Forensic Detection of Image Tampering Using Intrinsic Statistical Fingerprints in Histograms

Matthew C. Stamm and K. J. Ray Liu
Department of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742 USA
Email: { mcstamm,kjrliu }@umd.edu

***Abstract*—As the use of digital images has become more common throughout society, both the means and the incentive to create digitally forged images has increased. Accordingly, there is a great need for methods by which digital image alterations can be identified. In this paper, we propose several techniques for identifying digital forgeries by detecting the unique statistical fingerprints that certain image altering operations leave behind in an image's pixel value histogram. Specifically, we propose methods to detect the global and local application of contrast enhancement and to detect the addition of noise to a previously JPEG compressed image. These methods are tested through a number of experiments, and results showing the effectiveness of these algorithms are discussed.**

I. Introduction

In recent years, digital images have become increasingly prevalent throughout society. Many governmental, legal, scientific, and news media organizations rely on digital images to make critical decisions or to use as photographic evidence of specific events. This proves to be problematic, as the rise of digital images has coincided with the widespread availability of image editing software. At present, it is not difficult for an image forger to alter a digital image in a visually realistic manner. To avoid both embarrassment and legal ramifications, many of these organizations now desire some means of identifying image alterations so that the authenticity of a digital image can be verified. As a result, the field of digital image forensics has been born.

One of the primary goals of digital image forensics is the identification of images and image regions which have undergone some form of manipulation or alteration. Because of the ill-posed nature of this problem, no catchall method of detecting image forgeries exists. Instead, a number of techniques have been proposed to identify image alterations under a variety of scenarios. While each of these methods possess their own limitations, it has been posited that if a large set of forensic methods are developed, it will be difficult for a forger to create an image able to fool all image authentication techniques [1].

Previous image forensic work has dealt with the identification of computer generated objects within an image [2] as well as detecting lighting angle inconsistencies [3], [4]. Classifier based approaches have been proposed which identify image forgeries using a variety of statistical features [5], [6], [7]. While these methods are able to test for the use of a variety of image manipulating operations, they suffer the drawback of requiring training data to perform classification.

One set of digital forensic techniques aimed at detecting image tampering has grown out of research into imaging device identification. Forensic imaging device identification methods attempt to determine the type of device used to capture an image, ascertain the device manufacturer or model, and identify the particular imaging device used. These methods generally perform identification by estimating some device specific parameter such as color filter array (CFA) interpolation coefficients or sensor noise. Image forgery detection techniques have been proposed which operate by locating inconsistencies in these parameters [1], [8], or by using these parameters to estimate a tampering filter [9], [10], [11]. While these techniques are quite effective, they suffer the drawback of requiring either access to the imaging device, knowledge of the forensically significant device parameter, or access to a training database of images from which the device parameter can be inferred.

It is important to note that most image altering operations leave behind distinct, traceable "fingerprints" in the form of image distortions. Because these fingerprints, which can be deterministic or statistical in nature, are often unique to each operation, an individual test to catch each type of image manipulation must be designed. While detecting image forgeries using these techniques requires performing a large set of operation-specific tests, these methods enjoy the benefit of requiring no knowledge of a device specific parameter or training data from the device used to generate the image in question. Instead, these methods operate on the premise that the only information available is the image in question itself.

Prior work which identifies image tampering by detecting operation specific fingerprints includes the detection of resampling [12], double JPEG compression [13], [14], [15], as well as the parameterization of gamma correction [16]. Methods for detecting image forgeries have been proposed by detecting local abnormalities in an image's signal to noise ratio [13]. Inconsistencies in chromatic abberation [17] as well as the absence of CFA interpolation induced correlations [18] have been used to identify inauthentic regions of an image. Additionally, the efficient identification of copy and move forgeries has been studied [19], [13].

In this work, we propose a set of image forgery detection techniques which operate by detecting tampering fingerprints

in the form of statistical artifacts left in an image's pixel value histogram. By identifying the forensically significant properties of an unaltered image's pixel value histogram, we are able to identify the distinct fingerprints that each operation considered leaves behind. Specifically, we propose methods for detecting globally and locally applied deterministic contrast enhancement, as well as a method the global addition of noise to a previously JPEG compressed image. While much of this work focuses on detecting operations which alter the perceptual qualities of an image as opposed to its content, detecting these types of manipulations are still forensically significant. Operations such as contrast enhancement may be locally applied to disguise visual clues of image tampering. Localized detection of these operations can be used as evidence of cut and paste type forgery. Additive noise may be globally applied to an image not only to cover visual evidence of forgery, but also in an attempt to destroy forensically significant indicators of other tampering operations. Though the detection of these types of operations may not pertain to malicious tampering, they certainly throw in doubt the authenticity of the image and its content.

This paper is organized as follows. In Section II, we describe the forensically significant qualities of an unaltered image's pixel value histogram. We describe our proposed contrast enhancement detection techniques in Section III. Included are methods for detecting both globally and locally applied contrast enhancement.We develop a method for detecting the addition of noise to a previously JPEG compressed image in Section IV. Experiments designed to test the efficacy of each forensic scheme as well as simulation results are discussed after each detection method is proposed. We conclude this paper in Section V.

II. System Model and Assumptions

In this work, we consider digital images created by using an electronic imaging device to capture a real world scene. We adopt the following model of the digital capture process. Each pixel is assigned a value by measuring the light intensity reflected from a real world scene onto an electronic sensor over the area pertaining to that pixel. Inherent in this process is the addition of some zero mean sensor noise which arises due to several phenomena including shot noise, dark current, and on-chip amplifier noise [20]. For color images, it is often the case that the light passes through a color filter array so that only one color component is measured at each pixel location in this fashion. If this is the case, the color components not observed at each pixel are determined through interpolation. At the end of this process, the pixel values are quantized, then stored as the unaltered image.

When analyzing a digital image, a histogram $h(l)$ of the color or gray level values l recorded at each pixel can be generated by creating L equally spaced bins which span the range of possible pixel values, then tabulating the number of pixels whose value falls withing the range of each bin. Unless otherwise specified, we will hereafter assume that all color and gray level values lie in the set of integers between 0 and 255,

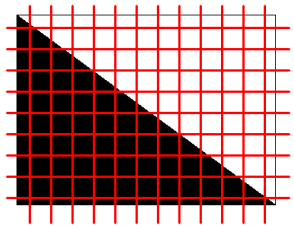


Fig. 1. Image sampling effects example.

and that all pixel value histograms are calculated using 256 bins so that each bin corresponds to a unique pixel value. We model all pixel value histograms as satisfying the following smoothness assumption; the entries of a pixel value histogram approximately conform to a smooth, low frequency envelope. It is worth explicitly noting that our model assumes that $h(l)$ contains no isolated spikes and that if $h(l) \gg 0$, then $h(l-1) \neq 0$ and $h(l+1) \neq 0$.

