Computational Aesthetics in Graphics, Visualization and Imaging (2005)
L. Neumann, M. Sbert, B. Gooch, W. Purgathofer (Editors)

Benford's Law for Natural and Synthetic Images

E. Acebo, and M. Sbert

Institut d'Informàtica i Aplicacions, Universitat de Girona, Spain

Abstract

Benford's Law (also known as the First Digit Law) is well known in statistics of natural phenomena. It states that, when dealing with quantities obtained from Nature, the frequency of appearance of each digit in the first significant place is logarithmic. This law has been observed over a broad range of statistical phenomena. In this paper, we will explore its application to image analysis. We will show how light intensities in natural images, under certain constraints, obey this law closely. We will also show how light intensities in synthetic images follow this law whenever they are generated using physically realistic methods, and fail otherwise. Finally, we will study how transformations on the images affect the adjustment to the Law and how the fitting to the law is related to the fitting of the distribution of the raw intensities of the image to a power law.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computing Methodologies]: Computer GraphicsPicture/Image Generation; I.4.8 [Computing Methodologies]: Image Processing and Computer VisionScene Analysis

1. Introduction

Image statistics is a growing field in image analysis [ERT01], [SO01], [AW05], with many potential applications. We contribute to this field by studying the fitting of images (both natural and synthetics) to Benford's Law. This law (also known as the First Digit Law) is well known in statistics of natural phenomena. It states that, when dealing with quantities obtained from Nature, the frequency of appearance of each digit in the first significant place is logarithmic. This law has been observed over a broad range of physical phenomena, from city populations to river lenghts and flows. It has been applied to fraud detection in tax paying, to the design of computers, to the analysis of roundoff errors and as a statistical test for naturalness. We will show in this paper how natural images, under certain constraints, follow closely this law. We also show how synthetic (computer generated) images follow this law whenever they are generated using physically realistic methods, and fail otherwise. On the other hand, a transformation on the image, such as filtering, will affect the adjustment to the law. We finally discuss how the fitting to Benford's Law is related to the fitting of raw intensities to a power law.

The rest of the paper is organized as follows. In section 2 we review the Benford's law, in section 3 we study its applicability to images, both synthetic and natural, in section 4 we present an explanation for the obtained results and finally in section 5 we present our conclusions and future research.

2. Benford's Law

2.1. The uneven distribution of digits in Nature

Suppose that we have a table with the populations of all the villages and towns in the world. Count how many of these numbers begin with the digit 1, how many with the digit 2 and so on. Contrary to what we can intuitively expect, the relative frequency of the populations starting with each of the nine digits is not the same. Actually, we will find that about 30% of populations begin with the digit 1, and only about 5% with the digit 9. That phenomenon is not particular of populations, a vast amount of "natural" quantities, ranging from river lengths and flows to molecular weights of chemical compounds and stock prices, exhibit the same uneven distribution of first digits. It is not new. It was observed by Simon Newcomb in 1881 and then rediscovered by Frank Benford in 1938.

Benford's Law (also known as logarithmic significant digit distribution), in its most general form (base 10, D_i means the ith significant digit) tells that for all positive integers k, all $d_1 \in \{1,2,\ldots,9\}$ and all $d_j \in \{0,1,2,\ldots,9\}$, $j = 2,\ldots,k$, probability P is given by

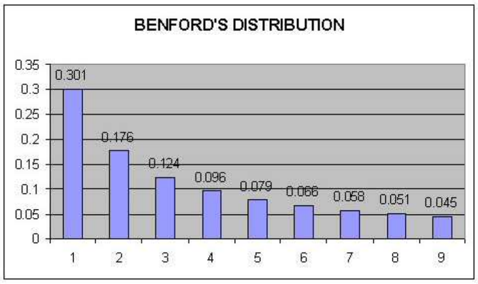

Figure 1: *First digit distribution according to Benford's Law.*

$$P(D_1 = d_1, \ldots, D_k = d_k) = \log_{10}[1 + (\sum_{i=1}^{k} d_i \cdot 10^{k-i})^{-1}] \quad (1)$$

or equivalently

$$P(\text{mantissa} \leq \frac{t}{10}) = \log_{10} t \quad (2)$$

for $t \in [1 \ldots 10)$.

Hence the first significant digit law:

$$P(\text{first significant digit} = d) = \log_{10}(1 + \frac{1}{d}) \quad (3)$$

We can see in the graph in Fig.1 the values of the probabilities of each of the nine possible first digits. A perhaps surprising corollary to the general law is that significant digits are not independent, that is, the probability for the nth significant digit to take a certain value depends on the values of the $n-1$ leading significant digits [Hil96].

