Ashwin Swaminathan, Min Wu, and K.J. Ray Liu

Component Forensics

Theory, methodologies, and applications

Visual sensor technologies have experienced tremendous growth in recent decades, and digital devices are becoming ubiquitous. Digital images taken by various imaging devices have been used in a growing number of applications, from military and reconnaissance to medical diagnosis and consumer photography. Consequently, a series of new forensic issues arise amidst such rapid advancement and widespread adoption of imaging technologies. For example, one can readily ask what kinds of hardware and software components as well as their parameters have been employed inside these devices? Given a digital image, which imaging sensor or which brand of sensor was used to acquire the image? How was the image acquired? Was it captured using a digital camera, cell phone camera, image scanner, or was it created artificially using an image-editing software? Has the image undergone any manipulation after capture? Is it authentic, or has it been tampered in any way? Does it contain any hidden information or steganographic data? Many of these forensic questions are related to tracing the origin of the digital image to its creation process. Evidence obtained from such analysis would provide useful forensic information to law enforcement, security, and intelligence agencies. Knowledge of image acquisition techniques can also help answer further forensic questions regarding the nature of additional processing that the image might have undergone after capture.

There are various ways to address the questions at hand. In this article, we survey one of the major class of techniques based on *component forensics* that aims to answer these forensic questions. Component forensics aims at identifying the algorithms and parameters employed in the various components of the device that was used in capturing the data. Component forensic analysis works by finding the *intrinsic fingerprint* traces that are left behind in a digital image when it goes through various processing blocks in the information processing chain, and uses such traces for estimating component parameters.

Digital Object Identifier 10.1109/MSP.2008.931076

1053-5888/09/$25.00©2009IEEE

Building upon component parameter estimation via intrinsic fingerprint identification, component forensics provides a framework to address a number of larger forensic issues such as in discovering device-technology infringement, protecting intellectual property rights, and identifying acquisition devices.

Protecting the intellectual property rights of imaging devices has been a primary concern in the recent times, and fierce competition in the electronic imaging industry has led to an increasing number of infringement cases filed in courts. The remunerations awarded to successful prosecution have also grown tremendously, sometimes in billions of dollars. Patents have been known as powerful tools for intellectual property protection. However, with the development of modern sophisticated tools, patent infringement of imaging devices has become easy to perform, difficult to detect, and even harder to prove in the court of law. A common way to perform infringement analysis is to examine the design and implementation of a product and to look for similarities with what have been claimed in existing patents through some type of reverse engineering. However, this approach could be very cumbersome and ineffective, and in several cases might involve a line-by-line comparison of low-level assembly language codes that go into the software module of the device. Component forensics provides a systematic methodology for infringement/licensing forensics by identifying the algorithms and parameters employed in each component of the information processing chain; thus protecting intellectual property rights.

Component forensics also serves as a foundation to establish the trustworthiness of images and imaging devices. With the fast development of tools to manipulate multimedia data, the integrity of both content and acquisition device has become particularly important when images are used as critical evidence in journalism, reconnaissance, and law enforcement applications. For example, information about hardware/software modules and their parameters in a device can help in building device identification systems. Such systems would provide useful acquisition forensic information to law enforcement and intelligence agencies about which device or which brand/model of the device was used to acquire an image. Additionally, component forensics helps establish a solid model for defining the characteristics of images obtained directly from a device, in turn facilitating tampering forensics to determine if there has been any additional editing and processing applied to an image since it leaves the device.

Component forensics can be performed in three major types of scenarios depending on the nature of available inputs. In intrusive forensics, the forensic analyst has access to the device; he/she can then break it apart, isolate each component, and present methods to compute the individual component parameters. In the case of semi nonintrusive forensics, the analyst still has access to the device but is not allowed to break it apart; he/she can then design appropriate inputs to be fed into the device so as to collect forensic evidence about the processing techniques and parameters of the individual components. In completely nonintrusive forensics, the forensic analyst estimates the component parameters just based on the sample data available from the device.

In this article, we will use visual sensors and images captured by digital cameras to demonstrate component forensics, while these techniques can be appropriately modified and extended to other types of acquisition models, and sensing technologies. We review methods to estimate the parameters of various camera components. We show that the computed parameters can be employed to assess the similarity in camera technologies for providing tell-tale clues on infringement/licensing, to identify the type of camera and the brand/model that was used to capture an image under question, and to a build ground-truth model to assist the detection of content manipulations.

SYSTEM MODEL FOR DIGITAL IMAGING DEVICES

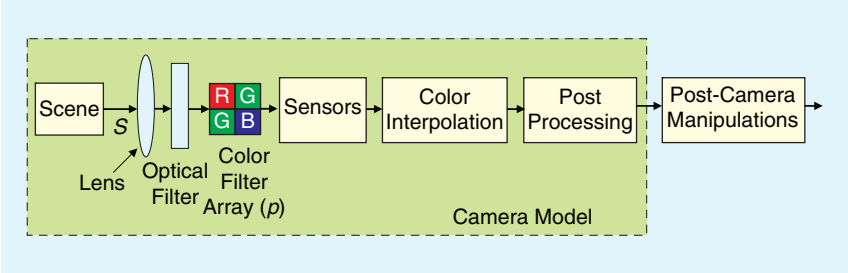
In this section, we review the image capture model in digital cameras to examine the various components in its information processing chain. As illustrated by the image capturing model in Figure 1, light from a scene passes through a lens and optical filters, and is finally recorded by an array of sensors. Most cameras employ a color filter array (CFA) to capture the information from the real-world scene. A CFA is a thin film on the sensors that selectively allows a certain component of light to pass through them to the sensors [1]. To facilitate discussions, let S be the real-world scene to be captured by the camera and let p be the CFA pattern matrix. $S(x, y, c)$ represents a three-dimensional (3-D) array of pixel values of size $H \times W \times C$, where H and W represent the height and the width of the image, respectively, and $C = 3$ denotes the number of color components (red, green, and blue). The CFA sampling converts the real-world scene S into a 3-D matrix S_p of the form

$$S_p(x, y, c) = \begin{cases} S(x, y, c) & \text{if } p(x, y) = c, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

After the data obtained from the CFA is recorded, the intermediate pixel values corresponding to the points where $S_p(x, y, c) = 0$ in (1) are interpolated using its neighboring pixel values to obtain $S_p^{(I)}$, by a processing operation popularly known as color interpolation or demosaicking [2]. After interpolation, the three images corresponding to the red, green, and the blue components go through a post-processing stage. In this stage, various types of in-camera processing operations such as white balancing, color correction, color matrixing, gamma correction, bit-depth reduction, and compression may be performed to enhance the overall picture quality and/or to reduce storage space and the final camera output S_d is created. S_d may then undergo additional processing operations using software such as Adobe Photoshop and Google Picasa to further improve the picture quality and/or tamper with the image. In our system model, we represent such post-camera processing as an additional manipulation block as shown in Figure 1.

