SOURCE CAMERA IDENTIFICATION WITH COMPUTATIONAL INTELLIGENCE TECHNIQUES

SHI CHAO

M.Phil

The Hong Kong Polytechnic University

2016

The Hong Kong Polytechnic University

Department of Electronic and Information Engineering

Source Camera Identification with Computational intelligence Techniques

SHI Chao

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Philosophy

Jan. 2016

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____(Signed)

SHI Chao (Name of student)

Abstract

Identifying the source camera of images is becoming increasingly important nowadays with the popularity of image capturing devices and easy access of image processing software. In this thesis, commonly used camera identification approaches have been reviewed. These methods rely on extracting features derived from different stages of image acquisition process so as to identify the source camera. Example features include lens distortion, pixel defects, CFA interpolation, image processing artifacts and pattern noise called Photo Response Non-uniformity (PRNU) and dark currents.

Among these methods, the pattern noise approach has recently emerged as a powerful tool for digital image forensics. It is because the pattern noise contains device specific features that can be used to uniquely identify each individual camera with a high accuracy while other methods can only identify the model of the source camera. Despite that, the PRNU estimation is sensitive towards scene content and image intensity. The PRNU estimation is poor in areas having low or saturated intensity, or in areas with complicated texture. Though applying distinct weightings to different image regions of image for camera detection may improve the accuracy, it is difficult to determine the appropriate weightings. If the weightings are assigned too aggressively, the detection accuracy may even drop.

In this thesis, the relation between the reliability of PRNU-based camera identification and various features are studied. To solve the scene content problem, two schemes have been proposed in this thesis. In the first scheme, we considered that the intensity and texture features can be used to indicate if the block is severely affected by the

scene content or not. Hence, they are inputted to the neural network so that the network can allocate different weighting to different image blocks. The neural network is trained to produce the weightings that better separate the positive and negative data. The second scheme utilizes the local variance to characterize the severeness of the scene content artifacts. The local variance is then incorporated to the framework of the general matched filter and Peak to Correlation Energy detector to provide an optimal framework for PRNU signal detection. A comparative study with existing start-of-the-art algorithms has been performed. Results show that the proposed scheme achieves the highest True Positive Rate (TPR) for different levels of False Positive Rate (FPR) with different image sizes. The future direction using the PRNU signal for face spoof detection is also discussed with some preliminary experiments.

Publications

1. **Edwin C. Shi**, Frank H.F. Leung, Bonnie N.F. Law, “Differential Evolution with Adaptive Population Size” International Conference on Digital Signal Processing 2014.

2. **Chao SHI**, Ngai-Fong Law, Hung-Fat Leung, Wan-Chi Siu, Weighting Optimization with Neural Network for Photo-Response-Non-Uniformity-based Source Camera Identification, Asia-Pacific Signal and Information Processing Associatie Annual Summit and Conference 2014

3. **Chao SHI**, Ngai-Fong Law, Hung-Fat Leung, Wan-Chi Siu, "A Local Variance Based Approach to Alleviate the Scene Content Interference for Source Camera Identification" submitted to *IEEE Transactions on Information Forensics and Security* pending for review

Acknowledgements

Graduate study is not just about absorbing knowledge from textbooks. The thing that is more important is the training of my mind in identifying and solving problems. It would never be possible without the help and guidance from people around me. I am deeply grateful to all these people.

Firstly, I would like to express my sincere gratitude to my chief supervisor Dr. Bonnie N.F. Law and co-supervisor Prof. W.C. Siu and Dr. Frank H.F. Leung. Dr. Law has always been patient and provides me with a log of insightful ideas in my research. Prof. Siu and Dr. Leung have taught me how to think critically and present my ideas clearly with their decades of experience. Their guidance will be of great benefit in my future career.

Special thanks are given to Mr. Zhang Hongbin, Mr. Huang Junjie, Miss Ginny Wong and other colleagues who have accompanied me during my study. They not only made the university life more enjoyable but also benefited my studies via a lot of interesting discussions on our research works.

Finally, I would like to thank my family for their continuous love, support and encouragement.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Introduction

In recent years, we have witnessed the overwhelming popularity of image capturing devices. According to the estimation from Sony Corporation [1], the production of the digital image sensor has a compound annual growth rate (CAGR) of 11% from year 2009 to 2016 as shown in Figure 1.1. The easy accessibility of digital image capturing devices such as smart phones makes the digital images more frequently presented as pieces of evidence in the court. The observers of a crime scene can record the details of the whole event with their smart phones or digital cameras by simply pressing a button. On the other hand, crimes related to digital images such as the child pornography and infringement of privacy occur more often in the current age of digitalization. However, digital images differ fundamentally from the traditional photography in the way it is created, stored and edited. The release of numerous sophisticated image editing software allows users to modify digital images without much expertise knowledge. This situation undermines the credibility of digital images as evidence in the court of law. Therefore, digital image forensics is becoming increasingly important nowadays. Advancement in the field of digital image forensics is required to restore people's trust in the digital data.

One major objective of digital forensics is camera source identification (CSI). In the cases involving illegal digital photos, the possessor of the device that captures the images can be a suspect. On the other hand, when presented as evidence, the trustworthiness of the digital images will be doubted if the images are not captured by the device they are claimed to be captured with. Another interest of digital image

forensics is forgery detection. It aims to discover malicious processing, examples of which are object removal or adding that often try to change the hidden message in the photo.

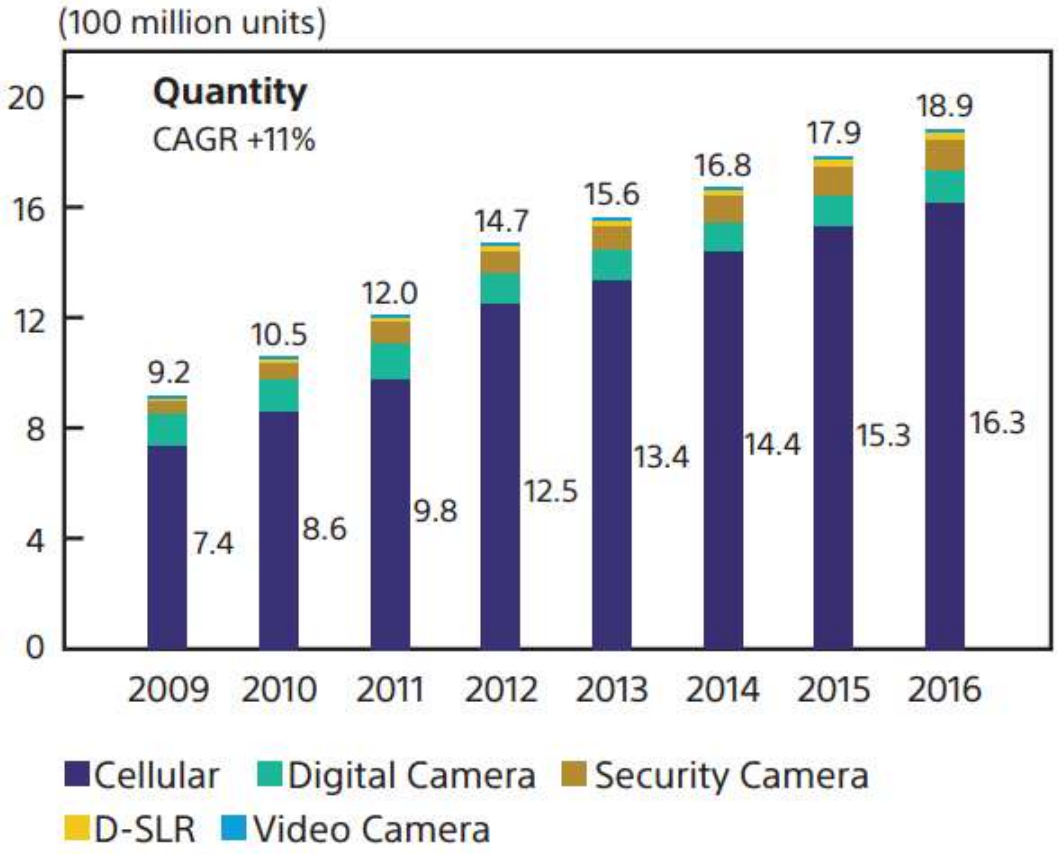


Figure 1.1 Image sensor market as estimated from Sony Annual Report 2012 [1]

There are several possible approaches for solving the camera source identification problem. The simplest one is to examine the information contained within the header of image file. One example is the Exchangeable Image File (EXIF) header that records the details about the digital camera model and the settings that were used to taken that photo. However, recompressing or changing the format of the photos may cause the information to be lost. Moreover, information in the header can be easily

modified which compromises the credibility of this method. Another approach to tackle the source camera identification problem is to embed digital watermark in the images that would carry information about the digital cameras[2][3]. However, adding watermarking feature in cameras incurs an additional cost and most consumer cameras do not have this feature [4]. So this method cannot be used in general. The artifacts introduced by post processing such as JPEG compression, color interpolation and demosaic algorithms can be used to determine the model of the source camera. In[5] [6], a vector of numerical features is extracted from each image, and the feature vectors are used to train a support vector machine (SVM) to classify image sources. However, the accuracy of this method does not meet the requirement of forensics applications in the court. Moreover, this method is not capable of distinguishing images taken by different cameras from the same model since they are using the same post processing algorithms. Pixel defects are also employed to identify the camera source. This method also has limitation because dead pixels may not exist in images or be eliminated by post processing operations. Apart from the aforementioned methods, one of the most popular approaches is based on the imaging sensor pattern noise, where each sensor pattern noise uniquely corresponds to an imaging device and serves as the intrinsic fingerprint. The sensor pattern noise method provides a reliable identifier and it is capable of distinguishing cameras of the same brand and model.

1.2 Motivation

While many different methods for source camera identification have been proposed, recently, a powerful approach based on the photo-response non uniformity (PRNU) emerges. The PRNU can be considered as a kind of intrinsic fingerprint of a specific digital camera sensor. It results from the imperfections and differences in the silicon wafer used to manufacture the imaging sensor [7]: pixels respond differently to the illumination of light. The pixel by pixel difference forms some artifacts which will remain unchanged for all the images taken by that specific camera. This unique artifact can be used as a fingerprint of the camera. Since the PRNU is distinct from one camera to another, it can identify not only different camera models but also different individual cameras of the same model. Another advantage of PRNU is its robustness. The PRNU signal will remain detectable after processing operations like lossy compression, cropping, printing, downsizing etc [8][9][10][11]. Apart from source camera identification, the PRNU based techniques can also be used for image forgery detection[12][13][14][15].

The PRNU signal is a very weak signal. The average signal to noise ratio is around 51db only [16]. In most of the papers, the PRNU signals are extracted by subtracting the denoised image from the original image. The noise residual obtained can be used for PRNU signal detection. With the denoising filter the scene content of the image can be suppressed. Nevertheless, due to the nonideal character of the denoising filter, the noise residual obtained from denosing usually contains a significant level of scene artifacts. This in turn compromise the accuracy of PRNU signal extraction since the scene artifacts lower the signal to noise ratio in the noise residual. If images contain

a lot of textures, have low or saturated intensities, the accuracy of source identification will drop. Several methods have been proposed to suppress the influence of scene content. The approaches which make use of the idea of reliable regions have been proposed in [17][18][19][20][21]. In[19][20], learning based methods are adopted, in which a training phase is required before camera identification. In [19], the author proposed to construct a correlation predictor and allocate the weighting for image blocks according to the predicted correlation values. However, the correlation predictor is constructed by simple polynomial regression. The simplicity of the model may not be capable of revealing the underlying relationship between the reliability of image block and features. Furthermore, this method requires gathering a considerable size of data for training the predictor. In [17], Li made a hypothesis that the stronger a signal component is, the more likely that it is associated with strong scene details, and thus the less trustworthy the component should be. Based on this hypothesis, Li proposed five different models to shrink the noise residuals with large magnitude. However, there is no theory showing that the five models give the optimal weightings.

There are some other methods trying to reduce the scene content artifact from different aspects. Kang et al. [22] proposed to suppress the scene content artifacts by using the phase component of the pattern noise. The noise residuals are normalized by its magnitude in the Fourier domain to eliminate the scene content artifact before they are used to estimate the reference fingerprint. Some other studies focus on the influence of denoising filters [23] [24]. Since no denoising filter can perfectly separate the scene content from noise, a good model which estimates the reliability and allocates

weightings for different regions are always necessary. From pervious analysis, there are two possible directions to further improve the PRNU-based detection accuracy. The first one is to use more powerful machine learning tools to establish the relation between image local features and the regional reliability. Therefore, we propose to use training an Artificial Neural Network to predict the reliability of image regions. The Artificial Neural Network has been proved to be a universal approximator which can represent any continuous functions with nonlinear activation function[25]. The other direction is to find an efficient scheme to allocate the weightings such that no training nor extra effort on implementing the learning algorithm are needed. To achieve this goal, a method which regards each pixel as a random variable and estimates its distribution is proposed. The General Matched Filter which is considered as the optimal detector in that it gives the best False Rejection Rate for any False Acceptance Rate.

1.3 Organization of the thesis

The remainder of the thesis is organized as follows. In Chapter 2, the image capture model will be firstly introduced. Then several common methods for camera identification will be presented. After that, details of using pattern noise for source camera identification will be described. In Chapter 3, the proposed method of using the neural network for weighting optimization will be discussed in details. Experimental results will be given and analyzed. In Chapter 4, the relationship between the reliability of each pixel and some local features are analyzed. .The proposed enhancement method utilizing the local variance of the noise residual and the General Matched Filter are described. Finally, the conclusion and the direction of future work will be described in Chapter 5.

Chapter 2 Literature Review

While different features from image capturing systems have been utilized to identify the image origin, there are several advantages which make the PRNU-based techniques the most powerful approach for source camera identification. First of all, the PRNU-based method is capable of identifying the individual source device while some other methods can only identify the source camera model. Secondly, the PRNU-based method is robust to various image processing operations e.g. cropping, JPEG compression, printing etc. Thirdly, it is a passive technique which does not require any change of digital camera and the PRNU signal exists in all the consumer cameras regardless of whether the image sensors are CMOS or CCD. The identification accuracy is higher as compared with other methods.

In this chapter, a simplified image capture model of digital camera will be described in Section 2.1. Then some of the existing source camera identification methods based on features from different phases of the digital camera system are described in Section 2.2. Specifically, the techniques based on lens distortions, pixel defects, CFA interpolation artifacts, composite image processing artifacts are reviewed. Then the general framework of the PRNU based source camera identification methods are introduced in Section 2.3. The various ways to enhance the PRNU based camera identification method are reviewed in Section 2.4. Finally, Section 2.5 gives a summary of this chapter.

2.1 Image capture model

The general structure of a digital camera is shown is Figure. 2.1. A typical camera consists of lens system, filters, color filter array (CFA), image sensor and digital image processor (DIP) [26]. The main use of lens is to focus the incident light onto the image sensor such that the image captured would be in focus. The lens system is also used to reduce the effects of chromatic aberration and spherical aberration. Chromatic aberration arises when the lights with different wavelengths cannot converge to the same position while spherical aberration arises when lights pass through the periphery of a spherical lens and converge to a point closer to the lens than the lights passing through the center of the spherical length. These aberrations can be minimized with special combinations of convex and concave lenses. The lens system may also include auto-exposure control, auto-focus control and stabilization unit. A set of filters are used to enhance the quality of image generated. They help to filter out the invisible part of the spectrum such as infrared and ultraviolet which ensures that sensor would only respond to the light that can be visualized by human visual system. The filters also help to reduce aliasing which happens when spacing between pixels cannot support the finer spatial frequency of the scene. As the sensors only record brightness of the light, color filter array needs to be used in front of the sensor to capture different color components in a single image sensor. The Green – Red – Green – Blue (GRGB) Bayer pattern of CFA is most commonly used in digital cameras. The image sensor is a matrix of photondiode elements which is also called pixels. When the sensor is exposed to light, each element in the sensor will generate an analog signal which is proportional to the intensity of light. The analog signal will then be

converted to digital signal. After the sensor output is obtained, the sensor output signal undergoes various in-camera processing such as CFA interpolation, white balancing, noise reduction etc.

In the whole image capture process, many features have been utilized to perform source camera identification in literature. In the next part, methods using the lens aberration, pixel defects, CFA interpolation and sensor pattern noise for camera identification will be introduced.

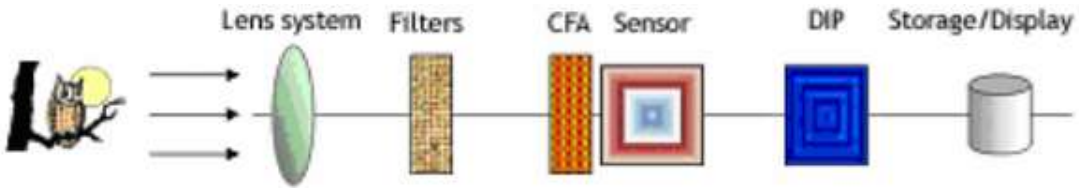


Figure 2.1 Digital camera output model from [17]

2.2 Introduction to camera identification approaches

2.2.1 Using lens aberration

Lens of camera can produce aberrations in images because of the design and manufacturing process. The lens radial distortion is the most severe component of the aberrations. It makes straight lines in the object space rendered as curved lines on the camera sensor. The radial distortion arises because the lens has different focal length and magnifications in different areas. The transverse magnification is a function of the off-axis distance instead of a constant. When the magnification of scene decreases with off-axis image distance, it is called barrel distortion. If the magnification of scene increases with off-axis image distance, it is named as the pincushion distortion. In [27], the radial distortion is modeled by the following equation:

$$r_u = r_d + k_1 r_d^3 + k_2 r_d^5 \qquad (2.1)$$

where r_u and r_d are the undistorted radius and distorted radius respectively, and k_1 and k_2 are the first and second order distortion parameters respectively. The radius is the radial distance of a point (x,y) from the center of the distortion which is assumed to be the center of the image. The distortion parameter k_1 and k_2 are estimated with an iterative process. It firstly extracts the distorted line segments and measures the distortion error between the distorted line segments and their corresponding straight lines. Then the distortion parameters k_1 and k_2, are tuned to minimize the distortion error. The process will be repeated until the relative change of distortion error is less than a predefined threshold. In [27], the two estimated parameters are used as input features of a classifier. They are also used together with other 34 features proposed

in [5] to classify images according to their sources. The accuracy obtained with the two methods in [27] and [5] are 91.54% and 91.39% respectively.

However, the lens aberration-based method requires that there must be straight lines in the image to measure the distortion. Otherwise the two distortion parameters cannot be estimated. Moreover, cameras from the same manufacturer or same model may have similar distortion property which would lower the accuracy of identification.

2.2.2 Using pixel defects

Since the sensors contain large number of elements, pixel defects will arise during the manufacturing process of the sensor. The pixel defects can be classified into point defects, hot point defects, dead pixels, pixel traps and cluster defects. Geradts et al [28] examine the defects of CCD pixels and use them to match target images to the source camera. They tested on 12 different cameras of the brand Trust and found that there were at least 5 pixel defects in each camera. However, the pixel defects are often compensated by the electronics and image processing operations in the camera. Therefore the pixel defects are not clearly visible in the final image. The visibility of pixel defects also depends on the temperature of the environment in which the images was taken. Furthermore, the pixel defects can be hardly seen in the cameras with high-end CCD. It is possible that there is no pixel defect in the camera sensor. Therefore, the method cannot be directly applied for all digital cameras.

2.2.3 Using CFA interpolation

As we discussed before, the camera sensor can only capture the intensity of the light. Therefore to obtain a color image, a mask named color filter array (CFA) is added in front of the sensor. With this approach, each sensor element can only sense one band of the wavelength. Therefore, the raw image collected from the sensor is a mosaic of different colors which are typically red, green and blue. Figure 2.2 shows a CFA pattern which is commonly used in cameras.

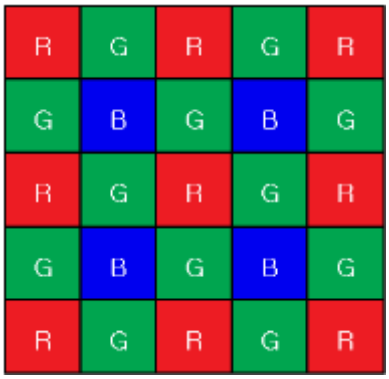


Figure 2.2 CFA pattern

Since each pixel only carries the information about one color band, the missing colors need to be interpolated for each pixel to generate a color image. The interpolation is usually done by applying a weighting matrix (kernel) to the neighborhood around a missing value. There are many different interpolation algorithms and different manufacturers may use different ones, i.e. they use kernels with different sizes and shapes. In [29], Bayram et al. used an iterative Expectation Maximization Algorithm (EM) to obtain two sets of features which are the interpolation coefficients from the images and the peak location and magnitudes in the frequency spectrum of the probability maps. Using these features, two cameras can be classified with an accuracy of 95.71%. However, the accuracy droped to 83.33%, when the experiment

was carried on three cameras. Large scale testing involving more cameras was required to test the effectiveness of the method. Again, this method may not be capable of identify cameras from the same model or manufacturer, because we can expect a similar interpolation algorithm will be used in such situation.

2.2.4 Using Composite features from Image

Since different manufacturers usually use different techniques on CFA configuration, demosaicing algorithm and color processing/ transformation, the image taken by certain camera may exhibit certain traits and patterns regardless of the original image content. In [5], Kharrazi extracted 34 features from images to classify a camera model. These features can be categorized into 3 groups which are color features, image quality metrics and wavelet domain statistics. Features are extracted from two cameras to train and test a SVM classifier. The result was as high as 98.73% for uncompressed image and 93.42% for compressed image with a quality factor of 75. The result drops to 88% when five cameras are involved in the experiments. However, this method still cannot identify cameras of the same model. In [30], Tsai reported that using image features, the accuracy was low (66.7%) when the cameras being classified have similar sensors. Moreover, this method requires all cameras to take the same content at similar resolution which is hard to achieve in practice.

2.3 Sensor pattern noise based approach

2.3.1 Sensor noise and the sensor output model

There are many sources of noises which enter into different stages of the image capturing process described in the previous section. Even when the cameras are exposed to an evenly illuminated scene, the output intensity slightly varies from one pixel to another. This difference comes from two sources. The first one is called the shot noise or photonic noise [31, 32], which is a random component. The second source is the pattern noise, which is a deterministic component that retains similar characteristics when images of different scenes are taken. Thus the pattern noise can be used for camera identification.

The pattern noise mainly consists of fixed pattern noise (FPN) and photo-response nonuniformity noise(PRNU). The FPN is caused by dark current. It is the noise pattern generated by the sensor when it is not exposed to light. The PRNU mainly arise from the inhomogenity of silicon wafers and imperfections during the sensor manufacturing process which is called pixel nonuniformity (PNU). The PRNU is unique for each individual sensor and thus can be used as a fingerprint for that camera sensor.

In [31], FPN has been used as fingerprint for camera identification. However, using FPN as fingerprint has limitation that FPN can only be extracted from dark frames. Furthermore, the FPN is usually suppressed by automatically subtracting a dark frame from the image they take in middle-to-high-end cameras. Therefore, FPN cannot be used as a reliable fingerprint for camera identification. The use of PRNU as camera fingerprint was firstly proposed by Lukas [33]. The PRNU is the dominant part of

pattern noise. It is much stronger than FPN and it can better survive from various image processing operations. Therefore, PRNU is more reliable for camera identification tasks.

