ISSN NO: 9726-001X
Volume 06 Issue2, April 2018

JOURNAL OF CURRENT SCIENCE

MULTI-LEVEL ERROR ANALYSIS FOR JPEG FORENSIC IMPROVEMENTS

Sumalatha
Mtech,cse,bcetw,A,p,India

Abstract

Many advanced picture editing tools in digital image processing make it simple to make significant changes to the original photographs. These programs take use of graphical editors and sophisticated image manipulation methods. The digital forensics sector now has a major problem with picture forgery. Using internal and external CMFD, the proposed study focuses on developing a forensic system to identify forgeries of various types. In order to determine which photographs are most likely to have been manipulated, the system employs an algorithm called Multi- Level Error Analysis (MLEA) to provide a qualifier for each image. In the experimental setting, many photographs were taken using several digital cameras (Canon, iPhone, and Samsung). A third party copied and rearranged some of the pictures in this collection at random. All three quality settings (75%, 85%, and 95%) of the MLEA are used to test the established ranking methods. The data shows that our techniques perform best at an MLEA quality level of 95%. The goal is to shorten the time it takes for an expert to confirm the photos' veracity. The experimental evidence demonstrates the low computational cost and high resistance to blurring and nosing of our suggested technique against multiple copy-move forgeries.

Image processing, JPEG compression, multi-level empirical analysis, correlated multiple-fractal dynamics,

Introduction

The significance of digital photographs in the contemporary world continues to grow. With the proliferation of affordable digital cameras and the advent of camera-equipped smartphones, anybody can take and share high-quality photographs in a flash. With the ability to take pictures and alter them at will, people often have trouble believing what they see. This is not usually seen as a major problem when pictures are utilized for humorous reasons. It is crucial, however, to be able to authenticate the validity and integrity of photos in situations when they are used as evidence. It is common practice to have a professional visually examine photos to ensure they have not been tampered with. When just a limited number of photographs are being considered, this is not an issue. Doing the same for huge picture collections, however, becomes a time-consuming and laborious task. The System is concerned with the efficiency and accuracy of rating big collections of photographs based on the possibility of being modified. Today, a wide variety of methods exist for modifying images. For instance, if an individual's pupils have a red glow, they may want to eliminate the red-eye effect [1] in the photograph. Alternately, you may boost the image's contrast to make the topic more recognizable. In most cases, one of the following methods—collectively referred to as "copy & move" manipulations—is used when altering an image: picture manipulation include subtraction, addition, and alteration of existing elements. For each of these editing methods, we distinguish between "internal" and "external" copy & move. An item may be duplicated and moved about within the picture itself using internal copy & move techniques. When performing an external copy & move operation, an item is instead lifted from one picture and pasted into the target image. Figures 1.1 and 1.2 show an example of an item being cropped out of the original and the final altered picture. The original photograph, which may be seen in Figure 1.1, shows Stalin's commissar of water transport, Nikolai Yezhov. Figure 1.2 is a digitally altered version of the original photograph in which Nikolai has been digitally erased from the scene, leaving no signs for the untrained eye to detect. Here, we're talking about some kind of internal copy/paste operation.

ISSN NO: 9726-001X
Volume 02 Issue1 March2014

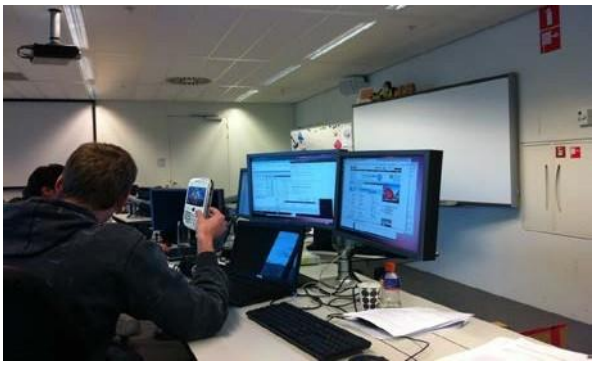

Figure 1.1: Original image

Figure 1.2: Stalin without Yezhov

Secondly Figure 1.3 illustrates an original image of two students of the University of Amsterdam. The image in Figure 1.4 was manipulated and illustrates the addition of a foreign object, namely a white-colored mobile phone that was not part of the original image. Since the mobile phone did not come from the original image itself, an external copy & move manipulation wasperformed.

