Hany Farid

Image Forgery Detection

A survey

We are undoubtedly living in an age where we are exposed to a remarkable array of visual imagery. While we may have historically had confidence in the integrity of this imagery, today's digital technology has begun to erode this trust. From the tabloid magazines to the fashion industry and in mainstream media outlets, scientific journals, political campaigns, courtrooms, and the photo hoaxes that land in our e-mail in-boxes, doctored photographs are appearing with a growing frequency and sophistication. Over the past five years, the field of digital forensics has emerged to help restore some trust to digital images. Here I review the state of the art in this new and exciting field.

Digital watermarking has been proposed as a means by which an image can be authenticated (see, for example, [21] and [5] for general surveys). The drawback of this approach is that a watermark must be inserted at the time of recording, which would limit this approach to specially equipped digital cameras. In contrast to these approaches, passive techniques for image forensics operate in the absence of any watermark or signature. These techniques work on the assumption that although digital forgeries may leave no visual clues that indicate tampering, they may alter the underlying statistics of an image. The set of image forensic tools can be roughly grouped into five categories: 1) pixel-based techniques that detect statistical anomalies introduced at the pixel level; 2) format-based techniques that leverage the statistical correlations introduced by a specific lossy compression scheme; 3) camera-based techniques that exploit artifacts introduced by the camera lens, sensor, or on-chip postprocessing; 4) physically based techniques that explicitly model and detect anomalies in the three-dimensional interaction between physical objects, light, and the

Digital Object Identifier 10.1109/MSP.2008.931079

camera; and 5) geometric-based techniques that make measurements of objects in the world and their positions relative to the camera.

I will review several representative forensic tools within each of these categories. In so doing, I have undoubtedly omitted some worthy papers. My hope, however, is that this survey offers a representative sampling of the emerging field of image forgery detection.

PIXEL-BASED

The legal system routinely relies on a range of forensic analysis ranging from forensic identification (Deoxyribonucleic acid (DNA) or fingerprint) to forensic odontology (teeth), forensic entomology (insects), and forensic geology (soil). In the traditional forensic sciences, all manner of physical evidence is analyzed. In the digital domain, the emphasis is on the pixel—the underlying building block of a digital image. I describe four techniques for detecting various forms of tampering, each of which directly or indirectly analyzes pixel-level correlations that arise from a specific form of tampering.

CLONING

Perhaps one of the most common image manipulations is to clone (copy and paste) portions of the image to conceal a person or object in the scene. When this is done with care, it can be difficult to detect cloning visually. And since the cloned regions can be of any shape and location, it is computationally impossible to search all possible image locations and sizes. Two computationally efficient algorithms have been developed to detect cloned image regions ([11], [34]; see also [23], [27], and [42]).

The authors in [11] first apply a block discrete cosine transform (DCT). Duplicated regions are detected by lexicographically sorting the DCT block coefficients and grouping similar blocks with the same spatial offset in the image. In a related approach, the authors in [34] apply a principal component analysis (PCA) on small fixed-size image blocks to yield a reduced-dimension representation. Duplicated regions are again detected by lexicographically sorting and grouping all of the image blocks. Both the DCT and PCA representations are employed to reduce computational complexity and to ensure that the clone detection is robust to minor variations in the image due to additive noise or lossy compression.

RESAMPLING

To create a convincing composite, it is often necessary to resize, rotate, or stretch portions of an image. For example, when creating a composite of two people, one person may have to be resized to match the relative heights. This process requires resampling the original image onto a new sampling lattice, introducing specific periodic correlations between neighboring pixels. Because these correlations are unlikely to occur naturally, their presence can be used to detect this specific manipulation ([36]; related approaches are described in [43],[22], [31], and [38]).

Consider the simple example of upsampling a 1-D signal $x(t)$ of length m by a factor of two using linear interpolation to yield $y(t)$. The odd samples of the resampled signal take on the values of the original signal: $y(2i-1) = x(i), i = 1, \ldots, m$, while the even samples are the average of adjacent neighbors of the original signal

$$y(2i) = 0.5x(i) + 0.5x(i+1). \qquad (1)$$

Since each sample of the original signal can be found in the resampled signal, the interpolated pixels can be expressed in terms of the resampled samples only:

$$y(2i) = 0.5y(2i-1) + 0.5y(2i+1). \qquad (2)$$

That is, across the entire resampled signal, each even sample is precisely the same linear combination of its adjacent two neighbors. In this simple case, a resampled signal can be detected by noticing that every other sample is perfectly correlated with its neighbors. This correlation is not limited to up-sampling by a factor of two. A large range of resamplings introduces similar periodic correlations. If the specific form of the resampling correlations is known, then it would be straightforward to determine which pixels are correlated with their neighbors. If it is known which pixels are correlated with their neighbors, then the specific form of the correlations can easily be determined. But in practice neither is known. The expectation/maximization (EM) algorithm is used to simultaneously solve each of these problems. The EM algorithm is a two-step iterative algorithm: 1) in the expectaton step, the probability of each pixel being correlated with its neighbor is estimated; and (2) in the maximization step, the specific form of the correlation between pixels is estimated. Assuming a linear interpolation model, the expectation step reduces to a Bayesian estimator, and the maximization step reduces to weighted least-squares estimation. The estimated probability is then used to determine if a portion of the image has been resampled.

[TODAY'S TECHNOLOGY ALLOWS DIGITAL MEDIA TO BE ALTERED AND MANIPULATED IN WAYS THAT WERE SIMPLY IMPOSSIBLE 20 YEARS AGO.]

SPLICING

A common form of photographic manipulation is the digital splicing of two or more images into a single composite. When performed carefully, the border between the spliced regions can be visually imperceptible. In [7] and [32], however, the authors show that splicing disrupts higher-order Fourier statistics, which can subsequently be used to detect splicing.

