Detection of fake images and camera attribution using PRNU

Master Thesis Dissertation

Jose C. Mendoza-Jimenez

21 agosto 2023

Abstract

MSc Dissertation project WIP

Contents

1. Introduction

Digital image forensics plays a pivotal role in detecting image manipulations and forgeries. This project centers around the development of an advanced technique leveraging Photo Response Non-Uniformity (PRNU) patterns. The goal is to create a robust method for identifying potential image tampering and ensuring the integrity of digital visual content. This report presents the methodology, experimental results, and analysis of our PRNU-based approach.

1.1 Main concepts

1.1.1 PRNU

Pixel output

Considering the impact of manufacturing defects coming from the wafer silicon a sensor is made of, the raw output of a sensor can be expressed as:

$$Y = I + IK + \tau D + c + \theta$$

Where τ is a scalar multiplicative factor whose value is determined by exposure, temperature, and ISO Settings. The matrix c is the matrix of offsets and the dark current factor D is a noise-like signal due to leakage of electrons into pixels' electron wells. K is the PRNU factor. The modeling noise θ is a collection of all other noise sources, which are mostly random in nature and therefore difficult to use for forensic purposes (readout noise, shot noise, also known as photonic noise, quantization noise, etc.). For spike pixel defects, such as hot or stuck pixels, at least one of the three parameters, K, D and c , becomes large. Even though the parameters could be accurately estimated by taking special test pictures, such as pictures of dark scenes with long exposure, in forensic setting the camera may not be available and the challenge becomes to estimate these three matrices from the limited set of images available to the analyst. We also need to design an estimator for the onset of a defect and an estimator for the acquisition time of an unknown image by detecting it in the presence of defects.

Photo Response Non-Uniformity (PRNU) is a fundamental characteristic inherent to individual image sensors, arising from imperfections and variations during the manufacturing process. It manifests as a unique noise pattern that is present in the output of each sensor pixel when exposed to the same uniform illumination. PRNU is mainly attributed to variations in pixel sensitivity, dark current, and other sensor-specific factors. This noise pattern can be considered akin to a "fingerprint" for the camera sensor and is often treated as a quasi-unique identifier for digital images. PRNU has found significant applications in the field of digital image forensics, where it is employed for source camera identification and authenticity verification. By analyzing the PRNU noise pattern extracted from a given image, experts can determine the camera that captured the image, facilitating the identification of image origins, tampering detection, and even establishing the credibility of digital evidence in legal cases. The ability of PRNU to reveal subtle sensor-specific noise characteristics has revolutionized the field of multimedia forensics, playing a pivotal role in ensuring the integrity and traceability of digital imagery.

$$\text{PRNU}(\mathbf{I}) = \frac{\mathbf{I} - \mathbf{K} \odot \mathbf{I}}{\mathbf{K} \odot \mathbf{I}}$$

The normalized correlation:

$$\text{corr}(\mathbf{X}, \mathbf{Y}) = \frac{(\mathbf{X} - \bar{\mathbf{X}}) \odot (\mathbf{Y} - \bar{\mathbf{Y}})}{||\mathbf{X} - \bar{\mathbf{X}}|| \odot ||\mathbf{Y} - \bar{\mathbf{Y}}||}$$

FOr an image I obtained by a digitarl camera, its noise residual is defined as $W_{\mathrm{I}} = I - F(I)$, where F is a denoising filter. THe major component of the sensor fingerprint is due to photo-response non-uniformity (PRNU), Which can be captured using a multiplicative factor K (the sensor *fingerprint*). Using the model, the noise residual can be written as:

$$W_{\mathrm{I}} = aIK + \theta$$

Where θ stands for all other random noise components, such as the shot noise or the redout noise, and a is an attenuation factor of the same dimension as K that, in general, depends on the image content.

The maximum likelihood estimator of the PRNU factor K from images $I^{I}, ..., I^{N}$ has the form:

$$\hat{K} = \frac{\sum_{\mathrm{i=1}}^{\mathrm{N}} W_{\mathrm{I}}^{(\mathrm{i})} I^{(\mathrm{i})}}{\sum (I^{(\mathrm{i})})^2}$$

Under the assumption that no geometric transform was applied to image J (cropping, scaling, and digital zooming), the presence of the camera fingerprint represented by an estimate $\hat{K}$ is established through the correlation detector:

$$\rho = corr(W_{\mathrm{J}}, J\hat{K})$$

1.1.2 PSNR

Peak Signal-to-Noise Ratio (PSNR) is a fundamental image quality metric used to assess the fidelity of a reconstructed or compressed image when compared to its original, reference image. PSNR quantifies the level of distortion or noise introduced during the image compression or transmission process and is widely employed in various fields such as image processing, computer vision, and multimedia technology.