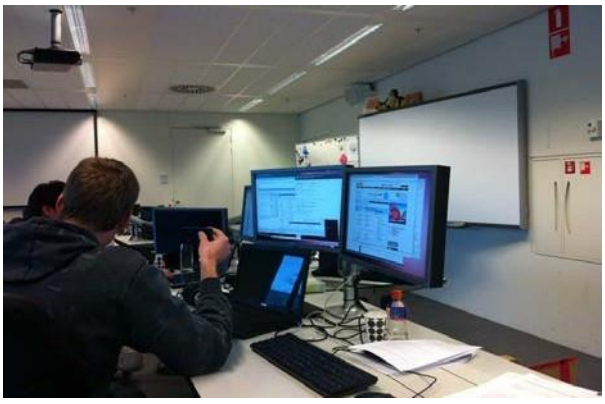

Figure 1.3: original image

ISSN NO: 9726-001X
Volume 02 Issue1 March2014

Thankfully, not only are there methods for modifying photographs, but there are also methods for identifying tampering. Joint Photographic Experts Group (JPEG), Graphics Interchange Format (GIF) [6, Tagged Image File Format (TIFF) [8, and RAW [9] are only few of the image formats used to save photographs. These picture formats store not just the information that makes up an image's apparent content, but also a plethora of supplementary data referred to as metadata. The Exchangeable Image File Format (EXIF) [10] is a common format for storing information associated with photos. The information about a picture, known as metadata, is sometimes just as valuable as the actual image itself. The image's metadata includes not only required data like its dimensions, but also optional data like the camera's make and model, the time and date of capture, and the color space used. For example, the camera model might be replaced with the software's own name in certain picture editors [11]. picture format analysis [12] is a proactive method of identifying picture alteration [13] that involves searching for such inconsistencies. The picture itself is ignored in favor of the image's information in image format analysis. Methods for detecting picture modifications, such as the ones described in [12], include Luminance Gradient (LG), Principal Component Analysis (PCA), and Wavelet Transformations (WT). These methods provide a "passive" way to detecting image tampering since they don't need any context about the picture being studied or where it came from [13]. The outcomes of the aforementioned methods are shown in a different graphic with emphasis on those methods alone. Image authenticity and integrity may be assessed by a specialist using the highlighted areas in such a visualization..

Related Work

As said, there exist different image manipulation detection techniques. Although these techniques are all related in the sensethat they share a common purpose, namely aiding in the detection of image manipulation, there is a completely different theory underlying each of them. For each of these techniques, in the order LG, PCA and WT respectively, some key aspects will be explained.

Luminance Gradient

LG is used to identify manipulation of an image by illustrating the general direction of light. To do so, LG utilizes the fact that light rarely hits an object with a uniform intensity. Instead, sections of an object that are closer to the light source will appear brighter. There are many variations of LG, however, they all aim to identify the light direction. In the simplest variation, the image is divided into squared blocks of a fixedsize, e.g. 3 by 3 pixels. For every block the direction of light is identified based on adjacent pixels in the block that appear brighter in a certain direction. This is done for every block whichresults in a set of directions, one direction per block pointing towards the brightest local light source. Finally, the color of every block is remapped based on the direction of the local light source. For example, a direction pointing to the right means all green, to the left means no green, upwards means all red, and downwards means no red. This results in another image that is then used by an expert to determine if the image wasmanipulated based on the transition of different colored surfaces.Smooth surfaces with even gradient transitions or no transition atall suggest digital manipulation or computer graphics [12].