Consider a 1-D signal $x(t)$ and its Fourier transform $X(\omega)$. The power spectrum $P(\omega) = X(\omega)X^*(\omega)$ is routinely used to analyze the frequency composition of a signal ($*$ denotes complex conjugate). Moving beyond the power spectrum, the bispectrum

$$B(\omega_1,\omega_2) = X(\omega_1)X(\omega_2)X^*(\omega_1+\omega_2) \qquad (3)$$

measures higher-order correlations between triples of frequencies ω_1, ω_2, and $\omega_1 + \omega_2$. Subtle discontinuities that result from splicing manifest themselves with an increase in the magnitude of the bispectrum and in a bias in the bispectrum phase, which are used to detect splicing in audio [7] and in images [32].

STATISTICAL

There are a total of 256^{n^2} possible 8-b gray-scale images of size $n \times n$. With as few as $n = 10$ pixels, there are a whopping 10^{240} possible images (more than the estimated number of atoms in the universe). If we were to draw randomly from this enormous space of possible images, it would be exceedingly unlikely to obtain a perceptually meaningful image. These observations suggest that photographs contain specific statistical properties. The authors in [9], [1], and [2] exploit statistical regularities in natural images to detect various types of image manipulation.

The authors in [9] compute first- and higher-order statistics from a wavelet decomposition. This decomposition splits the frequency space into multiple scale and orientation subbands. The statistical model is composed of the first four statistical moments of each wavelet subband and higher-order statistics that capture the correlations between the various subbands. Supervised pattern classification is employed to classify images based on these statistical features. In a complementary approach, the authors in [1] construct a statistical model based on local co-occurrence statistics from image bit-planes. Specifically, the first four statistical moments are computed from the frequency of bit agreements and disagreements across bit planes. Nine features embodying binary string similarity are extracted from these measurements. Another eight features are extracted from the histograms of these measurements. The sequential floating forward search algorithm is used to select the most descriptive features, which are then used in a linear regression classifier for discriminating authentic from manipulated images. In both cases, the statistical model is used to detect everything from basic image manipulations such as resizing and filtering [1] to discriminating photographic from computer-generated images [29] and detecting hidden messages (steganography) [30].

[THE FIELD OF IMAGE FORENSICS, HOWEVER, HAS MADE AND WILL CONTINUE TO MAKE IT HARDER AND MORE TIME-CONSUMING (BUT NEVER IMPOSSIBLE) TO CREATE A FORGERY THAT CANNOT BE DETECTED.]

FORMAT BASED

The first rule in any forensic analysis must surely be "preserve the evidence." In this regard, lossy image compression schemes such as JPEG might be considered a forensic analyst's worst enemy. It is ironic, therefore, that the unique properties of lossy compression can be exploited for forensic analysis. I describe three forensic techniques that detect tampering in compressed images, each of which explicitly leverages details of the JPEG lossy compression scheme.

JPEG QUANTIZATION

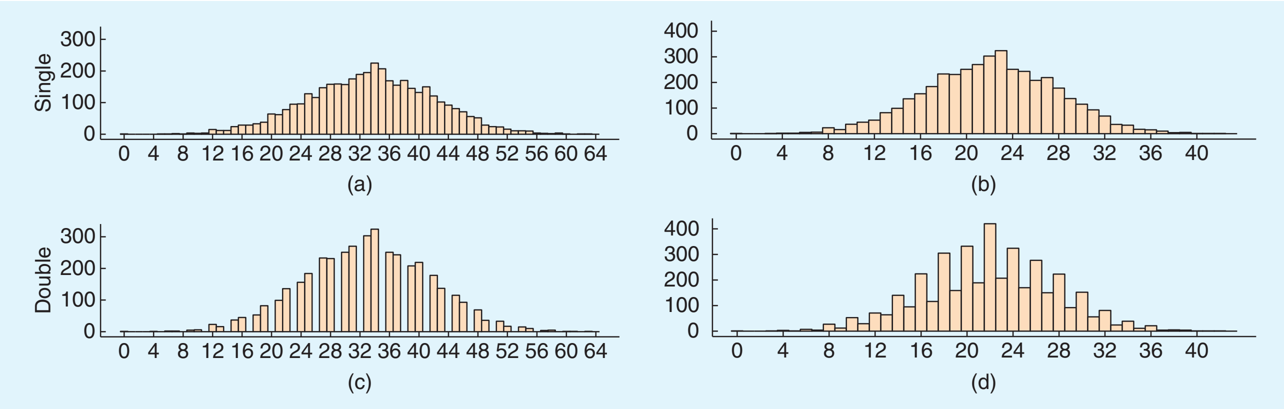
Most cameras encode images in the JPEG format. This lossy compression scheme allows for some flexibility in how much compression is achieved. Manufacturers typically configure their devices differently in order to balance compression and quality to their own needs and tastes. As described in [8] and [44], this difference can be used to identify the source (camera make/model) of an image.

Given a three-channel color image (RGB), the standard JPEG compression scheme proceeds as follows: The RGB image is first converted into luminance/chrominance space (YCbCr). The two chrominance channels (CbCr) are typically subsampled by a factor of two relative to the luminance channel (Y). Each channel is then partitioned into 8×8 pixel blocks. These values are converted from unsigned to signed integers (e.g., from $[0, 255]$ to $[-128, 127]$). Each block is converted to frequency space using a 2-D discrete cosine transform (DCT). Depending on the specific frequency and channel, each DCT coefficient, c, is then quantized by an amount q: $\lfloor c/q \rfloor$. This stage is the primary source of compression. The full quantization is specified as a table of 192 values—a set of 8×8 values associated with each frequency, for each of three channels (YCbCr). For low compression rates, these values tend toward a value of 1 and increase for higher compression rates. With some variations, the above sequence of steps is employed by JPEG encoders in digital cameras and photo-editing software. The primary source of variation in these encoders is the choice of quantization table. As such, a signature of sorts is embedded within each JPEG image. The quantization tables can be extracted from the encoded JPEG image or blindly estimated from the image, as described in [6].

Note that the quantization tables can vary from within a single camera as a function of the quality setting, and while the tables are somewhat distinct, there is some overlap across cameras of different makes and models. Nevertheless, this simple observation allows for a crude form of digital image ballistics, whereby the source of an image can be confirmed or denied.