Mathematically, PSNR is calculated as the ratio of the peak signal power (which represents the maximum possible pixel value of the image) to the mean squared error (MSE) between the original and the reconstructed images. The formula for PSNR is typically expressed in decibels (dB):

1.1.3 Flatfield and natural images

Flatfield Images Flatfield images, also known as flat-field calibration images or flat frames, are images taken with a uniform source of light, covering the entire field of view of a camera or imaging device. The purpose of flatfield images is to account for variations in sensitivity and other imperfections across the sensor or detector of the imaging device. These variations can result from differences in pixel response, lens vignetting, and dust or artifacts on the sensor's surface. By dividing the actual image by a flatfield image, these variations can be corrected, resulting in more accurate and consistent image data.

Natural Images Natural images refer to photographs or visual representations of real-world scenes or objects taken under natural lighting conditions. These images are not artificially manipulated or generated, but rather capture the visual information as it appears in the environment. Natural images are used in various applications, including photography, scientific imaging, computer vision, and machine learning. The complexity and diversity of natural scenes pose challenges for image analysis algorithms due to factors like lighting variations, occlusions, textures, and object shapes.

In the context of digital image forensics or PRNU analysis, flatfield and natural images can play a role in understanding the characteristics of images and their underlying noise patterns. PRNU, or Photo Response Non-Uniformity, refers to the unique noise pattern introduced by the imaging sensor of a camera. Analyzing PRNU involves differentiating between variations caused by the camera sensor and the natural variations present in a scene, and this analysis can benefit from understanding both flatfield and natural images.

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{\text{Peak Signal Power}}{\text{MSE}}\right)$$

In this equation, a higher PSNR value indicates better image quality, as it signifies a lower amount of distortion or noise present in the reconstructed image compared to the reference image. Conversely, a lower PSNR value suggests higher distortion levels and reduced fidelity.

During the course of my research, I have utilized PSNR as an objective measure to evaluate the performance of various image compression techniques, algorithms, or models that I have implemented. By calculating and analyzing the PSNR values, I am able to assess the effectiveness of different approaches in maintaining image quality while achieving compression or processing goals.

In summary, Peak Signal-to-Noise Ratio (PSNR) is a vital metric in that allows to quantitatively evaluate the quality of reconstructed or processed images. By understanding and applying PSNR calculations, I can make informed decisions about the performance of my implemented methods and their impact on image fidelity.

2. Literature Review and Related Work

The study of image forensics has led to various methods for detecting image manipulations. Notably, PRNU-based techniques have gained attention for their effectiveness in identifying tampered images. PRNU patterns, which capture unique sensor-specific noise in images, offer valuable insights for image authenticity verification. This project builds upon these concepts and explores novel approaches to PRNU-based forgery detection.

3. Description of the Task and Dataset

The primary objective of this project is to develop an accurate PRNU-based forgery detection algorithm. To evaluate our approach, we utilize a diverse dataset comprising real-world images, each containing varying degrees of manipulation. This dataset covers different forms of tampering, allowing us to thoroughly evaluate the performance of our PRNU-based method.

4. Methodology

Our approach employs PRNU patterns to detect potential image manipulations. The input image is processed using convolutional neural networks (CNNs) to extract and analyze PRNU features. PRNU patterns are captured from the reference images and processed using convolutional layers with ReLU non-linearities. The final output, the anonymized image, is obtained through an element-wise sum of the input image and the processed PRNU patterns.

5. Experimental Setting

To evaluate the effectiveness of our PRNU-based approach, we conducted comprehensive experiments using the benchmark dataset. Training parameters, data augmentation techniques, and evaluation metrics are defined in this section. We outline the training strategies and discuss how our method performs across various manipulation scenarios.

6. Results

Our experimental results demonstrate the efficacy of our PRNU-based approach in detecting manipulated images. We provide a detailed analysis of detection accuracy for different types of manipulations, showcasing the strengths and limitations of our method. The results highlight the potential of PRNU-based techniques to contribute significantly to the field of image forensics.

7. Analysis

In this section, we delve into a thorough analysis of the experimental results. We discuss the performance of our approach across different manipulation scenarios and provide insights into the factors influencing its accuracy. Additionally, we compare our method with existing techniques, emphasizing its advantages in terms of detection rates and computational efficiency.

8. Conclusion and Future Work

The project concludes by summarizing the contributions and outcomes of our work. We discuss the significance of our PRNU-based approach in the context of image forensics and its potential to ensure the authenticity and integrity of digital images. As future work, we propose investigating advanced architectures, exploring adversarial training, and extending the method to video forensics.

By harnessing the power of PRNU patterns, this project advances the field of image manipulation detection and enhances our ability to detect tampered images. Our approach holds promise for a range of applications that require the verification and authentication of digital visual content.

Appendix 1

Additional visualizations from exploratory data analysis

Citations