

MODELING THE EXIF-IMAGE CORRELATION FOR IMAGE MANIPULATION DETECTION

Jiayuan FAN, Alex C. KOT

Hong CAO

Farook SATTAR

School of Electrical and Electronic Engineering
Nanyang Technological University
{fanj0004, eackot}@ntu.edu.sg

Institute for Inforcomm Research
A*STAR Singapore
hcao@i2r.a-star.edu.sg

NEPTUNE Canada
University of Victoria
farooks@uvic.ca

ABSTRACT

EXchangeable Image File format (EXIF) is a metadata header containing shot-related camera settings such as aperture, exposure time, ISO speed etc. These settings can affect the photo content in many ways. In this paper, we investigate the underlying EXIF-Image correlation and propose a novel model, which correlates image statistical noise features with several commonly used EXIF features. By formulating each EXIF feature as a weighted combination of different image statistical noise features, we first select a compact image statistical noise feature set using sequential floating forward selection. The underlying correlation as a set of regression weights is then solved using a least squares solution. When applying our learned correlation to detect image manipulation, we achieve average test accuracies of 94.6%, 94.1% and 94.9% in three different cameras to detect the presence of common image brightness and contrast adjustment.

Index Terms— EXIF, Image manipulation, digital forensics

1. INTRODUCTION

With the advancement of photo editing tools, electronic alterations of digital images for deceiving purposes become an easy task. Existing works on image tamper detection have gained a lot of attention in the recent years. Different types of image regularities originated from different parts of digital still camera (DSC) system have been modeled and detected for forensic purposes, such as chromatic aberration [1] in optical system, photo response non-uniformity (PRNU) sensor pattern noise [2], demosaicing regularities [3, 4], and other statistical regularities [5, 6, 7, 8]. In this paper, we explore a novel correlation between image statistical noise features and EXchangeable Image File format (EXIF) header features for detecting image manipulation.

EXIF [9] is the DSC image file format standard for storing metadata, such as camera settings. It is widely supported by images in JPEG format and TIFF format as an international standard in the DSC system. As an example in Fig. 1,



(a) A Photo

EXIF Camera Settings	Value
ISO Speed Rating	100
Exposure Time	0.004
F-Number	5.6
...	...

(b) The EXIF header

Fig. 1. A photo with the corresponding EXIF header information.

shot settings such as ISO speed rating, exposure time and F -number are commonly recorded into the EXIF headers while the photo is captured. These shot settings can affect the photo content in many ways. For example, a large ISO speed rating generally provides good sensor's sensitivity especially under the low light condition and at the same time, leads to more visible image noise. To the best of our knowledge, this is the first forensic work that explores the potential EXIF-Image correlation for image forensic purposes, e.g. tamper detection.

The photon transfer curve (PTC) in [10] shows that the standard deviation of total sensor noise is correlated with digital camera signal intensity (DCSI), i.e. the amount of effective light impinging onto an image sensor during the shutter-opening period. Since DCSI is closely associated with some EXIF settings, e.g. aperture and exposure time, and the standard deviation of sensor noise can be evaluated from the photo content, potentially this correlation can be modeled and explored as EXIF-Image correlation for forensic purpose. Inspired by this new concept, we investigate and learn the correlation between several selected EXIF features and image statistical noise features as several sets of regression weights. When applied to detect image manipulation, our learned EXIF-Image correlation is demonstrated to be efficient in detecting the presence of common image brightness and contrast adjustment.

The rest of the paper is organized as follows. In Section 2, we model the correlation between images and EXIF headers. The learned correlation is used to detect image manipulation in Section 3. Section 4 concludes our work.