COMPONENT FORENSICS METHODOLOGIES

As can been seen in the system model discussed in the previous section, when a real-world scene is captured using a digital



[FIG1] **Information processing in digital cameras.**

camera, the information about the scene passes through the various device components before the final digital image is produced. Each of these components in the information processing chain modify the input via a particular algorithm using a specific set of parameters, and leave some intrinsic fingerprint traces on the output. In the following section, we present techniques to nonintrusively estimate the parameters of the various in-camera components using the intrinsic fingerprint traces.

ESTIMATING CAMERA RESPONSE FUNCTION

The camera response function (CRF) maps the incident light energy to image intensity values [3]. Knowledge of the CRF is useful in several applications such as in computer vision algorithms for shape from shading and photometric stereo, and in authentication algorithms where the CRF can be used as a natural watermark. Estimating the CRF just based on a single camera output is an under-constrained problem, and therefore most prior work tries to estimate the CRF by assuming a particular nonlinear model. In [4], Farid assumes the CRF to be of the form $f(r) = r^{\gamma}$, where r and $f(r)$ represent the incident light energy and the image intensity values, respectively, and γ is the transformation parameter. Farid shows that the transformations of the form r^{γ} introduce correlations in the frequency-domain that can measured using bicoherence (third order statistic), without the knowledge of the imaging device [4]. While this bicoherence method can estimate the value of γ within an average accuracy of 7.5%, this approach is limited to the use of the γ-curve CRF model that is insufficient for real-world CRFs.

In [5], Lin and Zhang propose a method for CRF estimation from a single red, green, and blue (RGB)-color image by measuring the effects of a nonlinear response on the image via edge color distributions. The authors assume that edge pixels are linear-blended and introduce a method to compute the inverse radiometric response that maps the nonlinear distributions of the edge colors to linear distributions. Experimental results suggest that the average root-means square error (RMSE) of the estimated inverse response curves is around 10^{-2}. This approach has been further extended for gray scale images in [6] by using higher-order distribution features along image edges giving an RMSE close to 10^{-2} for two camera sets. Ng et al. develop a constraint equation to identify the potential locally planar irradiance points and then use these points for CRF estimation under the assumption that the CRF function follows a first-order generalized gamma curve model [3]. The authors show through simulations over five camera models that the estimation algorithm performs well and can be used to estimate the CRF with an average RMSE close to 10^{-2} using a single image; the RMSE can be further lowered with additional camera outputs.

ESTIMATING COLOR FILTER ARRAY AND COLOR INTERPOLATION PARAMETERS

For components such as CFA and color interpolation modules, the knowledge of the component output gives complete information about the input because the input and the output correspond to the sampled and the interpolated data, respectively. In [7], the Popescu and Farid employ expectation/maximization (EM) algorithms to estimate the color interpolation coefficients for forensic analysis. The authors first assume that the image pixels belong to one of the two models: 1) the pixel is linearly correlated to its neighbors and is obtained by a linear interpolation algorithm, or 2) the pixel is not correlated to its neighbors. Based on this assumption, the authors propose a two-step EM algorithm to estimate the CFA coefficients [7]. In the expectation step, the probability of each sample belonging to the two models is estimated, and the specific form of the correlations is found in the maximization step. The EM algorithm generates two outputs: a two-dimensional probability map indicating the likelihood of the pixel belonging to the two models and the weighting coefficients. The authors show through simulation results that while the estimated probability maps can be efficiently used to detect if a color image is the result of color interpolation or not, the color interpolation coefficients can help distinguish between different interpolation algorithms [7].

Swaminathan et al. develop an algorithm to jointly estimate the color filter array pattern and the color interpolation coefficients [8]. This method is schematically illustrated in Figure 2. A search space $\mathcal{P}$ for the CFA patterns is first established based on common practice in digital camera design and the observation that most commercial cameras use a RGB type of CFA with a fixed periodicity of 2×2. For every CFA pattern p in the search space $\mathcal{P}$, the interpolation coefficients are computed separately in different types of texture regions by fitting linear models. Specifically, the image is divided into three types of regions based on the gradient features in a local neighborhood, and the image pixel at location (x, y) is classified into one of the three categories: Region $\mathcal{R}_1$ contains those parts of the image with a significant horizontal gradient; Region $\mathcal{R}_2$ contains those parts of the image with a significant vertical gradient; and Region $\mathcal{R}_3$ includes the remaining parts of the image that primarily contains the smooth regions.

Using the final camera output S_d and the assumed sample pattern p, the set of pixel locations in S_d that are acquired directly from the sensor array and the set of pixel locations that are

interpolated are identified. A linear model is assumed for color interpolation in each of the three types of regions $\mathcal{R}_m$ $(m = 1, 2, 3)$ and three color channels (R, G, B) of the image, and the interpolated pixels are represented as a weighted average of the pixels assumed to be directly obtained from the sensor. The coefficient weights are then obtained by solving these equations. Let the set of N_e equations with N_u unknowns for a particular region and color channel be represented as $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ of dimension $N_e \times N_u$ and $\mathbf{b}$ of dimension $N_e \times 1$ specify the values of the pixels captured directly and those interpolated, respectively, and $\mathbf{x}$ of dimension $N_u \times 1$ stands for the interpolation coefficients to be estimated. To cope with possible noisy pixel values in $\mathbf{A}$ and $\mathbf{b}$ due to other in-camera operations following interpolation (such as JPEG compression), singular value decomposition [8], [9] is employed to estimate the interpolation coefficients. After the coefficients are obtained, they are used to reinterpolate the camera output, S_d, to obtain $\hat{S}_d^{(p)}$, and the error term $e^{(p)} = \hat{S}_d^{(p)} - S_d$ is computed. These steps are repeated for all the patterns, p, in the search space, $\mathcal{P}$, and the pattern that gives the lowest interpolation error is chosen as the estimate for the CFA. The interpolation coefficients corresponding to the estimated CFA pattern are also obtained in this process. Further details can be found in [8].

ESTIMATING POST-INTERPOLATION PROCESSING

Such processing operations as white balancing and color correction are performed by the camera after color interpolation to ensure that a white object in the scene appears white in a photograph. White balancing operations are typically multiplicative and each color in the photograph is multiplied by an appropriately chosen constant in the camera color space. Due to this multiplicative nature of white balancing operations, they cannot be nonintrusively estimated with good accuracies just based on a single camera output [10]. However, they can estimated semi nonintrusively by a two-step process by first obtaining two images under different built-in white balance settings and then by solving a set of equations, formulated using the Von-Kries hypothesis [10].

