2. BENFORD'S LAW: INTERPRETATION AND GENERALIZATION

2.1. Interpretation

Benford's law was observed and introduced by physicist Frank L. Benford in 1938 [5]. It states that the frequency $B(d)$ of the first significant digit (FSD) (*i.e.*, the first non-zero digit) of an N-digit number (base 10) follows a certain logarithmic function in form

$$B(d) = \log_{10}\left(1 + d^{-1}\right) \text{ where } d = 1,\ 2,\ \ldots,\ 9. \tag{1}$$

It is observed that many empirical data in the real world follow Benford's law such as stock market indices, tax returns and Fibonacci numbers. Diaconis [6] provided an explanation to this phenomenon and pointed out that, for a data set whose statistical distribution is uniform in the log scale, the distribution of its FSD conforms to Benford's law.

We can formalize the above statement in terms of probability theory. Let us consider x as uniform distribution which has the pdf $p_X(x)$ is given by $1/N$ if we assume $0 \le x \le N$. Then, we can define a random variable y which conforms to log scale uniform distribution as $y = 10^x$. Through the transformation of random variable x, the pdf of y can be expressed as

$$p_Y(y) = \frac{1}{N \ln 10} \cdot \frac{1}{y}. \tag{2}$$

The random variable y, which is generated from the pdf $p_Y(y)$, is represented by

$$y = \sum_{i=1}^{N} d_i \cdot 10^{N-i} \tag{3}$$

and

$$d_i = \begin{cases} 1,\ 2,\ \ldots,\ 9, & i = 1 \\ 0,\ 1,\ \ldots,\ 9, & 2 \le i \le N \end{cases}$$

where d_i is the ith digit of y in base 10. Then, we can get the distribution of FSD $p_D(d_i = d)$ from the pdf $p_Y(y)$ by

$$p_D(d_i = d) = \frac{1}{N \ln 10} \left(\int_{d \cdot 10^{i-1}}^{(d+1)\cdot 10^{i-1}} \frac{1}{y} dy \right). \tag{4}$$

Generally speaking, the distribution of FSD $p_D(d)$ of y with $10^0 \leq y < 10^N$ can be written by

$$p_D(d) = \sum_{i=1}^{N} p_D(d_i = d), \quad d = 1,\ 2,\ \ldots\ 9. \tag{5}$$

As a result, the distribution of FSD $p_D(d)$ strongly conforms to the Benford's law expressed by

$$p_D(d) = B(d), \tag{6}$$

which means two probability density functions are identical.

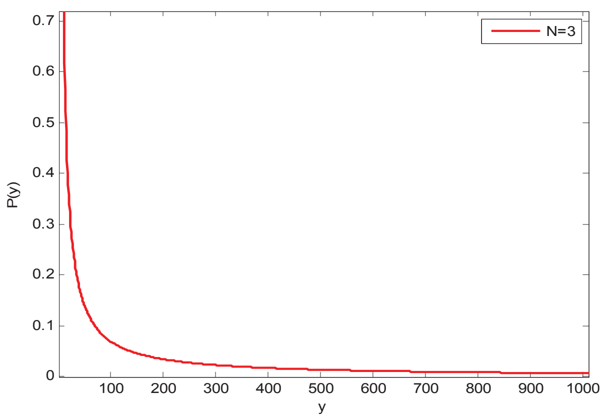

Fig. 1. The pdf $p_Y(y)$ when N is fixed to 3.

Consider the case $N = 3$ as an example. The pdf of $p_Y(y)$ is shown when $N = 3$ is given in Fig. 1. From Eq. (5), we can derive the distributions of FSD $p_D(1)$. The distribution of FSD $p_D(1)$ is

$$\begin{aligned} p_D(1) &= \frac{1}{3 \ln 10} \left(\int_{1}^{2} \frac{1}{y} dy + \cdots + \int_{100}^{200} \frac{1}{y} dy \right) \\ &= \log_{10} 2. \end{aligned} \tag{7}$$

We can recognize that $p_D(1)$ conforms Benford's law. Also, we can check that $p_D(i),\ i = 2,\ \ldots,\ 9$ conforms Benford's law from the same calculation with Eq. (7).