To justify our histogram model consider the simple case of imaging a real world scene consisting of two distinct color regions shown in Fig. 1. Instinctively, we might assume that the pixel value histogram of this image would consist of zeros everywhere except for two spikes located at the values corresponding to the colors present in the real world scene. Such a histogram would obviously violate our smoothness assumption. In this scenario, the border between the color regions does not align with the pixel boundaries on the sensor of the imaging device. Many pixels lying along the color border correspond to sensor areas containing both colors. The resulting values of each of these pixels will lie in the convex hull of the values corresponding to each of the two real world colors. The introduction of these new pixel values will effectively 'smooth out' the pixel value histogram.

Due to the complexity of real world scenes, it is exceedingly unlikely that the all color borders in an image will align directly with the pixel borders on an imaging device's sensor. Because of this, the effect described above should be present in virtually all real world images. Furthermore, additional factors contribute to the 'smoothness' phenomena. The complex nature of most natural and man-made lighting environments rarely result in a real world scene consisting of several distinct colors with no shading. Instead, a continuum of colors and illumination levels normally exist. Furthermore, the presence of observational noise will slightly change the value of several pixels during the image capture process, thus further smoothing the histogram.

III. Detecting Contrast Enhancement

In this section, we present a set of techniques designed to forensically detect the application of contrast enhancing operations to an image. While the detection of gamma correction has been previously examined [16], [13], this work assumes that the specific type of contrast enhancement which

may have been applied is known to the forensic examiner and that the contrast enhancement mapping can be described by a simple parametric equation. Here, we present a detection approach which can be used to detect more general contrast enhancement operations and which requires no a priori knowledge of the form of contrast enhancement potentially applied. We begin by discussing a method for detecting the global application of contrast enhancement [21]. We then adapt this scheme into one which can detect the local application of contrast enhancement and discuss how it can be used to detect cut and past forgeries in certain scenarios. Additionally we present a method for identifying the use of histogram equalization, a specific form of contrast enhancement.

A. Detection of Globally Applied Contrast Enhancement

Contrast enhancement operations seek to increase the dynamic range of pixel values in an image. For a deterministic globally applied contrast enhancement operation $T_{ce}(l)$, a nonlinear transformation $f(l)$ is applied to each pixel value l, then quantization is performed so that all pixel values lie in the allowable set. When coupled with the smoothness restriction placed on an image's pixel value histogram, the interaction between the nonlinear mapping used in each operation and quantization will leave behind a detectable statistical artifact.

A nonlinear mapping can be separated into regions where the mapping is locally contractive or expansive. For some distance measure $d(\cdot)$ such as the Euclidean distance, a mapping f is *contractive* if $d(f(u), f(v)) < d(u, v)$ and *expansive* if $d(f(u), f(v)) > d(u, v)$. When followed by quantization, the contractive regions in a contrast enhancement mapping can cause multiple unique input pixel values to be mapped to the same output value. This will result in the presence of an isolated peak in the histogram of the contrast enhanced image. Similarly, expansive regions in the contrast enhancement mapping can cause adjacent pixel values to be mapped one or more values apart, resulting in sudden gaps in the histogram of the enhanced image. These effects can be clearly seen in the top two plots of Fig. 2, which show the histogram of an image before and after it has undergone contrast enhancement.

Because of the smoothness restriction placed on the pixel value histograms of digital images, the peaks and gaps present in a contrast enhanced image's histogram are statistical artifacts of contrast enhancement which can be used to perform detection. These artifacts are particularly discernible when examining the frequency domain representation of an image's histogram. For an unaltered image, the discrete Fourier transform (DFT) of its pixel value histogram $H(k)$ should be a strongly low pass signal. The impulsive nature of the contrast enhancement artifacts present in an altered image's histogram, however, will result in the presence of a significant high frequency component in $H(k)$. The bottom two plots of Fig. 2 show the frequency domain representations of the histogram of a typical image before and after it has undergone contrast enhancement.

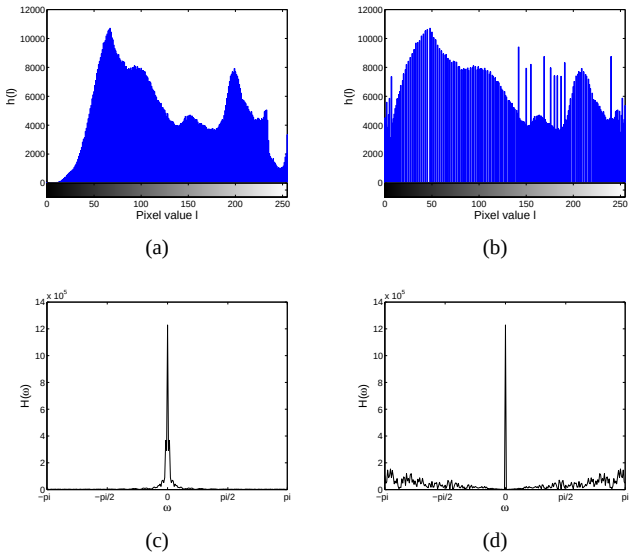


Fig. 2. Pixel value histogram of (a) an unaltered image and (b) the same image after contrast enhancement has been performed, as well as the magnitude of the DFT of (c) the unaltered image's histogram and (d) the contrast enhanced image's histogram.

While we expect $H(k)$ to be strongly low pass for unaltered digital images, there does exist one legitimately occurring phenomena which may violate this assumption. Consider the case of imaging a scene containing a bright background, such as the sky in the image shown in Fig. 3. In such a case, the light intensities recorded at pixels corresponding to the bright regions of the image may lie well above the cutoff for the highest quantization level. This will cause a substantial number of pixels to be assigned the value 255, thus creating impulsive spike in the image's pixel value histogram and adding a DC offset to $H(k)$. We refer to images which exhibit this behavior as *high end histogram saturated* images. Similarly, when capturing very dark scenes, a large number of pixels may be assigned the value 0, resulting in an impulsive spike at the low end of an image's histogram and the addition of a DC offset to $H(k)$. While this phenomena, which we refer to as *low end histogram saturation*, is less likely to arise due to the presence of sensor noise, we have observed it in several unaltered images.

We propose a detection scheme which operates by measuring the strength of the high frequency portions of $H(k)$, then comparing it to a threshold to determine if contrast enhancement has been performed. To mitigate the effects of high end and low end histogram saturation, we first obtain a modified histogram $g(l)$ free from saturation effects by performing the elementwise multiplication between $h(l)$ and a 'pinch off' function $p(l)$ so that

$$g(l) = p(l)h(l). \tag{1}$$

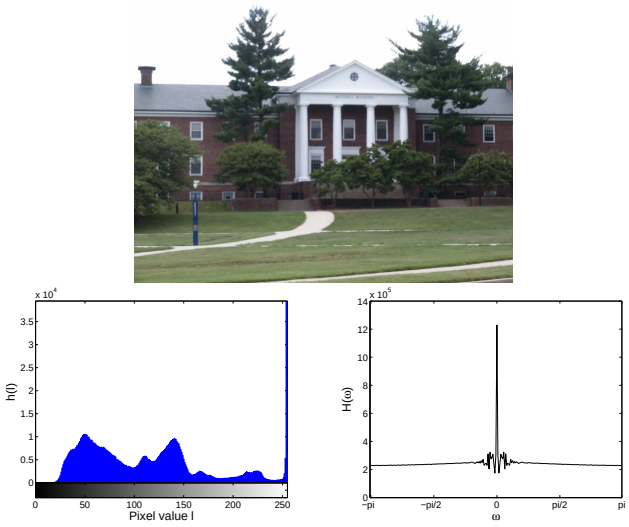


Fig. 3. Top: Image exhibiting high end histogram saturation. Bottom Left: Histogram of the image's green pixel values. Bottom Right: Magnitude of the DFT of the image's green pixel value histogram.

