J Ambient Intell Human Comput (2013) 4:265–274
DOI 10.1007/s12652-011-0070-2

ORIGINAL RESEARCH

Experimentations with source camera identification and Online Social Networks

Aniello Castiglione · Giuseppe Cattaneo · Maurizio Cembalo · Umberto Ferraro Petrillo

Received: 13 April 2011 / Accepted: 15 September 2011 / Published online: 5 October 2011

Abstract In this paper is presented an extended experimental evaluation of one of the most effective source camera identification techniques proposed so far, by Lukáš et al. (IEEE Trans Inf Forensics Security 1(2):205–214, 2006). This method uses the characteristic noise left by the sensor on a digital picture as a fingerprint in order to identify the source camera used to take the picture. The aim of the experiments is to assess the effectiveness of this technique when used with pictures that were previously modified using several common image-processing functions coming with photo-editing tools. Moreover, the technique is applied to photos passed through Online Social Networks or Online Photo Sharing websites, without any "human" explicit modification but only elaborated by such Web 2.0 tools. The results confirm that, in several cases, the method by Lukáš et al. (IEEE Trans Inf Forensics Security 1(2):205–214, 2006) is resilient to the modifications introduced by the considered image-processing functions. However, in the experiments it has been possible to identify several cases where the quality of the identification process deteriorated because of the noise introduced by the image-processing. In addition, when dealing with Online Social Networks and Online Photo Sharing services, it has been noted that some of them process and modify the uploaded pictures. These modifications make ineffective, in many cases, the method by Lukáš et al. (IEEE Trans Inf Forensics Security 1(2):205–214, 2006)

Keywords Online Social Networks · Online Photo Sharing · Image Forensics · Digital Forensics · Digital camera identification · Digital camera model identification · Digital investigations

1 Introduction

Nowadays, taking and sharing digital pictures is becoming a very popular activity. This is witnessed by the explosive growth of the digital cameras market: e.g., more than one billion digital cameras have been produced and shipped in 2010 (Camera Imaging Products Association 2011). A consequence of this trend is that also the number of crimes involving digital pictures increases, either because pictures are part of the crime (e.g., exchanging pedopornographic pictures) or because their analysis may reveal some important clue about the author of a crime. Image Forensics tries to help the investigators when in presence of digital photographic evidence. One of the many issues that the Image Forensics tries to deal with is the *source camera identification problem*, i.e., establish if a given image has

The project is a joint work with the CNCPO (*Centro Nazionale per il Contrasto della Pedopornografia Online*), part of the "Dipartimento della Pubblica Sicurezza" within the Italian Ministry of Interior. A preliminary version of this paper appeared in the International Workshop on Cloud, Wireless and e-Commerce Security (CWECS 2010) held in Fukuoka, November 4–6, 2010.

A. Castiglione · G. Cattaneo · M. Cembalo
Dipartimento di Informatica "R.M. Capocelli",
Università degli Studi di Salerno,
Via Ponte don Melillo, 84084 Fisciano, SA, Italy
e-mail: castiglione@ieee.org; castiglione@acm.org

G. Cattaneo
e-mail: cattaneo@dia.unisa.it

M. Cembalo
e-mail: maucem@dia.unisa.it

U. Ferraro Petrillo (✉)
Dipartimento di Scienze Statistiche, Università degli Studi di Roma "La Sapienza", P.le Aldo Moro 5, 00185 Rome, Italy
e-mail: umberto.ferraro@uniroma1.it

been taken by a given digital camera. Many identification techniques have been proposed so far in literature. All these techniques generally work by using the sensor noise left by a digital sensor when taking a picture as a fingerprint for identifying the sensor. These studies are generally accompanied with tests proving the effectiveness of these techniques, both in terms of False Acceptance Rate (FAR) and False Rejection Rate (FRR).

Unfortunately, most of these contributions do not take into consideration that, in practice, the images that are shared and exchanged over the Internet have often been pre-processed. Instead, it is a common practice to assume that the images to be examined are unmodified or, at most, to ignore the effects of the pre-processing.

Even without considering the case of malicious users that could intentionally process a picture in order to fool the existing identification techniques, this assumption is unrealistic for at least two reasons. The first is that, as previously mentioned, almost all current photo-managing software offers several functions for adjusting, sometimes in a "magic" way [see the "I'm feeling lucky" function on Google Picasa (2010)] different characteristics of a picture. The second reason can be found in the way the images are managed by some of the most important Online Social Network (OSN) and Online Photo Sharing (OPS) sites. These services usually make several modifications to the original photos before publishing them in order to either improve their appearance or reduce their size.

The contribution of this paper aims to understand how one of the most prominent source camera identification techniques performs when in presence of pre-processed images, either explicit modified by a user with photo management tools or by OSNs and OPSs services without user awareness.

