ANALYSIS OF NON ALIGNED DOUBLE JPEG COMPRESSION TECHNIQUES FOR IMAGE FORGERY DETECTION

Musolin H

University Institute of Engineering and Technology
Panjab University

University Institute of Engineering and Technology, Panjab University, Sector 25, Chandigarh
INDIA

ABSTRACT:
This paper focuses on analytical study of digital image forgery techniques for detecting Non Aligned Double JPEG Compression. If a region is randomly cropped and pasted within the same image, the blocks within an image get non aligned. The behavior study of such scenario is done using two important features IPM and DIPM to calculate minimum entropies. Weka classifier classifies these entropies and works on 5 cross validation and calculates the accuracy of the proposed scenario. Experiment results on a set of about 100 uncompressed and 100 tampered images, with various parameter settings is done to demonstrate the accuracy of the scenario.

Keywords: Image forensics, Active passive blind methods, JPEG artifacts, nonaligned double JPEG compression, WEKA classifier.

INTRODUCTION

Digital images and videos, widely used, play an important role in today's technical world. They act as photographic evidence by the government agencies, legal authorities and by many other. With the speedy advancement in editing software's, the integrity of a given digital image increases concerns.

Various methods were introduced to prevent tampering known as Active techniques. Watermarking systems, digital signatures were among them. Watermarking system assumes insertion of watermark at the source end verifying it at the receiver end[1]. Watermark limits to different instances as it was not possible to determine whether watermark was inserted after manipulations and was also not applicable on images or videos from surveillance cameras, military cameras etc. Digital signature is encoded at the sender side and decoded at receiver end to ensure its authenticity. These Active techniques had certain limitations. Therefore, nowadays the focus is on Passive techniques, popularly known as Blind methods, verify authenticity of images without using pre extracted or pre embedded information [2] [3]. It is based on the fact that tampering changes the
inherent statistical characteristics of an image. These inconsistencies act as a footprint for detection of tampering.

Various artifacts are introduced by different blind methods that are categorized as noise inconsistency[4], blur[5], sharpening [6], copy move[7][8][9], inpainting[10], image splicing[11],[12] etc. such artifacts are used to detect whether images are forged or not. Since, JPEG is widely used file format for storage and compression, various artifacts are introduced due to splicing are double Compression, blocking artifacts [13], JPEG ghosts[14], non aligned double JPEG compression[15].

In [16], the authors propose to detect areas which have gone through double JPEG compression with different quality factors, whereas in [17] double JPEG compression is detected analyzing the DCT coefficients where grids are perfectly aligned. The case of non aligned double JPEG compression has been investigated in [18], by considering blocking artifacts.

Our approach relies on features(IPM,DIPM) to detect misalignment of grids in double compressed images. We will be observing the behaviour of this algorithm on our scenario – copy paste without preserving grids.

**II] PREVIOUS WORK DONE**

Yi-Lei Chen and Chiou-Ting Hsu [19] proposed a method to discover traces caused by recompression. These compression artifact abnormalities, either in spatial or frequency domain, have been used to detect recompression in JPEG images. Zhang Yu-jin and Wang Shi-lin [20] presented a method for the detection of SD JPEG compressed image using Intra-Block and Inter-Block Correlations. Salma Hamdy and Haytham [21] proposed a method to estimate the quantization table from the peaks of the histogram of DCT coefficients. Tiziano Bianchi and Alessandro Piva [22] gave a method to discriminate between original and forged regions in JPEG images. This algorithm manually selects a suspected region in order to test the presence or the absence of double compression artifacts.

In [23] Jing Zhang proposed a method for the detection of double compression in JPEG 2000. This method calculates the difference between the sub-band DWT coefficient histograms between single and double JPEG compression. These coefficients have some visible artifacts in them. Double compression is implemented
using SSVM (Soft Margin Support Vector Machine). This method assumes high accuracy rate.

Fangjun Huang [24] pointed out that different quantization matrixes were used for primary and secondary compressions till date. So a method was proposed to detect double JPEG compression with the same quantization matrix. A JPEG image is recompressed with the same quantization matrix showing a sequential decrease in the quantized DCT coefficients. A proper ratio is calculated on the JPEG coefficient of recompressed test image via implementing a random perturbation strategy. This method is not only applicable for the detection of double compression but can also detect triple and sometimes for times compression and so on.

FangLing SHI [25] analyzed different features between JPEG double and single compression quantization histogram and estimated values and then proposed a method in which the extracted features were categorized into two classifier with the SVM technology, then a simulation experiment is conducted to differentiate between double compressed JPEG images and single compressed JPEG images. Method proposed is practical and applicable to real life experiments. It basically improves accuracy of checking whether images are double compressed or not. This method laid foundation for detecting image distortion and image computer forensics. Disadvantage of this method is that it is time consuming.

In [26], Yu Chen and Carmen Cheh method is given to improve the accuracy of JPEG image tampering detection. This paper presents a method that detect by differentiating between JPEG single and double compressed quantization histograms. It considers the characteristics of the random distribution of high value bins in the DCT histograms of real-world images. For this method we have used publicly available CASIA authentic and tampered image data set of 9501 JPEG images. 20 rounds of experiments with small set of images were performed leaving high number of images in each round for testing purpose showing robustness and accuracy of our method. Experimentally we prove, average improvement in the true negative and positive rate. Our method can supported automated and reliable digital image evidence authenticity verification. It had better performance.