The pinch off function, defined as

$$p(l) = \begin{cases} \frac{1}{2} - \frac{1}{2}\cos\left(\frac{\pi l}{N_p}\right) & l \leq N_p \\ \frac{1}{2} + \frac{1}{2}\cos\left(\frac{\pi(l-255+N_p)}{N_p}\right) & l \geq 255 - N_p \\ 1 & else \end{cases} \quad (2)$$

is designed to remove any impulsive components in $h(l)$ which may legitimately arise due to saturation effects, as well as minimize the frequency domain effects of multiplying $h(l)$ by $p(l)$, which behaves similar to a windowing function. In (2), N_p is the width of the region over which $p(l)$ decays from 1 to 0.

We use $g(l)$ to calculate a normalized measure of the energy in the high frequency components of the pixel value histogram F according to the formula

$$F = \frac{1}{N}\sum_k |\beta(k)G(k)| \quad (3)$$

where N is the total number of pixels in the image, $G(k)$ is the DFT of $g(l)$, and $\beta(l)$ is a weighting function which takes values between 0 and 1. The purpose of $\beta(l)$ is to deemphasize low frequency regions of $G(l)$ where nonzero values do not necessarily correspond to contrast enhancement artifacts. In this work, we use the simple cutoff function

$$\beta(k) = \begin{cases} 1 & c \leq k \leq 128 \\ 0 & else \end{cases} \quad (4)$$

where c is the entry of the 256 point DFT corresponding to a desired cutoff frequency. $\beta(k)$ is zero for all values greater than $k = 128$ because symmetry properties inherent in the DFT of real valued signals make it unnecessary to measure these values.

After F has been calculated, the decision rule δ_{ce} is used to classify an image as unaltered or contrast enhanced, such that

$$\delta_{ce} = \begin{cases} \text{image is } \textit{not} \text{ contrast enhanced} & F < \eta_{ce} \\ \text{image is contrast enhanced} & F \geq \eta_{ce} \end{cases} \quad (5)$$

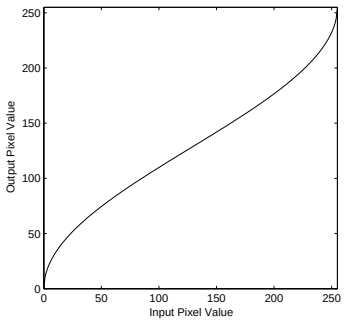


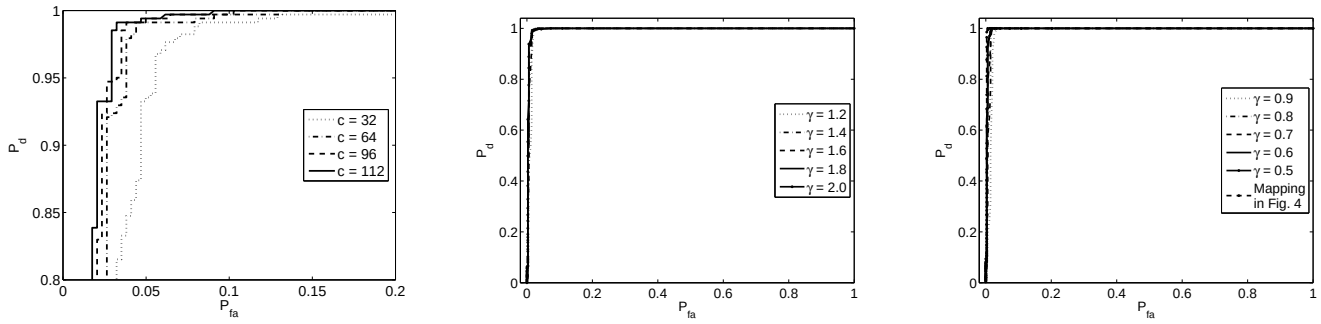
Fig. 4. Additional contrast enhancement mapping used.

To test the performance of our global contrast enhancement detection algorithm, we first compiled a database of 341 unaltered images consisting of many different subjects and captured under varying light conditions. These images were taken with several different cameras and range in size from 1500×1000 pixels to 2592×1944 pixels. To simplify our testing process, we used the green color layer of each of these images to form a set of unaltered grayscale images. Next, we created a set of contrast enhanced grayscale images by applying the power law transformation

$$T(l) = 255\left(\frac{l}{255}\right)^{\gamma} \quad (6)$$

to the pixel values of each of the unaltered grayscale images. This process was repeated for a variety of γ values ranging from 0.5 to 2.0, and the resulting images were saved as bitmaps. Additionally, contrast enhancement was performed on the unaltered images using the mapping displayed in Fig. 4 so that our database would include images which had undergone a nonstandard contrast enhancement transformation. The unaltered images were then combined with the contrast enhanced images to form a test database of 4092 grayscale images.

In order to choose an appropriate value of the cutoff parameter c for a large scale evaluation of our algorithm, we first conducted a small scale test using only unaltered images and those that had been altered using $\gamma = 0.6$. Each image was classified as unaltered or contrast enhanced by our detection scheme using values of c corresponding to angular frequencies ranging from $\frac{\pi}{4}$ to $\frac{7\pi}{8}$ and with $N_p = 4$. The probabilities of detection P_d and false alarm P_{fa} were determined for a series of decision thresholds η by respectively calculating the percent of contrast enhanced images correctly classified and the percent of unaltered images incorrectly classified. This information was used to generate the series of receiver operating characteristic (ROC) curves displayed in Fig. 5(a). The best performance was achieved for $c = 112$, which corresponds to the angular frequency $\frac{7\pi}{8}$. Furthermore,



(a) (b) (c)

Fig. 5. Contrast enhancement detection ROC curves for images altered by a power law transformation with (a) $\gamma = 0.6$ using several values of the cutoff parameter c, (b) $2.0 \geq \gamma \geq 1.2$, and (c) $0.5 \geq \gamma \geq 0.9$ as well as the mapping displayed in Fig. 4.

our detection algorithm's performance improved as the value of c was increased. This supports our assertion that $h(l)$ is a strongly lowpass signal and that contrast enhancement introduces into it an artificial high frequency component.

