Journal of Engineering and Applied Sciences 12 (1): 104-112, 2017
ISSN: 1816-949X

Current Techniques in JPEG Image Authentication and Forgery Detection

Nadeem Alherbawi, Zarina Shukur and Rossilawati Sulaiman
Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia,
43600 Bangi, Selangor, Malaysia

Abstract: Image authenticity is a real issue in the digital forensic field since the wide spread of images and the spread of low-cost image processing software which make it easy to alter images and change them using hard to detect techniques. This study highlights the importance of image authentication and on which level a JPEG image could be altered. Also, it covers some techniques used in making forged images. Finally, it discusses some current techniques used in detecting and localizing forgeries in JPEG images.

Key words: Forgery, JPEG, image authentication, Forgery detection, digital investigation

INTRODUCTION

With the facileness of digital image manipulation, image forgery has become a prevalent concern. The expeditious development of commercial image editing software such as Adobe Photoshop dramatically increases the amount of doctored photographs circulated every day. This phenomenon leads to earnest consequences and reducing trustworthiness in the various real-world application.

Image authentication: When a digital image is acquired, the information is stored in a storage medium, like a hard drive or memory card. This information can only be translated into a visual image; people can make sense of by displaying it on a monitor. However there are more characteristics of a digital image than just the image information. Digital images are a product of mathematics and computer language, both of which operate in an expected way.

Image authentication is about determining if any aspects of this order have been manipulated. Changes can be made to the media information, the digital file or inconsistencies in the events surrounding its capture. Therefore, the analysis of digital image files can be divided into four categories: file structure, global structure, local structure and source identification. The useful techniques in each category are illustrated in (Fig. 1).

File Structure analyzes investigate the format of the digital information such as the file type, EXIF, hex data, and MAC stamps. Digital cameras create files in a particular way, each with its unique structure. Information is embedded into image files which can be distinct between manufacturers and cameras. When computers or image processing software, interact with the file, this structure could be altered in some way. While this type of alteration does not necessarily mean that image content has been modified, it can raise concern about the authenticity of the file.

The next category of analysis focuses on the content of the digital file, the content that forms the actual image information. Global structure analyzes investigate the content that expresses the digital image information. In digital cameras, captured light coming through the lens is converted to electrical energy by a camera's sensor, and then the camera's system use specific image algorithms to convert the data to form an image file. As a result, the digital image is an outcome of a mathematical process. Digitalizing the pictures forms numerical relationships among the data of an image. Moreover, since three color space forms color images, color layers will be highly correlated to each other. Additionally when an image is manipulated, these correlations get altered and new relationships will be created. By comparing suspect images to exemplars taken from the suspected camera, inconsistencies or similarities can be identified. Adding to that, many images are saved with the JPEG compression standard. This standard can be used in a variety of image formats including JFIF and TIFF files. Since JPEG compression process introduces more relationships to an image which can be used to evaluate image authenticity. A good deal of authenticity analyzes take a general approach in determining alteration but cannot localize the exact part of an image that has been altered. To help recognize zones that have the new part, the local structure

Corresponding Author: Nadeem Alherbawi, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

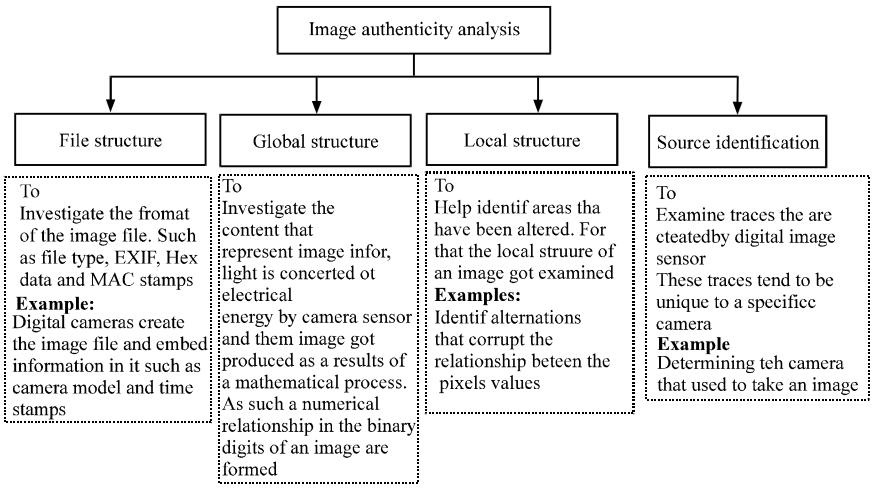

Fig. 1: Image authenticity analysis

is analyzed. These methods recognize adjustments that degenerate the relationship between the pixel values. While they are mathematically more complex than alternate examinations, these can be effective methods for distinguishing malignant forgery.

