Confidence Weighting for Sensor Fingerprinting

Scott McCloskey

Honeywell Labs

1985 Douglas Drive North, Golden Valley, MN 55422

scott.mccloskey@honeywell.com

Abstract

The use of photo-response non-uniformity (PRNU) has been proposed as the basis of a sensor fingerprint for common source camera identification. We perform tests of the PRNU-based fingerprint on a set of videos chosen to represent a wide range of potential inputs. Based on the results of these tests, we propose a confidence weighting scheme to address the problem of extracting a viable fingerprint from videos where high-frequency content (e.g. edges) persist at a given image location. We further show that the extended PRNU estimation algorithm with confidence weighting has improved performance on such problematic videos.

1. Introduction

In a world with an increasing amount of digital video content, the ability to group those acquired with the same camera has an increasing number of applications. Without access to the camera in question, this problem is referred to as blind common source camera identification. In the realm of law enforcement, the ability to group offending videos (e.g. child pornography) is needed to assist forensic investigators. Moreover, large video repositories such as YouTube offer a rich set of data from which this ability could be used to infer the topology of underlying social networks.

Several methods have been presented to solve the common source grouping problem on visual data. Readers are encouraged to refer to Sencar and Memnon [10] for an excellent overview of the current state of the art in camera identification and the wider field of digital image forensics. We briefly review some relevant work here. The method of Kurusowa *et al.* [7] uses the location of dead pixels and the level of dark current as the basis for its sensor fingerprints. Unfortunately, under proper illumination and after video-quality compression, dark noise is extremely difficult to measure. Methods that characterize lens properties [3] or image processing components [11, 5, 9], while useful in certain circumstances, do not provide a device level fingerprint, and are unable to distinguish between two different instances of the same model.

Building on previous work with still images [1, 8, 4], the method of Chen *et al.* [2] performs common source video identification using fingerprints based on the sensor's Photo-Response Non-Uniformity (PRNU). PRNU arises as a result of material and manufacturing imperfections of CCD and CMOS sensors and is manifested in slight sensitivity variations between pixels on the sensor. Because the PRNU characterizes the sensor rather than the lens or image processing components, it can distinguish between two different instances of the same model camera. Unlike dead sensor pixels, which are often identified in advance and interpolated in the camera's firmware, the effects of PRNU are difficult to characterize in advance, and are generally uncorrected in the final video data. In short, PRNU-based fingerprints are the most promising current option for device-level common source camcorder identification.

Though the method is very promising, the testing presented in [2] does not go into the level of detail necessary to give a reader a thorough understanding of the method's efficacy on different types of inputs. We address this need by presenting a detailed, independent testing of the method on a set of test videos chosen to span a wide range of potential inputs. While we find the method to be effective in many cases, we identify and analyze several conditions under which its performance is relatively weaker.

Based on the results of our testing and analysis of Section 2, we extend and improve the method of [2] to handle scenes with high-frequency content (e.g. edges) that persist at a given location throughout a video. We address such problematic videos with confidence weighting in Section 3. Whereas the method described in [2] does not discriminate between pixels within a frame, we use confidence mapping to identify parts of the scene that are relatively more helpful in building the PRNU estimate. This confidence map allows us to generate a more accurate PRNU fingerprint by minimizing the contribution from regions of the scene that are likely to distort the estimate. We experiment with different confidence weighting schemes, and demonstrate the improved performance of the extended algorithm in Section

3.1. We close the paper with concluding remarks and comments on future work in Section 4.

2. Common Source Camcorder Identification by PRNU Estimation

The method of common source camera identification, as outlined in [2], consists of two independent steps: an algorithm for PRNU estimation and a second for comparing estimated PRNUs. We review each of these two steps before describing our testing.