DOUBLE JPEG

At a minimum, any digital manipulation requires that an image be loaded into a photo-editing software program and resaved. Since most images are stored in the JPEG format, it is likely that both the original and manipulated images are stored in this format. In this scenario, the manipulated image is compressed twice. Because of the lossy nature of the JPEG image format, this double compression introduces specific artifacts not present in singly compressed images (assuming that the image was not also cropped prior to the second compression). The presence of these artifacts can, therefore, be used as evidence of some manipulation [25], [35]. Note that double JPEG compression does not necessarily prove malicious tampering.



[FIG1] **Shown along the top row are histograms of single-quantized signals with (a) Step 2 and (b) Step 3. Shown in the bottom row are histograms of double-quantized signals with (c) Step 3 followed by Step 2, and (d) Step 2 followed by Step 3. Note the periodic artifacts in the histograms of double-quantized signals.**

For example, it is possible to inadvertently save an image after simply viewing it.

As described in the previous section, quantization of the DCT coefficients c is the primary manner in which compression is achieved, denoted as, $q_a(c) = \lfloor c/a \rfloor$, where a is the quantization step (a strictly positive integer). Dequantization brings the quantized values back to their original range: $q_a^{-1}(c) = ac$. Note that quantization is not invertible, and that dequantization is not the inverse function of quantization. Double quantization that results from double compression is given by: $q_{ab}(c) = \lfloor \lfloor c/b \rfloor b/a \rfloor$, where a and b are the quantization steps. Double quantization can be represented as a sequence of three steps: 1) quantization with step b, followed by 2) dequantization with step b, followed by 3) quantization with step a. Now consider a set of coefficients normally distributed in the range $[0, 127]$. To illustrate the nature of the double quantization artifacts, consider four different quantizations of these coefficients. Shown in the top row of Figure 1 are the histograms of the coefficients quantized with steps 2 and 3. Shown in the bottom row are the histograms of the coefficients double-quantized with steps 3 followed by 2 and 2 followed by 3. When the step size decreases [Figure 1(c)], some bins in the histogram are empty, because the first quantization places the samples of the original signal into 42 bins, while the second quantization redistributes them into 64 bins. When the step size increases [Figure 1(d)], some bins contain more samples than their neighboring bins, because the even bins receive samples from four original histogram bins while the odd bins receive samples from only two. In both cases of double quantization, note the periodicity of the artifacts introduced into the histograms. It is this periodicity that the authors in [35] exploited to detect double JPEG compression. The work of [13] extended this approach to detect localized traces of double compression.

JPEG BLOCKING

As described in the previous sections, the basis for JPEG compression is the block DCT transform. Because each 8×8 pixel image block is individually transformed and quantized, artifacts appear at the border of neighboring blocks in the form of horizontal and vertical edges. When an image is manipulated, these blocking artifacts may be disturbed.

In [28], the authors characterize the blocking artifacts using pixel value differences within and across block boundaries. These differences tend to be smaller within blocks than across blocks. When an image is cropped and recompressed, a new set of blocking artifacts may be introduced that do not necessarily align with the original boundaries. Within- and across-block pixel value differences are computed from 4-pixel neighborhoods that are spatially offset from each other by a fixed amount, where one neighborhood lies entirely within a JPEG block and the other borders or overlaps a JPEG block. A histogram of these differences is computed from all 8×8 non-overlapping image blocks. A 8×8 "blocking artifact" matrix (BAM) is computed as the average difference between these histograms. For uncompressed images, this matrix is random, while for a compressed image, this matrix has a specific pattern. When an image is cropped and recompressed, this pattern is disrupted. Supervised pattern classification is employed to discriminate between authentic and inauthentic BAMs.

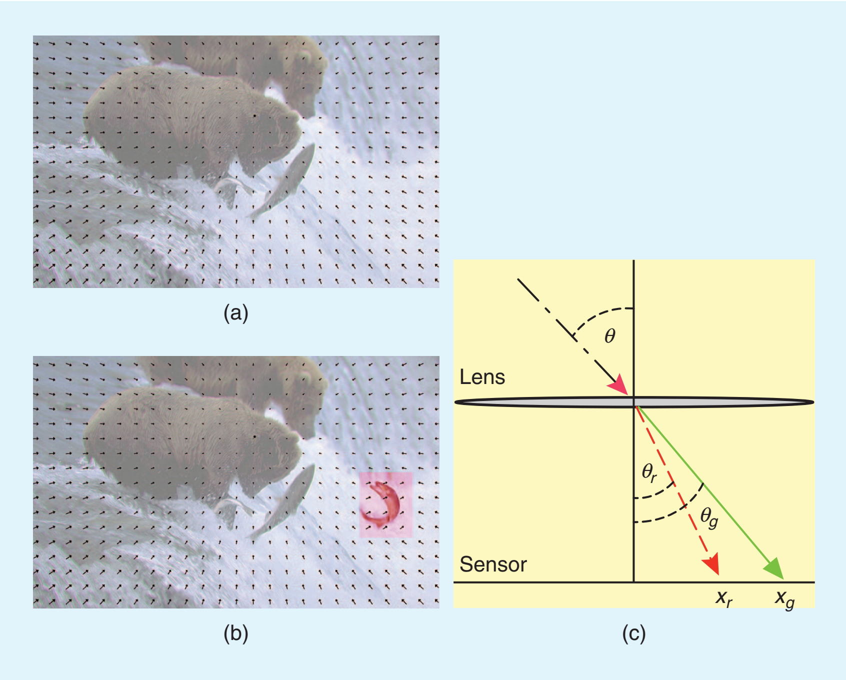
In [41], the authors describe how to detect more localized manipulations from inconsistencies in blocking artifacts. From a region of the image which is presumed to be authentic, the level of quantization is first estimated for each of 64 DCT frequencies. Inconsistencies between the DCT coefficients D and the estimated amount of quantization Q are computed as

$$B = \sum_{k=1}^{64} \left| D(k) - Q(k)\mathrm{round}\left(\frac{D(k)}{Q(k)}\right) \right|. \tag{4}$$

Variations in B across the image are used to detect manipulated regions.

