Distinguishing between Camera and Scanned Images by Means of Frequency Analysis

Roberto Caldelli, Irene Amerini, and Francesco Picchioni

Media Integration and Communication Center - MICC,
University of Florence, Florence, Italy
roberto.caldelli@unifi.it
http://lci.det.unifi.it/caldelli.html

Abstract. Distinguishing the kind of sensor which has acquired a digital image could be crucial in many scenarios where digital forensic techniques are called to give answers. In this paper a new methodology which permits to determine if a digital photo has been taken by a camera or has been scanned by a scanner is proposed. Such a technique exploits the specific geometrical features of the sensor pattern noise introduced by the sensor in both cases and by resorting to a frequency analysis can infer if a periodicity is present and consequently which is the origin of the digital content. Experimental results are presented to support the theoretical framework.

Keywords: digital forensic, source identification, scanner, sensor noise.

1 Introduction

Digital images are nowadays used in the majority of the application fields in place of "old" analog images because of their easiness of usage, quality and above all manageability. These favorable issues bring anyway an intrinsic disadvantage: digital content can be simply manipulated by ordinary users for disparate purposes so that origin and authenticity of the digital content we are looking at is often very difficult to be assessed with a sufficient degree of certainty. Scientific instruments which allow to give answers to basic questions regarding image origin and image authenticity are needed [1]. Both these issues are anyway connected and sometimes are investigated together. In particular, by focusing on assessing image origin, two are the main aspects to be studied: the first one is to understand which kind of device has generated that digital image (e.g. a scanner, a digital camera or it is computer-generated) [3,7] and the second one is to succeed in determining which kind of sensor has acquired that content (i.e. the specific camera or scanner, recognizing model and brand) [6,1,4]. The main idea behind this kind of researches is that each sensor leaves a sort of unique fingerprint on the digital content it acquires due to some intrinsic imperfections and/or due to the specific acquisition process. Various solutions have been proposed in literature among these the use of CFA (Color Filter Array) characteristics [5] is quite well-know, nevertheless two

M. Sorell (Ed.): e-Forensics 2009, LNICST 8, pp. 95–101, 2009.

seem to be the main followed approaches. The first one is based on the extraction, from images belonging to different categories (e.g scanned images, photos, etc.), of some robust features which can be used to train a SVM (Support Vector Machine). When training is performed and whether features grant a good characterization, the system is able to classify the digital asset it is asked to check. The second approach is based on the computation of fingerprints of the different sensors (this is particularly used in sensor identification) through the analysis of a certain number of digital contents acquired by a device (e.g. images scanned by a particular scanner, photos taken by a camera and so on). Usually fingerprints are computed by means of the extraction of PRNU noise (Photo Response Non-Uniformity) [1,2] through a digital filtering operation; PRNU presence is induced by intrinsic disconformities in the manufacturing process of silicon CCD/CMOSs. After that the PRNU of the to-be-checked content is compared with the fingerprints and then it is classified. In this paper a new technique to distinguish which kind of device, a digital scanner or a digital camera, has acquired a specific image is proposed. Because of the structure of CCD set, the (PRNU) noise pattern, left over a digital image, will have a completely different distribution: in the scanner case it should show a mono-dimensional structure repeated row after row in the scanning direction, on the other hand, in the camera case, the noise pattern should present a bi-dimensional template. On the basis of this consideration we construct a 1-D signal and by resorting to a DFT analysis, which exploits the possible existence of a periodicity, understanding which has been the acquisition device. The paper lay-out is the following: Section 2 introduces a characterization of the sensor pattern noise and the periodicity is discussed, in Section 3 the proposed methodology is presented and then in Section 4 some experimental results are brought to support theoretical theses; conclusions are drawn in Section 5.