PRNU Estimation - Given a video of an arbitrary scene, the extraction of the PRNU necessitates decomposing each video frame I_k into its scene $\hat{I}_k$ and noise $I_k - \hat{I}_k$ components. For a video sequence of K frames, the maximum likelihood estimate of the PRNU $\hat{P}$ is computed according to

$$\hat{P} = \frac{\sum_{k=1}^{K}(I_k - \hat{I}_k)\hat{I}_k}{\sum_{k=1}^{K}\hat{I}_k^{\,2}}, \tag{1}$$

where $\hat{I}_k$ is the de-noised version of video frame I_k, estimated using the method of Xiao *et al.* [6]. The PRNU is a zero-mean array on the same *M-by-N* lattice as the sensor. Conceptually, the PRNU value indicates the pixel's relative response to incoming light, with positive values indicating an increased sensitivity and negative values indicating a decreased sensitivity relative to other pixels on the sensor.

The separation of each frame into its scene and noise components is the classic source separation problem, which has been long studied in the field of image processing. Like most interesting problems, source separation is ill-posed in the sense that there are an infinite number of scene/noise combinations that could generate any given input. Due to the ill-posed nature of the problem, source separation methods make assumptions about the input and are, therefore, expected to erroneously classify some image content when conditions deviate from those assumptions. In particular, most denoising methods will mistakenly classify some high-frequency scene content as noise. As we will demonstrate in Section 2.1, one consequence of this misclassification is that highly-textured, stationary parts of the scene will contribute disproportionately to the estimated PRNU.

PRNU Comparison - In order to determine whether or not two videos came from the same camera, the PRNUs estimated from the two inputs must be compared. In [2], the comparison is done by computing the cross-correlation S of the two patterns

$$S(u,v) = \sum_{x,y} PRNU_1(x,y)PRNU_2(x+u, y+v), \tag{2}$$

computed over the range $u \in [-\frac{M}{2}, \frac{M}{2}]$, $v \in [-\frac{N}{2}, \frac{N}{2}]$. In the event that the two PRNUs come from the same camera, the cross-correlation should be similar to a delta function: a pronounced peak surrounded by a noise-like pattern. If the two PRNUs were unrelated, the cross-correlation function should have small, zero-mean values everywhere. In order to make a final decision, the method scores the cross-correlation for its similarity to a delta function and compares that score to a threshold. Several metrics exist to measure peak sharpness of a function. In [2], the authors use peak correlation energy, though their results are inconsistent with the definition of that metric. We use the peak-to-secondary ratio (PSR), defined as the ratio between the first and second highest peaks in S. This measure has a lower bound of 1, and has the advantage of a more intuitive meaning for the purpose of thresholding.

2.1. Performance Assessment Experiments

In order to asses the performance of this method, video data was collected from several cameras of different types (see Table 1). Each camera was used to shoot a fixed set of scenes covering a range of conditions (see Table 2). Videos were acquired with as little compression as possible. Those from the Sony DVCs cameras were acquired without compression, whereas videos from the Kodak and Canon DSCs were acquired with the camera's default compression (we could not disable it). In addition to these, a single video was taken from a Panasonic camera in order to verify that it would not be mistaken for any of the test cameras. This gives 666 exhaustive comparisons between different videos, of which 183 should be matched. The discrimination threshold of 1.505 was chosen to target a 0.5% false accept rate (FAR) over a larger experimental data set including different levels of compression.

In our first experiment, PRNU signatures were estimated from 40-second long video segments without compression. Figure 1 (left) shows our grouping results for uncompressed videos. The most obvious problem evident in these results is the method's inability to group videos taken with image stabilization. The `stabilization` videos from three of the cameras could only be grouped with themselves, while the video from the fourth camera (the Canon) was matched successfully. The distinction is that the three cameras whose videos could not be grouped employ *electronic* image stabilization, whereas the Canon camera does not. In order to understand this, it is necessary to understand the way in which image stabilization is performed. In cameras with *electronic* image stabilization, internal sensors are used to detect the direction and magnitude of camera shake. When present, camera shake is neutralized by shifting the image, using extra pixels around the perimeter of the sensor to maintain the same frame size. In so doing, electronic image stabilization *changes the relationship between pixels in the resulting*

