Comparison and analysis of Passive-blind techniques for detection of digital image
Priyashi gogoi, Shikha Choudhary
Lovely Professional University, Punjab, India

Abstract
Digital image forensics is a branch of image processing that analyzes an image in order to establish its truthfulness and authenticity. Digital image forensic techniques include two approaches: active and passive approaches. The active approach requires the digital image to be embedded with certain information in the form of digital watermarking or signatures during its acquisition. This embedded information is later checked to establish authenticity of the image and is usually robust and imperceptible against most of the attacks like compression, rotation, cropping, resampling, filtering, addition of noise etc. But, a major drawback of active approach is that specially designed digital cameras are required that are equipped with a watermarking chip or a digital signature chip which authenticates every image the camera takes before storing it on its memory card. So, digital signatures and watermarking (active approach) as image forgery detection tools are not widely used.

INTRODUCTION
Forgeries have been carried out by men for a long time and is not new to humankind. In the past, this problem was restricted to only craftsmanship and literature. But with the advancement in digital image processing soft wares and various editing tools, a digital image can be easily manipulated. It has become very difficult for people to identify an image’s authenticity.

In today’s modern era digital images play an important role as information carriers because of their immediacy and that the image content is easily understandable. Digital image forgeries are rapidly increasing in the internet and in the mainstream media resulting in decrease in the credibility of digital images. This calls for the requirement of robust techniques to verify the authenticity and integrity of digital images, especially as digital images are used and provided as evidences in the court of law, as information in news items, as legal documents and as financial and medical records.

In order to overcome this problem, a new method has been developed for image authentication that does not require digital images to be pre-embedded with any information during its creation. These techniques, referred to as passive approach are carried out based on the assumption that although digital image forgeries may or may not leave any visual traces indicating tampering, they may change the underlying statistics or parameters of an image. So the working of this method is
purely based upon analyzing the binary information of the digital image with no requirement of any extra information. This report focuses on the passive or blind techniques of digital image forgery detection, especially the pixel based techniques and format based techniques.

Fig 1- classification of digital image forensic techniques.

**DIGITAL IMAGE: IT’S FORMATION**

The life cycle of a digital image can be summed up into three phases: acquisition, coding and editing. During acquisition phase, the light coming from the real scene which is framed by the digital camera is focused onto the image sensors. The diaphragm controls the amount of light, the shutter speed is responsible for the time of exposure and the lens assembly then focuses the light rays on to the sensors to form a coherent image. The image sensor is the main component of a digital camera and most of them use either CCD (charge coupled device) sensor or CMOS (complementary metal oxide semiconductor) sensor. Each sensor is composed of an array of light sensitive diodes called photosites that convert photons falling on it into electrical charge that is proportional to the intensity of light. Each photosite captures the data for a single pixel or picture element in the image. This will produce greyscale images because the sensors cannot distinguish between colors. Normally, the colors present in an image are represented as a mixture or combination of varying percentages of the three primary colors- red, green, and blue. So, a thin film of primary color mosaic known as CFA (color filter array) is used for acquiring color information. This filter is laid over the image sensors and allows only one of the three primary
colors to pass through and blocks the other two colors for a single pixel. So, the brightness of one color per pixel is recorded. The arrangement of the CFA mosaic pattern depends on the manufacturer but Bayer’s mosaic filter is mostly preferred. Generally CFA contains double the number of green filters as opposed to red and blue filters because the human eye is more sensitive to green color. After a sensor records the brightness of a particular color, the color information of the neighboring pixels is used for interpolation of the other two colors not recorded by the sensor. Now these two interpolated colors are combined with the color directly measured by the site and the full color of each individual pixel is calculated. This process is known as interpolation or demosaicing. So, the camera converts the light received into proportional electric charge with each pixel corresponding to a color value.

During coding the signal generated is stored to the camera’s memory and in order to save storage the image is lossy compressed and JPEG is the most common and preferred format in commercial devices.

Finally, the image may undergo post-processing in order to enhance or modify its content. There are different types of image editing that can be applied to an image like geometric transformations which include rotation, scaling, etc., blurring, sharpening, image splicing, cloning and contrast enhancement.

Fig 1- General pathway of image acquisition
According to the different steps in the lifecycle of an image, the fingerprints developed can be of three types:

1. **Acquisition Fingerprints**: When an input passes through each of the components in a digital acquisition device, it gets modified introducing characteristic fingerprints in the final image output which is due to the specific optical system, imaging sensor, and camera software. The pathway of image acquisition is common for most of the commercially available devices. However, each step is performed according to specific manufacturer’s needs and choice because of which the traces depend on the brand and model of the camera. So, specific imperfections and irregularities are produced at every stage of the camera, leaving traces in the final image that represent the camera type. In addition, various inconsistencies are present in these artifacts that act as evidences for image forgery.

