IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 11, November 2015

A Novel Approach for Detecting Image Forgery

Nilu Treesa Thomas[1], Anju Joseph[2], Shany Jophin[3]

Student, Dep. of CSE, Amal Jyothi College of Engineering, Kanjirappally, India [1, 2]

Asst. Professor, Dep. of CSE, Amal Jyothi College of Engineering, Kanjirappally, India [3]

Abstract: To identify the traces of various forensic problems is an important issue in digital image forensic. The method propose a novel approach for detecting the traces of JPEG compression and image tampering using statistical feature extraction method. Discrete Cosine Transform Residual Features are used for extraction process. The method includes JPEG compression and that provides quantization noise based solution. Multiple-cycle JPEG compression is performed for noise analysis and define a quantity called forward quantization noise. The method analytically derive that decompressed image have lower variance of forward quantization noise. Using 64 kernels of DCT the quantized feature sets are generated and is so called as undecimated DCT. The proposed method solves the problems such as revealing the traces of JPEG compression history and identifies the tamped images using simple yet very effective detection algorithm. For chroma sub sampling and for small images image size the method is robust. The proposed algorithm can be applied in many practical applications, such as Internet image classification and forgery detection.

Keywords: Image Forgery, JPEG, DCT Features, Forward Quantization Noise, Forgery Detection.

I. INTRODUCTION

The popularization of imaging components equipped in personal portable devices, together with the rapid development of the high-speed Internet, makes digital images become an important media for communications. Various types of image compression standards, including lossy and lossless, coexist due to different kinds of requirements on image visual quality, storage, and transmission. Among them, JPEG is a very popular lossy compression format.

Knowledge about the JPEG compression history of images from unknown sources is of important interest to image forensics experts, whose aim is to trace the processing history of an image and detect possible forgeries. There are some reported works on identifying whether an image is uncompressed or has been compressed previously whether an image has been compressed once or twice whether an JPEG image [1] has been compressed again with a shifted JPEG grid position and on estimating the JPEG quantization table or quantization steps.

Focus on the problem of identifying whether an image currently in uncompressed form is truly uncompressed or has been previously JPEG compressed, and an image is tampered or non tampered. Being able to identify such a historical record may help to answer some forensics questions related to the originality and the authenticity of an image, such as where is the image coming from, whether it is an original one, or whether any tampering operation has been performed. For example, the solution facilitates the detection of image forgeries created by replacing a part of an image with a fragment from another image with a different compression historical record. The mismatch of historical records reveals the act of image tampering. The JPEG identification problem may also be the starting point for other forensics applications, such as JPEG quantization step estimation for that forensics experts can save time by only performing estimation on the decompressed images after filtering out the uncompressed images. There are also some techniques, called JPEG ant forensics aiming to fool the forensics detectors by concealing the traces of JPEG compression. However, as noted by removing the traces of JPEG compression is not an easy task. Some targeted anti-forensics detectors are designed to detect the traces left by anti-forensics operations. The method analytically derive that decompressed image have lower variance of forward quantization noise. using 64 kernels of DCT the quantized feature sets are generated and is so called as undecimated DCT.

The method consists of 2 modules
1. Quantization noise analysis.
2. DCTR Feature Extraction.

Here we are basically dealing with the above mentioned modules. These both modules are implemented by using two methods. Each method has its own output and a combined output is obtained to get the final result. The first method deals quantization noise analysis from the decompressed images, whereas the second method deals with DCTR Feature Extraction [2] from the quantized noise residuals obtained from the decompressed JPEG image using 64 kernels of the discrete cosine transform (DCT). Both the results are combined and a final result is attained. Enough graphs are shown to explain the analysis and accuracy obtained. The features are built as histograms of residuals obtained using the basis patterns used in the DCT. The feature extraction thus requires computing mere 64 convolutions of the decompressed JPEG image with 64 8 * 8 kernels and forming histograms. First, the analysis only uses a portion of the DCT coefficients that are close to 0. Hence, information is not optimally utilized.