In [34], a general image output model was given and it will be introduced here. Denote Y(i, j) to be the intensity of the incident light at pixel location (i, j). The sensor output signal with PRNU added is,

$$(1+K(i,j))Y(i,j) \quad (2.2)$$

where $K(i,j)$ is a constant which represents the PRNU characteristic of each pixel. The value of $K(i,j)$ varies from one pixel to another. It has a mean of 0 and variance σ_K^2 over all the pixels. Let $N_{DC}(i,j)$ and $N_s(i,j)$ be the fixed pattern noise and shot noise respectively. The sensor output signal now becomes,

$$(1+K(i,j))Y(i,j)+N_{DC}(i,j)+N_s(i,j) \qquad (2.3)$$

The charge at each pixel will be transferred to the output amplifier. The amplifier transforms the charge into a measurable voltage for readout. This process will incur additional noise with zero mean to the signal. Let the gain of the amplifier be A and the amplifier noise be N_R, the sensor output signal will become,

$$((1+K(i,j))Y(i,j)+N_{DC}(i,j)+N_s(i,j)+N_R)A \quad (2.4)$$

The signal is subsequently quantized by the analog to digital converter. With the quantization noise $Q(i,j)$, the final output signal will become,

$$((1+K(i,j))Y(i,j)+N_{DC}(i,j)+N_s(i,j)+N_R)A+Q(i,j) \qquad (2.5)$$

In [35], Chen et al. used a simplified camera output model which only captures the most relevant parts to the camera identification tasks. In this model, only white balance and gamma correction are considered. Thus the camera output becomes,

$$((1+K(i,j))Y(i,j)+N_{DC}(i,j)+N_s(i,j)+N_R(i,j))^{\gamma}A^{\gamma}+Q(i,j) \quad (2.6)$$

where γ denotes the gamma correction factor. This model is further simplified to be,

$$((1+K(i,j))Y(i,j)+\Lambda(i,j))^{\gamma}A^{\gamma}+Q(i,j) \qquad (2.7)$$

where $\Lambda(i,j)$ is the combination of the camera noise except PRNU. Dropping the pixel indices, the model becomes,

$$((1+K)Y+\Lambda)A^{\gamma}+Q \qquad (2.8)$$

Factor out *I* and use Taylor expansion approximation to keep the first two terms, the model is simplified to be

$$\begin{aligned} I &= (AY)^{\gamma}((1+K)Y+\frac{\Lambda}{Y})^{\gamma}+Q \\ &= (AY)^{\gamma}((1+\gamma K)Y+\frac{\gamma\Lambda}{Y})+Q \qquad (2.9) \\ &= I^{(0)}+I^{(0)}K'+Q' \end{aligned}$$

where $I^{(0)}=(AY)^{\gamma}$, $K'=\gamma K$ and $Q'=(AY)^{\gamma}(\frac{r\Lambda}{Y})+Q$. Hence, the output intensity contains three terms. They are the input intensity, a term containing the multiplication of the input intensity and the PRNU, as well as a sum of random noises.

2.3.2 Overview of camera identification procedure with pattern noise

The PRNU based source camera identification was firstly proposed in [7]. The general procedure of camera identification is shown in Figure 2.3. We have two sets of images which are set A, B. Set A are the test images which may be obtained from camera C or some other cameras, Set B is used to estimate the reference PRNU of camera C. Firstly, we need to extract the PRNU feature from all the images. A

denoising filter is applied to all the images and the PRNU feature will be the difference between the original image and the denoised image. In the original work [7], the reference PRNU was estimated with the noise residual obtained from set B using simple average. Later in [12], a maximum likelyhood estimator was proposed to estimate the reference PRNU. Then we can calculate the correlation between the noise residual of the test images and the reference PRNU. The correlation value will measure the similarity between the test images and the reference PRNU. A threshold can be determined by Neyman Pearson threshold decision approach to determine whether the image is taken by camera C or not. The detailed procedure will be introduced in the following sections.

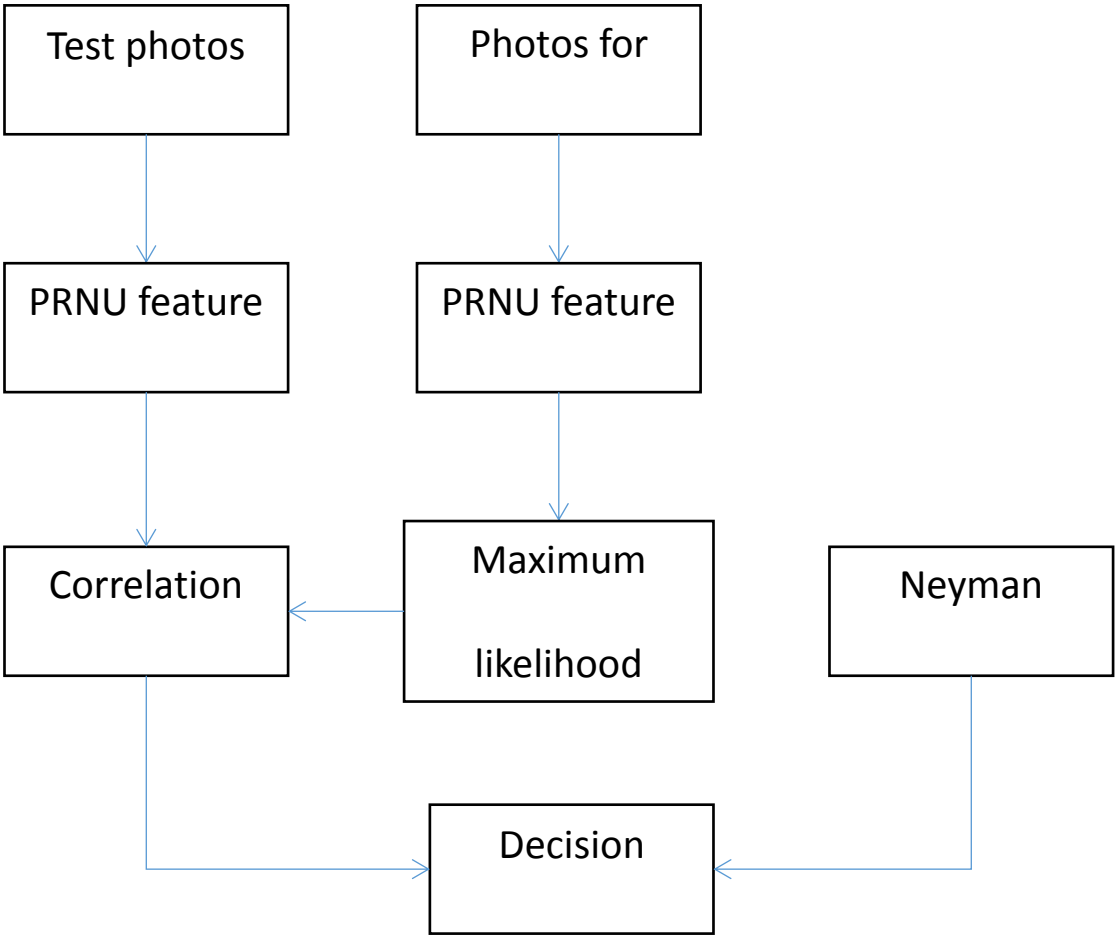


Figure 2.3 Camera identification procedures

2.3.3 Pattern noise estimation from images

The first step of camera identification is to estimate the reference PRNU feature $\hat{\mathbf{K}}$. To estimate $\hat{\mathbf{K}}$, a set of images from the camera is needed. We firstly suppress the image content by applying a denoising filter to each image and obtain the noise residual **W** by subtracting the denoised image from the original image **I**,

$$\begin{aligned}\mathbf{W} &= \mathbf{I} - \hat{\mathbf{I}}^{(0)} \\ &= \mathbf{I}\hat{\mathbf{K}} + \mathbf{I}^{(0)} - \hat{\mathbf{I}}^{(0)} + (\hat{\mathbf{I}}^{(0)} - \mathbf{I})\hat{\mathbf{K}} + \mathbf{Q} \\ &= \mathbf{I}\hat{\mathbf{K}} + \boldsymbol{\varepsilon}\end{aligned} \quad (2.10)$$

where ε is the combination of $\mathbf{Q}$ and two additional terms introduced by the denoising filter. $\boldsymbol{\varepsilon}$ is modeled as white Gaussian noise with variance σ^2. Then the noise residue can be used to estimate the PRNU. In [7], the reference PRNU is calculated by,

$$\hat{\mathbf{K}} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{W}_k \quad (2.11)$$

where N is the total number of images used to obtain the reference PRNU. In [35], a maximum likelihood estimator was derived to estimate $\hat{\mathbf{K}}$. For N images $k=1,\ldots,N$, equation (2.10) can be changed to,

$$\frac{\mathbf{W}_k}{\mathbf{I}_k} = \hat{\mathbf{K}} + \frac{\boldsymbol{\varepsilon}}{\mathbf{I}_k} \quad (2.12)$$

The log-likelihood of observing $\frac{\mathbf{W}_k}{\mathbf{I}_k}$ given $\hat{\mathbf{K}}$ is,

$$L(\hat{\mathbf{K}}) = -\frac{N}{2}\sum_{k=1}^{N}\log\left(\frac{2\pi\sigma^2}{\mathbf{I}_k^{\,2}}\right) - \sum_{k=1}^{N}\frac{\left(\frac{\mathbf{W}_k}{\mathbf{I}_k} - \hat{\mathbf{K}}\right)}{\frac{2\sigma^2}{\mathbf{I}_k^{\,2}}} \quad (2.13)$$

Taking partial derivatives, the estimated PRNU can be calculated as,

$$\frac{\partial L(\hat{\mathbf{K}})}{\partial \hat{\mathbf{K}}} = \frac{(\frac{\mathbf{W}_k}{\mathbf{I}_k} - \hat{\mathbf{K}})}{\frac{2\sigma^2}{\mathbf{I}_k^{\,2}}} = 0 \quad (2.14)$$

$$\Rightarrow \hat{\mathbf{K}} = \frac{\sum_{k=1}^{N} \mathbf{W}_k \mathbf{I}_k}{\sum_{k=1}^{N} (\mathbf{I}_k)^2} \quad (2.15)$$

To obtain a reliable PRNU $\hat{\mathbf{K}}$, the number of images *N* should be large enough. It is suggested that *N* should be larger than 50. Also, $\hat{\mathbf{K}}$ can be better estimated from bright images with smooth scene like the blue sky. However, the image cannot be saturated because saturated image carries no information about PRNU.

2.3.4 PRNU Preprocessing

It was found that the estimated PRNU $\hat{\mathbf{K}}$ contains systematic artifacts which present in every image[12]. The main source of the artifacts is color interpolation, on sensor signal transfer [36], and sensor design [37]. These kinds of artifacts are common among the cameras of the same manufacturer or sharing the similar sensor design. Due to the systematic artifacts, the estimated PRNU $\hat{\mathbf{K}}$ from different cameras might be weakly correlated. The weak correlation can increase the probability of false alarm. Thus, the systematic artifacts should be removed before calculating the correlation values. The color interpolation artifacts and the row-wise and column-wise operations can produce periodic patterns along the rows and columns of an image. Such patterns can be removed by a two step operation: 1) subtract the column average from each column of the estimated PRNU and 2) subtract the row average from each row of the output of step one. This operation is denoted as $\mathrm{ZM}(\hat{\mathbf{K}})$. Some other artifacts in the estimated PRNU such as the sensor design artifact appear to have

certain structure in the Fourier Domain. Therefore, a Wiener filter is applied in the Fourier Domain to remove such artifacts. The overall preprocessing operation can be expressed as,

$$\mathrm{WF}(\mathrm{ZM}(\hat{\mathbf{K}})) = \mathrm{IDFT}(\mathrm{DFT}(\mathrm{ZM}(\hat{\mathbf{K}})) - \mathrm{W}(\mathrm{DFT}(\mathrm{ZM}(\hat{\mathbf{K}})))) \qquad (2.16)$$

where W is a 3 by 3 Wiener filter with the variance obtained as the sample variance of the magnitude of the Fourier coefficient of $\mathrm{ZM}(\hat{\mathbf{K}})$. With preprocessing operation, the resultant PRNU are more reliable for the reason that PRNU from different devices have a correlation close to 0.

2.3.4 Correlation detector

Lukas [33] proposed to use the cross correlation measure to test whether a photo is taken by certain camera, and while the form of correlation slightly varies, it is still commonly used in PRNU based camera identification. In [35], the author calculates the correlation between the noise residue $\mathbf{W}_p = \mathbf{I}_p\mathbf{K}'_p + \boldsymbol{\varepsilon}_p$ obtained from equation (2.10) and $\mathbf{I}_p\widehat{\mathbf{K}}_C$ as,

$$\rho_{\mathrm{C}}(p) = corr(\mathbf{W}_p, \widehat{\mathbf{K}}_{\mathrm{C}}) = \frac{(\mathbf{W}_p - \overline{\mathbf{W}_p}) \cdot (\mathbf{I}_p\widehat{\mathbf{K}}_{\mathrm{C}} - \overline{\widehat{\mathbf{K}}_{\mathrm{C}}})}{\left\|\mathbf{W}_p - \overline{\mathbf{W}_p}\right\| \left\|\mathbf{I}_p\widehat{\mathbf{K}}_{\mathrm{C}} - \overline{\widehat{\mathbf{K}}_{\mathrm{C}}}\right\|} \qquad (2.17)$$

where $\widehat{\mathbf{K}}_{\mathrm{C}}$ is the PRNU estimated from camera C, the bar above the symbol denotes its mean value, $\cdot$ denotes dot product and ||.|| denotes the L_2 norm. The correlation value measures the similarity between the noise residue extracted from the testing image and the camera reference pattern. A high correlation value indicates a high probability that the testing image is taken by the same camera as the reference PRNU. Alternatively, the peak to correlation energy (PCE) [38][39] can also be used to

calculate the detection statistics. The PCE has been reported to be more suitable for camera fingerprint detection because the presence of hidden periodic signal will lower PCE and reduce the possibility of false alarm [39]. The PCE can be expressed as,

$$PCE(\mathbf{x},\mathbf{y}) = \frac{\rho(\mathbf{s}_{peak} = 0, \mathbf{x}, \mathbf{y})^2}{\frac{1}{MN - |A|}\sum_{s \notin A} \rho(s, \mathbf{x}, \mathbf{y})^2} \quad (2.18)$$

where $\rho(s, x, y)$ is the dot product between $x - \bar{x}$ and $y(s) - \bar{y}$, $y(s)$ is obtained by circularly shift y by a two dimension vector s, A is a small neighbor around the peak and M, N are respectively the width and height of the image.

A threshold is needed to make the decision whether the obtained PCE value is large enough that the noise residue of the testing image is considered to be identical to the reference PRNU. The threshold can be decided by the Neyman Pearson theorem which is described in the next section.

2.3.5 Neyman Pearson theorem for threshold decision

Given two sets of images, one set is taken by camera C, and the other set is taken by some other cameras. Then the problem can be formulated as a binary hypothesis test,

$$\begin{cases} H_0 : \mathbf{W} = \boldsymbol{\varepsilon} \\ H_1 : \mathbf{W} = \mathbf{I}_p \mathbf{K}_p + \boldsymbol{\varepsilon} \end{cases} \quad (2.19)$$

The hypothesis H_0 should be $W = \mathbf{I}_p \mathbf{K}_0 + \boldsymbol{\varepsilon}$, where $\mathbf{K}_0$ is the PRNU of other cameras, however, since the combined noise term ε is much larger than the $\mathbf{I}_p \mathbf{K}_0$ term, $\mathbf{I}_p \mathbf{K}_0$ is omitted in the equation. The distribution of $p(\rho \mid H_0)$ and $p(\rho \mid H_1)$ can be obtained with the correlation detector introduce in Section 2.3.4. The probability density

function of $p(\rho \mid H_0)$ and $p(\rho \mid H_1)$ can be modeled with a Generalized Gaussian model using the method of moment[40]. With the probability density function $p(\rho \mid H_0)$ we can then determine the threshold t_0 by setting the FAR toleration at a particular small value, says 10^{-3}. The threshold t_0 can be used to make the decision whether the photo is taken by the camera *C*. In addition, with the FAR set, together with probability density function of $p(\rho \mid H_1)$, FRR can be used to evaluate the system performance.

2.4 Various Issues related to PRNU-based camera identification

2.4.1 Scene content effect

As introduced in the previous section, a denoising filter is used to extract the PRNU feature from images. The most commonly used filter is the wavelet denoising filter which is described in Appendix A of [7]. However, the scene content can severely contaminate the extracted PRNU and make the PRNU signal very weak. This problem is illustrated in Figure 2.4. Figure 2.4 (a) shows a clean PRNU which is obtained by averaging 50 images of blue sky. Figure 2.4 (b) is an image of natural scene taken by the same camera. Figure 2.4 (c) is the PRNU extracted from Figure 2.4 (b). It can be observed that Figure 2.4 (c) contains a lot of scene details and the PRNU signal can be hardly seen as compared with Figure 2.4 (a).

(a) (b) (c)

Figure 2.4 The scene content problem. (a) The clean PRNU, (b) a natural scene and (c) the noise residue of (b).

In the process of camera identification, the reference PRNU can usually be obtained

by averaging a number of PRNU features extracted from smooth images. However, for the testing image to be identified, the noise residue is extracted from one image only. It is possible that the testing image contains complicated texture. In turn, the noise residue extracted will have a lot of scene details such that the accuracy of identification becomes low. This problem can be approached from different aspects. The first possible approach is on the denoising filter used to extract the PRNU and the other approach is on selecting more reliable regions from image for feature extraction. These two approaches will be introduced in the next two sections.

2.4.2 Using different denoising filters

Since the PRNU signal can be corrupted by the scene content, one approach to solve the scene content problem to improve the denoising filter used such that the scene content of images can be better suppressed. In the original publication [7], the author has tested a few different denoising filters and found that the wavelet-based filters [41] performs the best among all the filters tested.

There are more studies on the denoising filters. In [23] the author has studied two denoising filters operating in the wavelet domain and considered two different noise models. The first one was the original wavelet-based filter as suggested in [7], the other is a minimum mean square error (MMSE) filter operating in the undecimated wavelet domain [42]. The author made the assumption that the camera noise is dependent on the sensed signal and the MMSE filter will utilize this assumption. In contrast, the filter used in [7] used a signal-independent noise model. These two filters and a low-pass filter were tested with a set of 10 digital cameras. With the

experimental results, they concluded that when the noise model matched the actual situation, the filter gains better performance if the parameters of the filter are accurately estimated.

In [24], the author has used the BM3D filter based on the non-local multipoint approach for image forgery detection. The BM3D takes the advantage of both the context and the spatial correlation. The BM3D and the original wavelet filter were tested with two digital cameras. The experimental results showed that BM3D significantly improves the performance compared with the filter used in [7].

[43] simplified the filtering strategy by using a combination of adaptive wiener filtering and median fitering applied in the spatial domain. Three cameras from different manufacturers are used to test the performance of the filter. It is found that the distribution for the correlation of matching set and non-matching set were more widely separated as compared with the wavelet filter. Thus the author concludes that the proposed filter outperformed the wavelet filter adopted in [7].

2.4.3 Ideas of image regions reliability

Due to the fact that an ideal separation between the noise and scene content is not achievable, it is always necessary to evaluate the reliability of different regions. The other approach to suppress the influence of scene content is to analyze the content of image and make more use of the reliable areas which are smooth and bright but not saturated.

2.4.3.1 Correlation Predictor

In [12], a correlation predictor has been constructed with a polynomial regression

model. Test images were firstly divided into small blocks. The correlation predictor predicted the correlation value for each block with block intensity, texture and flattening as the the input. A high predicted correlation value implies that the block has strong PRNU signal and less undesired artifacts. The general matched filter was used to measure the similarity between the test image and reference PRNU.

Specifically, the source camera identification problem is formulated as,

$$\begin{cases} H_0 : \mathbf{W} = c\Xi \\ H_1 : \mathbf{W} = a\mathbf{TIK} + c\Xi \end{cases} \qquad (2.20)$$

where matrix **T** is a pixel-wise multiplicative attenuation factor, *a* and *c* are unknown multiplicative factor that are the same for all the blocks. For block *b*, the shaping factor **T** and σ_{Ξ}^2, are considered as constants and denote as $\mathbf{T}_b$ and σ_b^2 respectively. The optimal detector will be the normalized general matched filter which can be expressed as,

$$\rho = \frac{\sum_{b=1}^{M} \frac{\hat{T}_b}{\hat{\sigma}_b^2} \left(\mathbf{X}_b \cdot \mathbf{W}_b \right)}{\sqrt{\sum_{b=1}^{M} \frac{\hat{T}_b}{\hat{\sigma}_b^2} \left\| \mathbf{X}_b \right\|^2} \sqrt{\sum_{b=1}^{M} \frac{1}{\hat{\sigma}_b^2} \left\| \mathbf{W}_b \right\|^2}} \qquad (2.21)$$

where $\cdot$ represents the dot products, $\mathbf{W}_b$ and $\mathbf{X}_b$ are the noise residual and reference PRNU of block *b* respectively.. Under H_1 the shaping factor **T** and the variance σ_{Ξ}^2 can be estimated as,

$$\hat{\sigma}_b^2 = \frac{1 - \rho_b^2}{c^2 \left| B_b \right|} \left\| \mathbf{W}_b \right\|^2 \qquad (2.22)$$

$$\hat{T}_b^2 = \frac{\rho_b}{a \left\| \mathbf{X}_b \right\|^2} \left\| \mathbf{W}_b \right\|^2 \qquad (2.23)$$

where ρ_b is the normalized correlation between $\mathbf{X}_b$ and $\mathbf{W}_b$.and $|B_b|$ is the size of block *b*. Since $c^2|B_b|$ and a are the same for all the blocks, their value will not affect the identification result and can be skipped. The only value that is unknown is ρ_b. To address this problem, a correlation predictor is constructed to estimate ρ_b.

Three features that are considered highly influential to the correlation value were selected as the predictor input. They are the image intensity, the texture and signal flating.

The correlation value will be higher in the areas with high intensity due to the fact that the PRNU signal $\mathbf{IK}$ is proportional to the signal intensity of the image. However, the PRNU signal will not present in the saturated regions. The correlation value will be attenuated if the intensity of image is close to the maximum value. Therefore the intensity feature is defined as,

$$f_{\mathrm{I}} = \frac{1}{|B_b|}\mathrm{att}(\mathbf{I}[i, j])$$
$$\mathrm{att}(\mathbf{I}[i]) = \begin{cases} e^{-(\mathbf{I}[i]-I_{\mathrm{crit}})^2/\tau}, \mathbf{I}[i, j] > I_{\mathrm{crit}} \\ \mathbf{I}[i, j]/I_{\mathrm{crit}}, \mathbf{I}[i, j] \le I_{\mathrm{crit}} \end{cases} \quad (2.24)$$

where the constant parameters τ and I_{crit} are determined empirically.