2.2. Generalization

General image data can be represented in various transform domains, for example, the discrete cosine transform (DCT). The distribution of DCT coefficients can be modelled as the Laplacian distribution [7] or the Gaussian distribution [8].

The distribution of FSD $p_D(d)$ of the generalized Laplacian distribution or Gaussian distribution does not obey the original Benford's law exactly, and the generalization of Benford's law is required. In the following, we consider the generalization of Benford's law for the generalized Laplacian or Gaussian distribution.

The generalized Benford's law $B_g(d)$ is given by

$$B_g(d) = \alpha_1 \cdot log_{10} \left(\alpha_3 + \frac{1}{d^{\alpha_2}} \right) \tag{8}$$

where the set of model parameters $\alpha_i, i = 1,\ 2,\ 3$ are decided depending on the statistical distribution models.

Table 1. Model parameters for fitting two distributions.

Parameters	α_1	α_2	α_3
Standard Laplacian distribution	1.05	1.352	1.061
Standard Gaussian distribution	1.08	2.55	1.15

We test the fitting results of our proposed generalized Benford's law to the distortions of standard Laplacian and standard Gaussian. In order to show the difference clearly, we use a linear scale to plot their results in Fig. 2. According to the standard statistical model, those parameters are shown in Table 1. The statistical distribution results based on our proposed generalized Benford's law are shown in Fig. 2. It clearly shows that proposed generalized Benford's law fits the results of different statistical distributions very well.

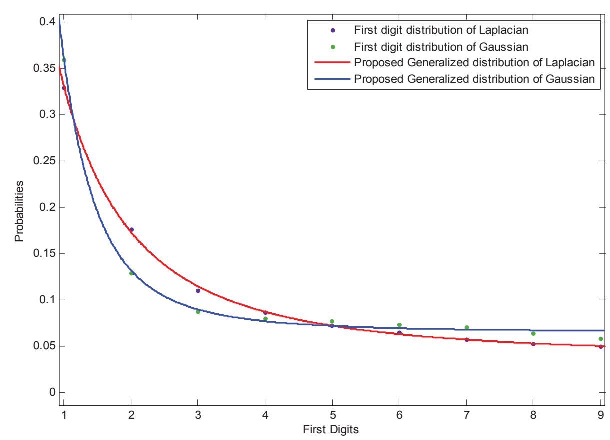

Fig. 2. The statistical distribution results based on our proposed generalized Benford's law in a linear scale.

With this precise statistical distribution representation, the generalized Benford's law can be a useful tool in image forensic applications. For example, we can differentiate cartoon images from natural images [2], apply for image steganogaphy which tells cover images apart from stego images [3], and distinguish uncompressed images from compressed images [4].

3. VULNERABILITY OF BENFORD'S LAW IN IMAGE FORENSIC APPLICATIONS

Although the generalized Benford's law can be widely used in image forensics, it has limitations when it encountered with the malicious attacks. Forensic applications aim at identifying malicious modifications from attackers. After some modifications are processed, we perform the forensic analysis to identify those modifications. In this situation, malicious attackers may have some knowledge about forensic mechanisms, which can be used for hindering forensic analysis. We point out the weakness of Benford's law in this point of view.

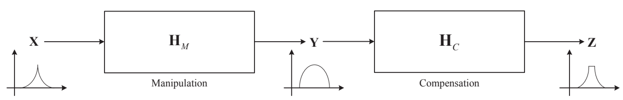

Fig. 3. The block diagram of system model for vulnerability.

We introduce the concepts of manipulation and compensation systems as shown in Fig. 3. In this system model, we have two systems represented by $\mathbf{H}_M$ and $\mathbf{H}_C$, respectively. $\mathbf{H}_M$ represents a manipulation system such as double compression or dithering, while $\mathbf{H}_C$ denotes a compensation system such as equalization or scaling.

A random signal vector $\mathbf{x}$ with the pdf $p_X(x)$ goes through the system $\mathbf{H}_M$, and we have a random signal vector $\mathbf{y}$ with the pdf $p_Y(y)$. Then, we have a random signal vector $\mathbf{z}$ with the pdf $p_Z(z)$ from a signal vector $\mathbf{Y}$ through the system $\mathbf{H}_C$. All signal vectors have different statistical distributions. If we observe the distribution of FSD of $\mathbf{y}$, this distribution does not follow the Benford's law, but the distribution of FSD of $\mathbf{z}$ follows the Benford's law.

