European Commission

JRC TECHNICAL REPORTS

Multimodal Fingerprinting of Imaging Devices – Inception Report

Project AVICAO – Authors and Victims Identification of Child Abuse On-line

Ferrara, P

Beslay, L

2016

Joint Research Centre

EUR 28402 EN

Contact information
Laurent Beslay
Address: Joint Research Centre, Via Enrico Fermi 2749, 21027 Ispra, Italy
E-mail: laurent.beslay@ec.europa.eu
Tel.: +39 0332 78 5998

JRC Science Hub
https://ec.europa.eu/jrc

JRC105300

EUR 28402 EN

PDF ISBN 978-92-79-64956-1 ISSN 1831-9424 doi:10.2760/751517

Luxembourg: Publications Office of the European Union, 2016

How to cite this report: Ferrara P. and Beslay L., *Multimodal Fingerprinting of Imaging Devices – Inception Report*, EUR 28402 EN , doi:10.2760/751517

Contents

Abstract

This report aims at surveying the state-of-art regarding source camera identification techniques, that can be conceived as complementary to the Sensor Pattern Noise (SPN) based method, that has been already developed within the framework of AVICAO (Authors and Victims Identification of Child-Abuse Online) JRC project in collaboration with EUROPOL European Cyber Crime Centre (EC3). The purpose of AVICAO project is to boost European LEAs capabilities in fighting against Child-Abuse Online. In such context, the topic of multimodal source camera identification is a new research field of the activities carried out within the named project and consists of the finding, evaluating, selecting and integrating of new features able to come up beside the SPN in identifying source cameras of still images and video recordings. Furthermore, JRC has identified such an application suitable for combatting another serious crime, namely terrorism, whose videos are shared on the World Wide Web for propaganda purposes. The report reviews the most promising techniques available in literature, analysing in depth their properties in term of uniqueness in identifying devices, accuracy performance and stability over time. Furthermore, decision fusion techniques are also explored to provide an effective practice to combine decisions from different classifiers.

1 Introduction

Source camera identification have given a valuable demonstration of helping Law Enforcement Agencies in identifying authors and victims of Child Sexual Abuse on-line [1][2]. Briefly, camera fingerprinting techniques aim at providing a way to associate multimedia contents as pictures and/or video recordings to its source camera, namely the device that was used to capture them. More in detail, the capability to recognize the source camera can enable linking across files coming from different cases or attributing untrusted unlawful material to its potential authors, and lead to an enhanced capability to identify perpetrators and victims of such crimes.

In the previous activity carried out by JRC staff within the Authors and Victims Identification of Child Sex Abuse On-line project [1], it has been shown that a robust cue that can be used to identify the source of digital images and video is the noise pattern left by the camera. In fact, such noise pattern is univocal of a camera sensor and can be seen as a unique "fingerprint" identifying an individual device, somehow close to what happen in ballistics, wherein the analysis enables to identify the gun that has fired a given cartridge case. Regarding digital cameras, the trace is the pattern of the noise left by the camera sensor [3], which is due to the unavoidable small differences in light response of each sensitive element (pixel); these ultimately result in a deterministic pattern of small pixel intensity variations that appear in the image, much like a noise. In the scientific literature, this noise pattern is commonly referred to as Sensor Pattern Noise (SPN).

SPN has been proved to have the desired characteristics of uniqueness and stability over the time that make it a proper fingerprint of a camera device, and it has been studied widely and tested not only on source device identification [4][5][6][7][8][9][10][11], but also in different forensic applications as image forgery detection as in [5][6] and image retrieval into social networks [12][13][14].

Although scientists and industry performed large effort to yield reliable [15] and efficient [16][17] SPN-based camera identification methods, or to fight against fingerprint-copy attack [18], the performance of such an approach in terms of accuracy in device identification are satisfying only within well circumscribed settings. As already mentioned in our previous report, SPN works reliably under certain hypothesis as low compression rate and in the case in which part of the image content (as edges) doesn't affect heavily the SPN extraction procedure. Moreover, in some operational scenarios as camera identification and camera verification, the method performs quite well, whereas in more challenging ones as picture-based retrieval and image clustering, the method might not reach a performance level such that this technique could be adopted in a real procedural workflow but only during the first step of the investigation.

To overcome these limitations, it is useful to remind that the image formation process involves not only camera sensors, but also a set of stages (as lenses and Colour Filter Array interpolation, just to name a few), each of them can be representative of a single device (as SPN does), or of a camera model as well as the manufacturer that has produced the given device. All these stages leave specific footprints that can be used by analysists to characterize the device, or at least the model or the manufacturer.

In addition to this, the possibility and the appeal of producing high quality video by means of low cost devices has risen the spread of video recordings. Such a multimedia format is composed not only of a sequence of frames (images), but it come up with a related audio recording. The audio stream is itself a signal that keeps the footprints left by the audio pipeline, in the same way as for visual contents.

With such perspectives, using a multimodal approach for camera identification might solve some of the constraints related to the adoption of a standalone SPN-based method, and doubtless could improve the overall capability of a camera identification scheme. In order to reach these results, we analysed in depth the properties of the methods appeared in literature in terms of accuracy, capability of device, model or manufacturer identification, and, last but not least, the stability over time of the features employed by the available tools.

The outline of the Chapters is as follows: in Chapter 2 we reviewed image/video-based camera identification techniques, whereas in Chapter 3 we summarized microphone-based device recognition techniques. In Chapter 4, a review of the available decision fusion techniques has been presented as well. Then, in Chapter 5 we shown some operational scenario in which multimodal camera identification can be applied, and finally, Chapter 6 concludes the report and provides recommendation for the next activities.

2 Camera fingerprinting

In this Chapter, we recall most of the camera identification techniques published over the past years, after a brief reminder on the camera pipeline useful to understand the nature of each type of footprint.

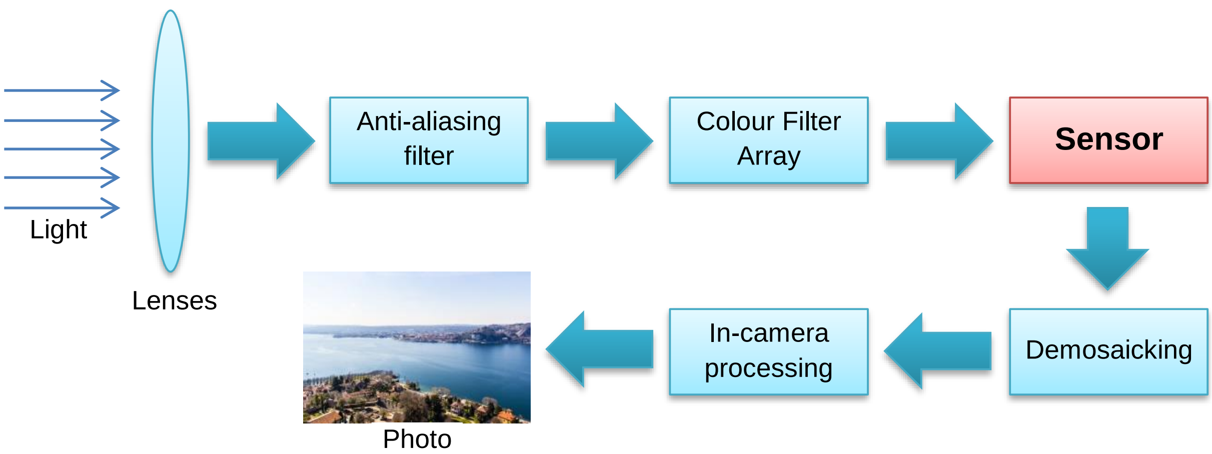
2.1 Camera pipeline

In order to more accurately understand the possible features alternative to SPN employed found in literature, it is useful to think back on how a digital camera works. Although much of the details on the camera pipeline are kept as proprietary information by the manufacturer, especially those concerning in-camera processing, the general structure and the sequence of stages within the camera pipeline is almost the same in all digital cameras, whether professional DSLR camera or low cost cameras as compact ones or embedded in mobile phones and smartphones.

