

Smartphone Video Source Identification Based on Sensor Pattern Noise

Raquel Ramos López, Anissa El-Khattabi, Ana Lucila Sandoval Orozco, Luis Javier García Villalba

Abstract—An increasing number of mobile devices with integrated cameras has meant that most digital video comes from these devices. These digital videos can be made anytime, anywhere and for different purposes. They can also be shared on the Internet in a short period of time and may sometimes contain recordings of illegal acts. The need to reliably trace the origin becomes evident when these videos are used for forensic purposes. This work proposes an algorithm to identify the brand and model of mobile device which generated the video. Its procedure is as follows: after obtaining the relevant video information, a classification algorithm based on sensor noise and Wavelet Transform performs the aforementioned identification process. We also present experimental results that support the validity of the techniques used and show promising results.

Keywords—Digital video, forensics analysis, key frame, mobile device, PRNU, sensor noise, source identification.

I. INTRODUCTION

IMAGES captured by electronic devices (i.e. smartphones) are often considered part of evidence in Court, and in a few minutes a video can communicate an enormous amount of information. According to the traffic meter “Alexa, The Web Information Company” [1], YouTube is currently the third most visited website in the world, which gives us a clear indication of the online popularity of videos. Video is widely used in everyday life due to the availability of a wide range of mobile devices that can reproduce and/or record it, such as mobile phones, tablets, portable game consoles and digital cameras or camcorders. As for mobile devices, Gartner Inc. [2], states that sales of smartphones grew by 36% in the fourth quarter of 2013, and represented 57.6% of the global sales of mobile phones in the fourth quarter, compared to 44% with respect to 2012. As digital cameras have swept away traditional film cameras in terms of popularity, nowadays mobile devices equipped with cameras have an important role in putting an end to the rapid growth that digital cameras previously experimented. A report by IC Insights [3] predicted that by 2016 the market rate of DSCs (Digital Still Camera) will drop from 47% in 2012 to 27%; it also predicts a rise in sales of digital cameras built into smartphones, PCs and tablets, from 31% in 2012 to 42% by 2016.

R. Ramos López, A. El-Khattabi and A. L. Sandoval Orozco are with Group of Analysis, Security and Systems (GASS), Department of Software Engineering and Artificial Intelligence (DISIA), School of Computer Science, Office 431, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases s/n, Ciudad Universitaria, 28040 Madrid, Spain.

L. J. García Villalba is with Group of Analysis, Security and Systems (GASS), Department of Software Engineering and Artificial Intelligence (DISIA), School of Computer Science, Office 431, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases s/n, Ciudad Universitaria, 28040 Madrid, Spain (e-mail: javiergv@fdi.ucm.es).

Due to the frequent use of mobile devices, in some cases there exist legal restrictions or limitations to their use in various locations, such as schools, universities, government offices, companies, etc. In parallel, videos are increasingly used, either directly or indirectly, in legal proceedings as evidence for law enforcement [4]. Therefore, given the increasing importance of video, digital video forensics are particularly relevant. Their main goal is the acquisition and analysis of digital video in order to find forensically sound evidence, generally while investigating a crime. Within this discipline, *Digital Video Integrity* aims to establish whether a digital video has been tampered with, *Digital Video Steganography* studies if a video contains hidden data and *Video Source Camera Identification* aims to identify which specific camera has been used to capture a video.

Video Source Camera Identification has many applications in real world scenarios, and its study is especially important and becoming more relevant with every passing day. For example, when a video is presented as evidence in a court of law, identifying the acquisition device of the video could be as important as the video itself. Not doing this in a forensically sound way can lead to legal challenges and render the evidence invalid [5]. Additionally, images or videos shared through social networks (Flickr, Instagram, Facebook, Twitter, etc.) or personal email can be authenticated and linked to the device (in this case, the smartphone or digital camera). This paper presents a combination of forensic analysis techniques for the identification of a video source device, but focusing on videos generated by mobile devices, mostly smartphones.

The paper is divided into six sections, the first being this introduction. Section II presents the differences between the pipeline in the creation of an image and a video. Section III introduces a state of the art for the forensic analysis of images and videos, regarding the issue of source acquisition identification. The proposed technique is presented in detail in Section IV. The supporting experiments are presented in Section V. Finally, Section VI shows the conclusions drawn from this work.

