TECHNIQUES FOR A FORENSIC ANALYSIS OF THE CASIA-IRIS V4 DATABASE

Luca Debiasi, Andreas Uhl

Multimedia Signal Processing and Security Lab
University of Salzburg
Salzburg, Austria

ABSTRACT

The photo response non-uniformity (PRNU) of a sensor can be useful to enhance a biometric systems security by ensuring the authenticity and integrity of images acquired with a biometric sensor, e.g. by performing a source device identification. Previous studies regarding the feasibility of this application have been conducted on the CASIA-Iris V4 database by studying the differentiability of the sensors PRNU fingerprints. The results showed a high variation among the different subsets of the database. It was assumed that this high variation could either be caused by correlated data or that different sensors may have been used for the acquisition of the subsets.

To investigate the latter case we perform a forensic investigation on the CASIA-Iris V4 database, since there is no specific documentation on the number of sensors used for the acquisition. We apply an existing forensic technique and we propose several novel forensic techniques to establish a ground truth of how many sensors have been used to a acquire a digital image data set in a blind manner and without any a priori knowledge.

1. INTRODUCTION

The photo response non-uniformity (PRNU) of a imaging sensor emerged as an important component to perform various forensic tasks such as device identification, device linking, recovery of processing history and the detection of digital forgeries.

The PRNU is an intrinsic property of all digital imaging sensors that emerges from to slight variations among individual pixels in their ability to convert photons to electrons. Consequently, every sensor casts a weak noise-like pattern onto every image it takes. This pattern, which plays the role of a sensor fingerprint, is essentially an unintentional stochastic spread-spectrum watermark that survives processing, such as lossy compression or filtering.

The PRNU fingerprint of a sensor can be used to improve a biometric systems security, for instance by using the PRNU fingerprint of a known sensor to check if the data presented to the system has been acquired with this specific sensor or by detecting manipulations which would also tamper the fingerprint. Previous feasibility studies on this application by Höller *et al.* [1] performed on the CASIA-Iris V4 database. The differentiability of the sensors in the CASIA Iris V4 database using sensor fingerprints has been tested with the conclusion, that the EERs and respective thresholds vary highly. Some sensors showed satisfying results while others showed EERs of over 20%. The question raised, that if the PRNU fingerprint is going to be applied as an authentication measure for iris databases, it is not clear if the poor differentiation results for some sensors come from the images special content with low variance between the images, or from the sensor properties.

It was assumed that this high variation could be caused by correlated data used to generate the sensors PRNU fingerprint. Further investigation from Debiasi *et al.* [2] showed that using uncorrelated data to generate the PRNU fingerprint does not improve the results for this data set and hence does not cause the high variation.

On the other hand Höller *et al.* [1] suspected that multiple sensors may have been used for the acquisition of the CASIA Iris-V4 subsets. Mixing images from more than one sensor during the generation of a PRNU fingerprint tampers the fingerprint estimate and could explain the high variances. Unfortunately neither the meta data of the images in the CASIA-Iris V4 database, nor the database description, denoting solely the sensor model without any additional information, can reveal number of sensors used. Even the researchers involved in the acquisition cannot determine the number of sensors any more.

Related work regarding blind classification of image source in an open set scenario makes use of Hierarchical Agglomerative Clustering (HAC) [3, 4] or Multi-Class Spectral Clustering (MCSC) [5, 6] by formulating the classification task as a graph partitioning problem. These approaches rely on a known training or test set or determine some criteria, e.g. the stop criterion for the clustering, on a ground truth. Other related work [7] relies on an iterative algorithm that consecutively "constructs" a sensor fingerprint from images with similar PRNU using a pre-calculated threshold function.

