Evaluation of Residual-Based Local Features for Camera Model Identification

Francesco Marra, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva(✉)

DIETI, University Federico II of Naples, Naples, Italy
{francesco.marra,poggi,carlosan,verdoliv}@unina.it

Abstract. Camera model identification is of interest for many applications. In-camera processes, specific of each model, leave traces that can be captured by features designed *ad hoc*, and used for reliable classification. In this work we investigate on the use of *blind* features based on the analysis of image residuals. In particular, features are extracted locally based on co-occurrence matrices of selected neighbors and then used to train an SVM classifier. Experiments on the well-known Dresden database show this approach to provide state-of-the-art performances.

Keywords: Camera model identification · Local features · Residuals

1 Introduction

Identifying which camera took a given photo can be of great interest in many instances [1]. Law enforcement agencies may be interested to trace back the origin of some unlawful or dangerous images. Likewise, linking a photo to a given camera may represent evidence before a court of law. On the other hand, someone may be interested in proving that a photo was taken by his/her camera to claim intellectual property. The growing interest towards this task is a direct consequence of the huge growth in the acquisition and diffusion of digital images, therefore it is bound to increase further.

On the technical side, a great impulse to research in this field came with the pioneering work of Lukàš et al. [2] proposing a simple and effective tool for source identification. The photo-response non uniformity (PRNU) pattern, also known as sensor pattern noise, is a distinctive pattern, unique for each camera and stable in time, originated by the unavoidable imperfections occurring during the sensor manufacturing process. Each photo taken by a given camera contains traces of the PRNU, which represents, therefore, sort of a camera fingerprint, which can be used for reliable identification. However, even these very promising methods have their drawbacks, which limit their field of application. In particular, to extract a camera's PRNU pattern, a large number of images taken by that camera are necessary. This is often not possible without the collaboration of the camera owner and, in any case, is extremely time-consuming. A more viable intermediate step is the identification of the camera model. Finding the model may significantly narrow the search, which can be completed also by more conventional methods.

V. Murino et al. (Eds.): ICIAP 2015 Workshops, LNCS 9281, pp. 11–18, 2015.
DOI: 10.1007/978-3-319-23222-5_2

Camera model identification is made possible by the distinctive traces left in images as a result of the unique combination of in-camera processing steps. In fact, in modern digital cameras, the final image is produced through a number of algorithms, each characterized by several free parameters. Popular examples are demosaicing, based often on complex adaptive nonlinear interpolation, and JPEG compression, where the quantization matrix can be defined by the user. It is highly unlikely that different camera models use the very same set of algorithms and parameters and, therefore, very likely that their traces allow reliable identification.

Indeed, in the very same literature on PRNU-based identification, the appearance of model-based artifacts in the estimated PRNU pattern was observed [3], and used to identify the camera model. However, this path was followed before by other researchers. In [4], inspired by the work of Popescu and Farid for image forgery detection [5], traces of different interpolation algorithms where sought and used as distinctive model features. In fact, each pixel is strongly correlated with its neighbors (both spatial and across color channels). The weights of the interpolation kernel are estimated and used as features for camera identification combined with frequency domain features, that take into account the periodic artifacts caused by the CFA pattern. The strong dependencies among pixels has been also explored in [6] [7]. In [6], in particular, and also in [8], weight estimation is conducted locally on each color band, and a different procedure is used based on the content of the region. This reflects the fact that often adaptive demosaicing techniques are used to reduce blurring artifacts. In [7], instead, partial second-order derivative correlation models are proposed to detect both the intrachannel and the cross-channel dependence due to demosaicing. Other methods seek to characterize JPEG compression artifacts [9], DCT coefficients statistics [10], or lens distortion artifacts like chromatic aberration [11]. A different approach is proposed in [12] where the heteroscedastic noise model valid for raw images and characterized by two parameters is used to identify camera models.

