INTER-CHANNEL DEMOSAICKING TRACES FOR DIGITAL IMAGE FORENSICS

John S. Ho, Oscar C. Au, Jiantao Zhou, and Yuanfang Guo

The Hong Kong University of Science and Technology
Email: ee_hjs@stu.ust.hk, eeau@ust.hk, eejtzhou@ust.hk, eeandyguo@ust.hk

ABSTRACT

Digital image forensics seeks to detect statistical traces left by image acquisition or post-processing in order to establish an images source and authenticity. Digital cameras acquire an image with one sensor overlayed with a color filter array (CFA), capturing at each spatial location one sample from the three necessary color channels. The missing pixels must be interpolated in a process known as demosaicking. This process is highly nonlinear and can vary greatly between different camera brands and models. Most practical algorithms, however, introduce correlations between the color channels, which are often different between algorithms. In this paper, we show how these correlations can be used to construct a characteristic map that is useful in matching an image to its source. Results show that our method employing inter-channel traces can distinguish between sophisticated demosaicking algorithms. It can complement existing classifiers based on inter-pixel correlations by providing a new feature dimension.

***Keywords*—** Image forensics, demosaicking, inter-channel correlation

1. INTRODUCTION

From the moment a scene is formed on the sensor, a digital image is subject to an intricate sequence of processing before it arrives at a form useable to the consumer. Characterization of these processes is of great interest to the forensic analyst because they are often unique to the camera brand, model, or even individual sensor. The demosaicking step arises from the fact that most digital cameras capture a scene with a single sensor in order to reduce cost and complexity. A color image, however, consists of three channels red (R), green (G), and blue (B). To address this, the sensor is overlayed with a pattern known as a color filter array (CFA) that spatially multiplexes the three channels. The most popular design used to achieve this is the Bayer pattern shown in Figure 1, where the green channel is sampled twice as often as red and blue. Demosaicking interpolates these missing samples in order to arrive at a full color image.

The simplest way to do this is to low-pass filter each color channel independently. While this may be satisfactory in smooth regions, simple filtering causes objectionable color artifacts in regions of detail [1]. Demosaicking methods of greater sophistication make use of other color channels in the reconstruction of any given sample. This is possible because color images are similar in high frequency components [2]. Practical algorithms will incorporate color channel dependencies in the demosaicking process, particularly in steps like gradient estimation [3]. Many other techniques exist, such as those employing wavelets and higher-order statistics, that dramatically improve over simple bilinear interpolation [4]. For these methods, the model of demosaicking as a linear filter becomes increasingly difficult to fit.

The problem of using detecting demosaicking traces has received much recent attention. The general idea of most methods is to characterize the correlation introduced between neighboring pixels when missing values are interpolated. An EM algorithm was proposed in [5] to determine if each sample pixel is either a linear combination of a fixed neighborhood or uncorrelated to its neighbors. Periodic patterns arising in the resulting probability map act as evidence of demosaicking. A similar idea was exploited in [6] by looking at the variance of values along the diagonals to determine if an image is photographic or computer-generated. In both cases, periodicity was established by observing peaks in the Fourier domain.

Such traces can be extended to perform source identification. The authors of [7] use the coefficients produced by the EM algorithm as one of the features to perform classification. In [8], an image is separated into three different gradient regions. By assuming that each has been interpolated by a separate linear filter, the coefficients can be obtained by solving the resulting linear equations. Quadratic pixel correlation was used in [9]. Although inter-pixel correlation arising from demosaicking is a major factor, the best results in source identification have been obtained by using a mix of features [10] [11] [7].

We propose using inter-channel, rather than inter-pixel, correlation as an additional feature to aid classification. Our observation is that in trying to take advantage of natural dependencies between the color channels, different demosaicking algorithms introduce their own correlations. Because the exact nature of a scene's inter-channel correlation is never known from the CFA image alone, demosaicking algorithms must impose their own model, leaving a distinct "fingerprint" for each algorithm on the image. This correlation can be readily observed in the frequency domain.

