University of Ibadan
Journal of Science and Logics in ICT Research

Blind Image Clustering Using Contaminated Sensor Pattern Noise

Joshua Olaniyi-Ibiloye, Khadijat T. Ladoja*, Joseph D. Akinyemi, O. F. W. Onifade
jolaniyiibiloye@yahoo.com, kt.bamigbade@ui.edu.ng, akinyemijd@gmail.com, ofw.onifade@ui.edu.ng

Department of Computer Science, University of Ibadan
* Corresponding Author

Abstract
With digital imagery fast becoming a part of our daily lives and the exponential development of image processing technologies, new challenges and problems are also rising. One such problem is that of identifying the source device of an image. Previous attempts to do this were focused on identifying sources for which there were some information about attacking the problem from a Supervised Learning standpoint. In this research, we present an alternative model for image source identification, in the absence of any information about the images, using properties generated during the image processing pipeline, which is the dominant Photo Response Non-Uniformity (PRNU), along with other impurities combined to form the contaminated sensor pattern noise or Polluted PRNU (POL-PRNU). Results showed a relatively low accuracy of 46% achieved by our model. It was also observed that there was a higher level of misclassification between cameras from the same manufacturer although the models were different and this affected the overall accuracy of the model. While Sensor pattern noise can be used to cluster images, it would require some more refinements in order to obtain a higher clustering accuracy.

Keywords: *Correlation clustering, Image forensics, Image processing, Image source identification*

1. INTRODUCTİON

Image clustering is a fundamental task in computer vision with various applications including image retrieval, image segmentation, and object recognition. Traditional clustering techniques rely heavily on visual feature extraction, such as colour and texture descriptors. However, these methods often suffer from high computational complexity and inadequate performance. The widespread use of digital imaging technology, coupled with advanced image processing and computer graphics capabilities, presents new challenges in ensuring the accountability and credibility of digital images due to concerns about authenticity, credibility, accountability, and privacy [12].

The problem of image source identification has attracted significant attention over the years as it serves as a viable means of authenticating images, especially in forensics image analysis [1]. To this effect, various methodologies have been developed to solve this problem, all of which are based on different aspects of the image processing pipeline.

In the analogue world, a photograph or an image has become generally acceptable as a proof of the occurrence of a depicted event. In this digital age, creating and manipulating digital images had been simplified by powerful digital processing tools that are readily available even on smart mobile phones. As a result of this, the authenticity of images, whether they are analog or digital, can no longer be assumed. This becomes even more important with legal photographic evidence. In this context, Image Forensics is therefore concerned with determining certain underlying facts about an image [10]. For instance, suppose there is an ongoing court case and the prosecuting counsel presents, as evidence, some images that allegedly place the defendant at the scene of the crime, or worse, holding the murder weapon. The defending counsel is then

Joshua Olaniyi-Ibiloye, Khadijat T. Ladoja, Joseph D. Akinyemi and O. F. W. Onifade (2023). Blind Image Clustering Using Contaminated Sensor Pattern Noise. *University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR),* Vol. 10 No. 1, pp. 85 – 95.

faced with the task of certifying the authenticity of that image. It is decided that, to achieve this, the source of the images should be ascertained. That is, identifying which camera or cameras took those images.

Image forensics comprises a range of techniques aimed at offering definitive responses to inquiries such as:

1. Is this image genuinely "original," or was it produced through copying and pasting elements from different images?

2. Which camera manufacturer produced the device that captured this image?

3. Is this image an accurate depiction of the original scene, or has it been digitally altered to mislead the viewer? For instance, was a coffee stain manipulated to appear as if it were a bloodstain that had been recolored?

There are different sources of imperfection and noise that enter an image at different stages of the image processing pipeline. One of these is a deterministic component that stays approximately the same if multiple pictures of the exact same scene are taken and is present in every image taken by a particular sensor. It is called the Sensor Pattern Noise (SPN) and it is made up of the Fixed Pattern Noise (FPN) and the Photo Response Non-Uniformity (PRNU). The FPN is caused by dark current and is basically the pixel-to-pixel differences when the sensor is not exposed to light; it is sometimes suppressed automatically in some middle to high-end cameras. The PRNU is caused by the heterogeneity of the silicon wafers used in manufacturing camera sensors and some imperfections during the manufacturing process. It is the dominant part of the sensor pattern noise and some research works have called it the “fingerprint” of a digital camera.

This research seeks to use this fingerprint to cluster images without any prior information about them. The premise is that the PRNU from images that come from the same sensor would exhibit more correlation between them than images from different cameras.

2. RELATED WORKS

There have been previous research efforts to identify image sources from camera noise patterns. Most previous works can be broadly categorized into Supervised and Unsupervised approaches.

Research in image source identification, over the years, has mostly been geared towards images about which there is sufficient information (supervised learning). But images in the real world do not come with any information about them asides from the metadata. While the metadata provides some information about the source of an image, it can easily be altered and thus cannot be trusted as proof of authenticity.

