The Hong Kong Polytechnic University

Department of Electronic and Information Engineering

Digital Camera Identification for Forensics Applications

Chan Lit Hung

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

August 2012

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)
_____ (Name of student)

Dedication

The original contributions reported in this thesis are as follows:

1. Performance evaluation for different types of images (Chapter 3, Section 3.4).

   We have performed a detailed study on the effect of image content on camera identification. Using a two-dimensional Gaussian model, the identification accuracy can be obtained for different types of image content. This study helps to quantify the seriousness of image content effect and relates that to the identification accuracy.

2. Two dimensional classification method (Chapter 3, Section 3.5).

   As the threshold used in the binary hypothesis test for camera identification depends on the image content, we introduced a 2D classifier to give a flexible threshold setting according to the image content. In particular, the predicted correlation is used in the 2D classifier to show the seriousness of the image content effect so that different thresholds are set for different types of image content. Experimental results show that the 2D classifier gives better identification accuracy than the traditional identification methods.

3. Confidence map and a pixel-based weighted correlation (Chapter 4, Section 4.4).

   We have extended the image content effect characterization from block level to pixel level. A non-linear regression model using the kernel principal component analysis is used to formulate the image content effect. Then, a confidence map is generated to indicate the reliability of each pixel in PRNU

estimation. By using the confidence map as a weighting function, a large weighting is assigned to reliable pixels and vice versa. Experimental results show that the proposed confidence map is able to achieve an accurate camera identification result.

Abstract

With the large number of digital imaging devices nowadays, the use of digital photos in court keeps on increasing. One might want to authenticate the origin of the photo (i.e., whether it is downloaded or produced from a certain camera). Source camera identification thus becomes important in digital forensic applications. Camera identification can generally be classified into two types, namely source model identification and individual source camera identification. Both of them try to extract device signature and check whether that signature can be found in a given photo. But the former can only determine the brand and the model of the camera while the latter can uniquely identify each individual camera. The focus of this thesis is on the individual source camera identification.

Existing source camera identification methods use a type of pattern noise called the photo response non-uniformities (PRNU) noise. It is caused by manufacturing imperfections and is presented in every image taken by the device. The PRNU is extracted through image denoising. In particular, it is obtained as the difference between the original image and its denoised version. One major problem is that the PRNU can be affected by image content. For example, the PRNU is completely absent in saturated area. Previous studies have found that the image content can seriously affect the identification accuracy. The objective of this study is to

investigate ways to compensate for the scene content effect in PRNU estimation.

A possible solution to deal with the image content is to use the correlation predictor. It tries to quantify the seriousness of the scene content effect on the pattern noise. We have performed a detailed study on the relation between the predicted correlation and the identification accuracy. Using 2D Gaussian modeling, the identification accuracy can be obtained for different image content as characterized by the predicted correlation. Using this result, a 2D classifier is proposed for individual source camera identification. The 2D classifier uses the predicted correlation as one of the features to quantify the scene content effect. It helps setting different correlation thresholds for different types of image content. Experimental result shows the identification accuracy increases by about 4% as compared to the traditional identification methods.

The correlation predictor is able to characterize the image content effect in a block-based manner only. We extended the characterization to a pixel level. In particular, a non-linear regression model is used to formulate the scene content effect. Then, a confidence map is generated which indicates the reliability of each pixel in the PRNU estimation. By using the confidence map as a weighting function in correlation calculation, the scene content effect can be compensated. Experimental results show that the proposed confidence map is able to achieve an accurate camera

identification result. As compared with state-of-the-art identification methods, our proposed method can achieve about 2%-5% and 4%-20% improvement in detection accuracy at JPEG quality factor of 90 and 70 respectively.

Publications

Conference Paper

1. L.H. Chan, N.F. Law and W.C. Siu, "A two-stage trademark retrieval system with invariant property", Proceedings of the Second APSIPA Annual Summit and Conference, pp280 – 283, Singapore, 14-17 December 2010.
2. L.H. Chan, N.F. Law and W.C. Siu, "A Two Dimensional Camera Identification Method based on Image Sensor Noise', IEEE International Conference on Acoustics, Speech and Signal Processing ,pp. 1741-1744, 25-30 March, 2012.

