Analysis of Benford's Law in Digital Image Forensics

Neetu Singh
Electronics and Communication Department
Jaypee Institute of Information Technology,
Noida, India
neetu.singh@jiit.ac.in

Rishab Bansal
Electronics and Communication Department
Jaypee Institute of Information Technology,
Noida, India
rishabrcf90@gmail.com

***Abstract*— Digital Image Forensics aims at uncovering and analyzing the underlying facts about the image. Its main objectives compromises: tampering detection (cloning, healing, retouching, splicing), hidden data detection and recovery, source identification with no prior measurement or registration of the image. In this paper, application of Benford' law in Digital Image Forensics has been analyzed. JPEG and JPEG2000 approximately follow this law and JPEG2000 found closer to the given law. The non-following of this law in different forensics set-ups can be used as fingerprint. Further, amount of deviation from Benford' law in compressed images can be used to find forgery. Simulations results are given to establish the theory.**

Keywords—Benford's law; compression; JPEG2000; digital image forensics;

I. INTRODUCTION

In this era of digital world, the tampering of digital images is not a difficult task. There are various images editing software's like Adobe Photoshop is available with which image can be forged or tampered. The images can be compressed and their quality can be changed which is hardly detectable. Digital images are sometimes given as a proof of evidence in courts and decision can be made in accordance so authenticity of image becomes the question to be asked. In order to establish the authenticity of an image various digital image forensics techniques are needed to be explored and evolved. Digital image forensics can be broadly classified into two categories:-

1.) Passive v/s Active Forensics [1, 2]: - In passive the investigator can't investigate the image generation process and the image is only in 'read only' mode. In active forensics the generation is purposely modified so that it can leave behind the traces.

2.) Blind v/s Non-Blind Forensics [1, 2]:- In Blind the investigator examines the final image without having any knowledge about the generation of the image and neither it's intermediate results. The investigator also doesn't know about the original scene at the time of investigation. But in Non-Blind forensics the original scene can be easily detected and the intermediate steps are also known to the investigator.

In this paper, one of the passive approaches knows as Benford's Law [3] is discussed for the analysis of the digital image forensics. Here, Benford's law is applied on images in frequency and/or wavelet domain. This law computes the probability of the first nine digits of the image coefficients in its simpler form. This technique can be used to detect compression and also to define new signatures for forgery or tampering in images.

Benford's Law was first implemented on JPEG images. The JPEG images completely follow the Benford's law curve. It says that the probability of the first 9 digits of the Discrete Cosine Transform (DCT) coefficients of an image follow the Benford's law. Then this law is tested and verified for JPEG and JPEG2000 images. When compression in JPEG or JPEG2000 images is increased, the mean deviation from Benford's law increases resulting into violation of Benford's Law.

In this paper, use of Benford's law in image forensics is discussed and analyzed. Firstly, it was applied to DCT Coefficients of the JPEG images and it was found to be followed as also established and explained in [4]. Further, effect of multiple compressions and different qualities on the Benford curve of DCT coefficients was also analyzed. Secondly, DWT was applied to the JPEG2000 images and analysis was repeated with DWT coefficients on JPEG2K images. Lastly, application of the Benford's law for the glare detection [5] using both DCT and DWT coefficients in JPEG2000 images was explored and analyzed. All the above analysis are done in MATLAB2013a 2GB RAM.

The paper is organized in five sections. In section II required background is discussed. In Section III, first digit distribution of DCT coefficients of JPEG images are analyzed. DWT coefficients of JPEG2000 images are verified on the Benford's law in section IV. Section V concludes the research work.

II. BENFORD'S LAW [3]

The Benford's Law, was introduced by Frank Benford in 1938 and was later developed by Hill for analysis of the probability distribution of the first digit (1-9) numbers obtained from natural data in statistics. A typical probability distribution of Benford's Law is shown in figure 1. In this Law, the first digit of natural numbers (1-9) can be classified in a specific way, that the occurrence of the smaller digit is more than the lager digit. Hill explained the law in terms of statistics, concluding that the nature of probabilities of first digits from 1 to 9 is logarithmic. The distribution for Benford's Law can be expressed as in (1).