In this paper, we will use four demosaicking algorithms: bilinear, edge-directed with constant hue [3], projection-onto-convex-sets (POCS) [2], and adaptive filtering (AF) [12]. Bilinear simply uses the bilinear kernel to independently interpolate each channel. The constant hue model will be described

978-1-4244-7493-6/10/$26.00

G	R	G	R	
B	G	B	G	$\ldots$
G	R	G	R	
B	G	B	G	$\ldots$
	$\vdots$		$\vdots$	$\ddots$

Fig. 1. Bayer sampling pattern. Note that there are twice as many G samples.

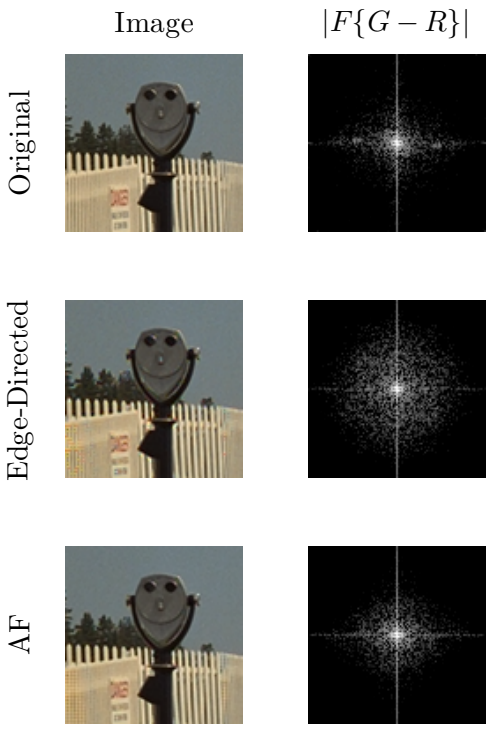

Fig. 2. Color spectrum difference of original image and the interpolated versions of the image after it has been sampled on the Bayer pattern.

in Section 2.1. POCS and AF are more recent state-of-the-art algorithms; a detailed description is beyond the scope of this paper.

2. INTER-CHANNEL SPECTRAL TRACES

2.1. Constant Hue Example

Consider a color image composed of color channels R, G and B. Given any natural image, an assumption made by most demosaicking methods is that high frequencies between luminance G and chrominance R, B components are largely identical [2] [13]. This is important because after sampling on the CFA, the G channel has twice the number of samples compared to R or B and is hence relatively free of aliasing. This assumption is often applied to the gradient estimation step, where by using other channels to estimate an edge, the algorithm implicitly assumes that the high frequencies are identical.

We will consider only G and R in this section, as B may be treated in the same manner. We will denote the R and G channels sampled at the R positions as in Figure 1 with R_s and G_s respectively, filling in zeros such that the image size is unchanged. Assume that low pass filtering R_s gives exactly the low frequencies of R, which we denote R^l.

A simple demosaicking scheme that introduces inter-channel correlation is the color difference interpolation scheme. It makes use of the constant hue assumption, which states that the color difference components are perfectly smooth within image objects [1]. An immediate consequence is that if one color channel is fully available, the others can be recovered by low-pass filtering the difference plane. The available channel is usually G, since it is sampled twice as often as either R or B and better preserves the high frequencies. More formally, if we assume that the full G is available by some interpolation process, R can be recovered via

$$R = \mathfrak{I}\{R_s - G_s\} + G \tag{1}$$

where $\mathfrak{I}\{\cdot\}$ denotes linear low-pass filtering. An interesting interpretation of (1) is that R "copies" the high frequencies of G [12]. This can be seen by decomposing G into high and low frequency components $G^h + G^l$ and noting that the low-pass nature of the filter implies that $\mathfrak{I}\{R_s - G_s\} = R^l - G^l$. The final sum cancels G^l, yielding $R = R^l + G^h$. In practice, the constant hue assumption does not hold perfectly so R and G are merely approximations of the original.

