Overview of State-of-the-Art in Digital Image Forensics

H. T. SENCAR and N. MEMON

Department of Computer and Information Science,
Polytechnic University,
Brooklyn, NY 11201, USA
E-mail: {taha,memon}@isis.poly.edu

Digital images can now be easily created, altered, and manipulated with no obvious traces of having been subjected to any of these operations. There are currently no established methodologies to verify the authenticity and integrity of digital images in an automatic manner. Digital image forensics is an emerging research field with important implications for ensuring the credibility of digital images. In an attempt to assist these efforts, this chapter surveys the recent developments in the field of digital image forensics. Proposed techniques in the literature are categorized into three primary areas based on their focus: image source identification, discrimination of synthetic images, and image forgery detection. The main idea of the proposed approaches in each category is described in detail, and reported results are discussed to evaluate the potential of the methods.

Keywords: Digital Image Forensics; Digital Cameras; Scanners; Source Identification; Tamper Detection.

1. Introduction

In the analog world, an image (a photograph) has generally been accepted as a "proof of occurrence" of the depicted event. In today's digital age, the creation and manipulation of digital images is made simple by low-cost hardware and software tools that are easily and widely available. As a result, we are rapidly reaching a situation where one can no longer take the authenticity and integrity of digital images for granted. This trend undermines the credibility of digital images presented as evidence in a court of law, as news items, as part of a medical record or as financial documents since it may no longer be possible to distinguish whether a given digital image is the original or a (maliciously) modified version or even a depiction of a real-life occurrences and objects.

This is especially true when it comes to legal photographic evidence. The Federal Rules of Evidence are shaped and drafted to deal with conventional (analog) photography. Digital photography, on the other hand, is fundamentally different from conventional photography in the way it is created, stored, and edited. Federal Rules do not currently set forth requirements for the admissibility of digital images, and, therefore, traditional notions of relevancy and authentication currently govern. Moreover, the problem becomes much more complicated (with possibly far more se-

vere consequences) when the digital image is synthetically generated to convey the depiction of a non-existent scene or object as the existing safeguards are not well suited to verify the integrity and authenticity of such visual evidence. The struck down of a 1996 child pornography law that prohibited the possession and distribution of synthetically generated images by United States Supreme Court, in April 2002, is very important in this context.[1] This ruling brought with it an immediate need for tools and techniques that can reliably discriminate natural images from the synthetic ones in order to be able to prosecute abusers. Another pressing issue concerning digital imagery is the ease with which processing tools and computer graphics algorithms can be used to modify images. The increasing appearance of digitally altered forgeries in mainstream media and on the internet is an indication of the serious vulnerability that cast doubt on integrity of all digital images. In,[2] some well known examples of digital tampering can be found.

To address these immediate problems, digital image forensics research aims at uncovering underlying facts about an image. For example digital image forensics techniques look for authoritative answers to questions such as:

- Is this image an "original" image or was it created by cut and paste operations from different images?
- Does this image truly represent the original scene or was it digitally tampered to deceive the viewer?
- What is the processing history of the image?
- What parts of the image has undergone processing and up to what extent?
- Was the image acquired by a source manufactured by vendor X or vendor Y?
- Did this image originate from source X as claimed?

The above questions are just a few examples of issues faced routinely by investigation and law enforcement agencies. However, there is a lack of techniques and methodologies that could determine the origin and potential authenticity of a digital image. Although digital watermarks have been proposed as a tool to provide authenticity to images, it is a fact that the overwhelming majority of images that are captured today do not contain a digital watermark. And this situation is likely to continue for the foreseeable future. Hence in the absence of widespread adoption of digital watermarks, we believe it is imperative to develop techniques that can help us make statements about the origin, veracity and nature of digital images.

The past few years have seen a growth of research on image forensics. The work has focused mainly on three types of problems:

(1) Image source identification to determine through what data acquisition device a given image is generated, *e.g.,* digital-camera or scanner. This entails associating the image with a class of sources that have common characteristics (*i.e.,* device model) and matching the image to an individual source device.
(2) Discrimination of synthetic images from real images to identify computer gen-

erated images which does not depict a real-life occurrence.

(3) Image forgery detection to determine whether a given image has undergone any form of modification or processing after it was initially captured.

To address these problems several techniques have been proposed. In this chapter, we will give an overview of state-of-the-art digital image forensics techniques. The outline of the chapter is as follows. In Section 2, we review various image source identification techniques. This is followed by an overview of techniques for differentiating synthetic images in Section 3. Image forgery (tamper) detection techniques are described in Section 4. Finally, our conclusions and open problems will be given in Section 5.

2. Image Source Identification

Image source identification research investigates the design of techniques to identify the characteristics of digital data acquisition device (*e.g.,* digital camera, camcorder, and scanner) used in generation of an image. These techniques are expected to achieve two major outcomes. The first is the class (model) properties of the source, and the second is the individual source properties. Essentially, the two outcomes refer two different operational settings. In determining the class properties, typically, a single image is available for evaluation and the source information is extracted through analyzing the image. In obtaining individual source properties, however, both an image and the potential source device or a number of images known to be acquired by the source is available for evaluation, and the analysis determines if the characteristics of the image in question matches to those of the source.

The success of image source identification techniques depend on the assumption that all images acquired by an image acquisition device will exhibit certain characteristics that are intrinsic to the acquisition devices because of their (proprietary) image formation pipeline and the unique hardware components they deploy regardless of the content of the image. (It should be noted that such devices generally encode the device related information, like model, type, date and time, and compression details, in the image header, *e.g.,* EXIF header. However, since this information can be easily modified or removed, it cannot be used for forensics purposes.) Due to prevalence of digital camera images, research has primarily focused on source digital camera identification and scanner identification research is just starting.

2.1. *Image Formation in Digital Cameras and Scanners*

The design of image source identification techniques requires an understanding of the physics and operation of these devices. The general structure and sequence of stages of image formation pipeline remains similar for almost all digital cameras and scanners, although much of the details are kept as proprietary information of each manufacturer. Below, we will describe the basic structure for a digital camera and scanner pipeline.