The final analysis is to examine traces that are created by the digital image sensor. These follows have a tendency to be novel to a particular camera and can consequently Another as an advanced unique finger impression to distinguish the source of the image or if two images were made by the same imaging gadget. These follows can be distinguished from flawed pixels or by utilizing the photograph reaction non-consistency of the image sensor. Attributes saw from these sorts of examinations are utilized to determine source image identification.

MATERIALS AND METHODS

Forgery techniques: One interesting point is to recognize which particular altering operations have been used in the manipulation of candidate image which provides more profound comprehension of the doctored image than just a plain black and white decision. Distinguishing proof of a particular tool utilized likewise permits an adaptable understanding of acceptable operations in practical applications. For instance, realizing that a picture has undergone a skin tone alteration helps the experts choose the picture as worthy in customer applications yet not in journalistic distribution. Every particular forgery is frequently composed in light of artifacts produced from the focused on operations, subsequently, may not be generalizable to distinctive types of manipulation.

A common form of digital image manipulation is splicing of two or more images into a single composite. Image splicing is a technology of image compositing by combining image fragments from the same or different images without further post-processing such as smoothing of boundaries among different fragments (Zhang *et al*., 2008). When performed carefully, the border between the spliced regions can be visually imperceptible. Another form of manipulation is re sampling. To create a convincing composite, it is often necessary to resize, rotate or stretch portions of an image rotate or stretch parts of an image. For example when creating a composite of two people; one person may have to be resized to match the relative heights. This process requires re-sampling the original image onto a new sampling lattice, introducing specific periodic correlations between neighboring pixels.

Furthermore, cloning is one famous form of image manipulation. That is done by cloning (copy and paste) portions of the image to conceal an object in the scene, and when this is done with patience, it can be so hard to detect this form of image manipulation by visual inspection. As cloning (copy-paste) tampering become more persuading, it is necessary to develop techniques which can still detect transformed areas and find such tampering (Kakar and Sudha, 2012).

RESULTS AND DISCUSSION

Image Forgery detection techniques and issues: This part of the study focus on reviewing the different forgery detection techniques in general. It also covers the drawbacks and issues facing the process of forgery detection in each proposed model.

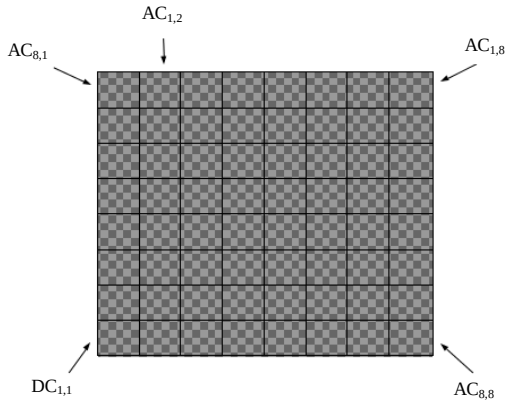

Fig. 2: DCT coefficient ordering. The DC coefficient dwells in grid area (1,1). The remaining AC coefficients are requested from the most minimal frequency to the most astounding frequency

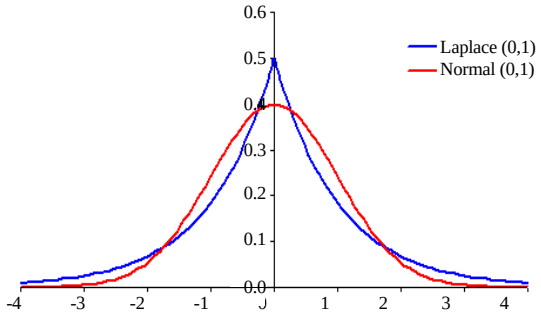

Fig. 3: The Laplacian distribution. Characterized by its sharp peak is modeled in blue. The normalized or Gaussian distribution, characterized by a smooth peak is modeled in red

DCT coefficient analysis: Investigation of the DCT coefficients can yield evidence of picture altering in the form of double JPEG compression. Commonly, when a picture is manipulated, it should first be loaded into a photograph editing software. At the point when the modifications have been made, the picture is then re-compressed. On account of a JPEG picture, this procedure groups the DCT coefficient values into products of the quantization step size (Luo *et al.* 2010).

