

IMPROVEMENTS ON SOURCE CAMERA-MODEL IDENTIFICATION BASED ON CFA INTERPOLATION

Sevinc Bayram ^a, Husrev T. Sencar ^b, Nasir Memon ^b

E-mail: sevincbayram@hotmail.com, taha@isis.poly.edu, memon@poly.edu

^aDept. of Electrical and Computer Eng., Uludag University, Bursa, TURKEY

^bDept. of Computer and Information Sci., Polytechnic University, Brooklyn, NY, USA

Keywords: Image forensics, digital camera, demosaicing/interpolation, color filter array

Abstract

The idea of using traces of interpolation algorithms, deployed by a digital camera, as an identifier in the source camera-model identification problem has been initially studied in [2]. In this work, we improve our previous approach by incorporating methods to better detect the interpolation artifacts in *smooth* image parts. To identify the source camera-model of a digital image, new features that can detect traces of low-order interpolation are introduced and used in conjunction with a support vector machine based multi-class classifier. Performance results due to newly added features are obtained considering source identification among two and three digital cameras. Also, these results are combined with those of [2] to further improve our methodology.

1. INTRODUCTION

The advances in digital technologies have given birth to very sophisticated and low-cost hardware and software tools that enabled easy creation, distribution and modification of digital images. This trend has brought with it new challenges concerning the integrity and authenticity of digital images. As a consequence, one can no longer take the authenticity of digital images for granted. Image forensics, in this context, is concerned with determining the source and potential authenticity of a digital image.

Although, digital watermarking technologies [3] have been introduced as a measure to address this problem, its realization requires that the watermark be embedded during the creation of the digital image. Essentially, this necessitates digital cameras to have built-in watermarking capabilities. However, this approach has not been adopted by digital camera manufacturers. Consequently, to determine origin, veracity and nature of digital images, alternative approaches need to be considered. The setting of this problem is further complicated by the requirements that the methods should require as little as possible prior knowledge on the digital camera and the actual conditions under which the image has been captured (*blind image authentication*). *At the present time, there is a severe lack of techniques that could achieve these goals.*

The underlying assumption for the success of blind image authentication techniques is that all images produced by a digital camera will exhibit certain characteristics regardless of the captured scene, which are unique to that camera, due to its proprietary

image formation pipeline. It should be noted that all digital cameras encode the camera model, type, date, time, and compression information in the EXIF image header. However, since this information can be easily modified or removed, it cannot be used for authentication.

In this paper, we concentrate on source camera-model identification problem by identifying the traces of proprietary interpolation algorithm deployed by digital cameras. For this, we improve our results in [2] by incorporating new methodologies to capture CFA interpolation artifacts due to low-order interpolation.

1.1 Prior Work

In our prior work [1], we studied the source camera-model identification problem by identifying and selectively combining a set of image features based on image quality metrics [4] and higher-order statistics of images [5]. This approach essentially requires the design of a classifier that is able to capture the variations in the designated image features, due to different digital cameras.

Another promising approach in this area is made by Lukas *et al.* [6]. In their work, sensor's *pattern noise* is characterized via wavelet-based image denoising. The reference noise pattern for a particular digital camera is obtained by averaging over a number of high quality JPEG images captured by that camera, and for a given image its source camera is verified by correlating the noise pattern of the particular camera (which is claimed to capture the image in question) with the individual noise pattern extracted from the image itself.

In [2], we exploit the fact that most state-of-the-art digital cameras, due to cost considerations, employ a single mosaic structured *color filter array* (CFA) rather than having different filters for each color component. As a consequence, each pixel in the image has only one color component associated with it, and each digital camera employs a proprietary interpolation algorithm in obtaining the missing color values for each pixel. Our approach in [2] was inspired by the technique proposed by Popescu *et al.* intended for image tamper detection [7]. The rationale for their technique is that the process of image tampering very often requires up-sampling operation which in turn introduces periodic correlations between the image pixels. To detect such phenomena they designated statistical measures. In a similar manner, we have applied variants of such measures to characterize the specifics of the deployed interpolation algorithm.