Journal Paper

1. L.H. Chan, N.F. Law and W.C. Siu, "A Confidence Map and Pixel-Based Weighted Correlation for Camera Identification using PRNU", Digital Investigation. (submitted)

Acknowledgements

Pursuing a postgraduate study is a valuable experience in my life, it is harsh but at the same time is very enjoyable. When I was getting into trouble, there were people supporting me and pushing me to step forward. I would like to give my many thanks to all these people.

First of all, I would like to gratefully and sincerely thank my chief supervisor, Dr. Bonnie N.F. Law and my co-supervisor Prof. W.C. Siu, for their guidance and encouragement. Without their support, this study could not have been finished in present level.

Special thanks are given to Dr. K.O. Cheng, Dr. Y.L. Chan and other colleagues and staffs who accompanied me for my study. They have given me an enjoyable university life.

Finally, I would like to dedicate this work to my family and my girlfriend for their love, patience and support.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Introduction

Due to the rapid development of technologies, digital imaging devices continuously replace their traditional analogue counterparts. In fact, the number of digital imaging devices is increasing all over the world. Even mobile phones nowadays are common to have a digital camera embedded in it. With the large number of digital imaging devices, digital photos can easily be used as an evidence of crime in court. For example, in child pornographic cases, digital photos involved would be one of the major evidences of the crime. In such cases, identifying the camera that took those photos would particularly be useful to prove whether the photos are downloaded or produced by the suspect. Despite that, the reliability of the digital photos is a major issue. One needs to authenticate the origin of the photo and prove that the photo is untampered. This makes digital forensics and digital camera identification getting important in recent years.

Although many digital photos nowadays contain the Exchangeable Image File (EXIF) headers that provide information such as the camera model and the setting used to took the photos, the headers can easily be modified or removed. Furthermore, EXIF only indicates the model of the camera taking the image but not the individual camera. Some more reliable methods should be used to authenticate the source of the

photos and prove the reliability of the digital photos.

Digital watermarking [57] has been proposed as a tool to provide authenticity to digital photos as well as to indicate the source camera. However, this technique is not popular. One of the main reasons is that this approach requires support of the camera manufacturer. The digital camera needs to insert the watermark into the photo at the time when the image is captured. Majority of the digital images nowadays do not contain watermark, and this situation is likely to continue in the future. Thus, passive techniques should be used to authenticate the source camera in the absence of watermark.

In recent years, many researchers started to use image processing techniques for digital camera identification [1-31]. The source camera identification can be roughly classified into two main groups: source model identification [15-31] and individual source camera identification [1-14]. For the source model identification, component forensics [15-31] is often used to examine the signature left in different components along the camera pipeline. For the individual source camera identification, the main trend is to use the sensor noise produced by the image sensor [1-14]. For most forensic applications, knowing the exact source camera is much more important than only knowing the model due to the fact that there are plenty of cameras of same model. As camera identification based on camera sensor noise has an edge on

identifying individual cameras even for the same brand and same model, this thesis focuses on the camera sensor noise for individual source camera identification.

1.2 Problems

The individual source camera identification using pattern noise was based on image denoising technique. The pattern noise is simply the difference between the original image and its denoised version. Recent researches [1-3] show that the pattern noise is capable of characterizing individual camera. Besides, the noise is stable in time and is robust against manipulations such as JPEG compression and filtering. Despite that, one of the main problems in the denoising-based pattern noise estimation is that the scene content was found to be left in the extracted pattern noise. For example, the saturated area (i.e., image intensity reaches its maximum value) or the textured area of the images contaminate the extracted pattern noise. As a result, the identification result could be affected. Despite the scene content is highly related to the identification accuracy, traditional identification method using pattern noise[1] did not consider the scene content of the testing images.