JPEG image format uses DCT coefficient in the compression process unlike, many another image format. The Discrete Cosine Transform converts the image from the spatial domain into the frequency domain. This transformation can be clarified by thinking about the forward DCT as frequency analyzer and the backward DCT as a frequency synthesizer. Every 8×8 squares is a 64-point discrete signalwhich is an element of the two measurements, width and height. The forward DCT breaks down this square into 64 orthogonal premise signals, everyone comparing to the spatial frequencies of the input's spectrum (Fig. 2). These values are alluded to as the DCT coefficients. The DCT coefficients are divided into two sorts of signals, DC and AC parts. The DC coefficient alludes to the mean estimation of the waveform and reflects the average of the input values. The DC part normally contains a huge segment of the aggregate vitality for every JPEG block. For every pixel block, there is stand out DC segment. The remaining 63 coefficients are alluded to as the AC coefficients.

Before being processed by the quantization table, the DCT coefficient values exhibit a Laplacian distribution, characterized by a curve with a pronounced sharp peak centered on a mean value as shown in Fig. 3. However, this distribution is disrupted when the quantization table divides the DCT coefficients and the resulting values are rounded to the nearest integer as shown in (Fig. 4). The DCT histogram of each AC coefficient for a first generation JPEG image shows periodic spikes and valleys at multiples of the quantization step size as shown in Fig 4a. It is important to note that the coefficient values are periodic in a first generation JPEG image. However, the periodicity of a first generation JPEG image is still evenly distributed by the Laplacian model as shown in Fig. 4b.

At the point when a JPEG image spared a second time, this image is alluded as a 2nd generation image. At the point when the JPEG compression is done again, the DCT coefficients experience further change and show attributes of double quantization; this is formally called the double quantization effect as shown in Fig. 4c. With slight compression, the quantization step will have a little impact on an image. Conversely, expanded compression will have a more discernible impact on the last image. It ought to be noticed that the quantization tables for the luminance channel and the chrominance channels are diverse.

There are three conceivable situations with second-generation JPEG images. The principal is that the auxiliary quantization of the re-compressed image, meant as Q2 is smaller than the essential quantization Q1 of the first packed image in another word (Q2<Q1). The second is the secondary quantization step size is more than the essential quantization quality or in another word (Q2>Q1). The third scenario when (Q2 = Q1). The estimations of the quantization matrices can have contrasting effects on the distribution of the DCT coefficients.

