Image Forgery Detection: Developing a Holistic Detection Tool

Andrew Levandoski and Jonathan Lobo

I. Introduction

In a media environment saturated with deceiving news, the threat of fake and altered images in our lives has become increasingly apparent. Top-grossing movies contain countless synthesized images that, not long ago, would have been extremely difficult to produce. Further, today's smartphones are capable of digitally manipulating even ordinary photographs with very little effort. Increasing capabilities in computer graphics and artificial intelligence have not only enabled new ways to analyze and create images and videos, but also have the ability to do so on an increasingly vast scale.

> *"Why did Stalin airbrush those people out of those photographs? Why go to the trouble? It's because there is something very, very powerful about the visual image. If you change the image, you change history. We're incredibly visual beings. We rely on vision – and, historically, it's been very reliable. And so photos and videos still have this incredible resonance. How much longer will that be true?"* [1]
>
> Hany Farid, photo-forensics expert, on the growing impact of forged images

With these dangers in mind, we see great value in a tool capable of identifying fake images and reporting to users the nature of an image's alterations. In this document, we will detail our progress so far in developing such a tool. Section II discusses the detection methods that we have examined along with their various implementations. Section III discusses the structure of the tool. Section IV concludes the document and previews future work for the tool.

II. Method Evaluation

The following methods presented are not a completely comprehensive set of detection algorithms; however, we have chosen them for the prolific nature of the forgeries that they detect. For each strategy (double JPEG compression detection, copy-move detection, color filter array (CFA) artifact detection, noise variance inconsistency detection), we have studied several methods to arrive at the most accurate algorithms to include in our tool. This process involved a comprehensive literature review – the methods presented here were initially chosen for their high reported accuracies, number of citations, and availability and ease of implementation. Upon testing each method against its peers on an identical set of data appropriate to each group, we arrived at a clear choice for each method in terms of accuracy and efficiency.

A. Double JPEG Compression Detection

Due to high compression ratio and good quality, the JPEG image format has been widely used in cameras and image processing software. Double compression in JPEG images occurs when a JPEG image is decompressed to the spatial domain and then resaved with a different (secondary) quantization matrix. Considering that double JPEG compression typically occurs when altering JPEG images, it is of great significance to us in detecting image forgeries.

During JPEG compression, the image is first divided into disjoint 8x8 pixel blocks B_{rs}, $r, s = 0, \dots, 7$. Each block is transformed using the discrete cosine transformation (DCT):

$$d_{ij} = \sum_{r,s=0}^{7} \frac{w(r)w(s)}{4} cos\frac{\pi}{16} r(2i+1) cos\frac{\pi}{16} s(2j+1) B_{rs}$$

where $w(0) = (1)/(\sqrt{2})$ and $w(r > 0) = 1$. The DCT coefficients d_{ij} are then divided by quantization steps stored in the quantization matrix Q_{ij} and rounded to integers

$$D_{ij} = round\left(\frac{d_{ij}}{Q_{ij}}\right), i, j \in \{0, \dots, 7\}.$$

The JPEG compression finishes by ordering the quantized coefficients along a zig-zag path, encoding them, and finally applying lossless compression. Decompression works in the opposite order. After reading the quantized DCT blocks from the JPEG file, each block of quantized DCT coefficients D is multiplied by the quantization matrix $Q, \hat{d}_{ij} = Q_{ij} \cdot D_{ij}$, and the inverse discrete cosine transformation (IDCT) is applied to $\hat{d}_{ij}$. The values are finally rounded to integers and truncated to a finite dynamic range (usually [0, 255]). The block of decompressed pixel values $\hat{B}$ is thus

$$\hat{B} = trunc\left(round\left(IDCT\left(Q_{ij} \cdot D_{ij}\right)\right)\right), i, j \in \{0, \dots, 7\}.$$

Due to the rounding and truncation involved in compression and decompression, $\hat{B}$ will, in general, differ from the original block B. We say that a JPEG image has been *double-compressed* if the JPEG compression was applied twice, each time with a different quantization matrix with the same alignment with respect to the 8x8 grid.