The majority of the methods look for artifacts related to some specific in-camera processing step, trying to estimate its unknown parameters. A "blind" approach is also possible, however, where no hypotheses is made on the origin of camera-specific marks, and the identification task is regarded simply as a texture classification problem. With this approach, the focus shifts on the definition of the most discriminative features, irrespective of their meaning. Both global and local features can be considered, drawing often from the vast literature of closely related fields, such as material classification or steganalysis. The aim of this paper is to evaluate a class of such features, based on co-occurrences of image residuals, and show their potential for the camera model identification task.

In the next Section we review "blind" feature-based methods, then the residual-based local features are described in Section 3, and evaluated experimentally in Section 4.

2 Related Work

The use of generic features was first considered in the work of Kharrazi et al. [13]. The Authors propose to use various global statistics, extracted from each individual color band, based on the correlation of couples of color bands, and also extracted from some wavelet subbands. In addition, some Image Quality Metrics (IQM), used in [14] for steganalysis, are evaluated on all the color bands, both in the spatial and transform domain. It is worth noting that these last features are computed on residual images (high-pass filtered versions of the original data).

Indeed, computing features on image residuals is common to the majority of methods proposed in the literature. This way, results become independent of the image content, and artifacts are more easily detected. In [15] IQM features are extracted from high-pass residuals of each color band. These features are then combined with BSM (Binary Similarity Measures), i.e., LBP (Local Binary Pattern) extracted from the least-significant bit planes, and with an enlarged set of features computed in the wavelet domain. Besides the features used in [13], other first-order statistics are computed, as well as some inter-band correlation indexes, inspired by [16]. Instead in [17] Gloe proposes to add some color features to the ones used in [13]. Experiments on the Dresden Image Database [18] prove this combination to guarantee a performance gain w.r.t. both [13] and [15].

The majority of the features recalled thus far are evaluated globally on the whole image (both original and high-pass filtered) or a decimated version of it, if wavelet subbands are considered. However, in order to capture subtle image patterns which may correspond to discriminative features, it is important to consider local features, extracted from a small neighborhood of each pixel of the image, as it happens for LBP. This is the main focus of the work of Xu et al. [19], where LBP features are evaluated both on the original image and on some residuals. Note that LBP is computed on two-pixel supports, and hence encodes only first-order spatial variations. However, combining it with a preliminary high-pass filter amounts to considering a larger support, and evaluating higher-order statistics [20]. We mention briefly other approaches [21] which look at the statistical differences in the DCT domain by computing Markovian transition probabilities. Also in this case, similar features had been already considered in steganalysis [22].

3 Residual-Based Local Features

The analysis of the state of the art shows that local descriptors can provide precious clues for camera model identification. Moreover, since such clues, related to the camera processing chain, are contained in the image micro-patterns and not in the scene content, it makes sense to remove the latter and work on image residuals. Even in this framework, however, two main open issues remain about *i)* how to extract informative image residuals and *ii)* how to process them in order to obtain an expressive camera-related feature. Given the complexity and variety of the in-camera processes involved, no conclusive answer can be hoped for. However, we will show that co-occurrence based local features, computed on image residuals and proposed originally in [23] for steganalysis, may represent a valuable tool for this task. A similar path was used successfully in digital image forensics [24] [25].

The feature vector associated with the image under test is extracted through the following steps:

- computation of residuals through high-pass filtering;
- quantization and truncation of the residuals;
- computation of the histogram of co-occurrences.

In [23] a number of linear and non-linear high-pass filters have been used for the computation of residuals, combining all resulting feature vectors by means of an ensemble classifier. In [25], instead, a few filters have been selected based on a preliminary performance analysis on the training set. In particular, the third order linear filter defined as

$$r_{i,j} = x_{i,j-1} - 3\,x_{i,j} + 3\,x_{i,j+1} - x_{i,j+2} \tag{1}$$

with x and r indicating original and residual images, and (i, j) the spatial coordinates, has been found to provide good and stable results. When filters are not row-column symmetric, they are applied also on the image transpose to augment data.

