Forensic Classification of Imaging Sensor Types

Nitin Khanna[a], Aravind K. Mikkilineni[b]
George T. C. Chiu[b], Jan P. Allebach[a], Edward J. Delp[a]

[a]School of Electrical and Computer Engineering
[b]School of Mechanical Engineering
Purdue University, West Lafayette, Indiana USA

ABSTRACT

Digital images can be captured or generated by a variety of sources including digital cameras and scanners. In many cases it is important to be able to determine the source of a digital image. Methods exist to authenticate images generated by digital cameras or scanners, however they rely on prior knowledge of the image source (camera or scanner). This paper presents methods for determining the class of the image source (camera or scanner). The method is based on using the differences in pattern noise correlations that exist between digital cameras and scanners. To improve the classification accuracy a feature vector based approach using an SVM classifier is used to classify the pattern noise.

Keywords: digital forensic, imaging sensor classification, flatbed scanner, sensor noise

1. INTRODUCTION

Advances in digital imaging technologies have led to the development of low-cost and high-resolution digital cameras and scanners. Both digital cameras and desktop scanners are becoming ubiquitous. Digital images produced by various sources are widely used in a number of applications from medical imaging and law enforcement to banking and daily consumer use. Forensic tools that help establish the origin, authenticity, and the chain of custody of digital images are essential to a forensic examiner[1]. These tools can prove to be vital whenever questions of digital image integrity are raised. Therefore, a reliable and objective way to examine digital image authenticity is needed.

There are various levels at which the image source identification problem can be solved. One may want to find the particular device (digital camera or scanner) which generated the image or one might be interested in knowing only the make and model of the device. As summarized in [2], a number of interesting and robust methods have been proposed for source camera identification [3–6].

One approach for digital camera identification is based on characterizing the imaging sensor used in the device. In [7], it is shown that defective pixels can be used for reliable camera identification even from lossy compressed images. In [6], an approach for camera identification using the imaging sensor's pattern noise was presented. The identification is based on pixel nonuniformity noise which is a unique stochastic characteristic for both CCD (Charged Coupled Device) and CMOS (Complementary Metal Oxide Semiconductor) sensors. Reliable identification is possible even from images that are resampled and JPEG compressed. The pattern noise is caused by several factors such as pixel non-uniformity, dust specks on the optics, optical interference, and dark currents[8]. The high frequency part of the pattern noise is estimated by subtracting a denoised version of the image from the original using a wavelet denoising filter [9]. A camera's reference pattern is determined by averaging the noise patterns from multiple images obtained from the camera. This reference pattern serves as an intrinsic signature of the camera. To identify the source camera, the noise pattern from an image is correlated with known reference patterns from a set of cameras and the camera corresponding to the reference pattern giving maximum correlation is chosen to be the source camera[6].

In[10] we extended the methods for source camera identification to scanners. A correlation based approach for authenticating digital cameras [6] was extended for source scanner identification. A SVM (Support Vector

Machine) classifier was also used to classify the images based on feature vectors obtained from the sensor pattern noise. We showed that this feature vector based approach gives much better classification accuracy than correlation based approaches.

The techniques used for both camera and scanner identification are dependent upon having prior knowledge of the class of devices (cameras or scanners) that the image was generated by. If the image was generated by a digital camera, then the digital camera identification methods must be used. Similarly if the image was generated by a scanner, the scanner identification methods must be used to obtain the best identification results.

In this paper we will present methods for determining whether an image was generated by a digital camera or a scanner. We use differences in the sensor pattern noise correlation that arise between the two classes due to inherent mechanical differences between their respective sensors. Features extracted from the pattern noise, similar to those used in [10] are used to determine the source class.