The basic structure of a digital camera pipeline is shown in Figure 1. After light enters the camera through a lens (or a system of lens), a set of filter are employed, the most important of which is the antialiasing filter. The sensor is the second main component of a digital camera. The sensor measures the amount of light at each pixel location on the detector surface. In the ideal case, a separate sensor should be used for each of the colour channel (i.e. 3 sensors for RGB colour space representation), making higher the manufacturing cost of the final product. A common approach is to use a single sensor, wherein each pixel is covered by a different spectral filter acquiring one of the colour channel and evenly distributed on the sensor surface. Such filters are called Colour Filter Array and allows to record just one colour channel for pixels, whereas the other colours are missing in the raw data. To obtain the remaining colour channels, an interpolation by a neighbourhood of pixels is applied. There are a number of different interpolation strategies (whether known or proprietary) that could be used for this purpose, so that different manufactures use different interpolation techniques.

After the colour decomposition has been operated by CFA, the light is converted in digital electric signal by the sensor. Next, a number of operations are done by in-camera processing of the device, including colour interpolation (also known as demosaicing/demosaiking), gamma correction, colour balancing and, as last step, compression.

Figure 1. Digital camera pipeline



Although the depicted stages of a camera acquisition process are rather standard, the exact processing details in each step varies from one manufacturer to the other, and even from different models of the same manufacturer.

The same generic pipeline can be adopted to describe the acquisition of a video sequence, with some modifications. Devices, conceived to acquire both images and videos, employ just a portion of the entire sensor frame to record a video and, eventually, geometrical transformations are applied to adapt the size frame to the final aspect ratio (i.e. from 3:2 standard format for images to 16:9 as in full HD videos).

In addition, the compression algorithms employed to store the video are much more complex than those used for images as JPEG, and video coding standards as MPEG leaves to the manufacturer a lot of degrees of freedom in their coder design. In this sense, video compression is more tied to the manufacturer design than image compression algorithm. Besides that, the audio component of a video is acquired in a parallel chain and finally encoded and embedded inside the video recording.

2.2 Lens based camera identification

Lenses are the first stage involved in the process of image formation, allowing to collect the light reflected (or emitted) by a scene and focusing it onto the sensor surface. The effects of a lens (or, more appropriately, a system of lenses) on the final image quality can be characterized by several parameters, often related each other. Leaving out by now the focal length, which defines the angular field of view and the aperture, the most important ones are the aberrations that such optical systems unavoidably introduce in the formed image, due to the design and the manufacturing process. There are different types of aberration, the majors of which are spherical aberration, coma, astigmatism, field curvature, lens radial distortion and chromatic distortion. Such aberrations can be interpreted as footprints left inside the visual content that can be exploited in order to identify a single device or its model.

Among the aberrations introduced above, just a few have been employed for camera identification. In [19], radial distortion has been exploited to identify the source camera of still images. The method is based on the fact that radial distortion maps straight lines into curves. This phenomenon is also known as barrel or pincushion distortions. Especially in low cost devices, wherein spherical lenses are mounted to curb manufacturing costs, radial (barrel) distortion is significant, although a compensation was often applied. Moreover, barrel distortion increases when the focal length becomes smaller, whereas for longer focal length the distortion tends to be of pincushion type. In [19], Devernay's straight line method [20] has been applied to estimate radial distortion coefficients, from the edges detected within the image. Such coefficients have been used as "signature" to retrieve the source camera. In the experimental evaluation, three cameras from different manufacturers equipped with different zoom lenses have been used to evaluate the performance of a Support Vector Machine (SVM) classifier trained and tested on those features. The average classification accuracy reported is 91.5%.

In [21], the same authors have proposed a deeper analysis of their technique by extending their distortion based feature vector with those proposed by [39], achieving better performance. Moreover, authors have made some additional tests on an extended set of cameras (5), so that to investigate the effect of the focal length on the overall performance. As theory could suggest, the performance of the system drops when a longer focal length is employed, since the distortion effects begin to be negligible and the related footprint become undetectable.

Although the performance seems to be promising, further investigations are needed to understand the real capability of this kind of footprints. Firstly, an extended database would be recommended and, secondary, cameras of the same model should be tested to understand if the manufacturing process have an impact on the radial distortions of different lenses, in such a way that the system would be able to identify cameras of the same model and brand. Moreover, the test should asses the performance also for smartphone/cell phone cameras, and not only on DSLR professional ones.

Finally, the stability over time of lens based fingerprinting techniques is questionable: although these methods are related to the design and physical properties of the lenses that remain constant for a long period of their life, a problem rises in case of DSLR professional cameras, whose lenses are interchangeable, whereas this doesn't happen in case of smartphones/cell phones.

2.3 Colour Filter Array based camera identification

Beside the Sensor Pattern Noise, the Colour Filter Array (CFA), and with that demosaicing traces, have been extensively used in the problem of source camera identification. The main idea underlying such approaches is that the use of demosaicing algorithm requires to reconstruct the colours filtered by CFA introduces statistical dependencies among near pixels and between the colour channels. Since often these algorithms are manufacturers' properties and move forward continuously to achieve better colour representation, the traces that these techniques leaves can be used to recognized the model of the source camera, or at least, the brand.

A first attempt to employ this type of feature has been made in [22], in which authors proposed to model image interpolation, due to the demosaicing step, by means of the Expectation-Maximization algorithm. Due to the non-linearity of the demosaicing in most of the cameras, the method is strongly dependent on the image content. Nevertheless, an SVM classifier, exploiting this features, has shown an average accuracy of 83.33% in recognizing images taken from 3 cameras of different brand.

A step further has been made in [23], by developing a quadratic inter-pixel correlation model, in which such correlation is expressed in a quadratic form. Starting from this model, a coefficient matrix is obtained from each colour plane, and the principal components are extracted and fed to a 3-layer feed-forward neural network for camera identification. The experimental evaluation on a set of 400 uncompressed images from 4 cameras of different brands shows an average accuracy more than 90%.

The milestone of this topic is represented by [24], where authors have formalized the approach based on CFA and demosaicing. In their work authors have proposed a model to blindly estimate the parameters of CFA (i.e. the RGB pattern employed) and the interpolation kernel that better approximate the demosaicing filter. Such estimates have been then used for camera and brand identification by means of an SVM classifier. The model has been tested on an extended setting, using 19 cameras belonging to 9 different brands. The average accuracy reached in brand identification is closed to 90%.

The approach of using inter pixel correlation has been resumed in [25], but with an interesting step forward: MaKay et al. suggest to combine traces from CFA with spatial (Gaussian) noise, wavelet and error prediction features within an SVM classifier. As starting point, the experimental evaluation aims at identifying the type of image acquisition device, namely if an image had been taken from a phone camera (5 devices), a standalone camera (5), a scanner (4) or edited by means of computer graphics, achieving an average classification accuracy of 93.75%; then, the same technique is applied to identify the model/brand of each type of device, reaching an average accuracy of 97.7% for cellphone cameras, 96.2% for scanner, 94.3% for standalone cameras.

The first attempt to go beyond brand/model camera identification, and to evaluate the performance in case of device recognition, has been made in [26], wherein 3 types of features has been developed to take into account the effects of demosaicing algorithm and post-demosaicing process. These features are then processed through an Eigenfeature regularization and a feature reduction, and exploited in a Probabilistic Support Vector Machine (PSVM) to identify images from 15 cameras, some of those coming from cameras of the same model. The average accuracy reaches 99.4% in case of camera brand recognition and 94.8% in case of camera model identification, whereas the performance in case of device identification narrows to 60%, in the worst case.