In this paper we perform a forensic analysis of the CASIA-Iris V4 database to investigate if multiple sensors

were used during the acquisition in a completely blind manner with no a priori knowledge of the data set. A brief description of the data sets is given in section 2, while the feature extraction and experiment set-up are described in section 3. In section 4 an already existing technique is applied to the test data set and evaluated. Several new forensic investigation techniques are proposed in section 5 and their performance is evaluated in section 6 in conjunction with the forensic investigation results of the CASIA-Iris V4 database. Finally, section 7 concludes the paper.

2. DESCRIPTION OF THE DATA SETS

The CASIA-IrisV4 data set has been used for the the actual blind forensic investigation with the assumption that all images in each sub set have been acquired with a single specific sensor unit. Additionally we generated a test data set with images from known sensors to evaluate the implementations and algorithms.

A brief description of the subsets of each data set is given in the following sections.

2.1. Test data set

The sensors used to generate this data set are a OKI IRISPASS-h and a Irisguard H100 IRT iris sensors, hence it is known which sensor instance has been used to acquire the images. The 1000 images in this data set have reliably been acquired with the two mentioned sensors, 500 with each one. All images are 8 bit grey-level JPEG files With these images we generated two subsets, each containing 1000 images.

In the *test-sequential* data set the first 500 images come from the *test-h100* and the latter 500 from the *test-irispassh* sensor.

The *test-mixed* data set contains alternated images from the *test-h100* and the *test-irispassh* sensor in blocks of 100 images.

2.2. CASIA Iris V4

The CASIA-IrisV4 contains a total of 54,601 iris images from more than 1,800 genuine subjects. All iris images are 8 bit grey-level JPEG files, collected under near infrared illumination. For this work we used images from five different subsets.

The five subsets with the corresponding sensors (as described in the database specification) are: CASIA-Iris-Interval (*intv*), CASIA-Iris-Lamp (*lamp*), CASIA-Iris-Twins (*twin*), CASIA-Iris-Distance (*dist*) and CASIA-Iris-Thousand (*thou*). The respective sensors are: CASIA close-up iris camera, OKI IRISPASS-h (1), OKI IRISPASS-h (2), CASIA long-range iris camera and Irisking IKEMB-100.

For the CASIA Iris V4 data sets it is not clear, whether the single data sets have been acquired with a specific sensor or if multiple instances of the same sensor model have been used. This question is substantiated by the fact that the same sensor model was used for two different data sets (*lamp* and *twin*).

3. FEATURE EXTRACTION AND EXPERIMENT SET-UP

All the the forensic investigation techniques in this work are based on the sensors PRNU. The extraction of the PRNU and the experiment set-up are described in the following section.

For an estimation of each sensors PRNU fingerprint, the algorithm described by Fridrich [8] was used to calculate the PRNU. The PRNU is represents the noise intrinsically inserted into an image during the acquisition process. For each image I the noise residual W_I is estimated

$$W_I = I - F(I) \tag{1}$$

where F is a denoising function filtering out the sensor pattern noise. We used the wavelet-based denoising filter as described in Appendix A of [9] , because it is producing good results in filtering out the PRNU.

In all of the proposed techniques the PRNU noise residual of an image is extracted from 4 patches located in the corners with a size of 128x128 pixels each, resulting in a total noise residual size of 256x256 pixels. This is done because the image size is varying between the data sets. The PRNU noise residual has been normalized in respect to the L_2-norm because its embedding strength is varying between different sensors as denoted by [1].

The PRNU fingerprint $\hat{K}$ of a sensor is then estimated using a maximum likelihood estimator for images I_i with $i = 1...N$.

$$\hat{K} = \frac{\sum_{i=1}^{N} W_I^i I^i}{\sum_{i=1}^{N} (I^i)^2} \tag{2}$$

The normalized cross correlation (NCC) is used to detect the presence of a PRNU fingerprint $\hat{K}$ in an Image J with

$$\rho_{[J,\hat{K}]} = NCC(W_J, J\hat{K}) \tag{3}$$

where ρ indicates the correlation between the PRNU residual W_j of the image J and the fingerprint $\hat{K}$ weighted by the image content of J.