Digital Camera Pipeline: Consumer level digital cameras consist of a lens system, sampling filters, color filter array, imaging sensor, and a digital image processor.[3] The lens system is essentially composed of a lens and the mechanisms to control exposure, focusing, and image stabilization to collect and control the light from the scene. After the light enters the camera through the lens, it goes through a combination of filters that includes at least the infra-red and anti-aliasing filters to ensure maximum visible quality. The light is then focused onto imaging sensor, an array of rows of columns of light-sensing elements called pixels. Digital cameras deploy charge-coupled device (CCD) or complimentary metal-oxide semiconductor (CMOS) type of imaging sensors. Each light sensing element of sensor array integrates the incident light over the whole spectrum and obtains an electric signal representation of the scenery. Since each imaging sensor element is essentially monochromatic, capturing color images requires separate sensors for each color component. However, due to cost considerations, in most digital cameras, only a single sensor is used along with a color filter array (CFA). The CFA arranges pixels in a pattern so that each element has a different spectral filter. Hence, each element only senses one band of wavelength, and the raw image collected from the imaging sensor is a mosaic of different colors and varying intensity values. The CFA patterns are most generally comprised of red-green-blue (RGB) and cyan-magenta-yellow (CMY) color components. The measured color values are passed to a digital image processor which performs a number of operations to produce a visually pleasing image. As each sub-partition of pixels only provide information about a number of color component values, the missing color values for each pixel need to be obtained through demosaicing operation. This is followed by other forms of processing like white point correction, image sharpening, aperture correction, gamma correction and compression. Although the operations and stages explained here are standard stages in a digital camera pipeline, the exact processing detail in each stage varies from one manufacturer to the other, and even in different camera models manufactured by the same company.

Scanner Pipeline: Conventional consumer scanners are composed of a glass pane, a bright light source (often xenon or cold cathode fluorescent) which illuminates the pane from underneath, and a moving scan head that includes lenses, mirrors, a set of filters, and the imaging sensor, whether CCD, CMOS or contact image sensors (CIS).[4] (Drum scanners which have been typically used for high-end applications use photomultiplier tubes.) To obtain color scans, typically, three rows (arrays) of sensors with red, green, and blue filters are utilized. During scanning, the imaging sensor and light source move across the pane (linear motion). The light strikes the image, reflects, and is then reflected by a series of mirrors to the scanner lens. The light passes through the lens and is focused onto imaging sensors to be later digitized. The resolution of a scanner depends on both the number of elements of the imaging sensor (horizontal resolution) and the step size of the scan head motor (vertical resolution). The hardware resolution of the scanner can be reduced down by either down-sampling or less commonly through activating only some elements

of the CCD array.

2.2. *Source Model Identification*

The work in this field has been primarily focused on digital cameras. The features that are used to differentiate camera-models are derived based on the differences in processing techniques and the component technologies. For example, the optical distortions due to a type of lens, the size of the imaging sensor, the choice of CFA and the corresponding demosaicing algorithm, and color processing algorithms can be detected and quantitatively characterized by analysis of the image. The deficiency of this methodology, in general, is that many models and brands use components by a few manufacturers, and processing steps/algorithms remain same or very similar among different models of a brand. Hence, reliable identification of a source camera-model depends on characterization of various model dependent features as briefly explained below.

2.2.1. *Image Features*

Inspired by the success of universal steganalysis techniques, Kharrazi et al.[5] proposed a similar approach to identify source camera-model. In essence, a select number of features designed to detect post-processing are incorporated with new features to fingerprint camera-models. The 34 features include color features (*e.g.,* deviations from gray world assumption, inter-band correlations, gamma factor estimates), image quality metrics, and wavelet coefficient statistics. These features are then used to construct multi-class classifiers. The results obtained on moderate to low compressed images taken by 4 different camera-models yielded an identification accuracy of 97%. When experiments are repeated on five cameras where three of them are of the same brand, the accuracy is measured to be 88%. Tsai et al.[6] later repeated this study using a different set of cameras and reported similar results. In their work,[7] Celiktutan et al. took a similar approach to differentiate between cell-phone camera-models by deploying binary similarity measures as features.[8] In this case, the identification accuracy among nine cell-phone models (of four different brands) is determined as 83%. There are two main concerns regarding this type of approaches. First is that as they provide an overall decision, it is not clear as to what specific feature enables identification which is very important in forensic investigations and in expert witness testimonies. Second concern is the scalability of performance with the increasing number of digital cameras in the presence of hundreds of digital cameras. Hence, in general, this approach is more suitable as a pre-processing technique to cluster images taken by cameras with similar components and processing algorithms.

2.2.2. *CFA and Demosaicing Artifacts*

The choice of CFA and the specifics of the demosacing algorithm are some of the most pronounced differences among different digital camera-models. In digital cameras with single imaging sensors, the use of demosacing algorithms is crucial for correct rendering of high spatial frequency image details, and it uniquely impacts the edge and color quality of an image. Essentially, demosaicing is a form of interpolation which in effect introduces a specific type of inter-dependency (correlations) between color values of image pixels. The specific form of these dependencies can be extracted from the images to fingerprint different demosaicing algorithms and to determine the source camera-model of an image. In,[9] Popescu et al. demonstrated that expectation/maximization (EM) algorithm can be used to estimate the (linear interpolation) filter coefficients by re-interpolating digital camera images (after down-sampling to remove existing traces of interpolation) with eight different CFA interpolation algorithms. The average accuracy in pair-wise differentiation over all pairs of interpolation algorithms is obtained as 97%. To fingerprint demosaicing algorithms used in different digital camera-models Bayram et al.[10,11] deployed EM algorithm, assuming a linear model for interpolation within a 5x5 window, and analyzed patterns of periodicity in second order derivates of rows and columns of pixels in moderately smooth and very smooth image parts, respectively. The estimated filter coefficients and the periodicity features are used as features in construction of classifiers to detect source camera-model. The accuracy in identifying the source of an image among four and five camera-models is measured as 86% and 78%, respectively, using images captured under automatic settings and at highest compression quality levels.