978-1-4799-6761-2/15/$31.00 ©2015 IEEE

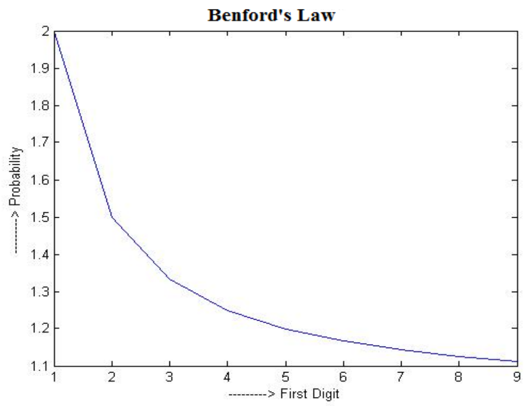

Fig. 1. Benford's law Curve

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

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

III. DCT COEFFEICENTS AND BENFORD'S LAW

A DCT is expressed a finite sequence of image points as a sum of cosine functions oscillating at different frequencies. The probability of the DCT coefficients of these low frequency components is calculated. Figure 2 shows the Cameraman and Lena image in JPEG format. The probability curve for these two images has been plotted and results are shown in figure 3 and 4.

Fig. 2. Cameraman.jpg (left) and Lena.jpg (right) [6]

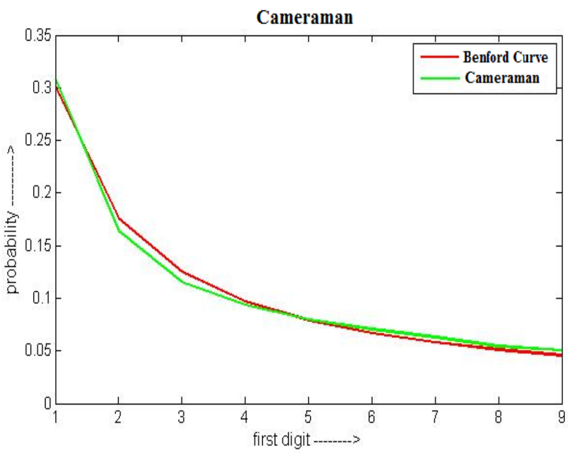

Fig. 3. Probability Curve of Cameraman.jpg

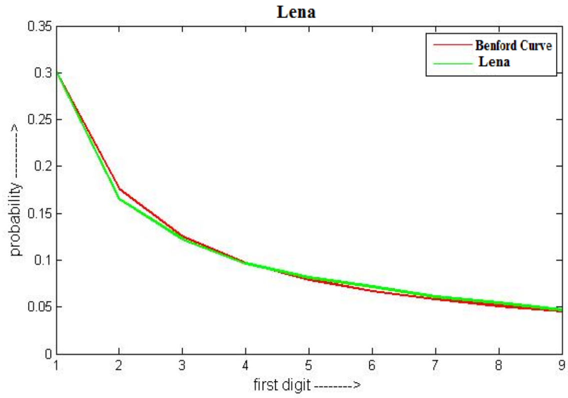

Fig. 4. Probability Curve of Lena.jpg

The system model for the analysis of JPEG image is as shown in figure 5. From figure 6 and 7, it can be found that the DCT coefficients of JPEG images at different quality factors almost follow the benford's law. This law is also tested and verified on 1338 UCID [7] images.