Based on this system model, we investigate cartoon images and compensation mechanisms. Then, we show the vulnerability of Benford's law when images are compensated by common image processing techniques. All simulations are performed on the block DCT domain in this section.

3.1. Vulnerability of Benford's Law in Cartoon and Natural Image

We study the generalized Benford's law for two cartoon images and one natural image to see different characteristics of three images, which are cartoon type I (left), cartoon type II (middle), and natural image (right) as shown in the bottom of Fig. 4. All three images can be distinguished clearly via the human eye. The distributions of FSD of block DCT coefficients for three images are shown in Fig. 4. We can see that the type I cartoon image does not follow the Benford's law.

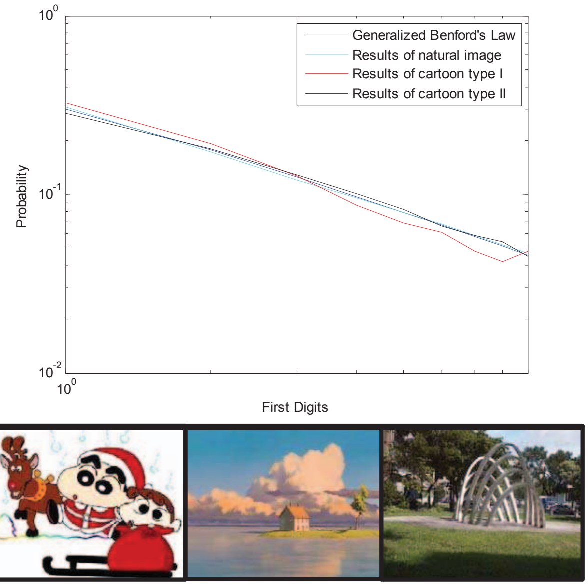

Fig. 4. The distributions of the corresponding FSD in a log scale for cartoon image type I (left), cartoon image type II (middle), and natural image (right).

However, type II cartoon image and natural image conform the Benford's law. We can give clear separation between type I cartoon image and natural image based on the distribution of FSD. However, if we manipulate type I cartoon image to type II cartoon image, the difference can no longer be distinguished by the Benford's law. If we assume there is a manipulation from distributions of natural image to those of type I cartoon image, it can be denoted by $\mathbf{H}_M$. Consequently, a compensation from type I cartoon image to type II cartoon image can be represented by $\mathbf{H}_C$.

3.2. Vulnerability of Benford's Law in Compression and Dithering

We study forensic applications with (1) JPEG double compression and (2) dithering. We observe the distributions of FSD of block DCT coefficients after both manipulations are applied, and then we observe them after certain compensation by equalization or scaling is performed.

The double compression is represented by $\mathbf{H}_M$, and the equalization in a log domain is denoted by $\mathbf{H}_C$. From Fig. 5, we can see that the Benford's law is violated after double JPEG compression. However, if we compensate the double compressed images using histogram equalization in a log domain, the compensated result conforms to the Benford's law. It indicates that the Benford's law can be directly affected by the manipulation of image histogram.

The dithering is represented by $\mathbf{H}_M$, and the re-scaling is denoted by $\mathbf{H}_C$. From Fig. 6, we can see that if we re-

