NEW FEATURES TO IDENTIFY COMPUTER GENERATED IMAGES

A. E. Dirik, S. Bayram

Polytechnic University
Electrical & Computer Engineering Dept.
Brooklyn, NY, 11201, USA

H. T. Sencar, N. Memon *

Polytechnic University
Computer & Information Science Dept.
Brooklyn, NY, 11201, USA

ABSTRACT

Discrimination of computer generated images from real images is becoming more and more important. In this paper, we propose the use of new features to distinguish computer generated images from real images. The proposed features are based on the differences in the acquisition process of images. More specifically, traces of demosaicking and chromatic aberration are used to differentiate computer generated images from digital camera images. It is observed that the former features perform very well on high quality images, whereas the latter features perform consistently across a wide range of compression values. The experimental results show that proposed features are capable of improving the accuracy of the state-of-the-art techniques.

***Index Terms*—** Digital image forensics,computer generated, Bayer filter, demosaicking, chromatic aberration

1. INTRODUCTION

A great deal of progress has been made in both fields of computer vision and computer graphics and these two fields have now begun to converge very rapidly. Consequently, more realistic synthetic imagery became achievable. However, this trend has brought with it new issues and challenges concerning the authenticity of digital images. The fact that digital images can now be easily created undermines the credibility of digital images presented as evidence in a court of law since now it is difficult to distinguish whether an introduced digital image is a depiction of real-life occurrences and objects or a synthetically generated one. Today, there is a severe lack of techniques and methodologies for addressing this sort of a problem. In this context, digital image forensics is concerned with determining the origin and potential authenticity of a digital image. Although determining the origins of an image involves multitude of problems, distinguishing photorealistic computer generated (PRCG) images from real images is the most challenging and immediate one. Therefore, several approaches have been proposed to address this problem. Essentially, almost all of the proposed approaches are based on machine learning methods in which relationships between variables (features), extracted from a set of sample images are generated automatically and expressed in the form of classifiers. These classifiers are later used to differentiate between the two types of images. Hence, the main difference among various proposed methods lies in the features they deploy in constructing the classifiers. Another difference is the image sets deployed in extraction of the features and how well their characteristics represent those of the corresponding class of images.

The first approach to differentiating natural images from PRCG images was proposed by Farid et al. [1] to model natural images. In this technique, the features are designed to capture the statistical regularities of natural images in terms of the first and higher-order statistics of wavelet transform coefficients. In [2], Ng et al. proposed another promising approach based on identifying the distinctive characteristics of PRCG images, as compared to natural images. Their technique takes into account the differences in surface and object models and differences in the acquisition process between the synthetic and real images. Another wavelet transform based method was proposed by Wang et al. [3] in which features are obtained through filtering sub-band histogram coefficients. More recently, Dehnie et al. [4] presented an approach that discriminates synthetic images from digital camera images based on the lack of artifacts due to use of a digital camera as an acquisition device. Among these approaches the former two has the best reported performance, both yielding roughly 80 percent accuracy rate in successful identification.

In this paper, first we attempt to detect the presence of the color filter array demosaicking from a given image. Most consumer cameras in the market use color filter array (CFA) which requires the involvement of a demosaicking operation in generating the RGB color values. The interpolation process can be approximated as a linear filtering and can be characterized in terms of the filter coefficients. In [5, 6], it is showed that source digital camera-model can be detected by estimating CFA demosaicking filter coefficients. Since most real images are created by digital cameras, it is expected that they exhibit traces of CFA interpolation, whereas PRCG images should not. Motivated by this, we deploy a similar approach

*This work is supported by National Institute of Justice grant 2006-92251-NY-IJ

to distinguish real and PRCG images. However, instead of estimating the CFA interpolation coefficents, we try to detect the presence of CFA interpolation. In our method, we first introduce 4 new features characterizing the presence of a CFA interpolation. Then we incorporate these with the features of [1].

Another feature we consider is the detection of the presence of chromatic aberration in a given image[7] . Chromatic aberration reveals itself as anomalous color shifts due to variation of refractive index of the optical glass formulation used in manufacturing lenses. Chromatic aberration essentially causes a misalignment between color channels of the image. Since PRCG images are not a product of an optical system, they won't inherently suffer from such an artifact. Therefore, it is possible to detect whether an unmanipulated image is PRCG or not based on these aberrations. The second set of features measure the misalignment between color channels to enable classification between camera images and computer generated ones. Our experimental results show that with the inclusion of the new features the performance of [1] is further improved.