Alternatively, Long et al.[12] considered analyzing the modeling error due to the linear interpolation model and identifying demosaicing algorithm based on the characteristics of this error, rather than using the estimated interpolation filter coefficients. (*I.e.,* the difference between the actual pixel values in the image and their reconstructed versions as a weighted sum of 13 neighboring pixels.) They realized this by computing the autocorrelation of the error over all image. Then, the (13x13) autocorrelation matrices obtained from many images are combined together and subjected to principal component analysis to determine the most important components which are then used as features in building a classifier. They reported that an accuracy of more than 95% can be achieved in identifying the source of an image among four camera-models and a class of synthetic images and studied the change in performance under compression, noise addition, gamma correction and median filtering types of processing.

Later, Swaminathan et al.[13] enhanced this approach by first assuming a CFA pattern, thereby discriminating between the interpolated and un-interpolated pixel locations and values—an advantage over EM algorithm, and estimating the interpolation filter coefficients corresponding to that pattern (assuming a linear model within a 7x7 window) for each of three activity regions, *e.g.,* smooth, horizontal gra-

dient, and vertical gradient. Then, the un-interpolated color values are interpolated with respect to the assumed CFA pattern with the obtained filter and the error with the resulting newly interpolated image and the actual image is computed. The CFA pattern of an image is determined by searching over all valid CFA patterns to minimize the resulting error, and the demosaicing algorithms are differentiated through the use of classifiers built based on estimated filter coefficients. The corresponding identification accuracy is determined by applying the method to images taken by 16 camera-models under different compression levels and it is reported to be 84%.

2.2.3. *Lens Distortions*

In their work,[14] Choi et al. proposed the utilization of lens radial distortion, which deforms the whole image by causing straight lines in object space to be rendered as curved lines. Radial distortion is due to the change in the image magnification with increasing distance from the optical axis, and it is more explicit in digital cameras equipped with spherical surfaced lenses. Therefore, manufacturers try to compensate for this by adjusting various parameters during image formation which yields unique artifacts. To quantify these distortions, the paper extends a first-order radial symmetric distortion model, which expresses undistorted radius (from optical axis) as an infinite series of distorted radius, to second order. These parameters are computed assuming a straight line model by first identifying line segments which are supposed to be straight in the scene and computing the error between the actual line segments and their ideal straight forms. Later, these parameters are used as features to build classifiers in a framework similar to.[5] The measurements obtained from images captured with no manual zooming and flash and at best compression level by three digital camera-models resulted with an identification accuracy of approximately 91%. These features are also incorporated with those of earlier proposed ones[5] and similar overall identification accuracy is reported.

2.3. *Individual Source Identification*

The ability to match an image to its source requires identifying unique characteristics of the source acquisition device. These characteristics may be in the form of hardware and component imperfections, defects, or faults which might arise due to inhomogeneity in the manufacturing process, manufacturing tolerances, environmental effects, and operating conditions. For example, the aberrations produced by a lens, noise in an imaging sensor, dust specks on a lens will introduce unique but mostly imperceptible artifacts in images which can later be extracted to identify the source of the image. The main challenge in this research direction is that reliable measurement of these minute differences from a single image is very difficult and they can be easily eclipsed by the image content itself. Another challenge is that these artifacts tend to vary in time and depend on operating conditions. Therefore, they may not always yield positive identification. Following approaches are proposed

to utilize such characteristics in image source identification.

2.3.1. *Imaging Sensor Imperfections*

This class of approaches to source matching aims at identifying and extracting systematic errors due to imaging sensor, which reveal themselves on all images acquired by the sensor in a way independent of the scene content. These errors include sensor's pixel defects and pattern noise which has two major components, namely, fixed pattern noise and photo response non-uniformity noise. The initial work in the field has been done by Kurusowa et al.[15] in which fixed pattern noise caused by dark currents in (video camera) imaging sensors is detected. Dark current noise refers to differences in pixels when the sensor is not exposed to light and it essentially behaves as an additive noise. Therefore, it can be easily compensated within the camera by first capturing a dark frame and subtracting it from the actual readings from the scene, thereby hindering the applicability of the approach. Geradts et al.[16] proposed matching the traces of defective pixels, *e.g.,* hot pixels, cold/dead pixels, pixel traps, cluster defects, for determining the source camera. Their experiments on 12 cameras showed the uniqueness of the defect pattern and also demonstrated the variability of the pattern with operating conditions. However, ultimately, such defects also cannot be reliably used in source identification as most cameras deploy mechanism to detect such defects and compensate them through post-processing.

The most promising and reliable approach in this field is proposed by Lukas et al.[17] to detect the pixel non-uniformity noise, which is the dominant component of the photo-response non-uniformity pattern noise arising due to different sensitivity of pixels to light. The main distinction of this approach as compared to earlier ones is that the correction of this noise component requires an operation called flat-fielding which in essence requires division of the sensor readings by a pattern extracted from a uniformly lit scene before any non-linear operation is performed. Since obtaining a uniform sensor illumination in camera is not trivial, most digital cameras do not flat-field the resulting images. The key idea of the method is to denoise the image by wavelet based denoising algorithm so that the resulting noise residue contains the needed noise components. However, since the underlying image model used in denoising is an idealistic one the residue signal also contains contributions from the actual image signal. Hence to eliminate the random component of the noise, denoising is applied to a set of images (captured by the same camera) and the corresponding noise residues are averaged to obtain the reference pattern of a given digital camera. Later, to determine whether a given image is captured by a digital camera, the noise pattern extracted from the individual image is correlated with the reference pattern of the digital camera. A decision is made by comparing the measured correlation statistic to a pre-determined decision threshold. The results obtained from (high quality) images taken by 9 cameras yielded 100% identification accuracy.

To determine the false-positive and true-detection performance of the scheme proposed in[17] under a more realistic setting Sutcu et al.[18] performed experiments on an image dataset with roughly 50K randomly selected images and observed that some of the tested cameras yield false-positive rates much higher than the expected values. To compensate for false-positives the authors proposed coupling the approach of[17] with source-model identification methodology. In this case, during the extraction of the pattern the demosaicing characteristics of the source camera-model are also determined as described in.[11] When a decision is to be made in matching an image to a potential source camera, it is also required that the class properties of the camera extracted from the individual image is also in agreement with those of the source camera. It is shown that this approach is very effective in reducing the false-positive rate with a marginal reduction in the true-detection rate. In,[19] Fridrich et al. proposed enhancements to the noise extraction scheme by deploying pre-processing techniques to reduce the contributions of image noise and to gain robustness against compression.