CAMERA BASED

Grooves made in gun barrels impart a spin to the projectile for increased accuracy and range. These grooves introduce somewhat distinct markings to the bullet fired, and can therefore be used to link a bullet with a specific handgun. In



[FIG2] Chromatic aberration: (a) Superimposed on the original image is a vector field showing the pixel displacement between the red and green channels. (b) The fish, taken from another image, was added to this image. Its chromatic aberration is inconsistent with the global pattern. (c) Polychromatic light enters the lens at an angle θ and emerges at an angle that depends on wavelength. As a result, different wavelengths of light, two of which are represented as the red (dashed) and the green (solid) rays, will be imaged at different points, x_r and x_g, giving rise to chromatic aberration.

the same spirit, several image forensic techniques have been developed that specifically model artifacts introduced by various stages of the imaging process. I describe four techniques for modeling and estimating different camera artifacts. Inconsistencies in these artifacts can then be used as evidence of tampering.

CHROMATIC ABERRATION

In an ideal imaging system, light passes through the lens and is focused to a single point on the sensor. Optical systems, however, deviate from such ideal models in that they fail to perfectly focus light of all wavelengths. Specifically, lateral chromatic aberration manifests itself as a spatial shift in the locations where light of different wavelengths reaches the sensor. In [16], the authors show that this lateral aberration can be approximated as an expansion or contraction of the color channels with respect to one another. Shown in Figure 2(a), for example, is an image overlayed with a vector field that represents the misalignment of the red channel relative to the green channel. Shown in Figure 2(b) is this same image when the fish was added to the image. Note that in this case, the local lateral aberration in this tampered region is inconsistent with the global aberration. The authors of [16] describe how to estimate lateral chromatic aberration in order to detect this type of manipulation.

In classical optics, the refraction of light at the boundary between two media is described by Snell's law: $n\sin(\theta) = n_f \sin(\theta_f)$, where θ is the angle of incidence, θ_f is the angle of refraction, and n and n_f are the refractive indices of the media through which the light passes. The refractive index of glass, n_f, depends on the wavelength of the light that traverses it. This dependency results in polychromatic light being split according to wavelength as it exits the lens and strikes the sensor. Shown in Figure 2(c), for example, is a schematic showing the splitting of short-wavelength light (solid green ray) and long-wavelength light (dashed red ray). Denote the position of the red and green rays on the sensor as (x_r, y_r) and (x_g, y_g). In the presence of chromatic aberration, these positions can be modeled as

$$x_r = \alpha(x_g - x_0) + x_0$$
$$\text{and} \quad y_r = \alpha(y_g - y_0) + y_0, \qquad (5)$$

where α is a scalar value and (x_0, y_0) is the center of the distortion. The estimation of these model parameters is framed as an image registration problem. Since the aberration results in a misalignment between the color channels, the model parameters are estimated by maximizing the alignment of the color channels. Specifically, the mutual information between the red and green color channels is maximized (a similar estimation is performed to determine the distortion between the blue and green channels). Local estimates of the chromatic aberration are then compared to the estimated global aberration to detect tampering.

COLOR FILTER ARRAY

A digital color image consists of three channels containing samples from different bands of the color spectrum, e.g., red, green, and blue. Most digital cameras, however, are equipped with a single CCD or CMOS sensor and capture color images using a color filter array (CFA). Most CFAs employ three color filters (red, green, and blue) placed atop each sensor element. Since only a single color sample is recorded at each pixel location, the other two color samples must be estimated from the neighboring samples in order to obtain a three-channel color image. The estimation of the missing color samples is referred to as CFA interpolation or demosaicking. The simplest demosaicking methods are kernel-based ones that act on each channel independently (e.g., bilinear or bicubic interpolation). More sophisticated algorithms interpolate edges differently from uniform areas to avoid blurring salient image features. Regardless of the specific implementation, CFA interpolation introduces specific statistical correlations between a subset of pixels in each color channel. Since the color filters in a CFA are typically arranged in a periodic pattern, these correlations are periodic.

At the same time, it is unlikely that the original recorded pixels will exhibit the same periodic correlations. As such these correlations can be used as a type of digital signature.

[OVER THE PAST FIVE YEARS, THE FIELD OF DIGITAL FORENSICS HAS EMERGED TO HELP RESTORE SOME TRUST TO DIGITAL IMAGES.]

If the specific form of the periodic correlations is known, then it would be straightforward to determine which pixels are correlated with their neighbors. On the other hand, if it is known which pixels are correlated with their neighbors, the specific form of the correlations can easily be determined. In practice, of course, neither is known. In [37], the authors describe how to simultaneously determine both the form of the correlations and which pixels are and are not CFA-interpolated (see also [3] and [4]). The authors employed the expectation/maximization (EM) algorithm. The EM algorithm is a two-step iterative algorithm: 1) in the expectation step the probability of each pixel being correlated with its neighbors is estimated; and 2) in the maximization step the specific form of the correlations among pixels is estimated. By modeling the CFA correlations with a simple linear model, the expectation step reduces to a Bayesian estimator and the maximization step reduces to weighted least squares estimation. In an authentic image, it is expected that a periodic pattern of pixels will be highly correlated with their neighbors; deviations from this pattern are therefore evidence of localized or global tampering.

CAMERA RESPONSE

Because most digital camera sensors are very nearly linear, there should be a linear relationship between the amount of light measured by each sensor element and the corresponding final pixel value. Most cameras, however, apply a pointwise nonlinearity in order to enhance the final image. The authors in [24] describe how to estimate this mapping, termed a response function, from a single image. Differences in the response function across the image are then used to detect tampering (a related approach is described in [14]).