Yang et al. analyze the error block in JPEG compression, showing the statistical differences of the error blocks between singly and doubly compressed images and proposing a set of features to characterize such differences. They adopt a support vector machine (SVM) to learn the discriminability from the 13 extracted features for detecting double JPEG compression with the same quantization matrix. Hou et al. propose a more powerful JPEG compression detection method based on the extended first digit features of DCT coefficients. They first assume that the value 0 is the first digit of the coefficient with value zero, and then use the probabilities of the first digits of quantized DCT coefficients including value 0 form individual alternating current modes to detect double compressed JPEG images. Their experimental results show very robust performance that outperforms the other existing algorithms available.

Thing et al. introduce a method where each round of classifier is generated from a unique, non-overlapping and small subset composing …

Wang and Zhang propose a double JPEG compression algorithm based on a convolutional neural network (CNN). The CNN is designed to classify histograms of DCT coefficients. The histograms were extracted as the input, and then a one-dimensional CNN is designed to learn features automatically from these histograms and perform classification. Their method produced encouraging reported results; however, it has some limitations. The computational complexity of the CNN is considerably high, thus generating a tradeoff between the localization accuracy capability and the computational effort required. Finally, Pevny and Fridrich present a method for the detection of double JPEG compression using SVM classifiers with feature vectors formed by histograms of low-frequency DCT coefficients. The double compression detector is implemented with a soft-margin SVM with the Gaussian kernel $k(x, y) = \exp\left(-\gamma \parallel x - y \parallel^2\right.$. An important feature of the method is its ability to detect double JPEG compression not only for cover images but also for images processed using steganographic algorithms. They built a maximum likelihood estimator of the primary quality factor in double compressed images. Since the main application of their work is steganalysis, the estimator was constructed to work for both cover and stego images. They evaluate the accuracy of their estimator on a large set of images with 34 primary quality factors, achieving a reported accuracy better than 90%.

B. *Copy-Move Detection*

Copy-move image forgery involves using spliced areas from the same or different image or images to produce new objects or hide areas in the forged image. Very often this is performed with the intention to make an object “disappear” from the image by covering it with a segment copied from another part of the image. Textured areas, such as grass, foliage, gravel, or fabric with irregular patterns, are ideal for this purpose because the copied areas will likely blend with the background and the human eye cannot easily discern any suspicious artifacts. Because the copied parts can come from the same image, their noise components, color palettes, dynamic ranges, and other important properties will be compatible with the rest of the image and thus will not necessarily be detectable using methods that look for incompatibilities in statistical measures in different parts of the image. To make the forgery even harder to detect, one can use feathered crop or retouching tools to further mask any traces of the copied-and-moved segments.

Any copy-move forgery where copied regions come from the same image introduces a correlation

between the original image segment and the pasted one. This correlation can be used as a basis for a successful detection of this forgery. Because the forgery will likely be save in a lossy JPEG format, and because tools to further hide the manipulation may have been used, the segments may not match exactly but only approximately. Thus, the requirements for a detection tool include:

1) The detection algorithm must allow for an approximate match of small image segments.

2) The detection algorithm should work in a reasonable amount of time while introduction few false positives.

3) The tool should be designed with the assumption that forged segments will likely appear in connected components rather than as a collection of very small patches or individual pixels.

4) The detection algorithm should be robust to rotation, scaling, and blurring of the copied region.