Make	Model	Sensor	Video Res.	Description	Picture
Kodak	V550	5MP CCD	480x640	Ultra-Compact DSC w/ Video Mode	
Canon	A610	5MP CCD	480x640	Compact DSC w/ Video Mode	
Sony	DCR DVD-200	1MP CCD	480x704	Hand-held DVD-R DVC	
Sony	DCR PC-350	3MP CCD	480x720	Hand-held MiniDV DVC	

Table 1. Cameras used in performance assessment

Code	Name	Description
A	`outdoor`	Hand-held walk around a suburban neighborhood
B	`flatField`	Defocused video of blank wall
C	`interview`	Tripod-mounted capture (fixed position) of an interview
D	`houseInterior`	Hand-held walk around a house
E	`kitchen`	Hand-held walk around a kitchen
F	`stabilization`	Shaky, outdoor acquisition with image stabilization
G	`lightingVariation`	Hand-held walk with extreme lighting variations
H	`zooming`	Hand-held zooming and panning
I	`flatField_nightmode`	Defocused video of a blank wall in low light using night mode
J	`indoor_nightmode`	Hand-held walk in low lighting using night mode

Table 2. Scenes used for performance assessment. Note that the night mode videos were only acquired with the Sony cameras, as neither the Kodak nor Canon cameras have night modes.

video and the sensor pixels from which they were recorded. Because the PRNU-based identification algorithm assumes that all intensities at a given location in the video have been recorded by the same sensor pixel, the estimated PRNU signature becomes inaccurate when electronic image stabilization is employed. Readers should note that some cameras use *optical* image stabilization, which involves a floating lens element instead of shifts in the acquired image, and does not present the same problem.

Besides the issues with image stabilization, the only misclassification was a false negative when comparing the Kodak V550-acquired `interview` and `outdoor` videos. Upon further inspection, we find that the PRNU estimated from the `interview` video is dominated by edges from the scene, as shown in Figure 2. Because the denoising algorithm classifies some of the edges as noise, and because the edges appear in relatively fixed locations in each frame, this error accumulates in the estimated PRNU and outweighs the desired sensor fingerprint. In addition to causing false negatives, as it did in our test, such errors may result in false positives between videos acquired with different cameras if they happen to have persistent edges in the same location. The same `interview` scene was classified appropriately for the uncompressed video from the three other cameras, which is likely due to the fact that they capture video at a higher quality level than the Kodak V550, as a result of its in-camera compression (at a lower bitrate than the Canon A610's). If this analysis is accurate, it is expected that the `interview` video will be misclassified

for other cameras at increasing levels of compression.

With respect to the videos acquired using night mode, we note that they were grouped as well as the videos taken in the daylight mode. Found on some digital video cameras, night mode enables capture of low-light scenes by removing an infra-red filter element from the optical path and employing a matched illuminator. Because neither of these changes alter the properties of the sensor, it is not surprising that they do not impact the PRNU estimation.

Figure 1 (right) shows the comparison matrix for videos that have been compressed to 1Mbps using the XVID codec. Overall, the performance of the PRNU-based signature matching is significantly degraded, as the noise features that comprise it are lost in compression. Readers should note that, while a larger number of frames can be used to compensate for this loss, we have kept that number constant in order to illustrate which videos are relatively harder to match. With that in mind, we see that stabilization remains a problem in the compressed videos. Also, as expected, the `interview` video was increasingly problematic, with two thirds (22) of the 33 new false negative results involving it.