1.1 Organization of the paper

In Sect. 2 some basic definitions about the source camera identification problem are provided as well as the current literature on this topic briefly reviewed. In Sect. 3 details of the identification technique presented by Lukáš et al. (2006) are introduced, based on the Photo-Response Non-Uniformity (PRNU) of both CCD (Janesick and Blouke 1995) and CMOS sensors (Holst 1996). In Sect. 4 are explained the conditions under which the tests have been conducted, while in Sect. 5 the results of several tests of this technique, performed on a test-sample of nearly 2,500 images taken from eight different cameras, are presented. In the tests, the performance of this technique, when applied to the identification of the cameras used to take both modified and unmodified pictures, is compared. The core part of the work is Sect. 6 where the tests aim to understand if, and how, the OSNs and OPSs services modify the pictures uploaded by a user on such websites, and which are the consequences of these modifications on the identification process. Finally, Sect. 7 presents some concluding remarks.

2 Digital camera identification

Every digital picture contains a random component of noise as well as a deterministic component, the *pattern noise*, that depends on the sensor used to shoot the picture. The pattern noise is "very" similar within all the pictures taken by the same sensor.

The problem of digital camera identification concerns with the identification of the camera that has been used to generate a digital picture, by examining the pattern noise in the picture. This technique is called *source camera identification* and should not be confused with the more general *digital camera model identification* problem, in which there is only interest in establishing which camera model has been used to take a certain picture.

Up until now, three main approaches were proposed in literature to deal with the source camera identification problem. These approaches differ in the type of pattern noise used.

The first approach uses the PRNU noise, i.e., the noise produced by the sensor due to the inhomogeneity of the silicon wafers used to build it. Lukáš et al. (2006) and Chen et al. (2008) proposed two methods to identify the source camera based on the PRNU. These techniques can be used to isolate and extract the noise pattern from a set of pictures taken with the same camera, using this pattern to match or not the cameras with the photos under investigation. Their results show that these methods have high detection rates. Goljan et al. (2009) used a refinement of the method of Chen et al. to run a digital camera identification test on a massive database of digital pictures downloaded from the Internet.

The second approach uses the lens radial distortion that causes straight lines to appear as curved lines on the output images. Choi et al. (2006) have tried this method and discovered that it failed to measure the radial distortion except when there are explicit straight lines in the picture to be processed.

The last approach relies on the *Color Filter Array* (CFA) interpolation, which is a technique used by digital cameras after a picture has been taken in order to determine the colors of the scene. This technique produces small non-uniform color zones that can be seen as a noise source. Every camera has its own interpolation algorithm and produces a small degree of noise that, generally, changes slightly from one camera to another. Bayram et al. (2005) explored the CFA interpolation process to determine the correlation structure present in each color band which can be used for image classification. In this direction, Kharrazi et al., and Long and Huang proposed two methods. The first method (Kharrazi et al. 2004) identifies a set of image features that can be used to uniquely classify a camera model. The accuracy of this method decreases as the number of cameras increases. The second method (Long and Huang 2006) obtains a coefficient matrix from a quadratic pixel correlation model where spatially periodic inter-pixel correlation follows a quadratic form. The results seem to suggest that these two methods work better for the problem of camera model identification rather than for that of source camera identification.

3 The approach by Lukáš et al.

Attention has been focused on the approach proposed by Lukáš et al. (2006) due to it being one of the most effective method, as well as inexpensive in terms of hardware resources unlike other similar methods such as the one proposed by Goljan et al. (2009).

The approach by Lukáš et al. (2006) works in two stages. In the first stage, the PRNU associated with a CCD sensor is determined by analyzing a batch of images taken with the sensor. In the second stage, given a picture, the procedure evaluates the correlation between the noise in the picture and the pattern noise evaluated in the previous stage in order to distinguish whether the picture has been taken using that CCD sensor or not.

The extraction of the PRNU from an image is performed by denoising the image using a wavelet-based algorithm. The denoised image is subtracted from the original image giving as output a new image containing several components: the CCD sensor noise, the random noise and various contributions from the image signal. Thus, in order to eliminate the random component of the noise, the denoising procedure is applied to a set of images (captured by the same camera) and the corresponding noise residues are averaged to obtain the reference pattern of a given digital camera.

Afterwards, to determine whether a given image is captured by a digital camera, the noise pattern extracted from the given image is correlated with the reference pattern of the camera. If the correlation value exceeds a pre-determined threshold, then the image was taken with that camera. In order to estimate the accuracy of the method as well as to compute the thresholds, the Neyman–Pearson criterion was used, specifying a bound on the FAR.

3.1 Implementation details

The method proposed by Lukáš et al. (2006) was implemented using the Matlab software (MathWorks 2010). This software was used due to it being efficient as well as providing several pre-implemented components (e.g., wavelets functions) which are useful in the implementation of the various identification techniques.