Mahmood et al. divide images into overlapping square blocks and use DCT components to represent each block. Gaussian RBF kernel PCA is applied to each block to produce a lower-dimensional feature vector representation that increases the efficiency of feature matching. Their results demonstrate high precision in detecting multiple copy-move forgeries in the same image, even in the presence of blurring, noise, and compression. Evdokimova and Kuznetsov propose a method for copy-move forgery detection using Local Derivative Pattern (LDP) based features. The LDP feature is computed by applying n-order derivatives in the neighborhood of a central pixel and assigning a code to each pixel by comparing the derivatives along the same direction for two adjacent pixels. A hash value is calculated from the LDP and used to construct a histogram in which frequency of appearance is used to identify forged regions. The LDP-based method is robust to distortions of the duplicated area and has lower computational complexity than competing algorithms.

Cozzolino et al. propose a method for image forgery detection using local descriptors based on the image noise residual that adapts methods commonly applied in steganalysis. Local residual features are extracted using a CNN closely associated with the Bag-of-Words paradigm before being passed through a linear SVM for classification. Their method achieves high accuracy for forgery detection in the presence of median filtering, Gaussian blurring, additive noise, resizing, and JPEG compression. Ulutas and Muzaffer utilize AKAZE features and nonlinear scale space to detect copy-move forgery, with specific focus on detecting object removal forgeries. AKAZE features use nonlinear diffusion filtering to preserve object boundaries that are often lost due the Gaussian blurring utilized by other features such as SIFT and SURF. Their method achieves high precision and is robust to rotation, blurring, additive white noise, and JPEG compression.

Alberry et al. propose an algorithm to detect copy-move forgery using Scale Invariant Feature Transform (SIFT) and Fuzzy C-means (FCM). They construct local feature descriptors using SIFT and perform fuzzy clustering on feature keypoints using FCM, in which each keypoint can belong to multiple clusters. The resemblance between descriptors is used to determine whether forgery is present. The authors achieve speedup of over 15% over the traditional SIFT algorithm by clustering only on central keypoints of different regions instead of identifying all keywords in the picture.

C. CFA Artifact Detection

Among the numerous fingerprints that can be left behind during an image forgery, CFA artifact detection involves those left by the interpolation process. Image interpolation is the process of estimating values at new pixel locations by using known values at neighboring locations. During the image life cycle, there are two main phases in which interpolation is applied:

1) Acquisition process to obtain the three color channels (red, green, and blue). The light is filtered by the CFA before reaching the sensor so that for each pixel only one particular color is gathered. Thus, starting from a single layer containing a mosaic of red, green, and blue pixels, the missing pixel values for the color layers are obtained by applying the interpolation process, also called demosaicking.

2) Geometric transformations to obtain a transformed image. When scaling, rotation, translation, and shearing, are applied to an image, pixels within the image are relocated to a new lattice, and new intensity values must be assigned to such positions by means of interpolation of the know values. This is also called resampling.

The artifacts left in the image by the interpolation process can be analyzed to reveal image forgery.

Ideally, an image captured with a digital camera, in the absence of post-processing, will show demosaicking artifacts on every group of pixels corresponding to a CFA element. On the contrary, demosaicking inconsistencies between different parts of the image, as well as resampling artifacts in all or part of the analyzed image, will but image integrity in question.

Katre and Chandel focus on the study of demosaicking artifacts at the local level. By means of an analysis of such traces they localize image forgeries whenever the presence of CFA interpolation is not present. They propose a new feature that measure the presence of these artifacts even at the smallest (2x2) block level, thus providing a forgery map with very fine localization as output. The authors assume s(x, y) to be an observed image where $(x, y) \in Z^2$. The prediction error can be obtained as:

$$e(x,y) = s(x,y) - \sum K(u,v)s(x+u, y+v)$$

where $K_{u,v}$ is a bidimensional prediction filter. In the ideal case $K_{u,v} = h_{u,v}$ where $h_{u,v}$ is the interpolation kernel of the demosaicking algorithm. In general, they assume that $K_{u,v} \neq h_{u,v}$ since the in-camera demosaicking algorithm is usually unknown. By assuming that the local stationarity of the prediction error is valid in a $(sK = 1)X(2K + 1)$ window, it is possible to define the local weighted variance of the prediction error:

$$\sigma_e^2(x,y) = \frac{1}{c}\left[\left(\sum_{i,j=-k}^{k} \alpha_{ij} e^2(x+i, y+j)\right) - (\mu_e)^2\right]$$