In the present work, we further improve our approach in [2] by designating new features. Due to perceptual image quality considerations, designers have to tailor the interpolation algorithm to deal with different qualities in an image, *i.e.*, edges, texture features, etc. This essentially requires introducing strong non-linearities to the interpolation algorithm. However, in relatively smooth image parts, most well known interpolation algorithms (*e.g.*, bilinear and bicubic methods) will ensure satisfactory quality, and very *expensive* algorithms are not needed. Our premise in this work is that most proprietary algorithms in smooth image parts will deploy simpler forms of interpolation, and therefore, they can be captured more effectively (as opposed to *busy* image parts where interpolation requires more careful processing). For this purpose, we

utilize the results of [8] where the periodicity pattern in the second order derivative of interpolated signal is analyzed.

The rest of this paper is organized as follows. In section 2, we briefly describe the image formation process in digital cameras. In Section 3, the results of [2] are reviewed and, the details of the improved approach are provided. We present our experimental results in Section 4 and conclude in Section 5.

2. IMAGE FORMATION IN DIGITAL CAMERAS

The general structure and sequence of processing stages of image formation pipeline in a digital camera remains to be very similar in all digital cameras (despite the proprietary nature of the underlying technology). A typical digital camera pipeline is shown in Figure 1-(a) [9]. The light entering the camera through the lens is first filtered (the most important being an anti-aliasing filter) and focused onto an array of charge-coupled device (CCD) elements, *i.e.*, pixels. The CCD array is the main and most expensive component of a digital camera. Each light sensing element of CCD array integrates the incident light over the whole spectrum and obtains an electric signal representation of the scenery. Since each CCD element is essentially monochromatic, capturing color images requires separate CCD arrays for each color component. However, due to cost considerations, in most digital cameras, only a single CCD array is used by arranging them in a pattern where each element has a different spectral filter, typically one of red, green or blue (RGB). This mask in front of the sensor is called the color filter array (CFA). Hence, each CCD element only senses one band of wavelengths, and the raw image collected from the array is a mosaic of red, green and blue pixels. Figures 1-b and 1-c display a CFA pattern using RGB and YMCG color space respectively for a 6x6 pixel block. As each sub-partition of pixels only provide information about a number of green, red, and blue pixel values, the missing RGB values for each pixel need to be obtained through interpolation (demosaicing). The interpolation is typically carried out by applying a weighting matrix (kernel) to the neighboring pixels around a missing value. Most generally, each manufacturer uses a proprietary demosaicing algorithm *i.e.*, kernels with different sizes, shapes and different interpolation functions. This is followed by the processing block, shown in the Figure 1-a, which involves a number of operations like color processing and compression producing the final image.

Although the block diagram for image formation pipeline remains same for almost all cameras, the exact processing detail at all stages vary from one manufacturer to other, and even in different camera models manufactured by the same manufactures. It should also be noted that many components in the image formation pipeline of various digital cameras, (*e.g.*, lens, optical filters, CCD array) are produced by a limited number of manufactures. Therefore, due to this overlap, different cameras may exhibit similar qualities, and this should be taken into consideration in associating image features with the properties of digital cameras. However, interpolation (demosaicing) algorithm and the design of the CFA pattern remain to be proprietary to each digital camera manufacturer. In the next section we will describe how the variations in color interpolation can be exploited to classify the images either originating from one camera or the other.

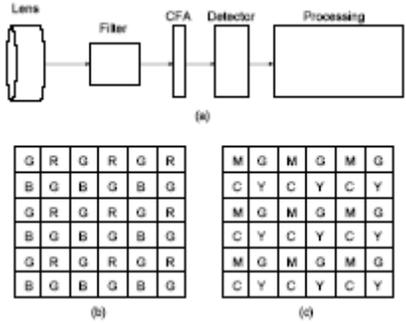


Figure 1. (a) The more important stages of a camera pipeline are shown. (b) CFA pattern using RGB values. (c) CFA pattern using YMCA values.