Interestingly, the `outdoor` video was problematic for the Kodak camera and, to a lesser extent, the Canon camera. In this case, the problem has less to do with the scene content than the camera settings. As the videos were taken during the day, the outdoor lighting provided a high level of exposure and the level of amplification necessary to produce a well-exposed video was relatively low. As a result, sensor noise was not amplified to the same degree as the videos
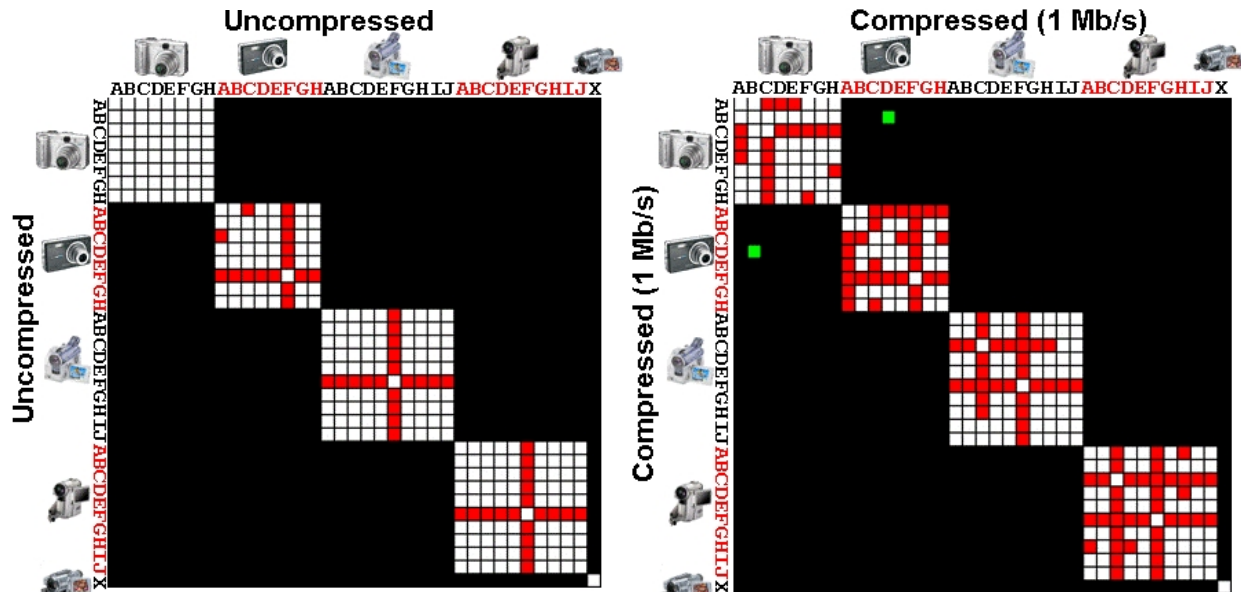
Figure 1. Comparison matrix for uncompressed videos (left) and compressed videos (right). White cells indicate videos that were correctly identified as coming from the same camera. Black cells indicate videos correctly identified as coming from different cameras. Red cells represent videos mistakenly identified as coming from different cameras. Green cells represent videos mistakenly identified as coming from the same camera.

taken indoors, so the sensor fingerprint is relatively weaker. This problem arises with the Canon and Kodak cameras because they use in-camera compression, by which the already low levels of noise are further reduced.

Summary of Testing - The results of these tests indicate several areas of potential improvement for the method of [2]: electronic image stabilization, persistent image content, and videos acquired with low levels of amplification. In the following section, we will address the issue of persistent image content, exemplified by the `interview` video. As we have noted in Section 2.1, the failures of classification result from the accumulated errors of the denoising algorithm. We will address this using confidence weighting, by which we attempt to predict the image locations at which the denoising method will fail, and consider the output of that method accordingly.

3. Confidence Weighted Fingerprints

At a high level, the misclassifications involving the `interview` video result from the fact that the algorithm ignores the differences between pixels within each frame. The overall objective of our confidence weighting scheme is to identify the regions of a frame where we have high confidence in the noise estimate that is used as a basis for the PRNU fingerprint. With that, equation 1 is modified so that areas of low confidence do not significantly contribute errors to the PRNU estimation. By weighting against such areas, we avoid artifacts of the sort that we have shown in Figure 2 which have been shown to cause false mismatches,

and may also lead to false matches between videos taken with different cameras.