2. IMAGE CAPTURE DEVICES

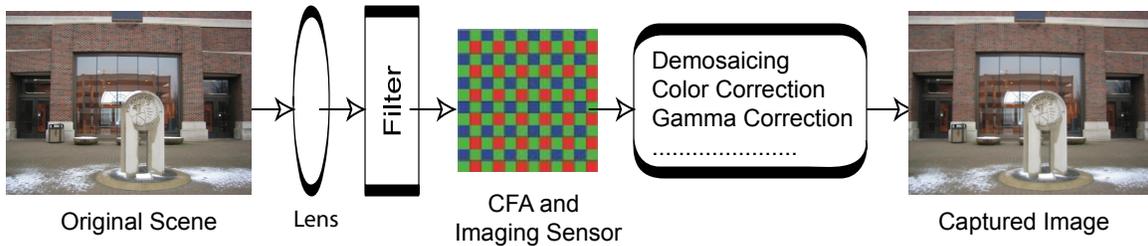### 2.1. Digital Camera Imaging Pipeline

Figure 1. Image Acquisition Model of a Digital Camera.

The imaging pipeline of most digital cameras are similar, irrespective of manufacturer or model[2]. The basic structure of a digital camera pipeline is shown in Figure 1. Light from a scene enters the camera through a lens and passes through a set of filters including an anti-aliasing filter. Next the light is "captured" by a sensor. The sensors, typically CCD or CMOS imaging sensors, are color blind in the sense that each pixel captures only intensity information. To capture color information, the light first passes through a color filter array (CFA) which assigns each pixel on the sensor one of three (or four) colors to be sampled. Shown in Figure 2 are CFA patterns using RGB and YMCG color spaces, respectively, for a 4×4 block of pixels. The individual color planes are filled in by interpolation using the sampled pixel values.

Finally, a number of operations are performed by the camera which include, but are not limited to, color interpolation, white point correction, and gamma correction. The image is then written into the camera memory in a user-specified image format (e.g. RAW, TIFF or JPEG). Although the operations and stages described in this section are standard in a digital camera pipeline, the exact processing details in each stage vary from one manufacturer to another, and between different camera models from the same manufacturer. This variation
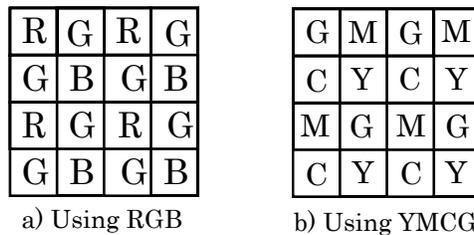

a) Using RGB   b) Using YMCG

Figure 2. CFA Patterns.

from one camera model to another can be used to determine the type of camera from which specific image was obtained. The features which remain same from camera to camera can be used to distinguish images generated by digital cameras from those generated by flat-bed scanners.

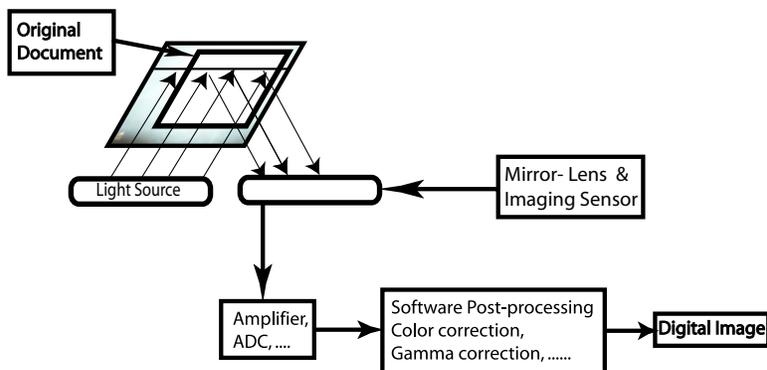## 2.2. Flatbed Scanner Imaging Pipeline

Figure 3. Flatbed Scanner Imaging Pipeline.

Figure 3 shows the basic structure of a flatbed scanner's imaging pipeline[11, 12]. The document is placed in the scanner and the acquisition process starts. The lamp used to illuminate the document is either a cold cathode fluorescent lamp (CCFL) or a xenon lamp, older scanners may have a standard fluorescent lamp. Using a stabilizer bar, a belt, and a stepper motor, the scan head slowly translates linearly to capture the image. The purpose of the stabilizer bar is to ensure that there is no wobble or deviation in the scan head with respect to the document. The scan head includes a set of lenses, mirrors, a set of filters, and the imaging sensor. Most desktop scanners use charge-coupled device (CCD) imaging sensors. Other scanners use CMOS (complementary metal oxide semiconductor) imaging sensors, Contact Image Sensors (CIS), or PMTs (photomultiplier tube) [11, 12]. The maximum resolution of the scanner is determined by the horizontal and vertical resolution. The number of elements in the linear CCD sensor determines the horizontal optical resolution. The step size of the motor controlling the scan head dictates the vertical resolution.