3. IDENTIFYING TRACES OF INTERPOLATION

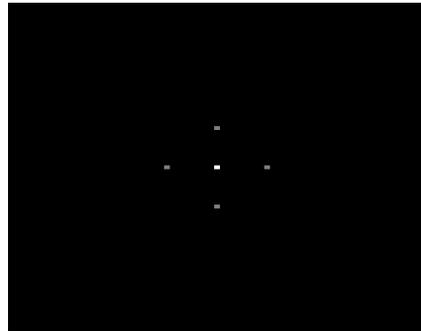
In [7], Popescu *et al.* presented a methodology to detect traces of up-sampling to identify images (or parts of images) that have undergone resizing by analyzing the correlation of each pixel value to its neighbors. Since in a typical digital camera RGB channels are heavily interpolated, we proposed to apply a similar procedure to determine the correlation structure present in each color band and classify images accordingly [2]. Our initial experimental results [1] indicate that both the size of interpolation kernel and the demosaicing algorithm vary from camera to camera. Furthermore, the interpolation operation is highly non-linear, making it strongly dependent on the nature of the depicted scenery. In other words, these algorithms are fine-tuned to prevent visual artifacts, in forms of over-smoothed edges or poor color transitions, in busy parts of the images. On the other hand, in smooth parts of the image, these algorithms exhibit a rather linear characteristic. Therefore, in our analysis we treat smooth and non-smooth parts of images separately.

3.1 Non-smooth image parts

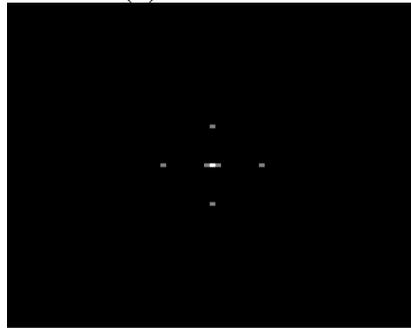
We employ Expectation/ Maximization (EM) algorithm to detect traces of interpolation [7]. The EM algorithm consists of two major steps: an expectation step, followed by a maximization step. The expectation is with respect to the unknown underlying variables, using the current estimate of the parameters, and conditioned upon the observations. The maximization step then provides a new estimate of the parameters. These two steps are iterated until convergence [10]. The EM algorithm generates two outputs. One is a two-dimensional data array, called *probability map*, in which each entry indicate the similarity of each image pixel to one of the two groups of samples, namely, the ones correlated to their neighbors and those ones that are not, in a selected kernel. On this *map* the regions identified by the presence of periodic patterns indicate the image parts that have undergone up-sampling operation. The other output is the estimate of the *weighting (interpolation) coefficients* which designate the amount of contribution from each pixel in the interpolation kernel.

Since no *a-priori* information is assumed on the size of interpolation kernel (which designates the number of neighboring components used in estimating the value of a missing color component) probability maps are obtained for varying sizes of kernels. When observed in the frequency domain, these probability maps yield to peaks at different frequencies with varying magnitudes indicating the structure of correlation

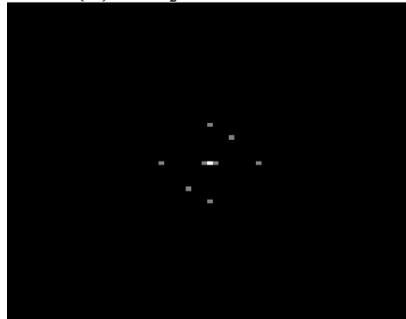
between the spatial samples. In designing our classifier we rely on two sets of features: The set of weighting coefficients obtained from an image, and the peak location and magnitudes in frequency spectrum. In Figure 2, sample magnitude responses of frequency spectrum of the probability maps for three cameras (Sony, Nikon and Canon) are given. The three responses differ in peak locations and magnitudes.