After choosing $c = 112$ in accordance with the results of our small scale experiment, we then used our contrast enhancement detection scheme to classify each image in the test database as unaltered or contrast enhanced. The detection and false alarm probabilities were then calculated for each type of contrast enhanced images in our database. A series of ROC curves, shown in Figs. 5(b) and (c), were generated to evaluate the performance of our detection algorithm. As can be seen in Fig. 5, our global contrast enhancement detection scheme performed very well against each of the forms of contrast enhancement included in our test database. It is worth explicitly noting that in each case, a P_d of 0.99 was achieved at a P_{fa} of approximately 0.03 or less.

B. Detection of Locally Applied Contrast Enhancement

The technique developed in Section III-A can be extended into a method that can be used to locate regions in an image to which contrast enhancement has been locally applied. Locally applied contrast enhancement can be modeled as the application of a nonlinear deterministic contrast enhancement mapping to the values of a contiguous set of pixels J within an image. Provided that the number of pixels in J is sufficiently large, a pixel value histogram of J should exhibit the same behavior as the histogram of a globally contrast enhanced image. The identification of localized contrast enhancement within an image can be performed by selecting a set of pixels J' comprising a region of interest, then applying the test discussed in Section III-A to the pixel value histogram of J'.

In some scenarios, a forensic examiner may wish to investigate a particular region within an image for evidence of contrast enhancement. In such cases, the testing set J' can be specified manually. It is more likely, however, that the entire image must be examined for evidence of local contrast enhancement. To accomplish this, the image can be segmented into fixed sized blocks, where each block constitutes a separate region of interest. Detection can then be performed on each block individually and the results can be aggregated to identify image regions which exhibit evidence of locally applied contrast enhancement.

In this technique, it is critical that the testing blocks are of sufficient size to yield a forensically useful histogram for contrast enhancement detection. If the blocks are too small, they may not contain enough pixels for the smooth histogram model to hold valid. In order to determine which block sizes are sufficient to perform reliable detection and examine the effectiveness of the local contrast enhancement detection scheme, the following experiment was performed. The unaltered images from the test database described in Section III-A along with the power law transformed images corresponding to $\gamma = 0.5$ through 0.9 were each segmented into square blocks. This process was performed for blocks of size 200×200, 100×100, 50×50, 25×25, and 20×20 pixels. Each block was then classified as contrast enhanced or unaltered using our contrast enhancement detection scheme. False alarm and detection probabilities were determined for each value of γ at all block sizes by calculating the percent of incorrectly classified unaltered blocks and the percent of correctly classified contrast enhanced blocks respectively. A set of ROC curves, shown in Fig. 6 were then generated showing the detector performance for each block size considered.

The ROC curves shown in Fig. 6 indicate that local contrast enhancement can be reliably detected using testing blocks sized least 100×100 pixels. At a P_{fa} of approximately 5%, a P_d of at least 95% was achieved using 200×200 pixel blocks and a P_d of at least 80% was achieved using 100×100 pixel blocks for each form of contrast enhancement tested. These results improved markedly when the contrast enhancement applied was stronger than the relatively mild power law transformation using $\gamma = 0.9$. In such cases, a P_d of roughly 98.5% and 96% was achieved with a P_{fa} of aproximatley 5% for blocks sized 200×200 pixels and 100×100 pixels respectively. It should also be noted that testing blocks sized 25×25 pixels and smaller appear to contain an insufficient number of pixels to perform reliable contrast enhancement detection.

In many situations, the detection of locally applied contrast

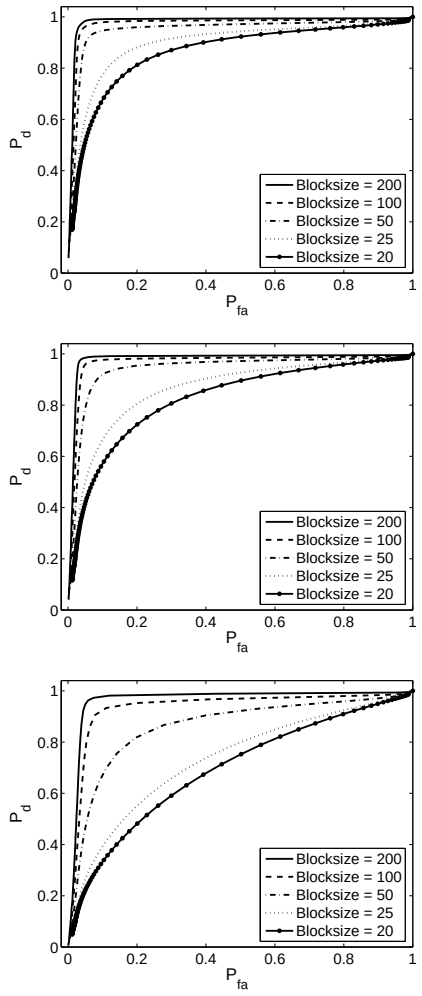


Fig. 6. ROC curves obtained using different testing block sizes for images altered by a power law transformation with $\gamma = 0.5$ (top), $\gamma = 0.7$ (middle), and $\gamma = 0.9$ (bottom).

enhancement is of greater forensic significance than the detection of its globally applied counterpart. One common scenario of particular importance is when contrast enhancement is coupled with cut and paste forgery. During cut and paste forgery, a forger creates a composite image I_f using two separate images I_1 and I_2. To accomplish this the forger replaces a set of pixels J_1 in I_1 with a contiguous set of pixels J_2 corresponding to an object in I_2. If I_1 and I_2 were captured under different lighting environments, as is often the case, it may be necessary for the forger to perform contrast enhancement on J_2 in order to make the composite image appear realistic. In such a scenario, the identification of localized contrast enhancement can be used to detect cut and paste forgery as well as identify J_2.

Fig. 7 shows an example of both a forged image generated in this manner, as well as the results of using our local contrast enhancement detection scheme to identify the inauthentic region. The forged image, shown in Fig. 7(c), was created from the images shown in Figs. 7(a) and 7(b) using Adobe Photoshop. This image was then segmented into 100×100 pixel blocks and our detection algorithm was used to identify blocks which contained contrast enhanced image regions. Figs. 7(d)-(f) show the results of performing localized contrast enhancement detection on the red, green, and blue color layers of the forged image. The blocks outlined in red represent local contrast enhancement detections and in each case contain pixels that correspond to the inauthentic object.

IV. Detecting Additive Noise in Previously JPEG Compressed Images

In this section we discuss the problem of detecting noise which has been globally added to an image that has previously undergone JPEG compression. At first glance, the addition of noise to an image may seem like a fairly innocuous operation. It can be used, however, to disguise visual clues of forgery or in an attempt to mask statistical artifacts left behind by other image altering operations. While prior work has dealt with the detection of locally added noise by estimating the variations in an image's signal to noise ratio (SNR) [13], this method fails when noise has been added to the entire image. Here, we present a method of noise detection that does not rely on estimating an image's SNR. We begin by discussing a simple operation which factors heavily into our detection method, continue by discussing a hypothesis testing scenario which relates to our problem, then finally proceed to develop a test for detecting additive noise in images which have previously undergone JPEG compression.