The correlation will be low in the high textured regions because the scene artifacts contaminate the PRNU signal and part of the PRNU will be removed by the denosing filter yielding a smaller **T**. Therefore, the texture feature is defined as,

$$f_{\mathrm{T}} = \frac{1}{|B_b|}\sum_{i\in B_b}\frac{1}{1+\mathrm{var}_5(\mathbf{F}[i], j)} \quad (2.25)$$

where **F** is the high pass filtered image generated from the high frequency band of the wavelet transform and $\mathrm{var}_5(\mathbf{F}(i,j))$ is the variance of **F** in the 5 × 5 neighboor of *i*.

The image processing operations like JPEG compression tend to flatten the image due to its low pass filtering nature. As a result, the predictor will overestimate the correlation if the signal is too flat. For this reason, a signal flating feature is added which is defined as,

$$f_{\mathrm{S}} = \frac{1}{|B_b|} \sum_{i \in B_b} \left| \{(i, j) \in B_b \mid \sigma_I[i] < cI[i, j]\} \right| \quad (2.26)$$

where c is a constant depending on the variance of the PRNU signal **K** and $\sigma_I^2[i, j]$ is the variance within a 5 × 5 neighboor of i, j.

Since the correlation value strongly depends on the mutual effects of texture and intensity, the texture-intensity feature is also included,

$$f_{\mathrm{TI}} = \frac{1}{|B_b|} \sum_{i \in B_b} \frac{\mathrm{att}(\mathbf{I}[i])}{1 + \mathrm{var}_5(\mathbf{F}[i])} \quad (2.27)$$

With these features f_{I}, f_{T}, f_{S} and f_{TI}, a simple multivariate polynomial regression is used to predict the correlation values. The predictor is formulated as,

$$\rho[k] = \theta_0 + \theta_1 f_{\mathrm{I}}[k] + \theta_2 f_{\mathrm{T}}[k] + \theta_3 f_{\mathrm{S}}[k] + \theta_4 f_{\mathrm{TI}}[k] + \theta_5 f_{\mathrm{I}}[k] f_{\mathrm{I}}[k] + \theta_6 f_{\mathrm{I}}[k] f_{\mathrm{T}}[k] + \cdots + \psi[k] \quad (2.28)$$

where $\psi[k]$ is the model noise and $\boldsymbol{\theta} = [\theta_0, \theta_1, ..., \theta_{14}]$ are parameters to be estimated. With a large training set, the parameters can be found using the least square estimator. Then the predicted correlation value can be used to calculate the weighting of each block accordingly.

2.4.3.2 Li's PRNU Enhancement Model

In [17], Li made the assumption that the stronger a signal component in noise residue is, the more likely that it is associated with strong scene details, and thus the less

trustworthy the component should be. Thus, he proposed five models to attenuate the pattern noise with large magnitude. Thereby the scene content artifacts are suppressed as well. The five models will assign less weighting to the strong components in the noise residue. Figure 2.5 shows the shrinkage function for the five models where the x-coordinate is the magnitude of the original noise signal and the y-coordinate is the magnitude of the attenuated attenuated noise signal. Among these models, model 1 and 2 are linear transformation and model 3-5 are nonlinear exponential transformations. The proposed five models were tested with six digital cameras. The experimental results showed that all of the five models can improve the performance of camera identification. Furthermore, model 3-5 are more preferable because they are more stable with the setting of the parameters in shrinkage function.

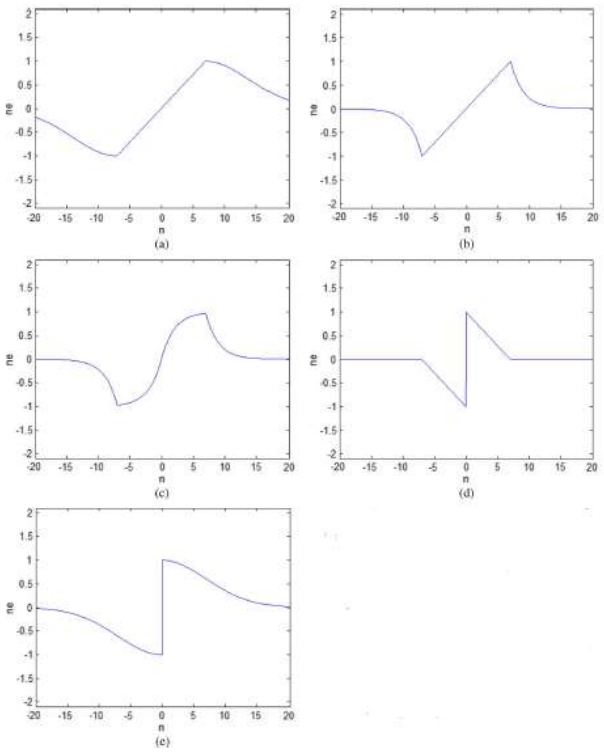


Figure 2.5 Shrinkage function for five models

2.4.3.3 Phase Approach

In [22], it has been found that the large magnitude values of the noise residual in the Fourier domain are usually caused by the texture artifact in the image. Therefore, to reduce the scene content effect, the phase component of the pattern noise was to be used to perform camera identification tasks. Before estimating the reference PRNU, the noise residual $\mathbf{W}_k$ will be transform to Fourier domain using Discrete Fourier Transform, i.e,

$$\mathbf{w}_k = \mathrm{DFT}(\mathbf{W}_k) \quad (2.29)$$

Then the signal will be normalized by its magnitude as,

$$\mathbf{w}_{\phi k} = \frac{\mathbf{w}_k}{|\mathbf{w}_k|} \quad (2.30)$$

where $|\mathbf{w}_k|$ is themagnitude of $\mathbf{w}_k$. The reference PRNU is then computed by averaging the phase component $\mathbf{w}_{\phi k}$ in the frequency domain which is then inversely transformed into spatial domain,

$$\hat{\mathbf{K}}_{\text{phase}} = \mathrm{real}(\mathrm{IDFT}(\frac{\sum_{k=1}^{N} \mathbf{w}_{\phi k}}{N})) \quad (2.31)$$

where N is the number of image used to estimate the phase PRNU and real(.) denotes the real part of the input. In this way, the scene content artifacts are suppressed. The estimated phase PRNU can be used to replace the conventional reference PRNU for computing the detection statistics with PCE or cross correlation.

In [44], Hu et al. proposed to only compare the larger components of the noise signal and the reference one instead of using all the signals. They assume that the large component of PRNU is more reliable and thus should be used in correlation detection

while the other components should be discarded. The values of the PRNU are sorted firstly, then, a certain percent of top values are recorded with their location information. The correlation is calculated between these points and their corresponding points in the noise residue of test image. The method was tested with several cameras. Identification results of two cameras were shown in the paper. In particular, the proposed method can better separate the two data sets and thus the accuracy of detection was expected to be increased. Moreover, since only a portion of image blocks were used to calculate the correlation, this method has less computational complexity than the traditional method.

Similar studies have been done in [45]. In this paper, images pixels are classified into four classes according two features,i.e. brightness and texture, using a fuzzy system. Then only a portion of blocks that is closest to the "high brightness, low texture" class are selected for correlation calculation. The experiments with 5 cameras showed that the detection rate was improved with this method.

In [46], Liu has developed a model to estimate the signal-to-noise ratio (SNR) for different regions of an image. The signal refers to the PRNU and the noise refers to the combination of other noise sources. A high SNR value implies that the strength of the PRNU signal is comparatively strong and the region with high SNR is more reliable for camera identification. The regions with low SNR value are usually corrupted by noise or scene content and they are not reliable for camera identification. In this paper, the image blocks are sorted according to the value of SNR and only a portion of the blocks with the largest SNR are selected to calculate the correlation. The proposed algorithm was tested with three different digital cameras. The

experimental results showed that the proposed method could better separate the distribution of the data under the two hypothesizes and decrease the false rejection rate.

In [47], Li proposed a method that utilizes the Pinciple Component Analysis (PCA) and Linear Discriminant Analysis (LDA) to extract features from the pattern noise. With a training set of images from different cameras, a PCA transform matrix can be calculated. The transformed coefficient will be the PCA features. Only the top few PCA features with largest variance are selected for camera identification because they are believed to contain the most of the PRNU information. The remaining PCA features discarded as they mainly consist of scene artifacts. In this way, a compact representation of the PRNU can be obtained and the scene content artifacts can be removed. Then the LDA will utilize the label information of the training data and transform the PCA features into a space where different classes are more separated. The identification accuracy was improved with this method. However, the false identification rate will increase if the camera is not involved in the training stage because the important features of the PRNU of the new camera may be discarded. Besides, if a new camera is to be added into the training set, the PCA and LDA transfer function must be recalculated.

Chan[19] et al. proposed to use the Kernel Principal Component Analysis Regression to predict the correlation for each individual pixel and build a confidence map. The confidence map is then used as pixel-wise weights for calculating the weighted correlation.

2.4.4 Other works related to PRNU based camera identification

There are some works that try to improve the performance of camera identification from other aspects. In [48], Li et al. proposed to firstly decompose each color channel into 4 sub-images and then extracts the PRNU noise from each sub-image. This method can improve the interpolation effect in the PRNU feature extraction. The experimental results showed that the performance was improved.

In summary, current research on PRNU-based camera identification mainly focuses on three aspects. The first one is to extract accurate PRNU from images by using good denoising filter and eliminating different kinds of artifacts. The second one is to select reliable regions in images for camera identification. The last one is to different detectors other than the cross correlation one.

2.5 Artificial neural network

Neural networks mimic the biological central nervous system. They are designed to nonlinearly map a set of inputs to a set of outputs. Neural networks are usually presented as a system of interconnected simple processing elements (PEs) which are analogous to neurons. They are adaptive information processing system that can adaptively adjust the weighting of each connection in response to the information environment. Hence, neural networks are capable of performing machine learning and pattern recognition tasks. The behavior of neural networks depends on its structure and weights. The weights are referred as the strength of connections between PEs. Neural networks bring some advantages including generalization capability, distributed memory, parallelism, redundancy, and learning compared with conventional processing technique[49].

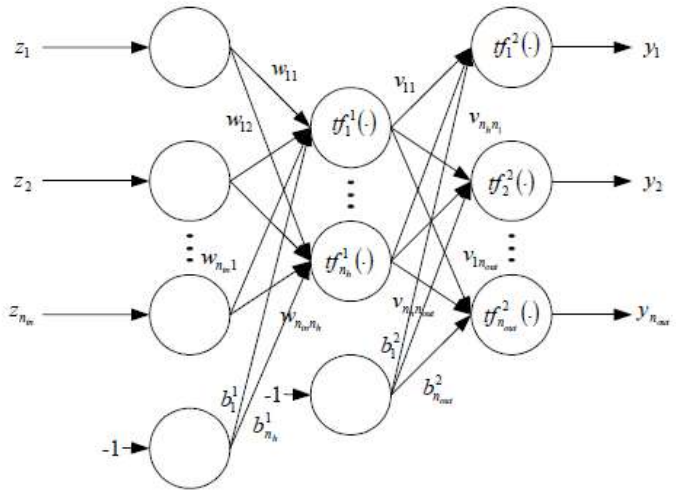


Figure 2.6 Architecture of a three layer feed-forward neural network

While many neural networks with different structures have been developed, the basic

three layer feed-forward neural network will be introduced here. Figure 2.6 shows the architecture of a three layer feed-forward neural network. The input and output relationship is given by,

$$y_h = tf_h^2\left(\sum_{g=1}^{n_h} v_{gh}\, tf_g^1\left(\sum_{i=1}^{n_{in}} w_{ig} z_i - b_g^1\right) - b_h^2\right),\ h = 1, 2, \ldots, n_{out} \tag{2.32}$$

where $z_i, i = 1,2,..., n_{in}$ are input variables; n_{in} denotes the number of inputs; n_h denotes the number of hidden nodes; $w_{ig}, g = 1,2,..., n_h$ denotes the weights of the connection between *i*-th input and *g*-th hidden node; v_{gh} denotes the weights of the connection between *g*-th hidden node and *h*-th output node; b_g^1 and b_h^2 denotes the bias for hidden nodes and output nodes respectively. $tf_g^1(\cdot)$ and $tf_h^2(\cdot)$ denote the transfer function in the hidden nodes and output nodes respectively. The commonly used transfer functions are the logarithmic sigmoid transfer function (logsig), hyperbolic tangent sigmoid transfer function (tansig), and linear transfer function (pureline). They are defined as,

$$\text{logsig}(\eta) = \frac{1}{1+e^{-\eta}} \in [0 \quad 1],\ \eta \in \Re \tag{2.33}$$

$$\text{tansig}(\eta) = \frac{2}{1+e^{-2\eta}} - 1 \in [-1 \quad 1],\ \eta \in \Re \tag{2.34}$$

$$\text{pureline}(\eta) = \eta,\ \eta \in \Re \tag{2.35}$$

The transfer function determines the output of the node given the inputs. The total number of parameters can be calculated by,

$$n_{para} = (n_{in} + 1) \times n_h + (n_h + 1) \times n_{out} \tag{2.36}$$

When the architecture of the neural network is determined, the parameters can be tuned by some optimization algorithms to meet certain requirement.

2.6. Discussion

As introduced in the previous sections, there are mainly two directions to improve the performance of camera identification. They are the denoising filter design and scene content suppression approach. In [23], a denosing filter based on a signal dependent noise model was proposed, however, its performance is similar to the one using the original wavelet filter. In [24], BM3D filter was proposed to be used for PRNU extraction for forgery detection. Besides, the predictor was used in [35] to predict the reliability of each image block. Comparative studies between the BM3D filter and the original wavelet filter with and without the predictor were carried out. In the case of no predictor, the performance of BM3D was better than the original filter. But in the case when predictor was added, the performance of two filters became similar. The predictor actually utilized the image features in an attempt to suppress the scene content problem.

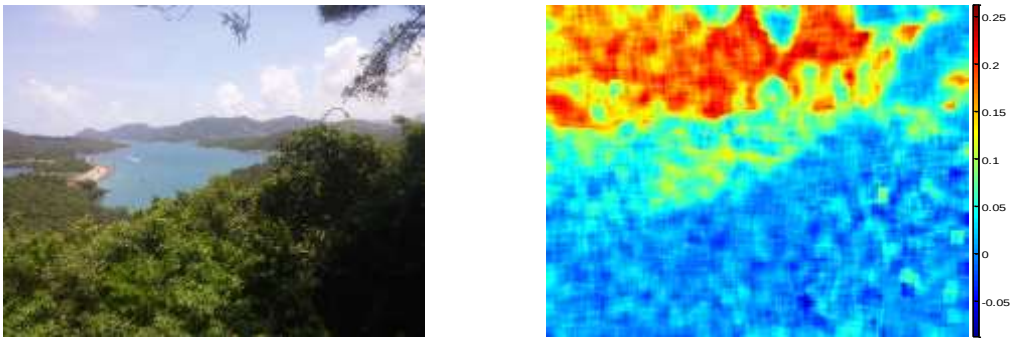
For the scene content suppression approach, five models were proposed to attenuate the influence of scene details[17]. With the proposed scheme, the performance was significantly improved. For a block size of 512 by 512, the true positive rate was improved by 12.21%. In [45], the proposed method has decreased the false rejection rate by 16%. The improvement was much larger than that reported in [23] and [24]. As discussed in [24], the predictor (which falls in the category of scene content suppression approach) can improve the performance with a poorer denoising filter such that its performance is similar to that with a better denoising filter. Therefore, it is likely that adopting scene content suppression method is more promising than improving the denosing filters. Therefore, further research works on improving the

scene content suppression approach for camera identification were carried out.

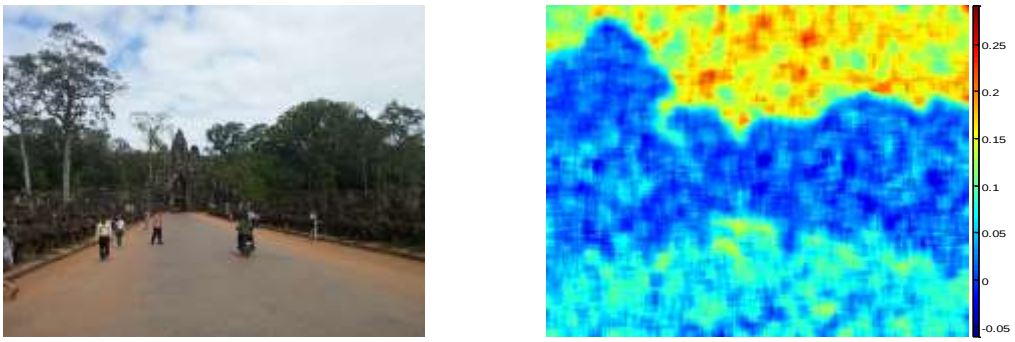
Chapter 3 Weighting optimization with Neural Network

3.1 Motivation

The influence of scene content to PRNU based camera identification can be illustrated in Figure 3.1.



(a) (d)



(c) (d)

Figure 3.1 Natural images and their correlation map

Figure 3.1 (a) and (c) are two images of the natural scene and Figure 3.1 (b) and (d) show the local correlation energy map with its reference PRNU calculated within a 64 by 64 window for each pixel for image (a) and (c) respectively. It can be observed

that the blue sky regions have the highest correlation energy due to the fact that those areas are smooth and have high luminance. In Figure 3.1 (a), the tree area is of complicated texture and its correlation is much lower than that in the blue sky region. However, for the tree region, though highly textured, it can still be seen that the correlation on the left is higher than that in the shaded area on the right. This is because the luminance is higher in the left than that in the right. As described in last chapter, the scene content-based method tries to make more use of the reliable regions and avoids the regions with complicated texture and low luminance.

While different schemes have been proposed to attenuate the influence of scene details, it is difficult to find a method to determine how much should we bias to the reliable regions. For example, in [17], the author made the assumption that the stronger a signal component in noise residue is, the more likely that it is associated with strong scene details, and thus the less trustworthy the component should be. Therefore the strong signals in the noise residue should be attenuated. However, it is difficult to determine how much we should attenuate the strong signal. Thus five different models for assigning the weighting factors to the noise residue were proposed. Despite that, the five models may not include the best model for assigning weighting factors. Moreover there is one parameter α needs to be tuned in each model. In [44], the author proposed to use only a portion of the signal with largest reference PRNU magnitude for the detection. However, deciding the exact number of pixels to be used is a problem. Using too few noise signals may lead to an increase of the standard deviation of the correlation value. But using too many, the improvement will be small. Hence the portion of signal to be used is determined empirically in [44]. In view of this, a neural network is proposed to find the weighting factors adaptively. Some

features of images will be extracted first and they will be used as input of the neural network. A set of training data will be used to train the neural network. Thanks to the good learning capability of neural network, optimal weighting factors to each region will be assigned.

3.2 Proposed weighting optimization with Neural Network

The overall framework of the proposed system for camera identification is shown in Figure 3.2. It consists of two main parts: training phase and testing phase. In the training phase, the neural network was trained using a set of training images. By optimizing the objective function, a set of optimal weighting factors is expected to be obtained. In the testing phase, the testing image would use the weights from the neural network and then decide whether the testing image is obtained from a particular source camera or not.

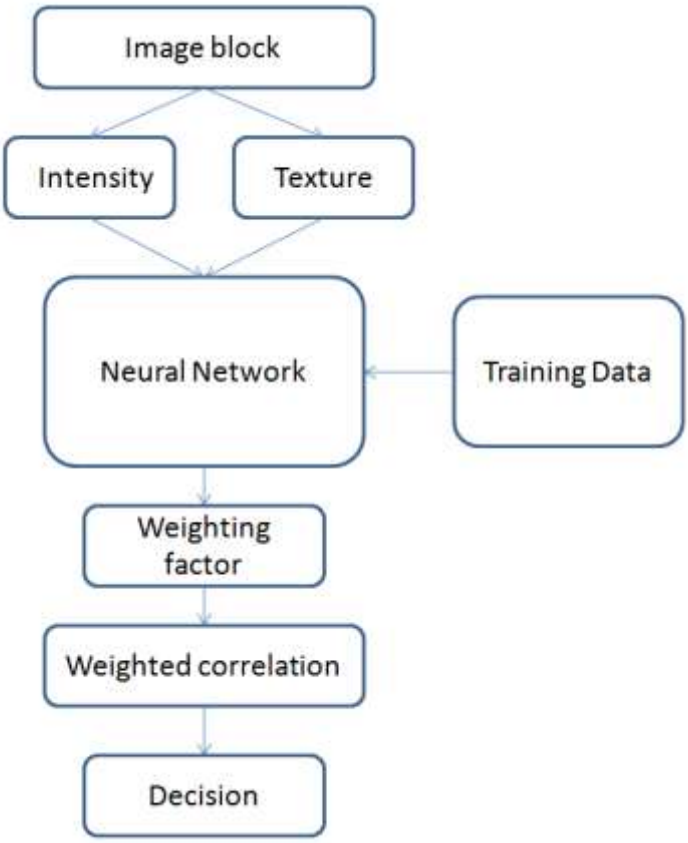


Figure 3.2 Overall framework of the proposed system

In both training and testing phases, an image is divided into a number of blocks.

Features are then extracted from each block. PRNU signal is lost in dark or saturated regions and complicated textures can contaminate the PRNU signal [35]. Hence, intensity feature and texture complexities feature are extracted from each block. The intensity feature and texture feature are defined respectively as,

$$f_{\text{int},I_i}^{j} = \frac{1}{N_b} \sum_{(x,y)\in jth\ \text{block}} I_i(x,y) \qquad (3.1)$$

$$f_{\text{text},I_i}^{j} = \frac{1}{N_b} \sum_{(x,y)\in jth\ \text{block}} \frac{1}{1+\text{var}(I_i(x,y))} \qquad (3.2)$$

where $I_i(x,y)$ is the intensity of the *i*th image I_i at pixel (x,y), N_b denotes the total number of pixels in the *jth* block and $\text{var}(I_i(x,y))$ is the variance within a 3 by 3 neighborhood of the pixel (x,y).