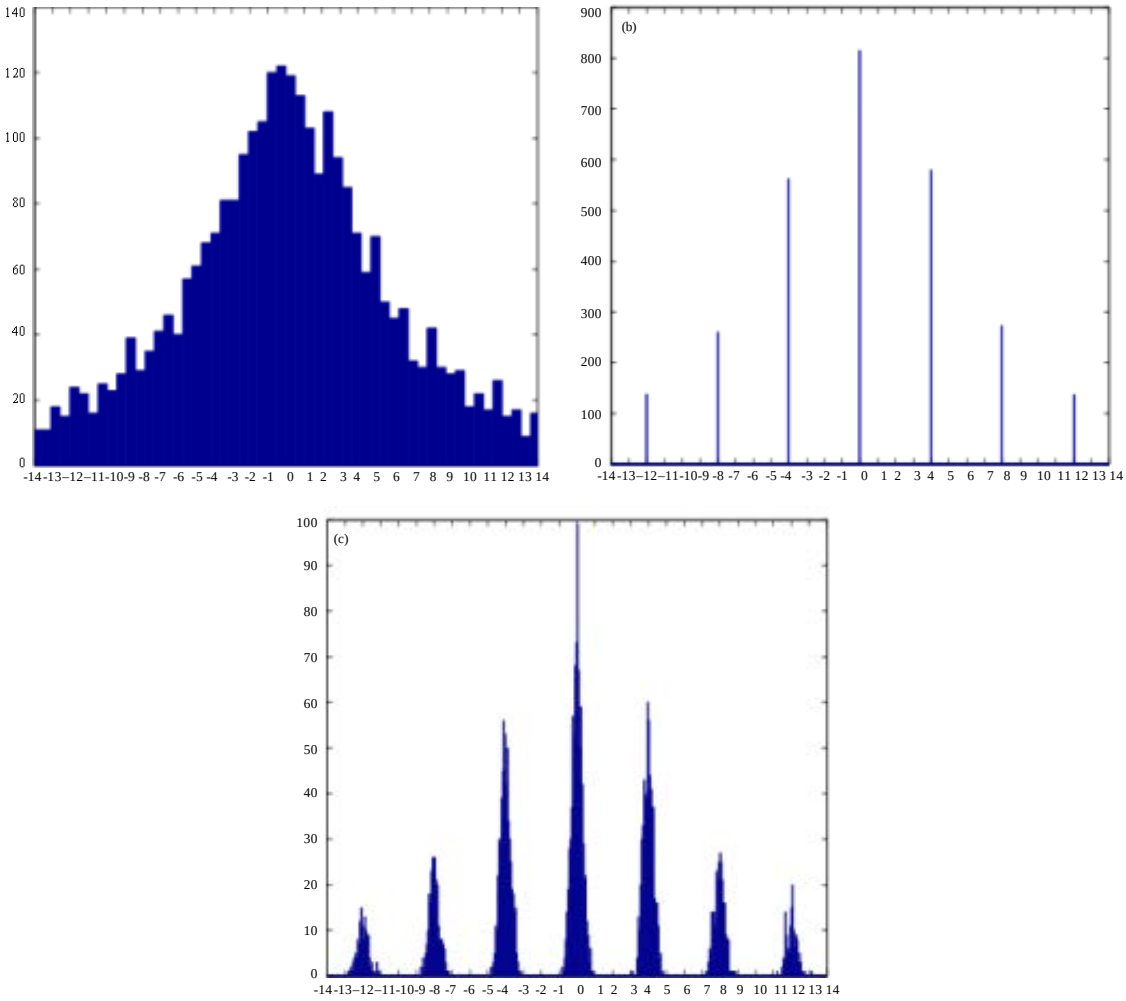

Fig. 4: DCT coefficient distribution. Distribution of AC coefficients at (1,1) for all blocks before quantization: a) Distribution of AC coefficients at (1,1) after quantization with step size = 4; b) Distribution of AC coefficients at (1,1) after being double quantized (Luo *et al*., 2010)

Figure 5 represent a scenario where tow images has been singly quantized with different quantization step the first has quantization step equal to four shown in Fig. 5a. The second image where quantized with quantization step equal to three as shown in Fig. 5b. If the secondary quantization step is more than the primary value, the DCT histogram can exhibit a periodic pattern of peaks and valleys as shown in Fig. 5c. If the secondary quantization step is less than the primary value, the histogram of the double compressed image will exhibit peaks with periodic missing values in Fig. 5d. If the two compression settings are the same, then the DCT histogram will not exhibit such a clear indication because the values of the coefficients will not be redistributed based on a secondary quantization value. These figures express the histograms of signal and double quantized signal. The first row is a signal quantized by a step size of 3 as shown in Fig. 5a and next to it is a signal quantized by a step size of 4 as shown in Fig. 5b. The bottom row is a double quantized signal by an initial step size of 3 and second of 4 as shown in Fig. 5c, next to it is a double quantized signal by an initial step size of 4 and second of 3 as shown in Fig. 5d (Redi *et al*., 2011). While the past cases speak to short-sighted perspectives of the quantization and double quantization impacts in all actuality the DCT coefficients in real images are much more complex. Since quantization step sizes can be somewhere between 0 and 255, indications of double compression may not be so promptly evident as shown in Fig. 6. The artifacts of double compression are clear in the

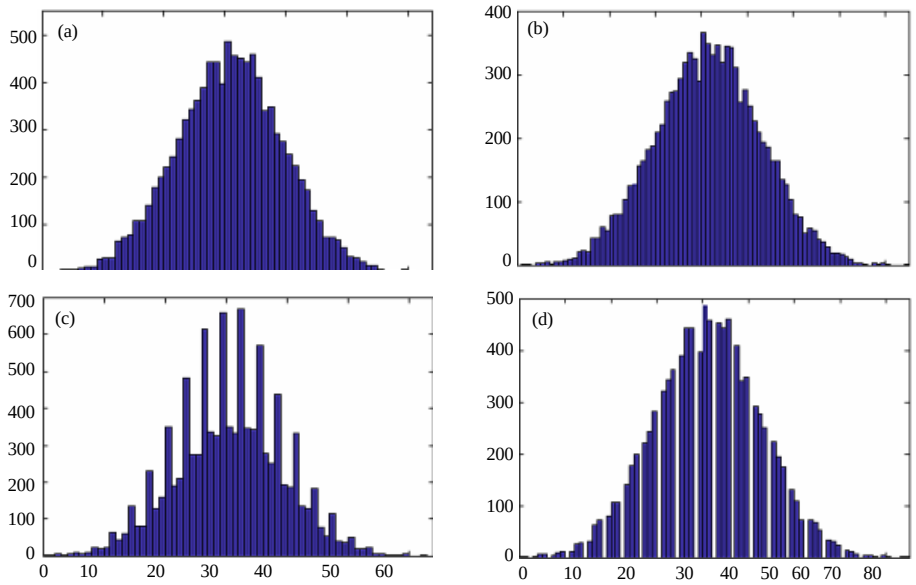

Fig. 5: Double quantization effect: a) Single quantized single, q = 4; b) Single quantized single, q = 3; c) Double quantized single, q1 = 3, q2 = 3; d) Double quantized single, q1 = 4, q2 = 3