Consider an edge where the pixels below the edge are of a constant color C_1 and the pixels above the edge are of a different color C_2. If the camera response is linear, then the intermediate pixels along the edge should be a linear combination of the neighboring colors. The deviation of these intermediate pixel values from this expected linear response is used to estimate the camera response function. The inverse camera response function that brings the pixel colors back to a linear relationship is estimated using a maximum a posteriori (MAP) estimator. In order to stabilize the estimator, edges are selected such that areas on either side of the edge are similar, the variances on either side of the edge are small, the difference between C_1 and C_2 is large, and the pixels along the edge are between C_1 and C_2. Constraints are also imposed on the estimated camera response function: the function should be monotonically increasing with at most one inflexion point and should be similar for each of the color channels. Since the camera response function can be estimated locally, significant variations in this function across the image can be used to detect tampering.

SENSOR NOISE

As a digital image moves from the camera sensor to the computer memory, it undergoes a series of processing steps, including quantization, white balancing, demosaicking, color correction, gamma correction, filtering and, usually, JPEG compression. This processing introduces a distinct signature into the image. The authors in [12] model this processing with a generic additive noise model and use statistics from the estimated noise for image forensics. In [40], the authors model camera processing with a series of in-camera processing operations and a second filtering. The parameters of this camera processing are then used to determine if an image has undergone any form of subsequent processing.

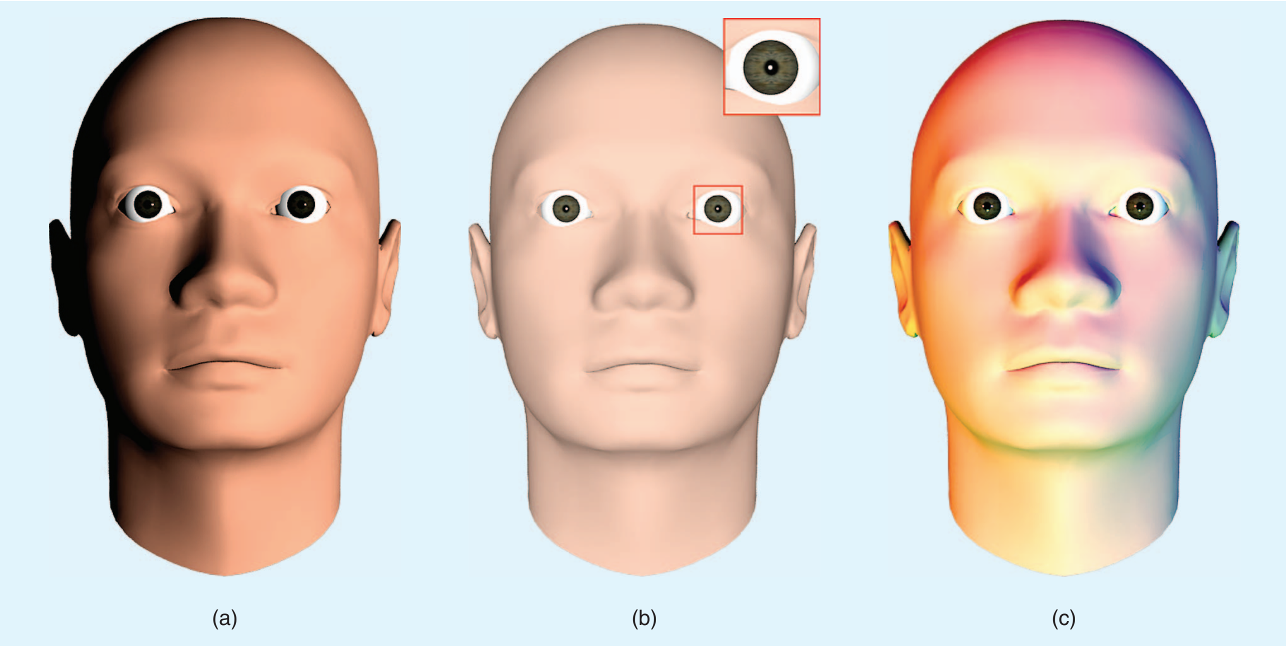
The authors in [10] model camera processing with an additive and multiplicative noise model. The parameters of the noise model are estimated from the original camera or a series of images originating from the known camera. Correlations between the estimated camera noise and the extracted image noise are then used to authenticate an image. The effect of the in-camera processing is modeled as follows:

$$I(x,y) = I_0(x,y) + \gamma I_0(x,y)K(x,y) + N(x,y), \qquad (6)$$

where $I_0(\cdot)$ is the noise-free image, γ is a multiplicative constant, $K(\cdot)$ is the multiplicative noise (termed photo-response nonuniformity noise (PRNU), and $N(\cdot)$ is an additive noise term. Since the PRNU varies across the image, it can be used to detect local or global inconsistencies in an image. The PRNU is estimated from a series of authentic images using wavelet-based noise removal techniques. As a general rule, at least 50 images are required to obtain accurate estimates of the PRNU. Fewer images could be used if they were generally low-pass in nature (e.g., images of the sky). Authentication is performed through a blockwise correlation between the estimated PRNU and an image whose authenticity has been called into question. The PRNU has also been shown to be distinct to a specific sensor [26]. As such, the PRNU can also be used for camera ballistics: identifying the specific camera from which an image originated.

PHYSICS BASED

Consider the creation of a forgery showing two movie stars, rumored to be romantically involved, walking down a sunset beach. Such an image might be created by splicing together individual images of each movie star. In so doing, it is often difficult to exactly match the lighting effects under which each person was originally photographed. I describe three techniques for estimating different properties of the lighting environment under which a person or object was photographed. Differences in lighting across an image can then be used as evidence of tampering.



[FIG3] The direction to a single light source can be determined from (a) the lighting gradient across the face or (b) the position of the specularity (white dot) on the eye. More complex lighting environments consisting of multiple colored lights (c) can be modeled as piecewise continuous functions on the sphere.