In a forensics setting, the interpolated R and G color channels are available. An interesting effect shown in Figure 2 arises when the color difference plane $G - R$ is computed. The spectrum of the difference plane for an image interpolated with two algorithms is clearly distinct, even though the original images are almost visually indistinguishable. Note that this is true only because the selected algorithms introduce different inter-channel correlations. It would not be possible, for example, to reliably distinguish between bilinear and bicubic interpolation.

This effect can be seen by rearranging (1) and applying the two dimensional Discrete Fourier Transform (DFT), which yields

$$\hat{G} - \hat{R} = \alpha(\hat{G}_s - \hat{R}_s) \tag{2}$$

where α is the frequency response of $\mathfrak{I}$, and $\hat{G}, \hat{R}$ is the DFT of G, R respectively. Define $\Delta \triangleq \hat{G} - \hat{R}$, which we refer to as the color difference spectrum.

Consider now the case that each frequency component of the image is random. Let Δ_{uv} be the uv^{th} entry of Δ and ρ_{uv} be the correlation coefficient between $\hat{R}_{uv}$ and $\hat{G}_{uv}$, which are assumed to have the same underlying variance σ^2_{uv}. We are interested in the effect the constant hue scheme has on the underlying statistics of the image. From (2), the frequency component of the difference plane at position (u, v) has variance given by

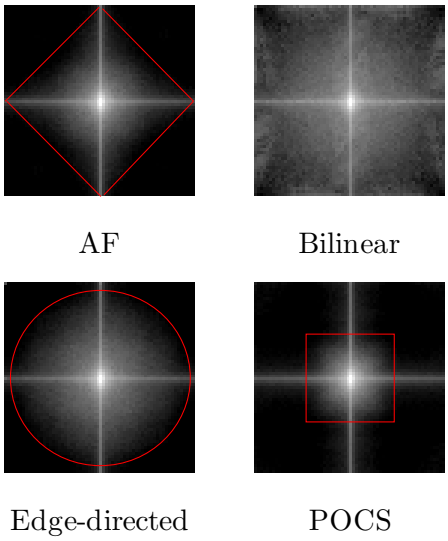

Fig. 3. Variance maps (v-maps) obtained from the same original image sampled on the Bayer pattern and interpolated using four different demosaicking algorithms. The distinct shapes of the v-maps have been outlined for clarity.

$$\begin{aligned}\text{Var}\{\Delta_{uv}\} &= \alpha_{uv}^2 \text{ Var}\{(\hat{G}_s - \hat{R}_s)_{uv}\} \\ &= 2\,\alpha_{uv}^2\,(1-\rho_{uv})\,\sigma_{uv}^2 \end{aligned} \tag{3}$$

where we have used the fact that $\text{Var}\{x-y\} = \sigma_x^2 + \sigma_y^2 - 2\sigma_x\sigma_y\rho_{xy}$. The effect of constant hue demosaicking is to introduce an extra factor α_{uv} to the underlying variance of the color difference plane. Because α corresponds to a low-pass filter, it leaves a salient trace on the variance of the color spectrum difference as a function of (u,v) at the high frequencies where $\alpha_{uv} \approx 0$. By measuring the variance of each component across many difference planes, such as the one in Figure 2, an estimation of Δ, which contains this trace term α, can be obtained.

Although other demosaicking algorithms may not be so easily modeled, the "copying" of frequencies between color channels can still be interpreted in terms of (1) for a inter-channel correlation term α. As such, the effect introduced on the image variance map can be described and measured in a similar manner.

2.2. Variance Map Estimation

Given any input image, we measure a $N \times N$ map of the variance of each frequency component. We compute the two color difference planes $G-R$ and $G-B$ for the demosaicked color image of interest, and then divide each plane into a set of subimages consisting of size $N \times N$ blocks. This constitutes as a trade-off between number of available sample images and the dimension of variance map. B is assumed to have the same set of variances σ as G and R, which allows blocks from both planes to be combined. This results in a sequence of M blocks.

