International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 06 | June 2021 www.irjet.net p-ISSN: 2395-0072

Source Device Attribution for Video Sequences using Photo Response Non-Uniformity

Darsana B S[1], Subha V[2]

[1]*M.Tech Signal Processing, Department of ECE, Sree Chitra Thirunal College of Engineering, Thiruvananthapuram*
[2]*Assistant Professor, Department of ECE, Sree Chitra Thirunal College of Engineering, Thiruvananthapuram*

---***---

Abstract - Source device attribution has gained wide attention due to its importance in multimedia forensics. Attribution of a multimedia content to its source device provides evidence before court of law in solving legal and security issues. Many people share illegal contents through anonymous profiles. Source device attribution helps in linking these contents to its source device, which helps to trace the owner of the device, which tackles many cases such as copyright as well as uploading of other illicit materials (For eg: women and child exploitation clips) in Social Media Platforms. Out of the many researches, one important technique is the Photo Response Non Uniformity (PRNU) traces. In this paper, the challenging problem of attributing videos to their source device is been dealt with. The proposed methodology proposes the characteristic fingerprint estimation of device for both stabilized as well as non-stabilized video sequences followed by matching of query video fingerprint with reference fingerprint using multiclass SVM. It exploits only videos to study the device from which they both come

***Key Words*: Video forensics, Device attribution, Digital fingerprint, Photo Response Non-Uniformity (PRNU), Video frames, Classification**

1. INTRODUCTION

Now-a-days, millions of people use smartphones and shares the multimedia contents through different social media platforms. Frequent use of these smart phones and platforms has increased the number of gruesome crimes such as acts of terrorism, child exploitation, and copyright infringement cases. Since these cases are related to digital forensics[1], solid digital evidences are required to solve the cases legally. Since the suspect uses anonymous profiles to share these contents, it might not be able to find the culprit through his social media accounts. Digital forensics can simply be defined to be the discipline that mixes elements of law and computing to gather and analyze data from computer systems, networks, wireless communications and storage devices during a way that's admissible as evidence during a court of law.

In particular, digital forensics science emerged in the last decade in response to the escalation of crimes. It is committed by the use of electronic devices as an instrument for committing a crime or as a repository of evidences related to a crime .One of the emerging techniques is to use the noise patterns from the multimedia contents, which are unique for each sensor [3]. Photo Response Non-Uniformity, or PRNU for brief, is one source of pattern noise in digital cameras. PRNU pattern can be considered as the characteristic noise fingerprint of any image or video acquisition device. When a steady uniform amount of light falls on the sensor cells in a camera, each cell within the camera should output precisely the same voltage. However due to variety of factors including small variations in cell size and substrate material like dishomgeneties in silicon wafers and imperfections in the sensor manufacturing process, this is not true. When a consistent light is shined within the cells in a camera, the cells output slightly different voltages. Because of the non homogeneity naturally present in the silicon, all the pixels of a sensor will never have the same photo response characteristics. The difference in response to a uniform light source is referred to as PRNU. PRNU is almost impossible to eliminate and it is inevitable for all type of camera sensors.

By testing the presence of a specific fingerprint in the image, one can achieve reliable device identification [2](eg: prove that a certain camera took a given image) or prove that two images were taken by one device (device linking) [9]. The presence of camera fingerprint in an image is also indicative of the fact that image under investigation is natural and not a computer rendering. By establishing the absence of the fingerprint in individual image regions, it is impossible to discover maliciously replaced parts of the image. This task pertains to integrity verification.

By detecting the strength or form of the fingerprint it is impossible to reconstruct some of the processing history. For eg: one can use the fingerprint as a template to estimate geometrical processing such as scaling, cropping or rotation. Non-geometrical operations are also going to influence the strength of fingerprint in the image and thus can be potentially detected. To identify the camera model or distinguish between a scan and a digital camera image, the spectral and spatial characteristics of fingerprint is used.

Source attribution for videos is quite challenging than that of images because the noise contents in video frames tend to be subtle than images [12]. In the proposed work, we use reference fingerprint from videos only and not from images. It is based on the scenario that no information is available in the case of images captured by the device under analysis. In such cases, where only videos are available, the reference fingerprint will also be taken from the video itself. Then, this reference fingerprint will be compared with that of the video under query or analysis.