A. Scale and Round Operation

Consider the effect of the following operation, which we shall refer to as the *scale and round* operation,

$$v = \text{round}(cu) \tag{7}$$

where $u, v \in \mathbb{Z}$ and c is a fixed scalar. We define $\mathcal{U}_c(v)$ as the set of u values mapped to each distinct v value by (7), where

$$\mathcal{U}_c(v) = \{u | v = \text{round}(cu)\}. \tag{8}$$

The cardinality of this set, denoted by $|\mathcal{U}_c(v)|$, depends on the values of both c and v. It can be proven that if $c = \frac{p}{q}$ such that $p, q \in \mathbb{Z}$ are relatively prime, $|\mathcal{U}_c(v)|$ is periodic in v with period p. To see why this is so, consider first the following two lemmas:

Lemma 1: Given $a \in \mathbb{Z}$ and $b \in \mathbb{R}$

$$a = \text{round}(b) \Leftrightarrow a + k = \text{round}(b + k), \forall k \in \mathbb{Z}. \tag{9}$$

Lemma 2: Given $u, v \in \mathbb{Z}$ and $c = \frac{p}{q}$ such that $p, q \in \mathbb{Z}$ are relatively prime

$$v = \text{round}(cu) \Leftrightarrow v + p = \text{round}(c(u + q)) \tag{10}$$

Proof : By Lemma 1, $v = \text{round}(cu)$ implies that $v + p = \text{round}(cu + p)$. The right hand side of this equation can be rewritten as

$$\begin{aligned} v + p &= \text{round}(cu + p) \\ &= \text{round}(cu + \tfrac{p}{q}q) \\ &= \text{round}(c(u + q)). \end{aligned}$$

■

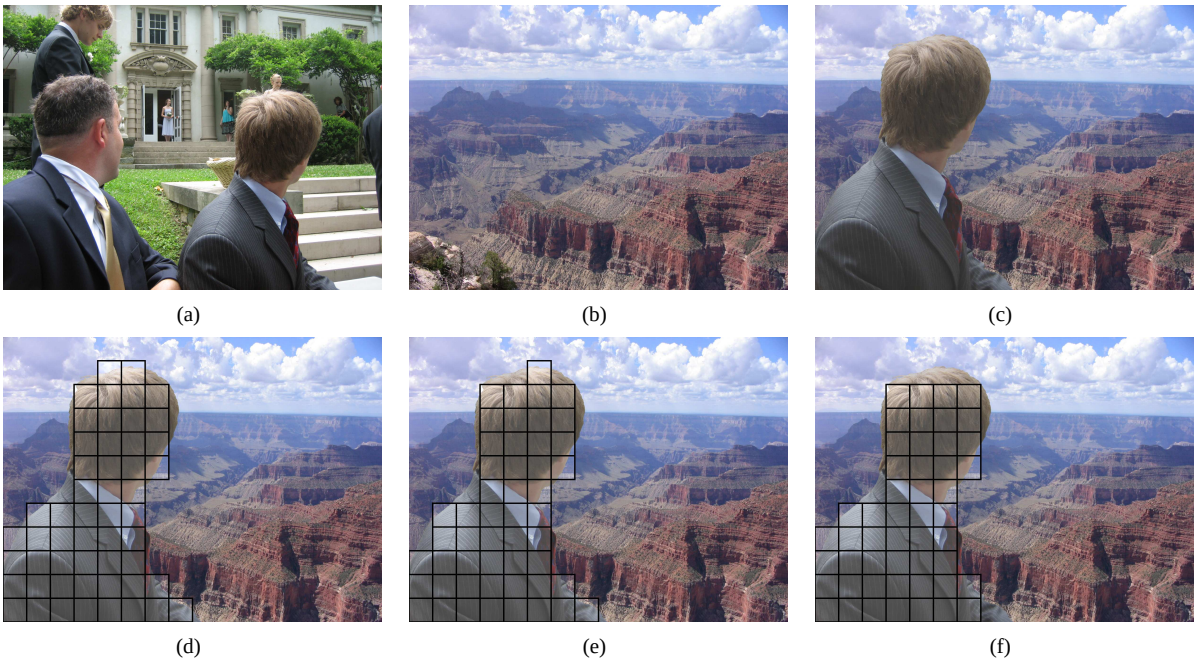


Fig. 7. Cut and paste forgery detection example showing (a) original image I_1, (b) original image I_2, (c) composite image I_f, (d) red layer blockwise detections, (e) green layer blockwise detections, and (d) blue layer blockwise detections. Blocks detected as contrast enhanced are highlighted and boxed.

Now using Lemma 2, we can state that for all $u \in \mathcal{U}_c(v)$, there exists some $\tilde{u} \in \mathcal{U}_c(v+p)$, namely $\tilde{u} = u + q$, which implies that $|\mathcal{U}_c(v)| = |\mathcal{U}_c(v+p)|$. This proves that the number of u values mapped to each v value is periodic with period p. We will make use of this property when developing our additive noise detection scheme.

B. Hypothesis Testing Scenario

When a color image is compressed as a JPEG, each pixel in the original image is first transformed from the RGB to the YCbCr color space using a linear transformation. Next, each color layer is segmented into blocks, and the discrete cosine transform (DCT) of each block is computed. The DCT coefficients are quantized by dividing each coefficient by its corresponding entry in a quantization matrix, then rounding the result to the nearest integer value. Finally, the resulting sequence of quantized DCT coefficients is rearranged and losslessly compressed.

Decompression is performed by first losslessly decoding the stream of quantized DCT coefficients and reconstituting it into blocks. The DCT coefficients are dequantized by multiplying each quantized DCT coefficient by its corresponding entry in the quantization matrix. Next, the inverse DCT (IDCT) is performed on each block of dequantized DCT coefficients, resulting in a set of pixels where each pixel is in the YCbCr color space. It is important to note that even though these values are not integers, they are still elements in a countable set due to the fact that the dequantized DCT coefficients are integers and the IDCT is a fixed linear transformation. Finally, the pixels are converted back to the RGB color space.

Several steps in this process are forensically significant to the detection of the addition of noise to an image that has been previously JPEG compressed. To develop a test to determine if noise has been added to such an image, let us first consider the following simplified hypothesis testing scenario.