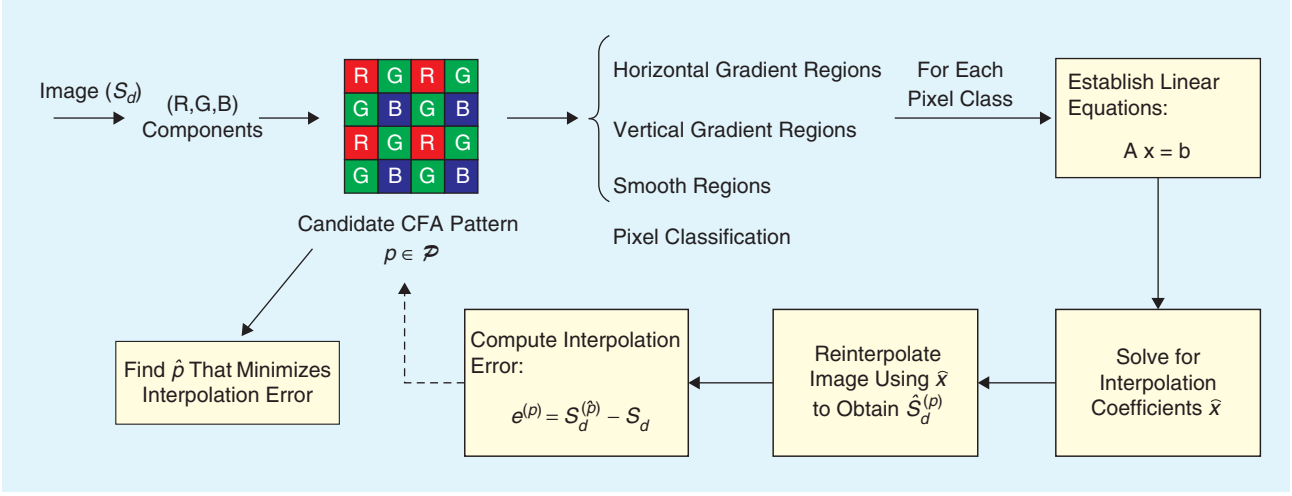
JPEG compression is another popular post-interpolation processing component in digital cameras. JPEG compression can be considered as quantization in the discrete cosine transform (DCT) domain. In this case, the knowledge of the component output does not give complete information about the corresponding input; but provides a rough estimate of its input within the range of the quantization step size. Based on this observation, statistical analysis based on binning techniques have been used to nonintrusively estimate the quantization matrices in literature [11], [12]. These algorithms have been shown to provide good accuracies in estimating the quantization step sizes in the low-low, low-high, and the high-low bands of images where there are a significant number of nonzero quantized values. In the case of high-high bands, a significant number of coefficients are quantized to zero resulting in larger estimation errors.

APPLICATIONS OF COMPONENT FORENSICS

We now look at several applications of component forensics to see how the intrinsic fingerprint traces left behind in the digital image provide tell-tale clues to help answer a number of questions related to the origin and authenticity of digital images.

CAMERA IDENTIFICATION FORENSICS

The estimated camera component parameters can be used as features for camera identification forensics to identify the camera brand and model utilized to capture the digital image. Bayram et al. developed a camera identification method [13] employing the weighting coefficients from the EM algorithm [7] and the peak location and magnitudes of the frequency spectrum of the probability map as features. Images captured from two cameras under controlled input conditions along with randomly acquired images from the internet for the third camera were used in the experiments, and the authors reported accuracies close to 84% on three brands [13] when 20% of the images were used in training and the remaining 80% employed in testing. Further improvements to this algorithm were made



[FIG2] Algorithm to estimate color filter array and color interpolation coefficients.

[TABLE 1] CAMERA MODELS USED IN THE EXPERIMENTS REPORTED BY [8].

NO.	CAMERA MODEL	NO.	CAMERA MODEL	NO.	CAMERA MODEL
1	CANON POWERSHOT A75	8	NIKON E5400	15	CASIO QV 2000UX
2	CANON POWERSHOT S400	9	SONY CYBERSHOT DSC P7	16	FUJIFILM FINEPIX S3000
3	CANON POWERSHOT S410	10	SONY CYBERSHOT DSC P72	17	FUJIFILM FINEPIX A500
4	CANON POWERSHOT S1 IS	11	OLYMPUS C3100Z/C3020Z	18	KODAK CX6330
5	CANON POWERSHOT G6	12	OLYMPUS C765UZ	19	EPSON PHOTOPC 650
6	CANON EOS DIGITAL REBEL	13	MINOLTA DIMAGE S304		
7	NIKON E4300	14	MINOLTA DIMAGE F100		

in [14] by separately considering smooth and non-smooth regions in the image to obtain accuracies close to 96% for three camera brands.

Swaminathan et al. proposed a joint CFA pattern and interpolation algorithm estimation technique for camera forensics and reported extensive camera identification results over a larger database of 19 different camera models in [8]. The list of camera models included in the experiments are shown in Table I. For each of the 19 camera models in the dataset, the authors collected about 200 different 512×512 images under uncontrolled conditions–different sceneries, different lighting situations, and compressed under different JPEG quality factors. These images in the database were grouped into different classes based on the brand or model, and the 7×7 filter coefficients estimated in each type of region and color channel (a total of 441 coefficients per image) were employed for camera identification. The authors showed through simulations that an average identification accuracy was around 90% for classifying the images from the nine different camera brands, and close to 86% for distinguishing between images from the 19 camera models from nine different brands [8]. Further, the authors demonstrated that these results were robust to post-interpolation processing operations inside the camera such as JPEG compression, additive noise, and nonlinear point operations like gamma correction [8].

Component forensics techniques, reviewed in this article, provide better accuracies over a larger and diverse dataset compared with other related work on camera identification [15]–[17]; they were later extended for cell phone cameras in [35] with 98% accuracy. Kharrazi et al. [15] proposed a set of 34 features for camera identification aiming to model the image-capture process in digital cameras. The set of features include: average pixel value, RGB pairs correlation, neighbor distribution center of mass, RGB energy ratio, wavelet domain statistics [18], and image quality metrics [19]. The authors employed support vector machines (SVM) for classification and reported accuracies close to 88% when tested with pictures captured under controlled input conditions from five camera models of three different brands. The same set of features were also tested for camera identification in [16] where accuracies close to 95% were obtained over four different camera models from two different models again under controlled input conditions. Another related work on camera identification estimated the pixel non-uniformity noise, which is a dominant component of the photo-response nonuniformity noise, inherent to an image sensor to distinguish between two cameras of the same brand, model, and set [17]. In the training phase of the algorithm, a wavelet based denoising algorithm is employed to obtain an estimate of the pixel non-uniformity noise and the random component of this noise is eliminated by averaging the estimates from a number of images. In the testing phase, to determine whether a given image is captured by a digital camera or not, the noise pattern from the image is obtained and correlated with the average noise pattern (also called the "reference pattern") of the given digital camera. A correlation value greater than the prechosen threshold suggested that the given image is from the digital camera. The authors showed that such an approach can identify the source camera with 100% accuracy when tested with high quality images. Dirik et al. presented a camera identification for digital single lens reflex (SLR) cameras by modeling the dust characteristics in [20], and reported accuracies close to 92% over three different camera brands. This work's focus is on extracting representative features for camera identification but does not explicitly estimate the various components of the information processing chain.