Second, the method requires the quantization step to be no less than 2 to be effective. As a result, this method fails on high-quality compressed image such as those with a

DOI 10.17148/IJARCCE.2015.41160

quantization table containing mostly quantization steps being ones. It is based on the assumption that the obtained statistics will be small for an uncompressed image, while the statistics will become large for an image with anti-forensics operations. The detector is effective to detect anti-forensics operations and may also be directly applicable to detect decompressed images. Built on a theoretical model on multi-cycle JPEG compression in previous work, which try to reveal the high-quality compression traces in the noise domain. The proposed method define a quantity, called forward quantization noise, and develop a simple yet very effective algorithm to judge whether an image has been JPEG compressed based on the variance of forward quantization noise. The method fully utilizes the noise information from DCT coefficients; therefore, it is neither restricted to large image size nor limited by the quantization step being no less than 2. Shows that the method outperforms the previous methods by a large margin for high-quality JPEG compressed images which are common on the Internet and present a challenge for identifying their compression history.

The proposed system is implemented to identify the uncompressed JPEG image and tampered image. The method includes two steps and through which features are generated then using SVM classifier the results are obtained.
The two steps are
1. Introducing Quantization noise and Undecimated DCT
2. DCTR Feature Generation
The main agenda here is to identify whether the image has undergone any type of compression or tampering in its previous stage.

II. SYSTEM DESCRIPTION

The proposed system is implemented to identify the uncompressed JPEG image and tampered image. The method includes two steps and through which features are generated then using SVM classifier the results are obtained.

The two steps are
1. Introducing Quantization noise and Undecimated DCT
2. DCTR Feature Generation
The main agenda here is to identify whether the image has undergone any type of compression or tampering in its previous stage. JPEG Quantization Noise Analysis

A. JPEG Quantization Noise Analysis

The first method that is employed for detection of various types of images. A JPEG compression cycle consists of an encoding phase and a decoding phase. In the encoding phase, irreversible information loss occurs due to quantizing DCT coefficients. The decoding phase is essentially the reverse of the encoding phase. An integer rounding and truncation operation occurs when JPEG coefficients are restored into image intensity representation. In a recent work presented a framework for analyzing multiple-cycle JPEG compression based on a complete JPEG compression model, in contrast to the simplified models that are commonly used. The analysis focused on information losses in JPEG compression which can be characterized by two types of noise, i.e., quantization noise (in DCT domain) and rounding noise (in spatial domain). The truncation error is ignored in the model due to its fairly low impact and hard-to-model nature. Distributions of the two types of noises at different compression cycles are derived.

An uncompressed image, by performing the JPEG encoding phase can obtain its quantization noise of the _rst compression cycle. On the other hand, the image which is previously compressed cannot retrieve the quantization noise. However, it can compute the quantization noise of the next cycle. To be unified, the quantization noise obtained from an image for the current available upcoming compression cycle as forward quantization noise. Forward quantization noise is the subject of analysis and it is a function of its quantization step. In this section, describes the undecimated DCT and its properties relevant for building the DCTR feature set in the next section. Since here uses only luminance component, the paper limit to grayscale JPEG images. For easier exposition, the size of all images is taken as multiple of 8. The features are built with histograms of residuals obtained using the basis patterns used in DCT. The feature extraction thus requires computing mere 64 convolutions of the decompressed JPEG image with 64 8×8 kernels and forming histograms. The features can also be interpreted in the DCT domain. Symmetries of these patterns are used to further compactify the features and make them better populated. The proposed features are called DCTR features (Discrete Cosine Transform Residual). An M × N grayscale image, the undecimated DCT is defined as a set of 64 convolutions with 64 DCT basis patterns. The DCT basis patterns are 8×8 matrices. When the image is stored in the JPEG format, before computing its undecimated DCT it is first decompressed to the spatial domain without quantizing the pixel values to 0, . . . , 255 to avoid any loss of information.

B. DCTR Features

The DCTR features are built by quantizing the absolute values of all elements in the undecimated DCT and collecting the first-order statistic separately for each mode (k, l) and each relative position (a, b), 0 ≤ a, b ≤ 7. Formally, for each (k, l) define the matrix 2 U (k,l) a,b R(M -8)/8*(N -8)/8 as a submatrix of U (k,l) with elements whose relative coordinates w.r.t. the upper left neighbor in the grid G8×8 are (a, b).