Depending on the available prior knowledge, the task of identifying a digital camera from photographs could take different forms. However, there is often little or no prior information on the devices possibly involved let alone on their PRNU pattern. Usually, an analyst is given a few images without much useful metadata associated with them and with little or no relationship between the images making device identification more difficult with such approaches. Nevertheless, it may still be very useful to understand, even in such cases, which images are from the same camera and which are not [5].

Due to technological advancements, image acquisition devices are becoming more pervasive and image editing tools are becoming more common and easy to use. It therefore, becomes very important to be able to reliably verify the integrity of images. Image forensics involve techniques for preventing the malicious tampering of images for illegitimate benefits. Of particular interest in multimedia forensics is source camera identification and its purpose is to trace the sources of images. This can indeed help identify the sources of images emanating from crime scenes or terrorist attack scenes and thus ensure the security and integrity of such digital information.

Most of the basic methods of individual camera identification focus mostly on Sensor Pattern Noise (SPN). Due to manufacturing imperfections, the difference between the parameters of the MOSFET (Metal–Oxide–Semiconductor Field-Effect Transistor) and the photodiode in each pixel can cause minor distortions in the final output signal of each pixel. Researchers believe that these indelible marks left by a camera as a result of slight manufacturing defects can be exploited as "fingerprints" for source camera identification from pictures. Thus, individual camera identification is often carried out using SPN-based methods.

Shen *et al.,* [13] applied direct SPN-based clustering and achieved promising results in terms of both accuracy and stability. Similarly, employing SPN-based feature extraction combined with traditional clustering algorithms has demonstrated improved clustering accuracy and stability. Meng et al., [11] utilized ICA to extract SPN-based features and applied k-means clustering on the feature vectors, achieving superior performance compared to direct SPN-based clustering. However, the limitation of this approach lies in the selection of appropriate feature extraction algorithms and parameter optimization

In Timmerman and Alegre [15] work, a constrained convolution net was developed to identify the source camera of a video based on the specific sensor noise patterns extracted from video frames. using this approach, their network classified individual images (video frames) and then used majority voting to identify the source camera of the video. They reported a 93.1% accuracy on the VISION dataset containing 1539 videos from 28 different cameras.

In Hui *et al.,* [3], the objective was to tackle the challenge of device identification based on Sensor Pattern Noise (SPN) and introduce an innovative approach called the Multi-Scale Feature Fusion Network (MSFFN) to enhance the accuracy of attributing source cameras based on sensor patterns. The MSFFN, structured as a multi-scale encoder-decoder, plays a pivotal role in suppressing image content, thereby improving source identification. Following this, the content-independent SPN features from different scales were combined, and ultimately, these fused features were employed for the purpose of identifying the source of the image.

In Kirchner and Johnson [7] work, a Sensor Pattern Noise Convolutional Neural Network (SPN-CNN) was developed for improving the process of extracting sensor pattern noise from images. They reported an average identification accuracy of 82% on the VISION dataset.

In Bernacki [2] work, the robustness of digital camera based on a Convolutional Neural Network (CNN) was discussed. In this context, robustness pertained to the network's ability to identify a camera even when presented with visually distorted images. To evaluate this, the CNN was trained on "normal" images captured by certain cameras and subsequently tested on images from the same cameras that had undergone degradation via Gaussian blur, Poisson noise, random noise, and the removal of the Least Significant Bit (LSB) of pixel intensities. The outcomes of this analysis revealed that the CNN could effectively attribute significantly altered images to their respective camera sources.

In Freire-obregon *et al.,* [4] work, the authors proposed a Convolutional Neural Network (CNN) architecture that could deduce the noise patterns inherent in mobile camera sensors. The primary objective was not only to accurately detect and identify the mobile device that captured an image (achieving a 98% accuracy rate) but also to determine which embedded camera within the device was responsible for capturing the image.

In Lukas *et al.,* [6] work, it was stated that the camera identification problem need to be approached from multiple dimensions by combining evidences from different methods. Hence, this research explores a different approach to camera identification which can be used as a means of validating the authenticity of an image or images.

Li *et al.,* [8] introduced a methodology aimed at distilling the reference Photo Response Non-Uniformity (PRNU) by eliminating interference noise through the utilization of Principal Component Analysis (PCA) technology. Specifically, the reference PRNU noise was represented as white Gaussian noise, whereas the interference noise induced correlations between pixels and their neighboring elements within the reference

PRNU noise. In the context of local pixel regions, a pixel and its neighboring pixels were treated as a vector and employed block matching to select PCA training samples that shared similar content. Subsequently, PCA transformation was utilized to estimate the interference noise within the local pixel region and implemented coefficient shrinkage in the PCA domain to enhance the accuracy of interference noise estimation.

In Shokunbi *et al.,* [14] a machine learning approach inspired by the Bayer filter and the demosaicing procedure employed in digital color cameras to enable the colorization of grayscale images was discussed. Their method entailed training a multilayer perceptron model on a dataset of color images that share semantic similarities. Subsequently, the model demonstrated the ability to add color to grayscale images that share semantic similarities with those in the training dataset.