(a) Nikon E-2100



(b) Sony DSC-P51



(c) Canon Powershot S200

Figure 2. Frequency spectrum of probability maps obtained for three models of digital cameras.

3.2 Smooth Image Parts

In [8], Gallagher showed that low-order interpolation introduces periodicity in the variance of the second order derivative of an interpolated signal which can be subsequently used to determine the interpolation rate and algorithm of the signal. The proposed interpolation detection algorithm first obtains the second order derivative of each row and averages it over all rows. When observed in the frequency domain the locations of the peaks reveal the interpolation rate and the magnitude of the peaks determine the interpolation method. We employed a similar methodology to characterize the interpolation rate and the method employed by a digital camera. It should be noted

that most digital cameras encode and compress images in JPEG format. Due to 8x8 block coding, the DC coefficients may also introduce peaks in the second-order derivative implying the presence of some form of interpolation operation at a rate of 8. Therefore, in detecting the interpolation algorithm, the peaks due to JPEG compression have to be ignored. Figure-3 displays the magnitude frequency response for the three models of digital cameras. The variation in magnitude and indicates that there are differences in the deployed interpolation algorithm. Therefore, the features extracted from each camera include the location of the (peaks except for the ones due to JPEG compression), their magnitudes, and the energy of each frequency component with respect to other frequency components at all color bands.

3. EXPERIMENTAL RESULTS

An SVM classifier was used to test the effectiveness of the proposed features. There are a number of publicly available SVM implementations. Our work is based on the LibSvm package [11]. We have also used the sequential forward floating search (SFSS) algorithm to select the best features from a given set of features.