The performance of the methods based on inter-pixel correlation drop dramatically when a JPEG compression is applied. Since this correlation is placed in high frequency components of the image, they are filtered out by JPEG compression, which discards high frequencies in order to compress the size of the image file. Better performance in presence of compression can be obtained by using inter-channels correlation introduced by the demosaicing algorithms, as done in [27]. Hu et al. have developed a boosted version of Ho et al.'s algorithm [28], which starts dividing the images in sub-blocks and, for each colour channel, the spectrum (DFT) is calculated for each block. Then, the difference spectra G-R and G-B are evaluated and used to estimate the variance map of these spectra. A set of shape and texture variance descriptors have been flanked to the previous feature in [27]. Finally, an AdaBoost classifier [29] is tested on a set of 50 images coming from 7 digital cameras of different models, providing an average classification accuracy of 91.90% in case of a compression factor of 70%.

In order to tackle with the cutting of the performance due to JPEG compression, in [30] an eigenalgorithms based approach has been described: the image under analysis is re-processed with a set of different strategies (called eigenalgorithms), which enable to characterize the interpolation strategy adopted by the camera that have acquired the given image, even in the presence of JPEG compression. Also in this work, an SVM classifier has been adopted to classify source camera image. Despite the relevance of the method, the performance, assessed on a data set comprising only three cameras of different models, are far from being satisfactory for a real operational workflow.

The efficiency of such kind of methods has been tackled in [31], by optimizing the techniques proposed in Swaminathan et al. work [24], without any significant improvement of source camera identification capability.

From the review of the state-of-art it is possible to preliminary conclude that this class of camera identification techniques are limited by JPEG compression: most of them are not tested or show limited performance when JPEG quality factor is less than 80%. Moreover, although in laboratory settings the methods are able to recognize the brand and the model of the source camera, device identification still remains an open issue. Finally, regarding to the stability over time, this approach seems to be quite stable, since in-camera demosaicing doesn't change during any device life, but just the response of CFA might change because of aging or damages.

2.4 Sensor dust based camera identification

Among all possible imaging devices available on the market, digital single-lens reflex (DSLR) cameras differ from the other ones in various aspects: larger and higher quality sensor, parallax-free optical viewfinder, less shutter leg, a better control over the depth field and interchangeable lenses. This last aspect characterizes the true nature of DSLRS cameras, that is the users' possibility to work with multiple interchangeable lenses. However, this attractive feature produces an undesired problem. During the mounting/unmounting of the interchangeable lens, very small particles in the environment are attracted to the camera and settle on the protective elements (as dichroic mirror or antialiasing filter) placed in front of the sensor. These tiny specks of dust, lint, or hair form a dust pattern that later affects the acquired images.

Due to the unrepeatable nature of such dust sediment process, these traces can be exploited to uniquely identify the device that have taken a given picture. This approach has been introduced in [32], and later extensively studied in [33] by the same authors. Their works begin characterizing the sensor dust pattern as almost circular spots in which the local pixel intensity decrease with respect the surrounding mean pixel intensity. More deeply, they model a single dust spot as a bi-dimensional Gaussian-shape intensity loss over image coordinates, whose parameter (mean and standard deviation) are strictly influenced by two camera parameters. While the mean represents the coordinates of the dust spot within the image and it is related to the focal length (i.e. different focal lengths maps real world points on different image points), the standard deviation represents the amount of pixel intensity loss and the spread of the spot, and it is influenced by the aperture: high apertures imply that the spots are much more out-of-focus, so that they appear more spread and the intensity loss is lower (i.e. a soft blemish); conversely, small aperture bring to strong shadow and the radius of the spot is reduced.

In order to detect the dust pattern, a normalized cross-correlation between the image and a Gaussian kernel is performed by means of a sliding convolution. The procedure is repeated by varying the standard deviation of the Gaussian kernel, to tackle with the case in which the aperture is not known. Then, these maps are combined together and a binary map is created by means of a threshold. Finally, the binary map is filtered by a means of spatial analysis of such specks in order to reduce the number of false correlations between the kernel and the image content.

The experimental assessment confirm that the method can be used for source camera identification at low false positive rates, even under compression and scaling. The average accuracy achieved by the method is 99.1% when the source camera is not available (worst case). However, just few cameras have been tested in [33] .

Although the method seems to be promising, especially because its workflow and the type of footprint recall in mind SPN based techniques, some limitations still remain. First of all, the method has been conceived and works only for DSLR cameras, and not in case of mobile devices, which nowadays are more spread than DSLR cameras. The second point is related to the need of further extensive tests to assess the capability and the limitation of such a fingerprinting method. Strictly related to this, is the stability over time, which is critical in this approach. Firstly, because some DSLR camera models might have an automatic sensor cleaning system. Secondary, user may clean the sensor to remove dust traces. Finally, some other particle can be deposited on the sensor over time. These facts limit the reliability of such methods over time. However, this doesn't mean that the method is not useful at all for camera fingerprinting purposes: an eventual matching between the dust patterns of an image and that produced by a given camera still remains an evidence.

2.5 In-camera processing based camera identification

Once the light, filtered at the previous stages, has been projected onto the sensor and converted in a digital signal, a processing chain is applied to maximize the quality of the content, comprising (automatic) colour balance, gamma correction, sharpening filtering. The sequence, the type of the algorithms employed are manufacturers' properties most of the time. Due to the variety of possible designs for in-camera processing, the footprints left by such operations can be also exploited to discriminate cameras of different models or brands. Among the operations mentioned above, the one that has been used to identify the source camera is the automatic white balance.

Lights are not equal each other: depending on the type of source, the light emitted by different sources shows distinct spectra that results in a peculiar dominant colour of the light. When a white light, which has a flat spectrum over all visible wavelengths, is reflected by a scene, the colours represented by a camera are depending only on the objects reflectance, generating an "objective" representation of the colour. Unfortunately, light sources as the Sun, incandescent or neon lamps don't emit a white light since their spectra are characterized by peaks and valleys at certain wavelengths, affecting the spectra of the light reflected by the scene and recorded by the camera. In order to achieve a consistent colour reproduction for images (or videos as well), white balance is adopted in all digital cameras. Most of the time, white balance is an automatic process in consumer-level cameras. A white balance process consists of two steps: the first one is the estimation of the source light, making some assumptions to make the process feasible when the light source is whether unknown or a combination of different sources, while the second step is the rotation and normalization of colours coordinates within a colorimetric reference system. This latter procedure is also known as colour adaptation.

In [34], automatic white balance has been used by exploiting its idempotence property. An operation having such property, produces the same output if applied once or multiple times. This means that, if the auto-white balance is applied as last step within the in-camera processing chain, and we applied the same auto-white balance method employed by the camera, then the output image would be exactly the same. It is worth to note that the assumption that auto-white balance is applied at the end of the processing chain is not true, since at least JPEG compression is applied after this operation. However, as authors highlight, in case of low JPEG compression (image quality more than 98%), the effects of compression on colours are negligible. The second assumption is that, although the auto-white balance method is not available in many cases, authors have tested various methods to approximate those that might be used inside the camera. The features employed are Image Quality Metrics [35][36] by means of which an SVM classifier has been trained and tested, after a sequential backword feature selection step. The performance has been evaluated using the "Dresden Image Database" [37], up to 29 different devices from 17 diverse models and 8 different brands. Authors tested their method in different operating scenario as brand identification (achieving an average accuracy of 99.265), model identification (98.61%) and device identification (98.57% in case of a specific camera model). The robustness of the method has been also tested, under attacks as double JPEG compression, additive Gaussian noise and resizing, so that to investigate strengths and limits of the method.

Based on the same idempotence property of auto-white balance process, in [38] a comparison between SVM and Neural Network classifiers for camera identification has been shown. The method, which operates mostly as that described above but on an extended version of the "Dresden Image Database", suggest the use of SVM (average accuracy 96.72%) instead of Neural Network (92.92%) classifiers.

Although the results provided by auto-white balance based techniques are quite impressive, some aspects need to be studied in depth. In [34], authors claim that their method is able to discriminate source cameras belonging to the same model. This fact needs to be further investigated because intuitively no difference would be present in the auto-white balance algorithm among cameras of the same model. Does the content or the focusing affect this method? If so, how much these effects make the performance lower? Moreover, in case of manual white balancing as well as in presence of post-processing operation, what are the limitation of such methods?