Figure 3.3 shows the architecture of the proposed neural network for weighting optimization. The two features in equation (3.1) and (3.2) are used as inputs to a three layer feed forward neural network to calculate the weightings for each block of the image. Let $z_1^k = f_{\text{int},I_i}^{k}$ and $z_2^k = f_{\text{text},I_i}^{k}$ where *k* stands for the *k*th image block. The output of the neural network which is also the weighting factor is governed by:

$$y^k = \text{tansig}\left(\sum_{j}^{n_h} \text{tansig}(S_j)\cdot w_j - b^2\right) \qquad (3.3)$$

where

$$S_j = \sum_{i=1}^{n_{in}} v_{ij} z_i^k - b_j^1 \qquad (3.4)$$

and

$$\text{tansig}(\eta) = \frac{2}{1+e^{-2\eta}} - 1 \qquad (3.5)$$

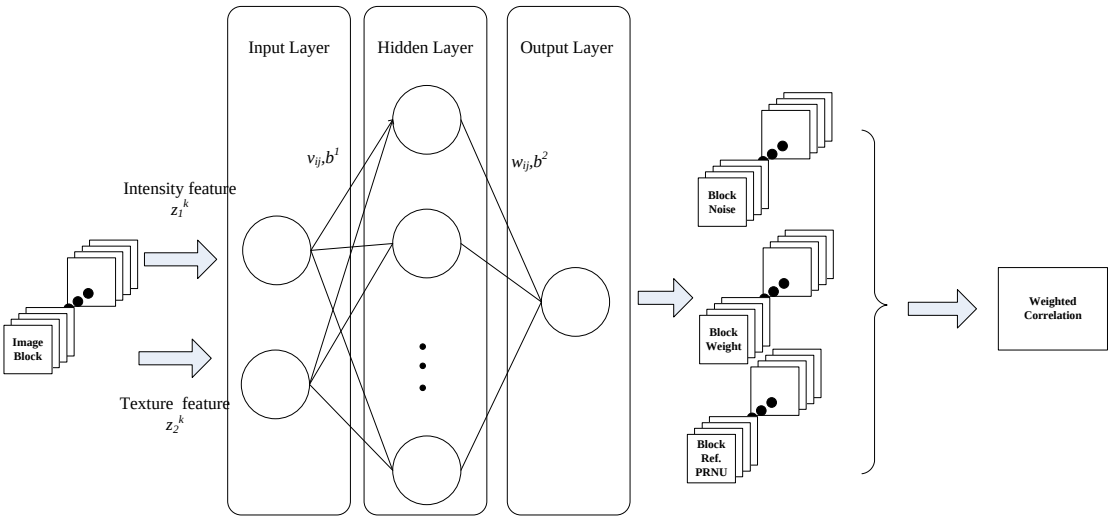


Figure 3.3 Neural Network for weighting optimization

$v_{ij}, i = 1,2,..., n_{in};\ j = 1,2,..., n_h$ denotes the weight of the link between the i-th input node and the j-th hidden node; n_{in} and n_h represent the number of inputs and the number of hidden nodes respectively; w_j denotes the weight of the link between the j-th hidden node and the single output; b_j^1 and b^2 denote the biases for the hidden nodes and output node, respectively. In this case, $n_{in} = 2$ and $n_h = 7$. The parameters to be tuned are v_{ij}, w_j, b_j^1 and b^2. So the total number of parameters is 29.

To let the neural network is trained so that desired outputs can be generated. An objective function should be defined appropriately for training the neural network to produce desired outputs. In the conventional neural network, the objective function (that is going to be minimized) is the mean squared error (MSE) of the Neural Network

outputs. However, in this situation, the ground truth output i.e. the optimal weighting is not available. Therefore to guide the Neural Network towards generating appropriate weightings, the objective function is defined as:

$$f_{obj} = \frac{1}{N}\sum_{i=2}^{N} weightedcorr^{i} \quad (3.6)$$

where

$$weightedcorr^{i} = (\frac{1}{M}\sum_{k}^{M} y^{k} corr^{k}) \quad (3.7)$$

where N is the number of images used to train the neural network and M is the number of blocks in an image. $weightedcorr^{i}$ calculates the weighted correlation for the ith image and $corr^{k}$ is the correlation of the kth image block. By maximizing this objective function, we are trying to find the parameters which maximize the average weighted correlation of the training images based on the two input features. The training process of the neural network is shown in Figure 3.4. Training images are first divided into blocks and the outputs for each of the block are obtained. The objective function is evaluated using each block output, the reference PRNU and the noise image at function input.

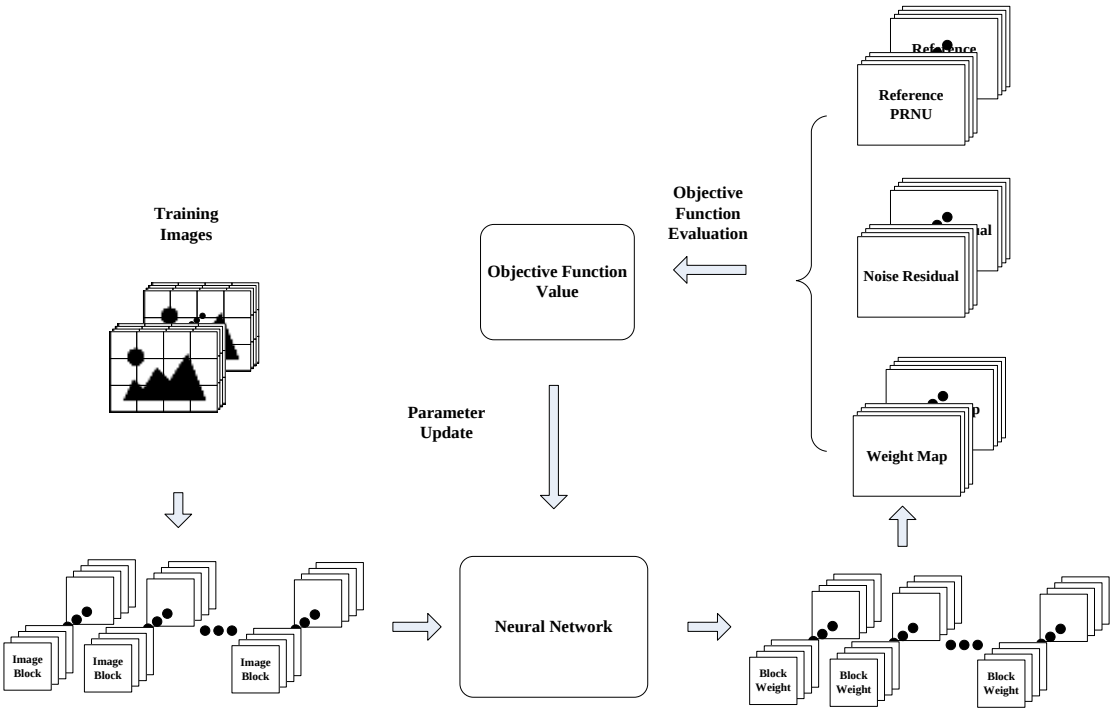


Figure 3.4 Training Process of the proposed Neural Network

The objective function is optimized with Differential Evolution (DE) [50] which is a simple yet efficient population-based stochastic method for global optimization problems. DE is easy to implement, robust, requires fewer parameters to be tuned and has a fast convergence speed[51]. Thus DE has been employed in many industrial applications in the last few years. Example areas of application include power system optimization[52], digital filter design[53, 54], mechanical engineering[55], etc.

3.3 Experimental results and analysis

Images from the Dresden Image Database[56] are used for the experiments. Six cameras from three different models are considered. Hence, we have three pairs of cameras that are of the same model. Table 3.1 summarizes the camera model, sensor type, resolution and picture format of the testing cameras. The size of the image is cropped to 256 x 256. The block size for applying the weighting is set to be 32 x 32. So there are a total of 64 blocks and 64 weighting factors for each image. 150 images from each camera are used. These images are randomly divided into 3 sets. 50 images are used to extract the PRNU of the camera, 50 images are used to train the neural network and another 50 images are used to test the performance of the proposed algorithm. Both the original images and the compressed images with a quality factor of 70 are tested. The color images are converted to gray level image before PRNU extraction.

The classification problem is formulated as a binary hypothesis problem. Let H0 be the hypothesis that the image under test and the reference PRNU are from the same camera while H1 represents the hypothesis that they are from different cameras. Therefore, for each camera, there will be 50 images for hypothesis H0 and 250 images for hypothesis H1. The performance of our proposed method is compared with the basic algorithm [7], the MLE method [12], the phase pattern noise method [22] and Li's enhancement pattern noise method [17]. For Li's method, model 3 and model 5 are selected for comparison because of their superior performance as compared with other models in [17].

Camera model	Sensor	Resolution	Format
Canon Ixus70	1/2.5”	3072 x2304	JPEG
Nikon Coolpix S710	1/1.72”	4352 x 3264	JPEG
Sony DSC-H50	1/2.3”	3456 x2592	JPEG

Table 3.1 Camera details for the experiment

The Receiver Operating Curve (ROC) curve shows the true positive rate (TPR) versus the false positive rate (FPR). Figure 3.5 and Figure 3.6 show the averaged ROC performance over the 6 cameras for different methods for the original and the compressed images respectively. Table 3.2 and Table 3.3 show the averaged TPR when the threshold is set such that the FPR equals 0.01 and 0.05 for the original image and compressed image respectively.

Figure 3.5 shows that the proposed method outperforms other methods at low FPR regions. In Table 3.2, the proposed method is the second best for FPR at 0.05 and 0.01.

Figure 3.7 shows the distribution of correlation for one camera. The left figure is the distribution without applying weighting and the right figure is the weighted correlation. The blue color represents that the image under test is not from the reference camera while the red color represents the contrary. Table 3.4 lists the averaged correlation values with and without applying the proposed weighting scheme. It can be seen that with the weighting scheme, the averaged correlation value has become larger than that without weighting scheme. This means that the proposed weighting scheme puts emphasis on reliable regions in source camera identification.

Figure 3.8 shows the weighting map derived from the proposed method for some images. The left figures are natural images and the right figures are their corresponding weighting map. A white block indicates a weighting factor of 1 and a black block indicates a weighting factor of 0. From these figures, we can see that the area with high luminance and low texture complexity will be assigned high weighting factors and low weightings will be given to the dark areas or highly textured regions.

In Figure 3.6, it can be observed that for the compressed images, the basic method performs the best and our proposed method performs the second best among all the methods. With some experiments, we have found that all the other methods have adopted a wiener filter after the denoising filter to remove random noise from the PRNU. Experimentally, it was found that the wiener filter may lower the identification accuracy for compressed images. We removed the wiener filters for all the methods and redo the experiments. Figure 3.9 shows the results without wiener filtering. It can be seen that the MLE and the proposed method will have comparable performance with the basic method. The wiener filtering can enhance the performance of the basic method, but not all the other methods.

In order to further test the performance of the proposed scheme, 5-fold cross validation are adopted to select the training and testing images. In this experiment, for each camera, 50 images out of 150 images are selected randomly for estimating the reference PRNU. The remaining 100 images are randomly divided into 5 groups of 20 images each. As shown in Fig. 3.10, for each run, 1 group of images are used for testing and the other 4 groups are used for training the neural network. The overall result for all the test images is shown in Table 3.5. In the 5-fold cross validation test, the

proposed method outperforms other methods at all the FPR levels. The reason for this improvement may be due to the increase of training data. In this experiment the number of training images has been increased from 50 to 80 for each camera.

Recently the deep learning architectures, especially the convolutional neural network (CNN) has been proven to be very powerful for image processing tasks. Convolutional filters are used to extract features at different abstract levels. In the proposed method, the texture and intensity are used as hand crafted features for the application. For these rough features, deep model may not be necessary. However detailed features are missing which can be captured by CNN. CNN may be a promising approach for estimating the weight for each image pixels. The training algorithm need to be redesigned, since there is no ground truth for the optimal weighting. Speed is concern because the complexity of CNN is usually very high.

The characteristic of PRNU varies with the camera model. If the neural network is trained with images from different models. The performance may be compromised.

However, the PRNU based approach requires the suspect camera or a set of images taken by the camera is available for estimating the PRNU. A portion of the images can be used as training set for the neural network. However, training the neural network for each camera model is inconvenient in practice. Therefore, another approach which is easy to implement and does not require a training phase will be introduced in the next chapter.

	FPR=0.05	FPR=0.01	FPR=0.001
Basic	0.8925	0.7667	0.4866
MLE	0.9183	0.8525	0.6016
Phase	0.9125	0.8575	0.5850
Model3	0.9000	0.8425	0.6050
Model5	0.8592	0.7733	0.5200
Proposed Method	0.9150	0.8550	0.6733

Table 3.2 True Positive Rate at different False Positive Rate for original images

	FPR=0.05	FPR=0.01
Basic	0.6825	0.5308
MLE	0.5242	0.3492
Phase	0.5042	0.3200
Model3	0.9000	0.3692
Model5	0.4975	0.3008
Proposed Method	0.5817	0.3842

Table 3.3 True Positive Rate at different False Positive Rate for compressed images

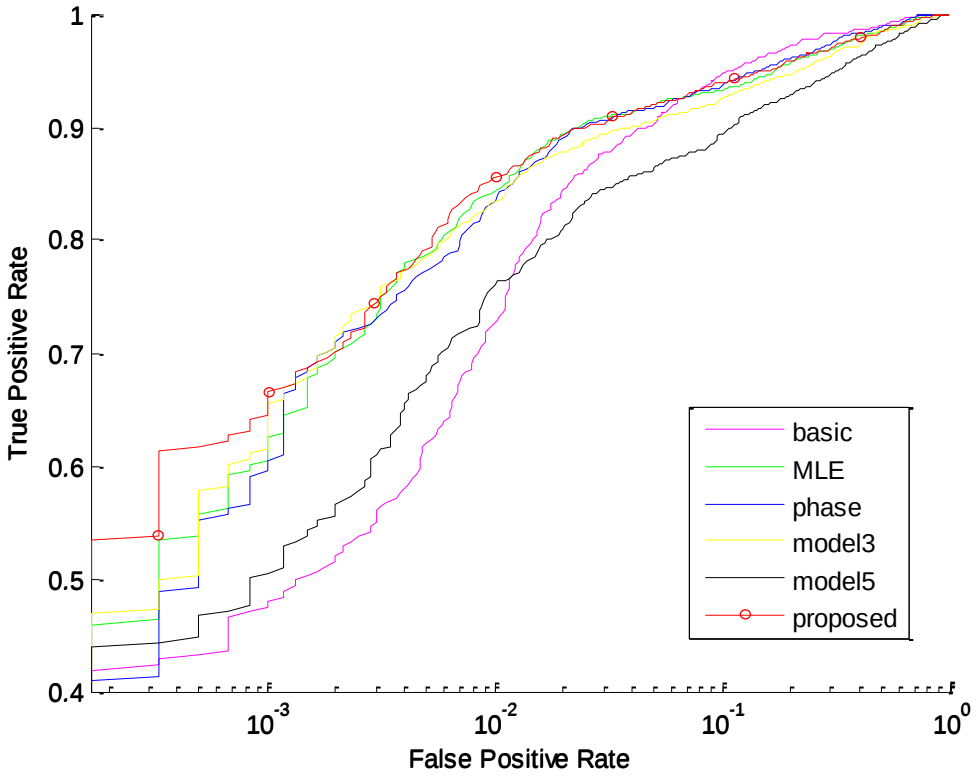


Figure 3.5 ROC for various algorithms for original images

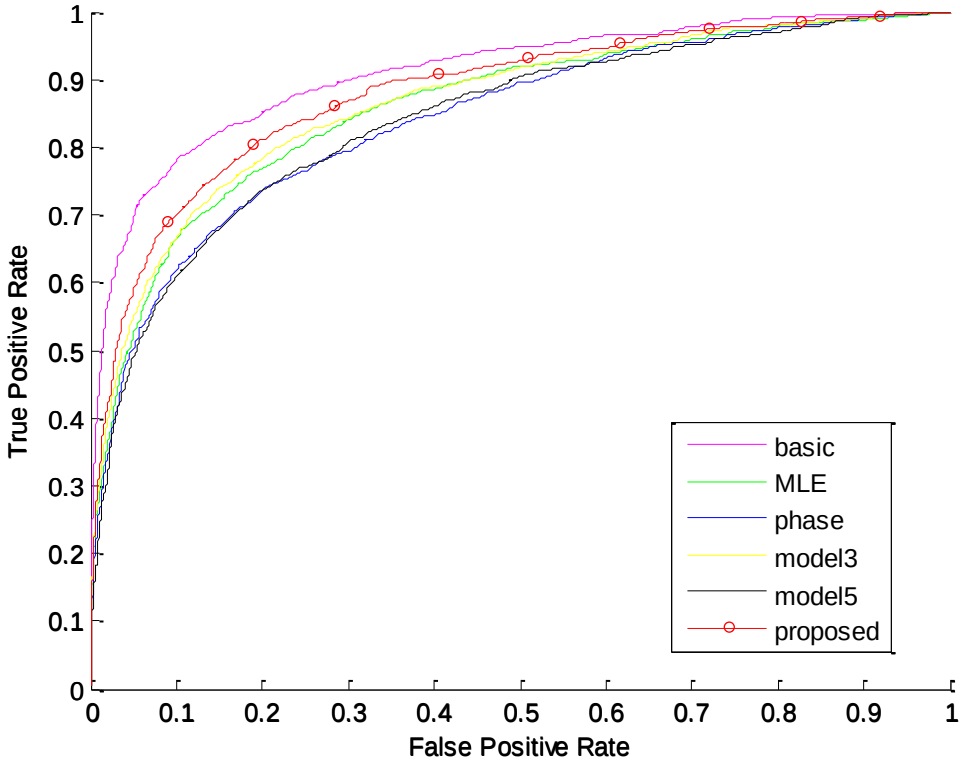


Figure 3.6 ROC for various algorithms for compressed images

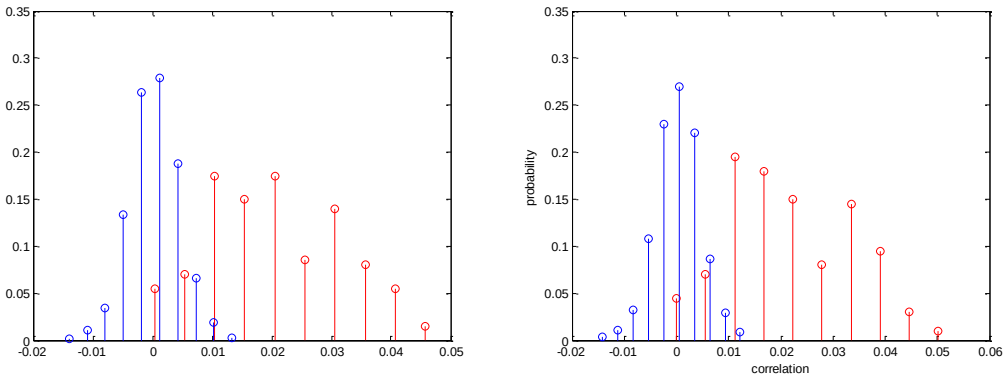


Figure 3.7 Distribution of correlation for (1) correlation without weighting and (2) weighted correlation

Camera No.	Averaged weighted correlation	Averaged correlation
1	0.0355	0.0312
2	0.0315	0.0278
3	0.0167	0.0158
4	0.0215	0.0204
5	0.0585	0.0564
6	0.0344	0.0327

Table 3.4 Averaged correlation values with/without weighting

Figure 3.8 weighting map for given images

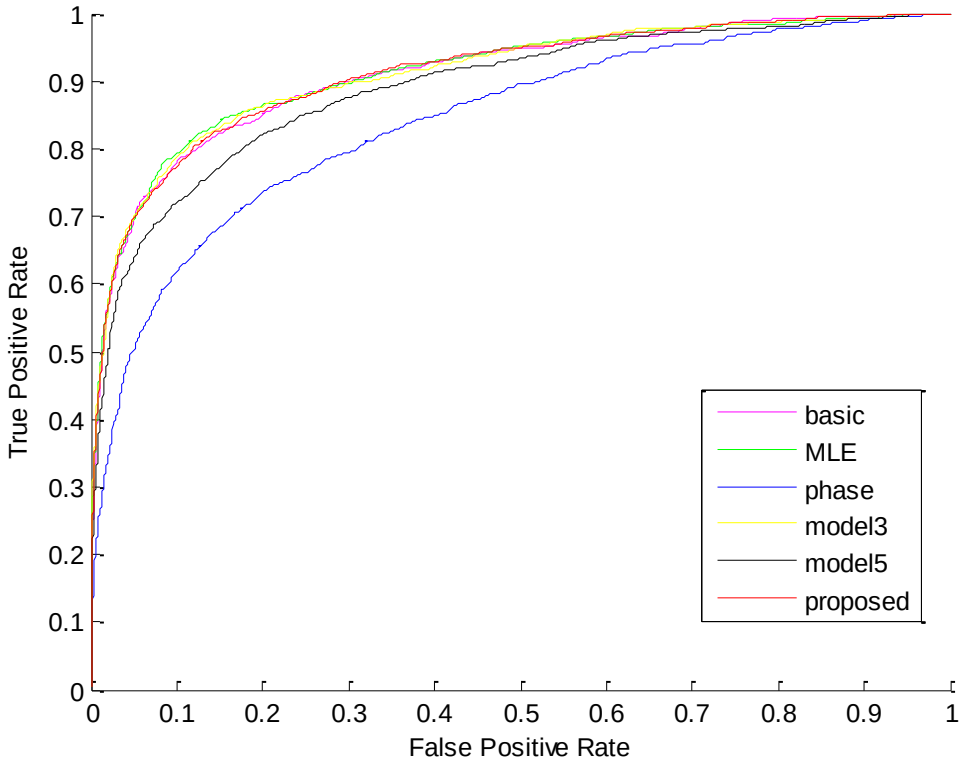


Figure 3.9 ROC curves for various algorithms for compressed images without wiener filter

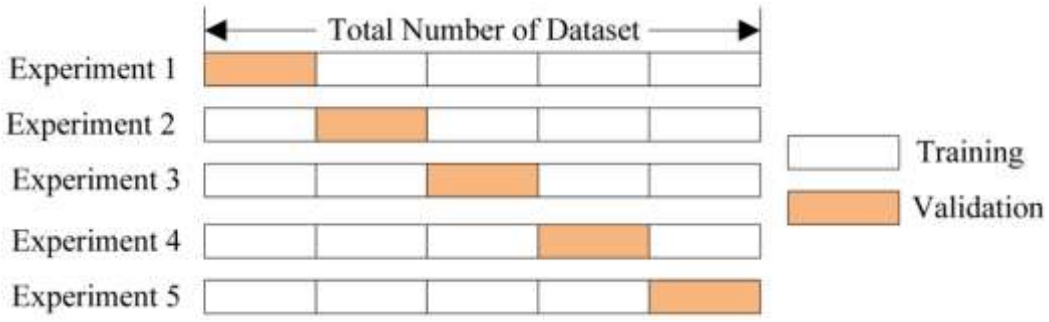


Figure 3.10 Illustration of 5-fold cross validation

	FPR=0.05	FPR=0.01	FPR=0.001
Basic	0.8850	0.7567	0.4867
MLE	0.9167	0.8550	0.6016
Phase	0.9117	0.8700	0.5850
Model3	0.8967	0.8517	0.6050
Model5	0.8667	0.7667	0.5200
Proposed Method	0.9200	0.8833	0.6850
Rank of Proposed	1	1	1

Table 3.4 True Positive Rate at different False Positive Rate for 5-fold cross validation test

3.4 Chapter Summary

Though PRNU has been proved as a powerful tool for source camera identification, the scene content effect can severely deteriorate the performance of PRNU-based camera identification. While some methods have been proposed to give higher

weighting to reliable regions in images, it is difficult to find the optimal weighting values. In this chapter, we have developed a new scheme to obtain the weighted correlation for source camera identification using an artificial neural network. Since minimize the mean squared error is applicable for this weighting determination application, the neural network is trained to maximize the separation between the positive and negative data sets. The resultant neural network can be used to find the optimal weightings for different image regions. The proposed method is compared with several state-of-art methods. The experiments show an encouraging result in terms of the ROC curve, the true positive rate and the false positive rate.