There are two basic methods for scanning an image at a resolution lower than the hardware resolution of the scanner. One approach is to sub-sample the output of the sensor. Another approach involves scanning at the full resolution of the sensor and then down-sampling the results in the scanner's memory. Most good quality scanners adopt the second method since it yields far more accurate results.

2.3. Sensor Noise

The manufacturing process of imaging sensors introduces various defects which create noise in the pixel values [8, 13]. There are two types of noise which are important. The first type of noise is caused by array defects. These include point defects, hot point defects, dead pixels, pixel traps, column defects and cluster defects. These defects cause pixel values in the image to deviate greatly. For example, dead pixels show up as black in the image and hot point defects show up as very bright pixels in the image, regardless of image content. Pattern noise refers to any spatial pattern that does not change significantly from image to image and is caused by dark current and photoresponse nonuniformity (PRNU). Dark currents are stray currents from the sensor substrate into the individual pixels. This varies from pixel to pixel and the variation is known as fixed pattern noise (FPN). FPN is due to differences in detector size, doping density, and foreign matter trapped during fabrication. PRNU is the variation in pixel responsivity and is seen when the device is illuminated. This noise is due to variations between pixels such as detector size, spectral response, thickness in coatings and other imperfections created during the manufacturing process. Frame averaging will reduce the noise sources except FPN and PRNU. Although FPN and PRNU are different, they are sometimes collectively called scene noise, pixel noise, pixel nonuniformity, or simply pattern noise. In [14] a method of estimating sensor pattern noise is successfully used for source camera

identification. The method uses a wavelet filter in combination with frame averaging to estimate the pattern noise in an image. We developed an extension of that method for scanners in[10] using LPA-ICI[15] to estimate the noise and a feature vector based classification that we believe is more robust than correlation methods.

Both digital cameras and scanners work on similar principles in terms of their imaging pipeline. However, digital cameras use a two dimensional sensor array while most scanners use a one dimensional linear sensor array. This difference can be used to distinguish between the two image sources. In a digital camera all the sensor elements are used to generate an image, while in scanners, only a portion of the sensor generates the complete image. In the case of flatbed scanners, the same linear array is translated to generate the entire image. It is expected to find periodicity between rows of the fixed component of the sensor noise of a scanned image. There is no reason to find a similar periodicity between columns of the sensor noise of the scanned image. Neither the rows nor the columns of the fixed component of the sensor noise of an image generated by a digital camera are expected to exhibit such periodicity. This difference can be used as a basis for discriminating between the two image source classes.

3. STATISTICAL FEATURES FOR IMAGING SENSOR CLASSIFICATION

First we will describe the feature vector used to discriminate between scanned and non-scanned classes. We will then, describe details of using a SVM to classify the images as scanned or non-scanned (taken from a camera).

3.1. Feature Vector Selection

The fixed component of the sensor noise can be used to discriminate between the two classes of images, scanned and non-scanned.

Let I denote the input image of size $M * N$ pixels, that is with M rows and N columns. Let I_{noise} be the noise corresponding to the original input image I and let $I_{denoised}$ be the result of denoising I using LPA-ICI[15]. Then,

$$I_{noise} = I - I_{denoised}$$

The noise, I_{noise}, can be modeled as a sum of two components, a random component I_{noise}^{random}, and a constant component $I_{noise}^{constant}$. For scanners, $I_{noise}^{constant}$ will depend only on the column index because the same linear sensor array is translated to generate the complete image.