INFRINGEMENT AND LICENSING FORENSIC ANALYSIS

Component forensic analysis can be used for identifying the common features tied to imaging devices for applications in identifying infringement and licensing of device components. In [8], a classification based methodology was employed to study the similarities in interpolation algorithms used by different cameras. The authors first trained a classifier by omitting the data from one of the camera models and tested it with the coefficients of the omitted camera, to find the nearest neighbor in the color interpolation coefficient space. The results reported in the paper show that when the SVM is trained using all the 200 images from 18 cameras except Canon Powershot S410, and then tested using the 200 images from Canon Powershot S410, 66% of the Canon Powershot S410 images were classified as Canon Powershot S400. Furthermore, of the remaining images, 28% of the pictures were classified as one of the remaining Canon models; this suggests that there is a considerable amount of similarity in the kind of interpolation algorithms used by various models of the same brand. The results in [8] also indicated a similarity between Minolta DiMage S304 and Nikon E4300 as around 53% of the Minolta DiMage S304 pictures are designated as Nikon E4300 camera model in the classification tests. Building upon these results, a new metric is defined to study the similarities between two camera brands/models in [8]. Such an analysis has applications in identifying the similarities in the estimated camera component parameters to determine potential infringement or licensing.

DETECTING CUT-AND-PASTE FORGERIES BASED ON INCONSISTENCIES IN COMPONENT PARAMETERS

Creating a tampered image by cut-and-paste forgery often involves obtaining different parts of the image from pictures captured using different cameras that may employ a different

set of algorithms/parameters for its internal components. Inconsistencies in the estimated sensor pattern noise obtained from different regions of the image [17] or the inconsistencies in the estimated intrinsic fingerprint traces (such as color interpolation coefficients [21] or CRFs [22]) left behind by camera components can be used to identify such digital forgeries as cut-and-paste operations. In this survey article, we review the case study presented in [21] for illustration. In [21], the authors created a tampered picture of size 2048×2036 by combining parts of two images taken using two different cameras. Figure 3(a) and (b) shows the tampered picture and its individual parts marked with different colors. The regions displayed in white in Figure 3(b) are obtained from an image taken with the Canon Powershot S410 digital camera, and the black parts are cropped and pasted from a picture shot using the Sony Cybershot DSC P72 model. The combined image was then JPEG compressed with quality factor 80%.

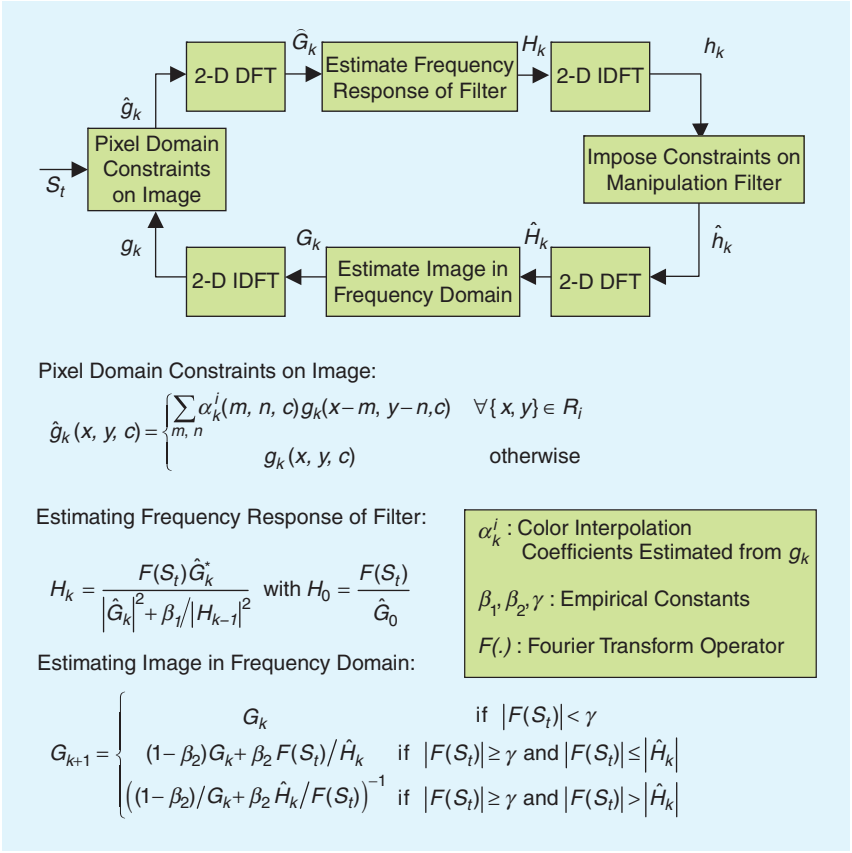
To identify the intrinsic camera fingerprints in different parts of the picture, the image is examined using a sliding window of 256×256 with step size 64×64, and the color interpolation coefficients are estimated in each 256×256 block [21]. The detection results from the 19-camera model classifier are shown in Figure 3(c). In this figure, the regions marked black denotes those classified as the Sony Cybershot DSC P72 model and the white areas correspond to the parts correctly classified as the Canon Poweshot S410 model. The remaining regions represented in grey correspond to the blocks that were misclassified as one of the remaining 17 camera models. As shown in Figure 3(c), the results indicate that the correct camera can be identified with a very high confidence in most of the regions in the tampered picture using the data obtained from each 256×256 macro-block. In this particular case, the manipulated picture has distinct traces from two different cameras and is therefore tampered.

COMPONENT FORENSICS AS GROUND-TRUTH MODEL FOR TAMPERING DETECTION

Post-camera processing operations include content-preserving and content-altering manipulations such as tampering. Post-camera processing are generally hard to detect and estimate due to the lack of knowledge about the manipulation type which leads to incorrect choice of model. To circumvent this problem, some prior work in the tampering detection literature try to detect tampering by defining the properties of a manipulated image in terms of the distortions it goes through; and using such analysis, present methods to both detect manipulated images and identify the manipulation type along with its parameters. In [18] and [23], features based on analysis of variance approaches [23] and higher order wavelet statistics [18] have been used to detect the presence/absence of image manipulations without focussing on identifying the manipulation type and/or its parameters. These methods require samples of tampered images (under each type of manipulation) for classification to distinguish such manipulated images from genuine camera-captured ones. Further, these methods may not be able to efficiently identify manipulation types that are not modeled or considered directly when building the classifier.

In [24], the authors extend component forensics approach to detect the presence/absence of image manipulations by modeling all post-camera processing as a manipulation block [24]. The algorithm works by first assuming that the given test image, S_t, is a manipulated camera output obtained by processing the actual camera output, S_d, via a manipulation operation. Any post-camera processing applied on S_d is then modeled as a linear shift-invariant filter and its coefficients are estimated by the iterative constraint enforcement algorithm [24]. Figure 4 shows the schematic diagram of the algorithm introduced in [24] to estimate the coefficients of the manipulation filter. The test image, S_t, is used to initialize the iterative process. In each iteration, the estimated camera output, g,and the estimated filter coefficients, h,are updated by repeatedly applying known constraints on the image and the filter, both in the pixel domain and in the Fourier domain [25], [26]. In the kth iteration, pixel-domain constraints are enforced on the image g_k to obtain $\hat{g}_k$; the pixel domain constraints represent the camera constraints where the camera component parameters α_k^i are estimated via component forensic techniques presented in the section "Estimating Color Filter Array and Color Interpolation Parameters." After the image $\hat{g}_k$ is obtained, it is transformed

[FIG3] Applications to source authentication showing (a) sample tampered image, (b) regions obtained from the two cameras, and (c) CFA interpolation identification results (black: Sony Cybershot DSC P72; white: Canon Powershot S410; grey: regions classified as other cameras).