Chapter 4 A Local Variance Based Approach to Alleviate the Scene Content Interference for Source Camera Identification

4.1 Introduction

To obtain the PRNU signal, denoising is applied to those images under test and the resultant noise residual forms the PRNU signal. One problem of using PRNU for source camera identification is that the scene content can severely contaminate the extracted PRNU and noise residues. If images contain a lot of textures, have low or saturated intensities, the accuracy of identification will drop. Several methods have been proposed to suppress the influence of scene content. In [35], a maximum likelihood method is proposed to estimate the camera reference PRNU. Kang et al. [22] proposed to whiten the PRNU in the frequency domain to estimate the reference PRNU. The approaches which make use of the idea of reliable regions have been proposed in [17][18][19][20]. In[19][20], learning based methods are adopted, in which a training phase is required before camera identification. In [17], Li made a hypothesis that the stronger a signal component is, the more likely that it is associated with strong scene details, and thus the less trustworthy the component should be. Based on this hypothesis, Li proposed five different models to shrink the noise residuals with high magnitude. However, there is no theory showing that the five models give the optimal weightings and the parameters for these models are chosen empirically. It is very important to allocate appropriate weightings to different regions of the image. If the weightings are assigned too aggressively, for example

assign a weighting of 1 to the smoothest regions and assign 0 to other regions, the decision statistics obtained will be highly unstable because only small portion out of the whole image is utilized for the detection. On the contrary, if we assign the weightings too conservatively, the improvement of accuracy will not be significant. This problem is illustrated in Figure 4.1. Figure 4.1 shows the distribution of the decision statistics for positive and negative cases. Figure 4.1 (a) is the distributions when no weighting is applied and (b) is the distributions when weighting an over aggressive weighting is adopted. In the case of (b), although the mean correlation for matching case becomes larger, the chance of false detection is higher because the standard deviation of correlation becomes larger which results in a larger overlapping between the positive and negative situations. Hence, the accuracy of detection can only be improved if the weighting is selected appropriately.

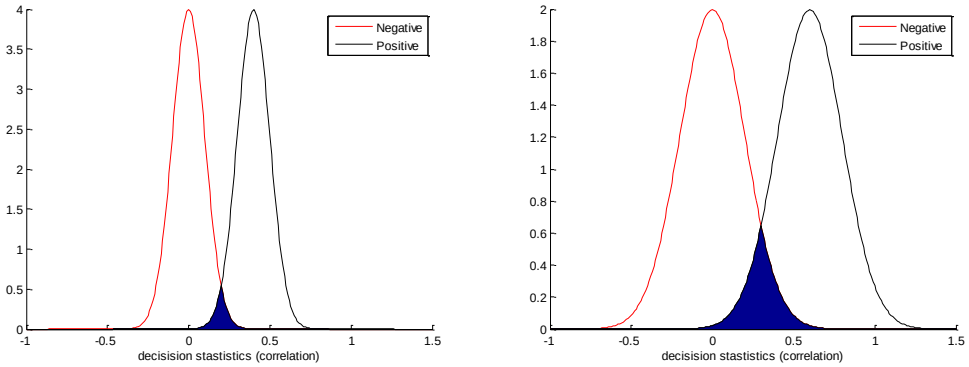


Figure 4.1 Illustration of aggressive weighting problem

In this chapter, on one hand, we study how the noise residual strength affects the source camera detection accuracy. On the other hand, we proposed a method to obtain the weighting for each pixel based on the general matched filter which has been proved to be the optimal detector.

According to Chen and Fridrich et al. [12], the noise residual can be expressed as,

$$\mathbf{W} = \mathbf{I} - F(\mathbf{I}) = \mathbf{TIK} + \Xi. \qquad (4.1)$$

where Ξ contains a combination of other independent noise and scene content artifact due to the imperfection of the denoising filter $F(.)$ and **T** is an attenuation factor which indicates how much the PRNU signal is retained in **W** because the denoising filter may remove part of the PRNU signal . In the source camera identification problem, we may refer the term **IK** as the signal of interest and Ξ as the undesired noise. Because of the scene content artifacts, Ξ is not stationary within images. In areas with complicated texture, the noise Ξ will be likely to have large variance.

To resolve the scene content problem, Li [17] has proposed five models to suppress the noise residual whose magnitude is large for the reason that the large the noise magnitude is, the more likely that it is caused by the scene content artifacts. From equation (4.1), it can be seen that Li [17]'s assumption is plausible as the scene content artifact may lead to a large value for Ξ and if **IK** is much smaller than Ξ, the noise residual **W** will be large as well. If we examine the correlation detector in (2.17) or (2.18) carefully, we can find that the element with large magnitude in **W** contribute more to the resultant correlation value, despite that the large noise residual may not be reliable. Therefore attenuating large values in noise residual should be an effective method to compensate for the scene content problem.

However, equation (4.1) also indicates that the large magnitude of **W** may also arise from the strong PRNU signal. That means a large **IK** may also produce a large value of **W** even when it does not contain any scene content.

Motivated by this, for each pixel, instead of using its own value, we propose to utilize its neighbor pixels to estimate the amount of scene content artifacts. *Let* $I_{i,j}$, be the

intensity value of the pixel (i,j) in an image. The local variance and the mean of the noise residual can be calculated coarsely by,

$$\sigma^2_{\text{local}i,j} = \frac{1}{|A|}\sum_{(m,n)\in A}\left(W_{n,m} - \overline{W}_{\text{local}i,j}\right)^2, \overline{W}_{\text{local}i,j} = \frac{1}{|A|}\sum_{(m,n)\in A} W_{n,m} \qquad (4.2)$$

where A is the neighbor of the pixel $I_{i,j}$, $\overline{W}_{\text{local}}$ is the mean of the noise residual within A. As scene content artifact usually occurs in the complicated texture areas, the noise residual with scene content artifact is likely to have large variance. Therefore, for each pixel, the variance of its neighbor pixels (not including itself) is calculated. The variance value can be used as a measure of the scene content artifact. The local variance is a more reliable estimate of the amount of scene content artifact than the magnitude of the noise residual. Since pixel of interest is not involved in the calculation, the large PRNU component **IK** will not increase the value of the variance. Experiments were conducted to reveal the effectiveness of the local variance on characterizing the scene content artifact. A total of 300 natural images of six cameras from the Dresden image database [56] are selected. The camera details are summarized in Table 4.1. Each image is cropped into a size of 256×256. Another 50 images are used to estimate the reference PRUN **K** of each camera using equation (2.15). For each pixel, the product of the noise residual and the PRNU is calculated i.e. $C_{i,j} = W_{i,j} \cdot I_{i,j}K_{i,j}$ which is the covariance between the noise residual and PRNU for that pixel because the mean of $W_{i,j}$ and $K_{i,j}$ over an image will approach 0 if the image size is large. To study the relationship between $C_{i,j}$ and the noise residual magnitude $|W_{i,j}|$, the domain of $|W_{i,j}|$ is divided into 20 equal intervals in which the mean value μ and standard deviation σ of $C_{i,j}$ in each interval is calculated. The result is shown in Figure 4.2(a), where the red dots represent the mean value μ and

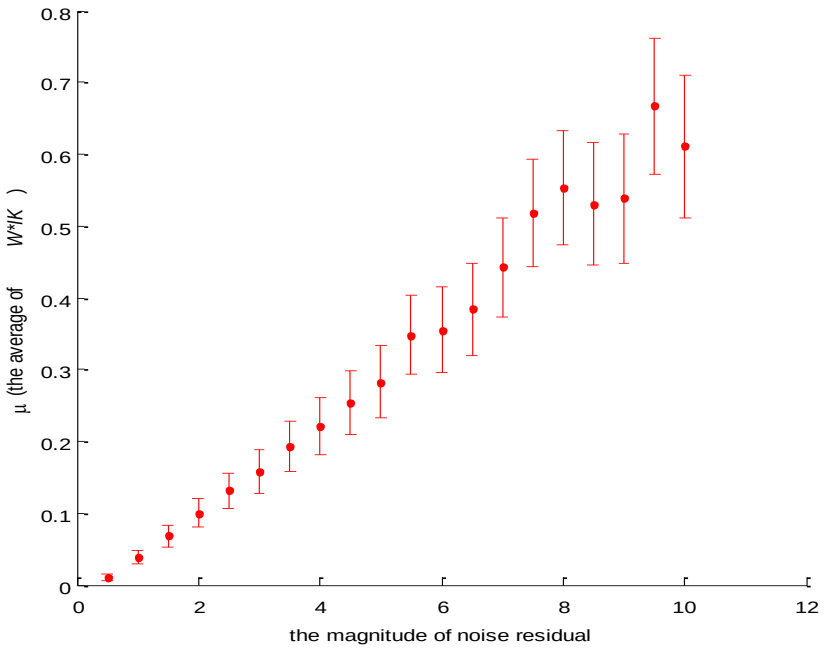
the red bars are of the length of 0.1σ. The standard deviation is scaled in the figure because it is too large compared with μ. A large mean value indicates that the strength of the PRNU signal is strong for the pixels in that bin. On the other hand, if the standard deviation of the covariance is large, the undesired noise like the scene artifact associated with that pixel should be large since the undesired noise is independent to the PRNU signal. Hence, a reliable pixel should have a large mean covariance and small standard deviation. In Figure 4.2 (a) it can be observed that both the mean value μ and the standard deviation σ increase with the magnitude of noise residual $|W_{i,j}|$. Hence the reliability of pixels can be hardly decided from Figure 4.2 (a). Therefore, Figure 4.2 (b) plots the signal to noise ratio μ/σ which can better measure the reliability of the data. From Figure 4.2 (b), it can be seen that μ/σ will not increase with $|W_{i,j}|$ when $|W_{i,j}|$.becomes large. Similarly, the relation between $C_{i,j}$ and the local standard deviation σ_{local} are plotted in Figure 4.3. Although Figure 4.2 (a) is similar to Figure 4.3 (a) in that both the mean value μ and standard deviation σ increase with the local variance, it is noticeable that the value of μ/σ drops with σ_{local} increasing which indicates that when σ_{local} gets larger, the corresponding pixel is less reliable. Therefore σ_{local} should be a more sensitive measure for the reliability of the pixel. Hence the local variance can be used as a measure of the severeness of the scene content artifact. The advantage of using the local variance of the noise residual over the original image is that, the textures in the original image might be removed by the denoising filter. Some textures, though with large variance, can be completely removed by the denoising filter and result in a small variance in the noise residual. Such regions will be considered as unreliable if the local variance of the original image is used. However, these regions should be of high reliability since

the scene artifacts have been removed. Therefore, using the local variance of the noise residual is a better measurement than that of the original image.

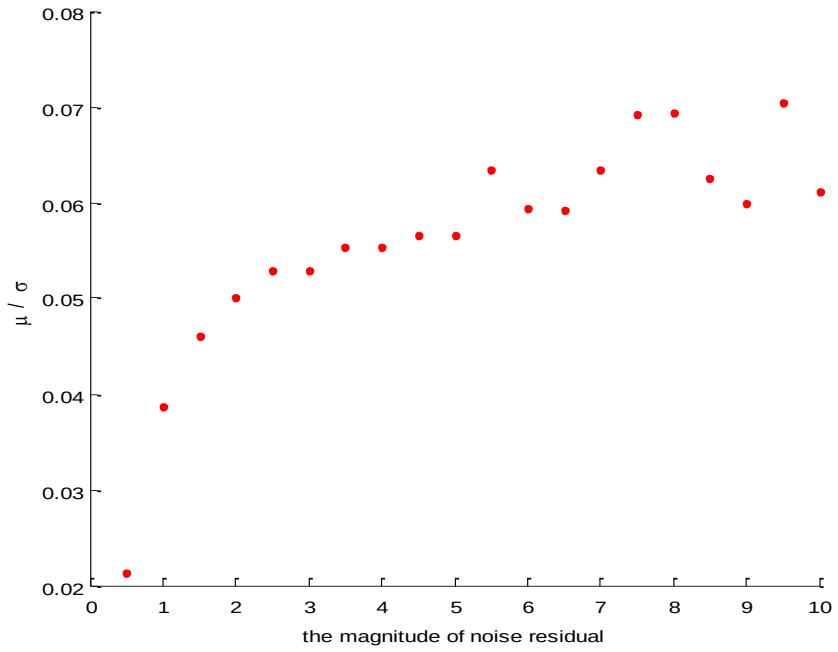
In the next section we will demonstrate how to weight the pixels according to the local variance.

Camera model	Number	Sensor	Resolution	Format
Canon Ixus70	2	1/2.5”	3072 ×2304	JPEG
Nikon Coolpix S710	2	1/1.72”	4352×3264	JPEG
Sony DSC-H50	2	1/2.3”	3456×2592	JPEG

Table 4.1 Camera details for the experiment.

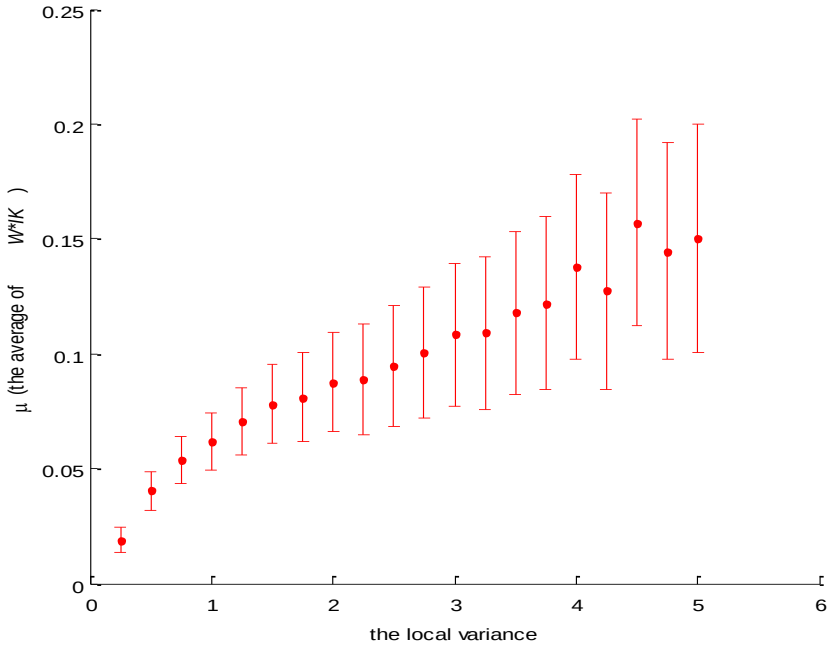


(a)

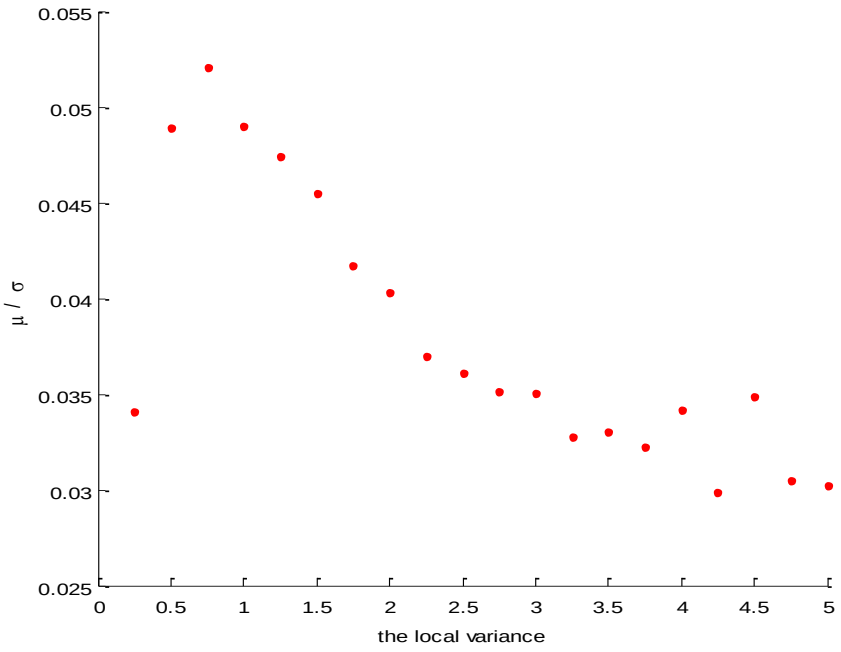


(b)

Figure 4.2 The relation of (a) C*i,j* and (b) μ/σ with respect to the magnitude of the noise residual



(a)



(b)

Figure 4.3 The relation of (a) C*i,j* and (b) μ/σ with respect to the magnitude of the local variance

4.2 The Proposed Method

The source identification problem is formulated in the framework of hypothesis testing. In the general signal detection problem, the two hypothesis are defined as,

$$\mathrm{H}_0: x_j = w_j$$

and

$$\mathrm{H}_1: x_j = s_j + w_j$$

where x_j is the observed noisy signal, s_j is the noise free signal, w_j is the white Gaussian noise (WGN) and j is the signal index. For such signal detection problem, the optimal detector will be the general matched filter [57][58] in that it gives the largest True Positive Rate for any given False Positive Rate.

Assuming that $\mathbf{w} = \{w_j | j = 1,2 \ldots N\} \sim \mathbf{N}(0, \mathbf{C})$, where **N** denotes the multivariate Gaussian distribution, N is the length of the signal and **C** is the covariance matrix of **w**. The distribution of **x** can be expressed as,

$$p(\mathbf{x}; \mathrm{H}_1) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp[-\frac{1}{2}(\mathbf{x}-\mathbf{s})^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{x}-\mathbf{s})] \qquad (4.3)$$

$$p(\mathbf{x}; \mathrm{H}_0) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp[-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{x}] \qquad (4.4)$$

where det(.) is the determinant of input matrix. According to the Neyman-Pearson theorem, the detection rate can be maximized for a given false alarm rate α by deciding H_1 when

$$L(\mathbf{x}) = \frac{p(\mathbf{x};\mathrm{H}_1)}{p(\mathbf{x};\mathrm{H}_0)} > \gamma \qquad (4.5)$$

where γ is the threshold that results in the False Alarm rate α. Taking log for both sides the expression will become,

$$l(\mathbf{x}) = \ln \frac{p(\mathbf{x};\mathrm{H}_1)}{p(\mathbf{x};\mathrm{H}_0)} > \ln(\gamma) \qquad (4.6)$$

Using equation (4.3) and equation (4.4), $l(\mathbf{x})$ can be expressed as,

$$l(\mathbf{x}) = -\frac{1}{2}[(\mathbf{x}-\mathbf{s})^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{x}-\mathbf{s}) - \mathbf{x}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{x}]$$

$$= -\frac{1}{2}[\mathbf{x}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{x} - 2\mathbf{x}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{s} + \mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{s} - \mathbf{x}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{x}]$$

$$= \mathbf{x}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{s} - \tfrac{1}{2}\mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{s} > \gamma' \quad (4.7)$$

Since the second term $-\frac{1}{2}\mathbf{s}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{s}$ is not data dependent, it can be moved to the right hand side and combines with γ'. The optimal detector then becomes,

$$y = \mathbf{x}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{s} > \gamma'' \quad (4.8)$$

If the noise is uncorrelated and $w_j \sim \mathbf{N}(0, \sigma_j^2)$, the general matched filter can be reduced to,

$$y = \sum_{j=1}^{N} \frac{s_j x_j}{\sigma_j^2} \quad (4.9)$$

where σ_j^2 is the expected variance of w_j, N is the length of the signal and y is the output of the filter.

Under the problem of source camera identification, $x_{i,j}$ is the noise residual $W_{i,j}$ of pixel $\mathrm{I}_{i,j}$, $s_{i,j} = T_{i,j}I_{i,j}K_{i,j}$ and $w_{i,j} = \Xi_{i,j}$ [12].

Hence, the detection problem will become,

$$\mathrm{H}_0: x_{i,j} = \Xi_{i,j}$$

$$\mathrm{H}_1: x_{i,j} = T_{i,j}I_{i,j}K_{i,j} + \Xi_{i,j}$$

Therefore the optimal detector will be

$$y = \sum_{i=1,\,j=1}^{i=M,\,j=N} \frac{T_{i,j}I_{i,j}K_{i,j}W_{i,j}}{\sigma_{i,j}^2} = \sum_{i=1,\,j=1}^{i=M,\,j=N} \frac{T_{i,j}}{\sigma_{i,j}^2} I_{i,j}K_{i,j}W_{i,j} \qquad (4.10)$$

From equation (4.10), it can be seen that to achieve the optimal detection performance, the value of the attenuation factor $T_{i,j}$ and $\Xi_{i,j}$ must be known. However, their values cannot be estimated easily, due to the fact that they depend on the scene content and the characteristic of the denoising filter. Therefore, usually the cross-correlation or the peak to correlation energy (PCE) is used as detection filter instead of general matched filter. Nevertheless, neither cross-correlation nor PCE is a good detector for the reason that they do not take the distinct distribution of each pixel into consideration. Hence, the distribution of the test statistics is difficult to be modeled due to the fact that the expected variance of PCE and cross correlation vary from one image to another. The value of $T_{i,j}$ and $\Xi_{i,j}$ can be calculated by learning based approached as in [12]. The learning based approach requires a large data set for training and the parameters value will be camera model dependent. Therefore, we propose a simple method from which we can estimate the distribution parameter from the single image and take advantage of the optimal detector.