Khanna et al.[20,21] extended sensor noise extraction methodology to also include scanned images and to enable source scanner identification. The main difference between the imaging sensors deployed in digital camera and (flatbed) scanners is that in the former sensor is a two-dimensional array, whereas in the latter it is a one-dimensional linear array, and a scan is generated by translating the sensor over the image. As a result the noise pattern extracted from a scanned image is expected to repeat itself over all rows. Therefore, a row reference noise pattern can be obtained from a single scanned image by averaging the extracted noise (via denoising) over all rows. In [20], the authors showed that this difference in the dimension of the array can be used to distinguish between digital camera and scanner images. In realizing this, classifiers are built based on (seven) statistics computed from averaged row and column reference patterns extracted from both scanned images at hardware resolution (*e.g.,* no down-sampling) and digital camera images. In experiments, various training scenarios are considered and an average accuracy of more than 95% is achieved in discriminating digital camera images from scanned images. The methodology is also applied to source scanner identification problem with the inclusion of new features in classifier design.[21] When identifying the source scanner of an image among four scanners an average classification accuracy of 96% is achieved and when the images are compressed with JPEG quality factor 90 an accuracy of 85% is obtained.

Gou et al.[22] proposed another approach to fingerprint the scanning noise associated with different models of (flatbed) scanners. The method characterizes scanning noise by three sets of features. The first set of features are obtained by denoising the scanned images and obtaining first and second order moments of the log-absolute transformed version of the noise residue. The second set of features are obtained as the mean, variance and error due to fitting normal distributions to high frequency sub-band coefficients of one-level wavelet decomposed version of the (normalized) scanned image. The third set consists of features extracted from the first two mo-

ments of prediction error applied to smooth regions. The most distinctive of the resulting 60 features are used to construct classifiers, which yielded an identification accuracy of 90% among seven scanner models with relatively smaller size of datasets (27 uncompressed images per model). The distinguishability of the features are also compared to wavelet coefficient statistics[23] and image quality metrics[24] and shown to be better. In the context of scanner identification, one issue that needs to be further studied is the variability of scanner noise among individual scanners and determining the corresponding false-alarm rates in identifying the source scanner.

2.3.2. *Sensor Dust Characteristics*

Dirik et al.[25] proposed another method for source camera identification based on sensor dust characteristics of single digital single-lens reflex (DSLR) cameras which are becoming increasingly popular because of their interchangeable lenses. Essentially, the sensor dust problem emerges when the lens is removed and the sensor area is opened to the hazards of dust and moisture which are attracted to the imaging sensor due to electrostatic fields, causing a unique dust pattern before the surface of the sensor. Sensor dust problem is persistent and most generally the patterns are not visually very significant. Therefore, traces of dust specks can be used for two purposes: to differentiate images taken by cheaper consumer level cameras and DSLR cameras and to associate an image with a particular DSLR camera. However, it should be noted that the lack of a match between dust patterns does not indicate anything since the dust specks might have been cleaned. Devising an empirical dust model characterized by intensity loss and roundness properties; the authors proposed a technique to detect noise specks on images through match filtering and contour analysis. This information is used in generation of a camera dust reference pattern which is later checked in individual images. In the experiments, ten images obtained from three DSLR cameras are used in generating a reference pattern which is then tested on a mixed set of 80 images (20 taken with the same camera and 60 with other cameras) yielding an average accuracy of 92% in matching the source with no false-positives.

3. Identification of Synthetic Images

A great deal of progress has been made in both fields of computer vision and computer graphics and these two fields have now begun to converge very rapidly. Consequently, more realistic synthetic imagery became achievable. Today, generative algorithms are able to produce realistic models of natural phenomena, *e.g.,* waves, mountains, sky, plants, objects with geometric structure, stimulate the behavior of light, *e.g.,* ray tracing and subsurface scattering methods, and take into consideration sensitivities of human perceptual system. Moreover, the sophistication of these algorithms parallels the increasing computation power. These advances in a way defeat the whole purpose of imagery and put the credibility of digital imagery at stake. Therefore, distinguishing photo-realistic computer generated (PRCG) images

from real (natural) images is a very challenging and immediate problem. Several approaches have been proposed to address this problem. Essentially, all proposed approaches are based on machine learning methods, which express the relations between features extracted from a sample set of PRCG and real images in the form of classifiers. These classifiers are later used to differentiate between the two types of images. Hence, the main difference between the proposed approaches lie in the features they use in constructing the classifiers. Another concern with this class of methods is the image sets used during training and test phases as the true performance of the method will depend on how well their characteristics represent the overall class of images they belong to.

The first approach to differentiating natural (photographic) images from PRCG images was proposed by Lyu et al.[26] based primarily on a model of natural images. In this technique, the features are designed to capture the statistical regularities of natural images in terms of statistics of three-level discrete wavelet transform coefficients. The features include first order statistics (*e.g.,* mean, variance, skewness, and kurtosis) of both sub-band coefficients at each orientation and scale and of the errors in a linear predictor of coefficient magnitude (of all spatial, scale, orientation, and color neighbors) to capture higher order statistical correlations across space, scales and orientations, resulting with 72 features in each color band. The experiments are done on 40K real and 6K PRCG images of which 32K+4.8K images were used for training the classifiers and the rest for the testing which yielded an identification accuracy of 67% at 1% false-alarm rate.

In their work,[27] Ng et al. proposed another promising approach based on identifying the distinctive (geometry-based) characteristics of PRCG images, as compared to natural images. Their technique takes into account the differences in surface and object models and differences in the acquisition process between the PRCG and real images. The selection of their features is motivated by the observations that generation of PRCG images, mostly due to issues of computational complexity, is based on polygonal surface models and simplified light transport models, and does not exhibit acquisition characteristics of hardware device *e.g.,* cameras and scanners. The 192 features used in the design of the classifier are extracted by analyzing local patch statistics, local fractal dimension, and (normalized) differential geometry quantities, *e.g.,* surface gradient, quadratic geometry, and Beltrami flow. The authors used 800 PRCG images and 1.6K real images to test their features and obtained an average identification accuracy of 83% in comparison to an accuracy of 80% by Lyu et al.'s features.[26] It is also shown that when classifiers are trained to identify the CG images that are captured by digital cameras (*i.e.,* recapturing attack), a similar performance can be achieved by both feature sets.