Given the observation random variable $\mathbf{y} \in \mathbb{R}^m$, we wish to differentiate between the following two hypotheses:

$$\begin{aligned} H_0 &: \mathbf{y} = \mathbf{T}\mathbf{x} \\ H_1 &: \mathbf{y} = \mathbf{T}\mathbf{x} + \mathbf{n}. \end{aligned} \tag{11}$$

where $\mathbf{T} \in \mathbb{R}^{m \times m}$ is a known, invertible linear transformation, $\mathbf{n} \in \mathbb{R}^m$ is an independent random variable representing additive noise, and $\mathbf{x} \in \mathbb{Z}^m$ is a random variable whose probability mass function (PMF) is unknown but satisfies the smoothness restriction placed upon image pixel value histograms in Section II. In this scenario, we can view $\mathbf{y}$ as a pixel in the RGB color space, $\mathbf{x}$ as the same pixel in the YCbCr color space, and $\mathbf{T}$ as the linear transformation between the YCbCr and RGB color spaces. It should be noted that using traditional Bayesian techniques, we cannot differentiate between the two hypotheses because the distribution of $\mathbf{x}$ is unknown. Nonetheless, a means of differentiating between the hypotheses can still be achieved by examining the effect of applying the mapping

$$\mathbf{z} = \text{round}(c\mathbf{T}^{-1}\mathbf{y}) \tag{12}$$

to $\mathbf{y}$, where the constant $c = \frac{p}{q}$ is such that $p, q \in \mathbb{Z}$ are relatively prime.

Under hypothesis H_0, $\mathbf{z}$ can be written as

$$\begin{aligned} \mathbf{z} &= \text{round}(c\mathbf{T}^{-1}(\mathbf{T}\mathbf{x})) \\ &= \text{round}(c\mathbf{x}), \end{aligned} \tag{13}$$

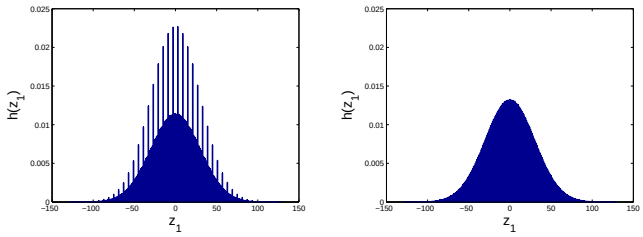


Fig. 8. Examples of normalized z_k histograms corresponding to hypothesis H_0 where no noise has been added (left) and hypothesis H_1 to which unit variance Gaussian noise has been added (right). In both cases the scaling parameter was chosen to be $c = \frac{6}{7}$.

therefore the k^{th} entry of $\mathbf{z}$ can be expressed using the formula

$$z_k = \text{round}(cx_k). \tag{14}$$

We can see that (14) is of the same form as (7), therefore the number of distinct x_k values mapped to each z_k value will occur in a fixed periodic pattern. Because an unequal number of x_k values are mapped to each z_k value and the PMF of $\mathbf{x}_k$ is smooth in the sense defined in Section II, a discernible periodic pattern will be present in the envelope of the PMF of z_k. Furthermore, the results of Section IV-A state that the period of this pattern is completely determined by our choice of c. This phenomenon can be clearly observed in the example shown in Fig. 8, where a set of periodically spaced spikes are present the histogram of z_1 under hypothesis H_0.

Under hypothesis H_1, we find that the PMF of z_k exhibits a different behavior. When $\mathbf{y}$ is subjected to the mapping described in (12), $\mathbf{z}$ can be rewritten as

$$\begin{aligned} \mathbf{z} &= \text{round}(c\mathbf{T}^{-1}(\mathbf{T}\mathbf{x} + \mathbf{n})) \\ &= \text{round}(c\mathbf{x} + c\mathbf{T}^{-1}\mathbf{n}). \end{aligned} \tag{15}$$

By defining the matrix $\mathbf{W}$ as the inverse of $\mathbf{T}$ such that

$$\mathbf{T}^{-1} = \mathbf{W} = \begin{bmatrix} W_{1,1} & \cdots & W_{1,m} \\ \vdots & \ddots & \vdots \\ W_{m,1} & \cdots & W_{m,m} \end{bmatrix}, \tag{16}$$

we can now express the k^{th} entry of $\mathbf{z}$ as follows:

$$\begin{aligned} z_k &= \text{round}\left(cx_k + \sum_{j=1}^{m} cW_{k,j}n_j \right) \\ &= \text{round}\left(\sum_{j=0}^{m} c_j s_j \right), \end{aligned} \tag{17}$$

where $c_0 = c$, $c_j = cW_{k,j}$, $s_0 = x_k$, and $s_j = n_j$ for $j = 1, \ldots, m$. Now, letting $r_j = \text{round}(c_j s_j)$, z_k can be approximated as

$$\begin{aligned} z_k &\approx \sum_{j=0}^{m} \text{round}\left(c_j s_j\right) + d_k \\ &= \sum_{j=0}^{m} r_j + d_k, \end{aligned} \tag{18}$$

where d_k is an independent random variable modeling the error induced by moving the summation outside of the round operation. Using this approximation, we can express the PMF of z_k, denoted by f_{z_k}, as

$$f_{z_k}(z_k) \approx (f_{r_0} * \cdots * f_{r_m} * f_{d_k})(z_k). \tag{19}$$

Because each r_j is generated by performing the scale and round operation on s_j, a periodic signal will be present in the envelope of the PMF of r_j. If the set of scaling constants c_j are such that the periodic components introduced share no common period, then the convolution of the PMFs in equation (19) will effectively smooth out the PMF of z_k. This will result in the absence of the periodic signal present in the envelope of the PMF of z_k under hypothesis H_0, as can be seen in Fig. 8.

Differentiating between the two hypotheses now becomes only a matter of detecting the presence of a signal of known period embedded in the PMF of z_k. The PMF of z_k cannot be directly observed, however, because the PMF of $\mathbf{x}$ is unknown. Instead, the PMF of z_k can be approximated by a normalized histogram of z_k values computed from a set of observations of $\mathbf{y}$. Detection of the periodic signal present in the histogram of z_k values can be accomplished using the frequency domain approach which we describe in Section IV-C.

C. Additive Noise Detection in Images

The detection of additive noise in a previously JPEG compressed image will differ slightly from the simplified hypothesis testing scenario previously described. In reality, a pixel $\mathbf{y}$ in the RGB color space lies in the set $\{0, \ldots, 255\}^3$, while a pixel $\mathbf{x}$ in the YCbCr color space lies in a countable subset of $\mathbb{R}^3$. When a pixel in the YCbCr color space is mapped back to the RGB color space during decompression, it must be projected into the set of allowable $\mathbf{y}$ values according to the equation

$$\mathbf{y} = \text{truncate}(\text{round}(\mathbf{T}\mathbf{x})) \tag{20}$$

where the operation $\text{truncate}(\cdot)$ maps values of its argument less than 0 to 0 and values greater than 255 to 255. By defining $Q(\mathbf{T}\mathbf{x}) = \text{truncate}(\text{round}(\mathbf{T}\mathbf{x})) - \mathbf{T}\mathbf{x}$, we can properly formulate the detection of this noise as the following hypothesis testing problem:

$$\begin{aligned} H_0 &: \mathbf{y} = \mathbf{T}\mathbf{x} + Q(\mathbf{T}\mathbf{x}) \\ H_1 &: \mathbf{y} = \mathbf{T}\mathbf{x} + Q(\mathbf{T}\mathbf{x}) + \mathbf{n}. \end{aligned} \tag{21}$$

This problem is similar in form to the simplified hypothesis test in (11) and can be solved using a refinement of the previously developed methods.