2. **Coding Fingerprints**: Lossy compression scheme of images leaves behind characteristic footprints that are related to the specific coding architecture. Since JPEG is the most widely used compression method, most of the studies and proposals are based on the processing history of JPEG-compressed images because each compression of an image introduces a different fingerprint.
3. **Editing Fingerprints** - Each and every post-processing operation applied to the digital image modifies its properties that leave specific traces accordingly to the type of operation used.

**PASSIVE/BLIND IMAGE FORGERY DETECTION TECHNIQUES**

Passive or blind image detection techniques can be classified into five categories\(^2\). These are-

i. **Pixel-based techniques** - These techniques emphasize on the pixels, which are the underlying building blocks of a digital image. There are four types of pixel-based techniques which analyzes correlations at the pixel level arising from certain types of tampering (fig. 3).

![Fig 3- Categories of pixel based techniques](image)

ii. **Format-based techniques** - These techniques are based on the format of the image, mainly the JPEG format. This detection technique is based upon certain statistical correlations introduced by specific lossy compression schemes. These can be divided into three categories (fig 4).

![Image diagram](image)
iii. **Camera based techniques**- An image after being captured by a digital camera moves from the camera sensor to the memory storage and then it undergoes a series of processing steps like quantization, color correlation, white balancing, gamma correction, filtering, and JPEG compression. Depending upon the camera model and its type, these processing steps starting from capturing to saving the image in the memory varies. These camera based techniques work on the principle that certain artefacts or footprints are introduced at various stages of image processing and can be divided into four categories (fig 5).

iv. **Physics based techniques**- These techniques are based on 3-dimensional interactions between the camera, the physical object and light. Lighting is a very important component for capturing an image and any differences in lighting across an image can indicate manipulation or tampering. These are further divided into three categories (fig 6).
v. **Geometry based techniques** - Specific artifacts are introduced by the various image processing steps and several image forensic techniques have been proposed that specifically model these artefacts. Geometry-based techniques work by taking measurement of objects in the world and their position relative to the camera.

1. **PIXEL-BASED TECHNIQUES**

A pixel is the smallest component of a digital image. These pixels are tiny and are packed tightly which produces perfectly smooth and sharp images. The pixels of an image can be seen
as rectangular components when the image is zoomed in sufficiently and each of these pixels are made up of a single color component. Pixels don’t have a fixed size rather their size is relative to the resolution of the screen\(^9\).

Normally pixels in a digital image are arranged in a rectangular array. The dimensions of this array determines the size of the image. The width of the image depicts the number of columns and the image height depicts the number of rows in the array. So, the image pixel matrix contains \(M\) columns and \(N\) rows. The pixel count of an image is identified by the no. of columns and rows. For example, if an image has 2048 columns and 1536 rows of pixels in an array, then the image contains 3,145,728 pixels or 3.1 megapixels (2048X1536). Also, with the increase of megapixels in the capturing device, the maximum possible size of an image produced also increases i.e. a camera having 5 megapixel will be able to capture a larger image than a 3 megapixel camera\(^{10}\).

The size of an image specifically refers to the number of pixels in a digital image. The representation of an image in an output or displaying device depends on the resolution of the device. Resolution is expressed in terms of ppi (pixels per inch). For example, if an image is of 4500X3000 pixels and set at 300 ppi, then the real world size of the image will be 15X10 inches. However, at 72 ppi the image will be 62.5X41.6 inches. The greater the value of ppi, the sharper and more defined the image will be\(^9\).

Each pixel has a variable intensity or brightness value. Otherwise, the image will be of a uniform shade. Black and white images vary in intensity from black to white. While color images vary in intensity from the darkest and lightest of the three different primary colors - red, blue and green. A mixture of these color in varying intensity produces a color image. Black and white images are known as grayscale image while colored images are known as RGB images.

Along with the intensity type of each pixel, its intensity value also varies. These intensity values of a digital image are defined by bits. A single bit has two values- 0 or 1 while with an 8-bit intensity range there are 256 possible values- 0 to 255. These values grow exponentially and mathematically can be expressed as \(2^n\) where \(n\) = no. of bits. So, for 1 bit image there are \(2^1=2\) possible values and for 8 bit image there are \(2^8= 256\) possible values. Usually 8-bit intensity range is used in standard digital photos. RGB images or colored images uses 8-bit intensity range for each of the three color and are therefore referred to as 24-bit color images(3X8-bit). Black and white images or grayscale images have a single 8-bit intensity range\(^9,^{10}\). 
The four types of pixel based techniques used for digital image forgery detection are as follows:

1.1 **COPY-MOVE OR CLONING**

Copy-move or cloning is the most common type of image forgery. In this, a portion of the image is copied or cloned and then pasted in some other location within the image. This is usually done to conceal or hide an object or a person in an image\(^\text{11}\). It is very difficult to detect cloning visually. Also, since the cloned portion belongs to the same image, some of the components like color and noise will be similar or compatible with the rest of the image due to which this type of forgery is not detectable using normal forensic methods and tools. Another complication lies in the fact that the cloned region can be of any size and shape and in any location which makes it computationally impossible to search all the possible locations of the image and sizes. This calls for the requirement of properly designed methods in order to cope with is type of forgery.
Two robust and computationally efficient algorithms have been developed for detection of cloned forgery - block based methods and key point based methods. The general process of detection of copy move forgery is as follows:

1) **Pre-processing** - In order to enhance the important features of a digital image preprocessing is very important. If needed, pre-processing can be applied to both block based and key point based methods at the initial stages.