Finally, white balancing approaches seems to be stable over time, or at least over a long range of time. The performance may decrease due to the physical and chemical ageing of filters (in particular Colour Filter Array) over time. This aspect has not been studied yet.

2.6 Format based camera identification

Format based camera fingerprinting refers to the prediction of the source camera by exploiting the format in which the recorded image/video has been stored. Although compression is a standard process among digital cameras, the size and the quality trade-off is at the manufacturers' and users' discretion. As reference case we consider how JPEG compression, which is one of standard process in most consumer-level cameras, works: briefly, it only provides the mechanism for discarding high frequency contents while keeping low and medium frequency contents by means of a block-wise Discrete Cosine Transform (DCT) coefficients quantization, in order to reduce the size of the raw file. After the quantization stage, an integer rounding is performed and a run-length encoding is performed. This last step allows to efficiently store a long sequence of zero values. As the compression factor climbs, the quantization steps do the same and the amount of zero values increases, reducing the number of bits needed to store the image. It is worth to note that JPEG allows manufacturers to design their own quantization matrices to achieve the best trade-off between visual quality and the resulting file size. From these premises, it appears clear that quantization matrices could be used to discriminate different models or manufacturers of digital cameras. Although lots of papers have been published on JPEG quantization matrix estimation [40][41][42] in various applications, the work in [43] represent the most valuable outcome in camera fingerprinting using image formats. The method employ the JPEG quantization estimation method proposed by Friedrich et al. [44], which consists of counting, for each DCT frequency, the percentages of zero values produced by different compression schemes. Then, a SVM classifier using this feature has been trained and applied to predict the source cameras of given set of images. The experimental setting comprises 4 different cameras, whose models and manufacturers are unfortunately not mentioned to the reader. The average classification accuracy is 92%.

As done for JPEG compression in case of still images, the same approach can be adopted with video sequences. Also in this case, the lossy compression format used to store video recordings might change among cameras of different models and manufacturers. Furthermore, due to the complexity of the architecture and the degrees of freedom (spatial and temporal correlation are used), associated to a fast development over time, video compression format might be much more representative of a device model than image compression formats. Unfortunately, to the best of our knowledge, no relevant outcome has been published, except for the work presented in [45], wherein a camera identification method from video recordings has been developed. In a nutshell, the method is based on Conditional Probabilities (CP) Features, borrowed from steganalysis and already employed in [46] for still images camera identification. CP features are extracted from DCT coefficients for each 8x8 block of pixels, generating 72 statistics then converted in the CP Features. Finally, these features are used to train an SVM classifier in order to recognize the source camera. The method has been tested on 4 video sequences (1 encoded in H.264 and the remaining 3 in MPEG format) from 4 cameras of different models, achieving an average classification accuracy of 97.2% in case of moving sequences.

In terms of accuracy, format based camera fingerprint techniques have been tested on very limited and well circumscribed operating scenarios, so that the overall performance assessed in literature are most likely not representative of a real working condition. Moreover, this kind of techniques can distinguish different camera models or manufactures, but not a specific device (intra model classification). However, these fingerprinting techniques are expected to be stable over time, since coders are defined at production time and embedded within the in-camera software for all the life-cycle of the device. The main challenges are due to multiple compressions and format changing: both of them limit the accuracy of this kind of methods because a first compression estimation is needed, with a consequentially loss of performance.

As last consideration, storing capability is not a big deal in the modern digital devices, so that the majority of the images are compressed to the maximum quality (100%). This means that all JPEG quantization steps are set to 1, making undistinguishable the quantization matrices employed by different encoders. On the contrary, in case of video recordings, this trend is less accentuated, since larger frame sizes are demanded by customers so that the compression design of video sequences is rapidly evolving right now.

2.7 Blind camera identification

By using an approach similar to steganalysis, blind camera identification techniques consider a source camera as a black-box wherein all stages (lenses, CFA, sensor and in-camera processing) contribute in leaving footprints as a whole. These traces are combined together following an unknown model and, thanks to the degree of freedom in designing each stage of a camera, they might be employed in recognizing source digital devices.

The first work based on this approach is [47], starting from the assumption that the output colour image is affected mainly by CFA configuration, the demosaicing algorithm and the following colour processing. In order to capture the different patterns in the underlying colour characteristics, 34 features have been used as candidates, namely average pixel value, for each colour channel, RGB pairs correlation, neighbour distribution centre of mass, RGB pairs energy ration and wavelet domain statistics. In addition to these colours based statistics, Image Quality Metrics (IQM) have been used to model the image quality, which depends from the source cameras. Then, an SVM classifier has been employed to recognize the source camera of still images. In order to assess the performance of this method, authors have measured the classification performance over images taken from 5 cameras of different models, achieving an average accuracy of 88.02%, demonstrating the feasibility of the approach.

Over the past years, many other works [48][49] have been published trying to heighten the average accuracy achievable with this class of methods. A first attempt was that proposed in [50], where lens radial distortion coefficients have been concatenated to the 34 features proposed in [47], to reach a better camera identification. Authors made also a further step, introducing a feature selection method, to reduce the dimensionality of the problem and to select the most informative features. The method is based on a stepwise discriminant analysis, in which features are chosen iteratively to enter or leave the model according to the significance level of an F-test (that measures the ratio between the intra and inter-groups variances) from an analysis of covariance. The experimental assessment of this latter process shows that the ten most significant features in discriminating source cameras are in order from the best to the worst:

1. lens radial distortion coefficient,
2. lens radial distortion coefficient,
3. spectral phase error,
4. Czenkonowski correlation,
5. spectral magnitude error,
6. mean square error,
7. mean absolute error,
8. mean of vertical subband on green channel,
9. mean of diagonal subband on blue channel,
10. centre of mass of neighbouring distribution on red channel.

Regarding the camera identification performance, a SVM based classifier has been tested over a set of images acquired from only 3 cameras of different models, and the conclusions suggest that, by means of the data reduction, the method reaches 96.67% of average accuracy in device identification, against the 92.6% of average accuracy when all features are employed

A further extension of this kind of methods is represented by [51]: authors have provided a method based on statistical moments of 1D and 2D features aiming to describe the overall general pipeline of a camera, including JPEG compression, using a 390-D descriptor. A SVM classifier has been employed to identify the 8 camera models employed to acquire a set of 40000 images (note that some images are acquired from different cameras of the same model), achieving an average accuracy of 85.9% in camera model identification, that rises to 96.3% in case of camera brand identification. Further tests show that the use of information from all stages in image acquisition pipeline performs better than using features describing each stage separately.

As already introduced in [47], Wavelet Transform based features have been resumed in [52], wherein Wavelet features are extracted, reduced by means of a Sequential Forward Feature Selection (SFFS) algorithm and then classified using a multi-class-SVM classifier. The method has been tested on a dataset of images generated from 6 cameras of different models. A comparison with the method in [47] has been provided, showing an average accuracy of 98% against 90.9% of Kharrazi et al. in model identification.

From a different perspective, Liu et al. proposed in [53] and then extended in [54] a blind camera identification method, in which images are treated as tensor (multidimensional array) subjected to a composite signal processing, wherein all camera fingerprints (noise, linear and non-linear dependencies) introduced by each acquisition stages are viewed as a whole. The approach is based on Tucker decomposition, which allows to extract a residual matrix representing the camera footprints embedded within the image. The final features are given from the spectrum (Discrete Fourier Transform) of the residual matrix. The method has been tested on images coming from 6 cameras of different models and, as shown in [54], it shows an accuracy of classification in the range from 81.4% to 93.0%, depending from the camera model.

More recently, a camera model identification method using local binary patterns has been developed in [55]. Local binary patterns are calculated from the prediction error in the spatial domain and from HH-subband of Wavelet Transform coefficients, generating a 354-D feature comprising each colour channel. Then, also in this case, an SVM classifier has been trained and tested to identify the source cameras of an uncategorized set of images from the Dresden Database. The experimental setup comprises 18 cameras of different models. The average identification accuracy in terms of true positive rate reaches 98%, although in some cases (models of the same brand) it drops significantly.