The average of I_{noise} over all the rows can be used as the row reference pattern, $\widetilde{I}_{noise}^{constant}(1, j)$, of a scanner because the random components of I_{noise} will cancel each other while at the same time enhancing the constant part.

$$\widetilde{I}_{noise}^{constant}(1, j) = \frac{\sum_{i=1}^{M} I_{noise}(i, j)}{M}; 1 \leq j \leq N$$

To detect the similarity between different rows of the noise, we find the correlation of each of the M rows with the estimated row reference pattern $\widetilde{I}_{noise}^{constant}(1, j)$. Let $row\_corr(i)$, $1 \leq i \leq M$ denote the correlation between i^{th} row of I_{noise} and the row reference pattern. Correlation between two vectors $X, Y \in \mathbf{R^N}$ is defined as

$$correlation(X, Y) = \frac{(X - \bar{X}).(Y - \bar{Y})}{||X - \bar{X}||.||Y - \bar{Y}||}$$

Similarly, we obtain $col\_corr(j)$, $1 \leq j \leq N$, as a measure of similarity between different columns of I_{noise}. The first order statistics mean, median, mode, maximum and minimum and the higher order statistics variance, kurtosis and skewness of the row_corr and col_corr are used to generate the feature vector for every image. Ratio of the averages of row_corr and col_corr, which indicates the relative similarity among the rows or columns of the pattern noise, is used as another feature. Thus eleven first order and six higher order statistical features form the complete feature space. Two sets of experiments are performed. First set of experiments used only mean, median, mode and ratio of the averages of row_corr and col_corr (total seven). For the second set of experiments all the seventeen features are used to generate feature vector for every image.

Since cameras have a 2-D image sensor, no correlations are expected to be present between the rows or columns of the sensor noise. On the other hand, the fixed component of the noise should be nearly identical for all the rows of a scanned image. Thus, for cameras, the statistics of *row_corr* and *col_corr* will be similar; whereas for scanners the statistics of *row_corr* will differ from those of *col_corr*.

3.2. Support Vector Machine

Suppose we are given training data $(\mathbf{x}_1, y_1), ..., (\mathbf{x}_n, y_n)$ where $y_i \in \{1, -1\}$. The vectors $\mathbf{x}_i, \forall i$ represent the feature vectors input to the SVM classifier and y_i represent the corresponding class labels. Assuming that the class represented by the subset $y_i = 1$ and the class represented by $y_i = -1$ are "linearly separable", the equation of a decision surface in the form of a hyperplane that does the separation is $\mathbf{w}^T \mathbf{x} + b = 0$; where, $\mathbf{x}$ is an input vector, $\mathbf{w}$ is an adjustable weight vector, and b is a bias.

For a given weight vector $\mathbf{w}$ and bias b, the separation between the hyperplane and the closest data point is known as the margin of separation, denoted by M. The goal of a support vector machine is to find the particular hyperplane for which the margin of separation M is maximized[16]. Under this condition the decision surface is referred to as the optimum separating hyperplane (OSH) ($\mathbf{w}_o^T \mathbf{x} + b_o = 0$).

The pair $(\mathbf{w}_o, b_o)$ with appropriate scaling, must satisfy the constraint:

$$\mathbf{w}_o^T \mathbf{x} + b_o \geq 1 \quad \forall y_i = +1 \tag{1}$$

$$\mathbf{w}_o^T \mathbf{x} + b_o \leq -1 \quad \forall y_i = -1 \tag{2}$$

The particular data points $(\mathbf{x}_i, y_i)$ for which $y_i[\mathbf{w}^T \mathbf{x}_i + b] = 1$ are known as support vectors, hence the name "Support Vector Machine." The support vectors are the data points that lie closest to the decision surface and are therefore the most difficult to classify. As such they have the direct bearing on the optimum location of the decision surface. Since the distance to the closest point is $\frac{1}{\|\mathbf{w}\|}$, finding the OSH amounts to minimizing $\| \mathbf{w} \|$ with the objective function: $min \; \phi(\mathbf{w}) = \frac{1}{2} \| \mathbf{w} \|^2$ subject to the constraints shown in Equations 1 and 2.