Another wavelet transform based method was proposed by Wang et al.[28] where features are obtained from characteristic functions of wavelet-coefficient histograms. The features are obtained by first applying three-level wavelet decomposition at each color channel and further decomposing the diagonal sub-band into four second-level sub-bands, yielding a total of 48 sub-bands, and then by obtaining the normalized

histograms in each sub-band. The DFT transform of the normalized histograms are filtered by three filters (two high-pass filters and a band-pass filter) to determine their energy at different frequency component ranges. Hence, a total of 144 features are obtained. The classifier trained on half of the 4.5K natural and 3.8K PRCG images yielded detection and false-positive results comparable to those of.[26] However, it is reported the classifier did not perform uniformly (much higher false-alarm rate) on the dataset used by in.[27]

Motivated by the fact that majority of the real images are captured by digital cameras, Dehnie et al.[29] presented an approach that aims at discriminating synthetic images from digital camera images based on the lack of artifacts due to acquisition process by focusing on the imaging sensor's pattern noise. Although each digital camera has a unique noise pattern,[17] since the underlying sensor technology remains similar, it is very likely that pattern noise introduced by different digital cameras may have common statistical properties. On the other hand, to avoid lack of real-life details, such as textures and lighting, generation of PRCG requires methods that add noise to simulate such phenomena in a physically consistent manner, *e.g.,* ray tracing algorithms. Similarly, it is very likely that the noise introduced by these methods to have certain statistical properties. To test the discriminative ability of the approach, a 600 PRCG images and more than 600 digital camera images have been denoised and the statistics of the resulting noise residues are analyzed. It is shown that the first-order statistics, like skewness and kurtosis, for the two noise components are distinct and the two types of image can be discriminated with an average accuracy of 75%.

Later, Dirik et al.[30] extended this approach to also include demosaicing artifacts[11] by proposing new features to detect the use of Bayer color filter array during demosaicing and and to detect traces of chromatic aberration. These features are later incorporated with the features of[26] and tested on 1.8 K PRCG and digital camera images half of which were used for training. Test results obtained on high quality images show that the classifier designed based on only four demosaicing features perform as good as wavelet transform coefficient statistics based features alone.[26] The results obtained from both high quality and medium level compressed images show that On the other hand, the proposed single feature based on traces chromatic aberration is shown to perform slightly worse but with less sensitivity to compression in highh to medium compression levels. The results for combined features show that the proposed five features can further improve the performance of the existing methods.

4. Image Forgery Detection

Due to the ease with which digital images can be altered and manipulated using widely available software tools, forgery detection is a primary goal in image forensics. An image can be tampered in many ways and at varying degrees, like compositing, re-touching, enhancing, with various intents. Although, many of the

tampering operations generate images with no visual artifacts, they will, nevertheless affect the inherent statistics of the image. Furthermore, the process of image manipulation very often involves a sequence of processing steps to produce visually consistent images. Typically, a forged image (or parts of it) would have undergone some common image processing operations like affine transformations (*e.g.,* scaling, rotation, shearing), compensation for color and brightness variations, and suppression of details (*e.g.,* filtering, compression, noise addition). As a result, it is very likely that tampered image statistics will also exhibit variations due to such operations. In what follows, we briefly review various techniques proposed to determine whether the image has undergone any form of modification or processing after it has been captured.

4.1. *Variations in Image Features*

These approaches designate a set of features that are sensitive to image tampering and determine the ground truth for these features by analysis of original (unaltered) and tampered images. These values are stored as reference values and later tampering in an image is decided based on deviation of the measured features from the ground truth. These approaches most generally rely on classifiers in making decisions. For example, to exploit the similarity between the steganalysis and image manipulation detection, Avcibas et al.[31] proposed an approach similar to[24] by utilizing image quality metrics to probe different quality aspects of images, which could be impacted during tampering. In,[24] image quality metrics are used in cooperation with classifiers to differentiate between original and altered images based on measures obtained between a supposedly modified image and its estimated original (obtained through denoising) in terms of pixel and block level differences, edge distortions, and spectral phase distortions. To ensure that the features respond only to induced distortions due to tampering and not be confused by the variations in the image content, in[31] metrics are also measured with respect to a fixed set of images. Results obtained on 200 images by subjecting them to various image processing operations at a global scale yielded an average accuracy of 80%. When the same classifiers are given 60 skillfully tampered images, the detection accuracy is obtained to be 74%.

Based on the observation that non-linear processing of a signal very often introduced higher-order correlations, Ng. et al.[32] studied the effects image splicing on magnitude and phase characteristics of bicoherence spectrum (*i.e.,* normalized bispectrum which is the Fourier transform of the third order moment of a signal). The authors modeled the discontinuity introduced at the splicing point as a perturbation of a smooth signal with a bipolar signal and showed that bipolar signals contribute to changes in bicoherence spectrum of a signal. When tested the magnitude and phase features provided a classification accuracy of 62% which can be attributed to strong higher order correlations exhibited by natural images. Later,[33] the authors augmented the existing bicoherence features with newer ones that take

into consideration the sensitivity of bicoherence to edge pixel density and the variation in bicoherece features between the spliced and un-spliced parts of the image. With the inclusion of the new features the accuracy in splicing detection is reported to increase to 72%. Bayram et al.[34,35] compiled three fundamental sets of features that have been successfully used in universal steganalysis and rigorously tested their sensitivity in detecting various common image processing operations by constructing classifiers to identify images that have undergone such processing. The tested features include image quality metrics,[24] wavelet coefficient statistics,[23] binary similarity measures,[8] the joint feature set which combines all the three sets, and the core feature set which is a reduced version of joint feature set. Different types of classifiers built from these features are tested under various image manipulations, like scaling up/down, rotation, contrast enhancement, brightness adjustment, blurring/sharpening and combinations, with varying parameters. Results on 100 locally tampered images, obtained from Internet, show that joint feature set performs best with an identification accuracy of around 90%.