Since $\Xi_{i,j}$ contains the scene content artifacts, $\sigma_{i,j}^2$ is scene content dependent. Although $\Xi_{i,j}$is not stationary due to the variation of scene content artifact, we may assume $\Xi_{i,j}$to be locally stationary because of the local similarity property of images. Thereby, for each pixel $I_{i,j}$, we calculate the local variance $\sigma_{\text{local},i,j}^2$ from equation (4.2). Given $\Xi_{i,j}$is stationary around pixel $I_{i,j}$ and $\Xi_{i,j}$ is independent to the PRNU signal, we will have,

$$\sigma_{\text{local},i,j}^2 = \text{var}(\mathbf{W}_{\text{local}}) \approx \sigma_{i,j}^2 + \sigma_{PRNU,i,j}^2 \qquad (4.11)$$

where $\sigma_{i,j}^2$ is the variance of $\Xi_{i,j}$ and $\sigma_{PRNU,i,j}^2$ is the variance due to the PRNU.

$$\sigma_{PRNU,i,j}^2 = I_{i,j}^2 T_{i,j}^2 \text{var}(\mathbf{K}) \quad (4.12)$$

where $\mathrm{var}(\mathbf{K})$ is the variance of the PRNU factor **K.** Combining equation (4.11) and (4.12) the variance of $\Xi_{i,j}$ can be calculated as,

$$\sigma_{i,j}^2 = \sigma_{\mathrm{local},i,j}^2 - I_{i,j}^2 T_{i,j}^2 \mathrm{var}(\mathbf{K}) \quad (4.13)$$

Since PRNU **K** is independent to the image content and for different region of each sensor **K** should have similar properties, **K** can be considered stationary over the image and $\mathrm{var}(\mathbf{K})$ should be constant in the image. Though the value of $\mathrm{var}(\mathbf{K})$ varies from one sensor to another, it is small as compared with $\sigma_{i,j}^2$. Therefore, the value of $\mathrm{var}(\mathbf{K})$ does not have significant influence on the estimation of $\sigma_{\mathrm{local},i,j}^2$ and an accurate estimate of $\mathrm{var}(\mathbf{K})$ is not necessary. To calculate the optimal weighting we still need to know the shaping factor **T**. However, since the noise residual is obtained from a very complex process, it is difficult to obtain the value of **T.** In this paper, to simplify the problem, we assume the shaping factor to be constant 1 over the image. Whereas, it is worth to mention that any model for estimating **T** can be incorporated into our model. Note that σ_j^2 cannot be negative and it cannot be 0. Even at smooth regions $\Xi_{i,j}$ contains the random noise like the quantization noise, $\sigma_{i,j}^2$ cannot be zero and must be positive. Hence a lower bound for $\sigma_{i,j}^2$ is set. Therefore,

$$\sigma_{i,j}^2 = \max(\sigma_{\mathrm{local},i,j}^2 - I_{i,j}^2 T_{i,j}^2 \mathrm{var}(\mathbf{K}), b_l) \quad (4.14)$$

where b_l is the lower bound for σ_k^2. Assuming that σ_k^2 is Gaussian distributed, the following can be calculated,

$$\mathrm{E}(\mathrm{y} : \mathrm{H}_0) = 0 \quad (4.15)$$

$$\mathrm{E}(\mathrm{y} : \mathrm{H}_1) = \sum_{i=1,j=1}^{i=M,j=N} \frac{\left(T_{i,j} I_{i,j} K_{i,j}\right)^2}{\sigma_{i,j}^2} \quad (4.16)$$

$$\mathrm{var}(\mathrm{y}:\mathrm{H}_0) = \mathrm{var}(\mathrm{y},:\mathrm{H}_1) = \sum_{i=1,j=1}^{i=M,j=N} \frac{\left(T_{i,j} I_{i,j} K_{i,j}\right)^2}{\sigma_{i,j}^2} \quad (4.17)$$

where $\mathrm{E}(\mathrm{y}:\mathrm{H}_0)$ and $\mathrm{E}(\mathrm{y}:\mathrm{H}_1)$is the expectation of *y* under H_0 and H_1 respectively and $\mathrm{var}(\mathrm{y}:\mathrm{H}_0)$ and $\mathrm{var}(\mathrm{y}:\mathrm{H}_1)$ are the variance. The general matched filter is normalized by $\mathrm{std}(\mathrm{y},:\mathrm{H}_0)$ such that all the detection statistics will have variance equal to 1. Therefore, the normalized general matched filter will be

$$y = \frac{\sum_{i=1.\,j=1}^{i=M,\,j=N} \frac{T_{i,j}}{\sigma_{i,j}^2} I_{i,j} K_{i,j} W_{i,j}}{\sqrt{\sum_{i=1.\,j=1}^{i=M,\,j=N} \frac{\left(T_{i,j} I_{i,j} K_{i,j}\right)^2}{\sigma_{i,j}^2}}} \quad (4.18)$$

However, in practice, there are some weak correlations between the PRNU fingerprints estimated from different cameras due to the JPEG compression, color interpolation etc. The cameras from the same manufacturer usually have higher correlations than those from different manufacturers. This correlation is likely to increase the False Acceptance Rate since such inter camera correlation might be detected as positive samples. The general matched filter does not take this situation into consideration. The Peak to Correlation Energy (PCE) does well in reducing the False Acceptance Rate caused by inter camera correlation. As discussed in chapter 4, the PCE is defined as,

$$PCE(\mathbf{x},\mathbf{y}) = \frac{\rho(\mathbf{s}_{peak} = 0, \mathbf{x},\mathbf{y})^2}{\frac{1}{MN - |A|} \sum_{s \notin A} \rho(s, \mathbf{x},\mathbf{y})^2} \quad (4.19)$$

where $\rho(\mathbf{s}, \boldsymbol{x}, \boldsymbol{y})$ is the dot product between $\boldsymbol{x} - \bar{\boldsymbol{x}}$ and $\boldsymbol{y}(\mathrm{s}) - \bar{\boldsymbol{y}}$, $\boldsymbol{y}(\mathrm{s})$ is obtained by circularly shift $\boldsymbol{y}$ by a two dimension vector **s**, *A* is a small neighbor around the peak and *MN* is the width and height of the image. Since the image is assumed not to be scaled or translated, the peak should be at the location 0. The nominator of the PCE

is the square of the correlation between signals $\boldsymbol{x}$ and $\boldsymbol{y}$ without normalization. The correlation can then be normalized by the denominator which is the square of the correlation value for all the possible shift vector **s** except for the shift close to the original position. Owing to the fact that the patterns that are shared among different cameras usually have a periodic nature, the correlations between cameras sensor will still exist even though the images are shifted. Hence, if there is a strong pattern that is shared between the two cameras tested, the denominator of PCE is likely to be high and the PCE value will be lowered. Thereby, the chance of False Acceptance can be reduced. To utilize this great feature of PCE, we propose to incorporate the general match filter into the PCE detector. The unormalized general matched filter in equation (4.10) can be rewritten as,

$$y = \sum_{i=1,j=1}^{i=M,j=N} \frac{T_{i,j}}{\sigma_{i,j}^2} I_{i,j} K_{i,j} W_{i,j} = \sum_{i=1,j=1}^{i=M,j=N} \left(\frac{1}{\sigma_{i,j}} W_{i,j} \cdot \frac{T_{i,j}}{\sigma_{i,j}} I_{i,j} K_{i,j} \right) \quad (4.20)$$

The first term in the equation is the noise residual term which is normalized by the standard deviation of the undesired noise. The second term is the reference PRNU signal which is also normalized by the standard deviation of the undesired noise. We proposed to use the normalized signal as the input of the PCE detector. Hence, the final detector can be expressed as,

$$PCE(\mathbf{x},\mathbf{y}) = \frac{\rho(\mathbf{s}_{peak} = 0, \mathbf{x}, \mathbf{y})^2}{\frac{1}{MN - |A|} \sum_{s \notin A} \rho(s, \mathbf{x}, \mathbf{y})^2}, \quad \text{where } \mathbf{x} = \frac{\mathbf{w}}{\boldsymbol{\sigma}} \text{ and } \mathbf{y} = \frac{\mathbf{TIK}}{\boldsymbol{\sigma}} \quad (4.21)$$

where all the variables are written in matrix from.

4.3Local variance estimation

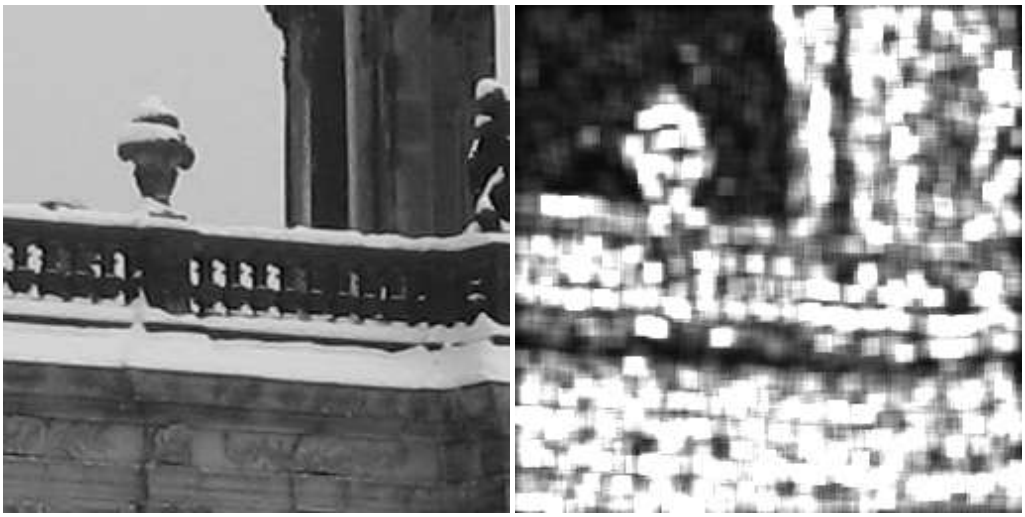


Figure 4.4 Local variance estimated using square window with equal weight

Since the weighting of each pixel is determined by the local variance in the proposed model, it is very important to estimate the local variance more precisely. A more accurate estimation of the local variance will improve the overall performance of the source identification. In section 4.2, a square window with equal weight is used to estimate the local variance. The resultant local variance is likely to be discontinuous with many blocks as show in Figure 4.4. A better option is to use a Gaussian kernel as weighting to estimate the local variance. The local variance estimated with a Gaussian kernel is expressed as,

$$\sigma^2_{\text{local}i,j} = \frac{1}{|A|}\sum_{(m,n)\in A} h_G(m,n,i,j)\left(W_{n,m} - \overline{W}_{\text{local}i,j}\right)^2 \qquad (4.22)$$

$$h_G(m,n,i,j) = \frac{1}{\sqrt{2\pi}\sigma_G} e^{\frac{-\mathrm{D}(m,n,i,j)^2}{2\sigma_G^2}} \qquad (4.23)$$

where function D(*m,n,i,j*) is the distance between pixel (*m,n*) and (*i,j*) and σ is the standard deviation parameter for the kernel. With the Gaussian kernel, pixels that are

far from the center will be assigned small weightings for estimating the local variance. Figure 4.5 (c) and Figure 4.6 (c) show the local variance estimated from the Gaussian kernel for two images. However, it is found that the large variance of the noise residual often appeared along the edges of the original image. The Gaussian kernel tends to underestimate this kind of local variance because the local variance along the edge is averaged by its nearby pixels. In order to precisely estimate the local variance, the bilateral kernel can be used. The bilateral kernel [59]penalizes the pixels that have large variation compared with the center pixel such that they will have smaller weightings. By using the standard deviation map of the original image as guide image, the joint bilateral filtering can be used to estimate the local variance of the noise residual. The joint bilateral kernel[60] is defined as,

$$\sigma^2_{\text{local}i,j} = \frac{1}{|A|}\sum_{(m,n)\in A} h_{\mathbf{B}}(m,n,i,j)\left(W_{n,m} - \overline{W}_{\text{local}i,j}\right)^2 \quad (4.24)$$

$$h_B(m,n,i,j) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-\mathrm{D}(m,n,i,j)^2}{2\sigma_s^2}} e^{\frac{-(g(m,n)-g(i,j))^2}{2\sigma_r^2}} \quad (4.25)$$

where σ_{s} and σ_{r} are the spatial and range parameters of the bilateral filter which determines how sensitive the filter is towards the change of distance and value difference to the centre pixel respectively and $g(.)$ is the local standard deviation of the original image for a given position. Since the variance for the edge region is large while that for the smooth region is small, with the guided bilateral filter, the local variance will be estimated along the edges if they are presented in the image. Figure 4.5(d) and Figure 4.6(d) show the local variance estimated with the guided bilateral filter. It can be observed that, compared with the estimation of Gaussian filter, the

edge regions have a larger local variance which is closer to the expected value. The joint bilateral filter can be implemented efficiently using the method described in [61]

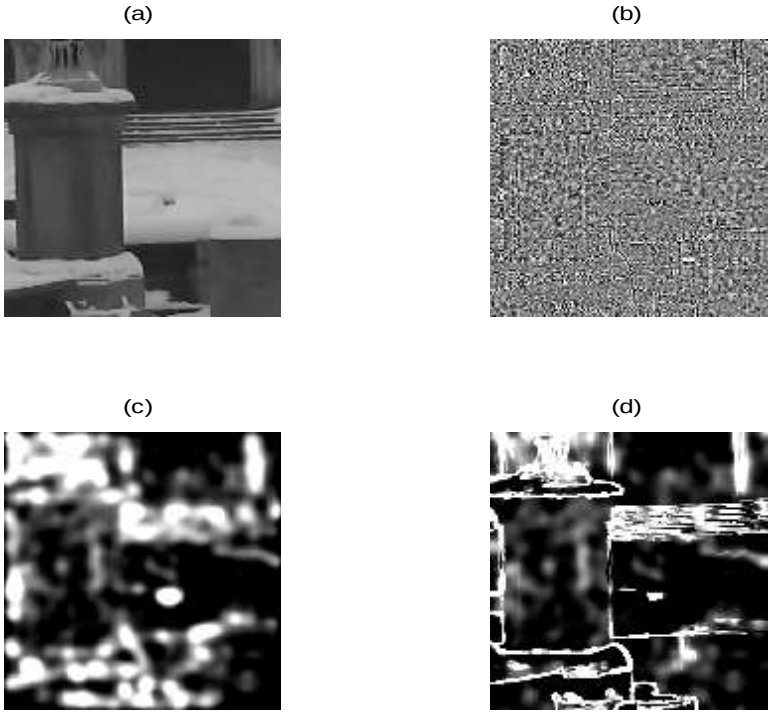


Figure 4.5 Local variance estimated with different methods. (a) the original image (b) the noise residual (c) the local variance estimated with the Gaussian kernel (d) the local variance estimated with bilateral kernel

In summary, the proposed algorithm for source camera identification will be as follows,

1. Estimate the reference PRNU with equation (2.15).
2. Extract the Noise residual **W** with equation (2.10).
3. Preprocess the reference PRNU and noise residual with zero mean operation and Wiener Filter in frequency domain.
4. Compute the local variance $\sigma^2_{\text{local},i,j}$ with equation (4.24) and (4.25).

5. Estimate the variance of the undesired noise $\sigma_{i,j}^2$ with equation (4.14).

6. Compute the detection statistics with equation (4.21).

7. Make decision based on the detection statistics obtained.

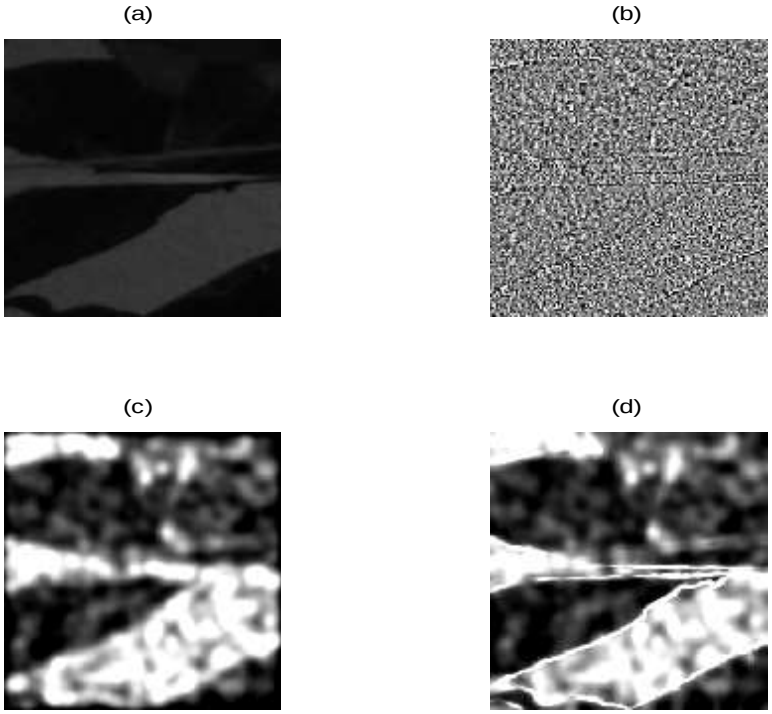


Figure 4.6 Local variance estimated with different methods. (a) the original image (b) the noise residual (c) the local variance estimated with the Gaussian kernel (d) the local variance estimated with bilateral kernel

4.4 Experimental Result

4.4.1 Experimental setup

Images from the Dresden Image Database [56] are used as testing data for the experiments. Nineteen different cameras are randomly selected from the database for testing. The cameras covered a wide range of brands and models. For most of the camera models selected, two cameras are used to test the distinguishing ability within the same model. For the cameras of the same make or model, similar post processing operations like JPEG compression and color interpolation are applied. Therefore, there should be small correlations among the noise residual of these devices and it is more difficult to distinguish photos from them. Table 4.2 summarizes the details about the cameras under test, including the camera model, the number of cameras in each model. The images cover a large variety of scenes and they are taken with different camera settings such as different ISO and different focal length which makes the source camera identification more difficult. Example photos are shown in Figure 4.7. For each camera, 50 images are selected randomly to estimate the reference PRNU, and another 50 images are used as testing image. Totally 950 images are tested. The color images are converted to gray level image before PRNU extraction.

Figure 4.7 Examples of Test Photos from Dresden Image Database

As the identification accuracy is usually very high for large images, it is difficult to compare the performance of different algorithms. To obtain a clear idea about the performance of different algorithms, the size of the image is cropped to 256 × 256 and 128 × 128. The experiment is conducted only on the two image sizes respectively.

Device ID	Camera model	Sensor	Resolution	Format	No. of Devices

1	Canon Ixus 55	1/2.5”	2592 × 1944	JPEG	1
2, 3	Canon Ixus 70	1/2.5”	3072 × 2304	JPEG	2
4, 5	Casio_EX-Z150	1/2.5”	3264 × 2448	JPEG	2
6, 7	Nokon CoolPix S710	1/1.72”	4352 × 3264	JPEG	2
8, 9	Nikon_D200	23.6×15.8	3872 × 2592	JPEG	2
10, 11	Olympus 1050 SW	1/2.33”	3648 × 2736	JPEG	2
12, 13	Panasonic DMC FZ50	1/2.5”	3648 × 2736	JPEG	2
14, 15	Samsung NV15	1/2.5”	3648 × 2736	JPEG	2
16, 17	Sony DSC T77	1/2.5”	3648 × 2736	JPEG	2
18, 19	Sony DSC W170	1/2.5”	3648 × 2736	JPEG	2

Table 4.2 Camera details for the experiment

4.4.2 Experiment Methodology

The performance of our proposed method is compared with the basic algorithm [7], the MLE method [12], the phase pattern noise method [22] and Li’s enhancement pattern noise method [17]. For Li’s method, model 3 is selected for comparison because of their superior performance as compared with other models in [17].

To extract the noise residual **W**, the wavelet denoising filter described in [3] is used since it has been reported as an effective method. For the basic algorithm [7], the reference PRNU for each camera is estimated from 50 images with equation (2.11). For our method, Li’s method [17] and the MLE method [12], the reference PRNU is estimated with equation (2.15) and the correlation for a particular image will be calculated by equation (2.17). For the phase pattern method [22], before estimating reference PRNU, the images are whitened in the frequency domain and transformed back to the spatial domain. To remove the periodical patterns, zero-mean operation

and Wiener filter in frequency domain are used to preprocess the image as described in [12].

For the proposed method, the detection statistics is calculated by the proposed detector as equation (4.21). For other methods, both the cross correlation and the peak to correlation energy (PCE) [38][39] are used as detection statistics as equation (2.18). The PCE has been reported more suitable for camera fingerprint detection because the presence of hidden periodic signal will lower PCE and reduce the possibility of false alarm [39].

The choice of the lower bound b_l is important in that it decides the largest possible weight for the image pixels. The weight of a pixel cannot be infinitely large since at least there will be some unpredictable random noise for each pixel. To study the influence of b_l, we have tested the proposed algorithm with different settings of b_l i.e. b_l = 1, 2, 3, 4, 5 and 6 for 128 ×128 images, the Receiver Operating Curve (ROC) and the detection rate for given false acceptance rates are shown in Figure 4.8 and Table 4.3 respectively. We can observe that the proposed algorithm is not sensitive to the choice of b_l. The algorithm performs slightly better when b_l=2, 3 or 4. The value of b_l is then set to 4 in the following experiment.

The guided bilateral filter is used to estimate the local variance of the noise image. The spatial parameter and the range parameter determine the characteristics of the bilateral filter. Since the parameter of the bilateral filter will only influence the estimation of the local variance, it does not directly influence the performance of the algorithm. The performance of the algorithm is not sensitive to the parameter of bilateral filter. Similar experiments are carried with a set of different parameter settings and the spatial

and range parameter are determined as $\sigma_s^2 = 5$ and $\sigma_r^2 = 3$ which achieve good detection accuracy.

The size of the window for estimating the local variance is set to 21. The size of the window cannot be too small because the estimation will be reliable for a small window size. Since the pixels far from the centre of the window contribute little to the estimation of the local variance due to the property of the bilateral filter, it is not necessary to make the window too large.

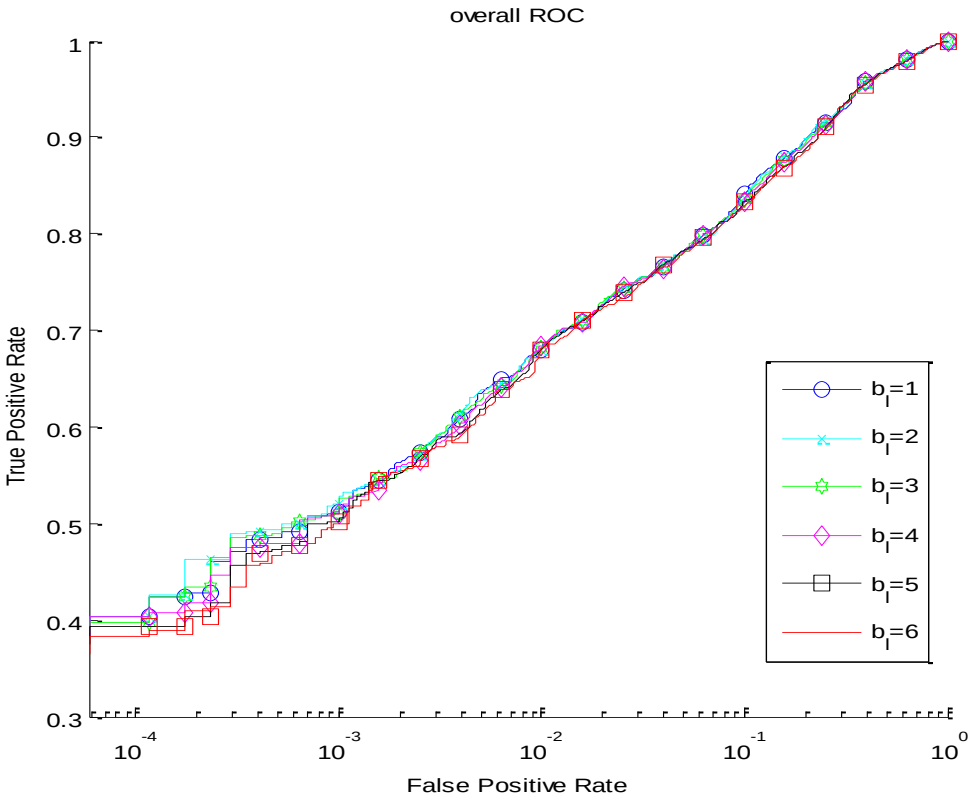


Figure 4.8 ROC curves using different b_l for image of size 128 ×128

b_l	FPR=10^{-3}	FPR=10^{-2}
1	0.51579	0.68000
2	0.52842	0.68105
3	0.52526	0.68211
4	0.51684	0.68421
5	0.50526	0.68000
6	0.49895	0.68211

Table 4.3 The TPR for given FPR using different b_l

4.4.3 Experimental result and analysis

The ROC curve illustrates the performance of the source camera identification system by plotting the True Positive Rate (TPR) versus the False Positive Rate (FPR). Figure 4.9 and Figure 4.10 show the overall ROC performance over the 19 cameras for different methods at different settings. In practice, a low False Positive Rate is desired to ensure a low probability of False Alarm. Hence we set the thresholds of detection such that the False Positive Rate equals to 10^{-3} and 10^{-2}. The True Positive Rates at the given False Positive Rate are shown in Table 4.4 and Table 4.5 for image of size image of size 128 ×128 and 256 × 256 respectively.