By choosing a rational constant $c = \frac{p}{q}$ and applying the operation $\mathbf{z} = \text{round}(c\mathbf{T}^{-1}\mathbf{y})$ to each pixel in the image, the hypothesis testing problem outlined in (21) becomes

$$\begin{aligned} H_0 &: \mathbf{z} = \text{round}(c\mathbf{x} + \mathbf{e}) \\ H_1 &: \mathbf{z} = \text{round}(c\mathbf{x} + \mathbf{e} + c\mathbf{T}^{-1}\mathbf{n}), \end{aligned} \tag{22}$$

where $\mathbf{e} = c\mathbf{T}^{-1}Q(\mathbf{T}\mathbf{x})$. Defining $\mathbf{W} = \mathbf{T}^{-1}$ and d_k as an independent random variable modeling the error summation

outside of the round operation as in Section IV-B, the k^{th} entry of $\mathbf{z}$ can be approximated under each hypothesis as

$$\begin{aligned} H_0 : z_k &= \text{round}(cx_k + e_k) \\ &\approx \text{round}(cx_k) + \text{round}(e_k) + d_k \\ H_1 : z_k &= \text{round}\Big(cx_k + c\sum_{j=1}^{3} W_{k,j}n_j + e_k\Big) \\ &\approx \text{round}(cx_k) + \sum_{j=1}^{3} \text{round}\left(cW_{k,j}n_j\right) \\ &\quad + \text{round}(e_k) + d_k. \end{aligned} \tag{23}$$

To differentiate between these hypotheses, we exploit differences in the PMF of z_k under each hypothesis.

Under hypothesis H_0, the PMF of z_k can be approximated as the convolution of the PMFs of $\text{round}(cx_k)$, $\text{round}(e_k)$, and d_k. Because the variances of the terms $\text{round}(e_k)$ and d_k are typically small, the term $\text{round}(cx_k)$ the dominates the behavior of the PMF of z_k and a periodic signal will be present in its envelope, as can be seen in Fig. 9. It should be noted that while the effects of terms $\text{round}(e_k)$ and d_k are minimal, they do produce a slight smoothing effect on the PMF of z_k. By contrast, the results of Section IV-B dictate that under H_1, this periodic signal will absent from the PMF of z_k. This phenomenon can be also be observed in Fig. 9, which shows the z_1 histograms of both an unaltered image and one to which unit variance Gaussian noise has been added.

Detecting the addition of noise to a previously JPEG compressed image can now be reformulated as detecting the presence of a signal of known period in the envelope of the PMF of z_k. While this PMF is unknown and cannot be directly observed, it can be approximated by creating a normalized histogram $h_{z_k}(l)$ of z_k values computed from each pixel in the image to be tested.

Because of its periodic nature, detection of the signal present in $h_{z_k}(l)$ under H_0 is particularly well suited for the frequency domain. To facilitate this, we obtain a frequency domain representation $G_{z_k}(b)$ of the histogram of z_k values free from any possible high or low end histogram saturation effects. This is accomplished by defining $G_{z_k}(b)$ as the DFT of $g_{z_k}(l)$, which we calculate using the equation

$$g_{z_k}(l) = h_{z_k}(l)p(l) \tag{24}$$

where $p(l)$ is the pinch off function denoted in (2). The signal we wish to detect will produce, if present, a peak in $G_{z_k}(b)$ centered at the b value corresponding to its fundamental frequency or an integer multiple thereof. The presence or absence of this peak under each hypothesis can be clearly seen in the example shown in Fig. 9.

To test for the presence of the peak in $G_{z_k}(b)$, we compare the value of $G_{z_k}(b)$ at the expected peak location b^* to its surrounding values. This is done using the following test statistic

$$S = \min\left\{ \frac{G_{z_k}(b^*)}{\frac{1}{|B_1|}\sum_{j\in B_1} G_{z_k}(j)}, \frac{G_{z_k}(b^*)}{\frac{1}{|B_2|}\sum_{j\in B_2} G_{z_k}(j)} \right\} \tag{25}$$

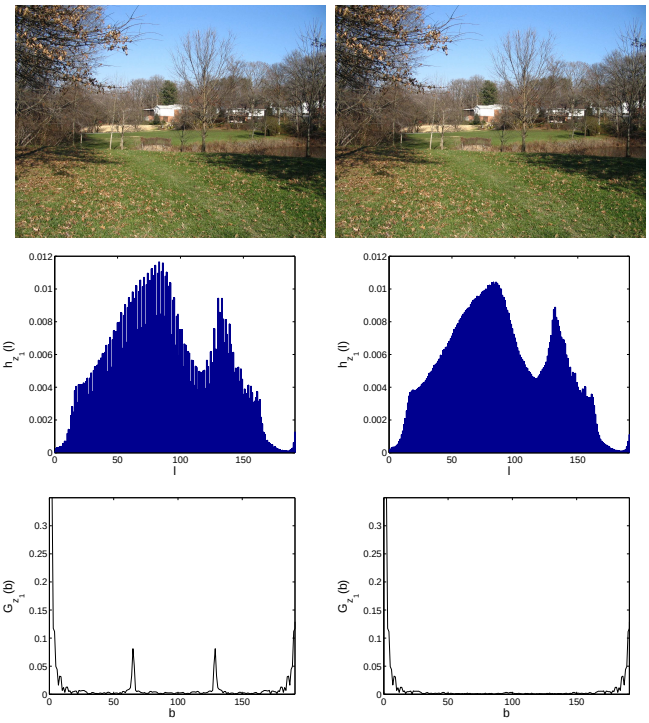


Fig. 9. Example showing an unaltered image (top left), its normalized z_1 histogram (middle left), and the magnitude of the DFT of its z_1 histogram (bottom left), as well as an altered version of the image to which unit variance Gaussian noise has been added (top right), its normalized z_1 histogram (middle right), and the magnitude of the DFT of its z_1 histogram (bottom right). In both cases, the scaling parameter was chosen to be $c = \frac{3}{4}$.

where B_1 and B_2 are sets of contiguous indices of G_{z_k} lying above and below b^* respectively. These sets should not include indices directly adjacent to b^*, because DFT windowing effects may result in higher values of $G_{z_k}(b)$ around the peak if it is present. Additionally, the smoothness restriction placed upon the histogram of pixel values in our image model implies that $G_{z_k}(b)$ will be strongly low pass in nature. This property suggests that to achieve better differentiability, c should be chosen such that it introduces a high frequency signal into $h_{z_k}(l)$.

A decision rule δ_n is then used to to determine the presence or absence of the peak in $G_{z_k}(b)$, and thus the presence or absence of additive noise in the image. This is accomplished by comparing the value of S to a predefined threshold T as follows:

$$\delta_n = \begin{cases} \text{noise has } \textit{not} \text{ been added} & \text{if } S < T \\ \text{noise has been added} & \text{if } S \geq T. \end{cases} \tag{26}$$

To test the performance of our additive noise detection algorithm, we compiled a set of 277 unaltered images consisting of a variety of different scenes. These images were taken using four different digital cameras, each from a different manufacturer, and were saved as JPEG compressed images using each camera's default settings. A set of altered images was created by decompressing each image and globally adding white Gaussian noise of unit variance to each pixel value.