II. SOURCE ACQUISITION IDENTIFICATION TECHNIQUES

Most research in the field of source identification has been focused on photographic images. However, there is increased need for research to find solutions to the forensic issues characteristic of video, as they have some peculiarities that need to be dealt with and a wide range of alterations that can be applied to them. Most of the forensic analysis techniques developed for images can also be applied to video, starting with operating over individual video frames [6].

In [7], a quite detailed comparison of the major techniques for source acquisition identification is presented. These are divided into five approaches, namely those based on metadata, image features, CFA array and color interpolation defects, sensor imperfections, and wavelet transforms.

The research area based around the study of metadata is largely dependent on the data inserted by the manufacturer when the image is created. The most widespread specification is Exif, and has two useful specific tags: "Make" and "Model". Unfortunately, adding metadata to the image is by no means mandatory.

Among the existing methods which are based on sensor imperfections, there are two main branches from which pixel or sensor pattern noise can be studied. In [8] it was shown that camera sensors generate pattern noise (Sensor Pattern Noise) which could be used as the sole method of identification.

In [9] it was shown that the extracted image sensor noise could be severely contaminated by the details of specific scenes. To deal with this problem, a new approach to mitigate the influence of the scene details was proposed, thus improving the success rate. In the experiments, 9 cameras, with 320 photos from each camera were used, with varying outdoor and indoor scenes.

Finally, in the area of wavelet transforms, there are various approaches, for example, [10] proposes a new identification technique based on conditional probability features. Such features were initially proposed for steganalysis purposes in [11]. The set of experiments was performed with 4 different iPhone cameras, proving that the technical proposal works well for different models of the same camera. Accuracies of 98.6%, 97.8% and 92.5% were obtained in the classification of 2, 3 and 4 iPhones, respectively, with an image crop of 800 by 600. This approach, unfortunately, does not look too promising when images are preprocessed.

In [12] it is determined that the use of sensor pattern noise together with the wavelet transform is an effective method for source identification, reaching an average success rate of 87.21%. This method is used to identify smartphones (based on their built-in cameras).

In the case of the development of techniques for video source acquisition identification, there are very few academic works in this area. Some are directly based on the encoding sequence, and others on frame extraction for later applying some classification method for still images.

Reference [13] proposes an algorithm based on motion vector information in the encoded stream. 100 video clips (20 of them coming from "Video Quality Experts Group" and 80 from DVDs) were used in the experiments. All the videos were encoded using different video editing software solutions. Through their experiments, a 74.63% accuracy in the identification of software used in encoding was obtained.

Reference [14] proposes an identification method using stills from videos. The characteristics of conditional probability are used and taken directly from the video frames. Tests used 4 different models of cameras and an SVM classifier, obtaining an 82.6% accuracy in the first experiment. In a second experiment using the same set of videos, taking the luminance value the average accuracy was 100%. In a third

experiment where a set of videos with major changes in the scenes was used, the accuracy was 97.2%.

III. TECHNIQUE DESCRIPTION

The proposed system has four main stages: The first divides input video into individual frames. Frame rate is generally about 15 to 30 frames per second. Next, a set of key frames are extracted.

To start the extraction process, the first frame is labeled a key frame. Then, the frame difference between the current frame and the last extracted key frame is computed. Color histogram correlation is used to choose frames with a significant scene change. If the frame difference satisfies a certain threshold condition, then the current frame is selected as a key frame. This process is repeated for all frames in the video until the whole set of key frames is extracted. This stage is crucial for the rest of the process.

The following stage extracts the sensor noise pattern from each key frame. The features are obtained by using a wavelet transform.

The final step is to use a Support Vector Machine (SVM) classifier.

It uses histogram correlation similarity to compare two frames, and proposes an improved key frame selection method to obtain more representative key frames.

The algorithm calculates and compares the frames contained in a video, the ones showing a significant change of scene will be used for classification and identification. This is because in [9] it was shown that the extracted noise in an image sensor may be severely contaminated by the details of the scene, in addition to video data containing temporal, spatial and spectral redundancy.