If $(\alpha_1, \alpha_2 ..., \alpha_N)$ are the N non-negative Lagrange multipliers associated with constraints in Equations 1 and 2, the OSH can be uniquely constructed by solving a constrained quadratic programming problem. The solution $\mathbf{w}$ has an expansion $\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i$ in terms of a subset of training classes, known as support vectors, which lie on the margin. The classification function can thus be written as

$$f(\mathbf{x}) = \text{sgn}(\sum_i \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b) \tag{3}$$

If the data is not linearly separable, SVM introduces slack variables and a penalty factor such that the objective function can be modified as

$$\phi(\mathbf{w}) = \frac{1}{2} \| \mathbf{w} \|^2 + C(\sum_{i=1}^N \zeta_i) \tag{4}$$

Additionally, the input data can be mapped through some nonlinear mapping into a higher-dimensional feature space in which the optimal separating hyperplane is constructed. Thus the dot product required in Equation 3 can be represented by $k(\mathbf{x}, \mathbf{y}) = (\phi(\mathbf{x}).\phi(\mathbf{y}))$, when the kernel k satisfy Mercer's condition[17]. Finally, the classification function is obtained as

$$f(\mathbf{x}) = \text{sgn}(\sum_i \alpha_i \mathbf{y}_i k(\mathbf{x}_i, \mathbf{x}) + b) \tag{5}$$

Because the SVM can be analyzed theoretically using concepts from statistical learning theory, it has particular advantage in problems with limited training samples in high-dimensional space.

4. EXPERIMENTAL RESULTS

Table 1 shows the digital cameras and scanners used in our experiments. Approximately 100 images are scanned for each of the three scanners (2 of which are the same model) at the native resolution of the scanners. The images are then sliced into blocks (sub-images) of size 1024x768 pixels and sub-images from the first two columns of the scanned images are used. For the cameras 350 images are captured by each of the three cameras at a resolution of 1024x768. In total, we have 1800 scanned sub-images and 1050 camera images. Figure 4 shows a sample of the images used in this study.

Table 1. Image Sources Used in Experiments

Device Model	Class	Sensor	Native Resolution	Image Format
Canon PowerShot SD200	Camera	1/2.5 inch CCD	2048 x 1536	JPEG
Nikon Coolpix 4100	Camera	1/2.5 inch CCD	2288 x 1712	JPEG
Nikon Coolpix 7600	Camera	1/1.8 inch CCD	3072 x 2304	JPEG
Epson Perfection 4490 Photo	Flatbed Scanner	CCD	4800 dpi	TIFF
HP ScanJet 6300c-1	Flatbed Scanner	CCD	1200 dpi	TIFF
HP ScanJet 6300c-2	Flatbed Scanner	CCD	1200 dpi	TIFF

Figure 4. Sample of Images Used in Experiments

The complete experimental protocol is shown in Figure 5. To check the effectiveness of our proposed scheme in classifying images based on their sources, a number of experiments are performed by varying the type of images and the number of features used for classification. The SVM^{light} package[18] is used with a radial basis function chosen as the kernel function.

4.1. Experiment 1

Out of the 2850 images, half are randomly chosen to train the SVM and rest are used for testing. Initially, the feature vectors are generated using only mean, median, mode and ratio of average values of row_corr and col_corr (total seven features). An average classification accuracy of 98.1% is obtained over multiple runs in this case, and the confusion matrix is shown in Table 2. To improve the classification accuracy, another SVM model is generated using feature vectors containing the first as well higher order statistics of row_corr and col_corr. In this case, an average classification accuracy of 98.6% is obtained. The corresponding confusion matrix is shown in Table 3.

4.2. Experiment 2

In completely white or completely black images (hence forth referred to as "saturated" images) the sensor noise is highly suppressed[8, 13]. Because the proposed method of imaging sensor classification utilizes features from
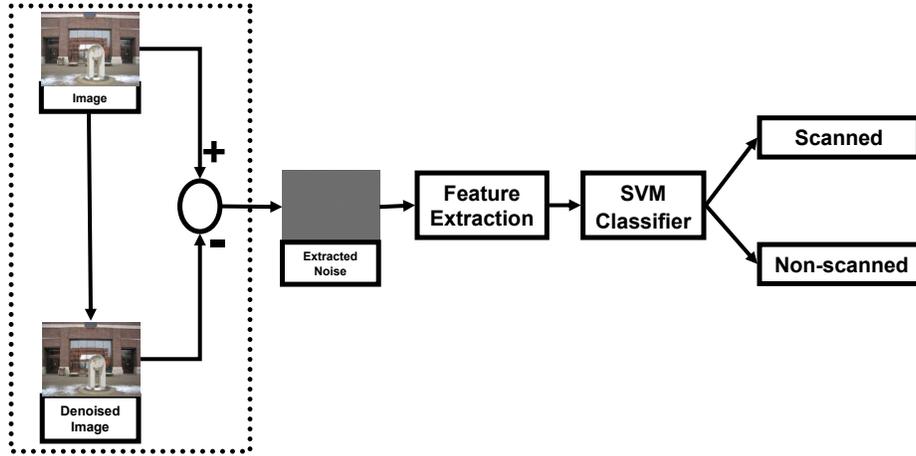
Figure 5. Image Source Classification