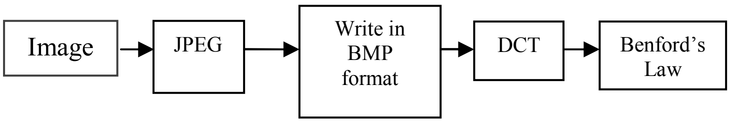

Fig. 5. System Model for analysis of JPEG images

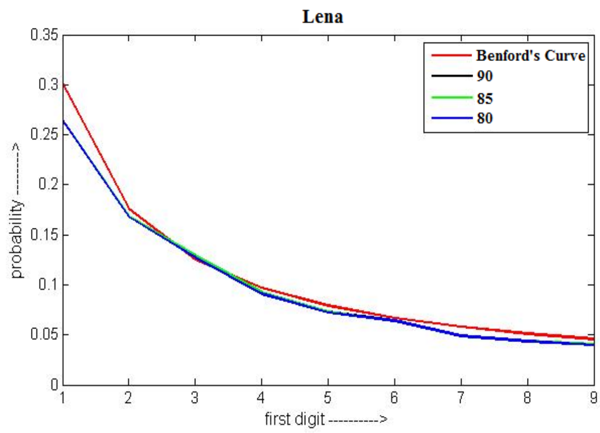

Fig. 6. Probability Curve of Cameraman.jpg at different Quality Factors

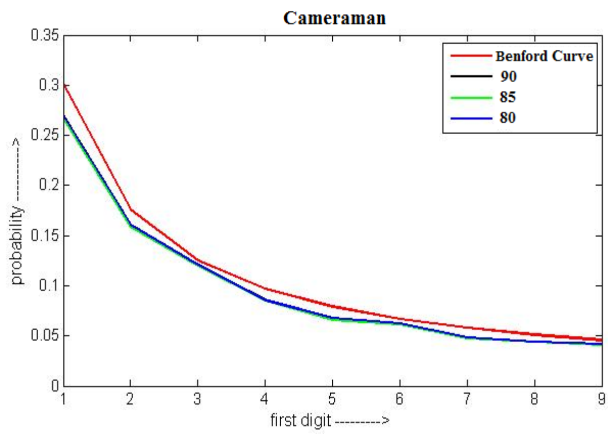

Fig. 7. Probability Curve of Lena.jpg at different Quality Factors

A. *DCT coefficients for Images with copy-paste forgery*

In this section, Benford's law is applied on DCT coefficients of tampered images and it is observed that these images also follow Benford's law. So, law in this form can't be used to detect this forgery but taking more places in consideration can help in detecting copy move forgery. This law used upto first place fails on the copy-paste images. In figure 8, the original image and the tampered JPEG images were written in TIFF and BMP format to verify the results. The Benford's curve for both the original and tampered image is as shown in figure 9 and 10.

Fig. 8. Original image (left) tampered image (right) [8]

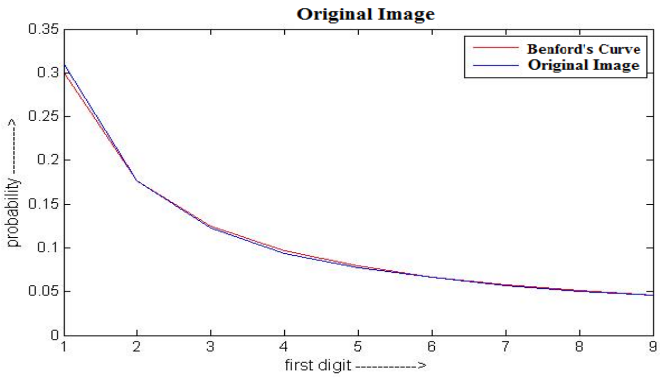

Fig. 9. Benford Curve of Original Image

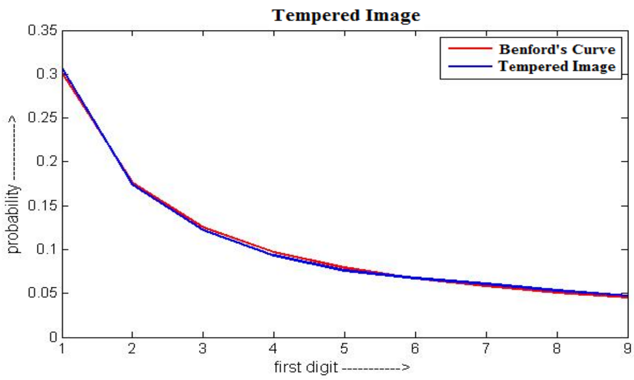

Fig. 10. Benford Curve of tampered Image

B Analysis of double compressed images

The DCT coefficients of double compressed JPEG images have been analyzed. Images were compressed twice with same or different Quality Factors (QF) and it is observed that double compressed images do not follow Benford's law but logarithmic distribution can be observed. This logarithmic law is similar to what was observed in single compressed images. Hence, this law can't be used for double compression detection in raw form but deviation ratio measures can be used as features to differentiate between single compressed and double compressed as one method proposed in [9].