Upon a thorough review of the existing literature related to camera identification and the analysis of sensor noise, several significant trends and findings have surfaced. Researchers have delved into a variety of methodologies, including the utilization of Sensor Pattern Noise (SPN)-based clustering, convolutional neural networks (CNNs), and conventional clustering algorithms to associate digital images with their originating cameras. While these methods have shown promise in accurately discerning the source of images and managing variations in them, a substantial research gap becomes evident in their ability to handle challenges arising from images originating in different domains or those that have been visually altered, whether through noise, blurriness, or other modifications. Our research endeavors to address this void by introducing an innovative approach that combines the merits of SPN-based feature extraction and photo response non-uniformity to augment the resilience and precision of camera identification across diverse domains and amidst a spectrum of image distortions.

3 METHODOLOGY

As earlier stated, this study aims to utilize PRNU, in conjunction with other image characteristics, to group images into clusters, by grouping together all images captured by the same camera within the same cluster. Typically, the PRNU itself is determined by averaging the noise residuals from several images known to be produced by the same camera. This also suppresses the other random low-frequency noise components resulting in a stronger estimate of the PRNU. Our methodology explores the correlation between the POL-PRNU (Polluted Photo Response Non-Uniformity) from random images in a dataset to create clusters such that, each cluster would represent images most likely taken by the same camera. Figure 1 shows an overview of our methodology.

Digital camera sensors often contain several photo-detectors (pixels) which are responsible for converting photons (light rays) into electrons. Each pixel in the sensor of a digital camera is designed to record the amount of incident light striking it and this together forms an image. However, slight manufacturing imperfections sometimes introduce small amounts of noise into the resultant image. This noise has a stochastic nature, is unique for each sensor and yet spatially variant and consistent over time, making it suitable for use for forensic purposes such as camera identification from photographs. Typically, most PRNU-based camera identification methodologies extract residual noise from images by subtracting the denoised version of an image from the original (noisy) image as shown in equation 1.

$$P = I - WF(I) \qquad (1)$$

where I is the image, $WF(I)$ is the denoised image, and WF is a wavelet filter.

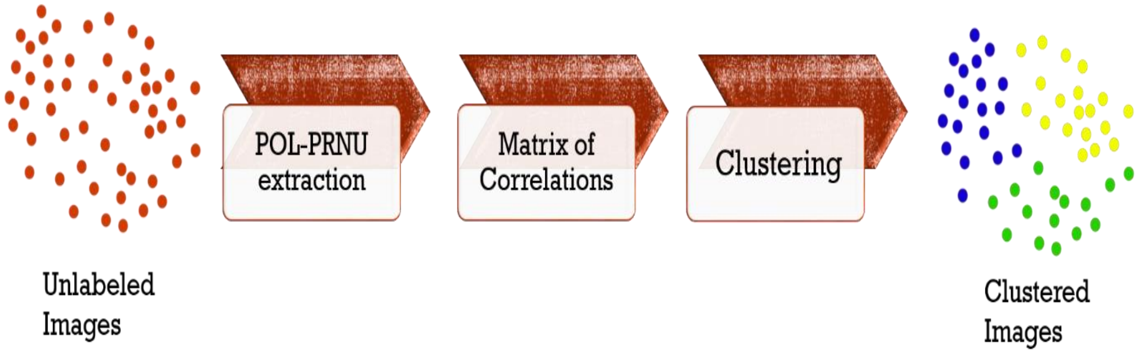


Figure 1: Overview of the methodology

A wavelet-based denoising filter is often recommended and it is used in most cases because it provides the least amount of traces of the scene. Residual noise is actually a sum of different noises, one of which includes the sensor pattern noise, PRNU. Other types of noise, which are also a part of the residual noise, such as image content, may also pollute the PRNU. An example of this is presented in Figure 1 showing an image and its residual noise or POL-PRNU.

3.1 Extraction of the POL-PRNU

First of all, each image is split or decomposed into its three color channels, and then, a central block of 1024 x 1024 pixels is extracted since a small block of pixels from the original image size can significantly reduce computational complexity and speed up processing. The extraction process is conducted individually on each of the three-color channels, namely Red, Green, and Blue.

To mitigate any artifacts arising from color interpolation and JPEG compression, a periodic signal known as the linear pattern L is derived by subtracting the average row from each row and the average column from each column of N, separately for each color channel. This yields three distinct linear patterns: Lr for the Red channel, Lg for the Green channel, and Lb for the Blue channel. Ultimately, these three patterns are merged using the grayscale conversion formula as outlined in equation 2. Using the recombined linear patterns will be more reliable because the three linear patterns (*Lr*, *Lg* and *Lb*) are highly correlated and they provide compact information.

$$P = 0.3Lr + 0.6Lg + 0.1Lb \qquad (2)$$

The extraction process is described in Figure 1 and the resulting POL-PRNU is shown in Figure 3.