Table 2. Confusion Matrix for Experiment 1 with 7 Dimensional Feature Vector

		Predicted	
		Scanner	Camera
Actual	Scanner	97.9	2.1
	Camera	1.6	98.4

Table 3. Confusion Matrix for Experiment 1 with 17 Dimensional Feature Vector

		Predicted	
		Scanner	Camera
Actual	Scanner	98.4	1.6
	Camera	1.2	98.8

the fixed component of the sensor noise, such saturated images are likely to be mis-classified. The images mis-classified in Experiment 1 show that this is indeed the case. In this experiment, the saturated images are removed from the dataset, which leaves a total of 2000 scanned and non-scanned images. Many sub-images from the scanned images come under this excluded category since they are portions of bright areas (sky) and dark areas (roads) of the full images.

Again, half the images are chosen randomly for training and the other half for testing. Using only the first order statistics of *row_corr* and *col_corr*, an average classification accuracy of 98.9% is obtained with the confusion matrix shown in Table 4. Using the first as well higher order statistics of *row_corr* and *col_corr*, an average classification accuracy of 99.3% is obtained. The corresponding confusion matrix is shown in Table 5.

4.3. Experiment 3

To check the robustness of the proposed scheme when the imaging device to be tested is unavailable for training, the SVM is trained using features from images captured by the HP ScanJet 6300c-1, HP ScanJet 6300c-2, Canon Powershot SD200 and Nikon Coolpix 4100, while the testing set includes images from the Epson Perfection

Table 4. Confusion Matrix for Experiment 2 with 7 Dimensional Feature Vector (excluding the "saturated" images)

		Predicted	
		Scanner	Camera
Actual	Scanner	98.7	1.3
	Camera	0.9	99.1

Table 5. Confusion Matrix for Experiment 2 with 17 Dimensional Feature Vector (excluding the "saturated" images)

		Predicted	
		Scanner	Camera
Actual	Scanner	99.2	0.8
	Camera	0.6	99.5

Table 6. Confusion Matrix for Experiment 3 with 17 Dimensional Feature Vector, Trained Without the Epson 4490 and the Nikon Coolpix 7600

		Predicted	
		Scanner	Camera
Actual	Scanner	98.1	1.9
	Camera	10.9	89.1

Table 7. Confusion Matrix for Experiment 3 with 17 Dimensional Feature Vector, Trained Without the HP Scanjet 6300c-1 and the Canon Powershot SD200

		Predicted	
		Scanner	Camera
Actual	Scanner	98.5	1.5
	Camera	11.2	88.8

4490 and the Nikon Coolpix 7600. Using the first as well as higher order statistics of row_corr and col_corr, an average classification accuracy of 93.5% is obtained with the corresponding confusion matrix shown in Table 6. In a similar experiment in which the HP Scanjet 6300c-1 and Canon Powershot SD200 are not used for training, an average classification accuracy of 93.67% is obtained with the corresponding confusion matrix shown in Table 7.

4.4. Experiment 4

The efficacy of the proposed scheme is also tested on images that have been JPEG compressed. An average classification accuracy of 93.5% is obtained for JPEG images compressed using quality factor 90, as shown by the confusion matrix in Table 8. Both the training and testing images are JPEG compressed at quality factor 90.