Most of the blind approaches listed in this review have been further studied in [56], wherein a deeper performance assessment of the methods available at that time has been done, in order to evaluate the impact of post processing as shearing, histogram equalization and contrast stretching, as well as the performance on larger image database (from 10 to 19 different cameras of different model) in case of unprocessed, compressed, cropped or scaled images. The conclusions underline that the performance of the methods is affected heavily by histogram equalization and contrast stretching, more than shearing operation. Then, the size of the database has an apparent effect on the overall performance of the classifier, usually decreasing when the database become larger and depending on the type of the post processing applied, especially in case of JPEG compression.

Blind source camera identification techniques have shown satisfying results in constrained scenario, whilst, as stated in [56], the performance of this kind of approach decreases in real working condition, or have not been tested at all. Moreover, because blind techniques take into account the image acquisition pipeline as a whole, the stability over time is difficult to predict without any targeted experimental assessment because several factors (lens, noise, dust, demosaicing and processing) are encountered at the same time.

3 Decision fusion techniques

The camera identification problem is a particular application of a wider field known as Pattern Recognition. As shown within the survey of the state of art concerning camera identification techniques, a variety of different classification schemes has been developed and applied, each of them by investigating some or all of the possible footprints left by the image acquisition chain within a multimedia content (whether images, or audios or videos). The results of the experimental assessment of different designs would be then the basis for selecting one of the classifiers instead of another. However, it is also true that many studies have observed that, although one of the methods would yield the best accuracy, the sets of the data misclassified by different classifiers would not necessary overlap, even though the same features have been used. This consideration has suggested that different classifiers potentially offer complementary information about the footprints to be classified, and the conjoint employment of them could improve the final performance.

Indeed, combining different sources of information (data fusion) will in principle rise accuracy and efficiency of classifiers. In literature, we can distinguish two types of decision fusion techniques [57] :

- Feature-based: features representing statistical properties of a signal are concatenated together to form a vector (i.e. a point in a multidimensional space). Usually, the dimensionality of the vector is reduced by means of data reduction techniques as PCA, LDA and similar, in order to lighten information redundancy and so to reduce the dimensionality of the problem for efficiency and stability purposes. Finally, a classifier (e.g. Support Vector Machine, K-Nearest Neighbour, Gaussian Mixture Model, Convolutional Neuronal Network, just to name the most adopted) are then trained and applied to make the final decision;
- Score-based: the opinions of different classifiers are combined together to derive a consensus decision. The fusion can be made at the score level (soft-decision), by combining the soft output of the classifier as probabilities, likelihoods and so on, or at the label level (hard decision) that means that the fusion is performed after that each classifier has expressed its preference.

Feature-based decision fusion techniques have been applied, although not explicitly mentioned, by the overwhelming majority of the works cited in this review, and represents the straightforward solution to the problem of decision fusion. Although it has been demonstrated that this approach improves the performance of each single pattern recognition system, it doesn’t affect or, sometimes, makes the efficiency worse.

On the counter side, to the best of our knowledge, score-based techniques have not received great attention in the field of camera identification, even though they have been applied in Biometrics and Image Forensics with fairly success.

In [58], Kittler et Al. have provided a theoretical framework for classifier combination, showing that many of the basic combination schemes can be considered, under different assumptions and approximations, as special cases of their proposed approach. Alongside that, authors have made a review of the basic classifier combination rules as sum, product, maximum, minimum, median, and majority vote rules. In order to justify the effectiveness of the framework, authors tested it in different applications as facial, voice and handwritten digit recognition. Remaining in the Biometrics, in [59] authors proposed a Q-stack classifier, a classifier stacking method in which feature similarity scores obtained from the first classification step are used in combination with the quality measures as features for the second classifier. The main idea is to develop a theoretical framework for combining classifiers based on the intuition, made in [60], that quality measures can be used to improve the performance of a multimodal biometric framework. Nevertheless, the framework has been applied only to synthetic data. The same intuition has been applied in [61], defining a two-stages quality-based decision approach, composed of a score normalization and a feature selection procedure, in order to recognize faces and speakers in audio and video sequences.

In addition to Biometrics, fusion classifier methods have been employed in Image Forensics as well. In [62] a Discriminative Random Field based fusion method has been proposed for integrating multiple cues for image tampering detection. The work addresses the challenge in combining evidences from diverse tampering detection tools that explore different physical characteristics of image formation and processing pipeline. Even though classifier fusion had been already proposed in Image Forensics, Fontani et al. [63] have made a step forward, introducing, and successfully employing, the concept of "reliability" of a tool and "background information" [57] in decision fusion. Somehow, it represents the extension to Image Forensics of quality measures-based fusion approach used in Biometrics. In their work, authors proposed a decision fusion framework that exploits the results provided by the available tools in order to yield a more reliable consensus about the authenticity or not of an image. Sources of information are modelled and merged applying Dempster-Shafer Theory of Evidence [63], which allows the analyst to handle the uncertainty (or reliability) concerning tools better than the classical Bayesian approach and, moreover, to exploit all the available information about the compatibility between footprints the tools use. The main constrained of such an approach is that it requires cues to be independent.

4 Evaluation protocol

This Chapter aims at setting the guidelines to define and evaluate a multimodal camera fingerprinting framework, in order to explore feasibility and limitations in a scenario close to a real operational workflow. To motivate the choice of the algorithms used for the proposed multimodal camera fingerprinting and to establish a meaningful evaluation protocol, we summarize in Table 1 the capabilities of the aforementioned methods in terms of brand, model and device identification as well as the supposed stability over time.

Table 1. A comparison between camera fingerprints properties

	Brand	**Model**	**Device**	**Stability**
Lens-based fingerprinting	Yes	Yes	To be checked	Limited in DSLR cameras
CFA-based fingerprinting	Yes	Yes	To be checked	Highly probable
Dust-based Fingerprinting	Yes	Yes	Yes	Limited
In-camera processing-based fingerprinting	Yes	Yes	To be checked	Yes
Format-based fingerprinting	Yes	Yes	No	Yes
Blind fingerprinting	Yes	Yes	Highly probable	To be checked

Even though lens-based fingerprinting techniques are not completely reliable in the specific case of DSLR cameras, past researches have shown their relevance in the camera identification problems and potentially enabling to identify a specific device, especially handheld devices such as cell-phones and smartphones. The same relevance has been proved for CFA footprints, except for the capability of identifying a single device which seems to be much more challenging in this latter case. Regarding to the dust traces, their application is mainly limited to DSLR cameras, and for a limited period of time. Instead, in-camera processing footprints are proved to be stable over the time, but the device identification capability is quite questionable, even if in [34] is asserted that this task is feasible. Format based device fingerprinting is not able to identify the device, and, to the best of our knowledge, no relevant contribution has been produced in case of video format fingerprinting. However, because high JPEG compression is the main factor that limits the performance of all the other techniques and, at the same time, high compression rates leave stronger footprints, this class techniques is somehow complementary to the others. Finally, blind fingerprinting techniques need to be further investigated and tested, since there's no an underlying comprehensive model so that their capabilities have to be tested "in place".

The choice of suitable techniques to build an enhanced camera identification framework is influenced also by the choice of the fusion method. In particular, feature-based fusion techniques are general and suitable for many applications, included the multimodal camera fingerprinting. Instead, a Dempster-Shafer fusion scheme is not applicable if blind fingerprinting techniques were adopted along with other camera based techniques, because such a theory assumes that the cues have to be independent information sources. On the other hand, a straightforward application of an SVM classifier to SPN is critical, due to the high dimensionality of SPN fingerprints.

By following these considerations, we choose to build a multimodal fingerprinting technique comprising:

- Sensor Pattern Noise
- Lens footprints
- CFA-based footprints
- Format based footprints

The most promising known techniques must be employed together, and eventually improved, to outperform the standalone SPN-based fingerprinting technique in [1]. The fusion technique adopted is the feature-based we introduced in Chapter 3.