4.2. *Image Feature Inconsistencies*

This class of techniques tries to detect image tampering based on inconsistent variations of selected features across the image. These variations may be in the form of abrupt deviations from the image norm or unexpected similarities over the image. One of the earliest methods in this class exploits the presence of double JPEG compression artifacts. Recompression of an (already compressed) image at a different quality factor distorts the smoothness of DCT coefficient histograms and creates identifiable patterns in DCT coefficient histograms. When the second quantization step size is smaller, some bins in the resulting histogram will be empty (zero valued) yielding a periodic peaks-and-valleys pattern. On the other hand, if the second quantization step is larger than the first one, all histogram values will be present but due to uneven splitting and merging of bins, histogram will show periodic peak patterns.

This phenomenon has been observed and studied in[36] and[37] to determine the initial compression parameters and to detect double compressed images. Essentially, the most common form of image tampering involves splicing of images which are very likely to be compressed at different quality factors. Therefore, the spliced parts in the recompressed image will have different double compression characteristics as compared to other parts. He et al.[38] developed a workable algorithm for automatically locating the tampered parts. In the method, the coefficient histogram of each DCT channel is analyzed for double compression effects and to assign probabilities to each (8x8) DCT block of its being a doctored block. The probabilities for each block are later fused together to obtain normality map of blocks, and tampering is decided based on presence and location of clusters on this map. Experiments performed on a small number of tampered images demonstrate the success of the algorithm. Further experiments are needed to determine how the method performs

under various types of image tampering.

Popescu et al.[39] proposed a method for detecting resized (parts of) images which might potentially indicate image tampering. The principle of their method is based on the fact that up-sampling (interpolation) operation introduces periodic inter-coefficient correlations (*i.e.,* all interpolated coefficients depend on their neighbors in the same manner) and re-sampling at arbitrary rates requires a combination of up-sampling and down-sampling operations to achieve the intended rate. Hence, the presence of correlation between pixels can be used to determine which parts of images underwent resizing. To extract the specific form of correlations, the authors assumed probabilistic models for the prediction errors of both interpolated and un-interpolated coefficients. The estimation of distribution parameters and grouping of coefficients are performed simultaneously by EM algorithm. Results obtained on high quality JPEG images by subjecting images to global transformations such as scaling, rotations and gamma correction yielded detection accuracy close to 100% in most cases. However, the accuracy of detecting locally tampered regions have to be further tested.

Johnson et al.[40] considered the use of lighting direction inconsistencies across an image to detect image tampering, as it is often difficult to ensure (physically) consistent lighting effects. The crux of the method lies in a technique that estimates the light source direction from a single image. Assuming a point light source infinitely far away, a surface that reflects light isotropically and has a constant reflectance, and the angle between the surface normal and the light direction is less than 90 degrees, the image intensity is expressed a function of surface normal, light source direction, and constants (*i.e.,* reflectance and ambient light terms). Surfaces of known geometry in the image (*e.g.,* plane, sphere, cylinder, etc.) in the image are partitioned into many patches and by solving the formulation for all patches and combining the results to obtain the light direction. Formulation is also applied to local light sources and multiple light sources by combining them into a single virtual light source. The applicability of the method is demonstrated on a smaller set of images.

Image tampering very often involves local sharpness/blurriness adjustments. Hence, the blurriness characteristics in the tampered parts are expected to differ in non-tampered parts. In,[41] Sutcu et al. proposed the use of regularity properties of wavelet transform coefficients to estimate sharpness/blurriness of edges to detect variations and to localize tampering. The decay of wavelet transform coefficients across scales has been employed for edge detection and quality estimation purposes previously. The proposed method first employs an edge detection algorithm to determine edge locations which is then followed by a multi-scale wavelet decomposition of the image. Edge locations are located by analyzing the edge image and corresponding maximum amplitude values of wavelet sub-band signals are determined. Then, a linear curve is fitted to the log of these maximum amplitude values and the goodness of the fit is used an indicator of sharpness/blurriness value. The potential of the method in detecting variations in sharpness/blurriness is demonstrated on

both globally blurred images and tampered images with local adjustments.

Another common form of forgery is content repetition which involves copying and pasting part(s) of an image over other parts of the same image to disguise some (contextual) details in the image. Although this type of tampering can be easily detected by exhaustive search and analysis of correlation properties of the image (autocorrelation function) due to introduced correlation by content repetition, these methods are not computationally practical and do not perform well when the copied pasted parts are smaller in region. To address this problem Fridrich et al.[42] proposed a better performing (faster and accurate) method. The method obtains DCT coefficients from a window that is slid over the whole image in an overlapping manner and quantizes them. The resulting coefficients are arranged and inserted into a row matrix. The rows (of quantized DCT coefficients) are then sorted in a lexicographical order and through row-wise comparisons similar blocks are determined. The main computational cost of this algorithm is due to sorting which requires significantly less time steps as compared to brute force search, *e.g.,* an O(nlogn) algorithm. To further improve the robustness of this method to possible variations Popesctu et al.[43] used an alternative representation of blocks based on principal component analysis to identify the similar blocks in the image. Similar to,[41] the coefficients in each block are vectorized and inserted in a matrix and the corresponding covariance matrix is computed. By finding the eigenvectors of the covariance matrix, a new linear basis for each image block is obtained and a new representation is obtained by projecting each image block onto selected basis vectors with higher eigenvalues to reduce dimensionality. Then, the representation of each block is lexicographically sorted and compared to determine the similar blocks. The robustness of the method in detecting tampered parts is demonstrated under ranging JPEG compression qualities and additive noise levels.

4.3. *Inconsistencies Concerning Acquisition Process*

As discussed earlier, image acquisition process introduces certain distinguishing characteristics in each acquired image which can be used for source identification. Since these characteristics will be fairly uniform over the whole image, their consistency across the image can also be used for detecting and localizing tampering. Hence, this group of techniques is extensions of source identification techniques with some minor differences. For example, Swaminathan et al.[44] used inconsistencies in color filter array interpolation to detect tampered parts of an image based on their approach in.[13] After estimating the CFA pattern and the interpolation filter, the demosaiced image is reconstructed and compared to the image itself. Modeling the linear part of the post-processing as a tampering filter, its coefficients are obtained by deconvolution. These coefficients are then used in design of a classifier to detect tampering by comparing the obtained filter coefficients with a reference pattern obtained from direct camera output (*i.e.,* unaltered images). Results obtained by subjecting test images to spatial averaging, rotation, compression and resampling

is reported to yield average detection accuracy of more than 90%.