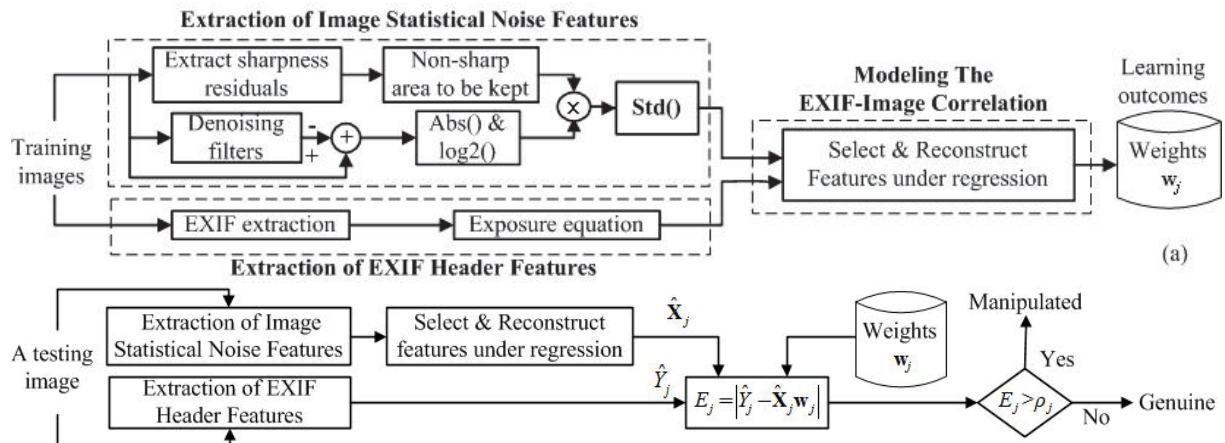


Fig. 2. Scheme of the proposed method. (a) Modeling the correlation between images and each of EXIF header features. (b) Detecting image manipulation using each of EXIF header features.

2. PROPOSED METHOD

Fig. 2(a) is the block diagram for modeling the EXIF-Image correlation. In the blocks labeled ‘Extraction of Image Statistical Noise Features’ and ‘Extraction of EXIF Header Features’, the procedures for extracting statistical noise features from the images and for extracting EXIF header features from the images are shown, respectively. The correlation of two types of features is learned in the block labeled ‘Modeling The EXIF-Image Correlation’. The detailed procedures are explained in the following section.

2.1. Extraction of image noise features

In view that the sharp area of the image tends to have high frequency energy which affects the noise estimation, we compute image statistical noise features in the non-sharp area of a given color image, A_l , where $l = 1, \dots, L$ and L is the total number of training images. A non-sharp area is defined by the sharpness residuals. By first converting A_l to a grayscale image, we convolute it with a 3×3 Gaussian filter to obtain its smoothed version. The sharpness residual image is computed as the absolute difference between the grayscale version of the image A_l and its smoothed version. Using a histogram, we plot the normalized distribution of the sharpness residual image in Fig. 3(a). The equation defining the sharp area and non-sharp area of the image is shown as follows:

$$B(p, q) = \begin{cases} 1 & \text{if } G(p, q) > r \\ 0 & \text{if } G(p, q) < r \end{cases} \quad (1)$$

where $G(p, q)$ is the sharpness residual at the location (p, q) and r is the threshold. In our work, r is empirically chosen to be the sharpness residual corresponding to 0.9 in the cumulative distribution of sharpness residuals. $B(p, q) = 1$ indicates that the location (p, q) is in the sharp area. Otherwise, (p, q) is in the non-sharp area.

We further expand the sharp area through dilation using a structuring element V of a unity mask of 3×3 as

$$M = B \oplus V \quad (2)$$

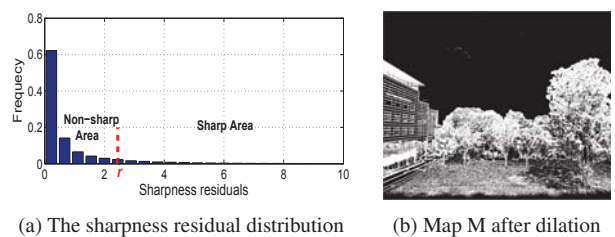


Fig. 3. An example of our computed sharp area and non-sharp area using the photo example in Fig. 1(a). Black area in (b) indicates non-sharp area to be included.

where M is the map indicating the sharp area as 1 and the non-sharp area as 0, as shown in Fig. 3(b). By expanding the sharp area through dilation, we ensure that our statistical noise features would not be affected by the image area with large sharpness residuals.

We then extract the noise features as follows. We use 4 different types of denoising filters including averaging filter, Gaussian filter, Median filter and Wiener filter [11]. All filter masks are of 3×3 pixels and for Gaussian filter, we choose 0 to be its mean and 0.5 to be its standard deviation. Different types of denoising filters are targeted to remove different types of noises. For instance, the median filter is effective for removal of salt and pepper noise.