The image was originally in TIFF format then the image is written using '**imwrite**' function twice in MATLAB, then the Benford's law is investigated on DCT coefficients. All the three possible cases i.e. firstly the image compressed with high quality factor then with lower and vice versa and thirdly with same quality factors. From figure 11 and 12, it can be found that compression operation disturbs probability distribution of first digits which can be seen in terms of deviation of probability curves from Benford's law. The degree of deviation is also changes with different set of quality factors used in double compression as given in Table 1 where (30, 60) represents first compression by quality factor of 30 (low quality) and then followed by 60 (high quality). Degree of deviation is mean square error between probabilities of first digits of original and double compressed images. Deviation ratio for both the Lena (left) and Cameraman (right) image is as shown in table 1.

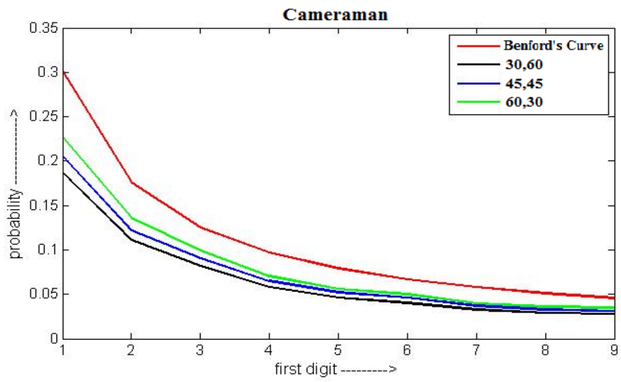

Fig.11. Probability curves of double compressed Cameraman.jpg

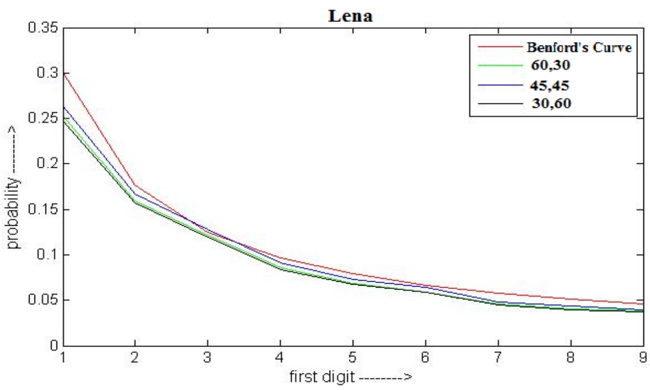

Fig. 12. Probability curves of double compressed Lena.jpg

C. *Analysis of Computer Generated JPEG Images*

Computer Generated Images applies computer graphics to generate photo-realistic images. This is very common these days to generates images from history, for visualisation of concepts, future products, buldings etc. Hence, thses images can also be used to regenerate crime scene.

TABLE I. DEVIATION RATIO

Quality factor	Deviation Ratio	Quality factor	Deviation Ratio
(30, 60)	0.1440	(30, 60)	0.3863
(45, 45)	0.0814	(45, 45)	0.3183
(60, 30)	0.1282	(60, 30)	0.2510

Fig. 13. CGI(1) (left) and CGI(2) right [6]

The above two figures shown in figure 13 are computer generated images. DCT is applied on these images and anaylsis is carried out further.

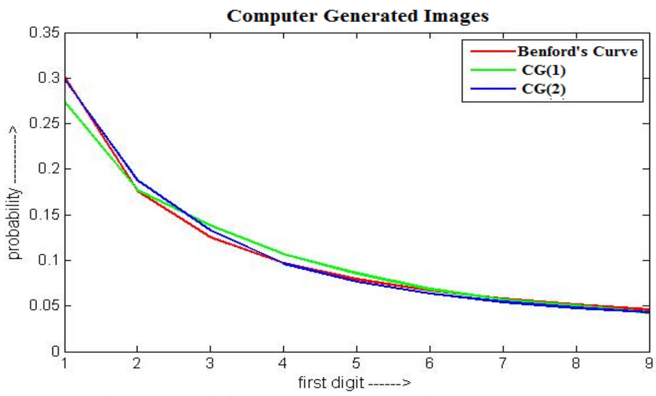

Fig. 14. Graph for Computer Generated Images

Here, from figure 14 it can be concluded that CGI also follow Benford's Law in general. But it is also showing deviation from Benford's law.