The outline of the paper is as follows: In Section 2, we describe in detail the newly proposed features used for distinguishing PRCG images. We provide implementation details and experimental results in Section 3 and conclude in Section 4.

2. DESCRIPTION OF FEATURES

2.1. Bayer Pattern Detection

Let $S_c(x, y)$ be the image intensity of color channel c at spatial location (x, y) and $c \in \{R, G, B\}$. Since each color filter array (CFA) is sensitive to a particular color band, the missing color components are interpolated by using the adjacent pixel colors. Let $\Psi_{k,c}$ represent the set of color filter array locations of channel c for a type of CFA pattern denoted by k. The color filter mask for a particular channel is defined as:

$$M_{k,c}(x, y) = \begin{cases} 1, & (x, y) \in \Psi_{k,c} \\ 0, & \text{otherwise} \end{cases}$$

Assuming 4×4 CFA conuration, there are 36 different filter arrangements. However, since human eye is more sensitive to the green channel, the number of green channel filter elements in a CFA array is usually twice as many as red and blue colors, and most of these CFA arrangements are not suitable for use. In this paper, we restrict the maximum value of k with 4 since most digital cameras use the Bayer CFA pattern in which the green color filters are placed on the diagonal. The 4 different CFA patterns used in this paper is given in fig. 1.

Denoting $I'_c(x, y)$ as the output of CCD sensors for three different channels and f as the demosaicking function, the interpolated image I before post processing operations is defined as: $I = f(I', M_k)$, and $I'_c(x, y) = S_c(x, y) \cdot M_{k,c}(x, y)$

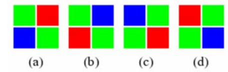


Fig. 1. Different Bayer CFA patterns, M_k

where $\cdot$ refers to element-wise multiplication. The pillar of the proposed method is that if a given image, that has been initially interpolated with a Bayer demosaicking filter, is re-interpolated with a different kind of CFA, the right CFA pattern should yield significantly smaller mean squared error than others patterns. Besides, this property should reveal itself consistently all over the image. Therefore, If an image is not acquired with Bayer sampling or heavily distorted with compression than we would expect that all re-interpoation MSE values for different CFA patterns should represent a uniform distribution.

To estimate the presence of Bayer interpolation, the image is divided into $D \times D$ sub-blocks. Since the proposed CFA pattern detection method is based on the mean square error (MSE) variations depending on different kind of demosaicking methods, the non smooth blocks whose high freq. components' energy (pixel values' standard deviation) is above a certain threshold are used solely. For the sake of ease in the analysis, in this paper, we choose the demosaicking algorithm, f, to be bilinear interpolation.

We denoted each non-smooth block with B_i, where $i = 1, ..., N$. N is the number of non smooth blocks in a given image. The corresponding re-interpolated blocks, using filter k, is denoted with $\hat{B}_{i,k}$. Essentially, $\hat{B}_{i,k}$ is computed as a convolution between the bilinear kernel and the re-sampled block B_i with the kth CFA pattern M_k. The re-interpolation error of ith sub block for the kth CFA pattern is defined as: $\hat{B}_{i,k} = f(B_i, M_k)$, and $k = 1, ..., 4$.

The MSE error between the blocks B and $\hat{B}$ is computed in non smooth regions all over the image as

$$E_i(k, c) = \frac{1}{D \times D} \sum_{x=1}^{D} \sum_{y=1}^{D} (B_i(x, y, c) - \hat{B}_{i,k}(x, y, c))^2$$

E_i is a 4×3 matrix which comprises the mean square errors for each channel and CFA pattern. To detect the relative error distances between color channels, a new error matrix $E_i^{(2)}$ is created by normalizing all the rows of the E_i, as

$$E_i^{(2)}(k, c) = 100 \times \frac{E_i(k, c)}{\sum_{l=1}^{3} E_i(k, l)}, \quad c = 1, ..., 3.$$