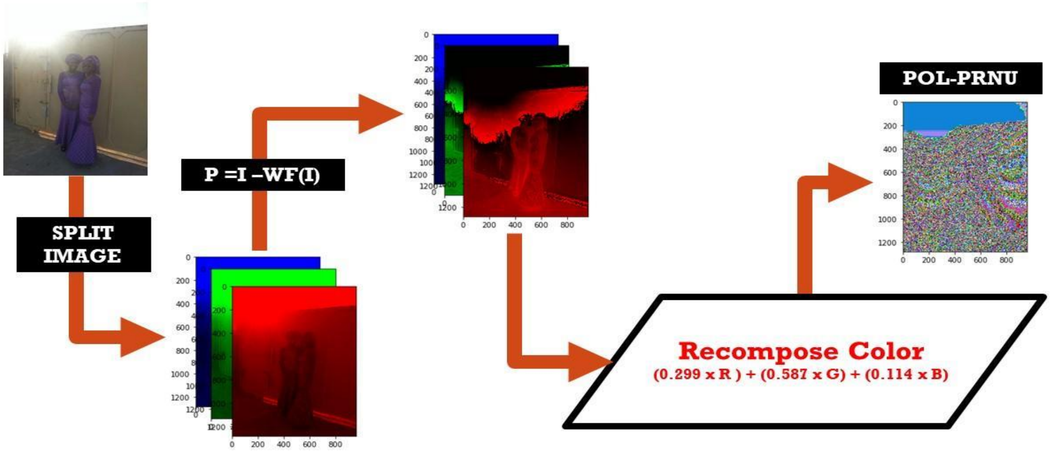
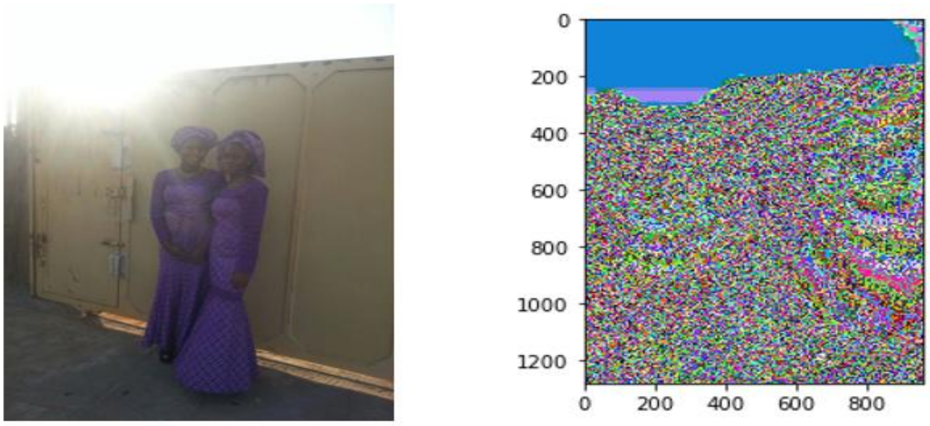


Figure 2. POL-PRNU extraction



Original Image **POL-PRNU**

Figure 3**.** Original image and its extracted POL-PRNU

3.2 Clustering

Typically, images are mapped to their source cameras by generating a reference pattern for each camera under observation and comparing the pattern extracted from each image to this reference pattern. Because the sensor pattern noise is a weak signal, the reference pattern is generated or, in other words, estimated, by averaging the sensor pattern noise from lots of images known to have been taken by a particular camera. The premise is that pattern noise from images taken by a particular camera would exhibit a much higher correlation with the reference pattern of that camera than with any other camera. Hence, we can say that the image was most likely taken by that camera.

Based on this premise we consider the correlation between each of the extracted POL-PRNU for each of the images under study. We assumed that since pattern noises from images would show a higher correlation with reference patterns from their source camera, then these images will also show a higher correlation with other images taken by the same camera. Our approach is to cluster images such that each cluster represents images most likely taken by the same camera.

To facilitate image clustering, we construct an M by M pairwise correlation matrix, which signifies the correlation between each image and every other image within the dataset, with M representing the total number of images in the dataset.

The degree of correlation between data sets serves as a metric for their relationship. Pearson's Correlation is the most widely used statistical correlation measure, revealing the linear association between two data sets. It yields a value within the range of (-1, 1), where -1 signifies a robust negative correlation, 1 indicates a strong positive correlation, and 0 denotes no correlation, often referred to as zero-correlation. The calculation of Pearson's correlation coefficient is detailed in equation 3.

$$r = \frac{n(\sum xy)-(\sum x)(\sum y)}{\sqrt{[n(\sum x^2)-(\sum x)^2][n(\sum y^2)-(\sum y)^2]}} \quad (3)$$

The resulting correlation matrix is then subjected to an agglomerative hierarchical clustering algorithm which uses the single link technique to update the correlation matrix. This is done by merging clusters with the highest correlation with every iteration. In other words, each image is regarded as a cluster on its own and then the clusters are merged based on the updated correlation matrix.