The feature vector formed by normalized histograms for 0 ≤ k, l ≤7, 0 ≤ a, b ≤ 7. And note that q could potentially depend on a, b as well as the DCT mode indices k, l, and the JPEG quality factor. Because U (k,l) = X B(k,l) and the sum of all elements of B(k,l) is zero (they are DCT modes (2)) each U(k,l) is an output of a high-pass filter applied to X. For natural images X, the distribution of u U(k,l) a,b will thus be approximately symmetrical and centered at 0 for all a, b, which allows us to work with absolute values of u U(k,l) a,b giving the features a lower dimension and making them better populated. The symmetries of projection vectors (7), it is possible to decrease the feature dimensionality by adding together the histograms corresponding to indices (a, b), (a, 8-b), (8-a, b), and (8-a, 8-b) under the condition that these indices stay within 0, . . . , 7*0, . . . , 7. Note that for (a, b) 1, 2, 3, 5, 6, 72, merge four histograms. When exactly one element of (a, b) is in 0, 4,only two histograms are merged, and when both a and b are in 0, 4 there is only one histogram. Thus, the total dimensionality of the symmetrized feature vector is 64 * (36/4 + 24/2 + 4) *(T + 1) = 1600 *(T + 1).

In the rest of this section, provide experimental evidence that working with absolute values and symmetrizing the features indeed improves the detection accuracy. Also experimentally determine the proper values of the threshold T and the quantization step q, and evaluate the performance of different parts of the DCTR feature vector w.r.t. the DCT mode indices k, l.

III.EXPERIMENTAL DETAILS

The images selected to be included in dataset is from the UCID dataset, which consists of images in the TIF format, BMP format and JPG format. The first is images are converted into gray-scale and then each images center cropped to generate images of smaller sizes, i.e., 256 × 256, 128 × 128, 64 × 64, and 32 × 32 pixels. The uncompressed images as well as their corresponding decompressed JPEG images are used for evaluation. Here subject the newly proposed DCTR feature set to tests on selected use four different settings, Firstly, test the methods on gray-scale images to show how the performance is on each designated compression quality. Secondly, run test on color images to show whether the methods are robust to chroma subsampling. Thirdly, conduct experiments on JPEG images from a publicly available database with random quality factors to verify the true positive rates. Finally, conduct experiments on uncompressed images from another database to verify the false negative rates. The computational complexity when extracting the feature vector using a Matlab code. The SVM (support vector machine) classifier, which is used for comparison. Since it is not as exible and time efficient as other three methods in performing forensics-related tasks, here uses only experimental setup. The (Gaussian) radial basis function kernel is used in the SVM and the parameters are optimized by grid-search.

The first use of this method checks for decompressed images and uncompressed images. Assume the decompressed images and uncompressed images respectively to be the positive class and the negative class, true positive rate and true negative rate respectively evaluate the percentage of correctly identified decompressed images and that of uncompressed images. False positive rate evaluates the percentage of wrongly identified uncompressed images. The second use of the method is image tampering detection. Once an image has inconsistency in JPEG compression history among different parts, possible forgery may be detected. Suppose an image forgery is composed of two parts as illustrated in below figure.

Part A is from a decompressed JPEG image; from another image part B is inserted.

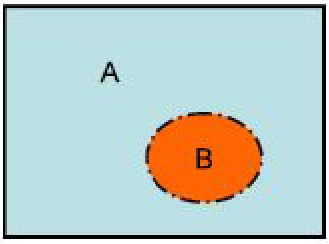


Fig.1: Feasible Forgery Detection Scenarios

From high-quality compressed JPEG image Part A is decompressed; our method is capable of detecting image forgery that belongs to one of the following cases.

1. Forgery Case A: Part B is from an uncompressed image.
2. Forgery Case B: Part B is synthesized through a computer graphics rendering or uncompressed image-based synthesis technique. All experiments in this section are carried out on BOSSbase 1.01 containing 10,000 grayscale 512 × 512 images.

All detectors were trained as binary classifiers implemented using the FLD ensemble, the ensemble by default minimizes the total classification error probability under equal priors PE. The dimensionality of random subspace and base learners is found by E00B. And also use EOOB to report the detection performance. In this section, experimentally validate the feature symmetrization. Denote by EOOB(X) the EOOB error obtained when using features X.