[FIG4] Iterative constraint enforcement algorithm to estimate manipulation filter coefficients [24].

by discrete Fourier transform (DFT) to give $\hat{G}_k$. The frequency response of the estimated manipulation filter, H_k, is then obtained from $\hat{G}_k$ and the Fourier transform of the test image, $\mathcal{F}(S_t)$, as shown in Figure 4. The estimated response, H_k, thus obtained is inverse Fourier transformed to give h_k; and filter constraints are enforced on h_k to obtain $\hat{h}_k$ to be the real part of h_k. The value of G_{k+1} is finally obtained as a function of two available estimates: a) previous value, G_k, and b) $\mathcal{F}(S_t)/\hat{H}_k$, where $\hat{H}_k = \mathcal{F}(\hat{h}_k)$. Complete details of this algorithm along with its properties can be found in [24]. Deviation of the estimated manipulation filter parameters from an identity transform, measured via a similarity score, indicates that the test image has been manipulated after being captured by the camera.

In [24], the authors use data from nine different camera models (corresponding to camera model numbers 1-7, 10, and 16 in Table 1) for testing the iterative constraint enforcement algorithm. This gives a total of 900 different 512×512 pictures in the camera-image database with 100 image per camera model. These images are then processed to generate 21 tampered versions per image to obtain 18,900 manipulated images. The type of manipulations considered along with the parameter settings are listed in Table 2. For each direct camera output and its manipulated version, the frequency domain coefficients of the estimated manipulation filter, H_t, are computed and its similarity with the chosen reference pattern, H_{ref}, is determined using a similarity score. The reference pattern, H_{ref}, is obtained a priori in the training stage from an authentic camera output using the same iterative constraint enforcement algorithm, and helps compensate for the minor deviations due to post-interpolation processing inside cameras. To compute the similarity score, the logarithm of the magnitude of the frequency response of the test image $\Theta_t = \log_{10}(|H_t|)$ is obtained and its similarity with log-magnitude coefficients of the reference image is computed as

$$s(\Theta_t, \Theta_{\text{ref}}) = \sum_{m,n} (\Theta_t(m,n) - \mu_t) \times (\Theta_{\text{ref}}(m,n) - \mu_{\text{ref}}),$$

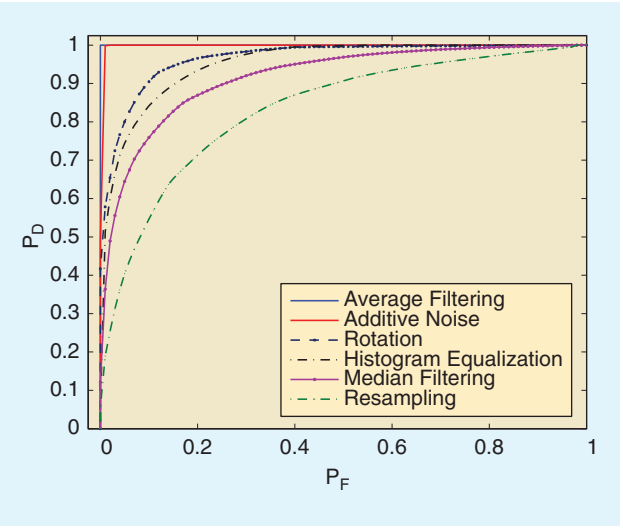
where μ_t denotes the pixel-wise mean of the Θ_t and μ_{ref} represents the pixel-wise mean of the Θ_{ref}. The test input is then classified as unmanipulated if the similarity to the reference pattern is greater than a suitably chosen threshold. If, on the other hand, the input image has undergone tampering or steganographic embedding operations, the estimated manipulation filter coefficients would include the effects of these manipulations and therefore be less similar to the reference pattern (obtained from an unmanipulated camera output); thus, resulting in a similarity score lower than the chosen threshold. The performance of the iterative constraint enforcement algorithm is shown in terms of the receiver operating characteristics (ROC) in Figure 5. The figure shows that for P_F close to 10%, P_D is close to 100% for such manipulations as spatial averaging and additive noise, and around 70%–80% for median filtering, histogram equalization, and rotation.

In [24], the authors showed that the estimated manipulation filter coefficients can also be employed to identify the

[TABLE 2] TAMPERING OPERATIONS INCLUDED IN THE EXPERIMENTS.

MANIPULATION OPERATION	PARAMETERS OF THE OPERATION	NUMBER OF IMAGES
SPATIAL AVERAGING	FILTER ORDERS 3-11 IN STEPS OF TWO	5
MEDIAN FILTERING	FILTER ORDERS {3, 5, 7}	3
ROTATION	DEGREES {5, 10, 15, 20}	4
RESAMPLING	SCALE FACTORS {0.5, 0.7, 0.85, 1.15, 1.3, 1.5}	6
ADDITIVE NOISE	PSNR 5 dB AND 10 dB	2
HISTOGRAM EQUALIZATION		1
TOTAL		21