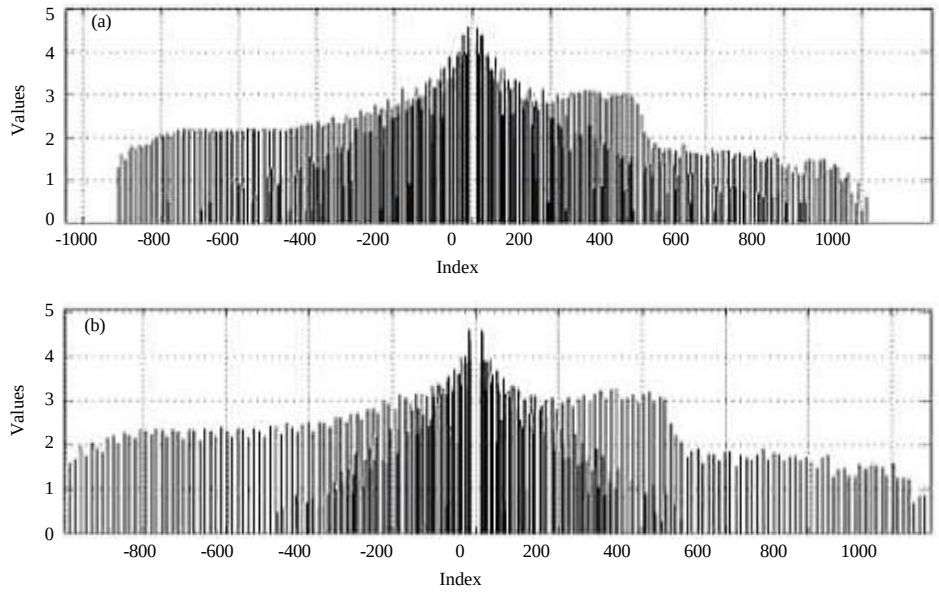

Fig. 6: DCT histogram of DC components: a) Single compression; b) Double compression

repeating pattern of 2 taller spikes, trailed by a somewhat smaller one which is most observable on the left half of the histogram in Fig. 5 and 6b). These artifacts are not shown in the original image originating from the camera as shown in Fig. 6 a. The quantization step for the DC part was 9 for the first image and 12 for the re-compressed image. In addition to the histograms, the periodicity of the DCT coefficients can also be viewed by computing the Fourier transform on the DCT histograms (Popescu and Farid, 2004). Artifacts are noticeable as sharp peaks in the Fourier domain in the higher frequencies. Figure 7a, a histogram of the AC coefficient values for position (2,2)

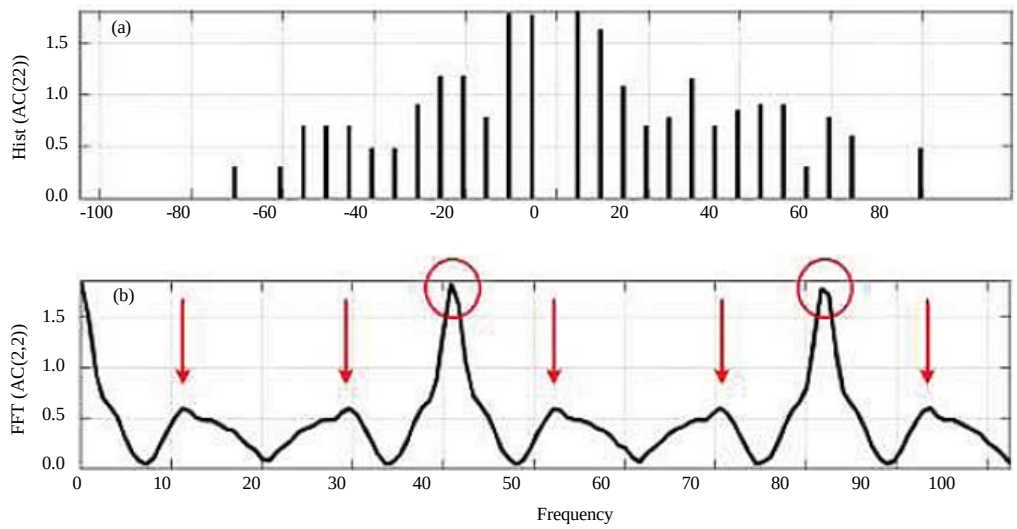

Fig. 7: Fourier transform on the DCT histograms (Popescu and Farid 2004) FFT of DCT coefficients the top panel: a) shows the AC (2,2) coefficient values for a double compressed image. The bottom panel; b) shows the FFT of the histogram. Periodicity in the histogram is shown by peaks in the FFT

is shown for a double compressed JPEG imag and the AC coefficient values are clearly distributed as The Laplacian distribution. Also an FFT of the AC coefficient values histogram shows periodicity. Figure 7b, two smaller peaks (marked with red arrows) separate the larger sharp peaks (circled in red).