4.1 Operational scenarios

In order to assess the performance of multimodal device fingerprinting, two operational scenarios are considered:

- **Scenario 1:** Camera identification;
- **Scenario 2:** Camera verification.

Briefly, camera identification is a scenario in which the task is to associate a given picture to the camera which has acquired it, by choosing the right one within a set of cameras available to the investigators. This scenario has two important requirements: The first is that the true camera is supposed to be in principle within the set of known cameras. The second is that the investigator should have direct access to each camera involved in the test, in order to generated reliable fingerprints characterizing the device. The accuracy in camera identification can be measured by using *Cumulative Matching Characteristic* (CMC) curves, which measure the cumulative probability, estimated over a test set, of finding the correct match within a given number of ranks (from the 1^{st} rank to the N^{th} rank).

Instead, camera verification scenario is a 1-vs-1 comparison between one picture and one camera. It corresponds to the operational case in which the analyst wants to verify whether a given device has been used in order to take a certain picture P. The answer provided is therefore binary (Yes/No). Note that scenario can also involve more than one camera: in this case, different one-vs-one comparisons will be performed. The difference with respect to camera identification is that in this case the true camera must be within the set of cameras involved in the test. The performance can be measured in terms of miss-detection (false negative) and false alarm (false positive) probabilities and can be described by means of *Receiver Operating Characteristic* (ROC) curves.

4.2 Image and video benchmark dataset

In order to achieve a fair comparison with the results provided by the SPN based device fingerprinting technique developed by JRC staff in collaboration with EUROPOL EC3 unit, a benchmark dataset close to that proposed in [1] should be employed. The dataset should be composed of at least thirty smartphones of different brands and models, comprising some devices of the same model/brand. Each device must be used to acquire 100 test images and 10 template images showing a clean background to extract SPN signature. Besides that, a corpus of video sequences must be recorded to assess the performance in a more challenging scenario as video recordings source identification.

Table 2. List of devices

ID	Brand	Model	OS	Native resolution
1	Apple	IPhone 4	iOS 4.3.3	2592 x 1936 pixels
2	Apple	IPhone 4	iOS 4.3.3	2592 x 1936 pixels
3	Apple	IPhone 4	iOS 5.0	2592 x 1936 pixels
4	Apple	IPhone 4	iOS 5.0	2592 x 1936 pixels
5	Apple	IPhone 4	iOS 5.0	2592 x 1936 pixels
6	Apple	IPhone 4S	iOS 6.1.2	3264 x 2448 pixels
7	Apple	IPhone 4S	iOS 6.1.3	3264 x 2448 pixels
8	RIM Blackberry	Bold 9900	BBOS	2560 x 1920 pixels
9	RIM Blackberry	Bold 9900	BBOS	2560 x 1920 pixels
10	RIM Blackberry	Torch 9800	BBOS	2592 x 1944 pixels
11	HTC	7 Mozart	Windows OS	3264 x 2448 pixels
12	HTC	One X	Android 4.0.3	3264 x 1840 pixels
13	HTC	One X	Android 4.0.3	3264 x 1840 pixels
14	HTC	One X	Android 4.0.4	3264 x 1840 pixels
15	HTC	One X	Android 4.0.4	3264 x 1840 pixels
16	HTC	One X	Android 4.1.1	3264 x 1840 pixels
17	Motorola	Milestone 2 Motoblur	Android 2.2	2592 x 1936 pixels
18	Samsung	Galaxy Nexus	Android 4.1.1	2592 x 1944 pixels
19	Samsung	Galaxy Nexus	Android 4.1.1	2592 x 1944 pixels
20	Samsung	Galaxy Nexus	Android 4.2.1	2592 x 1944 pixels
21	Samsung	Galaxy Nexus	Android 4.2.2	2592 x 1944 pixels
22	Samsung	Galaxy Nexus	Android 4.2.2	2592 x 1944 pixels
23	Samsung	Galaxy ACE	Android 2.2.1	2560 x 1920 pixels
24	Samsung	Galaxy S3	Android 4.1.2	3264 x 2448 pixels
25	Samsung	Galaxy S4	Android 4.2.2	4128 x 3096 pixels

26	Samsung	Nexus S	Android 2.3.3	2560 x 1920 pixels
27	Simvalley	SPX-5 dual SIM	Android 2.3.5	3264 x 2448 pixels
28	Sony	Xperia S	Android 4.0.4	4000 x 3000 pixels
29	Sony	Xperia S	Android 4.0.4	4000 x 3000 pixels
30	SonyID 21	Xperia S	Android 4.0.4	4000 x 3000 pixels
31	Sony	Xperia S	Android 4.0.4	4000 x 3000 pixels
32	Sony	Xperia Sole	Android 4.1	2592 x 1944 pixels

In **Error! Reference source not found.** the list of devices with their specifications is p rovided.

4.3 Testing protocols

For each operational scenario, a different testing protocol has been designed according to the type of the experiment and the performance measurements adopted. The details are described in the following.

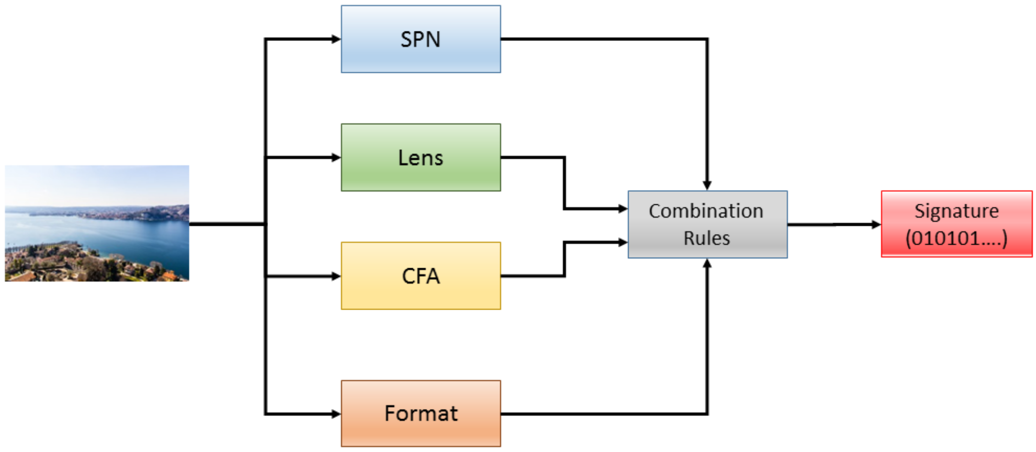
Camera Identification

For each camera, a template signature is extracted by combining the output of each tools (SPN, lens-based, CFA-based and format based), as depicted in Figure 1. To obtain a more reliable signature, the process should be repeated over at least 10 pictures for each devices. Then, the signature is extracted from the probes images (i.e. the images under investigation), to identify the right camera by means of a matching against all template signatures. The same procedure must be duplicates for video recordings, by extracting I-frames, which are less compressed and so more reliable than others, to extract the aforementioned signature. Devices are finally rank according the normalized cross-correlation value. A CMC curve is finally generated to summarize the performance.

Camera Verification

In this scenario, a signature is generated for each device and for each probe image, as explained in the previous scenario. After the matching procedure, the resulting score is stored and then compared with a moving threshold. So that, for each value of the threshold, false negatives and false positives are encountered and a ROC curve is generated.

Figure 2. Operational workflow of the multimodal fingerprinting method to extract the signature from an image



4.4 Further tests

The previous Section describes testing protocols for assessing multimodal device fingerprinting. Further tests can be made after this baseline procedure has been performed. Firstly, the aforementioned procedure should be repeated by varying the compression (both in case of still images and videos), which is the most common operation limiting the performance of the proposed tools. Then, the multimodal approach should be extended to more challenging operational scenarios, such as reference-based image retrieval, pictures-based image retrieval and clustering [1]. Moreover, the tests should be repeated to make a comparison between different score-based combination rules from Biometrics [58] as well as Dempster-Shafer Theory fusion framework, in order to assess the capability and the limitation of proposed approach in case of scored-based fusion strategies.