The correlation ρ is calculated between each image from a sensor S_i and the PRNU fingerprint $\hat{K}_i$ of the sensor S_i, where only images are used that have not been part of the PRNU fingerprint estimation. Additionally the correlation ρ between all images from the other sensors S_j, $i \neq j$, and the PRNU fingerprint $\hat{K}_i$ of the sensor S_i is also calculated.

4. APPLICATION OF EXISTING TECHNIQUES

Bloy [7] proposed a Blind Fingerprinting and Image Clustering (BFAIC) technique, which performs an agglomerative

clustering to construct PRNU fingerprints from a mixed set of images, enabling identification of each images source camera without any prior knowledge of source. This technique does not rely on a known training set, test set or ground truth. It solely depends on a pre-calculated threshold function. Because both scenarios from his and our work have a strong similarity we reimplemented his technique for our investigation.

We used different parameters for the threshold function in our experiment: the parameters proposed from Bloy (T_{Bloy}) and recalculated parameters from test data set in section 2.1 according to Bloy (T_{STSM}).

First of all we applied this technique to the test data set described in section 2.1 to evaluate its performance. Looking at the results in table 1 we realise that the algorithm produces clearly more than two clusters, regardless of which threshold function is used. Of the resulting clusters in most cases only two clusters remain with more than 100 associated images and almost all remaining clusters contain less than 10 images. There have been no unassociated images during the experiment and all clusters consisted only of images from the same sensor.

We can assert that this algorithm is not reasonably working well in this scenario since it produces highly fragmented clusters, even for a known number of sensors.

	T_{Bloy}		T_{STSM}	
BFAIC	TS	TM	TS	TM
images	1000	1000	1000	1000
partitions	10	12	41	34
partitions > 100	4	2	2	2
partitions < 10	4	7	35	27
unassociated images	0	0	0	0

Table 1: BFAIC experiment on *test-sequential* (TS) and *test-mixed* (TM) data sets using the threshold function from Bloy (T_{Bloy}) and the calculated function (T_{STSM}).

5. NOVEL FORENSIC INVESTIGATION TECHNIQUES

The results obtained from the existing technique in section 4 are not particularly satisfactory. Therefore, in the following section, we propose several novel techniques that all aim at a blind classification of the number of sensors used to acquire a data set with no a priori knowledge of the number of sensors in a forensic investigation context.

5.1. K-Means clustering

For this technique Lloyd's K-Means algorithm [10] has been adopted. In our experiment we define the PRNU noise residuals of the images in the investigated data set as the n objects. The clusters k represent the different sensors. Since we do not know how many sensors have been used to acquire the dataset we repeated the clustering for $k = 2 \dots 5$ with the assumption that not more than 5 sensors have been used. This limitation is not mandatory and can be extended if necessary, but increases the computational effort significantly.

To determine the distance between the objects the NCC was chosen and the K-Means clustering was repeated five times for each k to avoid local minima. In order to qualitatively evaluate the outcome of the clustering the Mean Silhouette Value (MSV) by Rousseeuw [11] was chosen. The silhouette value for each point is a measure of how similar that point is to points in its own cluster, when compared to points in other clusters.

The result for $k = 1$ has been determined by calculating the pairwise NCC between all point combinations i and j, where $i \neq j$, and then calculating the mean correlation over all points. For all $k >= 2$ the Mean Silhouette Value for the i-th point, S_i, is defined as

$$MSV = \frac{1}{N} \sum_{n=1}^{N} \frac{b_i - a_i}{max(a_i, b_i)} \tag{4}$$

where N is the number of noise residuals, a_i is the average distance from the i-th point to the other points in the same cluster as i, and b_i is the minimum average distance from the i-th point to points in a different cluster, minimized over clusters. The silhouette value ranges from -1 to +1. A high silhouette value indicates that i is well-matched to its own cluster, and poorly-matched to neighbouring clusters. If most points have a high silhouette value, then the clustering solution is appropriate. If many points have a low or negative silhouette value, then the clustering solution may have either too many or too few clusters.