To compare two frames, it is necessary to extract the histogram from each of them (color value frequency), and by correlating them, it can be found how much similarity exists. The correlation is calculated by (1):

$$\text{correlation}(H_1, H_2) = \frac{\sum_i H'_1(i) H'_2(i)}{\sqrt{\sum_i H'^2_1(i) H'^2_2(i)}} \quad (1)$$

Where $H'_k(i) = H_k(i) - \frac{1}{N}(\sum_j H_k(j))$ and N is equal to the number of gray levels for each RGB color channel.

There are several methods to calculate the difference between two-dimensional color histograms, but calculating their correlation is suggested, since it is a random vector (multi-dimensional random variable).

The first frame of the video is always taken as part of the set of selected key frames. The comparison is performed by taking the first frame and the second; if there is no significant difference between them based on the threshold, the next frame is taken and a new comparison with the first one is performed. This is done until the result of the correlation is less than the threshold, to take into account the frame for classification and identification.

If in the end the amount of scene changes based on the threshold is less than the amount needed, the comparison process is repeated by increasing the threshold until the number of scene changes is greater than or equal to the desired number. When the threshold is estimated, this is what is finally used for the extraction of video key frames.

To determine a possible initial threshold, several experiments were performed on the videos, and it was found that by comparing the histograms of a video, the lowest average correlation was -0.27, showing at least 1 or 2 scene changes, thus defining the initial threshold. For the increment value, experiments were made with different values, such as 0.1, 0.01, 0.001, 0.0001 and 0.0001. The 0.001 value was chosen as it proved to be an ideal value to reach the desired number of frames in less time and with more accuracy. These increments are made because, if the threshold is closer to the maximum value of direct correlation, i.e. to the value of 1, more scene changes can be found, thus extracting the number of frames defined by the user for classification and identification.

By analyzing the state of art works, it was found that the sensor pattern noise and wavelet transform help to define a fingerprint, these being effective methods for source identification. This article extends the use of sensor pattern noise and wavelet transform of [12].

Obtaining the sensor pattern noise of the images is based on the method described in [8] with the modifications of [12]. The next step is to get the features that characterize the sensor noise for the classification purposes. A total of 81 features (3 channels x 3 wavelet components x 9 central times) are obtained using the feature extraction algorithm described in [12].

IV. EXPERIMENTS AND RESULTS

To test the effectiveness of the frame extraction algorithm and the use of the fingerprints for source classification and identification, videos were captured without any consideration about the temporal or spatial characteristics, since they must represent real cases. As mobile phones currently show large improvements in video quality, it was considered to use 1080p quality videos (HD videos), i.e. with a resolution of 1920x1080 pixels. Table I shows the basic specifications and models of mobile phones considered for the experiments.

The classification was performed using an SVM with RBF kernel. The LibSVM package in which the SVM allows multiple class classification was used. It is also the most commonly used option by the most recent works of the state of the art and they show good results. The classifier was trained and tested with the feature vectors extracted from the frames.

TABLE I
SETTINGS USED IN MOBILE PHONE DIGITAL CAMERAS

Make - Model	FPS	Format	Codec
Apple - iPhone 5 (M1)	24	mov	H.264
Nokia - 808 Pureview (M2)	30	mp4	MPEG-4
Samsung - Galaxy S4 (M3)	30	mp4	MPEG-4
Wiko - Cink Slim (M4)	12	3gp	MPEG-4
Zopo - ZP-980 (M5)	15	3gp	MPEG-4

5 experiments were performed in which the 5 mobile devices in Table I were used.

Table II shows the average success percentages for each device for different crop sizes of frame, success percentage meaning the percentage of frames in a video whose source was correctly identified by the classifier. That is, for instance for a particular video from which 100 frames were extracted, the percentage of them that the classifier classified as belonging to the video in question is calculated. For each device, as discussed above, 5 videos were used for tests. Each video obtained a success percentage and the Table II shows the average success rate of 5 videos for each device and crop size.