5 Conclusions

This report offered an exhaustive review of the state-of-art about multimodal camera fingerprinting techniques and the available fusion strategies, by exploring their own potentialities and limitations. Moreover, it described the scene and set a preliminary evaluation protocol for a deeper understanding on the matter. Starting from the outcome of works already fulfilled by the JRC on SPN-based fingerprinting, the document discussed scenarios and challenges to employ such an approach within an operational workflow. Two operational scenarios have been stated to explore the feasibility of this approach, as well as two testing protocols have been detailed for a comprehensive assessment of the performance.

Future research activities are foreseen based on this preliminary report. After collecting a suitable images and videos corpus, an experimental evaluation will be performed, by following the guidelines mentioned in Chapter 4. From the analysis of the tests results, the most significant limitations will be collected and analytically discussed. Such a study will feed further research activities on device fingerprinting, in order to overcome the possible limitations, and to propose new effective solutions. Some research directions have been already intuited, as the employment of audio tracks associated to video recordings in order to add a new cue, which could be exploited for device fingerprinting from video recordings.

References

[1] R. Satta, L. Beslay, "*Camera fingerprinting as a tool for combatting Child Abuse on-line*" JRC Technical Report, JRC93821, European Commission – Joint Research Centre, 2015.

[2] R. Satta, J. Galbally, and L. Beslay, "State-of-the-art review: video analytics for fight against on-line child abuse," JRC Technical Report, JRC85864, European Commission – Joint Research Centre, 2013.

[3] J. Lukas, J. Fridrich, and M. Goljan, "*Digital camera identification from sensor pattern noise*," IEEE Transactions on Information Forensics and Security, 1(2):205–214, November 2006.

[4] C.-T. Li and R. Satta, "*Empirical investigation into the correlation between vignetting effect and the quality of sensor pattern noise*," IET Computer Vision, 6:560–566(6), November 2012.

[5] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, "*Determining image origin and integrity using sensor noise*," IEEE Transactions on Information Forensics and Security, 3(1):74–90, March 2008.

[6] J. Fridrich, "*Digital image forensic using sensor noise*," IEEE Signal Processing Magazine, 26(2):26–37, 2009.

[7] X. Kang, Y. Li, Z. Qu, and J. Huang, "*Enhancing source camera identification performance with a camera reference phase sensor pattern noise*," IEEE Transactions on Information Forensics and Security, 7(2):393–402, 2012.

[8] C.-T. Li, "*Source camera identification using enhanced sensor pattern noise*," IEEE Transactions on Information Forensics and Security, 5(2):280–287, June 2010.

[9] C.-T. Li and Y. Li, "*Color-decoupled photo response non-uniformity for digital image forensics*," IEEE Transactions on Circuits and Systems for Video Technology, 22(2):260–271, February 2012.

[10] C.-T. Li and R. Satta, "*On the location-dependent quality of the sensor pattern noise and its implication in multimedia forensics*," In Proceedings of the 4th International Conference on Imaging for Crime Detection and Prevention 2011 (ICDP 2011), London, UK, 2011.

[11] J. Lukas, J. Fridrich, and M. Goljan, "*Digital camera identification from sensor pattern noise*," IEEE Transactions on Information Forensics and Security, 1(2):205–214, November 2006.

[12] R. Satta and P. Stirparo, "*Picture-to-identity linking of social network accounts based on sensor pattern noise*," In Proceedings of the 5th International Conference on Imaging for Crime Detection and Prevention (ICDP 2013), London, UK, 2013.

[13] R. Satta and P. Stirparo, "*On the usage of Sensor Pattern Noise for Picture-to-Identity Linking through Social Network Accounts*," in Proceedings of the 9th International Conference on Computer Vision Theory and Applications (VISAPP 2014), Lisbon, Portugal, 2014.

[14] A. Castiglione, G. Cattaneo, M. Cembalo, U. F. Petrillo, "*Experimentations with source camera identification and online social networks*," Journal of Ambient Intelligence and Humanized Computing 4 (2)265-274, 2013.

[15] Z. Yu et al., "*Camera identification for very low bit rate time varying quantization noise videos*," 2014 9th International Symposium on Communication Systems, Networks & Digital Sign (CSNDSP), Manchester, 2014, pp. 208-212.

[16] D. Valsesia, G. Coluccia, T. Bianchi, E. Magli, "*Compressed fingerprint matching and camera identification via random projections*," Information Forensics and Security, IEEE Transactions on 10 (7) 1472-1485, 2015.

[17] G. Cattaneo, G. Roscigno, U. F. Petrillo, "*A scalable approach to source camera identification over hadoop*," in: IEEE International Conference on 430 Advanced Information Networking and Applications (AINA), IEEE, pp. 366-373, 2014.

[18] H. Zeng; J. Liu; J. Yu; X. Kang; Y. Q. Shi; Z. J. Wang, "*A Framework of Camera Source Identification Bayesian Game*," in IEEE Transactions on Cybernetics, vol.PP, no.99, pp.1-12

[19] K. S. Choi, E. Y. Lam, K. K. Y. Wong, "*Source camera identification using footprints from lens aberration*," Proc. SPIE 6069, Digital Photography II, 60690J (February 10, 2006)

[20] F. Devernay, O. Faugers, "Automatic calibration and removal of distortion from scenes of structured environments", in Investigative and Trial Image Processing, Proc. SPIE 5685, pp. 62-67, 1995.

[21] K. S. Choi, E. Y. Lam, K. K. Y. Wong, "*Automatic source camera identification using the intrinsic radial distortion,*" Opt. Express 14, 11551-11565 (2006)

[22] S. Bayram, H. Sencar, N. Memon and I. Avcibas, "*Source camera identification based on CFA interpolation*," IEEE International Conference on Image Processing 2005, 2005, pp. III-69-72.

[23] Y. Long and Y. Huang, "*Image Based Source Camera Identification using Demosaicking*," 2006 IEEE Workshop on Multimedia Signal Processing, Victoria, BC, 2006, pp. 419-424.

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

[25] C. McKay, A. Swaminathan, Hongmei Gou and M. Wu, "*Image acquisition forensics: Forensic analysis to identify imaging source*," 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, 2008, pp. 1657-1660.

[26] H. Cao and A. C. Kot, "*Mobile camera identification using demosaicing features*," Proceedings of 2010 IEEE International Symposium on Circuits and Systems, Paris, 2010, pp. 1683-1686.

[27] Y. Hu, C. T. Li, X. Lin and B. b. Liu, "*An Improved Algorithm for Camera Model Identification Using Inter-channel Demosaicking Traces*," 2012 Eighth International

Conference on Intelligent Information Hiding and Multimedia Signal Processing, Piraeus, 2012, pp. 325-330.

[28] J. S. Ho, O. C. Au, J. Zhou and Y. Guo, "*Inter-channel demosaicking traces for digital image forensics*," 2010 IEEE International Conference on Multimedia and Expo, Suntec City, 2010, pp. 1475-1480.

[29] http://www.boosting.org/tutorials

[30] S. Milani, P. Bestagini, M. Tagliasacchi and S. Tubaro, "*Demosaicing strategy identification via eigenalgorithms*," 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, 2014, pp. 2659-2663.

[31] X. Zhao and M. C. Stamm, "*Computationally efficient demosaicing filter estimation for forensic camera model identification*," 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 2016, pp. 151-155.

[32] A. E. Dirik, H. T. Sencar and N. Memon, "*Source Camera Identification Based on Sensor Dust Characteristics*," 2007 IEEE Workshop on Signal Processing Applications for Public Security and Forensics, Washington, DC, USA, 2007, pp. 1-6.