5.2. PCA K-Means Clustering

This forensic technique uses the same K-Means clustering method as in 5.1 with the exception that a principal component analysis (PCA) [12] is performed on the PRNU noise residual of each image. The PRNU was extracted from a 256x256 pixel patch in the centre of the image. From the PCA outcome we selected the first n principal components, where $n \in \{5, 10, 20, 50\}$. These represent the feature vector used to cluster the images and the squared Euclidian distance was chosen as distance metric. Briefly explained in this technique we cluster the PCA components of each images PRNU noise residual.

5.3. Sliding window fingerprinting

We propose an iterative algorithm where a window with a defined size moves over the data image after image and a PRNU fingerprint from the data within this window is calculated. After moving the window over the whole data set we evaluate the similarity of a PRNU fingerprint FP_i from the iteration i with all other fingerprints FP_j where $i \neq j$ by computing the

pairwise NCC. We applied this method with different window sizes.

5.4. Device identification on dataset partitions

For this forensic technique we perform a device identification experiment similar to Höller *et al.* [1]. We divide the data sets into n partitions with a the same size and treat the disjoint partitions as n different sensors.

For the device identification we need to calculate a PRNU fingerprint as described in section 3 for every partition. The images inside each partition are randomly shuffled, so that the PRNU fingerprint is not just computed from the first images in the partition. We use half of the images (up to a maximum of 50) to calculate the PRNU fingerprint and the remaining images in the partition to calculate the NCC scores (inter and intra partition scores). From these score we calculate the pairwise EER for two partitions P_i and Pj where $i \neq j$ as illustrated in figure 1. If the resulting EER score is low (e.g. 0%), the extracted PRNU and respectively the PRNU fingerprint is different for both partitions, hence they must have been acquired with different sensors. On the other hand, if the resulting EER score is high (e.g. 50%), the extracted PRNU very similar for both partitions and their images have all likely been acquired with the same sensor. We repeated this method with various partition sizes.

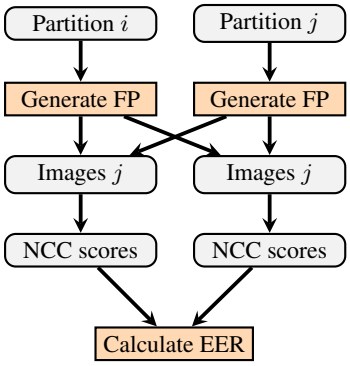

Fig. 1: Illustration of the calculation of the EER score of two disjoint partitions i and j.

6. FORENSIC INVESTIGATION RESULTS

The results of the previously described forensic techniques are presented in the following sections, first for the test data sets and then for the CASIA-Iris V4 subsets.

6.1. Test data sets

To begin with we have a look at the results from the K-Means and the PCA K-Means clustering in table 2. The best results were obtained using $n = 5$ principal components. As it can clearly be observed the two methods perform as expected and

	K-Means		PCA K-Means	
k	TS	TM	TS	TM
1	0.0230	0.0235	0.0112	0.0111
2	**0.0384**	**0.0388**	**0.0381**	**0.0370**
3	0.0161	0.0257	0.0271	0.0269
4	0.0030	0.0037	0.0150	0.0245
5	0.0042	0.0035	0.0140	0.0134

Table 2: Mean silhouette value (MSV) for K-Means and PCA K-Means clustering performed on the *test-sequential* (TS) and *test-mixed* (TM) data sets using $n = 5$ principal components.