2) **Feature extraction** - Feature extraction is the most important step in image analysis. It is the process by which specific features of interest are detected in an image and then represented for further processing. Extraction of feature vectors is carried out in both the methods. In case of block based methods feature vectors are extracted from each block and in case of key point based method the features are extracted from the SIFT key points.

There are different types of features that are extracted from an image. These can be - local features, global features and pixel-level features.

Local image features are also known as salient features or interest points and are specific patterns which are different and unique from its neighboring pixels. These are associated with one or more image properties such as edges, corners, regions, etc. Global features are those that are calculated over the entire image and gives information of the image as a whole. Pixel level features are those that are calculated at each pixel like the location and color of a particular pixel, etc.

3) **Feature matching** - Copy move forgery can be detected by searching for similar feature vectors.

4) **Filtering** - The presence or absence of a forgery cannot be predicted on the basis of a single similarity criterion. So in order to reduce false prediction filtering methods are used.

5) **Post-processing** - After the image has been proved to be manipulated, post-processing is carried out.
A. Block based technique

Fridrich et al.\textsuperscript{13} presented a block matching technique, which lead to the development of several other works based on the same technique. According to this technique, instead of looking for the whole forged area, the image is divided into smaller overlapping square blocks. These blocks are then compared to their neighboring blocks to find out the matching blocks. The region that is enclosed by the coordinating blocks is the forged region.

Assuming that the cloned region is bigger than the size of the blocks and thus that the cloned region is made up of several overlapping cloned blocks, each of the cloned block will move with the same shift and so the distance between each cloned pair will be the same as well. Forgery can be detected by looking for a minimum number of similar blocks within the same distance and that are connected to each other forming areas depicting the same shape. All the other developed methods follow the same block matching procedure—first, dividing an M ×N image into overlapping square blocks, and then sliding these blocks by one pixel from the upper left corner to lower right corner. From each of the blocks, a
set of features in extracted and based on the fact that similar blocks will have similar features, lexicographical sorting is applied to the feature vectors of the block in order to find the cloned blocks. The adjacent connected pair of blocks are considered to be tempered regions. All the methods following this technique of block matching differ on the type of features chosen to represent the blocks. Fridrich proposed a method that used DCT (discrete cosine transform) coefficients.

The authors, Popescu and Farid\textsuperscript{14} proposed to use PCA (principal component analysis) of pixels for description of the image blocks. PCA-based description is shown to bring higher discriminative power, together with a reduction of the computational cost achieved through the truncation of less significant eigenvectors. The number of false positives produced in this method is comparatively lower than DCT method. Also, this method is proven to be resistant to image degradation caused due to noise or lossy compression.

Bayram, Sencar & Memon\textsuperscript{15} proposed a method in order to distinguish copy-move image forgery by using FMT (Fourier-Mellin Transform) since FMT is generally invariant to blurring, lossy compression, scaling and noise. In this study, the image was divided into small sub-images and then Fourier transform of these sub-blocks was computed. Now, at that point they re-sampled, projected and then quantized in order to obtain the desired feature vectors. After this process, feature vectors that are similar in nature are identified using either lexicographic sorting or counting bloom filters. But, there might be numerous matching blocks even in an original image. In order to overcome this, an image can be authenticated as being forged only in cases having a number of connected sub-blocks within the same block distance. This reduces the probability of false positives, making this approach very effective and robust to JPEG compression.

The main limitation of block based technique is that it does not prove to be a robust technique in cases where the copied regions are subjected to scaling or rotation operations prior to being pasted. In this regard, key point based methods are exploited for the same.

**B. Key point based technique**

Hailing et al.\textsuperscript{16} proposed a technique of identifying copy-move forgery by finding the correlation between original region and the cloned and pasted region of the image. This approach is based on Scale Invariant Feature Transform (SIFT) local features for accurate detection. The basic concept is the usage of SIFT descriptors to find regions which match within the same image.
SIFT is a method for generating image features. This method is based on transforming an image into a collection of local feature vectors, which are invariant to operations like image translation, scaling and rotation. These features are identified using a staged filtering approach. In the first stage key locations are identified by finding out locations that are maxima or a minima of a difference-of-Gaussian function. Difference-of-Gaussian is a feature enhancing algorithm and can be used to increase visibility of edges and other such details of a digital image. Each of the key locations or points identified is used to generate a feature vector and these resulting feature vectors are known as SIFT keys. These are coordinated with each other in order to detect image forgery. The presence of any matching SIFT points indicate copy-move image forgery. The drawback of this approach is that it is not useful if the forged region is small in size and if signal to noise ratio value is very low.