It can be seen that for all the methods tested, the accuracy with the PCE as the test statistics is higher than that using the cross correlation detector. For the same methods, using the PCE as the detector can always improve the detection accuracy. Another observation is that, for the phase method, the accuracy of detection using the cross correlation and PCE is very close. Using the correlation as the detector, the detection accuracy is better than other methods except the proposed method.

Possible explanation is that the phase method can remove the common patterns caused by image post processing operations. Since the PCE can also reduce the effect of common patterns in the sensor noise, the two detectors will make little difference for the phase method.

From the experimental results we can see that the proposed method outperforms other state-of-the-art methods tested. The ROC curve obtained with the proposed methods is above those produced by other methods for both image size tested. The proposed method also gives the highest TPR at different levels of FPR among all the methods tested for both image sizes. For images of size 128 × 128, the TPR improvement is 3.58% - 9.96% at FPR = 10^{-3} and 3.58% - 9.16% at FPR = 10^{-2} as compared with other methods. The second best method is the phase method which performs well at FPR = 10^{-3} but it does not perform so well at the FPR = 10^{-2}. On the contrary, the proposed method, performs constantly well for both FPR levels. For images of 256 × 256, the TPR improvement is 2.42% - 7.15% at FPR = 10^{-3} and 1.26% - 4.10% at FPR = 10^{-2}. It can be seen that the improvement is larger for smaller images. The reason for which the proposed method outperforms other methods is that the proposed method can better handle the scene content artifacts problems. The local variance estimated from the guided bilateral filter provides an accurate measure of the reliability of each pixel. The scene content artifacts make the detection statistics unreliable while the general matched filter can allocate optimal weightings to all the pixels such that the resultant detection statistics becomes more reliable. The combination of general matched filter and PCE detector can also resolve the problem of correlation among different sensors.

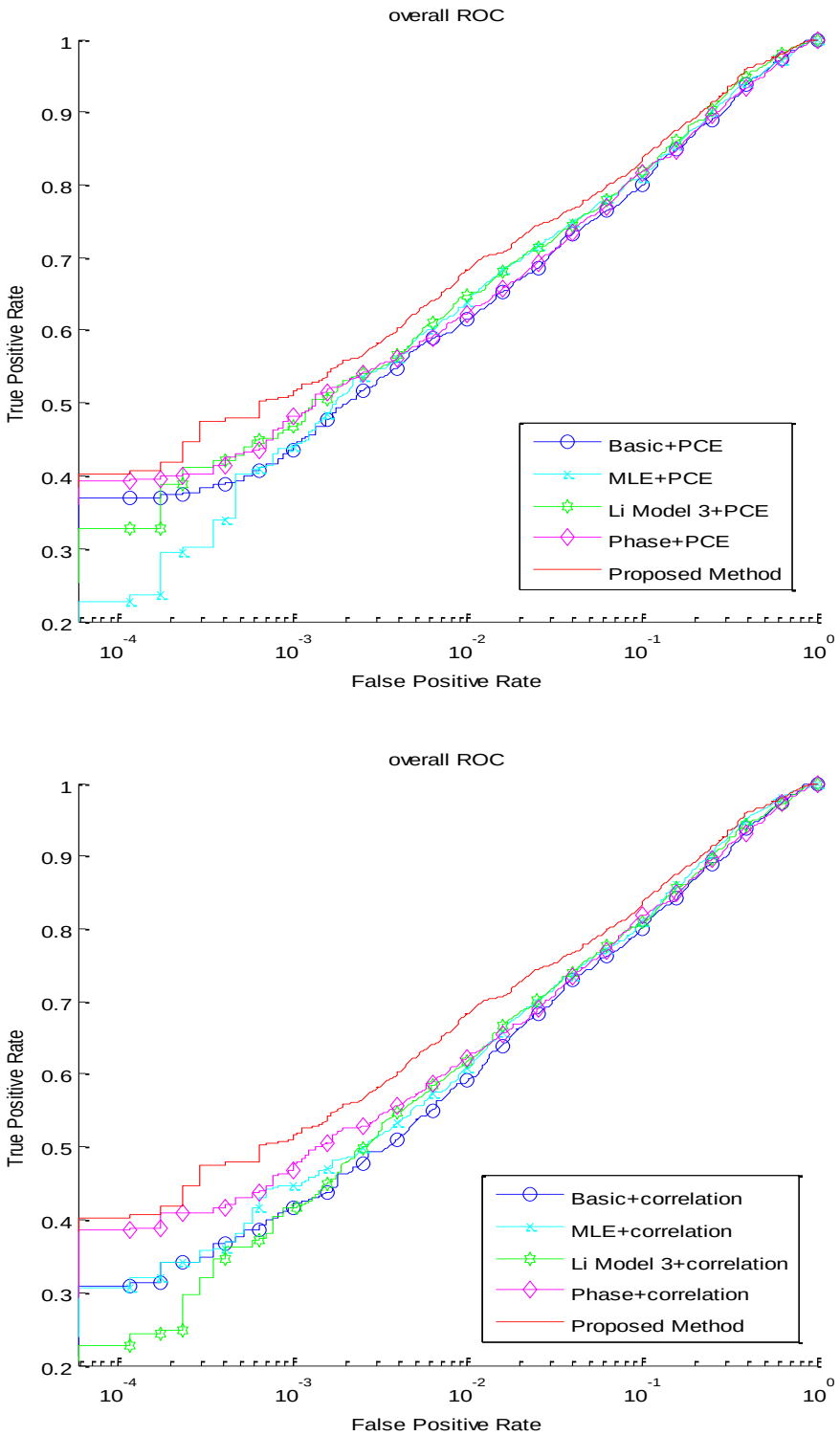


Figure 4.9 The overall ROC curve for different methods for image size of 128 × 128

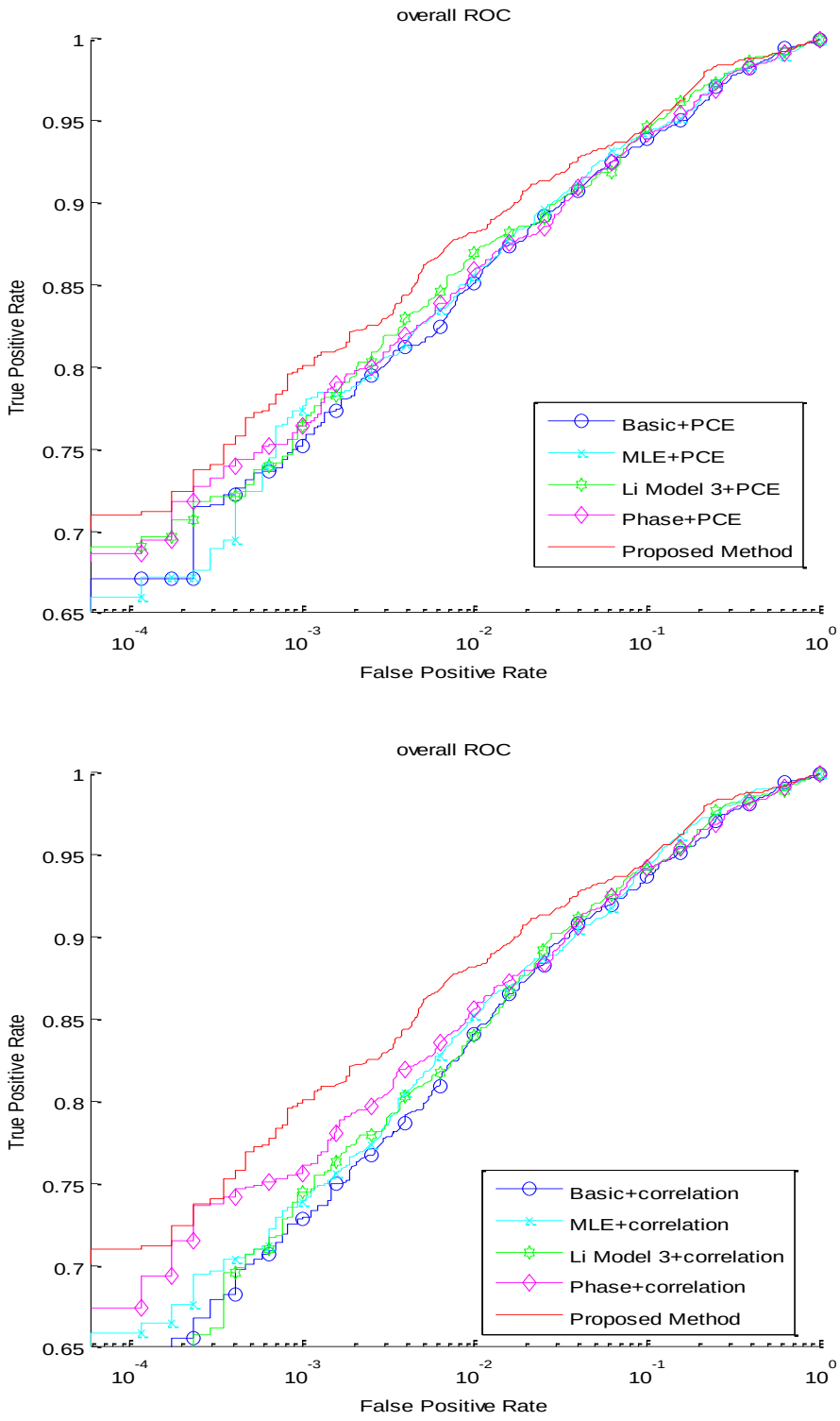


Figure 4.10 The overall ROC curve for different methods for image size of 256 × 256

	FPR=10^{-3}		FPR=10^{-2}	
	TPR	Improvement of proposed method	TPR	Improvement of proposed method
Basic + correlation	0.42421	+9.26316%	0.59263	+9.15789%
Basic + PCE	0.44316	+7.36842%	0.61474	+6.94737%
MLE + correlation	0.44784	+6.90023%	0.60906	+7.51484%
MLE + PCE	0.44000	+7.68421%	0.63895	+4.52632%
Model3 + correlation	0.41728	+9.95608%	0.61855	+6.56647%
Model3 + PCE	0.46842	+4.84211%	0.64842	+3.57895%
Phase + correlation	0.47684	+4.00000%	0.62316	+6.10526%
Phase + PCE	0.48105	+3.57895%	0.62211	+6.21053%
Proposed Method	**0.51684**	-	**0.68421**	-

Table 4.4 The True Positive Rate for given False Positive Rate for image size 128 × 128

	FPR=10^{-3}		FPR=10^{-2}	
	TPR	Improvement of proposed method	TPR	Improvement of proposed method
Basic + correlation	0.72947	+7.15789%	0.84105	+4.10526%
Basic + PCE	0.75579	+4.52632%	0.85158	+3.05263%
MLE + correlation	0.73973	+6.13266%	0.85248	+2.96290%
MLE + PCE	0.77684	+2.42105%	0.85684	+2.52632%
Model3 + correlation	0.74499	+5.60579%	0.83983	+4.22739%
Model3 + PCE	0.76737	+3.36842%	0.86947	+1.26316%
Phase + correlation	0.76105	+4.00000%	0.85579	+2.63158%
Phase + PCE	0.76421	+3.68421%	0.85895	+2.31579%
Proposed Method	**0.80105**	-	**0.88211**	-

Table 4.5 The True Positive Rate for given False Positive Rate for image size 256 × 256

For the source camera classification problem, all the images will be classified according to their source camera. To further evaluate the performance of the proposed method, the experiment for source camera classification is also conducted over the same set of data. In this experiment, each image will be attributed to the camera whose fingerprint produces the largest test statistics. The classification accuracy is tabulated in Table 4.6 which shows that the proposed method performs the best among all the methods for the source camera classification problem. For 128 × 128 image, the improvement in accuracy is 3.69% - 4.43% and for 256 × 256 images, the improvement of accuracy is 1.47% - 2.63%. The experimental results also indicate that the proposed method improves the accuracy more significantly when the image size become smaller. Table 4.7 shows the classification result for each camera where the row number is the device ID of real source camera and the column number is the device ID that the image is classified to. It is noticeable that the classification accuracy varies from camera to camera. The classification accuracy is low for devices 5 and 11. It can also be observed that the images are misclassified to different cameras and the misclassification within the same model is low. It indicates that the correlation within the same model has limited influence to the classification accuracy. Table 4.8 shows the classification result of each camera for all the methods tested from which we can see that the improvement of overall accuracy is attributed from different cameras. The accuracy of the proposed method is the highest for 12 cameras and for

other cameras the proposed method also has accuracy close to that of the highest method.

To show the merit of the joint bilateral filter, the performance for the proposed method using the Gaussian kernel and joint bilateral kernel for image of size 128 × 128 is shown in Table 4.9. It can be observed that, even using the Gaussian Kernel, the proposed method has a better performance compared with other state-of-the-art method. The joint bilateral filter can further improve the accuracy of the proposed method compared with the Gaussian Kernel.

In Figure 4.11, the experimental result of the two proposed method is shown. The experiment setup is the same with Chapter 3. It can be seen that the scheme 1 i.e. the neural network based method performs better at lower FPR and the local variance based method performs well at higher FPR.

In summary, the proposed method has the best performance for both the source camera identification and source camera classification problem among all the methods tested. The computation complexity for the two schemes proposed were also tested by measuring their computation time. All the method were implemented with MatLab on a PC with Intel i7 K6700 CPU. The computation time is tested on 300 images from 6 cameras. For each image, it computes the correlation with the PRNU from the 6 cameras. The computation time for each method is shown in Table 4.10. It can be seen that the first scheme, i.e. the neural network method, only increase the computation time by 0.6938%. This is because the neural network size is small (27 parameters) and it is applied to each block instead of each pixels. For the scheme 2, which is the local variance based method, the computation time increased by 14.2813%. The increase in time is mainly due to the joint-bilateral filter. However, in large scale test, each image

needs to be compared with a large amount of camera fingerprint in the database. The increase the in computation time will be reduced because the weighting only needs to be calculated once for each image.

	128 × 128		256 × 256	
	Accuracy	Improvement of proposed method	Accuracy	Improvement of proposed method
Basic + correlation	0.7189	+4.43%	0.9042	+2.21%
Basic + PCE	0.7232	+4.00%	0.9011	+2.52%
MLE + correlation	0.7263	+3.69%	0.9116	+1.47%
MLE + PCE	0.7284	+3.48%	0.9032	+2.31%
Model3 + correlation	0.7263	+3.69%	0.9063	+2.00%
Model3 + PCE	0.7326	+3.06%	0.9116	+1.47%
Phase + correlation	0.7263	+3.69%	0.9000	+2.63%
Phase + PCE	0.7263	+3.69%	0.9000	+2.63%
Proposed Method	**0.7632**	-	**0.9263**	-

Table 4.6 The source camera classification accuracy

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	38	2	0	1	0	0	1	0	2	1	0	1	1	0	2	0	1	0	0
2	0	40	2	1	1	1	2	1	0	1	1	0	0	0	0	0	0	0	0
3	0	1	37	0	1	1	0	3	2	1	0	1	0	1	1	0	0	0	1
4	0	3	3	33	0	1	0	0	0	2	1	1	0	0	2	1	1	2	0
5	3	0	1	2	28	0	1	1	1	1	4	1	1	1	1	2	1	1	0
6	4	1	2	1	4	31	3	2	2	3	1	3	1	1	3	2	2	1	1
7	0	2	2	1	1	2	36	0	1	1	1	0	1	1	0	1	0	0	0
8	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0
10	1	0	0	0	2	4	1	1	2	28	1	2	1	1	4	1	0	0	1
11	0	1	3	1	1	2	1	2	3	2	26	0	0	1	4	2	1	0	0
12	0	0	0	1	0	2	0	0	1	0	0	34	2	1	0	3	0	6	0
13	0	1	2	0	0	0	0	1	0	0	0	2	38	0	0	3	2	1	0
14	2	0	1	0	0	0	0	1	0	2	0	0	0	42	1	0	0	0	1
15	0	2	0	0	0	0	0	0	0	1	1	0	0	1	44	0	0	1	0
16	1	1	0	0	1	0	0	0	0	0	0	1	0	0	0	42	1	2	1
17	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	1	43	1	2
18	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	2	2	42	2
19	0	0	0	1	0	0	0	0	0	0	0	2	1	0	0	1	0	1	44

Table 4.7 Source Classification Matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Basic + correlation	36	41	34	28	23	27	35	**50**	**50**	17	24	33	35	38	**44**	44	**45**	40	39
Basic + PCE	37	41	34	29	23	28	33	**50**	**50**	17	25	**35**	37	38	43	44	44	40	39
MLE + correlation	34	36	32	31	25	27	**36**	**50**	49	25	23	33	34	**42**	**44**	42	42	40	**45**
MLE + PCE	35	41	35	30	23	28	34	**50**	49	22	23	32	36	40	**44**	**43**	44	41	43
Model3 + correlation	35	41	34	30	22	28	34	**50**	49	21	24	32	33	41	**44**	**43**	44	41	44
Model3 + PCE	34	38	35	31	25	28	**36**	**50**	**50**	26	23	34	36	41	**44**	42	41	40	43
Phase + correlation	35	42	32	29	26	**33**	**36**	**50**	**50**	19	**28**	29	31	36	43	41	**45**	**42**	43
Phase + PCE	35	**43**	32	29	26	**33**	**36**	**50**	**50**	19	27	29	31	36	43	41	45	**42**	43
Proposed Method	**38**	40	**37**	**33**	**28**	31	**36**	**50**	**50**	**28**	26	34	**38**	**42**	**44**	42	43	**42**	44

Table 4.8 Number of correct classification for each camera

	FPR=10-3	FPR=10-2
	TPR	TPR
Basic + correlation	0.42421	0.59263
Basic + PCE	0.44316	0.61474
MLE + correlation	0.44784	0.60906
MLE + PCE	0.44000	0.63895
Model3 + correlation	0.41728	0.61855
Model3 + PCE	0.46842	0.64842
Phase + correlation	0.47684	0.62316
Phase + PCE	0.48105	0.62211
Proposed Method (Gaussian Kernel)	0.49789	0.65368
Proposed Method (Joint Bilateral Kernel)	0.51684	0.68421

Table 4.9 Performance comparison between Gaussian Kernel and Bilateral Kernel for 128×128 images

Method	Basic	Li's method	Phase method	Proposed Scheme 1	Proposed Scheme 2
Speed (s/image)	0.361612	0.363233	0.366265	0.364121	0.413255
Complexity Gain compared with Basic	0%	0.45%	1.29%	0.6938%	14.2813%

Table 4.10 Computation time for the different approaches

	FPR=0.05	FPR=0.01	FPR=0.001
Basic	0.8925	0.7667	0.4866
MLE	0.9183	0.8525	0.6016
Phase	0.9125	0.8575	0.5850
Model3	0.9000	0.8425	0.6050
Model5	0.8592	0.7733	0.5200
Proposed scheme 1	0.9150	0.8550	0.6733
Proposed scheme 2	0.9333	0.8933	0.6667

Table 4.11 Performance comparison of the two scheme proposed

4.5 Chapter Summary

Though PRNU has been proved as an effective means for source camera identification, the scene content artifact can severely deteriorate the performance of PRNU-based camera identification. The detection accuracy will be low if the image contains dark areas and complicated texture, the image size is small and the quality of image is bad. In this paper, we have studied the pixel reliability of detection over the local variance

and the noise residual magnitude. We use the local variance of each pixel to estimate the distribution of the undesired noise signal. With the estimated distribution, the general matched filter which is an optimal detector theoretically is incorporated into the peak to correlation detector to perform the source camera identification tasks. The proposed method is compared with several state-of-art methods. The experiments show that the proposed method outperformed other state-of-the-art methods in terms of the ROC curve and detection accuracy for both source camera identification and classification problems.

Chapter 5 Conclusion and future work

5.1 Conclusion

While the advancement of technology brings countless benefit to the daily life of modern people, several pressing issues also come to surface due to the dramatic change of our life style. The digital image forensics is becoming increasingly important in the current age for three reasons. Firstly, the widespread use of digital imaging devices makes digital images more frequently presented in the court as pieces of evidence. Secondly, the dedicated image editing software allows a layman to modify the digital images without leaving any obvious traces. Last but not least, the digital images differ fundamentally from the traditional images in the way they are taken and stored which makes the conventional laws/practices for evidence admission not valid for digital images.

The Photo Response Non-uniformity (PRNU) has recently emerged as the most powerful tool for source camera identification. The PRNU based method is capable of identifying the source device with a high accuracy while most of the other methods can only identifying the model of the source camera. The PRUN based method can also be utilized to perform image forgery detection. The detection accuracy of PRNU based source camera identification method depends on the size of the image investigated. If the resolution of the testing image is low, the detection statistics obtained by the correlation detector will be not reliable. Improving the detection accuracy for small images is important because the images may be cropped or resized. It is also important for the forgery detection in that if a certain portion of the image is believed to be forged, the PRNU-based detection need to be performed in a small area

of the image. Apart from the resolution of the image, another factor that influences the accuracy of PRNU-based detection is the scene content. Due to the non-ideal property of the image denoising filter, the noise residual which contains the PRNU signal might be contaminated by the scene artifact if the testing image has complicated scene content. This thesis, therefore, studies how the scene content affects the performance of PRNU-based source camera identification and proposes some methods to alleviate the problem.

The scene content effects on the accuracy of PRNU based camera identification are analyzed. The PRNU-based source camera identification method can be formulated as a two hypothesis signal detection problem. H_0 represents the query image not coming from the camera under test. H_1 represents query image that is taken by the camera under test. It has been shown that the distribution of the correlation under H_1 is significantly influenced by the texture of the image. The mean correlation of the image block decreases with the variance of the image while the standard deviation of the correlation increases with the variance. Both changes increase the overlap between the distribution of the correlation under H0 and that under H1. Hence false identification is more likely to occur if an image contains more textures in its content. However, the relation between the texture and reliability of the image is very hard to model due to the complexity of denoising filters. Therefore, in Chapter 3, the artificial neural network is proposed to model the relation between the reliability of the image and different features. It has been proved as a universal approximator for different functions. The neural network is trained to improve the reliability of an image for PRNU-based detection by allocating the optimal weightings to each block of the image.

Though the proposed neural network improves the performance of PRNU based source camera identification, the training phase requires a large number of data which may not be accessible under some cases. Therefore, in Chapter 4, a simple method, that utilizes the characteristics of the noise residual to determine the weightings of each individual pixel of the query images, is proposed. The relation between image features and it is reliability is more complex to model since some texture patterns can be removed effectively by the denoising filter while some texture patterns cannot. Instead of using the image features, the noise residual features can better characterize the reliability of the image. The regions with large noise residual variance are less reliable for PRNU-based detection. Baseed on the general matched filter which is proved to be the optimal detector, we have derived the model which allocates the weighting for each individual pixels based on the local variance of the noise residual. In conclusion, this thesis studies the scene content problem of existing PRNU-based camera identification methods and two algorithms are proposed to resolve the problem. The first one is to build a neural network to predict the weightings of different blocks and the second method is to estimate the variance of the undesired artifacts and compute the optimal weightings based on the general matched filter. Experimental results show that the proposed methods can achieve better identification results as compared with the state-of-art methods.