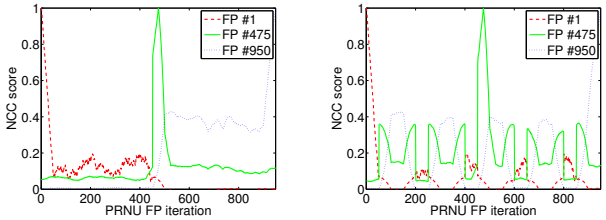

Fig. 2: Results of the SW experiment with a window size of 50 on the *test-sequential* data set (left) and the *test-mixed* (right). The different graphs represent the NCC scores of fingerprints from specific iterations with all other fingerprints.

show the correct number of clusters for the two sensors used to acquire the images in both test data sets.

Figure 2 represents the results from the Sliding Window (SW) experiment with a window size of 50. The left plot shows the results for the *test-sequential* data set, where the transition of the images from one sensor to the other can be seen at the fingerprint iterations 450 until 500. The right plot shows the results for the *test-mixed* data set, where the transitions between the images of different sensors are also observable. The high spikes with a peak value of 1 occur when fingerprints that have one or more common images in their generation are compared.

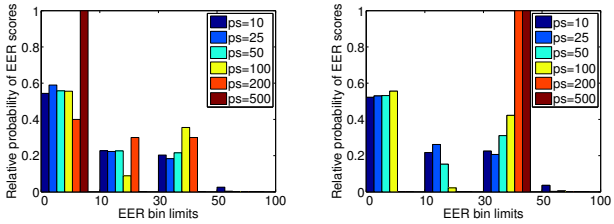

Fig. 3: Results of the DIODP experiment with different partition sizes (ps) on the test data sets *test-sequential* (left) and *test-mixed* (right) with the occurrence of EER scores in the different bins.

The Device Identification on Dataset Partitions (DIODP) experiments has been conducted for different partition sizes as represented in figure 3. To be able to clearly represent the resulting EER scores we performed a binning of the scores into four bins with the following limits: scores below 10%, scores between 10% and 29%, scores between 30% and 49% and scores above 50%.

T_{Bloy}	intv	lamp	twin	dist	thou
images	1307	6855	1095	1566	2000
partitions	36	64	31	1	3
partitions > 100	3	13	3	1	1
partitions < 10	17	21	16	0	1
unassociated images	0	0	0	0	0
T_{STSM}	intv	lamp	twin	dist	thou
images	1307	6855	1095	1566	2000
partitions	169	321	152	**1**	152
partitions > 100	2	12	1	1	1
partitions < 10	147	24	131	0	129
unassociated images	0	0	0	0	0

Table 3: BFAIC experiment results on the CASIA-Iris V4 data sets with Bloys threshold function (T_{Bloy}, top) and the calculated threshold function (T_{STSM}, bottom).

k	intv	lamp	twin	dist	thou
1	0.0024	**0.0391**	**0.0398**	**0.3821**	**0.0121**
2	0.0034	0.0100	0.0052	0.0632	0.0074
3	0.0031	0.0004	0.0063	-0.0076	0.0064
4	0.0032	0.0003	0.0057	-0.0069	0.0045
5	0.0030	0.0001	0.0053	-0.0071	0.0042
k	intv	lamp	twin	dist	thou
1	0.0003	0.0010	0.0024	0.0019	0.0006
2	0.0098	0.0185	0.0197	0.0234	0.0474
3	0.0090	0.0123	0.0141	0.0186	0.0427
4	0.0088	0.0097	0.0119	0.0180	0.0391
5	0.0089	0.0086	0.0105	0.0177	0.0405

Table 4: Mean silhouette value (MSV) for K-Means (top) PCA K-Means clustering (bottom) using $n = 5$ principal components.

In the left plot all EER scores for a partition size of 500 are below 10%, which is the expected result for the *test-sequential* data set. All other tested partition sizes meet the sensor boundaries in the data sets as well and show a relative EER score occurrence of approximately 0.5 below 10% because the images are distributed half half among the sensors, therefore the remaining EER scores are greater than 10% because PRNU noise residuals of the same sensor are compared with each other under the assumption that they belong to different sensors as described in section 5.4. Hence the resulting score distribution meets the expectation.