Some interesting properties of the logarithmic distribution are the following:

Let X_1, X_2 be random variables with logarithmic significant digit distribution, then

- $K \cdot X_1$ for $K > 0$
- $1/X_1$
- $X_1 + X_2$
- $X_1 \cdot X_2$
- X_1^K for $K > 0$

also have logarithmic significant digit distribution.

Also, let $X_1, X_2, \ldots, X_n$ be random variables, then, under certain (rather general) conditions

$$\prod_i X_i \quad (4)$$

tends to have logarithmic significant digit distribution when n tends towards infinity.

Benford's law has been so far applied to the fields of computer design:

- Data compression [Sch88]
- Floating point computing optimization [FT86], [Knu69], [BB85]

and to test for "naturalness":

- Model validation [Var72], [NW95]
- Tax fraud detection [Nig96], [NM97]

It has also been applied in gambling [Che81].

2.2. Explanations for Benford's Law

While Benford's law unquestionably applies to many situations in the real world, a satisfactory explanation has been given only recently through the work of Hill [Hil96]. An extensive account of the early efforts to explain the phenomenon can be found in [Rai76]. Perhaps the first important step in the right direction was the realization that Benford's law is applied to data that are not dimensionless (so the numerical values of the data depend on the units). If there is a universal probability distribution over such numbers, then it must be invariant under scale changes. If somehow Benford's Law has to be of universal applicability, it has to hold independently of the units used in the measurements. If you have a set of random variables following Benford's Law and you multiply them by some constant, the resulting values will also obey Benford's Law (after all, "God is not known to favor either the metric system or the English system" [Rai76]). It turns out that Benford's Law is the only scale-invariant probability distribution for significant digits (see [Hil96] for a deep analysis of this point). So, if natural quantities have to follow a given probability law for significant digits, it has to be Benford's Law. The question about why those quantities would have to follow such a fixed law remains open, however.

The most convincing explanation of the phenomenon, by now, comes from Hill [Hil96]. The main result of his work is a sort of central-limit-like theorem for significant digits which says, putting it in the author's own plain words, "if probability distributions are selected at random, and random samples are then taken from each of these distributions in any way so that the overall process is scale (or base) neutral, then the significant digit frequencies of the combined sample will converge to the logarithmic distribution". Two key concepts in the above statement are those of scale and base neutrality. Let's focus on the former. According to Hill, a sequence of random variables $X_1, X_2, \ldots$ has scale-neutral mantissa frequency if

$$n^{-1} \mid \#\{i \leq n : X_i \in S\} - \#\{i \leq n : X_i \in sS\} \mid \to 0 \text{ a.s.} \quad (5)$$

for all $s > 0$ and all $S \in \mathcal{M}$, where $\mathcal{M}$ is the sub-sigma algebra of the Borels where the probability measure P is defined (see [Hil96] for details):

$$S \in \mathcal{M} \Longleftrightarrow S = \bigcup_{n=-\infty}^{\infty} B \cdot 10^n \quad (6)$$

for some Borel $B \subseteq [1,10)$.

We will return to it later in the discussion section, where we will make use of the theorem to justify the fitting of a broad class of images to Benford's Law attending to the shape of their histograms.

3. Benford's Law and images

Our aim in this section is to study to what extent and in what aspects images, both syntethic and real, obey Benford's Law. This should allow us to design new methods for image analysis and classification based in this previous study. Application of Bendford's Law to image analysis is quite innovative. To our knowledge, there is only one other attempt in this way [Jol01]. In that paper, however, the author quickly discards the idea of pixel intensity values of general images to follow Benford's Law and focuses on the gradient of these values, instead. We will study the matter from another perspective. Obviously not all the images obey Benford's Law, but in this paper we will show how a great variety of natural and artificial images do, and will characterize them in terms of the shape of their histograms.

3.1. Benford's Law and synthetic physically realistic images

Radiosity and ray-tracing provide us with physically realistic images [DBB03]. In Radiosity three radiosity values (RGB) are computed for each patch (or polygon of the scene mesh) in the scene. In ray tracing three intensity values are computed for each pixel in the screen. These values closely follow the logarithmic first digit distribution in several scenes that we have tested. This can be seen in the set of images in Fig.2 and Fig.3. Scenes in Fig.2 contain 49124 (top) and 22718 (bottom) patches, respectively. The quality of the fitting is measured with the χ^2 divergence:

$$\sum_{i=1}^{9} \frac{(f_i - \log_{10}(1+\frac{1}{i}))^2}{\log_{10}(1+\frac{1}{i})} \qquad (7)$$

where f_i is the relative frequency of i as first digit in the set of values (radiosities or intensities).