Concerning the implementation, the main element of the method proposed by Lukáš et al. (2006) is the PNRU filter. This filter simulates the behaviour of the Wiener filter in the wavelet domain and it has been suggested by Kivanc Mihcak et al. (1999). The Wiener filter is based on a statistical approach and aims to filter the noise of an image.

There are several families of wavelets, each one suitable for different applications, differing in the number of coefficients they use. In the early stages of the tests, several combinations were tried, with the optimal choice being 4-levels and 8-levels Daubechies wavelets.

4 Experimental settings

The tests were organized in three phases. In the first phase, the effectiveness of the method by Lukáš et al. (2006) when applied to the camera identification for unmodified digital pictures was assessed. In the second phase, the original set of pictures were initially pre-processed using several types of image-processing functions, with identification process then being repeated, using the decision thresholds established during the first test according to non pre-processed images. In the third phase, the previous tests with pre-processed images were repeated, using the decision thresholds that have been recalculated from pre-processed images.

In all the tests, seven different camera models were considered, resulting in eight cameras (as shown in Table 1). In order to stress the identification method as well as cover a wider range of hardware, cameras belonging to different market sectors and different manufactures were chosen. Looking at Table 1, cameras with ID 1 and 2 were chosen because they have the same image sensor size as well as the same CMOS sensor (Camera Labs 2010). Cameras with ID 3 and 4 share the same brand and model. The other four cameras are a mix of common cameras.

For each camera model $c \in C = 1, 2,\ldots, 8$ two sets of images were collected: the *Images for Reference Pattern* (IRP) and the *Images for Testing* (IT). $\mathrm{IRP}_c/\mathrm{IT}_c$ denotes the IRP / IT sets for the camera c. The IRP_c set is composed of 128 images collected by taking pictures of a uniform white surface. The images were taken on a tripod, with no flash, auto-focus, no zoom, best JPEG compression quality, and with all the other options set to their default values. The IT_c set is made up of 180 images portraying different types of subjects. In this case, the images were taken using different types of settings, with the exception of the JPEG compression quality as well as the image size, which were always set to maximum.

Table 1 Cameras used in the tests

ID	Model	Sensor	Image size
1	Canon EOS 400D	CMOS	3,888 × 2,592
2	Canon EOS 1000D	CMOS	3,888 × 2,592
3	Canon powerShot A400 instance A	CCD	2,048 × 1,536
4	Canon PowerShot A400 instance B	CCD	2,048 × 1,536
5	Panasonic Lumix DMC-FZ20	CCD	2,048 × 1,536
6	Panasonic Lumix DMC-FS5	CCD	3,648 × 2,736
7	Kodak EasyShare CX 7530	CCD	2,560 × 1,920
8	HP PhotoSmart E327	CCD	2,560 × 1,920

The effectiveness of the identification was measured by counting the number of pictures erroneously rejected by the identification technique over the total number of pictures taken with a certain camera (FRR). Moreover, in all the tests, the decision thresholds were set in such a way to keep to 0 the total number of pictures erroneously classified as taken with a certain camera (FAR).

5 Experimental analysis

All the following tests have been conducted on a data set composed of images that have been knowingly modified in order to test the effectiveness of the identification tecnique by Lukáš et al. (2006).

5.1 Test 1

A preliminary problem to be faced when applying the method by Lukáš et al. (2006) is to ensure that the two images to be correlated (i.e., the image reference pattern (IRP) and the image to be identified) have the same size. This condition can be easily met in three different ways:

- extract from both images two sub-images of the same size (*Sub*). In this case, extract two images originating at point (0,0) and having size 512 × 512;
- crop the larger image to match the smaller image (*Crop*);
- resize the larger image to match the smaller image (*Resize*).

The original method proposed by Lukáš et al. (2006) uses the *Crop* approach. In this study, it was decided to also test the other two approaches in order to determine which one performs better. According to the tests, presented in Table 2, the best approach seems to be *Resize* while the worst is *Sub*. The reason for such a bad performance is likely to be due to the elimination of a large part of the original image, when processing large pictures. The average resolution of the pictures used in the tests is near 2,560 × 1,920. As a consequence of this, the cropped image, whose size is fixed to 512 × 512, retains only the 5% of the original image as well as its signature, and thus is subject to a worse correlation. In all the remaining tests presented in this paper, when needed, the *Resize* technique has been used.

Figure 1 presents the scatter plot of the correlations between all the images of the data set and the IRP of the Canon PowerShot A400 instance A (the camera with ID 3). It is interesting to note that this camera model is present twice in the experiments (instance A and instance B). The majority of the correlations between the IRP and images taken using camera with ID 3 is considerably above the decision thresholds, thus leading to a correct classifications. Pictures taken using different camera models exhibit a very small correlation value, close to zero. Finally, picture taken using camera with ID 4 (i.e., same camera model, different instance) feature higher correlation values which are, anyway, under the decision thresholds.