Where α_{ij} are of suitable weight and μ_e is a local weighted mean of the prediction error and c is a scale factor that makes the estimator unbiased for each pixel class. Using their proposed feature, it is possible to find an imbalance between the local variance of prediction errors when an image is demosaicked – if the local variance of the prediction error of acquired pixels is higher than that of the interpolated pixels, the expected value of L(k,1) is a nonzero positive amount. If the image is not demosaicked, the difference is zero.

Prakash et al. present a frequency domain method for image demosaicking detection. Frequency domain-based techniques are used to obtain the luminance and chrominance of the image which gives the accurate information about pixel distribution. In the next step, a bilinear interpolation approach is utilized to estimate the missing values of the pixels. To overcome existing issues of the efficient reconstruction of images, they propose a radial basis neural network approach for image reconstruction.

Fernandez et al. present a 4-step approach to CFA artifact detection:

1) Assuming that the configuration of the CFA pattern is known, a simple estimation of the interpolation kernel for the green channel is generated based on ordinary least squares. Only acquired pixels will be considered to get a better estimation.

2) An estimation of the image is obtained by using the interpolation kernel computed in the previous step on every pixel. Then, from the residuals between the estimation and the original image the standard deviation for interpolated and acquired pixels is computed.

3) Next, a probability map is generated to decide if a pixel belongs to the set of resampled data. However, the complimentary error function is used, which defined the probability of a pixel belonging to the resampled set.

4) Finally, the DCT is applied on blocks of size *BxB* to verify the presence of the CFA artifacts within the block. The DCT coefficient for the highest frequency is considered as an indicator to detect tampering. Unusual values (lower or higher than expected) in the coefficient for the highest frequency after applying DCT provide evidence for image tampering.

Singh et al. develop a CFA artifact detection method for videos; however, it is designed to be applied to individual frames and is therefore suitable to study in this work. The authors obtain the probability of the presence or absence of CFA artifacts in every block of a given image conditioned on the observed values of their chosen feature ***L*** (see Katre and Chandel) using a Bayesian approach. They denote the hypotheses of presence or absence of CFA artifacts in a given image by M_1 and M_2, and since for a tampered image both M_1 and M_2 are true, ***L****(k,l)* can be modeled as a mixture of two Gaussian distributions. Their model is used to generate a likelihood map that indicates the probability of every block of a given image as being authentic or forged based on the probability of the presence of CFA artifacts in that block. To further improve localization accuracy, the authors employ a low-pass spatial filter to better highlight the connected regions in the

forgery map since tampered regions are typically physically connected.

D. *Noise Variance Inconsistency*

A commonly used tool to conceal traces of tampering in the addition of locally random noise to the forged image regions. Typically, the amount of noise in an authentic image is uniform across the entire image. Adding locally random noise my cause inconsistencies in the image's noise. Therefore, detection of various noise levels in an image may signal tampering.

Noise degradation is the main cause of failure of most existing blind forgery detection methods. These methods are able to work correctly when the amount of present noise is small. For example, in CMFD, additive noise causes duplicated regions to not match closely. This causes a significant decrease in the performance of CMFD methods. The same phenomenon can be observed in resampling detection methods which are almost always necessary when two or more images are spliced together. In this case, noise degradation causes loss of detectable correlation among neighboring pixels. This correlation is brought into the signal by the interpolation step. Further, when two or more images are spliced together, the forged image may then contain several regions with various noise levels.

Pan et al. (2011) propose a detection method to effectively locate image forgeries based on inconsistency in image noise levels by first segmenting the image into blocks for initial noise estimation (Zoran and Weiss). They then cluster the blocks into clean and tampered blocks. The detected suspicious regions are further segmented into smaller blocks for refined noise estimation and classification in the second phase to obtain final detailed detection results.