3.2. Benford's Law and synthetic non-physically realistic images

When non-physically realistic rendering methods are used in addition with radiosity or ray tracing methods (ambient occlusion or obscurances [IKSZ03], extended ambient term [CNS00], textures) the resulting values diverge from Benford's Law. Thus, Benford's Law could be used as a test for physical realism in synthetic image rendering. In Fig.4 we see the discrepancy between Benford's Law and two texture images, obtained by PovRay [Pov].

3.3. Benford's Law and real images

We can't work with real images, only with pictures of real images. Digital pictures represent an imperfect measurement of natural phenomena. We face potential data corruption from:

- Noise
- Over/underexposure
- Discretization
- Interpolation
- Gamma correction
- Retouching

In order to overcome these problems as much as possible we will work with raw images (no interpolation, no gamma correction, no retouching, 16 bits per channel). Not all "real" images obey Benford's Law, see for instance Fig.5. However, they tend to do it quite often and quite well. (See Figure 6). In Fig.7 we show a photograph of a painting fitting also Benford's Law. In Fig.8 we show the effect on the fitting of a flash light, and in Fig.9 we show the effect of applying different filters to the images. Fig.9 top-left corresponds to the original raw, unfiltered image. Fig.9 top-right is the result of applying to the previous image the PhotoShop contrast and brightness autobalancing, in Fig.9 bottom-left we applied the PhotoShop high-pass filter, and finally in Fig.9 bottom-right we applied the PhotoShop equalizing filter.

4. Discussion

A broad class of synthetic and real images tend to obey Benford's Law. Those that better fit this law seem to have multiple heterogeneous interacting objects (from the point of view of lighting), see Fig.2,3& 6(bottom). Images reflecting only a small part of a scene or a detail do not fit so well or do not fit at all, see Fig.5. The logs image in Fig.6(middle) is obviously an exception to this rule. We can, however, try to give a more objective justification to the phenomenon. Recall from section 2 Hill's theorem, which states that if probability distributions are selected at random, and random samples are then taken from each of these distributions in any way so that the overall process is scale (or base) neutral, then the significant digit frequencies of the combined sample will converge to the logarithmic distribution. We can deduce from this that if we take samples from a single distribution and they result to be distributed in a scale neutral way in the sense of (5) then we can expect the samples to follow the logarithmic distribution. In other words, samples taken from a scale neutral probability distribution will follow Benford's Law. One such probability distribution is $p(x) = 1/x$. In effect we have

$$\int_a^b \frac{1}{x} dx = \int_{ka}^{kb} \frac{1}{x} dx \qquad (8)$$

for $a, b, k \in \Re$, $0 < a \leq b$, $k > 0$, and this implies, in lack of a rigorous proof, scale neutrality in the sense of (5).

Now consider the histograms of the images in the paper.