5.2 Test 2

The second test was intended to assess the resilience of the method by Lukáš et al. (2006) when used for classifying pictures that have been subjected to some sort of pre-processing. The test was organized by first applying six different commonly-used image processing operations to the data set used in the previous test. Then, the identification method on the resulting data sets was applied, using, for each camera, the same reference pattern and decision threshold established in the previous test. Finally, the resulting classification was compared with the results of the classification on the original (i.e., not pre-processed) pictures. The operations that were considered in the tests, as implemented by the Adobe Photoshop software (Adobe Systems Inc. 2010), are:

- Auto Level Adjustment (ALA): this function automatically corrects the highlights and shadows in a picture and adjusts the tones so that the lowest level in the picture is completely black and the brightest white is full white. Auto Levels tune each color channel individually, and this may remove or introduce color casts.
- Auto Contrast (ACS): this function adjusts the overall contrast and mixture of colors in an image, without introducing or removing color casts, and permits to create a more accurate tonal and color-correction.
- Auto Color (ACO): this function adjusts contrast and color of an image by neutralizing the midtones and clipping the white and black pixels.
- Resizing (R75, R50, R25): this operation rescales the image to match a smaller size; the interpolation

Table 2 Decision thresholds, FRR and number of images rejected on the red channel for the tests Sub, Crop and Resize

ID	Sub			Crop			Resize		
	Decision threshold	Images rejected	FRR	Decision threshold	Images rejected	FRR	Decision threshold	Images rejected	FRR
1	0.021	20	0.111	0.008	–	–	0.008	–	–
2	0.021	29	0.161	0.007	–	–	0.007	–	–
3	0.024	18	0.1	0.018	9	0.05	0.018	9	0.05
4	0.024	6	0.033	0.025	–	–	0.025	–	–
5	0.052	1	0.005	0.046	1	0.005	0.035	–	–
6	0.026	5	0.027	0.018	–	–	0.018	–	–
7	0.013	156	0.866	0.003	138	0.766	0.003	2	0.011
8	0.037	5	0.027	0.022	3	0.016	0.014	–	–

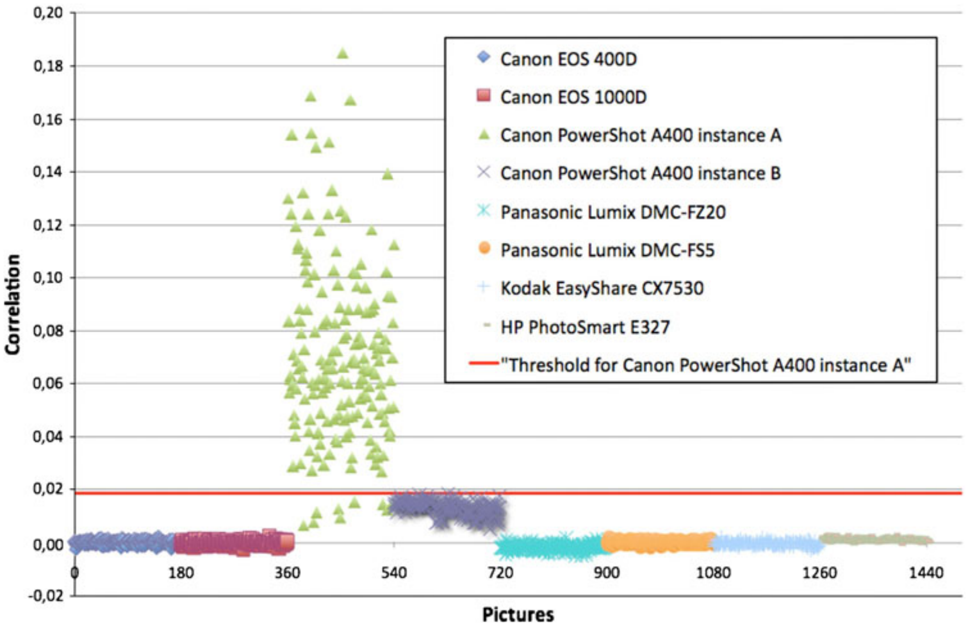

Fig. 1 Scatter plot of the correlations between all the images of the data set and the image reference pattern (IRP) of the camera with ID 3

algorithm is *bi-cubic* which produces noticeably sharper images than other methods such as *bilinear* or *nearest neighbour*, and it is a good balance between processing time and output quality. The images were processed with this operation by changing the scale factor. Pictures with the image size of 75, 50 and 25% of its original sizes were obtained.