1.2 RESAMPLING

While making a composite, certain geometric transformations like rotating, stretching or resizing needs to be applied. For example when making a composite of two people, resizing of one of the person is done in order to match the height of the other person. In this process the original image is resampled on to a new sampling lattice which introduces certain periodic correlations between the neighboring pixels. These correlations do not occur in a natural digital image and so, their presence can be used for the detection of this type of manipulation. The resampling detection techniques can be used for detecting both benign editing, for example rotating and resizing of the whole image and as well as malicious editing(resampling of a certain portion of the image and in order changing the information carried by it).

Popescu and Farid\textsuperscript{17} proposed a method for detection of periodic correlations introduced in the image by common resampling kernels. In this approach, the Expectation-Maximization algorithm is applied in order to estimate the interpolation kernel parameters. A probability map known as p-map is obtained for each pixel which provides its probability to be correlated to neighboring pixels, which is used for the detection of resampling. This method has a very high accuracy, provided that the image has not been compressed.

1.3 SPLICING

Splicing is a common type of image forgery. In this type of manipulation two or more images are spliced into a single composite. When done carefully, the border between the spliced regions is visually undetectable. In order to detect splicing the joining regions are analyzed by
a variety of methods. The presence of sharp edges or changes between different regions and their surroundings contain valuable clues for detection of splicing. Splicing detection methods can be roughly divided into two categories: region based and boundary based. In the boundary based methods, the irregular modifications at the splicing boundaries are detected. Region based methods include checking the consistency on the generative model of the image that is obtained, in case of a non-blind detection, from the original and the spliced image in order to identify the forgery.

Boundary based methods fall under the category of pixel based detection techniques. One such boundary based method has been proposed by Fang et al.\textsuperscript{18} that relies on the sharp boundaries in color images. The method analyzes and checks the consistency of color division in the neighboring pixels of the boundary. According to the authors the abnormality at the color edge gives valuable evidence of image tempering.

1.4 STATISTICAL

Every digital image consists of two types of properties—one, the visual perception properties like contrast and brightness and the other is the statistical properties. Natural images exhibit certain statistical regularities that distinguish them from the sea of all possible images and these statistical properties can be used for detection of image manipulation. Hany Farid and S. Lyu in their paper\textsuperscript{19} have described a statistical model for natural or original images and is based upon a multi-scale wavelet-like decomposition. This model contains first- and higher-order statistics that capture the regularities that are inherent to natural images. They then demonstrated how this statistical model differentiates between natural and un-natural images. This difference can be used in several digital forensic applications, specifically in detecting various types of digital image forgery. The decomposition used by the authors splits the frequency space into multiple scale and orientation sub bands. The statistical model is composed of the first four statistical moments of each wavelet sub band and higher-order statistics that capture the correlations between the various sub bands. Supervised pattern classification is employed to classify images based on these statistical features.

2. FORMAT-BASED TECHNIQUES

After capturing an image using a digital camera, it is saved in the camera’s memory and in order to save space, lossy compression schemes such as JPEG are implied. JPEG stands for joint photography experts group and is a standard method of compressing digital images. It supports a varying degree of compression which allows the storage size and the image quality
to be adjusted as required. Generally JPEG can achieve a 10:1 compression level with minimal loss of image quality. JPEG compressed files have two sub file formats- JPEG/Exif, which is most commonly used in digital cameras and other photography equipment and JPEG/JFIF, which often used for storing and transmitting digital images on the world wide web. The JPEG/JFIF file format supports a maximum image size of 65,535×65,535 pixels.

The compression method used in JPEG is usually lossy. It means that during compression some of the original image information is lost which cannot be restored and this effects the image quality to some extent. This lossy compression used by JPEG is based on the discrete cosine transform (DCT). It is a mathematical operation that converts each frame or field of the image source from the spatial (2D) domain into the frequency domain (transform domain).

There are three categories under format based techniques. These are-

2.1 JPEG QUANTIZATION

The lossy compression scheme used in JPEG allows for some flexibility in the amount of compression needed. Manufacturers configure their devices in order to balance compression and quality to their own needs. This difference in configurations of different manufacturers can be used to identify the source of the image, i.e. the camera model/make. This has been described by Farid in his two papers 20, 21.