Stamm and Liu (2011) point out a shortcoming in the DCT investigation when they alter the DCT coefficients of a JPEG image during the decompression of the image. A little measure of noise is added to the normalized DCT coefficients disguising the impacts of quantization error. The outcome is that the distribution of DCT coefficients in the manipulated image express those of an uncompressed image. To counter this anti-forensic attack, examination of the high-frequency sub-bands of an image can figure out whether the attack referenced above was utilized on the image or not (Lai and Bohme, 2011).

Copy and paste forgery detection: In image authentication, identification of malicious alteration in image content is largely important. One of the most common techniques is the use of copy and paste which takes information from within an image or another image and copies it over targeted image content. This kind of technique can be used in two ways. The first is to replace content that existed in the scene at the time the image was taken. The second is to add content into a scene that was not present in the original image. Accordingly, making a relationship that did not exist at the time the photo was taken. These sorts of methods can add or remove picture objects and leave no outwardly evident hints of change.

These kinds of modifications are effectively fulfilled utilizing the most image preparing software. Cloning is a particular procedure that adjusts picture content by utilizing within picture parts or objects to conceal different zones. For instanc if a white cat lies in a verdant field, the manipulator could remove the cat by utilizing grass from various regions of the picture and putting it on the cat as shown in Fig. 8. These sorts of adjustments can be difficult to recognize. The measure of the manipulated zone relies on the size of the object that the manipulator is attempting to hide. On the other hand if the area is large enough a visual assessment of the picture will uncover two objects which are rehashed in two unique parts of an image. The issue on the other hand when the region is sufficiently little or from another picture, the forgery may not be discernible by the human eye.

If the image incorporates within image change, the identification of cloning is a straightforward process. The algorithm searches for accurate pixel coordinates in a bunch of pixels to figure out whether two sections of an image are the same. This methodology takes a shot at the supposition that the copy and paste techniques will utilize vast adjoining areas of an image to modify content as opposed to numerous individual pixels. This comprehensive methodology looks at a predefined pixel group to each pixel piece of the same size to find a match. In any case, this kind of examination is computationally intensive and makes it impractical in all except small images. Also, the outcomes are reliant on the measure of correcting done to the cloned locales and the degree of

(a)

(b)

Fig. 8: Example of copy and paste forgery: a) Original picture; b) Doctored image

compression connected to the image after modification. To overcome those issues, the utilization of a robust matching method that uses lexicographically sorted DCT coefficients and sorting them by comparable spatial counterbalances was created (Fridrich *et al*., 2003).

A similar approach was used by applying the principal component analysis onto each image block (Popescu and Farid, 2004). These techniques were found to be slightly more robust when retouching was used to image content in the form of additive noise and lossy compression. Likewise, lateral color aberrations can be utilized to figure out whether bits of an image were altered (Johnson and Farid, 2006). Mostly every optical lenses contribute to color aberrations due to failures in the optics to consistently center the different wavelengths of light onto the sensor. The level of spatial shifting increases the further away the light is from the optical center of the lens.

Furthermore when an image is adjusted, these variations neglect to be reliable over the image matrix. Color aberrations are obvious at the edges of objects or in high contrast areas and are noticeable as green or red coronas that emanate outward from the optical center. At the point when a copy-paste modification happens, the bearing of the variation may be conflicting with the surrounding material. While JPEG compression did influence the exactness of this method, identification of color aberration was still discovered to be a helpful tool in recognizing the altered area.

JPEG ghosts analysis: Modifications in JPEG images can be recognized utilizing the quantization and rounding errors produced in the JPEG compression process. As specified earlier in this chapter an image is changed over from the spatial domain into the frequency domain utilizing the Discrete Cosine Transform (DCT). At that point, DCT coefficients will be quantized by quantization table. Therefore, reduction of the image quality because of the quantization dividing error or from converting a real number to the closest integer. JPEG ghost analysis focuses on the error generated as a result of quantization and dequantization process of the JPEG image.