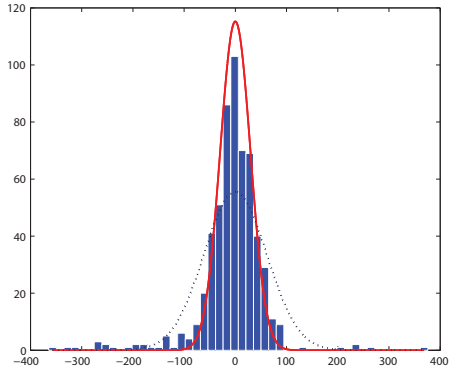

Fig. 4. Histogram of a typical high-frequency color difference spectral component across 576 sample blocks. A normal distribution fit before (dotted) and after (solid) mixture separation is shown.

Let $\Omega_i(u,v)$ be the magnitude of the DFT of the blocks $i = 1,\ldots,M$. We can compute a $N \times N$ sample variance map as follows

$$\Sigma_s(u,v) \triangleq \underset{i}{\text{Var}}\,\{\Omega_i(u,v)\} \tag{4}$$

For large M, Σ_s estimates the Δ given in (3). Such a measure of variance, however, is sensitive to outliers and is highly dependent on image content. To address this problem, we propose measuring the variance robustly by using a subset of the available samples. For notational convenience, we will consider a fixed location (u,v) at each block and denote the corresponding M samples at that location with x_i, dropping the indices u and v. It will be understood that operations should be repeated for all u and v (N^2 times).

We model the source of the samples x_i as a mixed distribution consisting of two classes: C_1 containing components that are correlated in the sense that samples are drawn from a zero mean normal distribution of variance σ_c, and C_2 containing uncorrelated "outlier" components. We assume that the uncorrelated components are drawn from a uniform distribution. This model well describes the distribution of the magnitude of high frequency components in natural images, as shown in Figure 4.

The parameter σ_c can be computed by mixture separation. At the n^{th} iteration the probability that a sample belongs to class C_1, is given by Bayes' theorem

$$P[x_i \in C_1 | x_i] = \frac{P[x_i | x_i \in C_1] P[x_i \in C_1]}{\sum_{k=0}^{2} P[x_i | x_i \in C_k] P[x_i \in C_k]} \tag{5}$$

We select the unbiased prior $P[x_i \in C_1] = {}^1/_2$ and $P[x_i \in C_2] = {}^1/_2$. According to the models, we also have

$$P[x_i | x_i \in C_1] = \frac{1}{\sigma_c^{(n)}\sqrt{2\pi}} e^{\left(x_i/\sigma_c^{(n)}\right)^2} \tag{6}$$

$$P[x_i|x_i \in C_2] = \frac{1}{\max_j\{x_j\} - \min_j\{x_j\}} \quad (7)$$

$\sigma_c^{(n)}$ is then updated as the weighted sample variance as follows

$$\sigma_c^{(n+1)} = \left(\frac{\sum_{i=0}^{M} \omega_i x_i^2}{\sum_{i=0}^{M} \omega_i} \right)^{1/2} \quad (8)$$

where $\omega_i \triangleq P[x_i|x_i \in C_1]$. The separation algorithm iterates between these two steps until the stopping condition $|\sigma_c^{(n+1)} - \sigma_c^{(n)}| \leq \epsilon$ is reached. Repeating this at each location (u, v), the output of the algorithm is the $N \times N$ variance map $\Sigma(u, v)$, which we refer to as the v-map of an image. As the v-map generally contains entries that differ by orders of magnitude, it is convenient to convert it to the log scale. Four of such v-maps are shown in Figure 3.