Hence, the selection of number of frames also matters. Either, we can take all the frames or only I-frames (intra-coded frames). I-frames contain more reliable PRNU information than P-frames or B-frames [12]. This is because P frames and B frames are usually compressed and may contain several artifacts. In the case of I frames, they are extracted in an uncompressed form as further compression can degrade the quality. For eg, when extracting I frames using FFMPEG (Fast Forward Moving Pictures Expert Group), we extract the frames in BMP(Bitmap), which is uncompressed rather than JPEG (Joint Photographic Experts Group), which is compressed. The accuracy of classification increases with the number of frames extracted for fingerprint extraction.

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 scheme presented consists of four stages: 1) Key frames extraction, 2) sensor pattern noise extraction, 3) feature extraction, and 4) classifier training and prediction.

2. RELATED WORKS

Many approaches were proposed for source identification by analyzing traces like sensor dust, defective pixels, and color filter array interpolation.

Jan Lukas et al. [3] first introduced this idea of using sensor noise for source identification. Jan Lukas et al. showed that, the Photo-Response Non-Uniformity (PRNU) noise in digital images can be used to perform digital image source attribution and forgery detection. This work was followed by many others which established the PRNU as one of the most promising and powerful imaging sensor characteristics which can be exploited for image source attribution.

PRNU based methods are used for many forensic tasks such as source device identification [2], forgery detection, detection of duplicate or spliced videos, authentication of smart phones [10]etc.

Massimo et al. in [4] proposed a "hybrid" approach to video source attribution. In this work, the video source identification is done using still images taken by the same device as reference. Massimo et al. also discussed how to link a Facebook account to a YouTube account by correlating two PRNU fingerprint estimates obtained from a query video downloaded from YouTube and images shared on a specific Facebook account, but the accuracy of by their method was very low.

The method in [4] has good identification results on native videos but source attribution accuracy for YouTube videos are not as high as for native camera outputs. Moreover, this method cannot be used to perform video-to-video device linking for cameras featuring video stabilization. Furthermore, in the case of Facebook-shared images, estimating a fingerprint using images from an unknown source is not realistic since it is assumed that they all come from the same device, which may not always be the case.

Chen et al. in [11] investigated the video source device attribution problem and showed that PRNU could be used to identify the source camcorder of a subjected digital video by estimating the PRNU fingerprint or sensor pattern noise (SPN) from individual video frames given that enough frames are available.

3. PROPOSED METHOD

The proposed model is demonstrated in Fig. 1. In the proposed work, we use reference fingerprint from videos only and not from images as mentioned earlier. In order to estimate the PRNU of camera, we need to take the frames first. Let us assume that 'N' denotes the number of available images or frames.

First step is to separate each original image from its noise. For that, a denoised version of the original image is obtained. This is done for obtaining the noise residues. There are many denoising filters for this process, the one used in this technique is the PCA denoising method. After this, the original image is subtracted from the denoised version of the image to obtain the noise residues.

PRNU can be estimated using the following equation, which is the maximum likelihood estimator: The reference pattern of a camera is first extracted from a series of images or frames of a reference video taken from known camera device. The reference pattern is then used to detect whether the camera used to generate the reference pattern was used to capture an unknown source image. Generally, for an image *I*, the residual noise is extracted by subtracting the denoised version of the image from the image itself as follows:

$$N = I - F(I), \qquad (1)$$

where F(I) is the denoised image, and F is a denoising filter. In order to extract the fingerprint of a camera, multiple images are denoised and averaged. The averaging of multiple images or frames reduces the random components and enhances the pattern noise. The PRNU fingerprint of a device can be estimated as equation (3):

$$k = \frac{\sum_{n=1}^{N} W_n I_n}{\sum_{n=1}^{N} I_n^2} \qquad (2)$$

Where W_n is the noise residual extracted from I_n and all operations are performed pixel wise.

Precisely, $W_n = I_n - i_n$, i_n is denoised version of I_n.