A. Working

Initially our image set is composed of 3,000 images, with 1,000 of them coming from BOSSbase ver 1.01 image database and 1,000 from UCID image database. These publicly available image sets are a reliable source of uncompressed images. The images are first converted into gray-scale and then center-cropped to generate images of smaller sizes.

These images are decompressed and served as the ground-truth JPEG decompressed images. And for forgery detection given a color test image, first extract its luminance channel, and then perform JPEG identification independently on non-overlapping B × B-pixel macro-blocks of the luminance channel. Considering a good trade-off between detection sensitivity and accuracy, use B = 32 for forgery detection.

And the features are generated using above DCTR Feature extraction method then using SVM classifier we identify the result. In the preprocessing step it is able to see how the original image is decompressed and color images are converted to luminance channel. In this process first the image is compressed by setting up a quantization table and the coefficients are generated during this preprocessing stage. Now see how preprocessing is done in the decompressed method. Then read function is performed for collecting all the saved data. Here extract the histogram of the image. The DCTR feature extraction method and feature selection process. The feature extraction is completed and the generated features are given for SVM training and the feature selection process and the SVM gives out the result for both compressed images and tampered images.

DOI 10.17148/IJARCCE.2015.41160

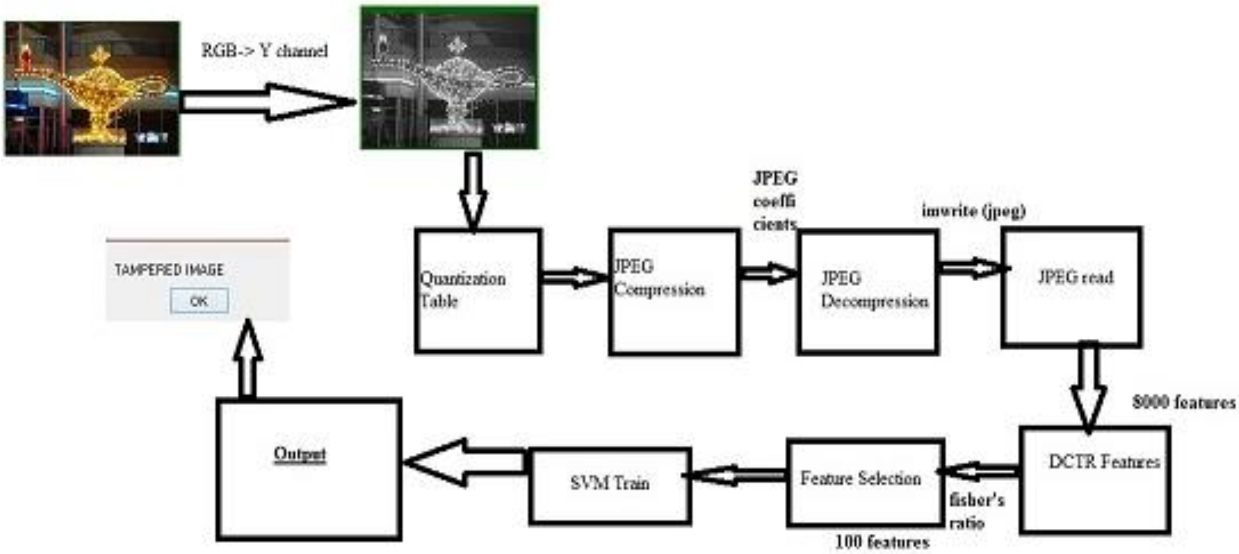
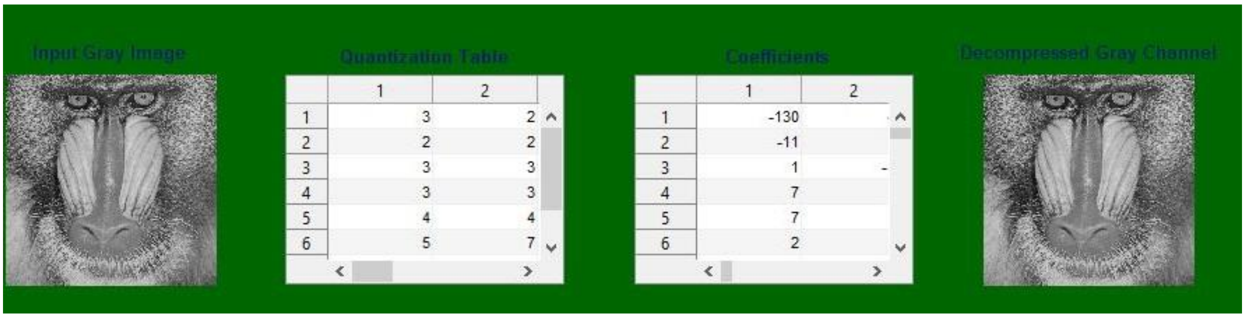