2 Sensor Pattern Noise Characterization

PRNU (Photo Response Non-Uniformity) noise is quite well-known as being an effective instrument for sensor identification because it is deterministically generated over each digital image it acquires. Such a noise is therefore an intrinsic characteristic of that specific sensor. The extraction of this noise is usually accomplished by denoising filters [8] and information it contains are used to assess something on the sensor characteristics. If we focus our attention on the acquisition process, it is easy to comprehend that when a photo is taken by a digital camera, basically a PRNU with a bi-dimensional structure is superimposed to it; on the contrary, when a digital image is created by means of a scanning operation the sensor array which slides over the to-be-acquired asset located on the scanner plate leaves its mono-dimensional fingerprint row by row during scanning. So in the last case, it is expected that a certain periodicity of the 1-D noise signal is evidenced along the scanning direction. This behavior should be absent in the camera case and this difference can be investigated to discern between images coming from the two different kinds of device. Being $R(i, j)$ with $1 \leq i \leq N$ and $1 \leq j \leq M$, the noise extracted by the scanned image of size $N \times M$, and assuming i (row) as scanning direction, it can, at least ideally, be expected that all the rows are equal (see Equation 1).

$$R(i,j) = R(k,j) \quad \forall \ 1 \leq j \leq M, 1 \leq i,k \leq N \tag{1}$$

So if a 1-D signal, $\mathbf{S}$ of $N \times M$ samples, is constructed by concatenating all the rows, it happens that $\mathbf{S}$ is a periodical signal of period M (Equation 2).

$$\mathbf{S} = [R(1,1), \cdots, R(1,M), \cdots, R(N,1), \cdots, R(N,M)] \tag{2}$$

It is also worthy to point out that if the 1-D signal is mounted along columns direction (i.e. this would be right assuming that j is the scanning direction), $\mathbf{S}$ is not periodical anymore, but it is constituted by diverse constant steps each of length M. A periodical signal such as $\mathbf{S}$, represented in Equation 2, contains a number of repetitions equal to N and therefore will have basically a frequency spectrum made by equispaced spikes. Such spikes will be spaced of $(N \times M)/M = N$ and will be weighted by the spectrum of the basic replica of the signal. So most of the energy of such a signal is located in these spikes. Obviously this is what should happen, in practice the 1-D signal will be corrupted and its periodical structure altered. Consequently the spectral spikes will be reduced and their magnitude partially spread over the other frequencies. If it is still possible to individuate such peaks, it will be simple to distinguish between a scanned image and a digital photo.

3 The Proposed Methodology

According to the idea presented in Section 2, let us describe in detail which is the proposed methodology to achieve that aim. The to-be-checked image I (size $N \times M$) is denoise filtered [8] obtaining I_d which is subtracted to the initial image to extract the sensor pattern noise R (see Equation 3).

$$R = I - I_d \tag{3}$$

To improve the possible presence of the deterministic contribution due to the 1-D PRNU pattern noise, R is divided into non-overlapping stripes (both horizontally and vertically, because both possible scanning directions have to be taken into account) and then all the different rows (columns) belonging to a stripe are averaged according to Equation 4 where L is the width of the stripe.

$$R_r(k) = \frac{1}{L} \sum_{i=1}^{L} R[i + (k-1)L] \quad 1 \leq k \leq N/L \tag{4}$$

After that two new noise images, named *bar codes*, respectively R_r (size $N/L \times M$) and R_c (size $N \times M/L$), have been obtained; R_r and R_c have the same number of samples. If an image has been scanned in the row direction, for instance, it is expected that R_r will be composed by equal (ideally) rows, on the

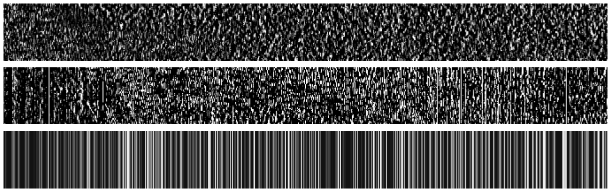


Fig. 1. Bar codes of size $N/L \times M$ (scanning direction = row): camera image (top), scanned image (center) and ideal bar code for a scanned image (bottom)

other side such a characterization can not be expected in the column direction for R_c and, above all, for an image coming from a digital camera (both directions): this circumstance is presented in Figure 1. *Bar codes* are then used to create the mono-dimensional signal by concatenating respectively rows of R_r and columns of R_c and then periodicity is checked. Sometimes to reduce randomness a low pass filtering operation (usually a median filter) is applied to bar codes, along the rows and the columns separately, before constructing 1-D signals.