IV. DWT COFFICIENTS AND BENFORD'S LAW ON JPEG2000

In this section, Benford's law is investigated for JPEG 2000 images.

A. JPEG2000 and Discrete Wavelet Transform

DWT putrefy or transforms the large signal into smaller bandwidth signals (subbands) and also at slower sample rates. JPEG2000 is a new compression standard which is based on the Discrete Wavelet Transform while JPEG coding standard was based on the Discrete Cosine Transform. JPEG2000 standard has very high quality as compared to JPEG standard. It is both lossy and lossless compression depending on the type of filter used in DWT. Generally, Le Gall 5/3 is used for lossless compression and Daubechies 9/7 filter is used for lossy compression. In our, investigation we are going for lossy compression. Now, the following procedure as shown in figure 15 has been adapted to analyze DWT coefficients.

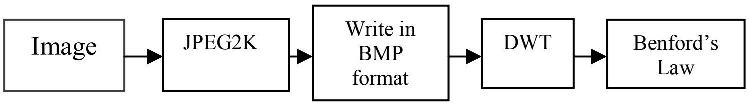

Fig. 15. Procedure for DWT coefficients on JPEG2000

In this, the input image is firstly written in JPEG2000 format in MATLAB by using '**imwrite**' function with different compression ratios using '**CompressionRatio**' in lossy mode. Then, the image's format is changed to BMP format and finally Discrete Wavelet Transform is applied at 5th level. It was observed that as the compression ratio increases the mean deviation ratio of the coefficient also increases. The deviation was calculated for the 3, 5, 7 on UCID database and the mean deviation ratio comes out be increasing. In this paper the results of Cameraman and Lena are being shown as an example. These images were applied the above proposed algorithm in MATLAB. Figure 16 and 17 shows the results of the algorithm.

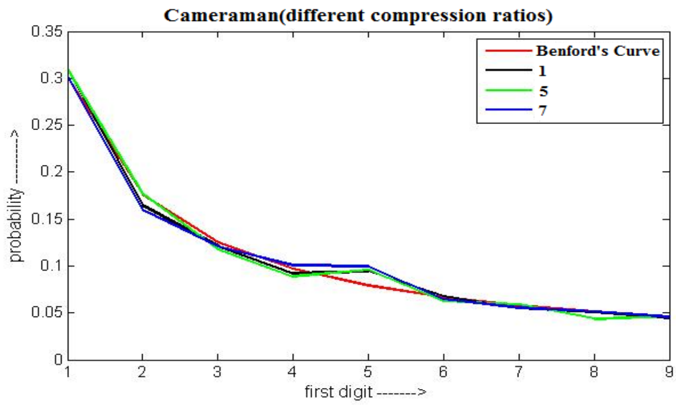

Fig. 16. Probability Curve with different Compression Ratios

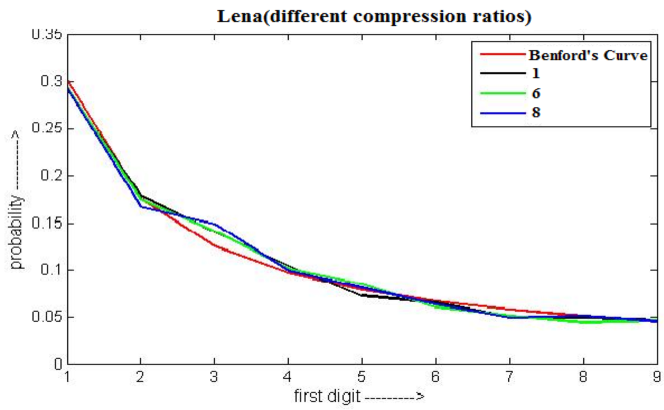

Fig. 17. Probability Curve with different Compression Ratios

From the above figures, we can conclude that as the compression ratio increases the deviation increases as there is significant increase in the coefficient number 5 for Cameraman image and 3 coefficient number 3 for Lena image.