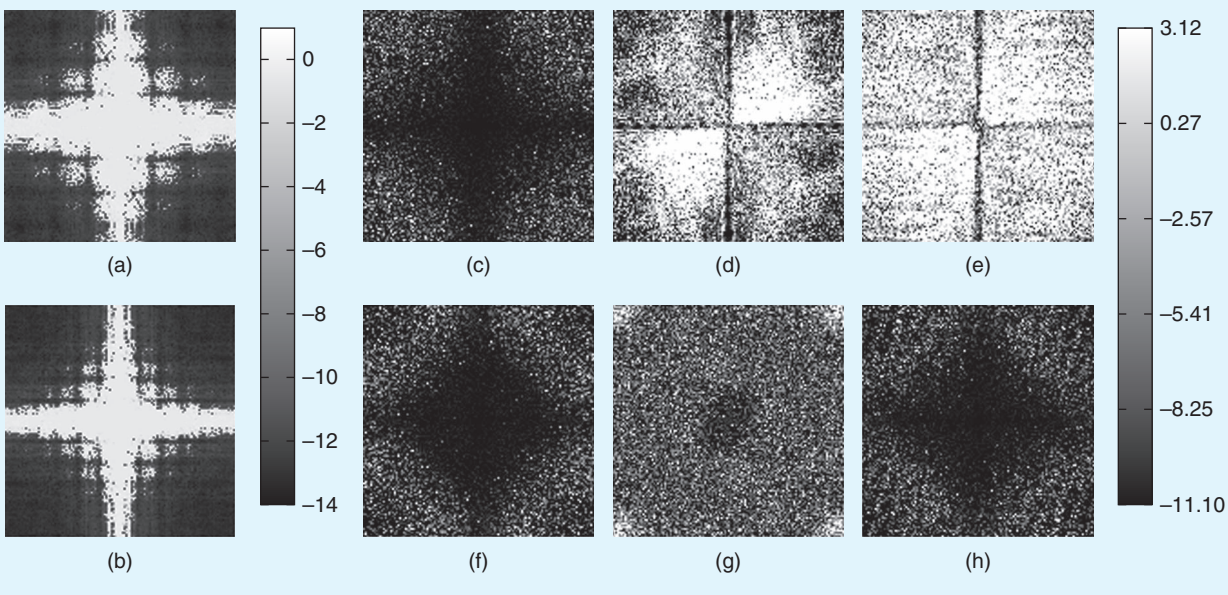
type and parameters of post-camera processing operations. Figure 6 shows the frequency response of the estimated manipulation filter coefficients for the different types of manipulations listed in Table 2. A closer look at the manipulation filter coefficients in the frequency domain suggests noticeable differences for the different kinds of tampering operations. For such manipulations as average filtering, distinct nulls in the frequency spectrum are observed and the gap between the nulls can be employed to estimate the order of the averaging filter and its parameters. Image manipulations such as additive noise result in a manipulation filter with a noisy spectrum as shown in Figure 6(g), and the strength of the noise can be computed from the filter coefficients. Rotation and downsampling can be identified from the smaller values in the low-high and the high-low bands of the frequency spectrum of the manipulation filter. Recently, Chuang et al. [27] built upon [24], and showed that many classes of linear shift invariant (LSI) and non-LSI image processing operations, such as resampling, JPEG compression, and nonlinear filtering, exhibit consistent and distinctive patterns in their empirical frequency response (EFR). The identification performance for manipulation types based on the EFR is around 93% for classifying six types of manipulation.



[FIG5] Receiver operating characteristics for tampering detection when tested with all images in the database with 200 images are used in training.

APPLICATIONS TO UNIVERSAL STEGANALYSIS AND IMAGE ACQUISITION FORENSICS

A common challenge for universal steganalysis and for image acquisition forensic analysis is how to model the ground truth original image data. Using a camera model and its component analysis, component forensics provides a framework to distinguish between camera-captured images and images with steganographic hidden information and images produced by other acquisition sources [24]. Image manipulations such as watermarking and steganography have been modeled as post-processing operations applied to camera outputs and the manipulation coefficients, estimated from the iterative constraint enforcement algorithm, have been employed to distinguish them from authentic data. Images produced by other types of acquisition sources would result in a manipulation filter that is significantly different from an ideal delta function expected for camera outputs and such images can be distinguished by comparing the coefficients of the manipulation filter with the reference pattern obtained from direct camera outputs. Further details can be found in [24].



[FIG6] Frequency response of the manipulation filter for camera outputs that are manipulated by (a) 7 X 7 averaging filter, (b) 11 X 11 averaging filter, (c) 7 X 7 median filter, (d) 20 degrees rotation, (e) 70% resampling, (f) 130% resampling, (g) noise addition with PSNR 20dB, and (h) histogram equalization. The frequency response is shown in the log scale and shifted so that the DC components are in the center.

THEORETICAL ANALYSIS OF COMPONENT FORENSICS

As shown in the previous sections, the intrinsic fingerprint traces left behind in the final digital image by the different components of the imaging device can be used as evidence to estimate the component parameters and provide clues to answer forensic questions related to the origin and authenticity of digital data. However, as the intrinsic fingerprint traces pass through the different parts of the information processing chain, some of them may be modified or destroyed and some others newly created. This observation leads to a number of fundamental questions as to what component traces are lost or modified? Which components in the information processing chain are identifiable and which ones are not? How does the identifiability of one component affect the estimation of the other? The theoretical framework for component forensics presented in [28] and [29] focuses on answering such questions and examines the conditions under which the parameters of a component can be identified or classified accurately. Next, we summarize the main results of this work.

THEORETICAL NOTIONS AND FRAMEWORK

In [28] and [29], the authors define a component as the basic unit of the information processing chain and represent a device to be a cascade of N_c components represented as $\{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_{N_c}\}$. The authors denote by θ^k, the set of parameters employed in kth component $\mathcal{C}_k$ of the system, and quantify the goal of component forensic analysis in terms of the identifiability or classifiability of the component parameter θ^k. The authors consider two possible scenarios for θ^k. In the first scenario, the authors assume that possible set of algorithm space for θ^k is known a priori, i.e., $\theta^k \in \Theta^k = \{\theta_1^k, \theta_2^k, \ldots\}$. In this case, the problem becomes a classification problem and pattern classification theory has been used to analyze this scenario [28]. In the second scenario, the authors assume that no knowledge about the possible algorithm space is known or assumed a priori, and present a framework based on estimation theory and Fisher information to analyze this case [29].

For the scenario where θ^k can take a finite number of possibilities such that $\theta^k \in \Theta^k$, the authors define a component $\mathcal{C}_k$ to be intrusively classifiable or i-classifiable if, for all inputs, the probability of classifying the component to employ the correct algorithm given the input and the output is greater or equal to the probability of classifying it to any other class. Furthermore, for at least one input x^* and its corresponding output, the probability of correct classification is strictly greater than the probability of misclassification [28], [30]. The goodness of the parameter identification algorithm under this scenario can be measured in terms of the confidence in making the right decision, and can be defined as the difference between the likelihood of the correct decision and the maximum of the corresponding likelihood of the making a wrong decision. Thus, the confidence score, $\gamma_i^k(x)$, for correctly deciding θ_i^k can be written as

$$\gamma_i^k(x) = f(\theta_i^k | y, x) - \max_{\theta \in \Theta^k \setminus \theta_i^k} f(\theta | y, x),$$

where $f(\theta | y, x)$ denotes the probability that the component employs the parameter θ conditional on the input x and its corresponding output y.

The confidence score $\gamma_i^k(x)$ is a function of the input x and can be improved by selecting proper inputs. For instance, consider an example of a component with parameters $\{\xi_0, \xi_1\}$ whose input-output relationship is given by

$$y(n) = \xi_0 x(n) + \xi_1 x(n-1).$$

Let $x^{(1)} = [\ \ldots, 1, 1, 1, \ldots\ ]$ and $x^{(2)} = [\ \ldots, 0, 1, 2, \ldots\ ]$ be two possible inputs to the system. The corresponding outputs would be $y^{(1)} = [\ \ldots, \xi_0 + \xi_1, \xi_0 + \xi_1, \xi_0 + \xi_1, \ldots\ ]$ and $y^{(2)} = [\ \ldots, -\xi_1, \xi_0, 2\xi_0 + \xi_1, \ldots\ ]$, respectively. Notice that $y^{(1)}$ is a constant sequence with each of its elements being equal to $(\xi_0 + \xi_1)$ and knowledge of the sum would not provide any indicative of the parameters ξ_0 or ξ_1. Therefore, $x^{(1)}$ is not a good input for evaluating the value of the component. On the other hand, observing the output $y^{(2)}$ of the system, one can formulate a system of linear equations to compute the value of ξ_0 and ξ_1; thus, $x^{(2)}$ is a good input to obtain the component parameter values. This example illustrates that the confidence score in parameter estimation can be improved by choice of inputs, and generalizing on this observation, [28] defines an optimal input as the one that maximizes the confidence score.