2.3. Classification with v-maps

The v-maps are normalized before comparisons are made between them. This avoids highly different scores between images with high and low detail. We do this by simply adding a constant term to each map such that the mean of the whole map is all equal to an arbitrarily chosen value. Because the dimension of the v-map is usually large (N^2), we opt for a simple nearest neighbor classification scheme. Let $\{\chi_{ij} \mid i = 1, \ldots, K, j = 1, \ldots, N\}$ be a set of training maps for N demosaicking algorithms, each with K training images. Given normalized input v-map Σ_N, the classifier uses the rule

$$(i^*, j^*) = \arg\min_{i,j} \left\{ \sum_{u,v} [\Sigma_N(u, v) - \chi_{ij}(u, v)]^2 \right\} \quad (9)$$

and assigns the input map Σ_N the class i^*. This corresponds to a single nearest neighbor classifier using Euclidean distance. While more sophisticated classifiers are certainly available, we have found that for small N, v-maps are usually sufficiently distinct for nearest neighbor to yield accurate classification.

3. RESULTS

We deploy the algorithm described in Section 2 for source identification. For v-map estimation, we use block sizes $N = 100$ with 50% overlap for synthetic tests and no overlap for the larger camera acquired images. The variance estimation algorithm is initialized with $\sigma_c^{(0)}(u, v)$ equal to half of the overall variance $\Sigma_s(u, v)$ at that position. The stopping condition is chosen to be $\epsilon = 0.01$.

3.1. Synthetic Classification

We test the v-map's ability to correctly classify a known demosaicking algorithm. We use the Kodak PhotoCD test set of 24 images, standard in demosaicking literature. They are each artificially sampled on the Bayer pattern, and then interpolated using four algorithms. We use four of the original images to build a 4×4 training set and the remaining 80 images are inputs to the classifier.

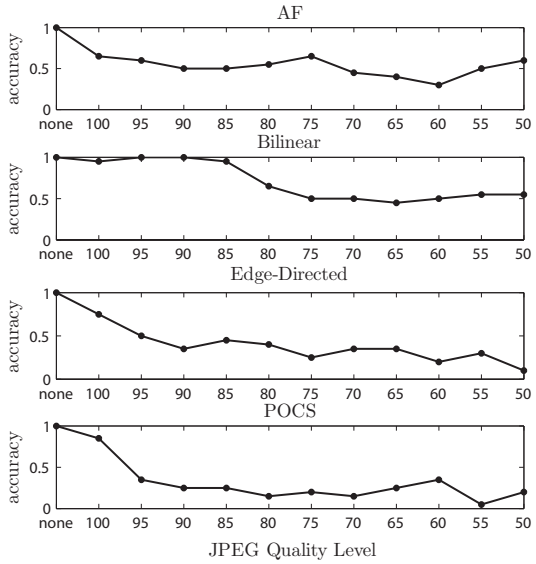

Fig. 5. Classification accuracy as a function of JPEG quality level, ranging from no compression to 100 through 50.

	Predicted			
Actual	Nikon	Sony	Canon 1	Canon 2
Nikon	**0.920**	0.040	0	0.040
Sony	0.040	**0.960**	0	0
Canon 1	0	0.040	**0.920**	0.040
Canon 2	0	0	0.040	**0.960**

Fig. 6. Confusion matrix for 4 commercial cameras: Nikon D70, Sony A100, Canon SD450 (1), and Canon 950IS (2).

In Figure 5, the classification accuracy is plotted as both the training and test images are subjected to various levels of JPEG compression. With no compression, classification is perfect. The results are shown to vary depending on the particular demosaicking method. While our method is able to distinguish bilinear from the methods that introduce correlation through moderate levels of JPEG compression, JPEG compression can cause confusion among the more sophisticated methods (AF and POCS). This is due to the fact that JPEG removes the high frequencies that make some of the maps distinct.