Pan et al. (2012) expand upon their initial research by estimating noise variances across different regions in an image to take advantage of a statistical regularity of natural images – that the kurtosis values of natural images in general band-pass filtered domains are positive and tend to be close to a constant. Then, approximating kurtosis of natural images across different band-pass filtered channels to be a positive constant, they construct an objective function using the relationship between the image kurtosis and noise variance in the band-pass filtered domain to estimate the global noise variance of the entire image. Their objective function is robust to infrequent outlying kurtosis values. More appealingly, the objective function has a closed-form optimal solution. Spliced regions are detected by segmenting the estimated local noise variances.

Kobayashi et al. exploit the nature of photon shot noise mixed into image signals (which depends on the cameral model). Photon shot noise results from the quantum nature of photons, where the variance of the number of photons coming into a camera is strongly correlated to the mean following a Poisson distribution. Thus, this correlation between the variance and the mean can be used as a powerful clue to detect inconsistencies in forged images. Given an image that contains some forged regions, the authors first analyze noise characteristics at each pixel. Once they obtain the per-pixel characteristics, the noise level functions (NLFs, Liu et al.) are fitted to the distribution using maximum a posteriori estimation. Likelihood is defined as the chi-square distribution to deal with the fluctuation in the noise characteristics resulting from a limited amount of sampled data. They simultaneously estimate the posterior probability of forgery and the parameters of the NFL using the expectation maximization algorithm. They represent an NLF as a linear combination of its basis functions by synthesizing a number of NLFs corresponding to various noise parameters to obtain a set of linear basis functions via PCA.

Finally, Mahdian and Saic propose a method based on a few main steps: wavelet analysis, tiling sub-band *HH* (which gives the diagonal details of the image with the highest resolution) with non-overlapping blocks, blocks noise variance estimation, and blocks merging.

III. Image Forgery Detection Tool

To avoid redundant operations and to increase the efficiency of our forgery detection tool, we evaluated each method against its peers on the same respective sets of images containing forged and natural images of the same nature to determine the detection accuracy of each algorithm. The resulting accuracies are reported in the following tables.

DOUBLE JPEG COMPRESSION	
YANG ET AL.	.953
HOU ET AL.	**.991**
THING ET AL.	.908
WANG AND ZHANG	.796
PEVNY AND FRIDRICH	.918

COPY-MOVE DETECTION	
ALBERRY ET AL.	**.982**
ULATAS AND MUZAFFER	.812
MAHMOOD ET AL.	.934
COZZOLINO ET AL.	.942
EVDOKIMOVA ET AL.	.879

CFA ARTIFACT DETECTION	
KATRE ET AL.	.752
PRAKASH	.786
FERNANDEZ ET AL.	.86
SINGH ET AL.	**.906**

NOISE VARIANCE INCONSISTENCY	
PAN ET AL. (2011)	.786
PAN ET AL. (2012)	**.81**
KOBAYASHI ET AL.	.745
MAHDIAN AND SAIC	.784

IV. Conclusion and Next Steps

In this paper, we have presented only a small subset of the techniques available for the detection of forgeries in images. Further, even the state-of-the-art in terms of detection accuracy may not offer a realistic solution for all image detection applications due to run time concerns. It is for these reasons that this tool and our research will be ongoing as new detection methods are discovered and as methods to evade these detection methods proliferate.

Currently, our immediate focus is on distribution and optimization of this tool so that it can be used efficiently by anyone, regardless of technical knowledge. Currently, the tool exists as a command line interface, available on GitHub. In a matter of clicks, the tool and its requisite libraries can be installed on any machine and run with a single script. Our next efforts will involve moving towards a web-based distribution, through which a graphical user interface can enable even easier use of the tool. Further, we hope to redesign some of our more time-consuming operations and algorithms to run in a distributed fashion so that the time constraints of using our tool become more realistic.