On the basis of a three-channel color image (RGB), the standard JPEG compression scheme is carried out as follows 22:

- **Color Transform (RGB → YCbCr)** - The first step is to transform the RGB image into luminance/chrominance space. The two chrominance channels Cb and Cr are subsampled by a factor of two, relative to the luminance channel, Y.
- **Image Partition** - Each of the channel is then portioned into 8×8 blocks of pixel.
- **Discrete Cosine Transform** - DCT is a mathematical operation that is converts each block from the 2-D domain (spatial domain) into the frequency domain.
- **Quantization** - Quantization reduces the number of bits per sample which makes this stage the primary source of compression and the main source for loss. The DCT coefficients obtained in the above step are then quantized to limited number of possible levels. The main aim of quantization is to reduce most of the less important high frequency DCT coefficients to zero because the more zeros the better the image will compress 23. Lower frequencies are used to reconstruct the image because human eye is more sensitive to them and higher frequencies are discarded. Quantization is viewed as a table containing 192 values- a set of 8×8 values which are associated each frequency
of each of the 3 channels-Y,Cb and Cr. These values are around 1 for low compression rates and increases in case of higher compression rate.

The above steps with some variations are employed by JPEG encoders in digital cameras and photo-editing softwares. These encoders vary according to the choice of quantization table hence embedding a kind of signature within each JPEG image. Fan and Queiroz\textsuperscript{24} have described a method to obtain the quantization tables from the encoded JPEG image and since these tables vary from one camera to another, they act as digital image ballistics using which the source of an image can be confirmed or denied.

### 2.2 DOUBLE JPEG

Any form of digital manipulation requires the image to be first loaded into a photo-editing software and then resaved. Now since most of the digital cameras and image processing softwares adopt the JPEG format, it is likely that both the original and the forged image is stored in this format. So, a manipulated image is required to be compressed twice and due to the lossy nature of the JPEG format, double compression of a single image introduces specific artefacts that are not present in the same image with single compression. Thus the presence of any tempering can be detected by analyzing the artefacts produced when the forged image is created. But the major drawback is that double JPEG compression does not necessarily indicate malicious tampering or forgery, for example saving an image after simply viewing it. The artefacts produced can be of two categories depending on whether the second compression uses a DCT grid aligned with the one used in the first compression or not. The former is called aligned double JPEG compression (A-DJPG) and the latter is referred to as non-aligned double JPEG compression (NA-DJPG).

#### A. A-DJPG detection

D. Fu, Y. Q. Shi, and W. Su\textsuperscript{25} proposed a method to detect A-DJPG that is based on the observation that in natural or non-tempered images the distribution of the first digit of DCT coefficients in single JPEG compressed images follows the generalized Benford’s law. Result from experiment show that with each compression step, the statistics of the first digit distribution gets altered. As a result, the accuracy of the fitting provided by the generalized Benford’s law decreases with increase in number of compression steps. However the performance of this method is not adequate when compared to the later proposed methods.

One of the promising idea was introduced by Lukáš and Fridrich\textsuperscript{26}. This method detects the presence of double-aligned JPEG compression based on the observation that
consecutive quantizations introduce periodic artifacts into the DCT coefficient’s histogram. These periodic artifacts are visible in the Fourier domain also known as frequency domain as strong peaks in medium and high frequencies in the Fourier domain are known as double quantization (DQ) effect. Depending upon the relationship between the quantization steps of the first and of the second compression, the peaks visible in the histogram assume different configurations. The presence of the so-called double peaks and missing centroids (those with very small probability) in the DCT coefficient histograms are said to be robust features that provide information about the primary quantization.

B. NA-DJPG detection

NA-DJPG compression detection method is also known as JPEG blocking which is the third category of format based techniques.

2.3 JPEG BLOCKING

One of the steps in the process of JPEG compression is the block DCT transform. Each of the 8×8 pixel block is individually transformed and quantized which results in the appearance of artifacts at the border of neighboring blocks in the form of horizontal and vertical edges. On manipulating an image, these blocking artifacts may be disturbed which can act as proof of manipulation.

Luo, Qu, Huang, Qiu\textsuperscript{27} proposed a method in which the blocking artifacts are characterized using pixel value differences both within and across the block boundaries. These differences are smaller within blocks as opposed to across blocks. Manipulation and recompression of an image introduces new set of blocking artifacts which do not necessarily align with the original boundaries. Pixel value differences of within and across the block boundaries are computed from 4-pixel neighborhoods that are spatially offset from each other by a fixed amount such that one of the neighborhood lies entirely within a JPEG block and the other one borders or overlaps a JPEG block. These differences are used to compute a histogram from all the 8×8 non overlapping blocks. The average difference between these histograms is computed by an 8×8 blocking artifact characteristics matrix. This matrix is not fixed and is random for uncompressed images, but it has a specific pattern for compressed images. So when an image is manipulated and recompressed, this pattern gets disrupted. This method has a disadvantage that it is reliable and robust only when the tampered region is very large, more than 500×500 pixels.

A fairly recent work was proposed by Bianchi and Piva\textsuperscript{28}. They proposed a forensic tool that can reveal tampering at a local level, without any prior information about the location of the manipulated area. The algorithm produces a map that gives the probability or the likelihood for each 8 × 8 image block to be tampered. The presence or the absence of NA-JPEG artifacts
At a local level can be interpreted as evidence of image tampering. This method is based on measuring how DCT coefficients cluster around a given lattice for any possible grid shift. This work has shown results that are more improved than the above method. Also, this method is effective even if the forged area is of 256×256 pixels.