6.2. CASIA-Iris V4

We first applied the existing Blind Fingerprinting and Image Clustering (BFAIC) technique to the CASIA-Iris V4 database to see how it performs in a real world scenario. The results in table 3 show a high clusters fragmentation for all subsets, except for the *dist* data set, where all images have been clustered together. The threshold function calculated from the test data set produces significantly more clusters, but they contain only a small amount of images. The results however are hard to interpret because this technique also generated a high amount of clusters for the known test data sets as noted in section 6.1.

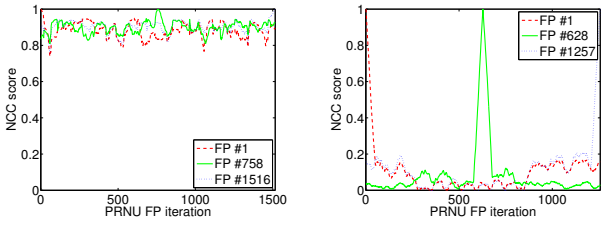

Fig. 4: Results of SW experiment with window size 50 for the *dist* (left) and *intv* (right) data sets.

The results of the K-Means and PCA K-Means experiments are presented in table 4. The K-Means results show that one sensor was used to acquire all subsets except the *intv* data set, for which no clear result can be established. The PCA K-Means results actually do not permit any statements to be made because of the insignificant differences of the obtained values.

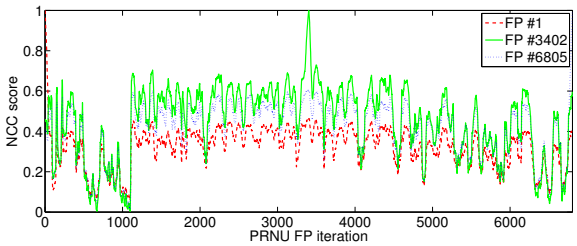

Fig. 5: Results of SW experiment with window size 50 for the *lamp* data set.

The figures 4, 5 and 6 show the results from the Sliding Window (SW) experiment with a window size of 50. No transitions like in the test data can be observed for the *dist*, *twin* and *thou* data sets. In the*lamp* and *intv* data sets such transitions can be observed at approximately iteration 700 and 1050 (*lamp*) and iteration 250 and 800 (*intv*), although not as clear as for the test data sets. The high spikes, again, occur when fingerprints having one or more common images in their generation are compared. This technique suggests that all data sets ,with the exception of the *lamp* and *intv*, have been acquired with a single sensor.

From the Device Identification on Dataset Partitions (DIODP) experiments in figure 7 it can be observed that for almost all data sets other than *intv* the EER scores are bigger than 30%, which indicates that these data sets might be acquired with one sensor. Having a closer look at the *intv* data set with different partition sizes indicates that this set might be acquired with more than one sensor, because the distribution of the EER scores is similar to the one from the two sensors in the test data set, only with higher EER scores.

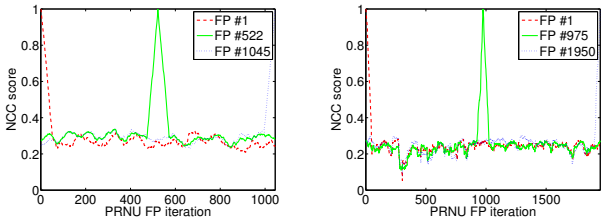

Fig. 6: Results of SW experiment with window size 50 for the *twin* (left) and *thou* (right) data sets.