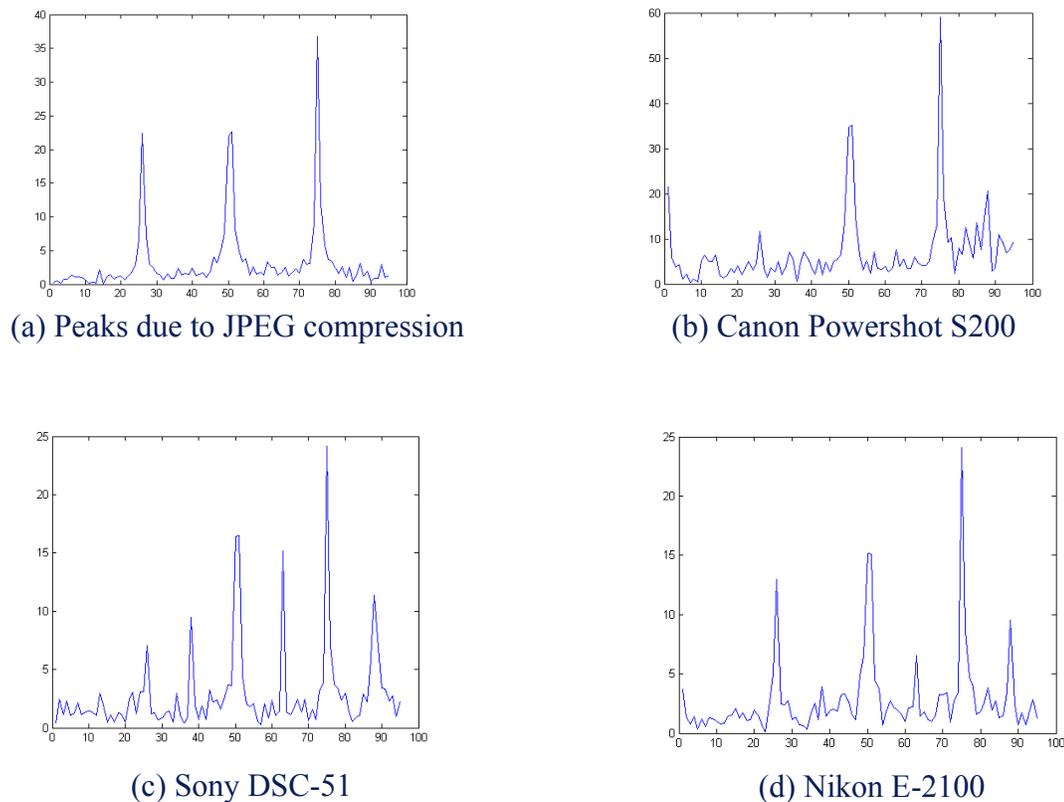


Figure-3. Frequency spectrum of averaged second order derivatives corresponding to JPEG compression and the three models of digital cameras with JPEG output images.

In the first part of our experiments, we have used two camera models: Sony DSC-P51 and Nikon E-2100. The two cameras have both a resolution of 2 mega-pixels. The pictures are of size 1600x1200 pixels and are obtained with maximum resolution, auto-focus, and other settings at default values. In order to reduce the dependency on the scenery being viewed, we used pictures that were taken from the same scene by two

cameras. A picture data set was made by obtaining 140 pictures from each model. One third of these images were used for training. Then the designed classifier is used in classifying the previously unseen 2/3 of the images. We used 75x75 pixel parts of the images for experiments. Based on the variance of each block the image is partitioned into smooth and non-smooth parts by an exhaustive search.

First we extracted features assuming a 3x3 interpolation kernel for both Sony and Nikon digital cameras. The accuracy is measured as 89.3%. Then, we extracted the features considering a neighboring 4x4 pixels. Correspondingly the accuracy in detection increased to 92.86%. The same experiment is repeated for 5x5 neighborhoods which lead to an accuracy of 95.71%. The corresponding confusion matrices are given in Tables 1, 2, and 3, respectively. As seen from the tables accuracy improves with larger kernel sizes. These results suggest that the actual size of the interpolation kernel used for CFA interpolation is not smaller than the considered sizes which were empirically known to be true [1]. Similar performance results are also obtained from smooth image parts using the features based on periodicity in the second order derivatives. Table 4 displays the accuracy for the two camera case. It is seen that the latter set of features do not prove as reliable as the former set of features.

Table 1. The confusion table for 2 cameras assuming a 3x3 interpolation kernel

| | | Predicted | |
|--------|-------|-----------|-------|
| | | Nikon | Sony |
| Actual | Nikon | 95.71 | 4.29 |
| | Sony | 17.14 | 82.86 |

Table 2. The confusion matrix for 2 cameras assuming a 4x4 interpolation kernel

| | | Predicted | |
|--------|-------|-----------|-------|
| | | Nikon | Sony |
| Actual | Nikon | 91.43 | 8.57 |
| | Sony | 5.71 | 94.29 |

Table 3. The confusion matrix for 2 cameras assuming a 5x5 interpolation kernel

| | | Predicted | |
|--------|-------|-----------|-------|
| | | Nikon | Sony |
| Actual | Nikon | 94.64 | 5.36 |
| | Sony | 3.57 | 96.43 |

Table 4. The confusion matrix for 2 cameras based on periodicity in the second-order derivative

| | | Predicted | |
|--------|-------|-----------|-------|
| | | Nikon | Sony |
| Actual | Nikon | 86.86 | 13.13 |
| | Sony | 23.33 | 76.66 |

In order to see how the proposed features perform for the case of three-cameras, we also obtained a set of images acquired by a Canon Powershot S200. In this case, the images were downloaded from internet and consist of different sceneries. In a similar manner, we extracted the features described in Sections 3.1-2 and used SVM and SFSS to classify three cameras. When features are extracted from 5x5 neighborhoods, the accuracy is measured as 83.33%, and corresponding confusion matrix is provided in Table 5. When attempted to discriminate cameras on the basis of features obtained from smooth image parts, the accuracy dropped to 74.3% as shown in Table 6.