Fig. 2: Proposed Method



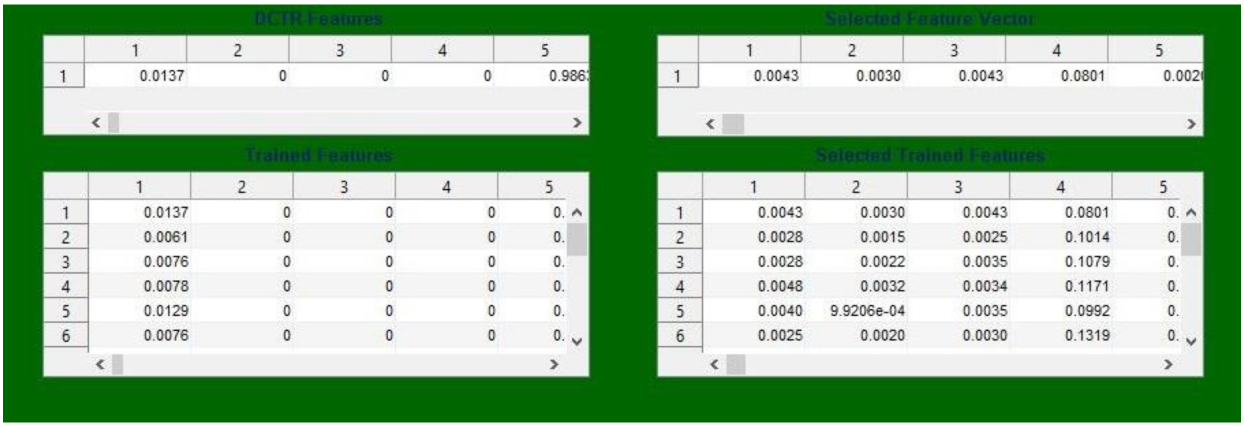
Quantization Table

	1	2
1	3	2
2	2	2
3	3	3
4	3	3
5	4	4
6	5	7

Coefficients

	1	2
1	-130	
2	-11	
3	1	-
4	7	
5	7	
6	2	

Fig.3: Preprocessing of Proposed Method



DCTR Features

	1	2	3	4	5
1	0.0137	0	0	0	0.986

Selected Feature Vector

	1	2	3	4	5
1	0.0043	0.0030	0.0043	0.0801	0.002

Trained Features

	1	2	3	4	5
1	0.0137	0	0	0	0.
2	0.0061	0	0	0	0.
3	0.0076	0	0	0	0.
4	0.0078	0	0	0	0.
5	0.0129	0	0	0	0.
6	0.0076	0	0	0	0.