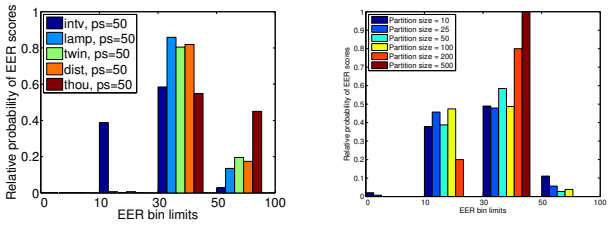

Fig. 7: DIODP experiment results with partition size 50 (ps) on the CASIA-Iris V4 data sets (left). Different partition sizes for the *intv* data set (right). The bars show the relative occurrence of EER scores in the different bins.

7. CONCLUSION

All newly proposed forensic techniques have consistently been able to detect the presence of multiple sensors in the investigated test data sets, for which the number of sensor was known, and outperformed the reimplemented related technique (BFAIC). Some of the novel techniques were also able to exactly point out the number of different sensors.

The investigation results on CASIA-Irsi V4 are not as clear as the test set results, hence only an assumption based on the trends in the results can be given because this is a completely blind approach without any a priori knowledge of the sensors. The results indicate that the *intv* data set might be acquired with more than one sensor, while the other subsets have been acquired with one sensor.

The *lamp* and *twin* data sets, suspicious of containing images from multiple sensors because of the same model denoted in the specification, seem both to be acquired with just one sensor. These data sets were also responsible for the high EER scores in the experiment of Höller *et al.* [1].

Also unknown factors could affect the quality of the PRNU noise residuals and hence tamper the results. Clearly further studies on factors that interfere with the PRNU have to be conducted to be able to use the PRNU fingerprint of a sensor to improve a biometric systems security by ensuring the authenticity and integrity of images presented to the system.

8. ACKNOWLEDGMENTS

This work has been partially supported by a COST 1106 Short Term Scientific Mission (STSM).

9. REFERENCES

[1] Andreas Uhl and Yvonne Höller, "Iris-sensor authentication using camera PRNU fingerprints," in *Proceedings of the 5th IAPR/IEEE International Conference on Biometrics (ICB'12)*, New Delhi, India, Mar. 2012, pp. 1–8.

[2] L. Debiasi, Z. Sun, and A. Uhl, "Generation of iris sensor PRNU fingerprints from uncorrelated data," in *Proceedings of the 2nd International Workshop on Biometrics and Forensics (IWBF'14)*, 2014.

[3] Chang-Tsun Li, "Unsupervised classification of digital images using enhanced sensor pattern noise.," in *ISCAS*. 2010, pp. 3429–3432, IEEE.

[4] R. Caldelli, I. Amerini, F. Picchioni, and M. Innocenti, "Fast image clustering of unknown source images," in *IEEE International Workshop on Information Forensics and Security (WIFS) 2010*, 2010, pp. 1–5.

[5] Shuhan Luan, Xiangwei Kong, Bo Wang, Yanqing Guo, and Xingang You, "Silhouette coefficient based approach on cell-phone classification for unknown source images.," in *ICC*. 2012, pp. 6744–6747, IEEE.

[6] I. Amerini, R. Caldelli, P. Crescenzi, A. Del Mastio, and A. Marino, "Blind image clustering based on the normalized cuts criterion for camera identification," *Signal Processing: Image Communication*, , no. 29, pp. 831–843, 2014.

[7] G. Bloy, "Blind camera fingerprinting and image clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 532–534, Mar. 2008.

[8] J. Fridrich, "Digital image forensic using sensor noise," *IEEE Signal Processing Magazine*, vol. 26, no. 2, March 2009.

[9] Jan Lukas, Jessica J. Fridrich, and Miroslav Goljan, "Digital camera identification from sensor pattern noise.," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, 2006.

[10] S.P. Lloyd, "Least square optimization in PCM," *IEEE Transactions on Information Theory*, vol. 2, no. IT-28, pp. 129–137, Mar. 1982.

[11] Peter Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, no. 1, pp. 53–65, Nov. 1987.

[12] George H. Dunteman, *Principal Components Analysis*, Number Nr. 69 in A Sage Publications. SAGE Publications, 1989.