3.3 Dataset

For this research, images were gathered from 3 different digital cameras, two of which are from the same manufacturers but of different models:

1. Nikon D1500
2. Nikon D1200
3. Canon EOS 1200D

Fifteen images were gathered for each camera and these were random images of different scenarios and different conditions, this was done so as to test the robustness of the model being developed. Thus, we had a total of 45 images obtained by taking shots of random scenes with each of these cameras.

After the POL-PRNU has been extracted from each image as previously described, the POL-PRNU is then cropped and flattened to a 1-Dimensional vector which is then put in a data frame structure as shown in Figure 3, so as to be able to calculate the correlation between the images and cluster them thereafter.

4 RESULTS

Utilizing a smaller pixel block from the original image size considerably diminishes computational complexity and expedites the matching procedure. In Li and Satta [9] , it was demonstrated by the authors that the false-positive rate (FPR) in camera identification diminishes as the pixel block's size increases, attaining its lowest point when the pixel block size reaches 1024 × 1024 pixels.

Based on this finding, two iterations of the experiment were carried out using:

1. The 1024 × 1024pixel block mentioned above.

2. A much smaller 32 × 32pixel block.

The remaining part of this article presents an analysis of the results of the experiments.

To evaluate the performance of both iterations of our model, we provide answers to two questions:

1. How well did our clustering do?

2. Did we predict the correct number of clusters in our dataset?

To answer the first question, the cophenetic correlation coefficient (or simply the Cophenetic coefficient) was calculated. The cophenetic correlation coefficient is a measure of how well a dendrogram preserves the pairwise distances between the original unmolded data points. Although it has been mostly applied in Biostatistics, it is also suitable for use in other fields where raw data tends to occur in clusters. The cophenetic Correlation Coefficient is simply the correlation coefficient between the distance matrix and the Cophenetic matrix. The closer it is to 1 or 100%, the better the clustering fits.

Figure 4 shows the cophenetic coefficient for the two iterations of our experiments and it can be clearly seen that our model performs better with the smaller block than with the much larger one as the experiment with the 32 × 32 pixel block obtains a higher cophenetic coefficient.

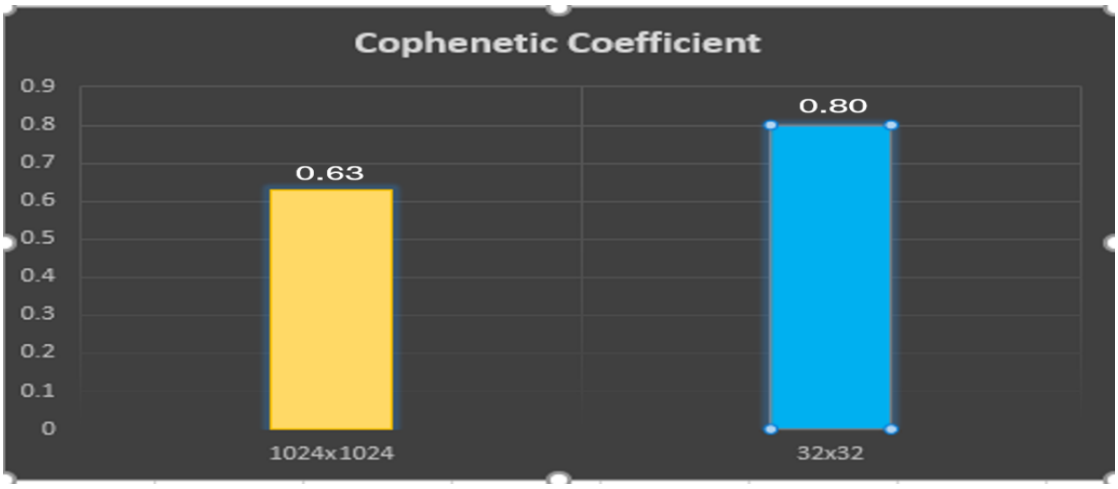


Figure 4: Copehenetic coefficient for both experiments

To answer the second question, we provide heatmaps to find out if the correct number of clusters have been predicted from the images. After clustering, it is observed that, plotting the heatmap of the clusters, as seen in Figure 5, shows that there are no visible clusters for the larger block. The two heatmaps in Figure 5 represents the level of correlation between the images taken by the different cameras. It is obvious that the smaller block-sized (32 × 32) images show much visible patterns in the heathmap and thus higher correlation than the larger block-sized (1024 × 1024) images. More precisely, three clusters are visible in the heatmap of the 32 × 32 block-sized images indicating that the images have come from three different sources. However, no such patterns are obvious in the heatmap of the 1024 × 1024 block-sized images.

The dendrograms in Figure 6a and 6b give a more detailed view of the clustering and it shows that for both iterations of the experiments, there were three clusters which correspond to the three different cameras used to populate the dataset. Again, once can see that the dendogram of the first iteration of the experiment where, 1024 × 1024 blocks (Figure 6a) were used, do not have well defined clusters like we have with the 32 × 32 blocks (Figure 6b). This is consitent with the observations in Figures 4 and 5. The dendograms of the larger block size identify all images as originating from nearly the same source, showing all lines in the same colour, while the dendograms of the smaller blocks show three distinct colours for the clusters indicating that the images originate from three different cameras.