After the extraction of frames, denoising of these frames is done using LPG-PCA (Local Pixel Grouping-Principal Component Analysis) [14] denoising technique to extract the residual values. Out of the many denoising techniques, LPG-PCA is found to have better PSNR value and hence will have better noise residuals. The illustration of LPG-PCA is given in Figure 2.

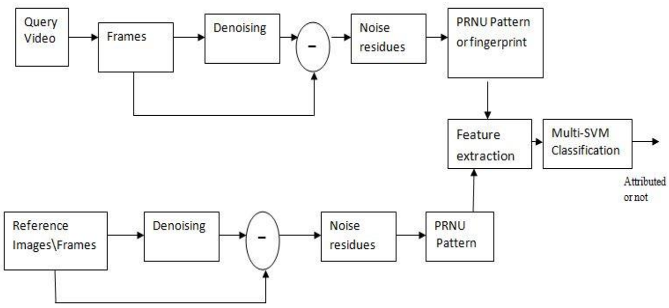

Fig -1: The framework of the proposed model

As shown in Fig 2, the algorithm has two stages. The first stage it gives an initial estimation of the image by removing most of the noise and the second stage will further refine the output of the first stage . The second stage has the same type of procedure except for the parameter of noise level is updated. Since the noise in the first stage is significantly reduced, the LPG accuracy will be much improved in the second stage so that the final denoising result is visually much better.

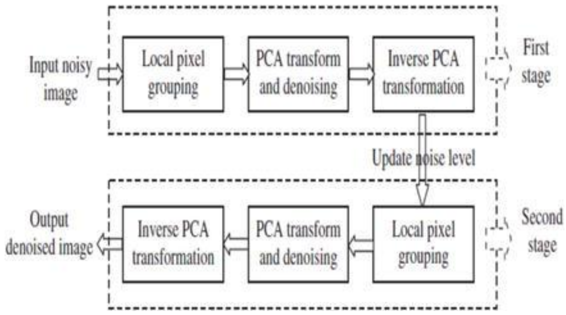

Fig -2: Illustration of LPG-PCA denoising technique

The comparison of values of different denoising techniques is given below:

Table -1: Comparison of Various DENOISING methods

Filters	image 1	image 2	image3	image4	image5	image6	Average
Median	23.7841	22.9481	20.6154	20.8301	28.5012	22.5348	23.2022
laplacian	25.7401	25.7506	24.8241	25.6951	28.5655	24.6743	25.8749
weiner	25.6134	26.0118	24.6756	25.6588	29.6443	23.9915	25.9325
gaussian	26.797	26.3957	25.2337	24.3735	29.3971	25.3792	26.2627
DWT	28.1815	28.9758	28.0575	28.3796	31.2737	26.2009	28.511
PCA 1st stage	28.766	30.204	29.6746	29.5114	26.8473	26.8473	29.3675
PCA 2nd stage	28.7573	30.5415	30.0384	29.7184	31.5945	26.7205	29.5617

From Table 1, it is clear that LPG-PCA based denoising technique is better in obtaining better noise residues.
After obtaining noise residues, the PRNU pattern or camera fingerprint is obtained using equation (2). At first reference fingerprint is estimated from the available frames. PRNU pattern for each frame is estimated. After this, the classification or attribution part comes. Multi-class SVM technique as given in Fig 3 is used for attribution of videos to its source device. Multi class SVM is used as it solves the problem of classification of more than two classes as it forms multiples of two class classifiers based on the feature vector derived from the input features and the class of the data. The feature vector is a file in matrix form which contains the variance, standard deviation and the mean of the SPN of each frame from the dataset. Now the fingerprint of each frame is divided into testing dataset and training dataset.

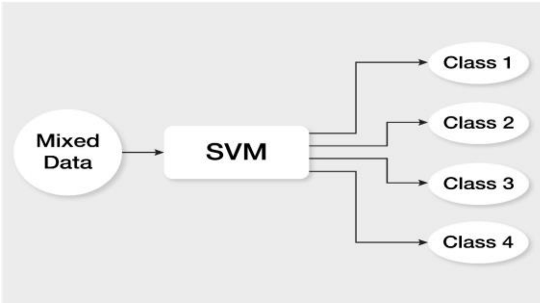

Fig -3: Multi-class SVM Block diagram

The algorithm uses two different datasets as mentioned: the training dataset and the testing dataset. The training dataset contains the tagged videos from known mobile devices models. The testing dataset is randomly sampled from the videos of mobile devices to be identified. After classification, the frames of videos under analysis or query video will be attributed to its source device. The accuracy of attribution depends on the performance of classifier.