The results of these tests, presented in Table 3, are noteworthy. A small increase in the number of erroneously rejected images can be observed when considering the pictures processed with the ALA, ACS and ACO operations. This increase is much more significant when considering the resized images. Here, the number of rejected images is high and grows linearly with the resize factor. By examining in details these results, it is worth noting that there are some camera models where the identification method performs very poorly when used with resized images. It is the case of models 3, 4, 6, and, especially, model 5. This seems to suggest either that the resize operation may have a very strong influence on the correlation between the picture and the reference pattern noise, and that this influence may vary greatly according to the camera being used, even for different cameras of the same model. Moreover, it can ben noted that if the decision thresholds are chosen using, as a reference, pictures that have not been previously pre-processed, the identification method may likely fail.

5.3 Test 3

In the previous test, it can be observed that when trying to classify pre-processed pictures using a classifier that has been tuned for unmodified pictures, the identification method by Lukáš et al. (2006) may fail, in some cases,

Table 3 Number of images rejected on manipulating pictures with thresholds computed as in Test 1 (Resize)

ID	Operation						
	No	ALA	ACS	ACO	R75	R50	R25
1	–	–	–	–	–	–	2
2	–	4	4	3	2	2	14
3	9	11	11	11	11	15	49
4	–	3	3	3	3	7	51
5	–	1	1	1	–	1	108
6	–	7	7	7	10	12	97
7	2	2	2	2	2	2	2
8	–	1	1	1	2	2	40
Total	11	29	29	28	30	41	363

with a very high probability. These failures are mostly due to the alteration of the pattern noise existing in a processed picture. This alteration implies a smaller correlation with the reference pattern noise. A possible solution to this problem is to lower the decision threshold used during the classification, so as to also correctly identify pictures with smaller correlations.

This is what has been done with Test 3. The results, documented in Table 4, show a significant improvement in the quality of the classification, with respect to the previous test. In this case, it has been possible to obtain FRR rates which are very similar to those experienced with the first test. However, such a result comes at a cost. The new decision thresholds are, in some cases, much lower than the original ones. For example, the decision threshold relating to camera 5 had to be lowered to more than 90% of its original value, thus raising the possibility of wrong classifications on larger data sets. So, this approach should be generally avoided as it tends to increase the FAR as well as the FRR.

The same behaviour can be noted when using R75, R50, and R25 operations. As shown in Table 5, even for these operations, the thresholds change without any correlation with the percentage of resize. In other words, what should have been expected with this test is that reducing the image size would decrease the correlation index. This happens only in some cases like, for example, in case of camera ID 8.

6 OSNs and OPSs experimentations

All the previous tests have been conducted on unmodified pictures or on pictures that have been intentionally modified by the end-user, using photo-editing tools. In the real world, it is quite usual to upload digital pictures on OPSs or OSNs websites, without any prior modification. This could lead to the wrong conclusion that the pictures found on these sites retain the same properties of their original counterparts and, so, that can be used for the digital identification process. Instead, OSNs and OPSs websites usually process uploaded pictures in order to reduce their size and to speed-up their handling. The arising question is: does this pre-processings puts at risk the effectiveness of the Lukáš et al. (2006) identification technique when applied to pictures retrieved from one of these sites?

Table 4 Decision thresholds, FRR and number of images rejected on the red channel for the tests ALA, ACS and ACO

ID	ALA			ACS			ACO		
	Decision threshold	Images rejected	FRR	Decision threshold	Images rejected	FRR	Decision threshold	Images rejected	FRR
1	0.008	–	–	0.008	–	–	0.009	–	–
2	0.004	2	0.011	0.004	2	0.011	0.005	2	0.011
3	0.022	11	0.061	0.019	11	0.061	0.019	11	0.061
4	0.019	2	0.011	0.02	2	0.011	0.02	2	0.011
5	0.035	–	–	0.035	–	–	0.0352	–	–
6	0.001	7	0.038	0.003	7	0.039	0.0019	7	0.039
7	0.003	2	0.011	0.003	2	0.011	0.003	2	0.011
8	0.014	–	–	0.014	–	–	0.014	–	–

Table 5 Decision thresholds, FRR and number of images rejected on the red channel for the tests R75, R50 and R25

ID	R75			R50			R25		
	Decision threshold	Images rejected	FRR	Decision threshold	Images rejected	FRR	Decision threshold	Images rejected	FRR
1	0.011	–	–	0.01	–	–	0.009	2	0.011
2	0.004	2	0.011	−0.001	–	–	0.0067	9	0.05
3	0.011	11	0.061	−0.002	–	–	0.009	11	0.061
4	0.013	2	0.011	0.009	2	0.0111	0.007	2	0.011
5	0.037	–	–	0.033	–	–	0.017	–	–
6	0.002	7	0.039	0.003	7	0.0389	0.005	13	0.072
7	0.004	2	0.011	0.006	2	0.0111	0.009	2	0.011
8	0.013	–	–	0.009	–	–	0.007	5	0.028