Athulya B and Manoj Ray D [27] provides detailed study
of digital image forgery on JPEG images. When a tampered JPEG image is double compressed, final image will have different compression properties than that of single compressed images. This difference in the blocking artifacts is used to detect recompression. To hide the information of the target image a portion of digital image is copied and pasted either on the same or different image in order to hide. BACM properties are used to detect whether an image is cropped or not. In [28], Zhenhua Qu estimated primary quantization matrix in double compressed JPEG images using ICA based identification algorithm. A convolutive mixing model is used for better interpreting the Shifted Double JPEG problem. The independency between BDCT coefficients weakens due to Shifted Double JPEG compression. The method used in this algorithm can be extended for color images and JPEG 2000. Babak Mahdian [29] proposed a double compressed detection method based on histograms of DCT coefficients and SVM. The method produces a significantly less number of false positives. In [30], Tomas Pevny introduced a method for detection of double compression in JPEG for application in steganography that is based on DCT coefficients. Primary quantization matrix is compared with secondary matrix. This method detects double-compression not only for cover images but also for images processed using steganographic algorithms. This is first complete solution to the problem of estimation of the primary quality factor in double-compressed JPEG images.

Zhang Ting [31] proposed a method based on SVM multi-class classification for detecting doctored JPEG image. Algorithm analyzes recompression statistical characteristics and concludes that the area not possessing double compression features are termed as suspected doctored region. SVM is used to located the doctored blocks and the doctored region consists of all the connected blocks. Method can detect copy paste forgery within or between JPEG images with low computational complexity and detects doctored region in an image by looking into the statistical characteristics of DCT coefficients and then analyze the differences of double compression effect between doctored and non-doctored region. Algorithm fails when the size of tampered area is small and also limits for detecting images with single details and similar texture.
In Non-aligned double JPEG compression, a portion of JPEG compressed source image is cut and pasted onto another JPEG compressed target image without preserving grid alignments. This kind of tampering is detected by a method that depends on the integer periodicity of the DCT coefficients. These coefficients cluster around a given lattice for any possible JPEG grid shift. This measure is compared with a threshold to decide whether grids are aligned or not. A method to detect such artifacts has been proposed by Bianchi [32].

Our work is in reference to the work proposed by Bianchi and Piva [32] where IPM and DIPM act as two important features to calculate the minimum entropy values of the images. We observe the behavior of a scenario that has not been discussed in this algorithm.

Scenario - Misaligned Double compressed copy paste
A region from a JPEG image is cropped and pasted onto the same JPEG image resulting in misalignment of blocks. Entropies of IPM and DIPM (which acts as features) are calculated for these images. These values are then stored and compared with that of double compressed images. Accuracy is calculated on Weka. Graphs are plotted.

A. PERFORMANCE ANALYSIS
In this section we will describe the experimental methodology we have followed in order to evaluate the performance of the proposed scenario. Then, the experimental results coming from such analysis have been discussed. The image dataset used for testing the algorithm is composed by 300 non-compressed TIFF and bmp images, having heterogeneous contents. Different image sizes 128x128, 512x512, 1024x1024.
are considered. Starting from these we then create various types of forged images by carrying out different operations as mentioned in above scenarios. An uncompressed image is compressed with quality QF1 (ranging from 50 to 94) resulting in single compressed JPEG image. A region from this JPEG image is cropped and pasted onto another JPEG image which is compressed again with quality QF2 (50, 53, 56, 60, 63, 66, 91, ..) to obtain a forged double compressed image. Forged images compressed with these qualities are obtained. Minimum entropy values for IPM and DIPM are calculated for these forged images. Further, double compressed images are obtained by simply compressing uncompressed images with quality factors, QF1 and QF2. The minimum entropies for IPM and DIPM are calculated on these double compressed images. Entropy values of these double compressed images and forged double compressed images are stored and given as input to weka classifier where it performs 5 cross validation strategy to calculates the accuracy, TPR and FPR. Graphs for the accuracy on different image sizes are plot.

The algorithm is also tested on a set of images representing realistic cases of forgery.

**B. Pseudo Code of the proposed scenario for detecting NA-DJPG compression.**

1: for i = 1 → n
2: for x = 50 → 94
3: for x1 = z
4: \( I_i = F(I_i, x) \)
5: \( r = R(I_i), c = C(I_i) \)
6: \( I_b = I_i (r:r+bs,c:c+bs) \)
7: \( I_i (r:r+bs,c:c+bs) = I_b \)
8: \( I_i = F2(I_i, x1) \)
9: end for
10: end for
11: end for

Where, \( n \) are the no. of original images

\( F(x) \) is single JPEG compression

\( x \) quality factor for first compression

\( F2(x1) \) is double JPEG compression

\( x1 \) quality factor for second JPEG compression

\( R() \) function returns random row value \( \not\equiv \mod(rows,8) \)

\( C() \) function returns random row value \( \not\equiv \mod(cols,8) \)

\( z \in \{50, 53, 60, 63, 69, 91, 97\} \)
C. Graph showing accuracy of the scenario with different image sizes

[V] CONCLUSION
From the scenario we conclude that the algorithm only predicts whether an image is forged or not but does not detect the tampered region. Entropy values are used as a benchmark to distinguish between classes in digital image forgery detection. Accuracy of the algorithm increases for larger image sizes. Our accuracy is maximum for 1024x1024 size images. Many other scenarios could be further taken up—If an image has 8x8 blocks, tampered region might not consist of 8x8 block and might have some other block size or in other case DCT of single and double compressed images is greater than that of double and triple compressed images.

ACKNOWLEDGEMENT
The authors are thankful to Dr. Savita Gupta Co-ordinator CSE, Librarian Mr Raj Kumar and also to his staff; University Institute of Engineering and Technology, Panjab University, Chandigarh, India for providing the necessary facilities for the preparation of the paper.

REFERENCES