A further alternative, that will be explored in this paper is the use of an image denoising algorithm to compute residuals. In fact, if the aim of this process is to remove the scene content from the image, keeping all the noise, this is best achieved by resorting to specialized filters that abound in the literature. In particular we will consider the nonlocal BM3D filter, among the best and well-known denoiser, already used for residual computation [26] in PRNU-based image forensics. The residual image will be therefore obtained as the difference between the original image x and its denoised version $\widehat{x} = f(x)$, with $f(\cdot)$ the denoising filter.

Co-occurrences are then evaluated on these residuals, which provide information on higher-order phenomena and involve a relatively large number of image pixels. In order to obtain a manageable co-occurrence matrix, residuals are quantized/truncated to a small number of values as:

$$\widehat{r}_{i,j} = \mathrm{trunc}_T(\mathrm{round}(r_{i,j}/q)) \tag{2}$$

with q the quantization step and T the truncation value. To limit the matrix size we consider only $T = 1$ or $T = 2$, leading to uniform mid-thread quantizers with 3 and 5 bins, respectively. At this point co-occurrences on four pixels along the filter direction (say, rows) are computed, that is

$$\begin{aligned} &C_1(k_0, k_1, k_2, k_3) = \\ &\sum_{i,j} I(\widehat{r}_{i,j} = k_0, \widehat{r}_{i,j+1} = k_1, \widehat{r}_{i,j+2} = k_2, \widehat{r}_{i,j+3} = k_3) \end{aligned} \tag{3}$$

where $I(A)$ is the indicator function of event A, equal to 1 if A holds and 0 otherwise, and all k indexes take values in $\{-T, \ldots, +T\}$. Another co-occurrence matrix is then computed working across the filter direction (say, columns)

Table 1. Performance comparison for various spam descriptors (512 × 512 cropping).

filter name	deriv. order	accuracy T=1 length 50	T=2 length 338
spam1_14hv	1	95.14	96.66
spam2_12hv	2	97.08	97.31
spam3_14hv	3	96.83	**98.99**
spam11(3x3)	2	95.80	98.23
spam11(5x5)	4	96.07	96.71

$$C_2(k_0, k_1, k_2, k_3) = \sum_{i,j} I(\widehat{r}_{i,j} = k_0, \widehat{r}_{i+1,j} = k_1, \widehat{r}_{i+2,j} = k_2, \widehat{r}_{i+3,j} = k_3) \quad (4)$$

Invoking left-right and positive-negative symmetries, the original $(2T+1)^4$ features can be reduced through pooling to just 338 and 50 for $T = 2$ and $T = 1$, respectively. The extracted features are eventually used to train an SVM classifier.

4 Experimental Results

To assess the performance of the features under analysis we carried out experiments on the Dresden database [18]. This is one of the most widespread databases in this field, used in many recent papers. In the Dresden database, 26 different camera models are available, each with several individual devices and hundreds or thousands photos. However, to speed up the experimental phase, we used a smaller version of this database, as done in [17], comprising only 10 models. We trained an SVM with linear kernel using all the images coming from one single device for each camera model. We run 20 times this procedure, each time randomly choosing the devices used for training, and averaged results. Notice that the images are not cropped, therefore they keep the original camera resolution, going from about 5 to about 15 Mpixels.

For comparison we implemented the features proposed in Celiktutan-2008 [15], Gloe-2012 [17] and Xu-2012 [19]. As for the features based on co-occurrences, some preliminary experiments are carried out to select the best configuration. In Table 1 we report results obtained on the selected database using linear high-pass filters computing derivatives of first to fourth order, called of type *spam* [23]. The residual are then quantized to 3 or 5 levels, changing the quantization step q accordingly. Since these features are relatively short, 50 or 338 components, they can be estimated reliably also on a portion of the image, so we worked on a 512 × 512 section to expedite the process. Results show that it is worth using longer features, associated with 5-level quantization,

Table 2. Performance of various descriptors on the 10-model Dresden subset. Note that the wavelet-based features used in Celiktutan-2008 cannot be extracted from very small images, like 64 × 64.

		accuracy			
Feature	length	whole image	512 × 512	128 × 128	64 × 64
Celiktutan-2008	592	96.89	91.29	73.16	-
Gloe-2012	82	97.60	82.32	64.25	55.60
Xu-2012	354	98.87	97.96	89.82	74.82
co-occurrences (BM3D)	338	99.06	97.03	84.03	68.12
co-occurrences (best spam)	338	**99.44**	**98.99**	**94.82**	**85.07**

while indication on the filter are more controversial. In any case, we used the best performing filter, spam3_14hv, in the rest of the experiments.