Table 6 Modifications performed by several OSN/OPS sites on a target image of resolution 3.888 × 2.592 pixels and size 2.275 kilobytes

Site	Updated	Resolution	Size (kB)
Facebook	Yes	720 × 480	53
Flickr	No	–	–
MySpace	Yes	600 × 399	33
PhotoBucket	Yes	1,023 × 682	131
Picasa	No	–	–
Twitpic	No	–	–

6.1 Test 4

This test was aimed at determining which OSNs and OPSs modify pictures uploaded by users. This has been done, first of all, by choosing a set of OSNs and OPSs according to their popularity. Then, several sets of pictures have been uploaded and downloaded from all these sites. The downloaded pictures have been analyzed in order to understand if, and how, they have been modified.

In Table 6 the results of one of these experiments, carried out with a sample picture of 3.888 × 2.592 pixels and size of 2.275 kilobytes, are presented. For each site, it has been checked if the picture was modified or not and, in the former case, it has been measured the size and the resolution of the modified picture.

These experiments show that only the following OSNs and OPSs, among the considered ones, process and modify uploaded pictures.

Facebook (FAC) is currently the most used OSN in the world, with more than 500 million active users. It offers a relatively simple support for uploading and sharing photos. No limit is apparently put on the size of the pictures that can be uploaded.

Currently, service administrators do not disclose any information about the way images are processed and stored on their servers. However the experiments revealed a strong compression, via downsampling, of all the pictures uploaded with a standard resolution of 720 pixels on the long edge. This is evidently done in order to cope with the huge amount of pictures uploaded daily.

Recently, service administrators have announced an upgrade on the maximum size of the images stored in the Facebook database. According to this new setting, it will be possible to upload also high-resolution images, with a maximum size of 2.048 pixels on the long edge.

Photobucket (PHB) is one of the most popular OPS with a massive audience of more than 23 million monthly unique users in the US, and over 4 million images uploaded per day from the web and smartphones (Photobucket Statistics 2011). Photobucket offers a simple support for uploading and dowloading and sharing photos. Like in the Facebook case, no information is disclosed about the way pictures are stored on their servers. However, the experiments revealed a compression process, albeit less aggressive than the one used by Facebook.

MySpace (MSP) is a OSN where users can share music, videos and pictures. No limit is put on the number and on the size of the uploadable pictures. However, even in this case, the experiments revealed a strong compression of the uploaded images, both in terms of size and downsampling.

6.2 Test 5

In this test, the Lukáš et al. (2006) identification technique has been experimented by applying it on pictures previously uploaded on the websites selected in the previous test. This experimentation has been first conducted by using the same decision threshold computed in Test 1 of the previous section (see Sect. 5.1). The results, presented in Table 7, show a substantial failure of the identification technique. On a side, this was expectable because, as already noted in the previous section, thresholds evaluated

Table 7 Number of images rejected on pictures previously uploaded on a OSN/OPS with thresholds computed as in Test 1 (Resize)

ID	OSN/OPS service			
	No	FAC	PHB	MSP
1	–	178	96	180
2	–	176	80	180
3	9	171	40	180
4	–	159	46	180
5	–	180	104	180
6	–	180	178	180
7	2	2	2	20
8	–	178	22	180
Total	11	1,224	568	1,280

using unmodified images imply bad performance when classifying modified pictures.

On the other side, however, it is worth to note that these bad performances are also much worser than the one experienced on resized pictures in Test 2 (see Sect. 5.2) and Test 3 (see Sect. 5.3), especially when processing pictures retrieved from FAC and MSP. This seems to suggest that the reason of this behavior is not only the resizing of the processed images but also to other factors such as, for example, compression tricks of the images retrieved from the considered OSNs/OPSs.

6.3 Test 6

In the last test, the Lukáš et al. (2006) identification technique has been experimented again on the same set of pictures of the previous test by using, this time, thresholds computed by means of images stored and retrieved from the considered OSNs/OPSs (a graphical representation of the updated decision thresholds is available in Fig. 2). The results, shown in Table 8, confirm a strong improvement of the identification technique with respect to the previous experiment when analyzing pictures retrieved from FAC and PHB. This is especially the case of pictures taken using camera with ID 3 and 4. A fair improvement is also evident for pictures taken using cameras with ID 5, 6 and 8. Instead, the identification is mostly uneffective when processing pictures retrieved from MSP. These differences are probably due to the different compression strategies employed by the considered OSNs/OPSs when uploading pictures. MSP is likely to be the service that adopts the most aggressive strategy, as witnessed by results presented in Table 6.