Table 5. The confusion table for 3 cameras assuming a 5x5 interpolation kernel

| | | Predicted | | |
|--------|-------|-----------|-------|-------|
| | | Nikon | Sony | Canon |
| Actual | Nikon | 85.71 | 10.71 | 3.57 |
| | Sony | 10.71 | 75 | 14.28 |
| | Canon | 0 | 10.71 | 89.28 |

Table 6. The confusion table for 3 cameras based on periodicity in the second-order derivative

| | | Predicted | | |
|--------|-------|-----------|-------|-------|
| | | Nikon | Sony | Canon |
| Actual | Nikon | 76.78 | 8.92 | 14.28 |
| | Sony | 12.5 | 76.78 | 10.71 |
| | Canon | 19.64 | 10.71 | 69.64 |

Finally, we have combined the two sets of features and repeated the same experiment. In this case the accuracy of discrimination has increased to 96% for the three camera case as shown in Table 7. The increase in the accuracy indicate that the two sets of features capture different characteristics of an image, thereby enabling better identification of the source camera-model.

Table 7. The confusion table for 3 cameras corresponding to combined set of features

| | | Predicted | | |
|--------|-------|-----------|-------|-------|
| | | Nikon | Sony | Canon |
| Actual | Nikon | 94.78 | 1.50 | 3.72 |
| Actual | Sony | 2.08 | 95.28 | 2.64 |
| | Canon | 0 | 2.26 | 97.74 |

4. CONCLUSIONS AND FUTURE WORK

In this paper, we attempt improve our previous approach to source camera-model identification problem. To detect traces of color interpolation (artifacts) in the RGB color channels, we incorporate a number of features tuned to capture the periodicity in the second-order derivatives with the features obtained through using EM algorithm [2]. A classifier is then designed using the combined set of features and tested to determine the reliability of the selected features in discriminating the source camera-model among two and three cameras. This method is, limited to images that are not heavily compressed as

the compression artifacts suppress and remove the spatial correlation between the pixels due to CFA interpolation.

6. REFERENCES

- [1] M. Kharrazi, H. T. Sencar, and N. Memon, "Digital Camera Model Identification," *Proc. of IEEE ICIP*, 2004.
- [2] S. Bayram, H. T. Sencar, and N. Memon, "Source Camera Identification Based on CFA Interpolation," *Proc. of IEEE ICIP*, 2005
- [3] Special Issue on Data Hiding, *IEEE Transactions on Signal Processing*, Vol. 41, No. 6, 2003.
- [4] I. Avcibas, N. Memon and B. Sankur, "Steganalysis using Image Quality Metrics," *IEEE Transactions on Image Processing*, Jan. 2003.
- [5] S. Lyu and H. Farid, "Detecting Hidden Messages Using Higher-Order Statistics and Support Vector Machines," *Proc. of Information Hiding Workshop*, 2002
- [6] J. Lukas, J. Fridrich, and M. Goljan, "Determining Digital Image Origin Using Sensor Imperfections," *Proc. of IS&T SPIE*, vol 5680, 2005
- [7] A. Popescu and H. Farid, "Exposing Digital Forgeries by Detecting Traces of Re-sampling," *IEEE Transactions on Signal Processing*, 2004.
- [8] A. C. Gallagher, "Detection of Linear and Cubic Interpolation in JPEG Compressed Images," *Proc. of CRV'05*, 2005.
- [9] J. Adams, K. Parulski and K. Sapulding, "Color Processing in Digital Cameras," *IEEE Micro*, Vol. 18, No.6, 1998.
- [10] Todd Moon, 'The Expectation Maximization Algorithm', *IEEE Signal Processing Magazine*, November 1996.
- [11] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," 2001, Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>