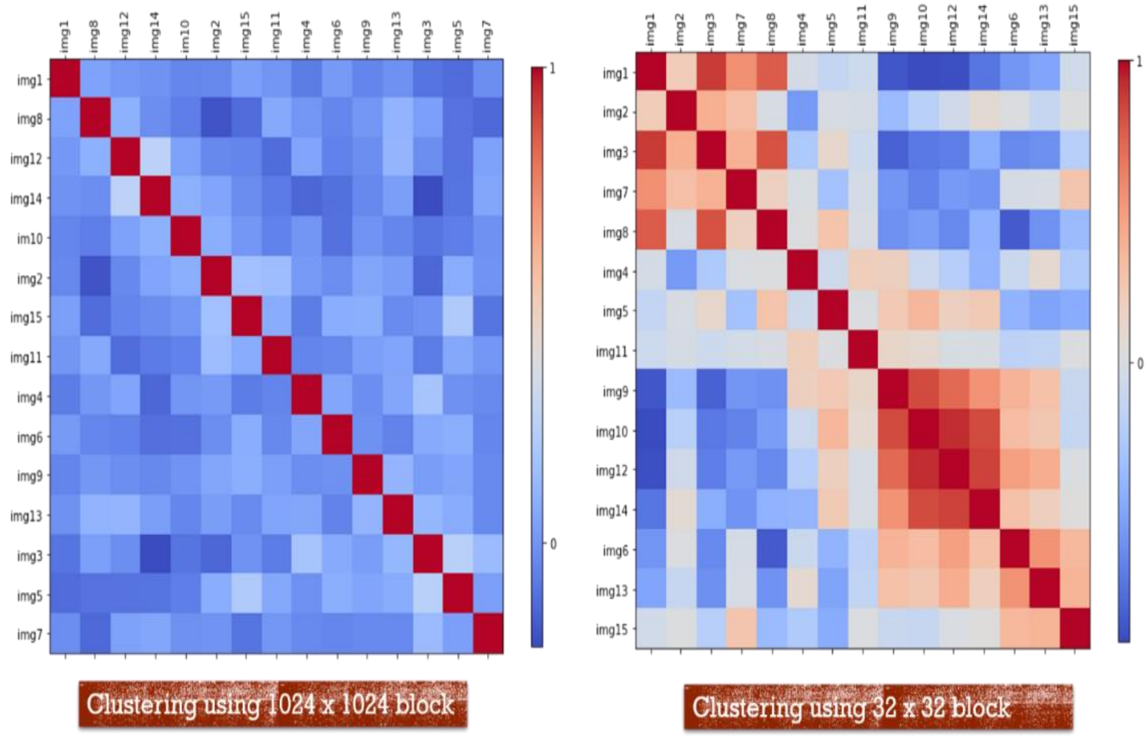


Figure 5: Heat maps showing Correlation-based clustering for both experiments

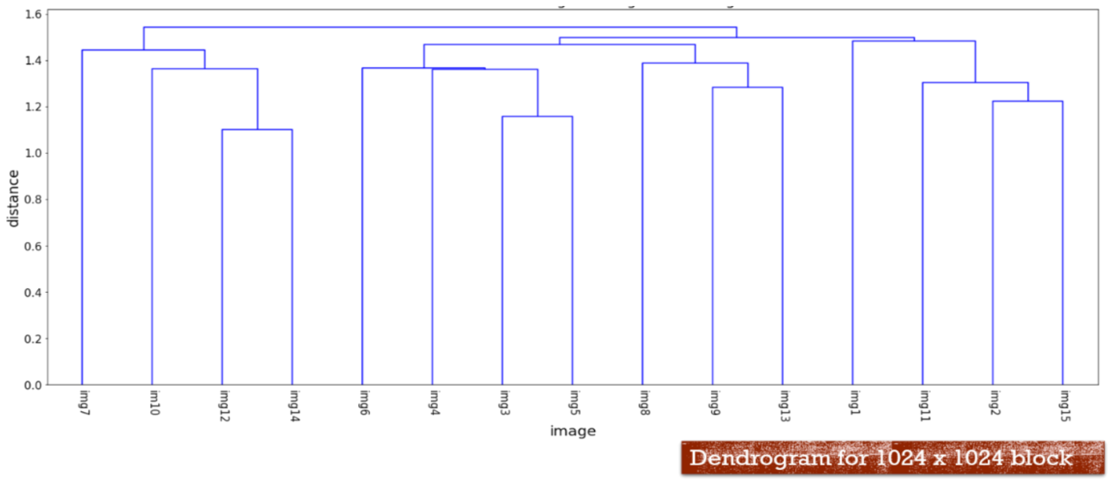


Figure 6a: Dendograms for both 1024 × 1024 block

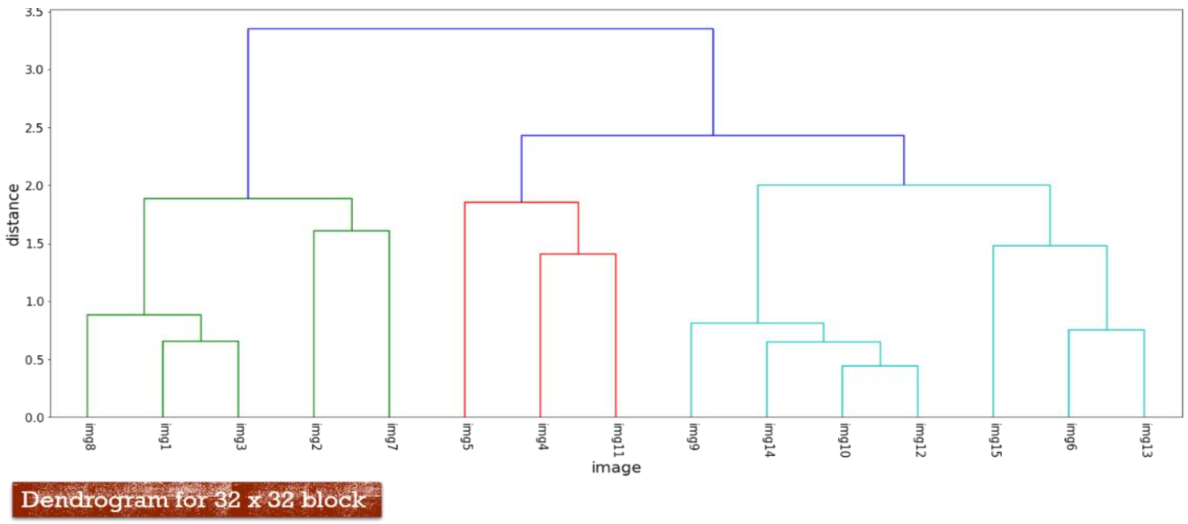


Figure 6b: Dendograms for both 32 × 32 block

Figure 7 further reinforced our findings showing that the smaller block-size of 32x32 pixels achieved better clustering accuracies across all three clusters and even overall than the larger block-size of 1024x1024 pixels which was supposed to have the best false positive rate. Overall, the smaller block-size achieved an accuracy of 46% as against 33% achieved by the larger block-size. This finding holds the promise of a more effective and less computational method of identifying image sources. It