For the sake of clarity, let us call S_r and S_c the two mono-dimensional signal, obtained as previously described, from R_r and R_c respectively. DFT (Discrete Fourier Transform) is applied to both these signals and the magnitude of the coefficients is considered. After that a selection is carried out on the basis of the following criterion: amplitude values above a threshold T (see Equation 5 where α is a weighting factor usually set to 0.4) and at the same time located in the expected positions within the spectrum (see Section 2) are taken.

$$T = \alpha * max(max(abs(DFT(S_r))), max(abs(DFT(S_c)))) \tag{5}$$

In the end all the values satisfying the previous selection criterion are added, separately for row and column cases, yielding to two energy factors, F_r and F_c respectively and their ratio $RATIO = F_r/F_c$ is computed. If the digital image has been scanned in the row direction, a high value of $RATIO$ is expected (if the scanning direction has been along columns $RATIO$ will be very small), otherwise if the image has been taken by a digital camera the two energy factors should be comparable and a value of $RATIO$ around one is foreseen. Doing so it is possible not only distinguishing between images coming from a scanner or from a camera but, in the scanner case, determining the scanning direction. To improve robustness, this technique is applied to all the three image channels (R, G, B) and three energy contributions are collected in each factor F_r and F_c.

4 Experimental Results

Experimental tests have been carried out to support the theoretical framework. Digital images coming from 4 different scanners (Epson Expression XL 10000 2400x4200 dpi, HP Scanjet 8300 4800x4800 dpi, HP Deskjet F4180 1200x2400 dpi, Brother DCP 7010 600x2.400 dpi) and from 7 commercial cameras (Canon

DIGITAL IXUS i ZOOM, Nikon COOLPIX L12, Fuji Finepix F10, HP Photosmart C935, Nikon D80, Samsung VP-MS11, Sony DSC-P200) have been acquired in TIFF and JPEG format. Because of the diverse size of the contents, the analysis have been done by dividing them into images of fixed dimension $N \times M$ (1024×768). Obtained results have confirmed theoretical assumptions as it can be seen in Figure 2 (a) where $RATIO$ values are plotted and a separate clustering is observed (for sake of clarity when $RATIO$ was over 1 the inverse was taken, due to this, information about scanning direction is lost). In Figure 2 (b), only scanned images, correctly detected, are figured: in this case inversion

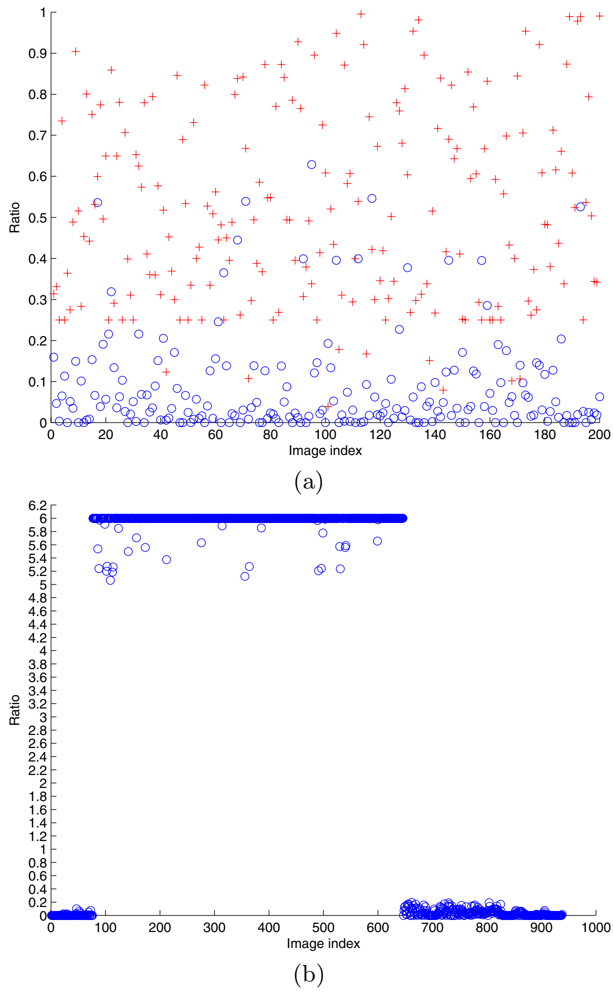
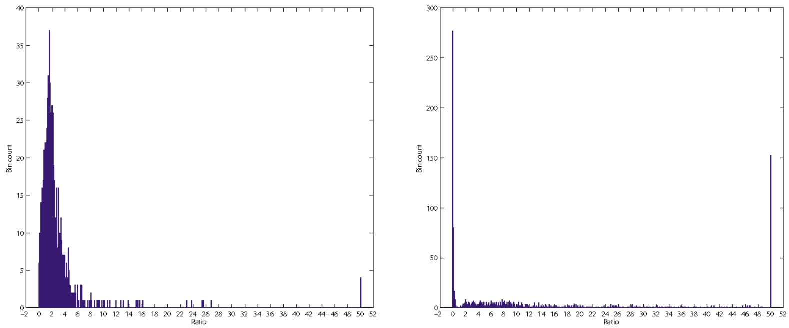