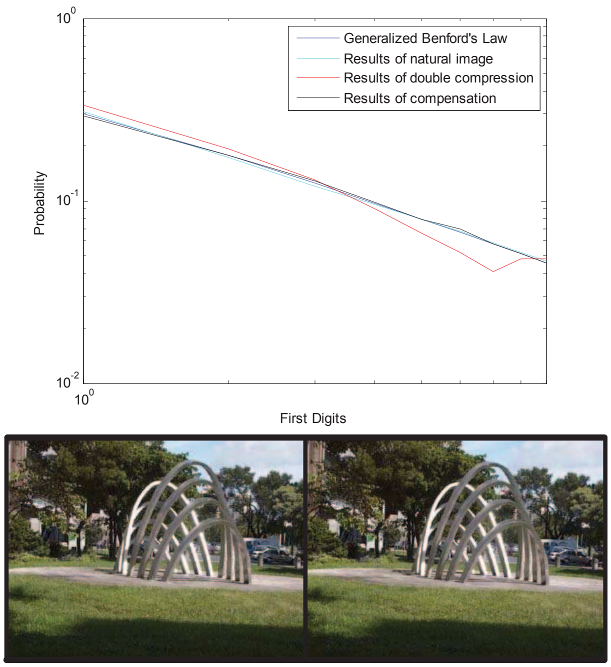

Fig. 5. The distribution of FSD in a log scale for double compressed image (left) and equalized image (right).

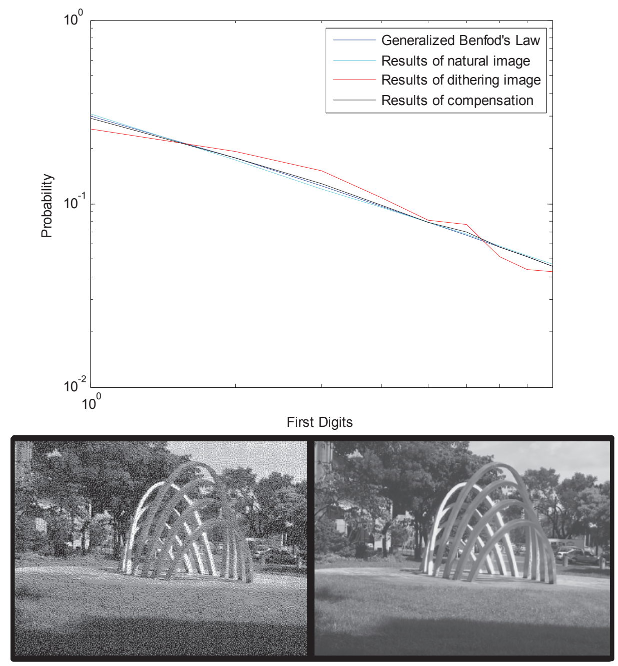

Fig. 6. The distribution of FSD in a log scale for dithered image (left) and re-scaled image (right).

cover the bit planes of the dithered image, the dithered image gradually conforms to the Benford's Law. If we take the image histogram into account, we can notice that the image histogram tends to be transformed into mixture Gaussian distributions. It can be known by the generalized formula we deducted above. The more scale an image has, the less gap between the Benford's law and distributions of images.

4. CONCLUSION AND FUTURE WORK

We investigated Benford's law for image forensic applications in this paper. We noticed that there is a relationship between Benford's law and image statistical models. To point out the vulnerability of Benford's law, we first derived a generalization of Benford's law. Finally, we performed exemplary simulations to show the weakness of Benford's law for image forensic applications. In the near future, we will extend our current research to compensate the weakness of Benford's law and develop the way to enhance the forensic performance for the forensic applications.

5. REFERENCES

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

[2] E. D. Acebo and M. Sbert, "Benford's law for natural and synthetic images," in *Proc. Int'l Workshop Computational Aesthetics in Graphics, Visualization, and Imaging*, Girona, Spain, May 2005, pp. 73–81.

[3] F. Perez-Gonzalez, G. Heileman, and C. T. Abdallah, "Benford's law in image processing," in *Proc. IEEE Int'l Conf. Image Processing*, San Antonio, TX, USA, September 2007, pp. 169–176.

[4] D. Fu, Y. Q. Shi, and W. 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*, San Jose, CA, USA, Janurary 2007.

[5] F. Benford, "The law of anomalous numbers," *Proc. Amer. Phil. Soc.*, vol. 78, pp. 551–572, 1938.

[6] P. Diaconis, "The distribution of leading digits and uniform distribution mod 1," *The Annals of Probability*, vol. 5, pp. 72–81, 1977.

[7] F. Muller, "Distribution shape of two-dimensional dct coefficients of natural images," *Electronic Letter*, vol. 29, pp. 1935–1936, October 1993.

[8] K. A. Birney and T. R. Fischer, "On the modeling of dct coefficients of natural images," *IEEE Tran. on Image Processing*, vol. 4, pp. 186–193, Feburary 1995.