was also observed, during clustering, that there was a higher level of misclassification between cameras from the same manufacturer (Nikon) even though they were of different models. This, along with the relatively small sample size, affected the overall accuracy of the model, but we have still been able to establish the fact that clustering can reveal the differences in image sources, especially at low pixel block sizes. The relatively-low accuracy is quite expected, given that we only tested with a relatively small number of images, but we consider the differences in the performance based on the different block-sizes a significant finding.

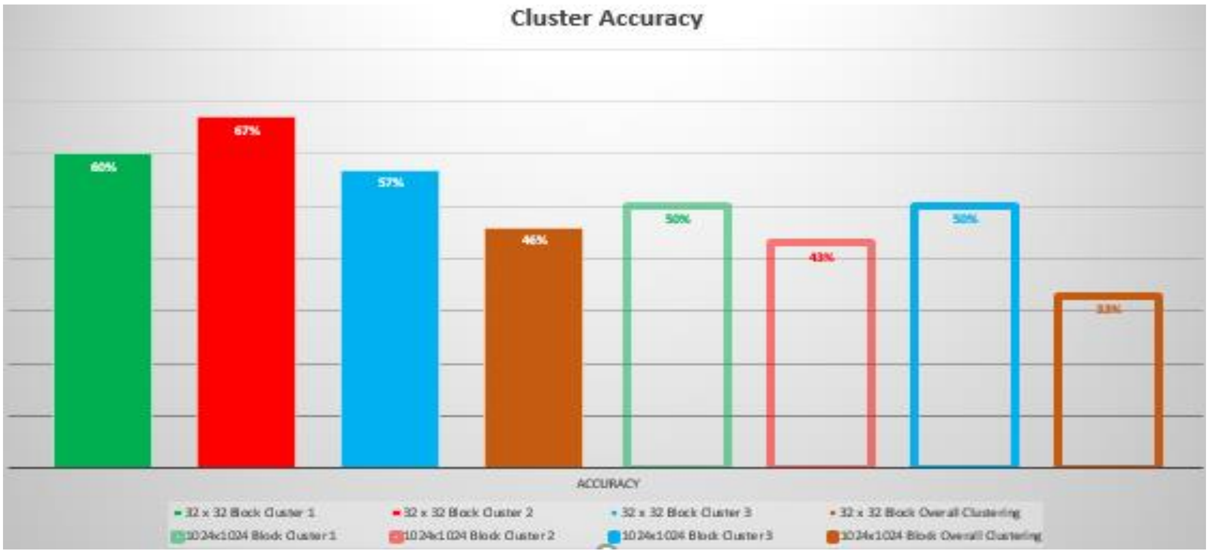


Figure 7: Clustering accuracies of the 32 x 32 block (solid bars) and the 1024 x 1024 block (outlined bars)