Each denoising filter is convoluted with each color channel F of an image A_l to obtain the denoised color channel D . The noise residual $f(p, q)$ is computed using

$$f(p, q) = \log_2 |F(p, q) - D(p, q)| \quad (3)$$

The image statistical noise feature z is computed below within the non-sharp area using

$$z = \left(\frac{1}{R} \sum_{M(p,q)=0} (f(p, q) - \mu)^2 \right)^{\frac{1}{2}} \quad (4)$$

where $\mu = \frac{1}{R} \sum_{M(p,q)=0} f(p, q)$ is the mean, R is the number of pixels in the non-sharp area, $z \in \mathbf{z}_l$ and \mathbf{z}_l is a feature vector of 12 image statistical noise features corresponding to a total

of 12 different combinations of color channels and denoising filter types.

2.2. Extraction of EXIF header features

Camera settings expressed in the log scale are used to represent the EXIF header. The EXIF header feature, Y_{lj} , $j = 1, \dots, 5$, is extracted from the EXIF header for a given training image A_l , as given in [9].

$$\begin{aligned} \text{Aperture value} & Y_{l1} = \log_2(N^2) \\ \text{Shutter speed value} & Y_{l2} = \log_2(1/t) \\ \text{ISO speed rating value} & Y_{l3} = \log_2(I/3.125) \\ \text{Brightness value} & Y_{l4} = Y_{l1} + Y_{l2} - Y_{l3} \\ \text{Exposure value} & Y_{l5} = Y_{l1} + Y_{l2} \end{aligned} \quad (5)$$

where N is the F -number, t is the exposure time and I is the ISO speed rating. The above EXIF header features share a common unit of *stop*, where increment by 1 *stop* indicates the amount of light falling on the digital sensor is doubled.

2.3. Modeling the EXIF-Image correlation

As mentioned earlier, the PTC in [10] shows a correlation between standard deviation of sensor noise and DCSI. We observe from the PTC plot that when DCSI is represented in the log scale, the large central portion of the PTC tends to be linear alike. The logarithm of DCSI is linearly associated with our selected EXIF header features and the standard deviation of sensor noises can be represented by our noise features. In this subsection, we correlate each EXIF header feature with the image statistical noise features. Each EXIF header feature is formulated as weighted combination of our reconstructed noise features. For instance, the aperture value Y_{l1} of the l -th image is written as follows:

$$Y_{l1} = \sum_{x_{lu} \in \mathbf{x}_l} x_{lu} s_{u1} + e_{l1} \quad (6)$$

where e_{l1} is the regression error, s_{u1} is the weight of the u -th feature x_{lu} , $x_{lu} \in \mathbf{x}_l$ and $\mathbf{x}_l = \{z_{l1}, \dots, z_{lC}, \dots, z_{l1}z_{l2}, z_{l1}z_{l3}, \dots, z_{l(C-1)}z_{lC}, \dots, z_{l1}^2, \dots, z_{lC}^2\}$ is the reconstructed feature vector which includes both the first order and the second order polynomial terms of statistical noise features $\mathbf{z}_l = \{z_{lc}\}$. The second order polynomials are added to cater for the non-linearity of the underlying correlation, which can be further affected by the postprocessing pipeline of a DSC system.

After reconstructing the statistical noise features, the dimensions of the new features are high. Since not all features are equally important and some can even adversely affect the result, we use sequential floating forward selection (SFFS) [12] to reduce the feature dimension. Starting from an empty feature set, SFFS performs stepwise feature inclusion and conditional feature removal to select a subset of features to get the optimal feature subset. The criteria we minimize in SFFS is the root mean square regression error (RMSRE) of the estimated aperture values on the training images. During each step, the regression weights are estimated using a standard least squares solution in order to estimate the RMSRE.

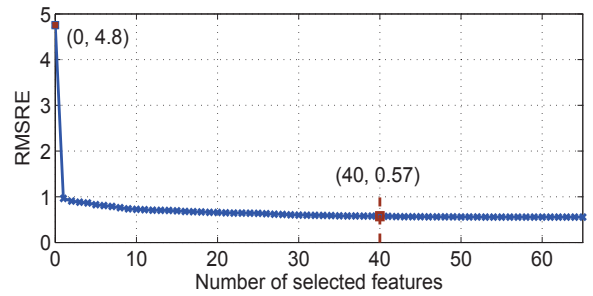


Fig. 4. RMSRE versus selected features number in SFFS for estimating aperture value.