Recent studies [2-3] used the correlation predictor to deal with the scene cotnent effect. The idea is to extract features such as image intensity and texture information from an image. These features are then used to predict the correlation between the

PRNU of the testing image and the reference PRNU from a particular camera. Existing method uses the predicted correlation to provide a weighting mechanism so as to reduce the scene content effect on camera identification. For image region with less scene content effect left in the pattern noise, a larger weighting is applied in the correlation calculation. As for image region with serious content contamination on the pattern noise, a lower weighting is applied. In this way, a reliable signal detection can be achieved. Despite some successes, this block based approach is not flexible enough as the image content within each block can have a large variation which limits the effectiveness of the weighting mechanism.

Another method proposed to deal with the scene content effect is to use the phase component of the pattern noise [4]. The scene content of this phase pattern noise is suppressed as compared with the original pattern noise. Due to the reduction of the scene content effect, the identification performance of the phase pattern noise is better than that of the original pattern noise. Besides modifying the pattern noise, different denoising techniques [35-38] have been proposed to remove the scene content effect in the pattern noise. However, all these approaches use a single feature for the source camera identification which is the correlation value for the threshold setting in the camera identification. The identification performance still has room for further improvement.

1.3 Objectives

Our focus is on the individual source camera identification using sensor pattern noise. As discussed in Section 1.2, one of the main problems in sensor pattern noise for individual source camera identification is the scene content effect on the pattern noise. The objective of this thesis is to investigate methods to compensate for the scene content effect on the source camera identification. We will first demonstrate how the scene content affects the pattern noise estimation and investigate how it relates to the camera identification accuracy. Then, we will investigate two ways to make use of the correlation predictor to incorporate the scene content effect into the camera identification problem. The first approach is to use the correlation predictor as one of the features in a 2D classifier. It helps setting flexible threshold values according to the type of scene content. The second approach is to extend the formulation of the scene content effect to pixel level so that the reliability of the pattern noise for each pixel in an image can be characterized. Using this pixel-level reliability information, a weighted correlation is formulated so that a large weighting is applied to reliable pixels and vice versa.

1.4 Organization of the thesis

The rest of the thesis is organized as follows. In Chapter 2, the image capture

model will first be introduced. Then, existing camera identification methods including lens distortion [27-31] and demosaicing artifacts [15-16] will be followed. Afterwards, the use of pattern noise for source camera identification [1, 2, 4] will be described in details.

In Chapter 3, the importance of the image content in the source camera identification will be introduced. This is followed by the correlation predictor which is used to quantify the characteristics of the image content effect. A detailed study on the relation between the predicted correlation and the camera identification accuracy will be described. Afterwards, our proposed 2D classifier is presented.

Chapter 4 introduces our second approach for compensating the scene content effect on camera identification. The formulation of the scene content effect using a non-linear regression model is given. Then, we will describe how the pixel-level characterization of the scene content effect is generated and used to indicate the reliability of each pixel in calculating the correlation between pattern noises. A conclusion and future direction will be given in Chapter 5. The contribution of the thesis will also be summarized.

Chapter 2 Literature Review

Source camera identification can be classified into two types, namely the source model identification and the individual source camera identification. Source model identification attempts to determine the brand/model of the image source while the individual source camera identification tries to find out the exact source camera used to take the images.

In this chapter, image capture model of digital camera will first be given in Section 2.1. Then, the source camera identification based on features from different parts of in-camera processing will be discussed. In particular, the use of lens distortion for camera identification will be introduced in Section 2.2. Then, the use of demosaicing artifacts will be followed in Section 2.3. Afterwards, the use of camera pattern noise, which is one of the most reliable approaches in this research area, will be described in detail in Section 2.4. Finally, Section 2.5 summarizes different approaches for camera identification.