LIGHT DIRECTION (2-D)

Because the right side of the face in Figure 3(a) is more illuminated than the left, we can infer that a light source is positioned to the right. This observation can be formalized by making simplifying assumptions: the amount of light striking a surface is proportional to the surface normal and the direction to the light. With knowledge of three-dimensional (3-D) surface normals, the direction to the light source can therefore be estimated [33]. Because 3-D surface normals usually cannot be determined from a single image, the authors in [15] consider only the two-dimensional (2-D) surface normals at the occluding object boundary. In return, they estimate two of the three components of the light source direction. Although there remains an ambiguity in the estimated light direction, these two components of light direction are still useful in a forensic setting.

In order to simplify the estimation of light source direction, it is assumed that the surface of interest is Lambertian (the surface reflects light isotropically), has a constant reflectance value, and is illuminated by a point light source infinitely far away. Under these assumptions, the image intensity can be expressed as $I(x,y) = R(\vec{N}(x,y)\cdot\vec{L}) + A$, where R is the constant reflectance value, $\vec{L}$ is a 3-vector pointing in the direction of the light source, $\vec{N}(x,y)$ is a 3-vector representing the surface normal at the point (x,y), and A is a constant ambient light term. Since only the direction to the light source is of interest, the reflectance term, R, can be considered to have unit-value. At the occluding boundary of a surface, the z-component of the surface normal is zero. In addition, the x- and y-components of the surface normal can be estimated directly from the image. Under this added assumption, the image intensity is now given by

$$I(x,y) = \vec{N}(x,y)\cdot\vec{L} + A \;=\; \begin{pmatrix} N_x(x,y) & N_y(x,y)\end{pmatrix}\begin{pmatrix} L_x \\ L_y \end{pmatrix} + A. \tag{7}$$

Note that in this formulation, the z-component of both the surface normal and light direction is ignored since $N_z(x,y) = 0$. With at least three points with the same reflectance, R, and distinct surface normals, $\vec{N}$, the light source direction and ambient term can be solved for using standard least-squares estimation. A quadratic error function, embodying the imaging model of (7), is given by (8), shown at the bottom of the page.

This quadratic error function is minimized using standard least-squares estimation to yield $\vec{v} = (M^T M)^{-1} M^T \vec{b}$. This process can be repeated for different objects or people in the image to verify that the lighting is consistent.

$$E(\vec{L},A) = \left\| \begin{pmatrix} N_x(x_1,y_1) & N_y(x_1,y_1) & 1 \\ N_x(x_2,y_2) & N_y(x_2,y_2) & 1 \\ \vdots & \vdots & \vdots \\ N_x(x_p,y_p) & N_y(x_p,y_p) & 1 \end{pmatrix} \cdot \begin{pmatrix} L_x \\ L_y \\ A \end{pmatrix} - \begin{pmatrix} I(x_1,y_1) \\ I(x_2,y_2) \\ \vdots \\ I(x_p,y_p) \end{pmatrix} \right\|^2 = \| M\vec{v} - \vec{b} \|^2. \tag{8}$$

LIGHT DIRECTION (3-D)

The estimation of light source direction in the previous section was limited to 2-D because it is usually difficult to determine 3-D surface normals from a single image. In [20], the authors describe how to estimate the 3-D direction to a light source from the light's reflection in the human eye [see Figure 3(b)]. The required 3-D surface normals are determined by leveraging a 3-D model of the human eye.

Shown in Figure 4 is the basic imaging geometry where the reflection of the light is visible in the eye. This reflection is termed a specular highlight. In this diagram, the three vectors $\vec{L}$, $\vec{N}$, and $\vec{R}$ correspond to the direction to the light, the surface normal at the point at which the highlight is formed, and the direction in which the highlight will be seen. The law of reflection states that a light ray reflects off of a surface at an angle of reflection θ_r equal to the angle of incidence θ_i, where these angles are measured with respect to the surface normal $\vec{N}$. Assuming unit-length vectors, the direction of the reflected ray $\vec{R}$ can be described in terms of the light direction $\vec{L}$ and the surface normal $\vec{N}$:

$$\vec{R} \;=\; \vec{L} + 2(\cos(\theta_i)\vec{N} - \vec{L}) \;=\; 2\cos(\theta_i)\vec{N} - \vec{L}. \qquad (9)$$

By assuming a perfect reflector ($\vec{V} = \vec{R}$), the above constraint yields

$$\vec{L} \;=\; 2\cos(\theta_i)\vec{N} - \vec{V} \;=\; 2(\vec{V}^T\vec{N})\vec{N} - \vec{V}. \qquad (10)$$

The light direction $\vec{L}$ can therefore be estimated from the surface normal $\vec{N}$ and view direction $\vec{V}$ at a specular highlight. This estimated light direction can be compared across several people in an image or the estimated light direction using the technique described in the previous section.

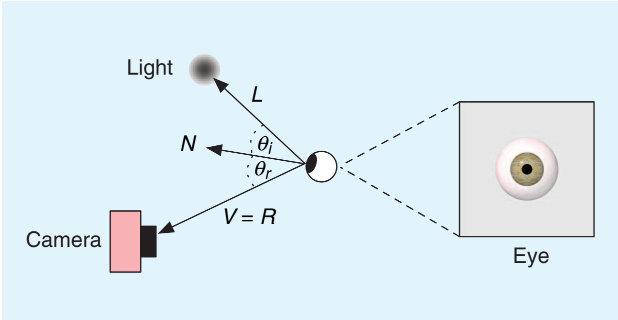
LIGHT ENVIRONMENT

In the previous two sections, a simplified lighting model consisting of a single dominant light source was assumed. In practice, however, the lighting of a scene can be complex: any number of lights can be placed in any number of positions, creating different lighting environments [see Figure 3(c)]. In [19], the authors describe how to estimate a low-parameter representation of such complex lighting environments.