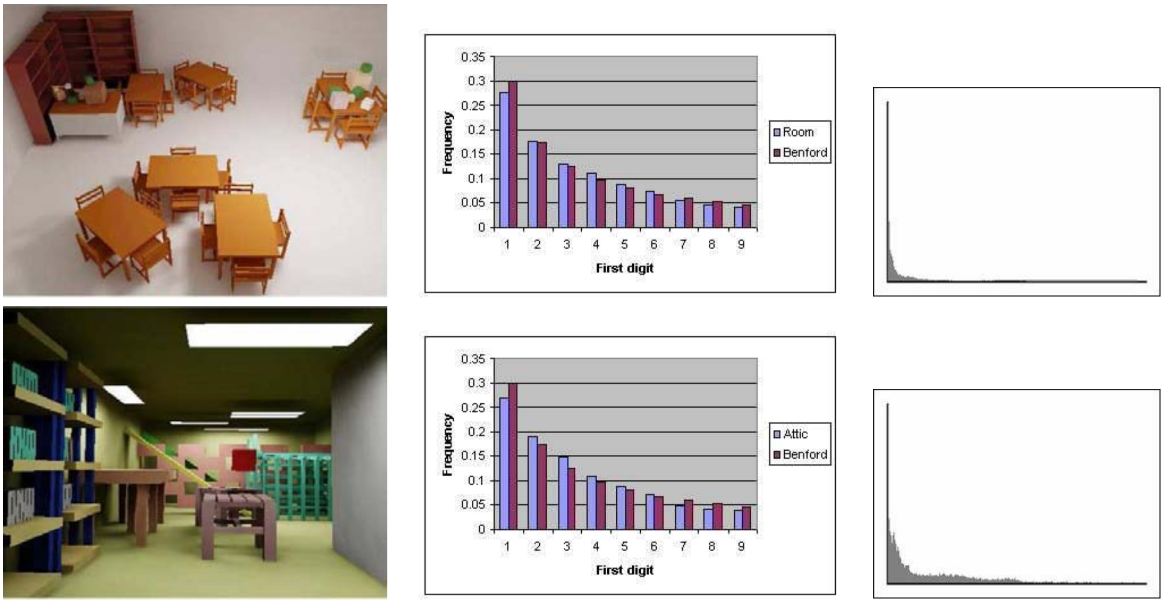

Figure 2: *Left: radiosity images. Middle: graphs demonstrate the fitting of left image with Benford's law. For the top image the fitting corresponds to* χ^2*=0.00703, and for the bottom one* χ^2*=0.01549. Right: Histograms. Images courtesy of Francesc Castro and Roel Martinez.*

They plot frequencies against intensity values of pixels (or patches in the case of radiosity), so we can see them as probability distributions of pixel (patch) intensities. By the above discussion, histograms with shape similar to the $1/x$ function will correspond to images whose pixels values are more scale-neutrally distributed. We expect those images to follow closely Benford's Law. And that is the case, indeed, as we can see in Figs.2,3, 6, 7& 8.

We can draw some additional conclusions. On the one hand, Benford's Law tends not to hold well in images obtained by means of non-physically realistic rendering methods. On the other hand, in real images, the fit with the law is very sensitive to the application of filters and retouching algorithms, see Fig.9.

5. Conclusions and future work

A broad class of synthetic and real images tend to obey Benford's Law. The images that fit the law better seem to be those with multiple heterogeneous interacting (from the point of view of lighting) objects. Benford's Law tends not to hold in images obtained by means of non physically realistic rendering methods. In real images, the fit with the law is very sensitive to the application of filters and retouching algorithms.

We envisage the following applications:

- Hardware/Software design: digital camera sensors, graphic cards, graphic algorithms
- Data compression. (Not much, about half a bit per floating point number in FP16)
- Test for "naturalness" in synthetic images

Further analysis of when and why images follow Benford's Law is needed. We intend to further study specialized image types, including astronomical and medical images, as well as other types of data (do natural sounds follow Benford's Law?).

References

[AW05] AWATE S. P., WHITAKER R. T.: Higher-order image statistics for unsupervised, information-theoretic, adaptive, image filtering. In *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition 2005* (2005). 1

[BB85] BARLOW J., BAREISS E.: On roundoff error distributions in floating point and logarithmic arithmetic. *Computing 34* (1985), 325–347. 2

[Che81] CHERNOFF H.: How to beat the massachusetts numbers game. *Math. Intel. 3* (1981), 166–172. 2

[CNS00] CASTRO F., NEUMANN L., SBERT M.: Extended ambient term. *Journal of Graphic Tools 4*, 5 (2000), 1–7. 3

[DBB03] DUTRE P., BALA K., BEKAERT P.: *Advanced Global Illumination*. AK Peters, USA, 2003. 3

[ERT01] ERIK REINHARD P. S., TROSCIANKO T.: *Natural image statistics for computer graphics*. Research