Having experimented with the de-noising method of [6] on images with artificial noise we find that, as in the example shown in Figure 2, edge and texture content is most often misclassified as noise. Moreover, due to the manipulation of a multi-scale wavelet representation, these errors are spread throughout a neighborhood around edge and texture content. Early experiments showed that edge detectors were a poor predictor of such failures, as they did not detect textured regions such as hair. In order to account for such regions, we use a confidence weight based on image gradient magnitudes, which we have found to be an excellent predictor of failures of the de-noising method. As a result, the weight associated with a pixel p is

$$w(p) = G(\sigma) * \frac{1}{(1 + \|\nabla I(p)\|)}, \qquad (3)$$

where I is the image intensity, $G(\sigma)$ is a Gaussian kernel, and ∇ denotes the gradient operator. The convolution with a Gaussian is used to mimic the spreading of errors due to the multi-scale wavelet denoising. The weighting matrix w is applied multiplicatively to the nose estimate in equation 1 to generate the confidence-weighted sensor fingerprint.

3.1. Experiments

Figure 2 (lower right) shows the PRNU signature estimated from the Kodak-acquired `interview` video using confidence weighting. This signature lacks the high-
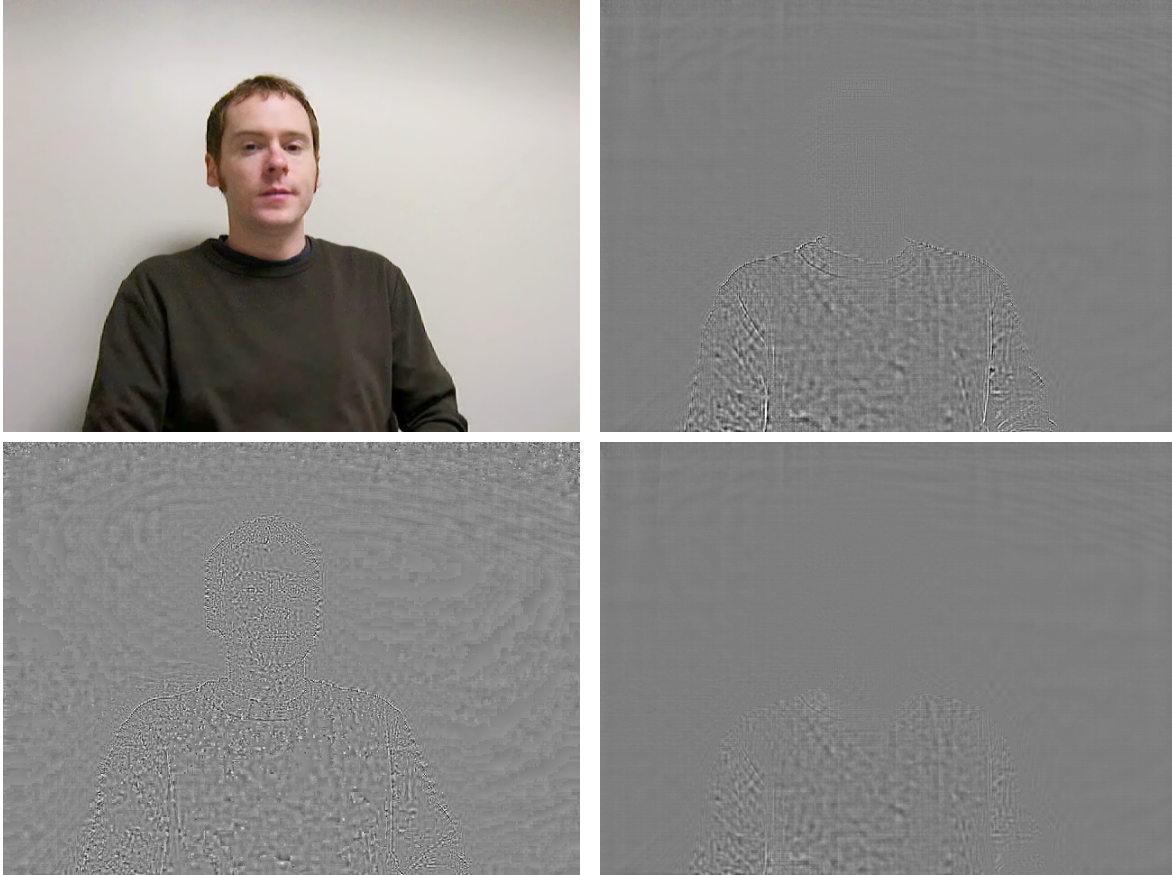
Figure 2. **Top Row:** Example frame (left) and estimated PRNU fingerprint (right) from `interview` video, acquired by the Kodak V550. Edge content from the video dominates the estimated PRNU due to failures of the denoising method. **Bottom Row:** estimated noise content of the example frame (left) and PRNU fingerprint (right), estimated using confidence weighting. Readers are encouraged to zoom in on the signatures in the electronic document, as fine details may not be visible in a printed copy.