Selected Trained Features

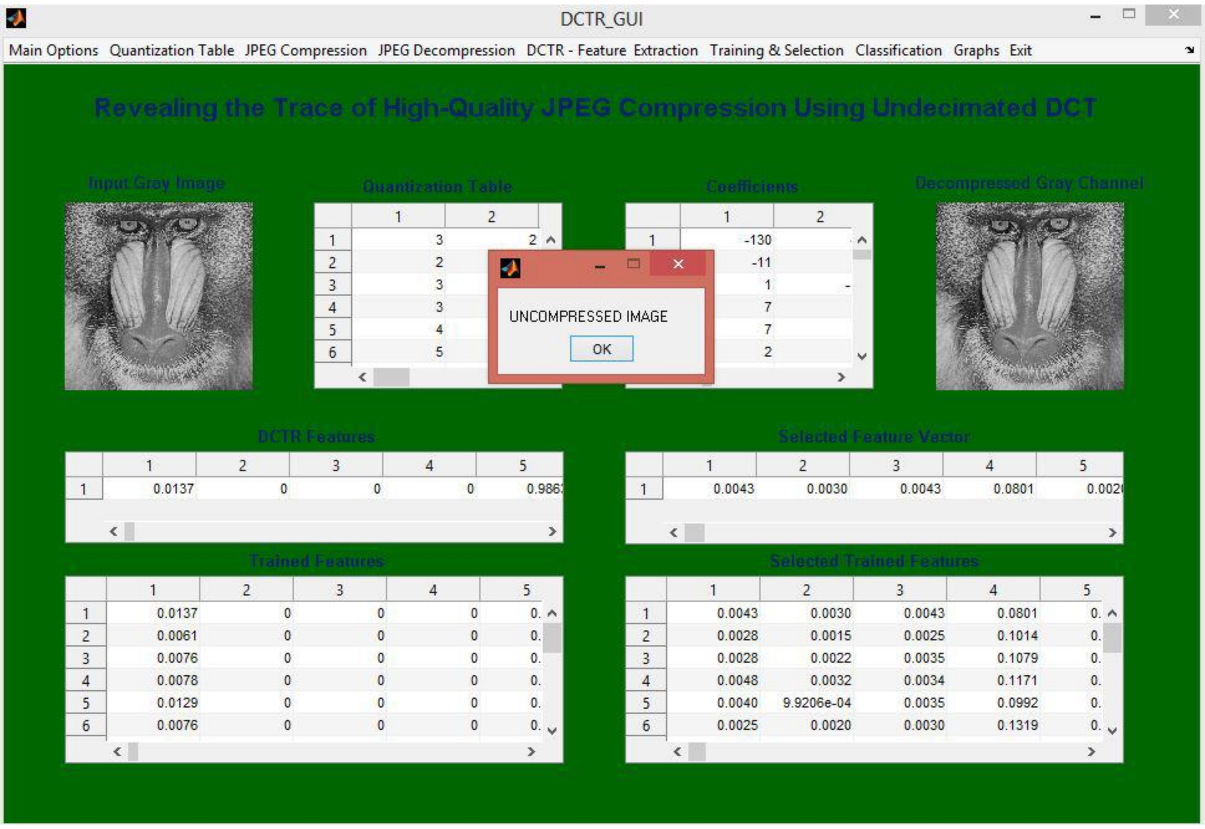
	1	2	3	4	5
1	0.0043	0.0030	0.0043	0.0801	0.
2	0.0028	0.0015	0.0025	0.1014	0.
3	0.0028	0.0022	0.0035	0.1079	0.
4	0.0048	0.0032	0.0034	0.1171	0.
5	0.0040	9.9206e-04	0.0035	0.0992	0.
6	0.0025	0.0020	0.0030	0.1319	0.

Fig.4: Feature Extraction Process

B. Analysis

The analysis of the work is done with the help of a correlation matrix which consist of true positive, false positive, true negative, false negative, and for EOOB for both compressed images and tampered images. In addition to this we compare the variance along with DCTR Features. As we fix a false positive rate for the whole image set, we can easily obtain threshold for our method. In this case, the performance can be evaluated based on the true positive rate, the higher the better. For the results reported in accuracy, we may need to tune the threshold or the parameters for each quality factor. For the results reported in true positive, we only need to set the threshold according to the uncompressed images, which bring us great flexibility. For tampered images,since the composite image is of size 512×512, there will be an amount of 256 macro-blocks of size 32 × 32. Among them, exactly 4 Macro-blocks are from the uncompressed image. When all 252 macro-blocks in the outer region of the composite image are identified as decompressed, and at least 2 out of the 4 macro-blocks in the inner tampered region are identified as uncompressed, regard the image as being correctly identified.