4. EXPERIMENTS AND RESULTS

In this section, we report the results and experiments conducted. First, we describe the number of videos used for the work. Then, we will discuss the evaluation metrics and finally the results achieved by the proposed method.

4.1 Dataset

In order to test the proposed method, 10 different camera sources were taken as given Table 2. This dataset has been taken from Vision dataset [6]. We selected devices which has Full-HD resolution (1920 x1080 pixel).

Table -2: Camera sources from Vision datset

No of Sources	sources	Digital stab	Video resolution
1	Samsung galaxy s3 mini	off	1280 X 720
2	Apple iphone6	on	1920 X 1080
3	Huawei_Honor5c	off	1920 X 1080
4	Huawei_P8	off	1920 X 1080
5	Lenovo_P70A	off	1920 X 1080
6	OnePlus_A3003	on	1920 X 1080
7	Samsung galaxy s5	off	1920 X 1080
8	Xiaomi_RedmiNote3	on	1920 X 1080
9	Lumia 640	off	1920 X 1080
10	Galaxy s4 mini	off	1920 X 1080

4.1 Results

In this section, the different results obtained will be described: First, the conversion of videos into intra-coded frames is shown in Fig 4.

Fig -4: Video to I-frames

The denoised version of frames obtained using LPG-PCA is given below. Fig 5 is the original frame. Fig 6 is the first stage of denoising and Fig 7 is the second stage of denoising. For better understanding one frame is given below.

Fig -5: Original frame

Fig -6: Denoising first stage output

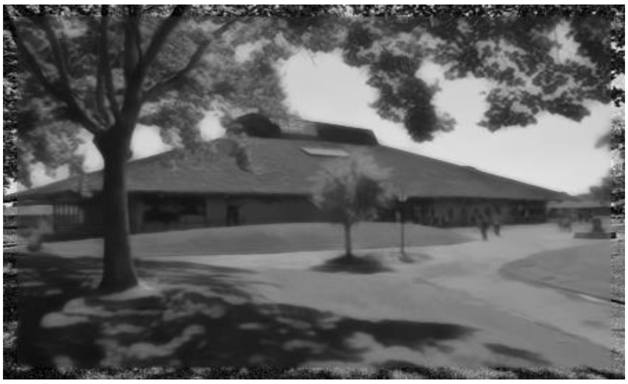

Fig -7: Denoising second stage output

After denoising of each frame, the fingerprint or PRNU pattern is obtained as in Fig 8 below.

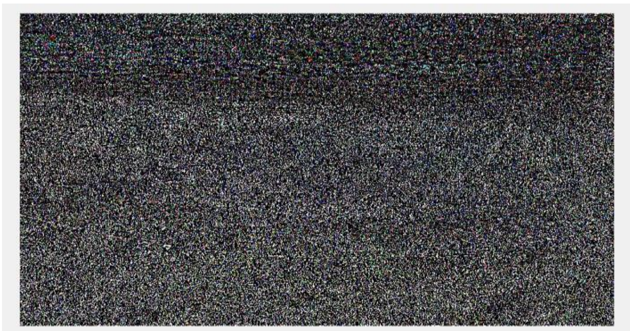

Fig -8: A PRNU pattern obtained from the noise residues

5. EVALUATION METRICS

In order to find the accuracy, we need to plot the receiver operating characteristic (ROC) curves. We consider all videos recorded with that camera as positive samples and the set of negative sample includes sequences not taken with that camera. ROC curve depicts the relation between true positive rate (TPR) and false positive rate (FPR). To calculate numerically, AUC (area under the curve) can be calculated.

Also, for measuring the performance of classification, confusion matrix is obtained.
The proposed work is performed on MATLAB platform. The classification part is mainly done on 3 classes for stabilized and non-stabilized videos and also on non-stabilized videos alone. Also, classification is performed on the basis of stabilized and non-stabilized videos. Non-stabilized videos are usually more correctly classified than stabilized videos. It is because stabilization reduces the noise content of the videos. For each video around 1000 I frames were taken.