5. CONCLUSİON

In the real world, images do not come with much information besides the metadata. With the availability of applications like Adobe Photoshop, it is quite easy to alter the metadata of images hence it cannot be trusted. This research provides such a clustering-based model for identifying images sources using the impurities inserted into images in the image sensor of digital cameras. It was observed that images obtained from different camera sources could be identified more accurately at a lower pixel block sizes of 32x32 than at a higher pixel block sizes of 1024x1024. This provides an advantage of less computational demand as well as improved identification accuracy. Future works could experiment with different pixel block sizes between the two as well as on a larger dataset.

References

1. Anthony T. S. Ho; Shujun Li, "Forensic Camera Model Identification," in *Handbook of Digital Forensics of Multimedia Data and Devices* , *IEEE*, 2015, pp.329-374, doi: 10.1002/9781118705773.ch9.

2. Bernacki, J. Robustness of digital camera identification with convolutional neural networks. *Multimed Tools Appl* **80**, 29657–29673 (2021). https://doi.org/10.1007/s11042-021-11129-y

3. C. Hui, F. Jiang, S. Liu and D. Zhao, "Source Camera Identification with Multi-Scale Feature Fusion Network," 2022 *IEEE International Conference on Multimedia and Expo (ICME)*, Taipei, Taiwan, 2022, pp. 1-6, doi: 10.1109/ICME52920.2022.9859965.

4. David Freire-Obregón, Fabio Narducci, Silvio Barra, Modesto Castrillón-Santana. Deep learning for source camera identification on mobile devices. *Pattern Recognition Letters*, Volume 126, 2019, Pages 86-91, ISSN 0167-8655, https://doi.org/10.1016/j.patrec.2018.01.005.

5. F. Marra, G. Poggi, C. Sansone and L. Verdoliva, "Blind PRNU-Based Image Clustering for Source Identification," in *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 9, pp. 2197-2211, Sept. 2017, doi: 10.1109/TIFS.2017.2701335.

6. J. Lukas, J. Fridrich and M. Goljan, "Digital camera identification from sensor pattern noise," in *IEEE Transactions on Information Forensics and Security,* vol. 1, no. 2, pp. 205-214, June 2006, doi: 10.1109/TIFS.2006.873602.

7. Kirchner, M., & Johnson, C. (2019). SPN-CNN: Boosting Sensor-Based Source Camera Attribution with Deep Learning. *2019 IEEE International Workshop on Information Forensics and Security (WIFS)*, 1-6.

8. Li J, Liu Y, Ma B, Wang C, Qin C, Wu X, Li S. A Novel PCA-Based Method for PRNU

Distillation to the Benefit of Source Camera Identification. *Applied Sciences*. 2023; 13(11):6583. https://doi.org/10.3390/app13116583.

9. Li, C. T., & Satta, R. (2012). Empirical investigation into the correlation between vignetting effect and the quality of sensor pattern noise. *IET Computer Vision*, *6*(6), 560-566. https://doi.org/10.1049/iet-cvi.2012.0044.

10. M. Kharrazi, H. T. Sencar and N. Memon, "Blind source camera identification," *2004 International Conference on Image Processing, 2004. ICIP '04.* Singapore, 2004, pp. 709-712 Vol. 1, doi: 10.1109/ICIP.2004.1418853.

11. Meng, D., Ding, W., Cheung, Y. and Pang, H., 2017. Blind image deblurring via spatially weighted total variation minimization. *IEEE transactions on image processing*, 26(6), pp.3006-3017.

12. Sevinc Bayram, Husrev T. Sencar, Nasir Memon. Classification of digital camera-models based on demosaicing artifacts. *Digital Investigation*. Volume 5, Issues 1–2, 2008, Pages 49-59, ISSN 1742-2876 (2008). https://doi.org/10.1016/j.diin.2008.06.004.

13. Shen, Y., Liu, T., Yang, X. and Shen, H.T., 2015. Blind image clustering exploiting high-order noise properties. *IEEE Transactions on Information Forensics and Security*, 10(1), pp.12-27.

14. Shokunbi, O.M., Akinyemi, J.D., Onifade, O.F.W. (2023). A Deep Multi-Layer Perceptron Model for Automatic Colourisation of Digital Grayscale Images. In: Chmielewski, L.J., Orłowski, A. (eds) Computer Vision and Graphics. ICCVG 2022. Lecture Notes in Networks and Systems, vol 598. *Springer*, Cham. https://doi.org/10.1007/978-3-031-22025-8_14.

15. Timmerman, D., Bennabhaktula, S., Alegre, E., & Azzopardi, G. (2020). Video camera identification from sensor pattern noise with a constrained convnet. *arXiv preprint arXiv:2012.06277*.