Similarly, based on,[17] Lukas et al. proposed to detect and localize tampering by analyzing the inconsistencies in the sensor pattern noise extracted from an image.[45] The noise patterns obtained from various regions are correlated with the corresponding regions in the camera's reference pattern and a decision is made based on comparison of correlation results of region of interest (potentially tampered region) with those of other regions. Along the same line Popescu et al.[46] proposed to detect the presence of CFA interpolation, as described in,[9] in overlapping blocks of an image to detect tampering. Experiments were performed on a limited number of digital camera images to identify traces of CFA interpolation in each block with no tampering.

Johnson et al.[47] proposed a new approach by inspecting inconsistencies in lateral chromatic aberration as a sign of tampering. Lateral aberration is due to inability of the lens to perfectly focus light of all wavelengths onto imaging sensor, causing a misalignment between color channels that worsens with the distance from the optical center. The method treats the misalignment between color channels as an expansion (or contraction) of a color channel with respect to one another and tries to estimate the model parameters (*e.g.,* center and aberration constant) to attain alignment. The estimation of these model parameters is framed as an image registration problem and a mutual entropy metric is used to find the exact aberration constant which gives the highest mutual entropy between color channels. To detect tampering, image is partitioned into blocks and the aberration estimated in each block is compared to global estimate. Any block that deviates significantly from the global estimate is deemed to be tampered. The threshold deviation is determined experimentally under varying compression qualities; however, further experiments are needed to generalize the results and determine the dependency on image content.

Alternatively, Lin et al.[48] proposed a method to recover the response function of the camera by analyzing the edges in different patches of the image and verifying their consistency. Camera response function defines the relation between radiance values from the scene and measured brightness values in each color channel and due to this non-linear response a linear variation of the pixel irradiance at the edges will be distorted. The main idea of the method is to utilize this phenomenon by computing the inverse response function and to determine its conformance to known properties of response functions (which should be monotonically increasing with at most one inflexion point and similar to each other in each color channel). The normality of the estimated functions, from each patch, is decided by comparing them to a database of known camera response functions. For this, classifiers are designed by extracting features from the computed and available response functions and tested on a few example images to demonstrate the feasibility of the idea. Although the success of the method requires images to be of high contrast so that the color range in each patch is wide enough, this assumption can be relaxed by applying the method to source camera-model identification problem.

5. Conclusions and Outlook

There is a growing need for digital image forensics techniques, and many techniques have been proposed to address various aspects of digital image forensics problem. Although many of these techniques are very promising and innovative, they all have limitations and none of them by itself offers a definitive solution. Ultimately, these techniques have to be incorporated together to obtain reliable decisions. However, there are still two major challenges to be met by image forensics research.

- *Performance Evaluation and Benchmarking.* Essentially the foremost concern that arises with respect to forensic use of proposed techniques is the achievable performance in terms of false-alarm and true-detection/identification rates and clear understanding of the factors that affect the performance. From this point of view, many of the proposed techniques can be more accurately defined as proof of concept experiments. To further refine these methods, performance merits have to be defined more clearly and proper test and evaluation datasets have to be designed and shared.
- *Robustness Issues.* The most challenging issue that image forensics research faces is the robustness to various common and malicious image processing operations. Proposed methods are not designed and tested rigorously to perform under the most difficult conditions, and, moreover, most techniques can be easily circumvented by a novice manipulator. Since the information utilized by the image forensics techniques is mostly in imperceptible detail, it can be easily removed. It is a matter of time for such tools to be available for public use. Techniques have to be designed and evaluated with this caveat in mind.

Overcoming these challenges requires the development of several novel methodologies and thorough evaluation of their limitations under more general and practical settings. This can be achieved in collaboration with forensics experts and through their continuous feedback on the developed methods. The research effort in the field is progressing well in these directions.

References

1. The US Supreme Court Ruling in Ashcroft v. Free Speech Coalition, No. 00-795.
2. B. Goldfarb, Digital Deception (Online, http://larrysface.com/deception.shtml.
3. J. Adams, K. Parulski, and K. Spaulding, Color Processing in Digital Cameras, *IEEE Micro*, vol. 18, no. 6, pp. 20-31 (1998).
4. How Scanners Work. (Online, http://www.extremetech.com/article2/0,1697, 1157540,00.asp.
5. M. Kharrazi, H. T. Sencar, and N. Memon, Blind Source Camera Identification, *Proc. of IEEE ICIP* (2004).
6. M.-J. Tsai and G.-H. Wu, Using Image Features to Identify Camera Sources, *Proc. of IEEE ICASSP* (2006).
7. O. Celiktutan, I. Avcibas, B. Sankur and N. Memon, Source Cell-Phone Identification, *Proc. of ADCOM* (2005).