Fig. 4 plots the RMSRE of the estimated aperture value versus the number of selected noise features. When no feature is selected, the RMSRE is about 4.8, which is also the root mean square of the aperture values of all the training images. With increasing number of selected features, the RMSRE gradually decreases. When the number of selected noise features exceeds 40, the RMSRE almost saturates at 0.57. Therefore, we choose $K=40$ features. In the SFFS selection, the reduction of the RMSRE from 4.8 to 0.57 shows that the aperture value can be efficiently predicted from the selected noise features using our proposed model.

After processing all training images, a set of compact features is obtained. Furthermore, considering the selected noise features, regression weights for the aperture value can be computed as follows:

$$\mathbf{w}_1 = (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{Y}_1 \quad (7)$$

where $\mathbf{X}_1 = \begin{bmatrix} x_{11} & \cdots & x_{1K} \\ \vdots & \ddots & \vdots \\ x_{L1} & \cdots & x_{LK} \end{bmatrix}$, $\mathbf{Y}_1 = \begin{bmatrix} Y_{11} \\ \vdots \\ Y_{L1} \end{bmatrix}$, $x_{lk} \in \mathbf{x}_l$ is the k -th

selected feature by SFFS and $k \leq K$. Using the same procedure described, we can learn the regression weights for other EXIF header features. Finally, regression weights \mathbf{w}_j , where $j = 1, \dots, 5$, are obtained and saved for the testing phase.

2.4. Detection of image manipulation

Fig. 2(b) shows the procedures of detecting image manipulation using the learned regression weights \mathbf{w}_j . The selected statistical noise features, $\hat{\mathbf{X}}_j$, are extracted from a testing image in the same manner as the training images. The estimated EXIF header feature is obtained as $\hat{\mathbf{X}}_j \mathbf{w}_j$. The estimation error, $E_j = |\hat{Y}_j - \hat{\mathbf{X}}_j \mathbf{w}_j|$, between the genuine EXIF header feature, \hat{Y}_j and the estimated header feature is computed. E_j is then compared against a threshold, ρ_j . If E_j is larger than the threshold, ρ_j , the testing image is regarded as manipulated. Otherwise, the image is genuine.

3. EXPERIMENTAL STUDY

In this experiment, we detect 6 types of brightness and contrast adjustment. We first collect 400 camera default JPEG

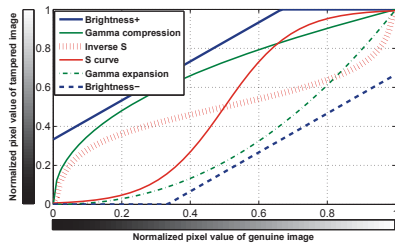


Fig. 5. Mapping curves used for brightness and contrast adjustment of the image.

photos from an Olympus E500 camera. These photos are captured in the automatic mode with different camera settings. These images cover a large variety of common scenes. We randomly select 200 photos to learn the regression weights and the other 200 images for testing. This experiment is repeated 100 times based on different combinations of training images and testing images. Each color channel of the testing image is manipulated separately using the curves in Fig. 5, which are commonly used by brightness and contrast adjustment. In total, our testing set contain 1400 images, which consists 200 genuine testing images and 1200 manipulated testing images.

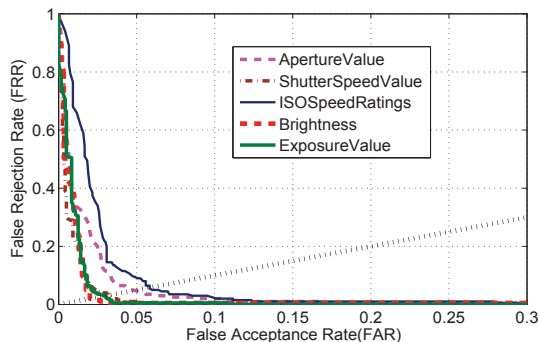


Fig. 6. ROC curve of various EXIF header features.