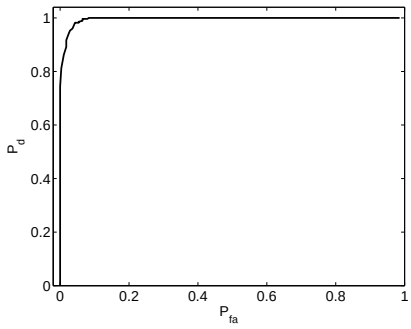


Fig. 10. Additive noise detection ROC curve for images altered by adding unit variance Gaussian additive noise.

These altered images were then saved as bitmaps, along with decompressed versions of the original images, creating a testing database of 554 images. Next we used our additive noise detection test to determine if noise had been added to each image in the database. When creating the histogram of z_k values, we chose $k = 1$ which corresponds to using the luminance or "Y" component of each pixel. The parameter c was chosen to take the value $c = \frac{7}{11}$ leading to an expected peak location of $b^* = 71$. The sets of B_1 and B_2 were chosen to be $B_1 = \{61, \ldots, 68\}$ and $B_2 = \{74, \ldots, 81\}$.

Detection and false alarm probabilities were determined by calculating the percentages of correctly classified images to which noise had been added and incorrectly classified unaltered images respectively. Using this data, an ROC curve showing the performance of our additive noise detection algorithm is displayed in Fig. 10. A P_d of approximately 80% was achieved at a false alarm rate less than 0.4%. When the P_{fa} was held less than 6.5%, the P_d increased to nearly 99%. These results indicate that our detection scheme is able to reliably detect additive noise in previously JPEG compressed images.

V. Conclusion

In this paper, we propose a set of digital forensic techniques which identify image alterations by detecting the unique tampering fingerprints these alterations leave in an image's pixel value histogram. We characterize the forensically significant properties of an image's pixel value histogram and provide a justification for their presence. We identify globally applied contrast enhancement by detecting the high frequency component that it introduces into an image's pixel value histogram. We extend this technique into a method for detecting locally applied contrast enhancement and demonstrate its usefulness in detecting cut and paste type forgeries. We detect the global addition of noise to an image by determining the presence of a specific periodic component in a histogram of values obtained by applying a specific nonlinear mapping to the image's pixel values. Simulation results indicate that each of these forensic methods are able to detect with a high degree of accuracy the application of the image altering operation for which they were designed.

References

[1] M. Chen, J. Fridrich, M. Goljan, and J. Lukáš, "Determining image origin and integrity using sensor noise," *IEEE Trans. on Inform. Forensics Security*, vol. 3, no. 1, pp. 74–90, March 2008.

[2] T.-T. Ng, S.-F. Chang, J. Hsu, L. Xie, and M.P. Tsui, "Physics-motivated features for distinguishing photographic images and computer graphics," in *Proc. ACM Multimdedia*, Singapore, 2005, pp. 239–248, ACM.

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

[4] M.K. Johnson and H. Farid, "Exposing digital forgeries by detecting inconsistencies in lighting," in *Proc. ACM Multimedia and Security Workshop*, New York, NY, USA, 2005, pp. 1–10.

[5] T.-T. Ng, S.-F. Chang, and Q. Sun, "Blind detection of photomontage using higher order statistics," in *Proc. IEEE Int. Symp. Circuits Systems*, Vancouver, BC, Canada, May 2004, vol. 5, pp. V–688–V–691.

[6] S. Bayram, I. Avcibaş, B. Sankur, and N. Memon, "Image manipulation detection," *Journal of Electronic Imaging*, vol. 15, no. 4, pp. 041102, 2006.

[7] I. Avcibas, S. Bayram, N. Memon, M. Ramkumar, and B. Sankur, "A classifier design for detecting image manipulations," in *Proc. ICIP*, Oct. 2004, vol. 4, pp. 2645–2648.

[8] J. Lukáš, J. Fridrich, and M. Goljan, "Detecting digital image forgeries using sensor pattern noise," in *Proc. SPIE, Electronic Imaging, Security, Steganography, Watermarking of Multimedia Contents*, San Jose, CA, USA, Feb. 2006, vol. 6072, pp. 362–372.

[9] A. Swaminathan, Min Wu, and K.J.R. Liu, "Digital image forensics via intrinsic fingerprints," *IEEE Trans. on Inform. Forensics Security*, vol. 3, no. 1, pp. 101–117, March 2008.

[10] A. Swaminathan, M. Wu, and K. J. R. Liu, "Intrinsic fingerprints for image authentication and steganalysis," in *Proc. SPIE Conf. Security, Steganography, Watermarking of Multimedia Contents*, San Jose, CA, USA, Feb. 2007, vol. 6505.

[11] A. Swaminathan, Min Wu, and K.J.R. Liu, "Image tampering identification using blind deconvolution," in *Proc. ICIP*, Atlanta, GA, USA, Oct. 2006, pp. 2309–2312.

[12] A.C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of resampling," *IEEE Trans. Signal Processing*, vol. 53, no. 2, pp. 758–767, Feb. 2005.

[13] A.C. Popescu and H. Farid, "Statistical tools for digital forensics," in *Proc. 6th Int. Workshop on Information Hiding*, Toronto, Canada, 2004.

[14] T. Pevný and J. Fridrich, "Detection of double-compression in jpeg images for applications in steganography," *IEEE Trans. on Information Forensics and Security*, vol. 3, no. 2, pp. 247–258, June 2008.

[15] J. Lukáš and J. Fridrich, "Estimation of primary quantization matrix in double compressed JPEG images," *in Proc. Digital Forensic Research Workshop*, pp. 5–8, 2003.

[16] H. Farid, "Blind inverse gamma correction," *IEEE Trans. on Image Processing*, vol. 10, pp. 1428–1433, Oct. 2001.

[17] M. K. Johnson and H. Farid, "Exposing digital forgeries through chromatic aberration," in *Proc. ACM Multimedia and Security Workshop*, Geneva, Switzerland, 2006, pp. 48–55.

[18] A.C. Popescu and H. Farid, "Exposing digital forgeries in color filter array interpolated images," *IEEE Trans. on Signal Processing*, vol. 53, no. 10, pp. 3948–3959, Oct. 2005.

[19] J. Fridrich, D. Soukal, and J. Lukáš, "Detection of copy-move forgery in digital images," in *Proc. Digital Forensic Research Workshop*, 2003.

[20] G.E. Healey and R. Kondepudy, "Radiometric ccd camera calibration and noise estimation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 3, pp. 267–276, Mar 1994.

[21] M. Stamm and K.J.R. Liu, "Blind forensics of contrast enhancement in digital images," in *Proc. ICIP*, San Diego, CA, USA, Oct. 2008, pp. 3112–3115.