2.1 Image capture model

Despite that cameras from different manufacturers always have different in-camera processing, the simplified model shown in Figure 2.1 provides an accurate model for digital output from most cameras in use nowadays. Light from the scene passes through the lens, optical filter and color filter array (CFA) before reaching the image sensor. The lens focuses the incident light onto the image sensor such that the image captured is in focus. Then, the optical filter is used to filter some high frequency light such as infrared and ultraviolet to ensure that the camera responds to the light that can be detected by the human visual system. Besides, in many digital cameras, CFA is used in order to capture different color components in a single image sensor. There are two main reasons in using CFA. The first reason is due to the high cost in producing a full resolution sensor for each color component. The second reason is due to the substantial difficulty in perfect matching of different color components to form the true color. After the light passing through the CFA, it illuminates the image sensor. The image sensor is a light sensitive device which converts the photons from the incident light into charge. For every pixel in the image sensor, the charges accumulated within the exposure time are converted to voltage which will be quantized. The sensor output signal then undergoes various in-camera processing such as CFA interpolation, white balancing and gamma correction etc.

In digital camera identification, researchers examined different components that might contain device characterization, and had developed different kinds of identification methodology. Next part of this section will first introduce the use of lens distortion as a feature for camera identification.

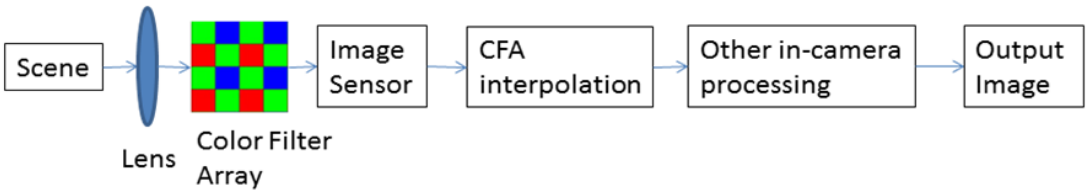


Figure 2. 1 Digital camera output mode.

2.2 Lens distortion

Due to the design and manufacturing process, lens produces radial distortion in the captured image. The radial distortion causes straight lines in the object space rendered as curved lines on camera sensor. There are two main types of radial distortions, namely the barrel distortion and pincushion distortion. Barrel distortion occurs when magnification of scene decreases with off-axis image distance while pincushion distortion occurs when magnification of scene increases with off-axis image distance. Figure 2.2 shows examples of barrel distortion and pincushion distortion.

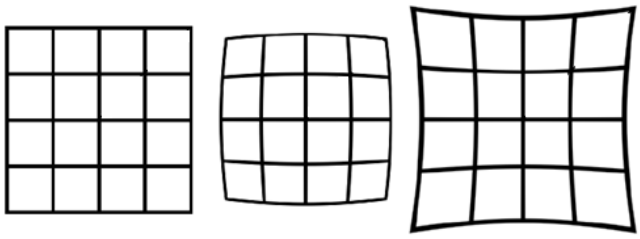


Figure 2. 2 Distortions of a rectangular grid. From left to right: undistorted, barrel distorted and pincushion distorted grids [27].

Studies [27-28] used these kinds of radial distortion as features to identify different camera brands or models. A second order radial symmetric model was used to quantify the distortion as follows,

$$\mathrm{r}_u = r_d + k_1 \mathrm{r}_d^3 + k_2 \mathrm{r}_d^5, \tag{2.1}$$

where r_u and r_d are the undistorted radius and distorted radius respectively, k_1 and k_2 are the first order and second order distortion parameters respectively. The radius is the radial distance $\sqrt{x^2 + y^2}$ of a point (x,y) from the center of distortion where the center of distortion is assumed to be the center of the image.

Study [27] used the straight line method introduced by Devernay [31] to compute the radial distortion. The method is based on a simple property that if a lens does not have radial distortion, every straight line in space should be projected as a straight line onto an image. By using this property, an iterative process can be used to estimate the

distortion parameters k_1 and k_2. These distortion parameters can be solely used as features for the classifier or combined with other statistical image features [23-26] to identify different cameras.

The main restriction of the method is that the captured images must contain straight lines in order to estimate the distortion parameters. Besides, lens from camera are manufacturer dependent. Cameras from the same model or same series may use the same types of lens. In this case, individual camera cannot be identified.