In order to classify the genuine class and the manipulated class of testing images, we find the threshold of equal error rate (EER), of these two classes from the Receiver Operating Characteristics (ROC) curve, as shown in Fig. 6 and obtain the accuracy rate. The average accuracy rate of each type of manipulated images for the 100 repeated experiments is shown in Table. 1. For instance, the true positive rate (detection of genuine images) of Aperture is 95.26%. On the other hand, the true negative rates (detection of manipulated images) of Aperture for Brightness+, Gamma compression, Inverse S, S, Gamma expansion and Brightness- are 91.48%, 96.97%, 96.96%, 97.1%, 96.96% and 91.98% respectively. Among all the EXIF header features, the true positive rate of ISO is the best with 95.37%. From the table, we can observe that the true positive rates and true negative rates remain consistently high for all EXIF header features which indicate that our model is fairly accurate. The average EER rate for 5 EXIF header features is 94.6%. We have conducted this model on several other camera models. Similar good average rates are achieved with 94.1% for Canon 10D and 94.9% for Canon 450D, respectively.

Table 1. Accuracy rate for different EXIF header features averaged over 100 repeated experiments.

Average Rate=0.946		Aperture	Shutter	ISO	Brightness	Exposure
True Negative	True Positive	0.9526	0.9424	0.9537	0.9446	0.9364
	Image brightness +	0.9148	0.9261	0.9285	0.9115	0.9232
	Gamma Compression	0.9697	0.9491	0.9646	0.9587	0.9480
	Inverse S	0.9696	0.9506	0.9684	0.9553	0.9427
	S curve	0.9710	0.9505	0.9665	0.9583	0.9460
	Gamma expansion	0.9696	0.9458	0.9667	0.9589	0.9446
	Image brightness -	0.9198	0.9311	0.9250	0.9185	0.9163

4. CONCLUSION

In the paper, we propose a novel method to model the EXIF-Image correlation for image manipulation detection. The model correlates the image statistical noise features with the EXIF header features. First, we use regression algorithms to obtain the correlation between the image statistical noise features and EXIF header features. Then, we make use of the model to detect brightness and contrast adjustment. By applying this model to detect brightness and contrast adjustment, we can achieve average test accuracies of 94.6%, 94.1% and 94.9% in three different cameras for 6 types of brightness and contrast adjustment. The good performance shows that our method effectively models the correlation between the image and its EXIF for detecting image manipulation.

5. REFERENCES

- [1] M. K. Johnson and H. Farid, "Exposing digital forgeries through chromatic aberration," in *Proc. of the 8th Workshop on Multimedia and Security*, pp. 48–55, 2006.
- [2] M. Chen, J. Fridrich, M. Goljan, and J. Lukás, "Determining image origin and integrity using sensor noise," *IEEE Trans. on Info. Forensics and Security*, vol. 3, pp. 74–90, 2008.
- [3] A. Swaminathan, M. Wu, and K. J. R. Liu, "Digital image forensics via intrinsic fingerprints," *IEEE Trans. on Info. Forensics and Security*, vol. 3, pp. 101–117, 2008.
- [4] H. Cao and A. C. Kot, "Detection of tampering inconsistencies on mobile photos," in *Int. Workshop on Digital Watermarking*, 2010.
- [5] M. C. Stamm and K. J. R. Liu, "Forensic detection of image manipulation using statistical intrinsic fingerprints," *IEEE Trans. on Info. Forensics and Security*, vol. 5, no. 3, pp. 492–506, 2010.
- [6] I. Avcibas, S. Bayram, N. D. Memon, M. Ramkumar, and B. Sankur, "A classifier design for detecting image manipulations," in *ICIP*, pp. 2645–2648, 2004.
- [7] S. Bayram, I. Avcibas, B. Sankur, and N. D. Memon, "Image manipulation detection," *J. Electronic Imaging*, vol. 15, no. 4, p. 041102, 2006.
- [8] A. Popescu and H. Farid, "Statistical tools for digital forensics," in *6th Int. Workshop on Info. Hiding*, (Toronto, Canada), 2004.
- [9] Japan Electronics and Information Technology Industries Association, "Exchangeable image file format for digital still cameras: Exif version 2.2." JEITA CP-3451, April 2002.
- [10] G. C. Holst and T. S. Lomheim, *CMOS/CCD sensors and camera systems*. Society of Photo Optical, 2007.
- [11] H. Gou, A. Swaminathan, and M. Wu, "Intrinsic sensor noise features for forensic analysis on scanners and scanned images," *IEEE Trans. on Info. Forensics and Security*, vol. 4, pp. 476–491, 2009.
- [12] P. Pudil, F. Ferri, J. Novovicova, and J. Kittler, "Floating search methods for feature selection with nonmonotonic criterion functions," in *Proc. of Int. Conf on Pattern Recognition*, vol. 2, pp. 279–283, 1994.