Table 2 shows results for the 10-model identification experiment. The co-occurrence based local feature with the best spam filter provides a 99.44% accuracy, better than both Gloe-2012 and than Xu-2012. When residuals are computed through denoising, the performance is slightly worse. Table 3 provides detailed model-by-model results for the co-occurrence based feature. The latter provides perfect on near-perfect accuracy in all cases, except for the Nikon D200 images, associated sometimes with the Kodak M1063 camera. It is worth pointing out that also in other investigations, including [17], this particular camera model has been found hard to identify, which raises interest on the in-camera processes it adopts.

In Table 2 we show also results obtained after cropping images to much smaller sizes. As expected, a 512 × 512 section allows one to obtain almost the same results as on the full image. With much smaller sections, even 64 × 64

Table 3. Mis-Classification matrix for the best co-occurrence feature.

	identified as									
device	I70	Z150	M10	S710	D200	μ	DCZ	7325	L74	NV
I70	100	-	-	-	-	-	-	-	-	-
Z150	0.6	98.6	-	-	-	0.8	-	-	-	-
M10	-	-	99.7	-	-	-	0.1	0.1	-	0.1
S710	-	-	-	100	-	-	-	-	-	-
D200	-	-	4.1	-	95.1	-	0.4	0.2	-	0.2
μ	-	-	-	-	-	100	-	-	-	-
DCZ	-	0.4	-	-	-	-	99.6	-	-	-
7325	-	-	-	-	-	-	-	100	-	-
L74	-	-	-	-	-	-	-	-	100	-
NV	-	-	0.4	0.4	-	-	-	-	-	99.2

pixels, only the co-occurrence features based on the best spam filter keep being reliable. If confirmed in other experiments, this might be a valuable property, not only for reducing analysis time when computational power is limited, but also for dealing with situations in which only a fragment is available, for example in image forgery localization.

5 Conclusions

Our analyses confirm for camera model identification the excellent performance exhibited in other domains by co-occurrence based features. Improved features can be certainly designed, for example by taking into account individually the three color bands. In any case, this is only a preliminary exploration, and much work remains to be done. Among the most important issues: more comprehensive experiments are necessary, with the full Dresden database, as well as other datasets; robustness to various forms of post-processing should be studied; the open set scenario should be also considered, dealing with unknown models.

Acknowledgments. This work was partially funded by the Italian Ministry of Education, University and Research (MIUR) within the framework of the project PAC02L1_00050 AETERNUUM.

References

1. Kirchner, M., Gloe, T.: Forensic camera model identification. In: Ho, T., Li, S. (eds.) Handbook of Digital Forensics of Multimedia Data and Devices. Wiley-IEEE Press (2015)
2. Lukàš, J., Fridrich, J., Goljan, M.: Digital camera identification from sensor pattern noise. IEEE Transactions on Information Forensics and Security **1**(2), 205–214 (2006)
3. Filler, T., Fridrich, J., Goljan, M.: Using sensor pattern noise for camera model identification. In: IEEE International Conference on Image Processing, pp. 1296–1299 (2008)
4. Bayram, S., Sencar, H., Memon, N., Avcibas, I.: Source camera identification based on CFA interpolation. In: IEEE Int. Conference on Image Processing, pp. 69–72 (2005)
5. Popescu, A., Farid, H.: Exposing digital forgeries by detecting traces of resampling. IEEE Transactions on Signal Processing **53**(2), 758–767 (2005)
6. Swaminathan, A., Wu, M., Liu, K.J.R.: Rich models for steganalysis of digital images. IEEE Transactions on Information Forensics and Security **2**(1), 91–105 (2007)
7. Cao, H., Kot, A.: Accurate detection of demosaicing regularity for digital image forensics. IEEE Trans. on Information Forensics and Security **4**(4), 899–910 (2009)
8. Bayram, S., Sencar, H., Memon, N.: Improvements on source camera-model identification based on CFA. In: Advances in Digital Forensics II, IFIP International Conference on Digital Forensics, pp. 289–299 (2006)