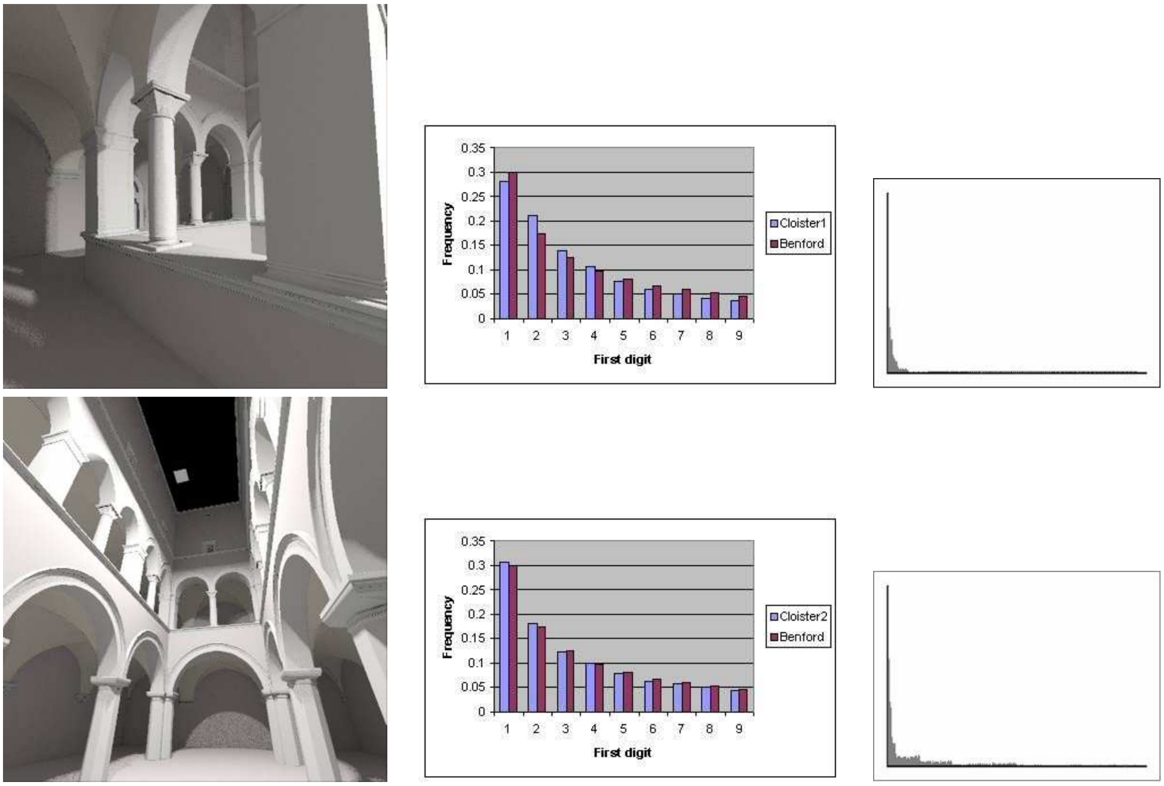

Figure 3: *Left: ray-tracing images. Middle: graphs demonstrate the fitting of left image with Benford's law. For the top image the fitting corresponds to* χ^2*=0.01714, and for the bottom one* χ^2*=0.00094. Right: Histograms. Images courtesy of Ignacio Martin.*

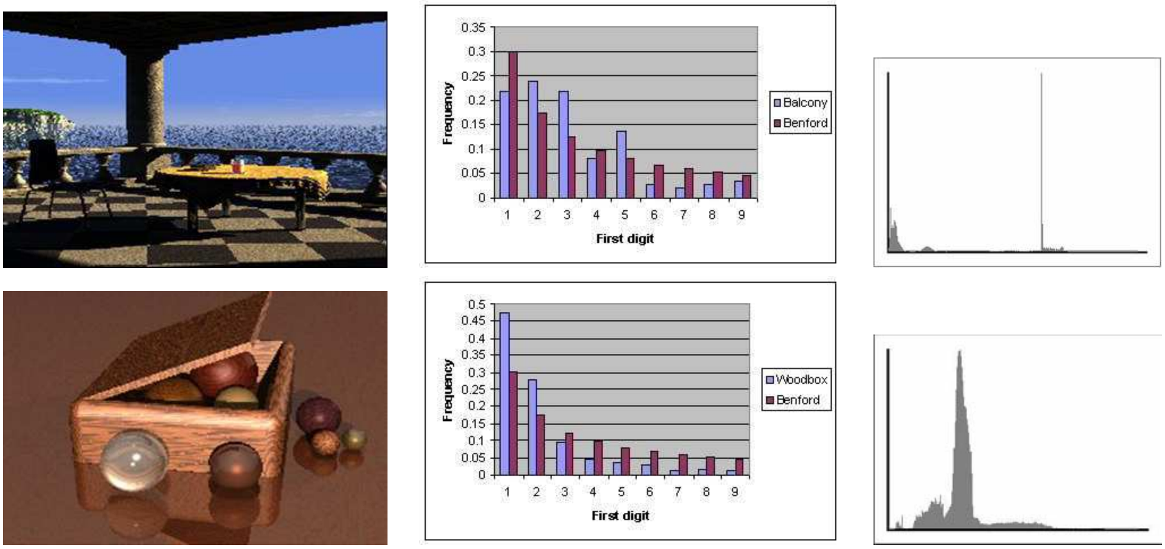

Figure 4: *Left: textured images. Middle: graphs demonstrate the unfitting of left image with Benford's law. For the top image the fitting corresponds to* χ^2*=0.31708, and for the bottom one* χ^2*=0.22037. Right: Histograms.*