3.2. Commercial Digital Cameras

Given that demosaicking algorithms are different between camera brands and models, we can also use the v-map to classify images by its source camera. Among commercial cameras, we have observed that v-maps are usually sufficiently distinct to be robust against moderate JPEG compression. We four cameras in our experiment: Nikon D70, Sony A100, Canon SD450, and

Fig. 7. Representative test digital camera images captured different brands and models with the corresponding normalized maps.

Canon 950IS. The first two are DSLR cameras. We capture 50 images at the default JPEG quality level from each camera and use 25 of them to form the training set. Interestingly, we have found that the demosaicking algorithms used by commercial cameras often treat the vertical and horizontal directions differently. This causes a non-symmetric v-map; several such examples can be seen in Figure 7. To correct for this, we rotate the images prior to v-map estimation to ensure that the orientations agree.

Figure 6 shows the resulting confusion matrix. Despite the presence of moderate JPEG compression, the classifier has an average accuracy of 94.5%, comparable to methods using only inter-pixel correlations. This suggests that commercial cameras employ demosaicking algorithms more dissimilar than those selected for the synthetic experiment.

A practical classifier will make use of more features in making the decision and we expect better results by incorporating other possible characteristics, such as in [11]. The results suggest, however, that inter-channel correlation is a highly capable feature for source identification.

4. CONCLUSION

We have demonstrated how this correlation can be introduced for the simple case of constant hue interpolation and developed a method to measure the it through the variance of color difference planes. Using the resulting v-maps for classification yields promising results, even when distinguishing between highly complex demosaicking schemes. We are currently extending our method to perform tamper detection as well. Inter-channel correlation offers a complementary approach towards past work exploiting inter-pixel correlations introduced by demosaicking. We explicitly measure inter-channel dependencies and use the resulting v-map to characterize the process. We expect our method to be a valuable addition to the forensic toolbox.

5. REFERENCES

[1] B. K. Gunturk, J. Goltzbach, Y. Altunbasak, R.W. Schafer, and R. M. Mersereau, "Demosaicking: Color filter array interpolation," *IEEE Signal Process. Mag.*, vol. 22, pp. 44–54, Jan. 2005.

[2] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Color plane interpolation using alternate projections," *IEEE Trans. Image Process.*, vol. 13, pp. 997–1013, Sept. 2002.

[3] J. E. Adams, "Design of practical color filter array interpolation algorithms for digital cameras," *Proc. SPIE*, vol. 3028, pp. 117–125, Feb. 1997.

[4] X. Li, B. Gunturk, and L. Zhang, "Image demosaicing: A systematic survey," *Proc. SPIE*, vol. 14, no. 6822, Jan. 2008.

[5] A.C. 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.

[6] A. Gallagher and T. Chen, "Image authentication by detecting traces of demosaicing," *Proceedings of CVPRW*, 2008.

[7] S. Bayram, H. T. Sencar, and N. Memon, "Improvements on source camera-model identification based on cfa interpolation," *Proc. of the WG 11.9 Intl. Conference on Digital Forensics*, Jan. 2006.

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

[9] Y. Z. Huang and Y. J. Long, "Demosaicking recognition with applications in digital photo authentication based on a quadratic pixel correlation model," *Proceedings of CVPR*, pp. 1–8, 2008.

[10] M. Kharrazi, H. T. Sencar, and N. Memon, "Blind source camera identification," *Proceedings of ICPC*, pp. 709–712, 2004.

[11] C. McKay, A. Swaminathan, H. Gou, and M. Wu, "Image acquisition forensics: Forensic analysis to identify imaging source," *Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing*, pp. 1657–1660, 2008.

[12] N. Lian, L. Chang, Y.P. Tan, and V. Zagorodnov, "Adaptive filtering for color filter array demosaicking," *IEEE Trans. Image Process.*, vol. 16, pp. 2515–2525, Oct. 2007.

[13] D. Alleyson, S. Susstrunk, and J. Herault, "Linear demosaicking inspired by the human visual system," *IEEE Trans. Image Process.*, vol. 14, pp. 439–449, Apr. 2005.