8. I. Avcibas, M. Kharrazi, N. Memon and B. Sankur, Image Steganalysis with Binary Similarity Measures, *EURASIP Journal on Applied Signal Processing*, vol. 17, pp. 2749-2757 (2005).
9. A. Popescu, Statistical Tools for Digital Image Forensics, Ph.D. Dissertation, Department of Computer Science, Darthmouth College (2005).
10. S. Bayram, H. T. Sencar and N. Memon, Source Camera Identification Based on CFA Interpolation, *Proc. of IEEE ICIP* (2005).
11. S. Bayram, H. T. Sencar and N. Memon, Improvements on Source Camera-Model Identification Based on CFA Interpolation, *Proc. of WG 11.9 Int. Conf. on Digital Forensics* (2006).
12. Y. Long and Y. Huang, Image Based Source Camera Identification Using Demosaicing, *Proc. of IEEE MMSP* (2006).
13. A. Swaminathan, M. Wu and K. J. Ray Liu, Non-Intrusive Forensics Analysis of Visual Sensors Using Output Images, *Proc. of IEEE ICIP* (2006).
14. K. S. Choi, E. Y. Lam and K. K. Y. Wong, Source Camera Identification Using Footprints from Lens Aberration, *Proc. of SPIE* (2006).
15. K. Kurosawa, K. Kuroki and N. Saitoh, CCD Fingerprint Method, *Proc. of IEEE ICIP* (1999).
16. Z. J. Geradts, J. Bijhold, M. Kieft, K. Kurusawa, K. Kuroki and N. Saitoh, Methods for Identification of Images Acquired with Digital Cameras, *Proc. of SPIE*, vol. 4232 (2001).
17. J. Lukas, J. Fridrich and M. Goljan, Digital Camera Identification from Sensor Pattern Noise, *IEEE Trans. Inf. Forensics and Security*, vol. 1, no. 2, pp. 205-214 (2006).
18. Y. Sutcu, S. Bayram, H. T. Sencar and N. Memon, Improvements on Sensor Noise Based Source Camera Identification, *Proc. of IEEE ICME* (2007).
19. M. Chen, J. Fridrich and M. Goljan, Digital Imaging Sensor Identification (Further Study), *Proc. of SPIE* (2007).
20. N. Khanna, A. K. Mikkilineni, G. T. -C. Chiu, J. P. Allebach and E. J. Delp, Forensic Classification of Imaging Sensor Types, *Proc. of SPIE* (2007).
21. N. Khanna, A. K. Mikkilineni, G. T. -C. Chiu, J. P. Allebach and E. J. Delp, Scanner Identification Using Sensor Pattern Noise, *Proc. of SPIE* (2007).
22. H. Gou, A. Swaminathan and M. Wu, Robust Scanner Identification Based on Noise Features, *Proc. of SPIE* (2007).
23. S. Lyu and H. Farid, Steganalysis Using Higher-Order Image Statistics, *IEEE Trans. Image Forensics and Security*, vol. 1, no. 1, pp. 111-119 (2006).
24. I. Avcibas, B. Sankur and N. Memon, Steganalysis of Watermarking and Steganography Techniques Using Image Quality Metrics, *IEEE Trans. Image Processing*, vol. 12. no. 2, pp. 221-229 (2003).
25. E. Dirik, H. T. Sencar and N. Memon, Source Camera Identification Based on Sensor Dust Characteristics, *Proc. of IEEE SAFE* (2007).
26. S. Lyu and H. Farid, How Realistic is Photorealistic?, *IEEE Trans. On Signal Processing*, vol. 53, no. 2, pp. 845-850 (2005).
27. T.-T Ng, S. -F. Chang, J. Hsu, L. Xie, M. -P. Tsui, Physics-Motivated Features for Distinguishing Photographic Images and Computer Graphics, *ACM Multimedia* (2005).
28. Y. Wang and P. Moulin, On Discrimination between Photorealistic and Photographic Images [corrected version], *Proc. of IEEE ICASSP* (2006).
29. S. Dehnie, H. T. Sencar and N. Memon, Identification of Computer Generated and Digital Camera Images for Digital Image Forensics, *Proc. of IEEE ICIP* (2006).
30. E. Dirik, S. Bayram, H. T. Sencar and N. Memon, New Features to Identify Computer Generated Images, *Proc of IEEE ICIP* (2007).

31. I. Avcibas, S. Bayram, N. Memon, B. Sankur and M. Ramkumar, A Classifier Design for Detecting Image Manipulations, *Proc. of IEEE ICIP* (2004).
32. T. Ng, S. -F. Chang and Q. Sun, Blind Detection of Photomontage Using Higher Order Statistics, *Proc. of ISCAS* (2004).
33. T. Ng and S. -F. Chang, A Model for Image Splicing, *Proc. of ICIP* (2004).
34. S. Bayram, I. Avcibas, B. Sankur and N. Memon, Image Manipulation Detection with Binary Similarity Measures, *Proc. of EUSIPCO* (2005).
35. S. Bayram, I. Avcibas, B. Sankur and N. Memon, Image Manipulation Detection, *Journal of Electronic Imaging*, vol. 15, no. 4 (2006).
36. J. Lukas and J. Fridrich, Estimation of Primary Quantization Matrix in Double Compressed JPEG Images, *Proc. of DFRWS* (2003).
37. A. C. Popescu and H. Farid, Statistical Tools for Digital Forensics, *Proc. of IHW* (2006).
38. J. He, Z. Lin, L. Wang and X. Tang, Detecting Doctored JPEG Images via DCT Coefficient Analysis, *Proc. of ECCV* (2006).
39. 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).
40. M. K. Johnson and H. Farid, Exposing Digital Forgeries by Detecting Inconsistencies in Lighting, *Proc. of ACM Multimedia Securty Workshop* (2005).
41. Y. Sutcu, B. Coskun, H. T. Sencar and N. Memon, Tamper Detection Based on Regularity of Wavelet Transform Coefficients, *Proc. of IEEE ICIP* (2007).
42. J. Fridrich, D. Soukal and J. Lukas, Detection of Copy-Move Forgery in Digital Images, *Proc. of DFRWS* (2003).
43. A. C. Popescu and H. Farid, Exposing Digital Forgeries by Detecting Duplicated Image Regions, Technical Report, TR2004-515, Dartmouth College, Computer Science.
44. A. Swaminathan, M. Wu and K. J. R. Liu, Image Tampering Identification Using Blind Deconvolution, *Proc. of IEEE ICIP*, (2006).
45. J. Lukas, J. Fridrich and M. Goljan, Detecting Digital Image Forgeries Using Sensor Pattern Noise, *Proc. of SPIE* (2006).
46. A. C. Popescu and H. Farid, Exposing Digital Forgeries in Color Filter Array Interpolated Images, *IEEE Trans. Signal Processing*, vol. 53, no. 10, pp. 3948-3959 (2005).
47. M. K. Johnson and H. Farid, Exposing Digital Forgeries through Chromatic Aberration, *Proc. of ACM Multimedia Security Workshop* (2006).
48. Z. Lin, R. Wang, X. Tang and H. -Y. Shum, Detecting Doctored Images Using Camera Response Normality and Consistency Analysis, *Proc. of CVPR* (2005).