Video Name	New Score	Old Score
outdoor	1.03	1.10
flatField	2.10	1.28
houseInterior	2.07	1.58
kitchen	1.60	1.11
stabilization	1.00	1.00
lightingVariation	2.08	1.72
zooming	1.06	1.07

Table 3. Signature matching scores for the 1Mbps-compressed Kodak V550 videos using confidence-weighted PRNU (middle column) and unweighted PRNU (right column). Red cells indicate false negative matches, whereas white cells represent true matches.

magnitude edges that were prominent in the signature shown in Figure 2 (upper right). Readers may note that the magnitude of the signature is higher in the torso region of the frame, due to the difference in overall intensity (recall the denominator in equation 1).

Table 3 shows the results of matching the 1Mbps-compressed videos with signatures extracted from the

`interview` video with and without confidence weighting. From it we can see that the signature generated using confidence weighting is correctly grouped with the `flatField` and `kitchen` videos that had been incorrectly identified as a mismatch without weighting. In addition, the two existing matches, one of which was quite close to the threshold, were strengthened as a result of confidence weighting. The remaining mismatches involve video clips from which the extraction of viable signatures is difficult for other reasons, as discussed in Section 2.1.

Figure 3 shows the comparison matrix for the signatures extracted from the 1Mbps-compressed videos using confidence mapping. From it, we can see that the comparisons involving the `interview` video are significantly improved. Of the 26 previous misclassifications involving that video, the confidence-weighted signature enables a match for 18. The remaining misclassifications involve videos, primarily `stabilization`, that are problematic for other reasons analyzed in Section 2.1. Given that we have not attempted to address these other shortcomings, we view this
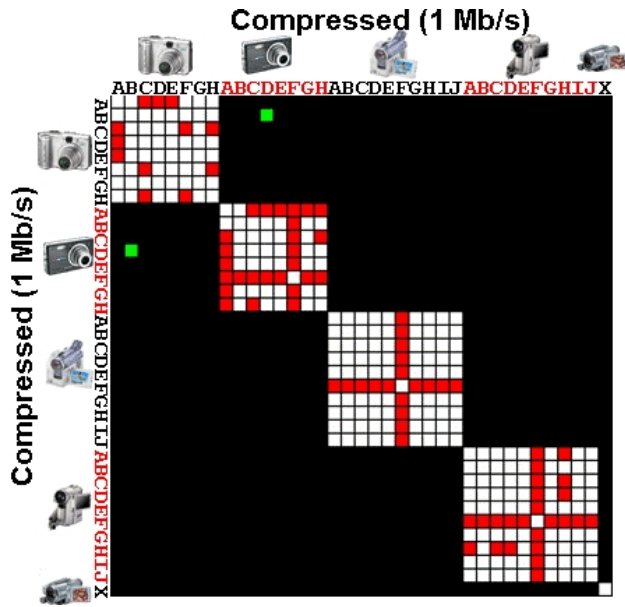
Figure 3. Comparison matrix for 1Mbps-compressed videos using confidence weighting. As before, white cells indicate videos that were correctly identified as coming from the same camera. Black cells indicate videos correctly identified as coming from different cameras. Red cells represent videos mistakenly identified as coming from different cameras. Green cells represent videos mistakenly identified as coming from the same camera.

as confirmation of the utility of confidence mapping as an extension to the PRNU-based sensor fingerprinting. This improvement was achieved without a change in the detection threshold, and without an increase in the number of false matches.