The bad performance of the Lukáš et al. (2006) identification technique on pictures retrieved from OSNs/OPSs may also be due to other reasons, apart from the image compression. As a matter of fact, it can not be excluded that OSNs/OPSs may add some kind of "watermarking" to all the photos that flow on their websites. Such possibility could have two opposite effects from the Image Forensics point of view. On the one hand, the inscription of a watermark on a picture could alter its inner structure and fool the identification process, thus leading to a wrong classification. On the other hand, if the pictures have been previously "watermarked" by the OSNs/OPSs, the eventual "discovery" of the adopted watermarking technique could give useful hints in the direction of establishing which is the OSN/OPS service that hosted the image under scrutiny. In this case, what will be assessed is the OSN/OPS that processed the image and not which camera model was used to shoot the photo.

Fig. 2 Thresholds values used according to the pre-processing operations being tested

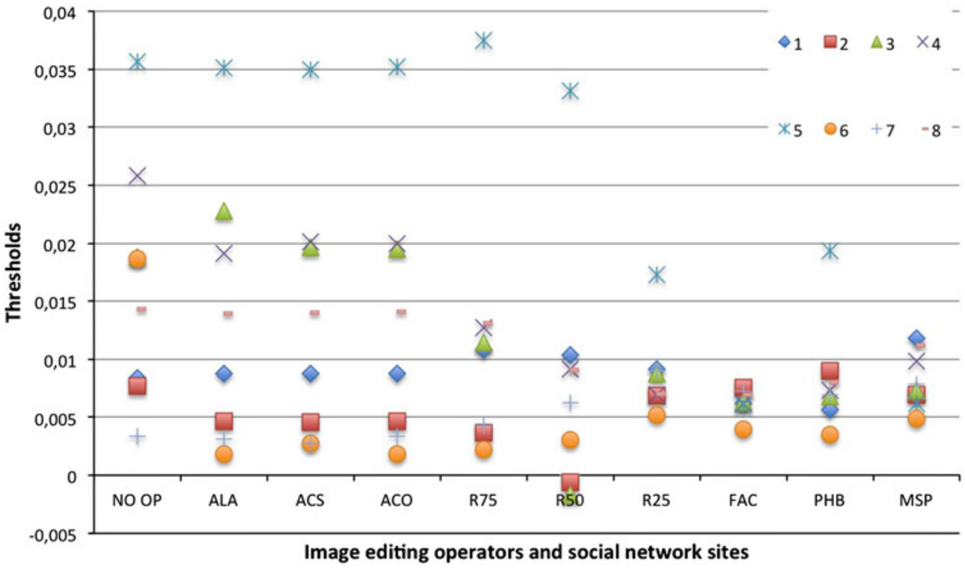

Table 8 Decision thresholds, FRR and number of images rejected on the red channel for the tests FAC, PHB and MSP

ID	FAC			PHB			MSP		
	Decision threshold	Images rejected	FRR	Decision threshold	Images rejected	FRR	Decision threshold	Images rejected	FRR
1	0.006	165	0.913	0.006	36	0.2	0.012	180	1
2	0.008	175	0.967	0.009	105	0.583	0.007	179	0.994
3	0.006	18	0.093	0.007	11	0.061	0.007	152	0.844
4	0.006	2	0.0133	0.007	2	0.011	0.01	168	0.933
5	0.006	14	0.06	0.0193	–	–	0.006	160	0.889
6	0.004	84	0.487	0.003	10	0.056	0.005	179	0.994
7	0.007	4	0.02	0.007	2	0.011	0.008	90	0.5
8	0.007	68	0.387	0.008	2	0.011	0.0112	180	1

7 Conclusions

In this paper, the effectiveness of the source camera identification technique from Lukáš et al. (2006) has been evaluated, when using, as input, pictures that have been altered by means of commonly used image processing operations. The results of the tests show, first of all, that the classification of the altered images may perform very poorly if the classifier has been tuned using unmodified images (see Sect. 5.2). A simple solution to this problem is to tune the classifier according to a data set of altered images and, consequently, by lowering the decision thresholds used to establish whether a picture has been taken with a given camera. In this new configuration, the Lukáš et al. (2006) method seems to confirm its ability to correctly identify pictures show with a given camera (see Sect. 5.3), even when processing altered images. However, this solution has an important side effect, that is it may increase the FAR as well. Moreover, there are processing operations, such as resizing and/or increasing the compression factor of a JPEG picture, which seems to have nevertheless a negative effect on the results of the classification. As an additional result of the tests, it was noted that the use of a *Resize* operation seems to be preferable to a *Crop* one when calculating the correlation between two images of different sizes.