In the majority of natural images, AC coefficient distribution can be modeled by a Laplacian curve (Reininger and Gibson, 1983). After change into the frequency space, AC coefficients group around a solitary mean worth. Anyhow, dissimilar to a Gaussian, or normalized distribution which has a smooth bell formed maxima, the AC coefficients show a maintained sharp peak at the mean value as illustrated in Fig. 3. At the point where these values are quantized utilizing the quantization table, the AC coefficients stop to be easily dispersed and get to be assembled into products of the quantization step size "Q". At the point when the compressed image is again changed over into the frequency domain, the DCT coefficients are no more equitably distributed around a solitary value but are spread with a Laplacian distribution around the multiples of the first quantization step size as indicated in Fig. 4b.

A method proposed by Luo *et al.* (2007) that can identify hints of JPEG compression in an uncompressed image. Their proposed strategy meets expectations by changing over a suspect image into the frequency area and examine the subsequent AC coefficients. On the off chance that the image had been beforehand compressed as a JPEG image, the distribution of the AC coefficients will show periodicity as clarified already. The proposed strategy claims accuracy of 98.6% on 256×256-pixel blocks, down to 95.08% on 8×8-pixel blocks. This technique could be utilized to distinguish if any piece of an uncompressed image originated from an already JPEG compressed image.

Farid (2009) proposed a technique to recognize manipulated regions of an image by recognizing the existence of "JPEG ghosts". Consider an image

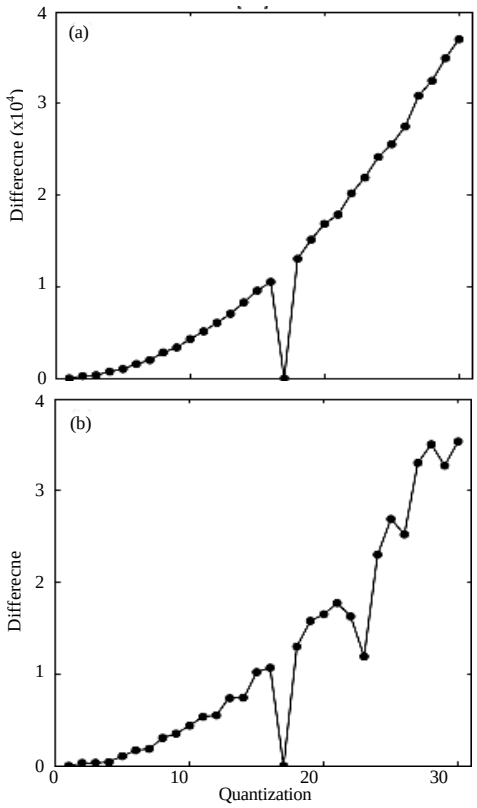

Fig. 9: Sum of the squared difference of a JPEG

compressed by quality variable q1 to make image a1. If the image is re-saved at quality factor q2, the process will result a new image. If we subtract the values of the DCT coefficients of the old image from the new one, we will have the difference between the two compressed images. If the sum of the squared difference between the DCT coefficients of the old and new image are plotted onto a chart, the distinctions will increment as the quality compression q2 builds as shown in Fig. 9a. It is important that the diagram will achieve a minimal difference when q1 = q2, demonstrating the first compression setting. Note in Fig. 9a, the minimal difference came to at the first compression quality factor equal to 17.

In another situation an image is compressed by quality variable q1 and again compressed by quality component q2 which is <q1, resulting a new image. On the off chance that the new image is compressed again with quality compression q3, the later new image will be created. Once more, plotting the total of the squared difference of the DCT coefficients for an increasing q3 will demonstrate a minimal difference at quality level q2. Likewise a second minimum will be uncovered demonstrating the degree of the first quantization level q1 Fig. 9b. This second minimum is

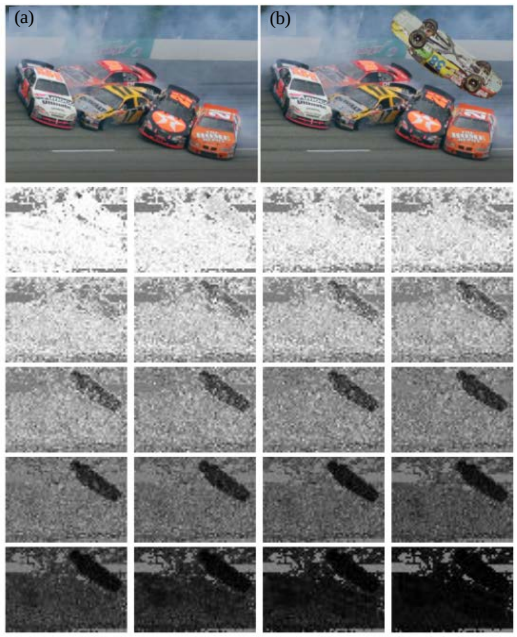

Fig. 10: The original image: a) and doctored image; b) image. Image a and b are the different images qualities range from 60-98