Table -3: Confusion matrix for 3 class non-stabilized videos

confMat
3x3 double

	1	2	3	4
1	1000	0	0	
2	0	1000	0	
3	0	0	1000	
4				

The confusion matrix for 3 class stabilized videos obtained is given in Table 4.

Table -4: Confusion matrix for 3 class stabilized videos

	Predicted			
	classes	1	2	3
Actual	1	910	58	20
	2	25	918	29
	3	12	8	949

Table -5: TP, TN, FP & FN values

	Class 1	Class 2	Class 3
TP	910	918	949
TN	1904	1891	1911
FP	78	54	20
FN	37	66	49

Table 4 shows the values of the true positive (TP), true negative (TN), false positive (FP) and false negative (FN). From this, we can calculate the values of accuracy, precision, recall and specificity. These values can be calculated from the following equations.

$$\text{Accuracy} = \frac{TP + TN}{Total} \qquad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \qquad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (5)$$

$$\text{Sensitivity} = \frac{TN}{TN + FP} \qquad (6)$$

The corresponding values obtained are given in Table 6 below.

Table -6: Values of accuracy, precision, recall and sensitivity

	class 1	class 2	class 3
Accuracy	0.938	0.9363	0.953
Precision	0.921	0.944	0.979
Recall	0.96	0.932	0.9509
Specificity	0.96	0.9722	0.9896

From the table above, it is clear that the accuracy is around 95 % for non-stabilized videos. The first class and third classes are from non-stabilized videos. The second class is from stabilized videos. Since, we have taken only 1 stabilized video, the accuracy is comparable with non-stabilized ones. As the number of classes increase, the accuracy of stabilized may decrease. The accuracy of stabilized videos is less than non-stabilized here.
The ROC curve for these 3 classes is given in Fig 9 and ROC of stabilized and non-stabilized videos is given in Fig 10.

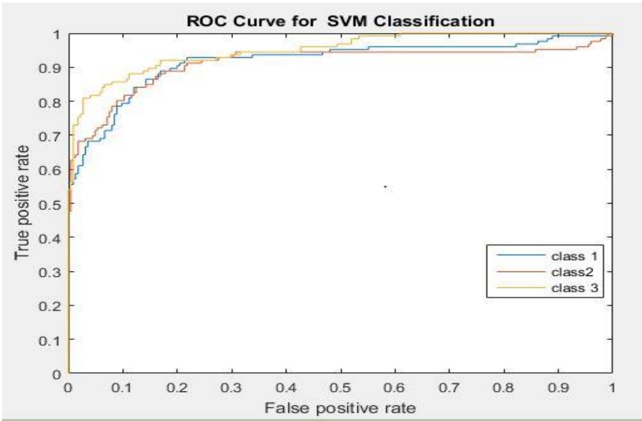

Fig -9: ROC for three classes of stabilized videos

From ROC curves, we can find the area under the curve (AUC). Ideally, AUC must be 1. The AUC values for the three classes is given in Table 7.

Table -7: AUC values

classes	Area under the curve (AUC
class 1	0.9173
class 2	0.9114
class 3	0.9498

From the above table, it is clear that all the values of AUC is above 0.9, which is close to 1.

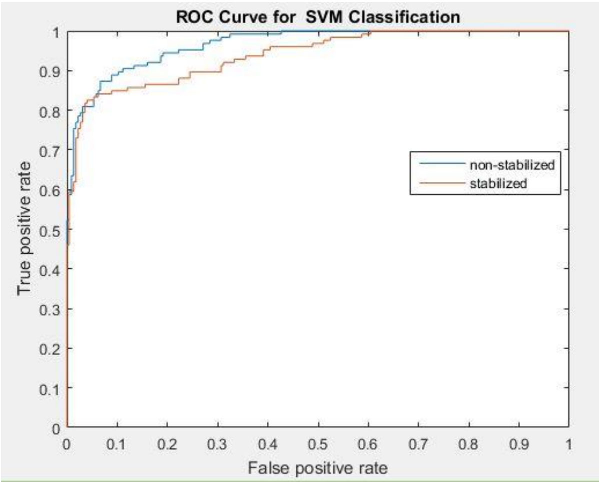

Fig -10: ROC for stabilized and non-stabilized videos

From Fig 10, it is clear that non-stabilized videos have accuracy slightly greator than stabilized videos. But, overall the accuracy is above 90%. For non-stabilized videos, the TPR IS 1.