9. Fan, N., Jin, C., Huang, Y.: Source Camera identification by JPEG compression statistics for image forensics. In: TENCON, pp. 1–4 (2006)
10. Thai, T., Retraint, F., Cogranne, R.: Camera model identification based on dct coefficient statistics. Digital Signal Processing **4**, 88–100 (2015)
11. Van, L., Emmanuel, S., Kankanhalli, M.: Identifying source cell phone using chromatic aberration. In: IEEE International Conference on Multimedia and Expo, pp. 883–886 (2007)
12. Thai, T., Cogranne, R., Retraint, F.: Camera model identification based on the heteroscedastic noise model. IEEE Transactions on Image Processing **23**(1), 250–263 (2014)
13. Kharrazi, M., Sencar, H., Memon, N.: Blind source camera identification. In: IEEE International Conference on Image Processing, pp. 709–712 (2004)
14. Avcibaş, I., Memon, N., Sankur, B.: Steganalysis using image quality metrics. IEEE Transactions on Image Processing **12**(2), 221–229 (2003)
15. Çeliktutan, O., Sankur, B., Avcibaş, I.: Blind identification of source cell-phone model. IEEE Transactions on Information Forensics and Security **3**(3), 553–566 (2008)
16. Lyu, S., Farid, H.: Steganalysis using higher-order image statistics. IEEE Transactions on Information Forensics and Security **1**(1), 111–119 (2006)
17. Gloe, T.: Feature-based forensic camera model identification. In: Shi, Y.Q., Katzenbeisser, S. (eds.) Transactions on DHMS VIII. LNCS, vol. 7228, pp. 42–62. Springer, Heidelberg (2012)
18. Gloe, T., Böhme, R.: The Dresden image database for benchmarking digital image forensics. Journal of Digital Forensic Practice **3**(2–4), 150–159 (2010)
19. Xu, G., Shi, Y.: Camera model identification using local binary patterns. In: IEEE International Conference on Multimedia and Expo, pp. 392–397 (2012)
20. Gragnaniello, D., Poggi, G., Sansone, C., Verdoliva, L.: An investigation of local descriptors for biometric spoofing detection. IEEE Transactions on Information Forensics and Security **10**(4), 849–863 (2015)
21. Xu, G., Gao, S., Shi, Y.Q., Hu, R.M., Su, W.: Camera-model identification using markovian transition probability matrix. In: Ho, A.T.S., Shi, Y.Q., Kim, H.J., Barni, M. (eds.) IWDW 2009. LNCS, vol. 5703, pp. 294–307. Springer, Heidelberg (2009)
22. Shi, Y.Q., Chen, C.-H., Chen, W.: A markov process based approach to effective attacking JPEG steganography. In: Camenisch, J.L., Collberg, C.S., Johnson, N.F., Sallee, P. (eds.) IH 2006. LNCS, vol. 4437, pp. 249–264. Springer, Heidelberg (2007)
23. Fridrich, J., Kodovsky, J.: Rich models for steganalysis of digital images. IEEE Transactions on Information Forensics and Security **7**, 868–882 (2012)
24. Kirchner, M., Fridrich, J.: On detection of median filtering in images. In: SPIE, Electronic Imaging, Media Forensics and Security XII, pp. 101–112 (2010)
25. Verdoliva, L., Cozzolino, D., Poggi, G.: A feature-based approach for image tampering detection and localization. In: IEEE Workshop on Information Forensics and Security, pp. 149–154 (2014)
26. Chierchia, G., Parrilli, S., Poggi, G., Sansone, C., Verdoliva, L.: On the influence of denoising in PRNU based forgery detection. In: 2nd ACM workshop on Multimedia in Forensics, Security and Intelligence, pp. 117–122 (2010)