Principal Component Analysis

PCA is used to identify the color spectrum within an image. PCA finds patterns in the image data and expresses this pattern data in a certain way to highlight their similarities anddifferences [14]. Assume an entire image is plotted in three dimensions based on the colors of the pixels: red is mapped to the X-axis, green is mapped to the Y-axis, and blue is mapped to the Z-axis. The resulting plot for most images has a narrowrange of colors that appear as a large cluster. Since the image's plot is three-dimensional, there are three Principal Components (PCs). Each PC defines a plane across the plot and emphasizes different sections of information. PC1 identifies the widest variance across the color set, PC2 the second widest variance with respect to PC1, and PC3 the smallest variance. For example, when areas of two different images and color sets are spliced together, they usually end up forming two distinct clusters. With PCA, areas within the image that come fromdifferent clusters will have noticeably different values. In theend, PCA is used to detect image manipulation by rendering the distance from each pixel to a PC in another image. Each different PC that is used for rendering will yield a differentimage that is further analyzed by an expert [12].

Wavelet Transformations

WT use wavelet characteristics to detect photo tampering. In signal analysis, a wavelet is a very specialized function. It is possible to approximate any signal by decomposing it into a collection of wavelets. The use of wavelets to create an approximation of a signal is analogous to compression. In the case of pictures, the image itself serves as the signal, and any wavelets may be used to approximate it. The total number of wavelets needed to reconstruct the picture exactly is proportional to the number of image pixels in each color channel. Although the picture becomes hazy when just a fraction of the wavelets are utilized to create it, the items inside it are still discernible. Sharpening and enhanced coloration occur when more wavelets are used. Ideally, the final result would have uniform sharpening over the whole picture. To help in the identification of picture tampering, WT makes use of this feature. various parts of a picture will sharpen at various rates [12] if the image is scaled or merged, i.e. if separate layers are used in an image editor.**Proposed**

ISSN NO: 9726-001X
Volume 02 Issue1 March2014

System

The proposed system Multi-Level Error Analysis (MLEA) isused to identify image manipulation by detecting areas in an image that have a different level of compression error compared to a given level. Essential is that MLEA makes use of the properties of image formats that utilize lossy compression. Just like the other image manipulation detection techniques explainedearlier, applying MLEA to an image results in a visual output in the form of a new image with dimensions equal to the processed image. In this newly generated image, manipulated areas with a different level of compression error stand out because they are visually contrasting in comparison to unmodified areas. An expert analyzes this image to determine if the processed image isauthentic.

Image Representation

The bulk of this section's content comes from [16], a concise introduction to picture representation. Many individual elements, or "pixels," are used to create an image. Simply said, a pixel is a square or rectangular area that is assigned a single numerical color value. The horizontal and vertical resolutions of a picture are defined by the number of horizontal and vertical pixels, respectively. If a picture has a greater resolution, it will have smoother-looking forms because more pixels will be used to create the image. A graphics format is required to define a location for each pixel. JPEG images employ the bitmap graphics format, sometimes called raster graphics format, to assign a pixel to a specific place in the picture.

Pixel color values can't be directly translated to real-world colors without a color model. A color model's color space defines the range of colors that may be represented by the model and is limited by the number of bits used to take a sample (one bit allows for black and white, two bits allow for four colors, and so on). RGB, which stands for "Red, Green, and Blue," is the most used color model in digital media. However, JPEG pictures almost invariably use another color model called YCbCr during storage. The letter Y represents the brightness, or luminance, of an image. Chroma, denoted by the letters Cb and Cr, describe the degree to which a picture leans toward blue and red, respectively.

In the YCbCr color model, the Y component is given far more weight than in the RGB color model, which gives equal weight to all three. By using more information from the Y component rather of the Cb and Cr components, JPEG picture compression may be improved. JPEG is a bitmap-based format, as was previously explained. One major downside of this graphic format is the exponential growth in file size brought on by improvements in picture quality and color space. As was previously said, picture compression is the answer.Compressing Images

At its most fundamental level, a picture is just a bunch of bits. Compression is a mathematical technique for reducing the amount of data required to describe a picture that is otherwise unchanged. This is accomplished by taking advantage of certain patterns within a data source. Compressing such a collection makes it more space- and bandwidth-efficient to store and send than an equivalent uncompressed set. Compression is a data reduction technique that requires computational resources to accomplish the task of data reduction. Thus, prior to its further usage, compressed data has to be decompressed. Both lossless and lossy compression techniques exist.To Compress Without Loss

Lossless compression algorithms lessen the quantity of data required to retain a picture while maintaining the quality of the original. In other words, after decompression, each bit retains the same value it had before compression. However, unlike lossy compression, it does not significantly decrease file size when using lossless compression. When the emphasis is not on minimizing file size at the expense of picture quality, lossless compression methods are often used.