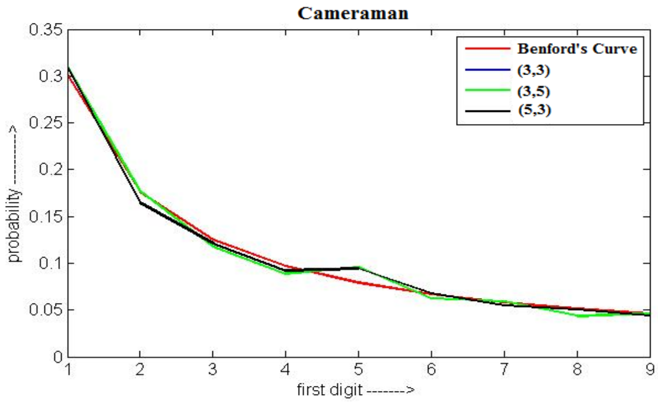

Fig. 18. Probability curves of double compressed Cameraman.jp2

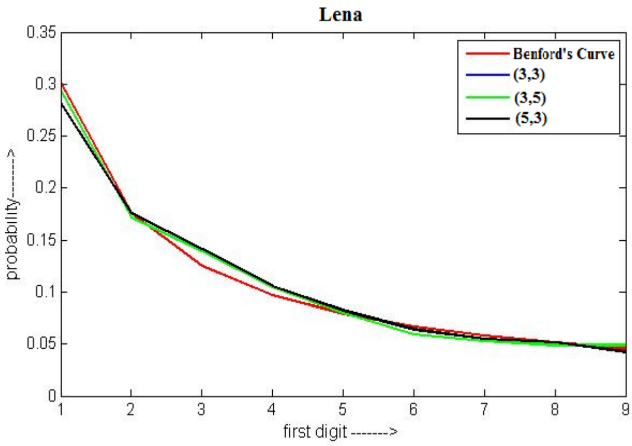

Fig. 19. Probability curves of double compressed Lena.jp2

B. Double compressed images in JPEG2000

In similar manner to JPEG, the DWT coefficients of double compressed JPEG2000 images have been analyzed.

Images were compressed twice with same or different Compression Ratio (or Quality Factor) and it is observed that double compressed images do not follow Benford's law at all. There is also a peak in logarithmic law similar to what was observed in single compressed images as shown in figure 18 and figure 19. Hence, this law can't be used for double compression detection in raw form but deviation ratio measures can be used as features to differentiate between single compressed and double compressed as one method proposed in [7]. Deviation ratios of Lena (left) and Cameraman (right) is as shown in table 2.

TABLE II. DEVIATION RATIO

Compression Ratio	Deviation Ratio	Compression Ratio	Deviation Ratio
(3,3)	0	(3,3)	2.2204e-16
(3,5)	0	(3,5)	0
(5,3)	0	(5,3)	2.2204e-16

C. Glare detection

When a glare is introduced or present in the image, DWT coefficients of such images is also investigated. Following steps were followed for glare detection. Firstly, the image was compressed to JPEG2000.Secondly, DWT at 1st level (although 2nd, 3rd level can also be done but information loss is . In the following two images as shown in figure20, glare is present.

Fig. 20 UCID00146 (left) UCID00181 (right) [4]

From figure 21 and 22 when DWT is applied it can be said that if glare is introduced in the image or present there is rise in the probability of the coefficient number 5. In figure 23, Benford's law is applied on DCT coefficients of Glare images. A similar peak at coefficient number 5 is observed. Hence, by applying both DCT and DWT we can get the same result.