The decrease of the threshold involves, nonetheless, several problems while choosing which one to use during a real investigation on a photographic exhibit. In fact, if the threshold computation is performed on a set of "unaltered" images, then the obtained threshold will be greater than the correlation index of a given, altered, image under scrutiny. Otherwise, if the computation of the decision threshold is computed on a set of altered images then the FRR is increased.

The research also investigates if and how OSN/OPS services modify the images that transit on their websites, and tries to establish if the method proposed by Lukáš et al. (2006) is able to correctly identify images modified in this way. The investigation has been conducted by means of three more tests. The Test 4 determines which OSN/OPS, among a set of them containing some of the most popular ones, alters the pictures. In the Test 5, the Lukáš et al. (2006) method is applied on photos that have been previously uploaded on the OSNs/OPSs scrutinized in the above step. The results show a significant inadequacy of the identification method when using the same decision threshold computed for the Test 1 (as in Sect. 5.1). Finally, the Test 6 has been conducted on the same data set of Test 2, but using threshold values precomputed on images stored and downloaded from the examined OSNs/OPSs. The results validate and confirm the expectations that the identification technique behaves better when using thresholds that have been extracted in the "correct" way. Despite that, the identification revealed to be unsuccessful in many cases, mostly because of the compression strategies employed by the considered OSNs/OPSs in order to reduce the size of uploaded images. This problem could be faced by analyzing in details the type of processing applied by the OSN/OPS services on the pictures they host and, consequently, develop *ad-hoc* image reference patterns able to cope with these transformations. Another question, worth to be investigated, is whether the other identification techniques would perform better than the one by Lukáš et al. (2006) or not in such a scenario.

Another factor, that can be further analyzed in a future work, is that the OSNs/OPSs could watermark in some way the images that flow on their websites. The watermarking operations while, from one hand, could contribute to the bad performance of the Lukáš et al. (2006) method, on the other hand could give useful hints in the direction of establishing the OSN/OPS that handled a given photo. In a few words, it would be possible to distinguish if a given photo under investigation has been posted to an OSN/OPS website just analyzing some "hidden" characteristics of the photo without relying on its "evident" origin.

Acknowledgments The authors would like to thank the chief of the CNCPO, V.Q.A. Dr. Elvira D'Amato, along with her group, for their valuable suggestions during the research phases. Their needs, questions and doubts coming from real and day-by-day investigations have encouraged the authors to further improve this work. A sincere thanks goes to the group of undergraduates who helped carry out the study.

References

Adobe Systems Inc (2010) http://www.adobe.com/products/photoshop/

Bayram S, Sencar HT, Memon ND, Avcibas I (2005) Source camera identification based on CFA interpolation. In: Proceedings of International Conference on Image Processing 2005, pp 69–72

Camera Imaging Products Association (2011) http://cipa.jp/english/index.html

Camera Labs (2010) http://www.cameralabs.com/reviews/Canon_EOS_1000D_Rebel_XS/verdict.shtml

Chen M, Fridrich JJ, Goljan M, Lukás J (2008) Determining image origin and integrity using sensor noise. IEEE Trans Inf Forensics Security 3(1):74–90

Choi KS, Lam EY, Wong KKY (2006) Source camera identification using footprints from lens aberration. In: Sampat N, DiCarlo JM, Martin RA (eds) Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol 6069, pp 172–179. doi:10.1117/12.649775

Goljan M, Fridrich J, Filler T (2009) Large scale test of sensor fingerprint camera identification. In: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol 7254. doi:10.1117/12.805701

Google Picasa (2010) http://picasatutorials.com/2009/04/picasa-tip-im-feeling-lucky/

Holst GC (1996) CCD arrays, cameras, and displays. JCD Pub, SPIE Optical Engineering Press, Bellingham. ISBN: 978-0819428530

Janesick JR, Blouke M (1995) Scientific charge-coupled devices: past, present, and future. Opt Photon News 6(4):16–20

Kharrazi M, Sencar HT, Memon ND (2004) Blind source camera identification. In: ICIP, pp 709–712

Kivanc Mihcak M, Kozintsev I, Ramchandran K (1999) Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising. In: ICASSP 1999: Proceedings of the Acoustics, Speech, and Signal Processing, IEEE International Conference, IEEE Computer Society, Washington, DC, pp 3253–3256. doi:http://dx.doi.org/10.1109/ICASSP.1999.757535

Long Y, Huang Y (2006) Image based source camera identification using demosaicking. In: 2006 IEEE 8th Workshop on Multimedia Signal Processing, pp 419–424. doi:10.1109/MMSP.2006.285343

Lukáš J, Fridrich JJ, Goljan M (2006) Digital camera identification from sensor pattern noise. IEEE Trans Inf Forensics Security 1(2):205–214

MathWorks (2010) http://www.mathworks.com/products/matlab/

Photobucket Statistics (2011) Service description. http://photobucket.com/about