Lossy Compression

Lossy compression algorithms make use of the limitations of the human eye, such as having a hard time to distinguish between nearly identical colors [16]. Some information can be discarded without losing much of the original visual structure. The compression levels in most lossy compression algorithms can be adjusted and as these increase, the file size is reduced, sacrificingimage quality due to image degradation. At the highest compression levels, image deterioration becomes more prominent, causing compression artifacts [17].

JPEG Image Format

Image files in the JPEG image format carry either the 'JPG' orthe 'JPEG' file extension. JPEG is an image format that utilizes a compression algorithm. In all cases described in this research report, JPEG is used as a lossy image format. A detailed description of how the JPEG compression algorithm works is beyond the scope of this document. However, in short, the compression algorithm essentially consists of the five steps listedbelow. For a more detailed description of JPEG you are advised to read [18].

1. Divide the image in blocks of 8 by 8 pixels: The first step of the compression algorithm is to divide the entire image intoblocks of 8 by 8 pixels. Each block is then further processed without any relation to the other blocks.
2. Transform the RGB color space to the YCbCr color space: Each pixel within a block is represented by RGB values and is called an RGB vector. The values in an RGB vector usually havesignificant amount of correlation and therefore need to be converted to something that has less correlation. This is done for better compression results.

ISSN NO: 9726-001X
Volume 02 Issue1 March2014

3. Apply Discrete Cosine Transformation (DCT): The heart of the JPEG compression algorithm is the Discrete Cosine Transformation. Basically, DCT transforms each block into a so called coefficient which can later be used in the process to decompress the image. DCT relies on the premise that pixels in an image exhibit a certain level of correlation with theirneighboring pixels. Consequently, these correlations can be exploited to predict the value of a pixel from its respective neighbors [19].

4. Apply quantization: The coefficients from the DCT process are stored as integers. Since integers are natural numbers, the coefficients need to be rounded before they can be saved. This is where the quantization comes in. The quantization process is the actual part of the compression algorithm that makes it a lossy compression algorithm since rounding the coefficients to integers will lead to losing some of the original data.

5. Apply Huffman encoding: The last step in the compression algorithm is to encode the transformed and quantized image. For this purpose, the Huffman encoding technique is used. The idea behind Huffman coding is to identify pixels that occur frequentlyin an image and assign them short bit representations. Pixels that occur infrequently in an image are assigned long bit representations.

1.1 Error Analysis

Lossy compression is used by JPEG, as previously indicated. This indicates that owing to quantization, some information is lost whenever a picture is stored in the JPEG format. Errors occur if critical data is lost. How much data is lost and how much inaccuracy is acquired depends on the quality setting used to save the JPEG. The quality of a JPEG file may be specified on a scale from 0 to 100. The general norm is that information loss decreases as the quality level increases in numbers. In addition, the quality of a JPEG picture is reduced when it is resaved, even if no modifications were made to the original. If a picture is compressed at 90% and then resaved at the same quality, it will take up the same amount of space as if it had been saved once at 81%. This is determined by multiplying 90% by 90%, therefore the nth resave at 90% should be about equal to 90%n [12], [20]. A similar drop in quality occurs when a picture with an original quality of 75% is resaved at a quality of 90%. It's worth noting that there is no direct relationship between the quantity of data lost and the number of times you resave. Each time you save, the JPEG compression process will only employ 8x8 blocks, limiting the amount of mistake that may be added. When a picture is (half) edited, however, the quality of the 8x8 blocks holding the alteration drops below that of the unaffected blocks. MLEA works by resaving a picture that might have been tampered with at a certain quality level, say 95%. Intentionally doing so creates a fixed rate of inaccuracy. The error state difference between the original and the resaved versions of a picture is determined by comparing identical 8x8 blocks from both photos. The lack of a significant difference between the two versions is evidence that the resaved image's block has attained its minimum error state at the specified quality level (95% in this case). If there is a substantial discrepancy, however, the block is not in a condition of minimum error. This data is then used to generate a picture with the same resolution as the suspect image, with the degree of inaccuracy for each block represented by a change in brightness. This block will seem darker in the final picture as the disparity between its quality and the chosen quality level decreases. Some examples will be provided to help clarify.The original picture quality in Figure1.5 is 96%. Images produced by the MLEA at 75%, 85%, and 95% quality are shown in Figures 1.6, 1.7, and 1.8, respectively. The MLEA results show that the error variance across all blocks is minimal.