[33] A. E. Dirik, H. T. Sencar and N. Memon, "*Digital Single Lens Reflex Camera Identification From Traces of Sensor Dust*," in IEEE Transactions on Information Forensics and Security, vol. 3, no. 3, pp. 539-552, Sept. 2008.

[34] Zhonghai Deng, A. Gijsenij and Jingyuan Zhang, "*Source camera identification using Auto-White Balance approximation*," 2011 International Conference on Computer Vision, Barcelona, 2011, pp. 57-64.

[35] I. Avcibas, N. Memon, B. Sankur, "*Steganalysis using image quality metrics*," IEEE Transactions on Image Processing, 12(2):221-229, 2003.

[36] A. Eskicioglu, P. Fisher, "Image quality measures and their performance," IEEE Transactions on Communications, 43(12):2959-2965,1995.

[37] T. Gloe, R. Boheme, "The 'dresden image database' for benchmarking digital image forensics," in ISAC'10: Proceedings of the 2010 ACM Symposium on Applied Computing, pp 1584-1590, New York, NY, USA, 2010.

[38] S. Arathy, D. S. Vidyadharan, C. Balan and T. Sobha, "*Auto White Balancing and comparison of Support Vector Machine and neural network classifiers in prediction of source camera*," 2013 International Conference on Control Communication and Computing (ICCC), Thiruvananthapuram, 2013, pp. 96-101.

[39] M. Kharrazi, H. T. Sencar, N. Memon, "*Blind source camera identification,*" in International conference on Image Processing, pp. 709-712, 2004

[40] Z. Fan, R.L. de Queiroz, "*Identification of bitmap compression history: JPEG detection and quantizer estimation*," IEEE Transactions on Image Processing, vol. 12, pp. 230-235, 2003.

[41] J. Lukas, J. Friedrich, "*Estimation of primary quantization matrix in double compressed JPEG images,*"in Proc. Of. DFRWS, 2003

[42] D. Fu, Y. Q. Shi, W. Su, "*A generalized Benford's law for JPEG coefficients and its applications in image forensics*," Proceedings of SPIE, vol. 6505, pp. 65051L-65051L-11, 2007.

[43] K. S. Choi, E. Y. Lam and K. K. Y. Wong, "*Source Camera Identification by JPEG Compression Statistics for Image Forensics*," TENCON 2006 - 2006 IEEE Region 10 Conference, Hong Kong, 2006, pp. 1-4.

[44] J. Fridrich, M. Goljanb, R. Dub, "*Steganalysis based on JPEG compatibility*", in Multimedia Systems and Applications IV, ser. Proc. SPIE, vol 4518, 2001, pp.275-280.

[45] S. Yahaya, A. T. S. Ho and A. A. Wahab, "*Advanced video camera identification using Conditional Probability Features*," IET Conference on Image Processing (IPR 2012), London, 2012, pp. 1-5.

[46] A. W. A. Wahab, P. Bateman, "*Image Source Identification by using Conditional Probability Features*," International Journal of Cryptology Research, volume 2, no. 1, pp. 63-71, (2010).

[47] M. Kharrazi, H. T. Sencar and N. Memon, "B*lind source camera identification*," Image Processing, 2004. ICIP '04. 2004 International Conference on, Singapore, 2004, pp. 709-712 Vol. 1.

[48] M.-J. Tsai and G.-H. Wu, "*Using Image Features to Identify Camera Sources*," 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, Toulouse, 2006, pp. II-II.

[49] M. J. Tsai, C. L. Lai and J. Liu, "*Camera/Mobile Phone Source Identification for Digital Forensics*," 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07, Honolulu, HI, 2007, pp. II-221-II-224.

[50] K. S. Choi, E. Y. Lam and K. K. Y. Wong, "*Feature Selection in Source Camera Identification*," 2006 IEEE International Conference on Systems, Man and Cybernetics, Taipei, 2006, pp. 3176-3180.

[51] Guanshuo Xu, Y. Q. Shi and Wei Su, "*Camera brand and model identification using moments of 1-D and 2-D characteristic functions*," 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, 2009, pp. 2917-2920.

[52] B. Wang, Y. Guo, X. Kong and F. Meng, "*Source Camera Identification Forensics Based on Wavelet Feature*s," 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Kyoto, 2009, pp. 702-705.

[53] M. Liu, N. Yu and W. Li, "*Camera Model Identification for JPEG Images via Tensor Analysis*," 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Darmstadt, 2010, pp. 462-465.

[54] W. Li, N. Yu and Y. Yuan, "*Identifying camera and processing from cropped JPEG photos via tensor analysis*," 2010 IEEE International Conference on Systems, Man and Cybernetics, Istanbul, 2010, pp. 3889-3895.

[55] G. Xu and Y. Q. Shi, "*Camera Model Identification Using Local Binary Patterns*," 2012 IEEE International Conference on Multimedia and Expo, Melbourne, VIC, 2012, pp. 392-397.

[56] Ying-Chu Chen, Yongjian Hu and Chang-Tsun Li, "*Further studies on forensic features for source camera identification*," 4th International Conference on Imaging for Crime Detection and Prevention 2011 (ICDP 2011), London, 2011, pp. 1-6.

[57] M. Fontani, E. Argones-Rúa, C. Troncoso and M. Barni, *"The watchful forensic analyst: Multi-clue information fusion with background knowledge,"* 2013 IEEE International Workshop on Information Forensics and Security (WIFS), Guangzhou, 2013, pp. 120-125.

[58] J. Kittler, M. Hatef, R. P. W. Duin and J. Matas, *"On combining classifiers,"* in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 3, pp. 226-239, Mar 1998.

[59] K. Kryszczuk, A. Drygajlo, *"Q-stack:UNi- and multimodal classifier stacking with quality measures,"* in Proc. Of the 7th International Workshop on Multiple Classifier Systems, MCS, 2007, pp.367-376.

[60] K.-A. Toh, W.-Y. Yau, E. Lim, L. Chen, C.-H. Ng, "*Fusion of Auxiliary Information for Multi-modal Biometrics Authentication*," ICBA, 2004.

[61] E. Argones-Rua, J. l. Alba-Castro and C. Garcia-Mateo, "*On the use of quality measures in face and speaker identity verification based on video and audio streams*," in IET Signal Processing, vol. 3, no. 4, pp. 301-309, July 2009.

[62] Y. F. Hsu and S. F. Chang, *"Statistical fusion of multiple cues for image tampering detection*," 2008 42nd Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, 2008, pp. 1386-1390.

[63] M. Fontani, T. Bianchi, A. De Rosa, A. Piva and M. Barni, "*A Framework for Decision Fusion in Image Forensics Based on Dempster–Shafer Theory of Evidence*," in IEEE Transactions on Information Forensics and Security, vol. 8, no. 4, pp. 593-607, April 2013.

[64] G. Shafer, "*A Mathematical Theory of Evidence*," Princeton, NJ, USA: Princetone Univ. Press. 1976

List of abbreviations and definitions

AVICAO (Project)	Authors and Victims Identification of Child Abuse On-line
JRC	Joint Research Centre
LEA(s)	Law Enforcement Agency(s)
SPN	Sensor Pattern Noise
CFA	Colour Filter Array
DSLR	Digital Single-Lens Reflex
HD	High Definition
RGB	Red-Green-Blue
DFT	Discrete Fourier Transform
DCT	Discrete Cosine Transform
SVM	Support Vector Machine
JPEG	Joint Photographic Expert Group
MPEG	Moving Picture Experts Group
H.264	MPEG-4 AVC (Advanced Video Codec)
IQM	Image Quality Metrics
SFFS	Sequential Forward Feature Selection
HH-subband	High-High frequencies in Wavelet Domain
PCA	Principal Component Analysis
LDA	Linear Discriminative Analysis
CMC (curves)	Cumulative Matching Characteristic (curves)
ROC (curves)	Receiver Operating Characteristic (curves)
EC3	European Cyber Crime Centre

List of figures

List of tables

KJ-NA-28402-EN-N

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office

doi:10.2760/751517

ISBN 978-92-79-64956-1