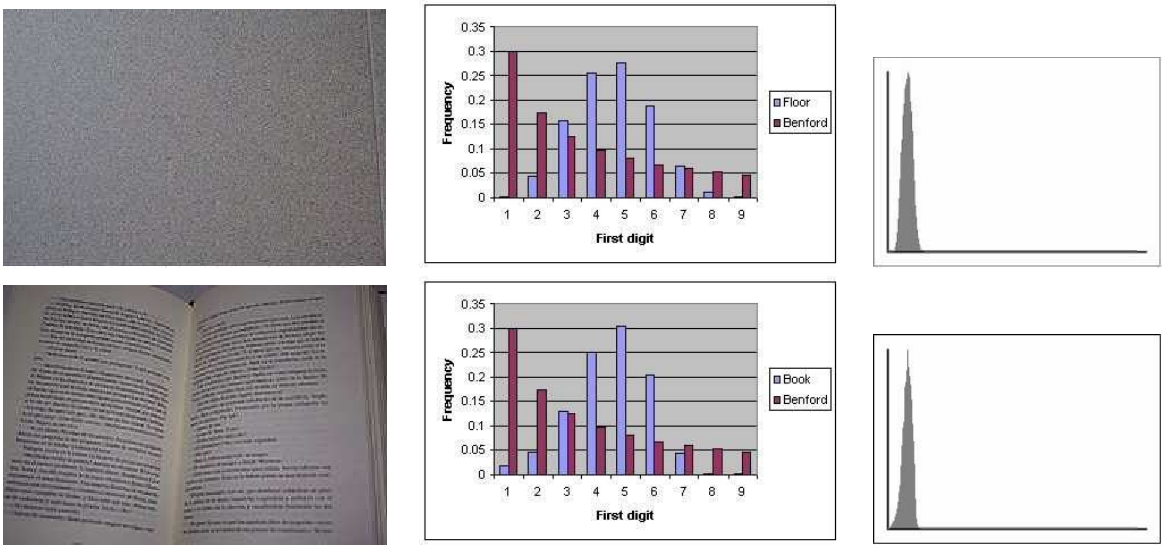

Figure 5: *Left: real images. Middle: graphs demonstrate the unfitting of left image with Benford's law. For the top image the fitting corresponds to* χ^2*=1.44835, and for the bottom one* χ^2*=1.62312. Right: Histograms.*