Figure:1.5 Original image

Figure: 1.6 MLEA at 75%

ISSN NO: 9726-001X
Volume 02 Issue1 March2014

Figure: 1.7 MLEA at 85% Figure:1.8 MLEA at 95%

4. Conclusion

Although the MLEA algorithm may be used to identify common picture editing methods, human verification by a specialist is still required.picture ranking for fakeness detection. There is not enough evidence to conclude that modified photographs are given a better rating than legitimate images, hence the ranking will not benefit an expert by cutting down on their workload. In conclusion, MLEA may be used to order a collection of images. Some of the modified photographs in the collection may be sorted using our approaches. However, when searching for all modified photographs in a collection, a human inspection of each image still seems to be the only practical choice.

5. References

[1] Red-eye effects and their scientific explanation, March 2011; S. Preston. View more information on the red eye effect here: http://www.cameratechnica.com/2011/03/14/ the-science- behind-the-red-eye-effect/.

[2] According to [2] "Dwt-pca (evd) based copy- move picture forgery detection," tech.rep., The School of Computer and Communications, Hunan University, January 2011.

[3] In November 2009, researchers T. Shahid and A. B. Mansoor published "Copy-move forgery detection algorithm for digital photos and a novel accuracy measure" as a technical report for the College of Aeronautical Engineering at the National University of Science and Technology.

[4] The disappearance of the commissar, September 1999 [4].Go to http://www.newseum.org/berlinwall/commissar_vanishes/vanishes.htm for more information.

[5] In 2006, D.Lucas wrote on Katie Couric's weight loss. Visit this link for more information: http://www.famouspictures.org/mag/index.php?title=Alter ed_Images#Katie_Couric.27s_weight_loss_-_2006.

[6] Technical report from the Portable Network Graphics (PNG) Development Group, 1999 [6].

[7] [7] Png (portable network graphics) spec_cation, version 1.0," tech. rep., CompuServe Incorporated, 1990, by M. Adler, T. Boutell, and J. Bowler.

[8] Tech. rep., Adobe Systems Incorporated, June 1992 [8] A. D. Association, "Ti_ revision 6.0."

[9] Raw data: an explanation, [9] M. Reichmann, "Understanding raw _les."Check out this lesson on u-raw files at http://www.luminous-landscape.com/tutorials/understanding-series/.

[10] Technical Report, Japan Electronics and Information Technology Industries Association, Exchangeable Image File Format for Digital Still Cameras, Version 2.2, April 2002 [10].

[11] May 2009 technical report by R. Tortorella entitled "Image doctoring: Jpeg encoding and analysis," published by the National Aviation Reporting Center on Unexplained Phenomena.

[12] An Example of Digital Image Analysis," Technical Report, 2007, by N. Krawetz.

[13] March 2009 edition of "Detecting picture forgery - state of the art" [13].

[14] According to [14] L. I. Smith's "A tutorial on principle components analysis," a technical report from the University of Otago published in February 2002.

[15] Tech. rep., Swiss Federal Institute of Technology, September 2000; D. Santa-Cruz and T. Ebrahimi, "An analytical evaluation of jpeg 2000 features."File Compression for Digital Photographs: JPEG, PNG, GIF, XBM, and BMP by J. Miano. Published by ACM Press in July 2009.

[16] Review of postprocessing strategies for compression artifact removal, technical report, March 1998, authors: M.-Y. Shen and C.-C. J. Kuo.

[17] Image compression: Perceiving absent details, by D. Austin [17]. Image compression may be found at http://www.ams.org/samplings/feature-column/fcarc.

[18] According to [18] S. A. Khayam, "The Discrete Cosine Transform (dct): Theory and Application," technical report, March 2003, Electrical and Computer Engineering Department, Michigan State University.