3. **CAMERA BASED TECHNIQUES**

The four techniques for analyzing and estimating the different camera artifacts are as follows-

3.1 **CHROMATIC ABERRATION**

Each digital camera model/type presents individual lens characteristics due to the design and manufacturing process and this is the reason that lenses produce several types of aberrations. These aberrations leave unique traces on the images being captured that can be used to link an image to a particular device or to discover the presence of image modifications. Johnson and Farid emphasized on lateral chromatic aberration. This lens aberration causes different light wavelengths to focus on shifted points in the image plane represented by the sensor, when the source light is off the optical axis, resulting in a misalignment between color channels. By assuming that the lateral chromatic aberration is constant within each of the three color channels and by using the green channel as a reference, the aberrations between the red and green channels and between the blue and green channels are estimated. In particular, the lateral chromatic aberration is represented as a low-parameter model consisting of three parameters, two for the center of the distortion and one for the magnitude of the distortion; the estimation of these model parameters is framed as an image registration problem. Johnson and Farid detect image forgeries by looking for the presence of local deviations or inconsistencies in these models with respect to the parameters obtained for the whole image: image tampering is then detected if an inconsistency is found.
3.2 COLOR FILTER ARRAY

A digital color image has three channels which contain samples from different bands of the color spectrum like red, green and blue. Digital cameras are equipped with a single CCD (charge coupled device) or CMOS (complementary metal oxide semiconductor) sensor that captures color images using a color filter array (CFA). Most CFAs contain filters of the three primary colors-red, blue and green and is placed atop each sensor element. Since each pixel location is attributed with only a single color sample, estimation of the other two color samples is done from the neighboring samples in order to obtain a three channel color image. The method of estimating the missing color samples is known as CFA interpolation or demosaicing. The simplest demosaicing methods are based on kernels and acts on each channel independently. CFA interpolation introduces some specific statistical correlations between the image pixels which can be detected. Detection methods based on CFA demosaicing as fingerprint can be classified as two main classes: algorithms aiming at estimating the parameters of the color interpolation algorithm and the structure of the pattern filter and algorithms aiming at evaluating the presence/absence of demosaicing traces. Algorithms within the first class are mostly intended to classify different source cameras, since each camera brand could adopt different CFA configurations and different interpolation schemes. The second class focuses on forgery detection. Normally an image coming from a digital camera, in the absence of any successive processing, will show demosaicing artifacts; on the contrary, demosaicing inconsistencies between different parts of the image, as well as their absence in all or part of the analyzed image, will put image integrity in doubt.
Popescu and Farid\textsuperscript{10} proposed a technique for detecting the presence of CFA interpolation in an image by assuming a linear interpolation kernel, a simplistic but effective hypothesis compared to complex methods adopted in commercial devices, and using an Expectation-Maximization (EM) algorithm to estimate its parameters (i.e., filter coefficients) as well as the pattern of the filter. The method determines a correlation map, which gives for each pixel the probability of being correlated to neighboring pixels, according to the currently estimated kernel. Depending on the actual CFA pattern, some pixels are interpolated, whereas others are directly acquired. Hence, the correlation map exhibits a periodic behavior, which is clearly visible in the Fourier domain. Such an analysis can be applied to a given image region, to detect the presence of tampering; however, a minimum size is needed for assuring the accuracy of the results: authors tested their algorithms on 256×256 and 512×512 sized areas. This approach is less robust to JPEG compression but gives a lower false positive rate.

Ferrara et al.\textsuperscript{31} worked on a method which uses local analysis of CFA and image forgeries are identified whenever the presence of CFA interpolation is not present. Starting from such an assumption, a new feature is proposed, that measures the presence/absence of these artifacts even at the smallest 2×2 block level, thus providing as final output a forgery map indicating with fine localization the probability of the image to be manipulated.

3.3 CAMERA RESPONSE

Most digital camera sensors are very nearly linear that is these sensors convert an optical image into an analog signal in a line by line fashion. So, the relationship between the amount of light measured by each sensor element and the corresponding pixel value should be linear. But most cameras, in order to enhance the final image apply a pointwise nonlinearity.

Lin et al.\textsuperscript{32} have described ways to estimate this mapping from a single image which is termed as response function. Image tempering can be detected by finding the differences in the response function across the image.

Considering an edge where the pixels below the edge are of a constant color R and the pixels above the edge are of a different color B, the intermediate pixels along the edge should be a linear combination of the neighboring colors, if the camera response is linear. Camera response function can be estimated from the deviation of these intermediate pixel values from this expected linear response. The pixel colors are brought back to a linear relationship using the inverse camera response function which is estimated using a maximum posteriori (MAP) estimator. The estimator is stabilized by selecting edges such that the areas on both sides of the edge are similar, the variance on both side of the edge are small, the difference between R and B is large, and the pixels along the edge are between R and B. Some constraints are also imposed upon the estimated camera response function such as the function should be
monotonically increasing with at most one inflexion point and must be similar for each of the color channels. This camera response function can be estimated locally and any significant variations observed in this function across the image can be used as proof for image manipulation.