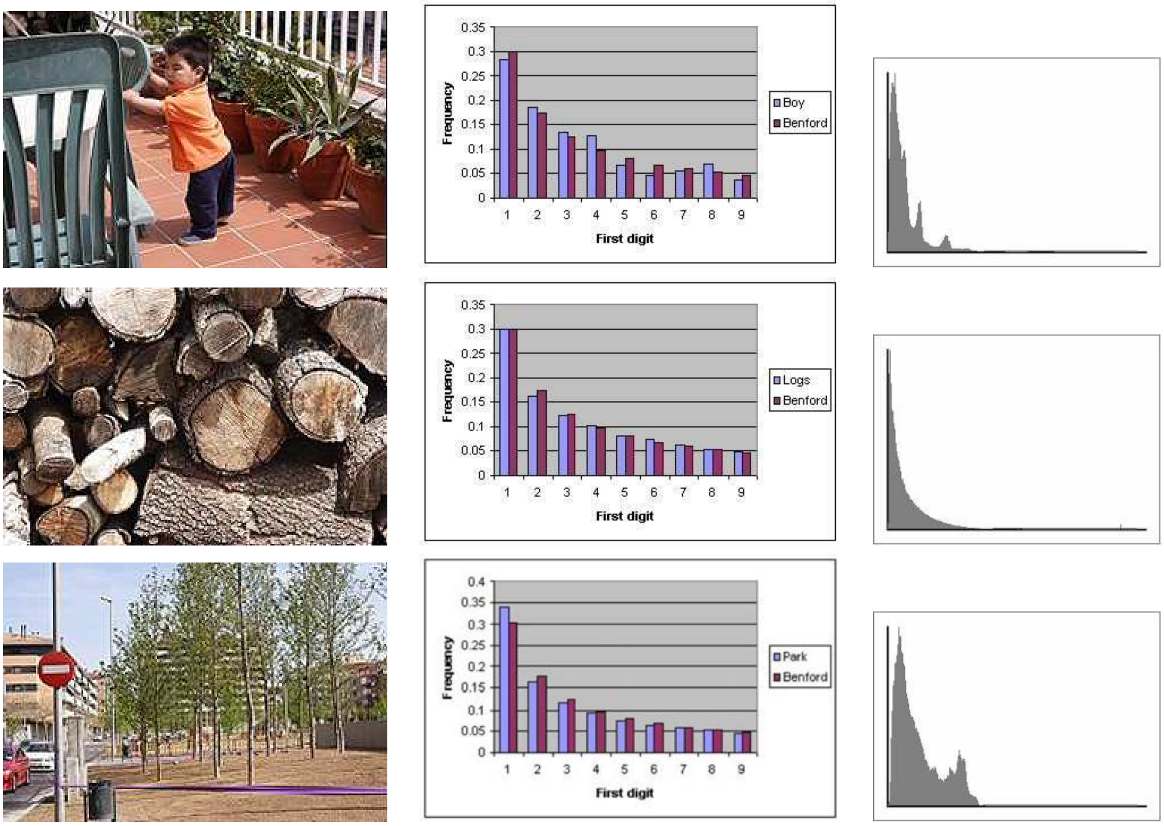

Figure 6: *Left: real images. Middle: graphs demonstrate the fitting of left image with Benford's law. From top to bottom the fitting corresponds to* χ^2*=0.02922, 0.00245, 0.00820, respectively. Right: Histograms.*

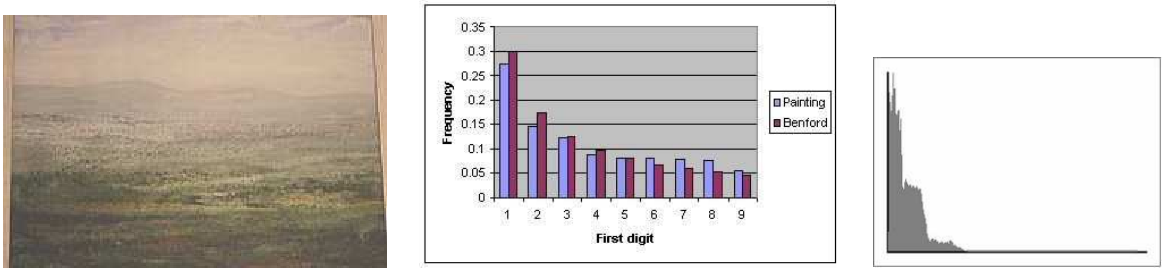

Figure 7: *Left: photograph of a painting. Middle: graph demonstrate the fitting of left image with Benford's law χ^2= 0.02998. Right: Histogram.*

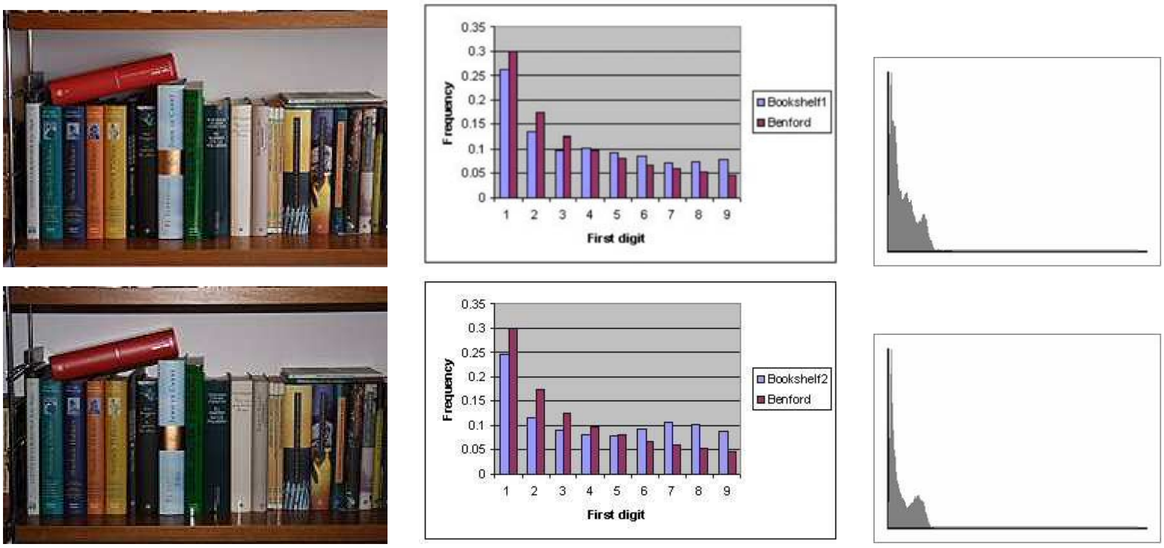

Figure 8: *Top/bottom rows correspond to a photograph (left images) taken without/with flash light, respectively. Middle: graphs demonstrate the fitting of left images with Benford's law. For the top image the fitting corresponds to χ^2=0.06433, and for the bottom one χ^2=0.18942. Right: Histograms.*