Table 8. Confusion Matrix for Experiment 4 with 17 Dimensional Feature Vector, for JPEG Images at Quality Factor 90

		Predicted	
		Scanner	Camera
Actual	Scanner	97.6	2.4
	Camera	7.1	92.9

5. CONCLUSION AND FUTURE WORK

In this paper we investigated the use of the sensor pattern noise for classifying digital images based on their originating device, a scanner or a digital camera. Selection of proper features is the key to achieve accurate results. The scheme presented here utilizes the difference in the geometry of the imaging sensors and demonstrates promising results. As shown by our results, the proposed scheme does not need the availability of the actual source device for training purposes. Thus, even images generated by a completely unknown scanner or digital camera can be classified properly. Although results demonstrate good performance, we would like to extend this technique to work with images scanned at resolutions other than the native resolutions of the scanners. The challenge in working with lower resolution is to somehow address the degradation in sensor noise pattern due to down sampling. Future work will also include, tests on images that have undergone various filtering operations such as sharpening, contrast stretching and resampling. We are also looking at extending this approach for forgery detection.

REFERENCES

1. S. O. Jackson and J. Fuex. (2002) Admissibility of digitally scanned images. [Online]. Available: www.iediscovery.com/news/AdmissibilityDigitalImages.pdf
2. N. Khanna, A. K. Mikkilineni, A. F. Martone, G. N. Ali, G. T.-C. Chiu, J. P. Allebach, and E. J. Delp, "A survey of forensic characterization methods for physical devices," *Digital Investigation*, vol. 3, pp. 17–28, 2006.
3. M. Kharrazi, H. T. Sencar, and N. D. Memon, "Blind source camera identification," *Proceedings of the IEEE International Conference on Image Processing*, 2004, pp. 709–712.

4. A. Popescu and H. Farid, "Exposing digital forgeries in color filter array interpolated images," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3948–3959, 2005.

5. S. Bayram, H. Sencar, N. Memon, and I. Avcibas, "Source camera identification based on cfa interpolation," *Proceedins of the IEEE International Conference on Image Processing*, 2005, pp. 69–72.

6. J. Lukas, J. Fridrich, and M. Goljan, "Determining digital image origin using sensor imperfections," *Proceedings of the SPIE International Conference on Image and Video Communications and Processing*, A. Said and J. G. Apostolopoulos, Eds., vol. 5685, no. 1.   SPIE, 2005, pp. 249–260.

7. Z. J. Geradts, J. Bijhold, M. Kieft, K. Kurosawa, K. Kuroki, and N. Saitoh, "Methods for identification of images acquired with digital cameras," *Enabling Technologies for Law Enforcement and Security*, S. K. Bramble, E. M. Carapezza, and L. I. Rudin, Eds., vol. 4232, no. 1.   SPIE Press, 2001, pp. 505–512.

8. G. C. Holst, *CCD Arrays, Cameras, and Displays, Second Edition.*   JCD Publishing & SPIE Press, USA, 1998.

9. M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters*, vol. 6, no. 12, pp. 300–303, 1999.

10. N. Khanna, A. K. Mikkilineni, G. T.-C. Chiu, J. P. Allebach, and E. J. Delp, "Scanner identification using sensor pattern noise," *Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents IX*, 2007, to appear in.

11. J. Tyson. (2001) How scanners work. [Online]. Available: http://computer.howstuffworks.com/scanner.htm

12. (2001, Nov.) Scanners. [Online]. Available: http://www.pctechguide.com/55Scanners.htm

13. J. R. Janesick, *Scientific Charge-Coupled Devices.*   SPIE-International Society for Optical Engine, Jan 2001.

14. J. Lukas, J. Fridrich, and M. Goljan, "Detecting digital image forgeries using sensor pattern noise," *Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents VIII*, vol. 6072, San Jose, CA, January 2006.

15. A. Foi, V. Katkovnik, K. Egiazarian, and J. Astola, "A novel local polynomial estimator based on directional multiscale optimizations," *Proceedings of the 6th IMA Int. Conf. Math. in Signal Processing*, vol. 5685, no. 1, 2004, pp. 79–82.

16. C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

17. N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines (and other kernel-based learning methods).*   Cambridge University Press, 2000.

18. T. Joachims, "Making large-scale support vector machine learning practical," *Advances in Kernel Methods: Support Vector Machines*, B. Schölkopf, C. Burges, and A. Smola, Eds.   MIT Press, Cambridge, MA, 1998.