The authors begin by leveraging the observation [39] that the appearance of a Lambertian surface can be well approximated by:

$$E(\vec{N}) \approx \sum_{n=0}^{2}\sum_{m=-n}^{n} \hat{r}_n l_{n,m} Y_{n,m}(\vec{N}), \qquad (11)$$

where $E(\vec{N})$ is the amount of lighting striking a surface (irradiance) with surface normal $\vec{N}$, $\hat{r}_n$ are known constants, $Y_{n,m}(\cdot)$ are the spherical harmonic functions, and $l_{n,m}$ are the unknown linear weights on these functions. Spherical harmonics form an orthonormal basis for piecewise continuous functions on the sphere and are analogous to the Fourier basis on the line or plane. Note that this expression is linear in the nine lighting



[FIG4] The formation of a specular highlight on an eye (small white dot on the iris). The position of the highlight is determined by the surface normal $\vec{N}$ and the relative directions to the light source $\vec{L}$ and viewer $\vec{V}$.

environment coefficients, $l_{0,0}$ to $l_{2,2}$ and can therefore be estimated using standard least-squares estimation. This solution, however, requires 3-D surface normals from at least nine points on the surface of an object. Without multiple images or known geometry, this requirement may be difficult to satisfy from an arbitrary image. The key observation in [19] is that by considering only the occluding boundary of an object, (11) simplifies to

$$\begin{aligned} E(\vec{N}) = {} & l_{1,-1}\frac{2\pi}{3}Y_{1,-1}(\vec{N}) + l_{1,1}\frac{2\pi}{3}Y_{1,1}(\vec{N}) \\ & + l_{2,-2}\frac{\pi}{4}Y_{2,-2}(\vec{N}) + l_{2,2}\frac{\pi}{4}Y_{2,2}(\vec{N}) \\ & + l_{0,0}\frac{\pi}{2\sqrt{\pi}} - l_{2,0}\frac{\pi}{16}\sqrt{\frac{5}{\pi}}, \end{aligned} \qquad (12)$$

where, most critically, the functions $Y_{i,j}(\cdot)$ depend only on the x and y components of the surface normal $\vec{N}$. That is, the five lighting coefficients can be estimated from only 2-D surface normals. In addition, (12) is still linear in its (now) five lighting environment coefficients, which can be estimated using standard least-squares estimation. The addition of a regularization term further improves the stability of the estimation. The estimated coefficients can be compared to detect lighting inconsistencies within an image.

GEOMETRIC-BASED

PRINCIPAL POINT

In authentic images, the principal point (the projection of the camera center onto the image plane) is near the center of the image. When a person or object is translated in the image, the principal point is moved proportionally. Differences in the estimated principal point across the image can therefore be used as evidence of tampering. In [18], the authors described how to estimate a camera's principal point from the image of a pair of eyes (i.e., two circles) or other planar geometric shapes. They showed how translation in the image plane is equivalent to a shift of the principal point. Inconsistencies in the principal point across an image can then be used as evidence of tampering.

[FIG5] (a) The original image. (b) A close-up of the license plate, which is largely illegible. (c) The result of planar rectification followed by histogram equalization.

The limbus, the boundary between the iris and the sclera, is well modeled with a circle. Now consider the projection of a pair of eyes (circles) that are assumed to be coplanar. In this case, the transformation from world to image coordinates can be modeled with a 3×3 planar projective transformation matrix H: $\vec{x} = H\vec{X}$, where the world points $\vec{X}$ and image points $\vec{x}$ are represented by 2-D homogeneous vectors. The transformation H can be estimated from the known geometry of a person's eyes and factored into a product of matrices that embody the camera's intrinsic and extrinsic parameters

$$H = \lambda \begin{pmatrix} f & 0 & c_1 \\ 0 & f & c_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \vec{r}_1 & \vec{r}_2 & \vec{t} \end{pmatrix}, \qquad (13)$$

where λ is a scale factor, the left-most matrix is the intrinsic matrix (where f is the focal length and (c_1, c_2) is the principal point), and the right-most matrix embodies the rigid-body transformation (rotation/translation) between world and camera coordinates. Once factored, the intrinsic matrix yields the desired estimate of the principal point. If in the creation of a composite of two or more people, a person was moved from his or her position in the original image, then the estimated principal points for each person will be inconsistent; this is evidence of tampering.

METRIC MEASUREMENTS

Shown in Figure 5 is an image of a license plate that is largely illegible. Also shown in this figure [Figure 5(c)] is the result of transforming the license plate as if it were viewed head-on. This rectified image clearly reveals the license plate number. In [17], the authors review several tools from projective geometry that allow for the rectification of planar surfaces and, under certain conditions, the ability to make real-world measurements from a planar surface. Three techniques for the rectification of planar surfaces imaged under perspective projection are described. Each method requires only a single image. The first method exploits knowledge of polygons of known shape (e.g., street sign, license plate, lettering on a billboard). The second method requires knowledge of two or more vanishing points on a plane and, for example, a pair of known angles on the plane. The third method requires two or more coplanar circles (e.g., car wheels). In each case, the world-to-image transformation is estimated, thereby allowing for the removal of planar distortions and metric measurements to be made on the plane.

THE FUTURE

Today's technology allows digital media to be altered and manipulated in ways that were simply impossible 20 years ago. Tomorrow's technology will almost certainly allow us to manipulate digital media in ways that today seem unimaginable. And as this technology continues to evolve, it will become increasingly important for the science of digital forensics to try to keep pace.

There is little doubt that as we continue to develop techniques for exposing photographic frauds, new techniques will be developed to make better fakes that are harder to detect. While some of the forensic tools may be easier to fool than others, some tools will be difficult for the average user to circumvent. For example, once disturbed, the color filter array interpolation can be regenerated by simply placing an image onto its original lattice and reinterpolating each color channel. On the other hand, correcting for inconsistent lighting is nontrivial in a standard photo-editing software program. As with the spam/antispam and virus/antivirus game, an arms race between the forger and forensic analyst is somewhat inevitable. The field of image forensics, however, has made and will continue to make it harder and more time-consuming (but never impossible) to create a forgery that cannot be detected.