Notions of semi nonintrusively classifiable and completely nonintrusively classifiable components can be defined similarly [28], and these definitions help establish a number of theoretical results. For instance, the authors show that if a component is nonintrusively classifiable, then its parameters can also be identified semi nonintrusively; and if a system is semi nonintrusively classifiable, then each of its components are also intrusively classifiable. Moreover, the average confidence values obtained using semi nonintrusive analysis is greater than or equal to the ones obtained via completely nonintrusive analysis, and lower than the ones achieved via intrusive analysis. These results follow from the fact that semi nonintrusive forensics provides more control to the forensic analyst who can design better inputs to improve the overall performance, and intrusive analysis provides the highest control over the experimental setup. The theoretical results also prove that intrusive, semi nonintrusive, and completely nonintrusive forensics can provide the same confidence scores only when all the components of the system are consistent, meaning that the knowledge of the input provides full information about its output and vice versa [28].

In the second scenario, if no prior knowledge is available about the possible algorithm space, then the component forensics problem becomes an estimation problem and bias and variance of the estimates can be employed as metrics to theoretically analyze such components as discussed in [29]. For this case, it can be shown that the component parameter estimation errors obtained via semi nonintrusive analysis are lower than that obtained via completely nonintrusive analysis and greater than the ones for intrusive analysis, and these

analysis techniques become equivalent only for consistent components. The details of the definitions and theorems along with the sketches of proofs can be found in [29].

CASE STUDY WITH DIGITAL CAMERAS

We now examine a few case studies that demonstrate the usefulness of the theoretical framework.

- *Color Filter Array and Color Interpolation Modules*: In the absence of noise or additional processing, for such components as CFA and color interpolation modules, the knowledge of the component output gives complete information about the corresponding input as the input and the output correspond to the sampled and the interpolated data, respectively. Therefore, under this scenario, both the CFA and color interpolation modules are consistent components. From the theoretical analysis presented earlier, it can be shown that semi nonintrusive forensics would provide the same accuracies as completely nonintrusive forensics, i.e., even under controlled input conditions and well-designed inputs, the component estimation accuracies cannot be improved compared to nonintrusive analysis.

However, in the presence of additional post-interpolation processing operations, the components are not consistent, and semi nonintrusive forensics would provide better accuracies than completely nonintrusive forensics [28], [29]. For this case, good test conditions and heuristic patterns have been designed for semi nonintrusive forensics based on common knowledge about these components [10], and these have been further optimized in [28] and [29] to provide better confidence and accuracy in parameter estimation.

KNOWLEDGE OF IMAGE ACQUISITION TECHNIQUES CAN ALSO HELP ANSWER FURTHER FORENSIC QUESTIONS REGARDING THE NATURE OF ADDITIONAL PROCESSING THAT THE IMAGE MIGHT HAVE UNDERGONE AFTER CAPTURE.

- *Post-Interpolation Processing Modules*: Operations such as white balancing and color correction are typically multiplicative in nature. Due to this multiplicative nature, they are not nonintrusively classifiable [10] as given the output as a product of two terms, the individual terms cannot be separately resolved unambiguously. In such scenarios, the authors show that the knowledge of the camera input can help address this issue and semi nonintrusive analysis can be employed to estimate the parameters with very good accuracies [10].
- *Post-Camera Processing Modules*: Several post-camera processing modules can be similarly analyzed via the theoretical framework. In literature, methods have been proposed to nonintrusively estimate such post-processing operations as resampling [31], irregular noise patterns [32], luminance or lighting directions [33], chromatic aberration [34], nonlinear point operations, and gamma correction [4]. For instance, when the image is upsampled, some of the pixel values are directly obtained from the smaller version of the image, and the remaining pixels are interpolated and thus highly correlated with its neighbors. Thus, resampling parameters can be identified by studying the induced correlations for a certain range of resampling values [31]. Image manipulations such as contrast changes, gamma correction, and other image nonlinearities have been modeled and higher order statistics such as the bispectrum have been used to identify its component parameters [32]. Some of these methods assume prior knowledge about the possible algorithm space and may require an exhaustive search over all the possibilities.

Thus, the theoretical analysis framework surveyed in this section provides a systematic methodology to answer what components and processing operations in the information processing chain are identifiable and what are not, and helps quantify the estimation accuracies. These frameworks can also be extended to study the interactions between different components in a general information processing chain.

CONCLUSIONS

This article considers the problem of component forensics and presents a survey of existing literature related to multimedia component forensics of visual sensors. The survey is organized in three parts. In the first part, several methodologies for component forensics of digital cameras are discussed and methods to estimate such in-camera components as the camera response function, color filter array and color interpolation parameters, and post-interpolation processing operations such as white balancing and JPEG compression are reviewed. The second part demonstrates that the estimated parameters can be employed for a wide range of forensics applications including device brand and model identification, infringement/licensing forensic analysis, building ground-truth model to detect global and local tampering including identifying steganographic embedding, and for image acquisition forensics to distinguish between images taken from different acquisition sources. The theoretical analysis framework for component forensics is presented in the third part focused at gaining a concrete understanding about component forensics and to answer a number of fundamental questions related to what processing operations can and cannot be identified and under what conditions. In summary, we believe that such component forensic analysis would provide a great source of information for patent infringement cases, intellectual property rights management, and technology evolution studies for digital media and push the frontiers of multimedia forensics to gain a deeper understanding of information processing chain.

AUTHORS

Ashwin Swaminathan (sashwin@qualcomm.com) received the B.Tech degree in electrical engineering from the Indian Institute of Technology, Madras, India in 2003, and the Ph.D. degree in electrical and computer engineering from the University of

Maryland, College Park in 2008. He is currently a senior engineer at Qualcomm Incorporated in San Diego, California. He was a research intern with Hewlett-Packard Labs in 2006 and Microsoft Research in 2007. His research interests include multimedia forensics, information security, authentication, and information discovery. He was the winner of the Student Paper Contest at the 2005 IEEE International Conference on Acoustic, Speech and Signal Processing and received the ECE Distinguished Dissertation Fellowship Award in 2008. He is a Member of the IEEE.

Min Wu (minwu@eng.umd.edu) received the Ph.D. degree in electrical engineering from Princeton University in 2001. She is an associate professor at the University of Maryland, College Park. She leads the Media and Security Team at the University of Maryland, with main research interests on information security and forensics and multimedia signal processing. She is a corecipient of two Best Paper Awards from the IEEE Signal Processing Society and EURASIP, respectively. She also received a U.S. NSF CAREER award, a TR100 Young Innovator Award from the *MIT Technology Review Magazine*, a U.S. ONR Young Investigator Award, and a Computer World "40 Under 40" IT Innovator Award. She is currently the area editor of *IEEE Signal Processing Magazine* for its "Inside Signal Processing E-Newsletter" and is the associate editor of *IEEE Transactions on Information Forensics and Security*.