DOI 10.17148/IJARCCE.2015.41160



Revealing the Trace of High-Quality JPEG Compression Using Undecimated DCT

UNCOMPRESSED IMAGE

Quantization Table

	1	2
1	3	2
2	2	
3	3	
4	3	
5	4	
6	5	

Coefficients

	1	2
1	-130	
	-11	
	1	-
	7	
	7	
	2	

DCTR Features

	1	2	3	4	5
1	0.0137	0	0	0	0.986

Selected Feature Vector

	1	2	3	4	5
1	0.0043	0.0030	0.0043	0.0801	0.002

Trained Features

	1	2	3	4	5
1	0.0137	0	0	0	0.
2	0.0061	0	0	0	0.
3	0.0076	0	0	0	0.
4	0.0078	0	0	0	0.
5	0.0129	0	0	0	0.
6	0.0076	0	0	0	0.

Selected Trained Features

	1	2	3	4	5
1	0.0043	0.0030	0.0043	0.0801	0.
2	0.0028	0.0015	0.0025	0.1014	0.
3	0.0028	0.0022	0.0035	0.1079	0.
4	0.0048	0.0032	0.0034	0.1171	0.
5	0.0040	9.9206e-04	0.0035	0.0992	0.
6	0.0025	0.0020	0.0030	0.1319	0.

Fig.5: DCTR Output

No matter which type (single compressed, aligned double compressed, or non-aligned double compressed) an image belongs to, the image is in the category of JPEG decompressed. A perfect detector would give a result indicating all images are positives. Then contrast the detection accuracy and computational complexity of DCTR which is used for detection of JPEG steganographic methods.

The following image shows the analysis of the above mentioned variants of the correlation matrix. Assume the decompressed images and uncompressed /tampered images respectively to be the positive class and the negative class, true positive rate and true negative rate respectively evaluate the percentage of correctly identified decompressed /tampered images and that of uncompressed /non tampered images.

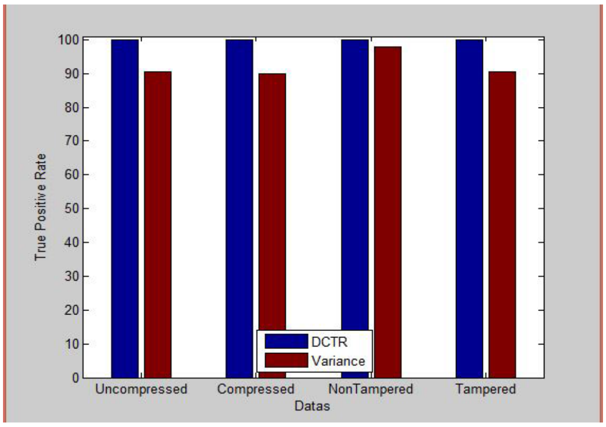


Fig.6: True Positive Rate

The graph shows the true positive rate of uncompressed/ decompressed images along with tampered and non tampered images plots DCTR vs variance function.

False positive rate evaluates the percentage of wrongly identified uncompressed / tampered images. As we report the results with accuracy, we always randomly split the images into the training set (4/5 of the overall images) and the testing set (1/5 of the overall images), and apply the threshold or the parameters, obtained on the training set with the best accuracy, to the testing set. The graph 4.6 shows the false positive rate with DCTR along variance.

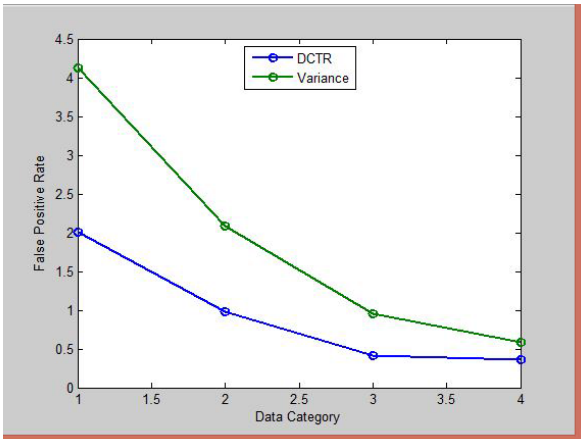


Fig.7: Detection Error

Detection Error is analyzed with various Quality Factor, DCTR and Variance is checked for each QF along with EOOB. Each plot shows 4.7 that there the only slight

DOI 10.17148/IJARCCE.2015.41160

difference with the various QF (ie, 95, 90, 80, 75) and that shows the better performance of the method.