3.4 SENSOR NOISE

Imaging sensors of a digital camera have been shown to introduce various defects and to create noise in the pixel values. This sensor noise is introduced due to imperfections of the image sensor and is the result of three main components - pixel defects, fixed pattern noise (FPN) and Photo Response Non Uniformity (PRNU).

Pixel defects include point defects, hot point defects, dead pixels, pixel traps, and cluster defects. These vary with different sensors and with different camera model. Geradts et al. attempted to reconstruct pixel defects patterns. The authors in order to determine pixel noise, took images with black or green background with 12 different cameras and then compared the defect points which appeared as white. According to the result of their experiments, each camera has distinct patterns of defect pixels also across the same model. However, the impact of defect pixels is closely dependent on the content of the image. Moreover, some camera models do not contain any defectives pixels or they are removed. Therefore, this method is applicable to only certain camera models.

Fixed pattern noise (FPN) and Photo Response Non Uniformity (PRNU) are the two components of pattern noise, and depends on dark currents in the sensor and pixel non-uniformities respectively. So, they are not dependent on the image content but very closely related to the physical characteristics of each single sensor. Lukas and others have proposed another method which analyze pattern noise for camera identification since it is a unique stochastic characteristic for both CCD and CMOS sensors. As such, the pattern noise extracted from images taken by the same camera should be more correlated than those extracted from different cameras.

4. GEOMETRY BASED TECHNIQUES

The geometry based techniques are based on the fact that image forgery create inconsistencies from the geometric point of view. The major advantage of both geometry based and physics based techniques is that since these techniques are independent of low-level image characteristics, they are extremely robust to operations like compression, filtering and other image processing operations which make these techniques applicable even in cases where the image quality is low.
Geometry based techniques are of two categories which are as follows-

4.1 PRINCIPAL POINT

Principal point is simply the projection of the camera center onto the image plane and this is the point from which the focal length of the lens is measured. In authentic images, the principal point is nearby the image center. When image manipulation is carried out by translating a person or object in the image, the principal point gets shifted proportionally. This difference in the position of principal point across the image can be used for proving image tempering. K. Johnson and H. Farid proposed a method to estimate a camera’s principal point from different planar geometric shapes. They depicted the relationship between translation in the image plane and shift of the principal point, which is proportional or equivalent. Thus, any inconsistencies in the principal point across an image can act as evidence of forgery.

4.2 METRIC MEASUREMENTS

M. K. Johnson and H. Farid in a study, reviewed several tools from projective geometry that can be used for the rectification of planar surfaces and in some cases, can make real-world measurements from a planar surface. There are three techniques that can be used for the rectification of planar surfaces and each method can be carried out using a single image. These are as follows-

- The first method depends upon the knowledge of polygons of different shape that are known to people. For example- street sign, license plate, lettering on a billboard, etc.
- The second method requires information of two or more vanishing points on a plane. A vanishing point is actually an abstract point on the image plane. At this point 2D projections of a set of parallel lines seem to converge in a 3D space.
- The third method involves the requirement of two or more coplanar circles like car wheels, etc.

In each of these methods, the 3D world to 2D image transformation is estimated which allows the removal of planar distortions and to make metric measurements on the image plane.

5. PHYSICS BASED TECHNIQUES

This type of techniques are not based on signal properties but on the inconsistencies that are introduced due to manipulation at the scene level such as lighting inconsistencies, colors, perspective, etc.

These techniques are divided into three categories which are based on the differences in lighting
across an image. These are as follows-

5.1 LIGHT DIRECTION-2D

Lighting inconsistencies from an image can be detected using information of 2D surface and radiance from an image. Recently, work has been done by Johnson and Farid\(^\text{36}\) on this technique based on the lighting inconsistencies using specular highlights in the eyes. It explained the method to estimate the direction of light from the specular highlights from eyes. Finding the light source direction in a scene is not easy because of the difficulty of extracting 3D surface normals from a single image. So, in order to simplify it, a solution is proposed by M. K. Johnson and H. Farid which considers only the 2D surface normals at the occluding object boundary which require estimation of only two of the three components of the light direction. In order to further simplify the estimation of light source direction, certain assumption are made such as the surface of interest is Lambertian (surface reflects light isotropically), it possesses a constant reflectance value, and is illuminated by a point light source that infinitely far away.

5.2 LIGHT DIRECTION-3D

In the above method, the estimation of light source direction was limited to 2D because Determining 3D surface normals from a single image is very difficult. The authors, M. K. Johnson and H. Farid\(^\text{37}\) proposed a method to estimate 3D light direction that is based on the spotlight reflections in human eyes to check if two persons in the same image originally belong to different images or not.