Due to the higher presence of green channel elements, most significant MSE variations due to different CFA patterns are observed in green channels. Therefore, the green channel column of the normalized error $E_i^{(2)}$, $V_i(k)$, is used in extraction of some features which are directly correlated with CFA demosaicking operation. From $V_i(k)$ vector, it is possible to get the index of CFA pattern which yields the minimum MSE, as $P_1(i) = \mathrm{argmin}_k \ V_i(k)$. The pattern number which yields the second minimum MSE in the error matrix is stored in a separate P_2 vector as well. It is also expected that the values of V_i vector is not distributed uniformly in the presence of Bayer filter demosaicking. Thus, another feature can be derived to capture the uniformity of V_i vector such as

$$P_3(i) = \sum_{l=1}^{4} |V_i(l) - 25|$$

$P_1(i)$, $P_2(i)$, and $P_3(i)$ values are computed for all non smooth blocks. If the given image is interpolated with any demosaicking algorithm, histograms of P_1, P_2, and P_3 should concentrate in particular values which can be consequently used to detect demosaicking operation. Since P_1 and P_2 contain the estimated CFA patter number k and it takes values only from one to four, their histogram vectors would be four element vectors. These vectors are defined as

$$H_1(k) = \frac{100}{N} \sum_{l=1}^{N} \delta(P_1(l) - k), \quad k = 1, ..., 4$$

The histogram vector H_2 of P_2 is computed as same as H_1. By using the definitions above, the 4 main features for Bayer interpolation detection are defined as follows:

$$F_i = \frac{1}{10} \sum_{l=1}^{4} |H_i(l) - 25|, \quad i = 1, 2 \text{ and } \quad F_3 = median(P_3)$$

To compute the $4th$ feature, all E_i error matrices are averaged and the feature is computed over the green channel column vector of $\bar{E}$ such as $V_g(k) = \bar{E}(k, 2)$:

$$\bar{E} = \frac{1}{N} \sum_{i=1}^{N} E_i, \quad F_4 = \textstyle\sum_{k=1}^{4} \bar{E}(k, 2)/min(V_g)$$

2.2. Chromatic Aberration Detection

When white light passes through a simple or complex lens system, the component wavelengths are refracted according to their frequency. As a consequence, following Snell's law, in an image blue, red and green lights would be refracted differently, and a misalignment will occur. Although efficient utilization of chromatic aberration requires the knowledge of the optical center in an image, for the sake of simplicity we will assume optical center is located in the center of an image. To measure the misalignment, we use the mutual information between color channels. The basic idea here is that when the alignment between color channels is attained the mutual information will be maximized. Based on this, the image is upsampled by a factor of 100000 and mutual information is computed (for 110 shift locations) by radially shifting color channels with respect to each other. Since we assume there is no misalignment for the color channels of PRCG images the mutual information is expected to be maximized when there is no shift and it should reduce with shifting. On the other hand, for real images the mutual information should be maximized when the two colors are aligned. In our experiments, we observed that in a range of shifts the mutual information values reduce suddenly for PRCG images but it is constant for real images. Therefore we propose to use the variance of mutual information values in that range of the shifts. Hence, if the mutual information value is close to constant we classify the image as real and if it reduces we classify it as PRCG. This method would be less sensitive to compression since the misalignment between color channels will be much or less preserved during compression.

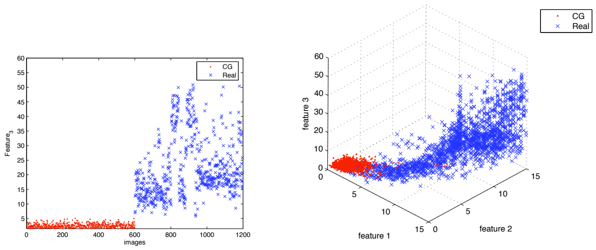
3. EXPERIMENTS

The Bayer features described above are expected to take relatively high values if the given image is obtained through a Bayer color filter array. On the contrary, if the given image is a PRCG image, these features tend to be zero. In a similar manner, the variances of the mutual information values are expected to be relatively high for PRCG images and close to zero for real ones.

To verify the performance of the proposed features 600 high quality JPEG images are taken from 5 different cameras. The brand and models of the cameras are Nikon D50, Canon Power Shot A80, Canon EOS Digital Rebel XT, Canon Power Shot S1 IS, and Sony DSC-S90. To build a PRCG image set 600 computer generated images are downloaded from the internet. Then proposed features are computed for every image in real and PRCG image sets. (The distribution of the F_3 for the high quality image set is given in fig. 2-a.) The image set is further expanded by including lightly and moderately compressed images. For this, the image sets are compressed at JPEG qualities of 95 and 90. At the end, 1800 different images are created for each class of PRCG and digital camera images. (The scattering diagram of F_1, F_2, and F_3 features for this larger image set is given in fig. 2-b.)