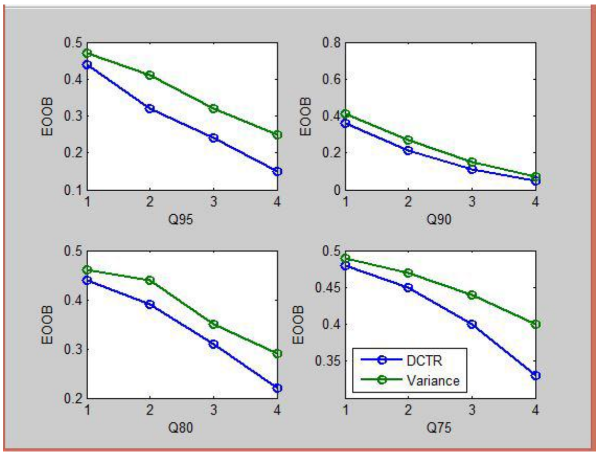


Fig.8: Detection Error

IV. CONCLUSION

This paper summarizes the findings of the study about various methods to reveal the traces of JPEG compression. It finds foundation on the implementation work done prior to this, of the existing JPEG image compression detection and tamper detection technique. The proposed method can be applied to Internet image classification and forgery detection with relatively accurate results. It should be noted that the proposed method is limited to discriminating uncompressed images from decompressed ones which have not undergone post-processing. Finally, we would like to mention that it is possible that the DCTR feature set will be useful for forensic applications. In conclusion, the foundation work for the proposed system was laid out and the implementation details were written.

ACKNOWLEDGMENT

First of all I submit all glory and honor to GOD ALMIGHTY for his eternal love and protection throughout the work. The tremendous amount of help that I received from a number of people gave me the incentive to complete this work to the best of my capability. I thank sincerely all my friends who gave me all the help in times of trouble.

REFERENCES

[1] Bin Li, Tian-Tsong Ng, Xiaolong Li, Shunquan Tan, and Jiwu Huang,``Revealing the Trace of High-Quality JPEG Compression Through Quantization Noise Analysis,"IEEE Transaction on Information Forensics and Security, VOL. 10, NO. 3, MARCH 2015.

[2] Vojtech Holub and Jessica Fridrich,\Low-Complexity Features for JPEG Steganalysis Using Undecimated DCT," IEEE Transaction on Information Forensics and Security, VOL. 10, NO. 2, FEBRUARY 2015.

[3] S. Sahami, M.G. Shayesteh, \Bi-level image compression technique using neural networks," IET Image Process, 2012.

[4] Jianquan Yang, Jin Xie, Guopu Zhu, Sam Kwong,and Yun-QingShi,``An Effective Method for Detecting Double JPEG Compression With the Same Quantization Matrix," IEEE Transactions , VOL. 9, NO. 11, Nov 2014.

[5] Fausto Galvan, Giovanni Puglisi, Arcangelo Ranieri Bruna, and Sebastiano Battiato, ``First Quantization Matrix Estimation From Double Compressed JPEG Images," IEEE Transaction on information forensics and security,2014.

[6] Zhouchen Lin, JunfengHe, XiaoouTang, Chi-KeungTang, ``Fast, automatic and fine-grained tampered JPEG image detection via DCT coefficient analysis," Elsevier-Pattern Recognition, 2009.

[7] Chunhua Chen, Yun Q. Shi, and Wei Su, ``A Machine Learning Based Scheme for Double JPEG Compresson Detection," IEEE Transaction 2008.

[8] Bin Li, Yun Q. Shi, Jiwu Huang, ``Detecting Doubly Compressed JPEG Images by Using Mode Based First Digit Features," IEEE Transaction, 2008.

[9] T. Bianchi, A. Piva, ``Analysis of Non-Aligned Double JPEG Artifacts for the Localization of Image Forgeries," IEEE Trans, 2011.

[10] Fangjun Huang, Fangjun Huang, Yun Qing Shi , \Detecting Double JPEG Compression With the Same Quantization Matrix," IEEE Transaction on information forensics and security, vol.5, No.4, Dec 2010.