ACKNOWLEDGMENTS

This work was supported by a gift from Adobe Systems, a gift from Microsoft, a grant from the National Science Foundation (CNS-0708209), and by the Institute for Security Technology Studies at Dartmouth College under grants from the Bureau of Justice Assistance (2005-DD-BX-1091) and the U.S. Department of Homeland Security (2006-CS-001-000001). Points of view or opinions in this document are those of the author and do not represent the official position or policies of the U.S. Department of Justice, the U.S. Department of Homeland Security, or any other sponsor.

AUTHOR

Hany Farid (farid@cs.dartmouth.edu) received his undergraduate degree in computer science and applied mathematics from the University of Rochester in 1989. He received his Ph.D. in computer science from the University of Pennsylvania in 1997. Following a two year post-doctoral position in brain and cognitive sciences at Massachusetts Institute of Technology, he joined the faculty at Dartmouth in 1999. Hany is the William H. Neukom Distinguished Professor of Computational Science, and the director of the Neukom Institute for Computational Science. He is the recipient of an NSF CAREER Award, a Sloan Fellowship, and a Guggenheim Fellowship.

REFERENCES

[1] S. Bayram, I. Avcibas, B. Sankur, and N. Memon, "Image manipulation detection with binary similarity measures," in *Proc. European Signal Processing Conf.*, Turkey, 2005.

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

[3] S. Bayram, H. T. Sencar, and N. Memon, "Source camera identification based on CFA interpolation," in *IEEE Int. Conf. Image Processing*, Genova, Italy, 2005, pp. 69–72.

[4] S. Bayram, H. T. Sencar, and N. Memon, "Improvements on source camera model identification based on CFA interpolation," in *Proc. IFIP WG 11.9 Int. Conf. Digital Forensics*, Orlando, FL, 2006, pp. 24–27

[5] I. J. Cox, M. L. Miller, and J. A. Bloom, Digital Watermarking. San Francisco, CA: Morgan Kaufmann, 2002.

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

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

[8] H. Farid, "Digital image ballistics from JPEG quantization," Dept. Comput. Sci., Dartmouth College, Tech. Rep. TR2006-583, 2006.

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

[10] J. Fridrich, M. Chen, and M. Goljan, "Imaging sensor noise as digital x-ray for revealing forgeries," in *Proc. 9th Int. Workshop on Information Hiding,* Sant Malo, France, 2007, pp. 342–358.

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

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

[13] J. He, Z. Lin, L. Wang, and X. Tang, "Detecting doctored JPEG images via DCT coefficient analysis," in *Proc. European Conf. Computer Vision*, Graz, Austria, 2006, pp. 423–435.

[14] Y.-F. Hsu and S.-F. Chang, "Image splicing detection using camera response function consistency and automatic segmentation," in *Proc. Int. Conf. Multimedia and Expo*, Beijing, China, 2007.

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

[16] 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.

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

[18] M. K. Johnson and H. Farid, "Detecting photographic composites of people," in *Proc. 6th Int. Workshop on Digital Watermarking*, Guangzhou, China, 2007.

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

[20] M. K. Johnson and H. Farid, "Exposing digital forgeries through specular highlights on the eye," in *Proc. 9th Int. Workshop on Information Hiding*, Saint Malo, France, 2007, pp. 311–325.

[21] S. Katzenbeisser and F. A. P. Petitcolas, Information Techniques for Steganography and Digital Watermarking. Norwood, MA: Artec House, 2000.

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

[23] G. Li, Q. Wu, D. Tu, and S. Sun, "A sorted neighborhood approach for detecting duplicated regions in image forgeries based on DWT and SVD," in *IEEE Int. Conf. Multimedia and Expo*, Beijing, China, 2007, pp. 1750–1753.

[24] Z. Lin, R. Wang, X. Tang, and H-V Shum, "Detecting doctored images using camera response normality and consistency," in *Proc. Computer Vision and Pattern Recognition*, San Diego, CA, 2005.

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

[26] J. Lukás, J. Fridrich, and M. Goljan, "Digital camera identification from sensor noise," *IEEE Trans. Inform. Forensics Security*, vol. 1, no. 2, pp. 205–214, 2006.

[27] W. Luo, J. Huang, and G. Qiu, "Robust detection of region-duplication forgery in digital images," in *Proc. Int. Conf. on Pattern Recognition*, Washington, D.C., 2006, pp. 746–749.

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

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

[30] S. Lyu and H. Farid, "Steganalysis using higher-order image statistics," *IEEE Trans. Inform. Forensics Security*, vol. 1, no. 1, pp. 111–119, 2006.

[31] B. Mahdian and S. Saic, "Blind authentication using periodic properties of interpolation," *IEEE Trans. Inform. Forensics Security*, vol. 3, no. 3, pp. 529–538, 2008.

[32] T.-T. Ng and S.-F. Chang, "A model for image splicing," in *Proc. IEEE Int. Conf. Image Processing*, Singapore, 2004, vol. 2, pp. 1169–1172.

[33] P. Nillius and J.-O. Eklundh, "Automatic estimation of the projected light source direction," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2001, pp. 1076–1083.

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

[35] A. C. Popescu and H. Farid, "Statistical tools for digital forensics," in *Proc. 6th Int. Workshop on Information Hiding*, Toronto, Canada, 2004, pp. 128–147.

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

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

[38] S. Prasad and K. R. Ramakrishnan, "On resampling detection and its application to image tampering," in *Proc. IEEE Int. Conf. Multimedia and Exposition*, Toronto, Canada, 2006, pp. 1325–1328.

[39] R. Ramamoorthi and P. Hanrahan, "On the relationship between radiance and irradiance: Determining the illumination from images of a convex Lambertian object," *J. Opt. Soc. Amer. A*, vol. 18, no. 10, pp. 2448–2559, 2001.

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

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

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

[43] A.C. Gallagher, "Detection of linear and cubic interpolation in jpeg compressed images," in *Proc. 2nd Canadian Conf. Computer and Robot Vision.*, Victoria, British Columbia, Canada, vol. 171, 2005, pp. 65–72.

[44] H. Farid, "Digital ballistics from jpeg quantization: A followup study," Dept. Comp. Sci., Dartmouth College, Tech. Rep. TR2008-638, 2008. [SP]