TABLE II
AVERAGE SUCCESS RATE BY DEVICE ACCORDING TO CROP SIZE

Resolution	Device					% of success
	M1	M2	M3	M4	M5	
1024x768	80.80%	97.40%	88.80%	85.40%	75.40%	85.56%
800x600	81.80%	96.80%	84.80%	86.00%	68.80%	83.64%
640x480	79.60%	95.60%	85.00%	85.60%	66.20%	82.40%
320x240	73.80%	88.20%	78.00%	78.80%	63.40%	76.44%
128x128	65.80%	79.20%	66.60%	75.00%	64.00%	70.12%

In most cases, the success percentages per device increase with larger crop size of frames (this occurs in all cases taking into account the average success rate). The highest resolution (1024x768) obtains the highest average success rate, 85.56%.

As can be observed, it exceeds the individual video rate of 50% in all cases. This indicates that in all cases, for all frames of a given video from a particular device, at least 50% of the frames are identified correctly. Finally, the source identification of a video should answer the specific question of to which acquisition source that video belongs. As a logical criterion, it can be estimated that the video belongs to the source with the highest number of frames classified with respect to the other sources (higher success rate compared to other sources). It could be possible that several sources had exactly the same number of frames and at the same time they were the highest number with respect to the other sources. In this unusual case, it could be said that the video source cannot be identified with determination and the doubt would be between those several sources.

The experiments obtain conclusive results which leave no doubt as to the identification of the video acquisition source considering the criteria defined above, since in all cases the success exceeds 50%. It can also be noted that success rates in many cases are much higher (in some cases reaching up to 100%). Therefore, according to this experiment, using the previously defined criterion and taking the video as a unitary entity (i.e. a video is either properly classified or not), it can be concluded that this technique identifies the video source with a 100% success rate.

Using the results of experiments, it is observed that there is a 2.92% improvement in the average success rate for a crop size of 1920x1080 with respect to a crop size of 1024x1024. In Table II can be seen that using the entire image for every case there is a higher average success rate in source identification, although the increase is small. In general, the larger the size

of the crop, the higher its success rate is. Also, with the results of the experiments is concluded that from a certain crop size the increase in the success rate is small, and in some cases there can be small decrements. We must also bear in mind that the larger the crop, the longer the run time of the feature extraction algorithm.

V. CONCLUSIONS

The general conclusion is that this technique presented is valid and obtains good results. The presented frame extraction algorithm takes into account the nature of a video and its frames, optimizing the extraction of key frames. i.e., it extracts frames taking into account that if obtained frames have greater scene variation between them (looking for scene changes), the future classification process will be better. However, for classification using SVM a number of frames is needed for training and testing, which is also taken into account by this algorithm, because there may be a case in which the video has little change of scenes and the algorithm has to obtain the most distant frames between existing scenes. Once the frames have been obtained we rely on the extraction of features obtained from the sensor pattern noise and wavelet transform as specified in [12].

Once the selected key frames have been classified, the question of what the video acquisition source is as a unitary entity must be answered. Our view has been that the video belongs to the source with the highest number of frames classified into this type. Taking this approach, the application of the proposed techniques has been successful, since the success rate has been 100%.

ACKNOWLEDGEMENTS

Part of the computations of this work were performed in EOLO, the HPC of Climate Change of the International Campus of Excellence of Moncloa, funded by MECD and MICINN.