Report UUCS-01-002, School of Computing, University of Utah, Salt Lake City, Salt Lake City, UT 84112 USA, 2001. 1

[FT86] FELDSTEIN A., TURNER P.: Overflow, underflow and severe loss of significance in floating-point addition. *IMA J. Numerical Analysis 6* (1986), 241–251. 2

[Hil96] HILL T.: A statistical derivation of the significant-digit law. *Statistical Science 10*, 4 (1996), 354–363. 2

[IKSZ03] IONES A., KRUPKIN A., SBERT M., ZHUKOV S.: Fast realistic lighting for video games. *IEEE Computer Graphics & Applications* (2003). 3

[Jol01] JOLION J. M.: Image and the benford's law. *J. Math. Imaging Vision 14* (2001), 73–81. 3

[Knu69] KNUTH D.: *The art of computer programming. Vol. 2*. Addison-Wesley. New York, 1969. 2

[Nig96] NIGRINI M. J.: Note on the frequency of use of the diferent digits in natural numbers. *The White Paper* (1996). 2

[NM97] NIGRINI M. J., MITTERMAIER L. I.: The use of benford's law as an aid in analytical procedures. *Auditing: A Journal of Practice and Theory 16* (1997), 52–67. 2

[NW95] NIGRINI N., WOOD W.: Assessing the integrity of demographic tabulated data, 1995. 2

[Pov] The persistence of vision ray-tracer, http://www.povray.com. 3

[Rai76] RAIMI R.: The first digit problem. *American Mathematical Monthly 83* (1976), 521–538. 2

[Sch88] SCHATTE P.: On mantissa distributions in computing and benford's law. *J. Inf. Process. Cyber. EIK 24* (1988), 443–455. 2

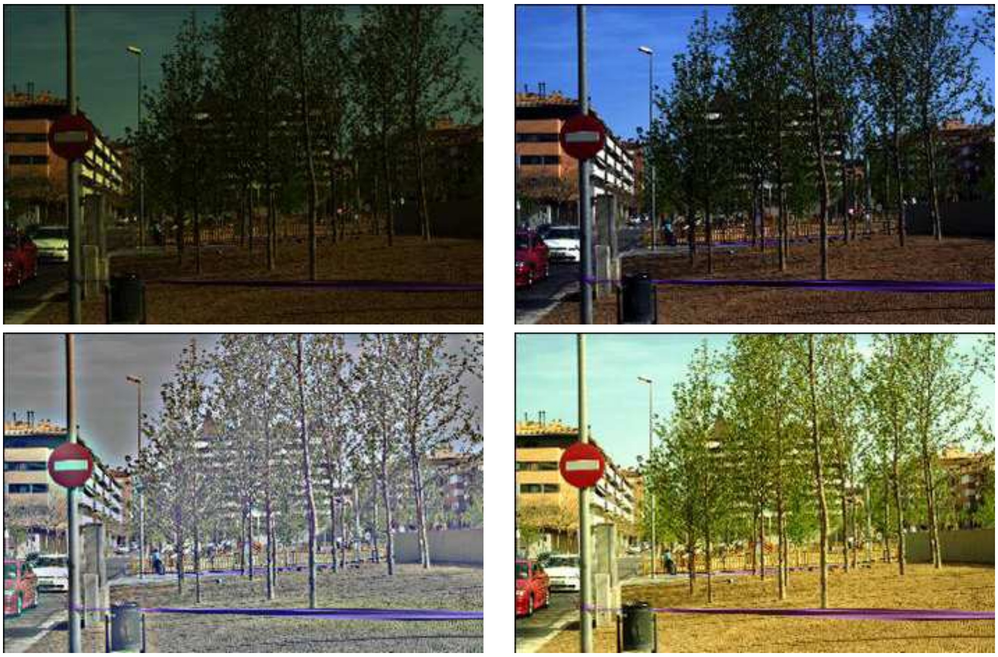

Figure 9: *Results of applying three different filtering methods to the same raw image (top-left). Top-right is the result of applying to the previous image the PhotoShop contrast and brightness autobalancing, in bottom-left we applied the PhotoShop pass-high filter, and in bottom-right we applied the PhotoShop equalizing filter.From left to right and top to bottom,* χ^2*=0.03254, 0.05648, 0.28120, 0.22083, respectively*

[SO01] SIMONCELLI E. P., OLSHAUSEN B. A.: Natural image statistics and neural representation. *Annu. Rev. Neurosci. 24* (2001), 193–216. 1

[Var72] VARIAN H.: Benford's law. *Amer. Statician 23* (1972), 65–66. 2