Fig. 2. Energy $RATIO$ for 200 scanned (circle) and 200 camera (cross) images (a). Energy $RATIO$ only for 950 scanned images, correctly detected: scanning directions are evidenced (b).



(a) (b)

Fig. 3. Statistical distribution of $RATIO$: camera (a) and scanned images (b)

Table 1. Confusion matrix for scanned and camera images over a data set of 2000 images (left) and scanning direction recovery for scanner correct answers (right)

	Camera	Scanner
Camera	89.74%	10.26%
Scanner	14.65%	85.35%

	Row	Column
Row	100.00%	0.00%
Column	0.00%	100.00%

of $RATIO$ has not been done and, to make visualization easier, high values are saturated at 6. It is simply to distinguish the two different scanning directions individuated by high and low values of $RATIO$; in particular it is interesting to note the left and the right side of the plot related to column scanning direction and the central part related to row direction. In Figure 3 the statistical distribution of $RATIO$ for 1000 camera images (a) and 1000 scanned ones (b) are pictured where, in this case, higher values have been saturated at 50; a strong concentration is evidenced on the tails of the graph for the scanner case. Finally, a massive test has been carried out on a data set of 2000 images (half scanned images and half photos) by setting a threshold at 0.2 with $RATIO$ normalized between 0 and 1 (as done for Figure 2 (a)): percentages are presented in the rows of Table 1 (left). In Table 1 (right) percentages related to the scanning directions in the scanner successful cases (85.35% of Table 1 left) are reported.

5 Conclusions

In this paper a new technique to distinguish between digital images acquired by a scanner and photos taken by a digital camera has been proposed. Sensor pattern noise periodicity along the scanning direction is checked for classification through a frequency analysis. Experimental results have been presented to support the theoretical framework. Future developments will regard the integration of this feature within a SVM.

References

1. Chen, M., Fridrich, J., Goljan, M., Lukas, J.: Determining Image Origin and Integrity Using Sensor Noise. IEEE Trans. on Information Forensics and Security 3(1), 74–90 (2008)
2. Mondaini, N., Caldelli, R., Piva, A., Barni, M., Cappellini, V.: Detection of malevolent changes in digital video for forensic applications. In: Proc. SPIE, vol. 6505, 65050T (2007)
3. Lyu, S., Farid, H.: How realistic is photorealistic? IEEE Transactions on Signal Processing 53(2), 845–850 (2005)
4. Gou, H., Swaminathan, A., Wu, M.: Robust scanner identification based on noise features. In: Proc. SPIE, vol. 6505, p. 65050S (2007)
5. Swaminathan, A., Wu, M., Liu, K.J.R.: Digital Image Forensics via Intrinsic Fingerprints. IEEE Transactions on Information Forensics and Security 3(1), 101–117 (2008)
6. Khanna, N., Mikkilineni, A.K., Chiu, G.T.-C., Allebach, J.P., Delp, E.J.: Scanner identification using sensor pattern noise. In: Proc. SPIE, vol. 6505, 65051K (2007)
7. Khanna, N., Chiu, G.T.-C., Allebach, J.P., Delp, E.J.: Forensic techniques for classifying scanner, computer generated and digital camera images. In: Proc. IEEE ICASSP, pp. 1653–1656 (2008)
8. Mihcak, M.K., Kozintsev, I., Ramchandran, K.: Spatially Adaptive Statistical Modeling of Wavelet Image Coefficients and its Application to Denoising. In: Proc. IEEE ICASSP, vol. 6, pp. 3253–3256 (1999)