6. CONCLUSIONS

In this paper, we propose a video source attribution technique using fingerprint derived from frames. The experiments are conducted on publicly available vision dataset composed of both stabilized and non-stabilized videos. The proposed method involves device fingerprint extraction, which is the PRNU pattern extraction and the fingerprints are classified using multi-class SVM. The proposed method can be applied for source device attribution, device linking and forgery detection as well. The best results were obtained if non-stabilized videos were used for attribution. Overall, the experimental results showed that the accuracy of non-stabilized videos (here for only 3 videos) is 100% with a TPR of 1 and for stabilized videos, accuracy is around 93%. Also, it should be noted that the accuracy may decrease with the increase in videos for stabilized cases. The future work is to work with more classes and to improve the performance of stabilized video sequences as well.

REFERENCES

[1] S. Milani *et al.*, “An overview on video forensics,” *APSIPA Trans. Signal Inf. Process.*, vol. 1, pp. 1–37, Dec. 2012.

[2] M. Kirchner and T. Gloe, “Forensic camera model identification,” in *Handbook of Digital Forensics of Multimedia Data and Devices*, 2015, pp. 329–374.

[3] J. Lukáš, J. Fridrich, and M. Goljan, “Digital camera identification from sensor pattern noise,” *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 2, pp. 205–214, Jun. 2006.

[4] Massimo Iuliani , Marco Fontani , Dasara Shullani and Alessandro Piva, “Hybrid reference-based Video Source Identification.” Sensors 2019

[5] C. Meij and Z. Geradts, “Source camera identification using photo response non-uniformity on WhatsApp,” Digit. Invest., vol. 24, Mar. 2018, pp. 142–154.

[6] D. Shullani, M. Fontani, M. Iuliani, O. A. Shaya, and A. Piva, “VISION: A video and image dataset for source identification,” EURASIP J. Inf.Secur., vol. 2017, no. 1, p. 15, 2017.

[7] F. Marra, G. Poggi, C. Sansone, and L. Verdoliva, “Blind PRNU-based image clustering for source identification,” *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 9, pp. 2197–2211, Sep. 2017.

[8] M. Goljan and J. Fridrich, “Camera identification from cropped and scaled images,” Proc. SPIE, vol. 6819, Feb. 2008, Art. no. 68190E

[9] M. Iuliani, M. Fontani, D. Shullani, and A. Piva, “Hybrid referencebased video source identification,” *Sensors*, vol. 19, no. 3, p. 649, 2019.

[10] M. Chen, J. Fridrich, M. Goljan, and J. Lukáš, “Determining image origin and integrity using sensor noise,” IEEE Trans. Inf. Forensics Security, vol. 3, no. 1, Mar. 2008, pp. 74–90.

[11] M. Chen, J. Fridrich, M. Goljan, and J. Lukáš, “Source digital camcorder identification using sensor photo response non-uniformity,” *Proc. SPIE*, vol. 6505, Mar. 2007, Art. no. 65051G.

[12] S. Mandelli, P. Bestagini, L. Verdoliva, and S. Tubaro, “Facing device attribution problem for stabilized video sequences,” IEEE Transactions on Information Forensics and Security, 2019.

[13] Bo Wang,Yiping Guo,Xiangwei Kong,Fanjie Meng, “Source camera identification forensics using wavelet features”, 2009.

[14] Sabita Pal, Rina Mahakud, “PCA based image denoising using LPG”, 2011.

BIOGRAPHIES

1) **Darsana B S,** PG Scholar is with the Sree Chitra Thirunal College of Engineering, NH 66, CTO Colony, Pappanamcode, Thiruvananthapuram, Kerala 695018
2) **Subha V** , Assistant Professor is with the Sree Chitra Thirunal College of Engineering, NH 66, CTO Colony, Pappanamcode, Thiruvananthapuram, Kerala 695018