	False Pos	True Pos	Max Accuracy
Bayer (4f)	0.023	0.986	0.981
Chromatic (1f)	0.164	0.951	0.8933
Wavelet (72f)	0.003	0.996	0.996
Bayer+Chromatic (5f)	0.031	0.996	0.982
Bayer+Wavelet (76f)	0.001	1.000	0.999
All (77f)	0.002	1	0.998

Table 1. Accuracy rates of the proposed method, wavelet features, and fusion of the two methods



(a) (b)

Fig. 2. Scatter plot for (a) Bayer feature F_3 for CG and Real images, and (b) Bayer features for 1800+1800 real and CG images.

3.1. Classifier design

To compare the accuracy of the proposed Bayer features with the existing wavelet features [1] a SVM classifier is built. In this experiment, 1800 cg and 1800 real images, taken from five different cameras, are labeled as negative and positive respectively. Similarly, we set a threshold for variance of compute mutual information values and label each image with 0 if it is below the preset threshold and 1 if it is above the threshold. These labels are used in classifier design. Half of the image sets are used for cross validation and training and the other half is used for testing. All real images have maximum resolution, and the CG image set is not resized or manipulated after downloading. After obtaining the detection accuracies for 4 demosaicking based features and 1 chromatic aberration based feature, we derived 72 wavelet features from gray level images. The classifier is trained with the combination of all proposed features and their fusion with the wavelet features of [1].

The measured accuracies are given in Table 1 for high quality JPEG image set. The detection accuracy for the three sets of features are measured as 0.981, 0.893, and 0.996, respectively for demosaicing based features, chromatic aberration based feature, and wavelet features. The best results are obtained when demosaicking based features are combined with wavelet features with false positive and false negative rates are 0.001 and 0, respectively. (That is, all 900 real images from 5 different cameras in the test set were identified correctly as real.) On the other hand, when all three features are fused together the performance dropped to 0.998 due to a classification error in one of the real images. However, since both wavelet and demosaicking based features are sensitive to compression, these performance results are expected to decrease with increasing compression levels.

To test the reliability of chromatic abberration based feature against compression we compressed 600 original real and PRCG images at quality 50 and 70, in addition to existing compressed images at quality factors 90 and 95. These images are then pooled together, generating image sets of 3000 images for each class. The accuracy results are obtained by radially shifting green channel with respect to the red channel and applying a hypothesis testing procedure to measured variance values. Corresponding results are given in Table 2.

True Detected PRCG	True Detected Real	Accuracy
2852/3000	2552/3000	0.9

Table 2. Accuracy of the proposed chromatic aberration method for the images compressed with 50, 70, 90,95 and 100 quality factors.

4. CONCLUSION

In this paper, we investigate the problem of distinguishing photo-realistic computer generated and real images. We introduce new features to detect the presence of CFA demosaicking and chromatic aberration from a single image. We incorporate these new features to the previously proposed wavelet features. Our results show that proposed CFA detection scheme can be used to distinguish real (digital camera) from PRCG images. The detection accuracy is very promising for high quality real images, and accuracy decreased with higher degrees of compression. However this phenomenon can be utilized for post processing and image forgery detection as well. Although wavelet based features [1] perform satisfactorily alone, with the incorporation of proposed CFA features its performance is further improved on the used image set. In addition, it is shown that the chromatic aberration based feature performs relatively lower but consistently over a wide range of compression levels.

5. REFERENCES

[1] H. Farid and S. Lyu, "Higher-order wavelet statistics and their application to digital forensics," *Conference on Computer Vision and Pattern Recognition Workshop*, vol. 8, 2003.

[2] N. Tian-Tsong, C. Shih-Fu, H. Yu-Feng, X. Lexing, and T. Mao-Pei, "Physics-motivated features for distinguishing photographic images and computer graphics," in *ACM Multimedia*, Singapore, November 2005.

[3] Y. Wang and P. Moulin, "On discrimination between photorealistic and photographic images," in *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, May 2006, vol. 2, pp. II–161– II–164.

[4] S. Dehnie, H. T. Sencar, and N. Memon, "Identification of computer generated and digital camera images for digital image forensics," in *Proc. of IEEE ICIP*, 2006.

[5] S. Bayram, H. T. Sencar, N. Memon, and I. Avcibas, "Source camera identification based on cfa interpolation," in *Proc. of IEEE ICIP*, 2005.

[6] A. Swaminathan, M. Wu, and K. J. Ray Liu, "Nonintrusive component forensics of visual sensors using output images," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 1, pp. 91–106, March 2007.

[7] Warren J. Smith, "Modern optical engineering," in *3rd ed., McGraw-Hill*, 2000.