5.2 Future works

The face-based authentication system has gained increasing popularity in recent years in the access control application. This kind of system can be used for security check, mobile phone unlocking and mobile payment. As a biometric authentication system similar to the fingerprint authentication system, the face recognition system does not require any additional sensor since all the smart phones are equipped with the digital image sensors. However, the face recognition system is vulnerable to the face spoof attacks. The face spoof attack refers to the attempts of using the copied photos or videos played on display devices like the tablet to deceive the face recognition system. Since it is relatively easy to gain access to the personal face images or videos, the security level of the face recognition will be greatly undermined by the face spoof attacks. Studies [62] have shown that the state-of-art Commercial Off-the-Shelf (COTS) face recognition system has poor capability of detecting the face spoof attacks. Several methods have been proposed in the literature to deal with the face spoof attacks. These methods can be classified into four categories i.e. motion based method[61][63][64][65], texture based methods[66][67][68][69], image quality analysis based methods [70]and methods based on other cues[71][72][73]. The state-of-art method [62] can achieves detection rate of 86.7% - 94.7% accuracy at a false alarm rate at 10%. The accuracy is not very high and it still has some room for improvement.

The printed photos or displayed videos might bring an additional noise to the face recognition camera due to the imperfection of printer and displaying device. This noise will in turn contaminate the PRNU signal and decrease the correlation between

the noise residual of spoof face image and the reference PRNU. Some preliminary experiments have been done on the camera of a LG G3 smart phone to verify the hypothesis.

Image No	PCE for Natural Image	PCE for corresponding Printed Image
1	752.8186	2.0642e+003
2	813.7045	3.6590e+003
3	1.1552e+003	1.6977e+003
4	3.5437e+003	2.5982e+003
5	4.1478e+003	3.2444e+003
6	4.5767e+003	3.9264e+003
7	4.4322e+003	4.4125e+003
8	8.5010e+003	5.4183e+003
9	6.3818e+003	3.9956e+003
10	1.0119e+004	6.7510e+003
11 (blue sky)	3.5230e+004	7.1038e+003

Table 5.1 PCE for natural images and it printed version

Figure 5.1 Examples of Natural Image (left) and its printed version (right)

In the first experiment, ten images of the natural scene and a blue sky image are selected and printed out. The PCE for the natural images and printed images are calculated. Table 5.1 shows their PCE values. It can be seen that for the images with large PCE, the PCE drops obviously. For the images with small PCE, the PCE may increase. However, the experiment is considered to be biased because the lighting condition is different for images in comparison. The printed images are taken indoors with a good lighting condition. Some of the natural scene images are taken in poor lighting condition.

Therefore the second experiment was carried on. In this experiment, the face recognition system was simulated. Ten face images were taken with a fixed background (smooth wall). Then one face images are printed out (all the face images are similar). The printed face images are placed in the same place where the face images are taken. Ten images containing the original background are also taken for comparison. Example of images is shown in Figure 5.1. In this way, all the images are guaranteed to be taken in the same lighting condition. The result of PCE value is shown in Table 5.2. It can be seen that the PCE for the background is the highest. The PCE for natural face image is lower than that of the background because the texture of

the face will reduce the correlation value. The PCE drops for over 200 for the printed face image as compared with the live face image. To further verify that the decrease of correlation is caused by the noise of printer, the variance of the noise residual for both the live image and spoof image were calculated. For the whole image the variances of the noise residual for live and spoof image is 2.33 and 2.56 respectively. For the smooth regions, the variance becomes 2.17 and 2.56 respectively. This data suggests that the printed image has a higher noise level as compared with the live captured one. The experiment shows that the additional noise caused by printing will decrease the correlation statistics in the PRNU-based detection process.

Therefore, for the live face detection of an authentication system, a PCE predictor can be built with machine learning algorithms using image features such as the intensity and texture as input features. Since the PCE predictor is built under the assumption that image is live captured, the PCE value will be lower than the predicted value if a printed image is used to deceive the face recognition system. Given that the camera is fixed and under the control of system owner, the predict value can be of high accuracy and the statistical distribution of predicting error can also be established. Then with Neyman Pearson theory, a threshold can be determined to achieve a certain FRR, e.g. 10^{-3}. If the PCE of an image is smaller than the threshold, it can be decided that it is not a live capture. To the best of our knowledge, the proposed method differs fundamentally to all the existing methods in the literature. Besides it is also possible to combine PRNU based features with other features such as the image distortion analysis features used in [62] to further improve the detection performance.

In summary, the face spoof attacks may affect the detection of PRNU signal of the captured image because of the additional noise introduced by the printing or displaying

device. Such interference can be utilized to detect the face spoof attacks. A face image can be considered not a live capture if the correlation between its noise residual and the reference value is lower than the value it should be.

Image No	PCE for background(Wall)	PCE for Natural Face image	PCE for printed Face image
1	1.3613e+004	1.1412e+004	8.4559e+003
2	1.3179e+004	1.0335e+004	7.7814e+003
3	1.2506e+004	1.0488e+004	7.3662e+003
4	1.2907e+004	1.0487e+004	7.9190e+003
5	1.3058e+004	1.0589e+004	7.9751e+003
6	1.3615e+004	1.0350e+004	7.8215e+003
7	1.3030e+004	1.0221e+004	8.0317e+003
8	1.3461e+004	1.0466e+004	7.5776e+003
9	1.2541e+004	1.0275e+004	7.6398e+003
10	1.3241e+004	1.0084e+004	8.0876e+003

Table 5.2 PCE for background, face image and printed face images

Appendix A

Proof of Neyman – Pearson Theorem

Neyman – Pearson Theorem

When performing a binary hypothesis test between H_0 and H_1, to maximize the detection rate P_D for a given false alarm rate $P_{FA} = \alpha$, H_1 is decided if

$$L(x) = \frac{p(x;H_1)}{p(x,H_0)} > \gamma$$

where the threshold γ is found from

$$P_{FA} = \int_{\{x:L(x)>\gamma\}} p(x;H_0)dx = \alpha$$

Proof:

The proof of Neyman – Pearson Theorem can be found in [74] and is shown here.

To maximize the detection rate P_D for a given P_{FA}, the Lagrangian multipliers can be used which is expressed as,

$$\begin{aligned} F &= P_D - \lambda(P_{FA} - \alpha) \\ &= \int_{R_1} p(x;H_1)dx - \lambda(\int_{R_1} p(x;H_0)dx - \alpha) \\ &= \int_{R_1} (p(x;H_1) - \lambda p(x;H_0))dx + \lambda\alpha \end{aligned}$$

where R_1 is the range of x that corresponds to H_1, λ is the Lagrangian multiplier and α is the false alarm rate. To maximize F, R_1 should contain all the x values that make $(p(x;H_1) - \lambda p(x;H_0)) > 0$. Hence for any value λ the region R_1 that maximize F can be defined as,

$$R_1(\lambda) = \{\mathbf{r} \mid (p(\mathbf{r};H_1) - \lambda p(\mathbf{r};H_0)) > 0\}$$

$$R_1(\lambda) = \{\mathbf{r} \mid \frac{p(\mathbf{r};H_1)}{p(\mathbf{r};H_0)} > \lambda\}$$

Therefore, H_1 can be decided if the following condition is satisfied,

$$L(\mathbf{r}) = \frac{p(\mathbf{r}; H_1)}{p(\mathbf{r}; H_0)} > \lambda$$

where $L(\mathbf{r})$ is the likelihood ratio which is always non-negative and the value of λ should be larger than zero. Otherwise, H_1 will be always be decided and the false alarm rate P_{FA} will be 1. The value of λ can be found from,

$$P_{FA} = \int_{R_1(\lambda)} p(x; H_0) dx = \alpha$$

Therefore, let $\gamma = \lambda$, the Neyman – Pearson Theorem is proved.

Appendix B

Wavelet Denoising Filter

The wavelet denoising filter proposed in [41] is used for PRNU feature extraction in this thesis. This filter has been reported in [7] to have the best performance among several denoising filters.

The wavelet filter [41] models the high frequency wavelet coefficients as an additive mixture of locally stationary noise i.e. signal with zero mean (the noise free image) and a stationary white Gaussian noise $N(0, \sigma_0^2)$. The local image variance is estimated firstly. Then a Wiener filter is used to obtain an estimate of the denoised image in the wavelet domain. The details are described as follows.

1. Compute the four level wavelet decomposition of the noisy image with the 8-tap Daubechies Quadrature Mirror Filters (QMF). For each level, the high frequency subbands, i.e. vertical, horizontal and diagonal subbands are denoted respectively as *h(i,j), v(i,j), d(i,j)* in which *i* and *j* are the coefficient index for the corresponding decomposition level.
2. Maximum a Posteriori (MAP) estimation is used to estimate the local variance of the noise-free image for all the wavelet coefficients in different subbands. The local variance is estimated with four different window sizes i.e. 3 × 3, 5 × 5, 7 × 7 and 9 × 9. Let the window size be *W* × *W*. The local variance in the horizontal subband can be estimated by,

$$\hat{\sigma}_W^2 = \max(0, \frac{1}{W^2}\sum_{(i,j)\in J} h^2(i,j) - \sigma_0^2), (i,j) \in J$$

where *J* is the index set for the corresponding subband. The minimum of the four estimated variance are selected as the final estimate i.e.

$$\hat{\sigma}^2(i,j) = \min(\hat{\sigma}_3^2(i,j), \hat{\sigma}_5^2(i,j), \hat{\sigma}_7^2(i,j), \hat{\sigma}_9^2(i,j))$$

3. Use the Wiener filter to remove the noise in the wavelet domain.

$$h_{den}(i,j) = h(i,j)\frac{\hat{\sigma}^2(i,j)}{\hat{\sigma}^2(i,j) + \sigma_0^2}$$

   For *d(i, j)* and *v(i, j),* the denoised coefficients are calculated in the same way.

4. Repeat 1-3 for color channel and each wavelet decomposition level. The denoised wavelet coefficients are transformed back to the spatial domain and form the denoised image.

The parameter σ_0 which is the standard deviation of the stationary white Gaussian noise is set to 5 as suggested in [7].

References

[1] Sony Corporation, "**Annual report 2012**," Sony Corporation, available at http://www.sony.net/SonyInfo/IR/library/ar/2012/common/docs/EAR.pdf;, Japan, 2012.

[2] P. Blythe and J. Fridrich. Secure digital camera. Presented at Digital Forensic Research Workshop. 2004, .

[3] P. W. Wong, "A watermark for image integrity and ownership verification," in *IS AND TS PICS CONFERENCE*, 1998, pp. 374-379.

[4] H. T. Sencar and N. Memon, "Overview of state-of-the-art in digital image forensics," *Algorithms, Architectures and Information Systems Security*, vol. 3, pp. 325-348, 2008.

[5] K. Mehdi, H. T. Sencar and N. Memon. Blind source camera identification. Presented at Image Processing, 2004. ICIP'04. 2004 International Conference on. 2004, .

[6] S. Lyu and H. Farid. Detecting hidden messages using higher-order statistics and support vector machines. Presented at Information Hiding. 2003, .

[7] J. Lukas, J. Fridrich and M. Goljan. Digital camera identification from sensor pattern noise. *Information Forensics and Security, IEEE Transactions on 1(2)*, pp. 205-214. 2006.

[8] J. Fridrich, "Digital image forensics," *Signal Processing Magazine, IEEE,* vol. 26, pp. 26-37, 2009.

[9] K. Rosenfeld and H. T. Sencar, "A study of the robustness of prnu-based camera identification," in *IS&T/SPIE Electronic Imaging*, 2009, pp. 72540M-72540M-7.

[10] M. Goljan, J. Fridrich and J. Lukáš, "Camera identification from printed images," in *Electronic Imaging 2008,* 2008, pp. 68190I-68190I-12.

[11] E. J. Alles, Z. J. Geradts and C. J. Veenman, "Source camera identification for low resolution heavily compressed images," in *Computational Sciences and its Applications, 2008. ICCSA'08. International Conference on,* 2008, pp. 557-567.

[12] M. Chen, J. Fridrich, M. Goljan and J. Lukás. Determining image origin and integrity using sensor noise. *Information Forensics and Security, IEEE Transactions on 3(1),* pp. 74-90. 2008.

[13] G. Chierchia, D. Cozzolino, G. Poggi, C. Sansone and L. Verdoliva, "Guided filtering for PRNU-based localization of small-size image forgeries," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on,* 2014, pp. 6231-6235.

[14] G. Chierchia, G. Poggi, C. Sansone and L. Verdoliva, "PRNU-based forgery detection with regularity constraints and global optimization," in *Multimedia Signal Processing (MMSP), 2013 IEEE 15th International Workshop on,* 2013, pp. 236-241.

[15] G. Chierchia, G. Poggi, C. Sansone and L. Verdoliva, "A Bayesian-MRF approach for PRNU-based image forgery detection," *Information Forensics and Security, IEEE Transactions on,* vol. 9, pp. 554-567, 2014.

[16] H. T. Sencar and N. Memon, "Digital image forensics," 2013.

[17] C. Li, "Source camera identification using enhanced sensor pattern noise," *Information Forensics and Security, IEEE Transactions on,* vol. 5, pp. 280-287, 2010.

[18] S. McCloskey, "Confidence weighting for sensor fingerprinting," in *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on,* 2008, pp. 1-6.

[19] L. Chan, N. Law and W. Siu, "A confidence map and pixel-based weighted correlation for PRNU-based camera identification," *Digital Investigation,* vol. 10, pp. 215-225, 2013.

[20] C. Shi, N. Law, H. Leung and W. Siu, "Weighting optimization with neural network for photo-response-non-uniformity-based source camera identification," in *Asia-Pacific Signal and Information Processing Association, 2014 Annual Summit and Conference (APSIPA),* 2014, pp. 1-7.

[21] X. Kang, J. Chen, K. Lin and P. Anjie, "A context-adaptive SPN predictor for trustworthy source camera identification," *EURASIP Journal on Image and Video Processing,* vol. 2014, pp. 1-11, 2014.

[22] X. Kang, Y. Li, Z. Qu and J. Huang, "Enhancing source camera identification performance with a camera reference phase sensor pattern noise," *Information Forensics and Security, IEEE Transactions on,* vol. 7, pp. 393-402, 2012.

[23] I. Amerini, R. Caldelli, V. Cappellini, F. Picchioni and A. Piva, "Analysis of denoising filters for photo response non uniformity noise extraction in source camera identification," in *Digital Signal Processing, 2009 16th International Conference on,* 2009, pp. 1-7.

[24] G. Chierchia, S. Parrilli, G. Poggi, C. Sansone and L. Verdoliva. On the influence of denoising in PRNU based forgery detection. Presented at Proceedings of the 2nd ACM Workshop on Multimedia in Forensics, Security and Intelligence. 2010, .

[25] M. Kubat, *Neural Networks: A Comprehensive Foundation by Simon Haykin, Macmillan, 1994, ISBN 0-02-352781-7.*, 1999.

[26] T. Van Lanh, K. Chong, S. Emmanuel and M. S. Kankanhalli, "A survey on digital camera image forensic methods," in *Multimedia and Expo, 2007 IEEE International Conference on,* 2007, pp. 16-19.

[27] K. San Choi, E. Y. Lam and K. K. Wong, "Source camera identification using footprints from lens aberration," in *Electronic Imaging 2006,* 2006, pp. 60690J-60690J-8.

[28] Z. J. Geradts, J. Bijhold, M. Kieft, K. Kurosawa, K. Kuroki and N. Saitoh, "Methods for identification of images acquired with digital cameras," in *Enabling Technologies for Law Enforcement*, 2001, pp. 505-512.

[29] S. Bayram, H. Sencar, N. Memon and I. Avcibas, "Source camera identification based on CFA interpolation," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, 2005, pp. III-69-72.

[30] M. Tsai and G. Wu, "Using image features to identify camera sources," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, 2006, pp. II-II.

[31] J. R. Janesick, *Scientific Charge-Coupled Devices.* SPIE press Bellingham, Washington, 2001.

[32] G. C. Holst, *CCD Arrays, Cameras, and Displays.* JCD publishing, 1998.

[33] J. Luk, J. Fridrich and M. Goljan. Detecting digital image forgeries using sensor pattern noise. Presented at Proceedings of the SPIE. 2006, .

[34] G. E. Healey and R. Kondepudy, "Radiometric CCD camera calibration and noise estimation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, pp. 267-276, 1994.

[35] M. Chen, J. Fridrich, J. Lukáš and M. Goljan, "Imaging sensor noise as digital x-ray for revealing forgeries," in *Information Hiding*, 2007, pp. 342-358.

[36] A. El Gamal, B. A. Fowler, H. Min and X. Liu, "Modeling and estimation of FPN components in CMOS image sensors," in *Photonics West'98 Electronic Imaging*, 1998, pp. 168-177.

[37] M. Chen, J. Fridrich and M. Goljan, "Digital imaging sensor identification (further study)," in *Electronic Imaging 2007*, 2007, pp. 65050P-65050P-13.

[38] M. Goljan, J. Fridrich and T. Filler, "Large scale test of sensor fingerprint camera identification," in *IS&T/SPIE Electronic Imaging*, 2009, pp. 72540I-72540I-12.

[39] M. Goljan, "Digital camera identification from images–Estimating false acceptance probability," in *Digital Watermarking*Anonymous Springer, 2009, pp. 454-468.

[40] J. A. Dominguez-Molina, G. González-Farías, R. M. Rodríguez-Dagnino and I. C. Monterrey, "A practical procedure to estimate the shape parameter in the generalized Gaussian distribution," *Technique Report I-01-18_eng.Pdf, Available through Http://www.Cimat.mx/reportes/enlinea/I-01-18_eng.Pdf*, vol. 1, 2003.

[41] M. K. Mıhçak, I. Kozintsev and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, 1999, pp. 3253-3256.

[42] F. Argenti, G. Torricelli and L. Alparone, "MMSE filtering of generalised signal-dependent noise in spatial and shift-invariant wavelet domains," *Signal Process,* vol. 86, pp. 2056-2066, 2006.

[43] A. J. Cooper, "Improved photo response non-uniformity (PRNU) based source camera identification," *Forensic Sci. Int.*, vol. 226, pp. 132-141, 2013.

[44] Y. Hu, B. Yu and C. Jian, "Source camera identification using large components of sensor pattern noise," in *Computer Science and its Applications (CSA'09). 2nd International Conference on,* 2009, pp. 1-5.

[45] F. Gharibi, F. Akhlaghian, J. RavanJamjah and B. ZahirAzami, "Using the local information of image to identify the source camera," in *Signal Processing and Information Technology (ISSPIT), 2010 IEEE International Symposium on,* 2010, pp. 515-519.

[46] B. Liu, Y. Hu and H. Lee, "Source camera identification from significant noise residual regions," in *Image Processing (ICIP), 2010 17th IEEE International Conference on,* 2010, pp. 1749-1752.

[47] R. Li, C. Li and Y. Guan, "A compact representation of sensor fingerprint for camera identification and fingerprint matching," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on,* 2015, pp. 1777-1781.

[48] Y. Li and C. Li, "Decomposed photo response non-uniformity for digital forensic analysis," in *Forensics in Telecommunications, Information and Multimedia*Anonymous Springer, 2009, pp. 166-172.

[49] L. C. Jain and N. Martin, *Fusion of Neural Networks, Fuzzy Systems and Genetic Algorithms: Industrial Applications.* CRC press, 1998.

[50] R. Storn and K. Price, "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces," *J. Global Optimiz.,* vol. 11, pp. 341-359, 1997.

[51] J. Vesterstrom and R. Thomsen, "A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems," in *Evolutionary Computation, 2004. CEC2004. Congress on,* 2004, pp. 1980-1987 Vol. 2.

[52] L. Lakshminarasimman and S. Subramanian, "Applications of Differential Evolution in Power System Optimization," *Advances in Differential Evolution,* pp. 257-273, 2008.

[53] R. Storn, "Differential evolution design of an IIR-filter," in *Evolutionary Computation, 1996., Proceedings of IEEE International Conference on,* 1996, pp. 268-273.

[54] R. Storn, "Designing nonstandard filters with differential evolution," *Signal Processing Magazine, IEEE,* vol. 22, pp. 103-106, 2005.

[55] J. Lampinen and I. Zelinka, "Mechanical engineering design optimization by differential evolution," in *New Ideas in Optimization,* 1999, pp. 127-146.

[56] T. Gloe and R. Böhme, "The dresden image database for benchmarking digital image forensics," *Journal of Digital Forensic Practice,* vol. 3, pp. 150-159, 2010.

[57] G. L. Turin, "An introduction to digitial matched filters," *Proc IEEE,* vol. 64, pp. 1092-1112, 1976.

[58] S. M. Kay, "Fundamentals of statistical signal processing, Vol. II: Detection Theory," *Signal Processing.Upper Saddle River, NJ: Prentice Hall,* 1998.

[59] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Computer Vision, 1998. Sixth International Conference on,* 1998, pp. 839-846.

[60] L. Caraffa, J. Tarel and P. Charbonnier, "The Guided Bilateral Filter: When the Joint/Cross Bilateral Filter Becomes Robust," *Image Processing, IEEE Transactions on,* vol. 24, pp. 1199-1208, 2015.

[61] Q. Yang, K. Tan and N. Ahuja, "Real-time O (1) bilateral filtering," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on,* 2009, pp. 557-564.

[62] D. Wen, H. Han and A. K. Jain, "Face spoof detection with image distortion analysis," *Information Forensics and Security, IEEE Transactions on,* vol. 10, pp. 746-761, 2015.

[63] L. Sun, G. Pan, Z. Wu and S. Lao, "Blinking-based live face detection using conditional random fields," in *Advances in Biometrics*Anonymous Springer, 2007, pp. 252-260.

[64] S. Bharadwaj, T. I. Dhamecha, M. Vatsa and R. Singh, "Computationally efficient face spoofing detection with motion magnification," in *Computer Vision*

and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on, 2013, pp. 105-110.

[65] K. Kollreider, H. Fronthaler, M. I. Faraj and J. Bigun, "Real-time face detection and motion analysis with application in “liveness” assessment," *Information Forensics and Security, IEEE Transactions on*, vol. 2, pp. 548-558, 2007.

[66] J. Yang, Z. Lei, S. Liao and S. Z. Li, "Face liveness detection with component dependent descriptor," in *Biometrics (ICB), 2013 International Conference on*, 2013, pp. 1-6.

[67] T. de Freitas Pereira, A. Anjos, J. M. De Martino and S. Marcel, "Can face anti-spoofing countermeasures work in a real world scenario?" in *Biometrics (ICB), 2013 International Conference on*, 2013, pp. 1-8.

[68] I. Chingovska, A. Anjos and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of the*, 2012, pp. 1-7.

[69] T. de Freitas Pereira, A. Anjos, J. M. De Martino and S. Marcel, "LBP− TOP based countermeasure against face spoofing attacks," in *Computer Vision-ACCV 2012 Workshops*, 2013, pp. 121-132.

[70] J. Galbally, S. Marcel and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," *Image Processing, IEEE Transactions on*, vol. 23, pp. 710-724, 2014.

[71] T. Wang, J. Yang, Z. Lei, S. Liao and S. Z. Li, "Face liveness detection using 3d structure recovered from a single camera," in *Biometrics (ICB), 2013 International Conference on*, 2013, pp. 1-6.

[72] J. Komulainen, A. Hadid and M. Pietikainen, "Context based face anti-spoofing," in *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, 2013, pp. 1-8.

[73] G. Chetty, "Biometric liveness checking using multimodal fuzzy fusion," in *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*, 2010, pp. 1-8.

[74] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. John Wiley & Sons, 2004.