REFERENCES

- [1] Alexa Internet, Inc., "Alexa Top 500 Global Sites," goo.gl/XyQWae, 2016.
- [2] I. Gartner, "Gartner Says Annual Smartphone Sales Surpassed Sales of Feature Phones for the First Time in 2013," <https://goo.gl/4GbxIB>, 2014.
- [3] I. IC Insights, "Embedded Imaging Takes Off as Stand-alone Digital Cameras Stall," goo.gl/lpuWg9, 2014.
- [4] C. Wen and K. Yang, "Image authentication for digital image evidence," *Forensic Science Journal*, vol. 5, no. 1, pp. 1–11, September 2006.
- [5] P. Brown, "Searches of Cell Phones Incident to Arrest: Overview of the Law as it Stands and a New Path Forward," *Harvard Journal of Law & Technology*, vol. 27, pp. 563–587, 2014.
- [6] P. Bestagini, M. Fontani, S. Milani, M. Barni, A. Piva, M. Tagliasacchi, and S. Tubaro, "An overview on video forensics," in *Proceedings of the 20th European Signal Processing Conference*, August 2012, pp. 1229–1233.
- [7] A. L. Sandoval Orozco, D. M. Arenas González, J. Rosales Corripio, L. J. García Villalba, and J. C. Hernandez-Castro, "Techniques for Source Camera Identification," in *Proceedings of the 6th International Conference on Information Technology*, May 2013, pp. 1–9.
- [8] J. Lukas, J. Fridrich, and M. Goljan, "Digital Camera Identification from Sensor Pattern Noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, June 2006.
- [9] C. Li, "Source Camera Identification Using Enhanced Sensor Pattern Noise," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 280–287, June 2010.
- [10] A. Wahab, A. Ho, and S. Li, "Inter-Camera Model Image Source Identification with Conditional Probability Features," in *Proceedings of IEEEJ 3rd Image Electronics and Visual Computing Workshop*, November 2012, pp. 1–4.
- [11] A. Wahab, J. Briffa, H. Schaathun, and A. T. S. Ho, "Conditional Probability Based Steganalysis for JPEG Steganography," in *Proceedings of the International Conference on Signal Processing Systems*, May 2009, pp. 205–209.
- [12] D. M. Arenas González, A. L. Sandoval Orozco, J. Rosales Corripio, L. J. García Villalba, J. Hernandez-Castro, and S. J. Gibson, "Proceedings of the Source Smartphone Identification Using Sensor Pattern Noise and Wavelet Transform," in *Proceedings of the 5th International Conference on Imaging for Crime Detection and Prevention*, 16–17 December 2013, pp. 1–6.
- [13] Y. Su, J. Xu, and B. Dong, "A Source Video Identification Algorithm Based on Motion Vectors," in *Proceedings of the Second International Workshop on Computer Science and Engineering*, vol. 2, October 2009, pp. 312–316.
- [14] S. Yahaya, A. Ho, and A. Wahab, "Advanced video camera identification using Conditional Probability Features," in *Proceedings of the IET Conference on Image Processing*, July 2012, pp. 1–5.



Raquel Ramos López She received a Computer Science Engineering degree from the University Complutense of Madrid (Spain) in 2006. She holds a M.Sc. in Research in Computer Science from the University Complutense of Madrid in 2016. She is currently a Ph.D. student at the University Complutense of Madrid and a member of the research group GASS (<http://gass.ucm.es>), in the Department of Software Engineering and Artificial Intelligence (DISIA) of the Faculty of Computer Science and Engineering. Her main research interests is Computer Security.



Anissa El-Khattabi received a Computer Science Engineering degree from the ETSIT Universidad de Granada (Spain) in 2009. She has worked as R&D Engineer in different R&D companies. She is currently a Ph.D. student at the Universidad Complutense de Madrid (Spain) and a Research Assistant at Complutense Research Group GASS (<http://gass.ucm.es>). Her main research interests are image processing and multimedia forensics.



Ana Lucila Sandoval Orozco She received a Computer Science Engineering degree from the Universidad Autónoma del Caribe (Colombia) in 2001. She holds a Specialization Course in Computer Networks (2006) from the Universidad del Norte (Colombia), and holds a M.Sc. in Research in Computer Science (2009) and a Ph.D. in Computer Science (2014), both from the Universidad Complutense de Madrid (Spain). She is currently a Research Assistant at Complutense Research Group GASS. Her main research interests are Computer Networks and Computer Security.



Luis Javier García Villalba He received a Telecommunication Engineering degree from the Universidad de Málaga (Spain) in 1993 and holds a M.Sc. in Computer Networks (1996) and a Ph.D. in Computer Science (1999), both from the Universidad Politécnica de Madrid (Spain). Visiting Scholar at COSIC (Computer Security and Industrial Cryptography, Department of Electrical Engineering, Faculty of Engineering, Katholieke Universiteit Leuven, Belgium) in 2000 and Visiting Scientist at IBM Research Division (IBM Almaden Research Center, San Jose, CA, USA) in 2001 and 2002, he is currently Associate Professor of the Department of Software Engineering and Artificial Intelligence at the Universidad Complutense de Madrid (UCM) and Head of Complutense Research Group GASS (Group of Analysis, Security and Systems) which is located in the Faculty of Information Technology and Computer Science at the UCM Campus. His professional experience includes research projects with Hitachi, IBM, Nokia and Safelayer Secure Communications. His main research interests are Computer Networks and Computer Security.