K.J. Ray Liu (kjrliu@eng.umd.edu) is a Distinguished Scholar-Teacher of University of Maryland, College Park, where he received university-level Invention of the Year Award and both Poole and Kent Senior Faculty Teaching Award and Outstanding Faculty Research Award from A. James Clark School of Engineering Faculty. He is the recipient of numerous best paper awards and was an IEEE Signal Processing Society Distinguished Lecturer. He was vice president-Publications, the editor-in-chief of *IEEE Signal Processing Magazine*, and the founding editor-in-chief of *EURASIP Journal on Applied Signal Processing*.

REFERENCES

[1] J. Adams, K. Parulski, and K. Spaulding, "Color processing in digital cameras," *IEEE Micro*, vol. 18, no. 6, pp. 20–30, Nov./Dec. 1998.

[2] J. E. Adams, "Interaction between color plane interpolation and other image processing functions in electronic photography," in *Proc. SPIE Cameras and Systems for Electronic Photography & Scientific Imaging*, San Jose, CA, Feb. 1995, vol. 2416, pp. 144–151.

[3] T.-T. Ng, S.-F. Chang, and M.-P. Tsui, "Using geometric invariants for camera response function estimation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Minneapolis, MN, June 2007, pp. 1–8.

[4] H. Farid, "Blind inverse gamma correction," *IEEE Trans. Image Processing*, vol. 10, no. 10, pp. 1428–1433, Oct. 2001.

[5] S. Lin, J. Gu, S. Yamazaki, and H.-Y. Shum, "Radiometric calibration from a single image," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Washington, D.C., June 2004, vol. 2, pp. 938–945.

[6] S. Lin and L. Zhang, "Determining the radiometric response function from a single grayscale image," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Diego, CA, June 2005, vol. 2, pp. 66–73.

[7] A. C. Popescu and H. Farid, "Exposing digital forgeries in color filter array interpolated images," *IEEE Trans. Signal Processing*, vol. 53, no. 10, part 2, pp. 3948–3959, Oct. 2005.

[8] A. Swaminathan, M. Wu, and K. J. R. Liu, "Non-intrusive component forensics of visual sensors using output images," *IEEE Trans. Inform. Forensics Sec.*, vol. 2, no. 1, pp. 91–106, Mar. 2007.

[9] C. F. van Loan, *Introduction to Scientific Computing: A Matrix-vector Approach Using MATLAB*. Englewood Cliffs, NJ: Prentice-Hall, 1999.

[10] A. Swaminathan, M. Wu, and K. J. R. Liu, "Optimization of input pattern for semi non-intrusive component forensics of digital cameras," in *Proc. IEEE Int. Conf. Acoustic, Speech, and Signal Processing*, Honolulu, HI, Apr. 2007, vol. 2, pp. 225–228.

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

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

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

[14] S. Bayram, H. T. Sencar, and N. Memon, "Improvements on source camera-model identification based on CFA interpolation," in *Proc. WG 11.9 Int. Conf. Digital Forensics*, Orlando, FL, Jan. 2006.

[15] M. Kharrazi, H. T. Sencar, and N. Memon, "Blind source camera identification," in *Proc. Int. Conf. Image Processing*, Singapore, Oct. 2004, vol. 1, pp. 709–712.

[16] M.-J. Tsai and G.-H. Wu, "Using image features to identify camera sources," in *Proc. IEEE Int. Conf. Acoustic, Speech, and Signal Processing*, Toulouse, France, May 2006, vol. 2, pp. 297–300.

[17] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, "Determining image origin and integrity using sensor pattern noise," *IEEE Trans. Inform. Forensics Sec.*, vol. 3, no. 1, pp. 74–90, Mar. 2008.

[18] H. Farid and S. Lyu, "Higher-order wavelet statistics and their application to digital forensics," in *IEEE Workshop on Statistical Analysis in Computer Vision*, Madison, WI, June 2003, vol. 8., pp. 94–101.

[19] I. Avcibas, B. Sankur, and K. Sayood, "Statistical evaluation of image quality metrics," *J. Electron. Imag.*, vol. 11, no. 2, pp. 206–223, Apr. 2002.

[20] E. A. Dirik, H. T. Sencar, and N. Memon, "Source camera identification based on sensor dust characteristics," in *Proc. IEEE Workshop Signal Processing Applications for Public Security and Forensics*, Brooklyn, NY, Apr. 2007, pp. 1–6.

[21] A. Swaminathan, M. Wu, and K. J. R. Liu, "Component forensics of digital cameras: A non-intrusive approach," in *Proc. Conf. Information Sciences and Systems*, Princeton, NJ, Mar. 2006, pp. 1194–1199.

[22] Y.-F. Hsu and S.-F. Chang, "Image splicing detection using camera response function consistency and automatic segmentation," in *Proc. IEEE Int. Conf. Multimedia and Expo*, Beijing, China, July 2007, pp. 28–31.

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

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

[25] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Mag.*, vol. 13, no. 3, pp. 43–64, May 1996.

[26] G. R. Ayers and J. C. Dainty, "Iterative blind deconvolution method and its applications," *Opt. Lett.*, vol. 13, no. 7, pp. 547–549, July 1988.

[27] W.-H. Chuang, A. Swaminathan, and M. Wu, "Tampering identification using empirical frequency response," in *Proc. IEEE Int. Conf. Acoustic, Speech, and Signal Processing*, Taipei, Taiwan, Apr. 2009.

[28] A. Swaminathan, M. Wu, and K. J. R. Liu, "A pattern classification framework for theoretical analysis of component forensics," in *Proc. IEEE Int. Conf. Acoustic, Speech, and Signal Processing*, Las Vegas, NV, Apr. 2008, pp. 1665–1668.

[29] A. Swaminathan, M. Wu, and K. J. R. Liu, "A component estimation framework for information forensics," in *IEEE Workshop on Multimedia Signal Processing*, Crete, Greece, Oct. 2007, pp. 397–400.

[30] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley-Interscience, 2000.

[31] 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.

[32] A. C. Popescu and H. Farid, "Statistical tools for digital forensics," in *Proc. 6th Int. Workshop Information Hiding & Lecture Notes in Computer Science*, Toronto, Canada, May 2004, vol. 3200, pp. 128–147.

[33] M. K. Johnson and H. Farid, "Exposing digital forgeries in complex lighting environments," *IEEE Trans. Inform. Forensics Sec.*, vol. 2, no. 3, part 1, pp. 450–461, Sept. 2007.

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

[35] C. E. McKay, A. Swaminathan, H. Gou, and M. Wu, "Image acquisition forensics: Forensic analysis to indentify imaging source," in *Proc. IEEE Conf. Acoustic, Speech, and Signal Processing*, Las Vegas, NV, Apr. 2008, pp. 1657–1660.

[SP]