what is alluded to as a 'JPEG ghost'. Since various minimal are normal when quantization tables share integer multiple values, looking at the differences specifically from the image's pixel values, rather than the DCT coefficients straightforwardly can mitigate this. Forged territories are just found by taking a suspected image, re-compressing it at consecutively distinctive JPEG quality settings and subtracting every re-compressed image from the first suspect image. Any zones that have been already adjusted will bring about a JPEG Ghost showing up in the image around a particular JPEG quality values as demonstrated in Fig. 10. JPEG ghosting is exceedingly noticeable and effortlessly obvious in many cases.

Unfortunately, this technique is just valuable if the manipulated region was taken from an image compressed at a lower quality element than the investigated image. Moreover any misalignment in the 8×8 JPEG cross section structure will keep the JPEG ghost from showing up. This issue can be overcome by moving the image on a level plane and vertically onto each of the 64 conceivable arrangements before re-compacting the image at the diverse quality settings.

CONCLUSION

This study covers the importance of image authentication and in which level manipulation of the

image could occur. It also covers a set of different types of forgeries and manipulation on JPEG images and discuss the different techniques used in detecting and localizing the various types of forgery. JPEG encoding depends on DCT transformation and for that reason most of the discussed methods above discussion feature and patterns produced by DCT coefficient analysis. When forgery or manipulation occur on JPEG image the forged part undergo compression for the first time and the original parts of the image undergo compression for the second time this called double compression effect. The above methods rely on double compression and single compression artifacts in the process of detecting forgeries in JPEG images.

REFERENCES

Farid, H., 2009. Exposing digital forgeries from JPEG ghosts. IEEE Trans. Inf. Forensics Sec., 4: 154-160.

Fridrich, J., D. Soukal and J. Lukas, 2003. Detection of copy-move forgery in digital images. Proceedings of the 3rd Annual Digital Forensic Research Workshop, August 6-8, 2003, Cleveland, USA., pp: 174-184.

Johnson, M.K. and H. Farid, 2006. Exposing digital forgeries through chromatic aberration. Proceedings of the 8th Workshop on Multimedia and Security, September 26-27, 2006, ACM, Geneva, Switzerland, ISBN:1-59593-493-6, pp: 48-55.

Kakar, P. and N. Sudha, 2012. Exposing postprocessed copy-paste forgeries through transform-invariant features. IEEE Trans. Inform. Forensics Secur., 7: 1018-1028.

Lai, S. and R. Bohme, 2011. Countering Counter-Forensics: The Case of JPEG Compression. In: Information Hiding, Filler, T., T. Pevny, S. Craver and A. Ker, (Eds.). Springer, Berlin, Germany, ISBN:978-3-642-24177-2, pp: 285-298.

Luo, W., J. Huang and G. Qiu, 2010. JPEG error analysis and its applications to digital image forensics. IEEE. Trans. Inf. Forensics Secur., 5: 480-491.

Luo, W., Z. Qu, F. Pan and J. Huang, 2007. A survey of passive technology for digital image forensics. Frontiers Comput. Sci. China, 1: 166-179.

Popescu, A.C. and H. Farid, 2004. Statistical tools for digital forensics. Proceedings of the 6th International Workshop on Information Hiding, May 23-25, 2004, Berlin, Germany, pp: 128-147.

Redi, J.A., W. Taktak and J.L. Dugelay, 2011. Digital image forensics: A booklet for beginners. Multimedia Tools Appl., 51: 133-162.

Reininger, R. and J. Gibson, 1983. Distributions of the two-dimensional DCT coefficients for images. IEEE. Trans. Commun., 31: 835-839.

Stamm, M.C. and K.J.R. Liu, 2011. Anti-forensics of digital image compression. IEEE Trans. Inform. Forens. Secur., 6: 1050-1065.

Zhang, Z., Y. Zhou, J. Kang and Y. Ren, 2008. Study of Image Splicing Detection. In: Advanced Intelligent Computing Theories and Applications With Aspects of Theoretical and Methodological Issues, Shuang, D.H., D.C. Wunsch, D.S. Levine and K.J. Hyun (Eds.). Springer, Berlin, Germany, ISBN:978-3-540-87440-9, pp: 1103-1110.