4. Conclusions and Future Work

The results of our validation experiments presented in Section 2.1 indicate that, while effective in many circumstances, the camera identification method suggested in [2] has several shortcomings. One of the shortcomings analyzed in Section 2.1 is the poor quality of signatures extracted from videos with high-frequency content, such as edges, that persist at a fixed location. Having observed and explained this behavior, we use a simple image-based measure to predict the underlying failures of the denoising and reduce the impact of those failures. As we have demonstrated, this leads to improved performance relative to our test set. We believe that the videos in the test set are representative, and that other videos with static high-frequency content will demonstrate improvements similar to those illustrated with the `interview` video. Given that most videos acquired from a fixed position (such as a tripod) will contain high-frequency content at fixed locations, this extends PRNU-based sensor fingerprinting to a large class of videos on which it had not been previously applicable.

Going forward, we intend to explore the use of a confidence weighting scheme to address the other shortcomings mentioned in 2.1. In particular, in light of our analysis of the results of the `outdoor` video, we anticipate that frames acquired with a higher gain will be relatively more important in signature extraction. Given a per-frame estimate of the gain, which is currently unavailable for our test data, we could exploit this within the framework of confidence weighting. In the future, we expect to further improve the camera matching performance by estimating the relative gain and using that as a further confidence weight for frames.

References

[1] M. Chen, J. Fridrich and M. Goljan. *Digital Imaging Sensor Identification (Further Study)*. Proc. SPIE 2007.

[2] M. Chen, J. Fridrich, M. Goljan, and J. Lukas. *Source Digital Camcorder Identification Using Sensor Photo Response Non-Uniformity*. Proc. SPIE 2007.

[3] K. S. Choi, E. Y. Lam, and K Wong. *Source Camera Identification Using Footprints from Lens Aberration*. Proc. SPIE 2006.

[4] M. Goljan, M. Chen, and J. Fridrich. *Identifying Common Source Digital Camera from Image Pairs*. Proc. of IEEE International Conference on Image Processing (ICIP) 2007.

[5] M. Kharrazi, H. T. Sencar, and N. Memon. *Blind Source Camera Identification*. Proc. of IEEE International Conference on Image Processing (ICIP) 2004.

[6] S. Xiao, I. Kozintsev, and K. Ramchandran. *Stochastic wavelet-based image modeling using factor graphs and its application to denoising*. Proc. of IEEE Conference on Acoustic, Speech and Signal Processing (ICASSP), vol. 4, May 2000.

[7] K. Kurusowa, K. Kuroki, and N. Saitoh. *CCD Fingerprint Method*. Proc. of IEEE International Conference on Image Processing (ICIP) 1999.

[8] J. Lukas, J. Fridrich, and M. Goljan. *Digital Camera Identification from Sensor Pattern Noise*. IEEE Trans. on Information Forensics and Security, vol. 1, no. 2, pp. 205-214, 2006.

[9] A.C. Popescu and H. Farid. *Statistical Tools for Digital Forensics*. 6th International Workshop on Information Hiding, 2004.

[10] H. T. Sencar and N. Memnon. *Overview of State-of-the-Art in Digital Image Forensics*. In *Statistical Science and Interdiciplinary Research*. World Scientific Press, 2008.

[11] A. Swaminathan, M. Wu, and K. J. Ray Liu. *Non-Intrusive Forensic Analysis of Visual Sensors Using Output Images*. Proc. of IEEE Conference on Acoustic, Speech and Signal Processing (ICASSP), vol. 5, pp. 401-404, May 2006.