Acknowledgments

We would like to thank our mentor for her guidance and support throughout the development of this research.

References

[1] Rothman, J. (2018). In the Age of A.I., Is Seeing Still Believing?. [online] The New Yorker. Available at: https://www.newyorker.com/magazine/.

[2] P. Ferrara, T. Bianchi, A. De Rosa and A. Piva, "Image Forgery Localization via Fine-Grained Analysis of CFA Artifacts," in IEEE Transactions on Information Forensics and Security, vol. 7, no. 5, pp. 1566-1577, Oct. 2012.

[3] A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of re-sampling," IEEE Trans. Signal Processing, vol. 53, no. 2, pp. 758–767, 2005.

[4] A.C. Popescu and H. Farid, "Exposing digital forgeries by detecting duplicated image regions," Dept. Comput. Sci., Dartmouth College, Tech. Rep. TR2004-515, 2004.

[5] B. Mahdian and S. Saic, "Detection of copy move forgery using a method based on blur movement invariants," Forensic Sci. Int., vol. 171, pp. 180–189, 2007.

[6] H. Farid, "Detecting digital forgeries using bispectral analysis," AI Lab, Massachusetts Institute of Technology, Tech. Rep. AIM-1657, 1999.

[7] H. Farid and S. Lyu, "Higher-order wavelet statistics and their application to digital forensics," in Proc. IEEE Workshop on Statistical Analysis in Computer Vision (in conjunction with CVPR), Madison, WI, 2003.

[8] H. Gou, A. Swaminathan, and M. Wu, "Noise features for image tampering detection and steganalysis," in Proc. IEEE Int. Conf. Image Processing, San Antonio, TX, 2007, vol. 6, pp. 97–100.

[9] J. Fridrich, D. Soukal, and J. Lukás, "Detection of copy move forgery in digital images," in Proc. Digital Forensic Research Workshop, Aug. 2003.

[10] J. Lukas and J. Fridrich, "Estimation of primary quantization matrix in double compressed JPEG images," in Proc. Digital Forensic Research Workshop, Cleveland, OH, Aug. 2003.

[11] M. Kirchner, "Fast and reliable resampling detection by spectral analysis of fixed linear predictor residue," in ACM Multimedia and Security Workshop, 2008, pp. 11–20.

[12] M. K. Johnson and H. Farid, "Exposing digital forgeries in complex lighting environments," IEEE Trans. Inform. Forensics Security, vol. 3, no. 2, pp. 450–461, 2007.

[13] M. K. Johnson and H. Farid, "Metric measurements on a plane from a single image," Dept. Comput. Sci., Dartmouth College, Tech. Rep. TR2006-579, 2006.

[14] S. Bayram, I. Avcibas, B. Sankur, and N. Memon, "Image manipulation detection," J. Electron. Imaging, vol. 15, no. 4, p. 41102, 2006.

[15] S. Lyu and H. Farid, "How realistic is photorealistic?" IEEE Trans. Signal Processing, vol. 53, no. 2, pp. 845–850, 2005.

[16] S. Ye, Q. Sun, and E. C. Chang, "Detecting digital image forgeries by measuring inconsistencies of blocking artifact," in Proc. IEEE Int. Conf. Multimedia and Expo, Beijing, China, 2007, pp. 12 15.

[17] W. Luo, Z. Qu, J. Huang, and G. Qiu, "A novel method for detecting cropped and recompressed image block," in Proc. IEEE Conf. Acoustics, Speech and Signal Processing, Honolulu, HI, 2007, pp. 217–220.

[18] Z. Fan and R. L. de Queiroz, "Identification of bitmap compression history: JPEG detection and quantizer estimation," IEEE Trans. Image Process., vol. 12, no. 2, pp. 230–235, 2003.