5.3 LIGHT ENVIRONMENT

The above two mentioned techniques are based on a simplified lighting model consisting of a single dominant light source. But the lighting of a scene, in reality, can be very complex since the lights placed can vary in number and can be placed in any position which can create different lighting environments. Johnson and Farid\(^\text{38}\) have proposed a method to develop a low level representation of such complex lighting environments. Lighting environment coefficients are estimated by these methods which can be compared in order to find out lighting inconsistencies in the image.
COUNTERFORENSICS

There has been immense growth in the development of robust and highly effective forensic techniques for passive-blind detection of malicious digital image forgeries. As a result many attempts have been made in order to bypass these forensic techniques. Counterforensics or anti-forensics is nothing but the countermeasures to the investigation activities. For example- wiping off fingerprint from the crime scene. The techniques that challenges the digital forensic tools and projects their weaknesses by removing, hiding or overwriting the fingerprints produced by manipulation are known as counterforensic techniques. Counter-forensic techniques are classified as targeted or universal. Targeted techniques are those that removes the detectable traces with a specific image forensic technique while universal techniques are those that does the same using any unknown technique. Also, these techniques are called integrated if they are used at the same time or parallel with the image tampering operations and are called post processing if they are used after the image manipulations. Most of the work has been carried out on targeted and post processing counter forensics, keeping image compression as the subject.

JPEG compression is the most popularly used method and produces two important artifacts introduced because of its lossy nature. First is the quantization artifact formed due to quantization and the second are the blocking artifacts introduced due to pixel value discontinuities across block boundaries.

An attempt to hide DCT quantization artifacts have been proposed by Sreelakshmi and Venkataraman\textsuperscript{39}. This method acts by addition of counter noise to the DCT coefficients at the cost of some distortion being introduced within the acceptable limits. The additive noise distribution is critically chosen so that the DCT coefficients are not clustered around integer multiples of the step size. An attempt to suppress the JPEG blocking artifacts by smoothing followed by addition of low power White Gaussian noise while preserving the Laplacian distribution of the DCT coefficients has been proposed by Matthew et al\textsuperscript{40}. It uses DCT coefficients histogram and blocking artifact measure to detect if an image under test has undergone JPEG compression. To fool the forensic examination, it was proposed to add anti-forensic dither to the DCT coefficients after initial JPEG compression but before the second pass of JPEG compression. This will reflect as if the image has been obtained directly from a digital camera and has undergone JPEG compression within the camera just once.

The method of JPEG compression has been used since a long period of time and it has been studied extensively for forensic and anti-forensic purposes. In order to avoid the easy detection of
fingerprints introduced by image compression, new compression methods are being developed to be used in place of JPEG compression. One such method is the wavelet-based compression technique which is getting popular in today’s time since it provides better compression without the loss of much detail. Also, this compression technique do not divide the image into blocks. It analyzes the entire image and hence no blocking artifacts are produced.

COUNTERING ANTI-FORENSICS/COUNTER-FORENSICS

In order to cope up with the counter-forensics used by the forgers to hide the image tampering, many studies and methods have been proposed to detect fingerprints that might be introduced due to the use of anti-forensic techniques. Most of the counter-forensic techniques that have been developed hides JPEG compression artifacts. Now these techniques will also certainly leave fingerprints which are studied and explored in order to counter these techniques. Counter anti-forensics for wavelet based compression techniques have also been developed.

CONCLUSION

The concise number of techniques reviewed in this survey provide important results and a base for development of further robust and efficient image forgery detection techniques which play a major role in multimedia security of image authentication with no reference and those with no previous security protection.

The passive-blind techniques have been reviewed and classified on the basis of the different types of footprints acquired. Editing footprints or fingerprints are the fundamentals of pixel based techniques, geometry based techniques and physics based techniques. Coding fingerprints are the fundamentals of format based techniques and acquisition fingerprints are the fundamentals of camera based techniques.

Although all the techniques and proposals reviewed are quite efficient in detecting image forgery, they have their own pros and cons. Firstly, only a few of these methods can detect the actual forged area. Secondly, these techniques can give false positive results in certain cases due to which a final decision about the authenticity of an image cannot be relied on only a single technique. Thirdly, these techniques need human interpretation and hence cannot be automated.

Also, these techniques are specific to certain types of image forgery and in practice an image forgery analyst may not have prior knowledge about the forgery technique used which can
complicate the process of forgery detection. So, there is a sheer need of versatile forgery detection techniques.

In most of the existing literature and work, the three type of fingerprints- acquisition, coding and editing have been taken into account separately and no such relationship have been studied between the three fingerprints and the stages in which they are developed. So, there is scope of more complicated studies and algorithms that connect these fingerprints with one another in order to produce more reliable and efficient detection of image forgery.

REFERENCES

8. www.cambridgeincolour.com/learn-photography-concepts.htm