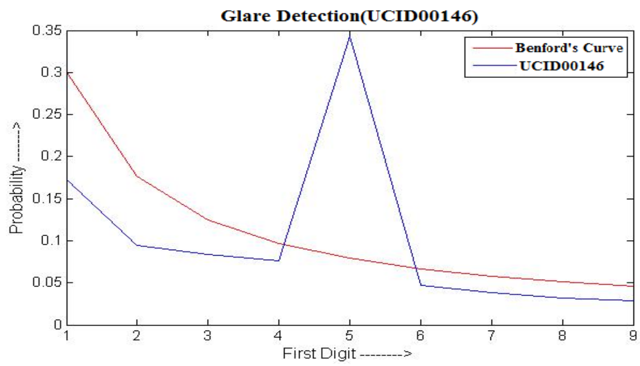

Fig.21. Glare detection for UCID00146

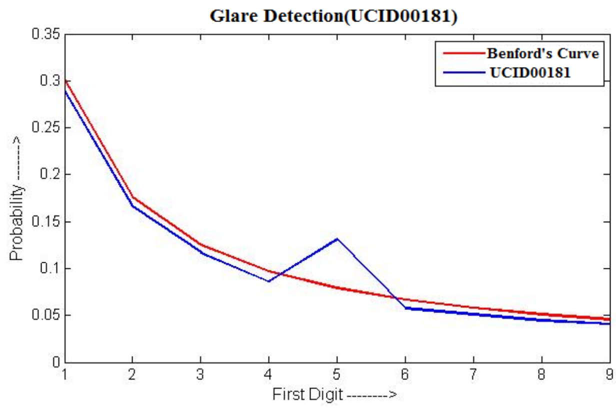

Fig.22. Glare detection for UCID00181

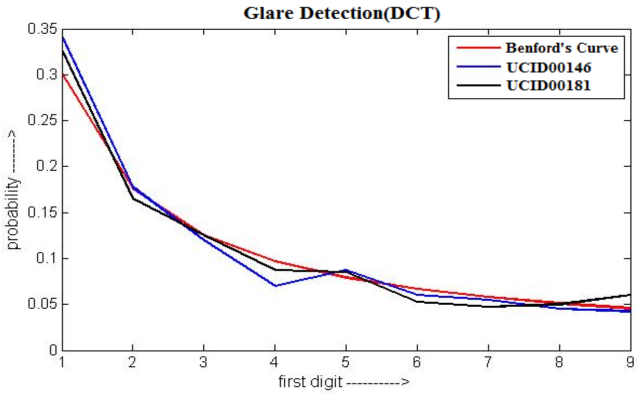

Fig. 23. Glare Detection by DCT

V. CONCLUSION

In this paper, Benford' law found to be image forensic tool for detecting multiple compression. Benford's Law is verified for JPEG and JPEG2000 images. Multiple compressions for JPEG images can be detected. Glare detections in UCID images also be investigated. Tampered images also follow Benford's Law. So, further tampering analysis is not done in the case of tampered images. Then, DWT coefficients were analyzed and it can be inference that as the compression ratio increases the mean deviation ratio also increases.

.

REFERENCES

[1] H. Farid, "Image forgery detection", Signal Processing Magazine, IEEE 26.2, pp. 16-25, 2009.

[2] J. A. Redi, T. Wiem and J. L. Dugelay, "Digital image forensics: a booklet for beginners", *Multimedia Tools and Applications 51.1*, pp. 133-162, 2011.

[3] A. Berger, T.P. Hill, "A basic theory of Benford's Law", *Probability Survey,* Vol. 8, pp. 1-126, 2011.

[4] Ghulam Qadir, Xi Zhao and Anthony TS Ho, on "Estimating JPEG2000 compression For Image Forensics Using The Benford's Law" Proc. of SPIE Vol. 7723 77230J-1, 2010.

[5] Qadir, Ghulam, Xi Zhao, Anthony TS Ho, and Matthew Casey. "Image forensic of glare feature for improving image retrieval using Benford's Law." In Circuits and Systems (ISCAS), 2011 IEEE International Symposium on, pp. 2661-2664. IEEE, 2011

[6] http://www.eecs.qmul.ac.uk/~phao/CIP/Images/, Oct, 2014.

[7] G. Schaefer and M. Stich, "UCID: An uncompressed color image database,"in Proc. SPIE: Storage and Retrieval Methods and Applications for Multimedia, 2003, vol. 5307, pp. 472–480.

[8] CASIA Image Tampering Detection Evaluation Database (CAISA TIDE) V1.0, http://forensics.idealtest.org.

[9] Fu, Dongdong, Yun Q. Shi, and Wei Su. "A generalized Benford's law for JPEG coefficients and its applications in image forensics." In Electronic Imaging 2007, pp. 65051L-65051L. International Society for Optics and Photonics, 2007.