2.3 Color Filter Array (CFA) and demosaicing artifacts

A digital image consists of three color channels, including red, green and blue. However, the digital camera nowadays are usually equipped with a single CCD or CMOS sensor. In such cases, CFA is always used to capture color images using a single imaging sensor. Most CFAs employ three color filters including red, green and blue. The filter pattern is usually 50% green, 25% red and 25% blue as human vision is more sensitive to green color. The most common CFA pattern used among digital cameras is the Bayer pattern which is shown in Figure 2.3.

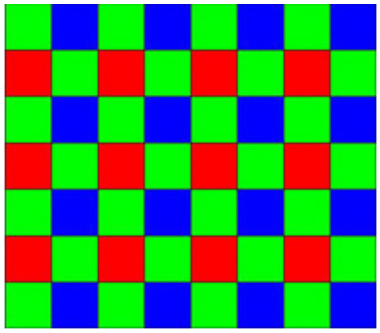


Figure 2. 3 Bayer pattern.

As each pixel of the imaging sensor only contains one color component and the other two color samples are missing, demosaicing or CFA interpolation has to be used to estimate the missing color components using information from neighboring pixels. Simple demosaicing methods treat different color channels separately and employ some kernel-based interpolation methods in each color component such as bilinear or bicubic interpolation [54]. Other sophisticated methods interpolate pixel along the direction of edges to achieve a better visual quality [53]. Regardless of the demosaicing methods used, images using same demosaicing method would introduce similar statistical correlation among pixels. This kind of correlation can be treated as a type of digital signature.

In [15], the authors used the expectationmaximization (EM) algorithm to estimate the interpolation coefficients by re-interpolating digital images after down-sampling. Experimental results show that the estimated interpolation coefficients can help to

distinguish different interpolation algorithms.

As some adaptive CFA interpolation methods may perform differently in different image regions, Swaminathan [16] proposed to divide an image into different types of regions based on the gradient feature. Let $V_{x,y}$ and $H_{x,y}$ be respectively the vertical and horizontal gradient level at a particular pixel (x,y) as follows,

$$H_{x,y} = \left|I_{x,y-2} + I_{x,y+2} - 2I_{x,y}\right| \tag{2.2}$$

$$V_{x,y} = \left|I_{x-2,y} + I_{x+2,y} - 2I_{x,y}\right| \tag{2.3}$$

where $I_{x,y}$ is the image intensity at pixel (x,y). Then for each pixel, the image region R1,R2 and R3 can be classified as follows:

$$\begin{cases} \text{R1: } \left(V_{x,y} - H_{x,y}\right) < -T \\ \text{R2: } \left(V_{x,y} - H_{x,y}\right) > T \\ \text{R3: } -\text{T} < \left(V_{x,y} - H_{x,y}\right) < T \end{cases}, \tag{2.4}$$

where T is a threshold which is chosen experimentally. Region R1 represents the region with significant horizontal gradient. Region R2 represents the region with significant vertical gradient and region R3 represents the region with smooth content.

For each type of region, the authors establish a linear model to solve for the interpolation coefficients. Let the linear model to be,

$$\boldsymbol{Ax=b}, \tag{2.5}$$

where A is the intensity values of neighboring pixels that are used to estimate the interpolation coefficients, x consists of interpolation coefficients and b is the actual

intensity value at a particular pixel . If there are M elements in a particular region type and the number of neighboring pixels used to estimate the interpolation coefficients are N, then A will be in the dimension of $M \times N$, x will be in the dimension of $N \times 1$ and b will be in the dimension of $M \times 1$.

Assuming that, $M \geq N$ which means the number of observations is larger than the unknown needed to be solved, then the interpolation coefficients can be estimated by using least squares method [16] or singular value decomposition (SVD) [18]. The CFA pattern of an image is determined by searching over all valid CFA patterns which minimizes the resulting error. The interpolation coefficients are then put into the support vector machine (SVM) for camera identification. The identification accuracy was larger than 85 percent in identifying images from 16 camera-models under different compression levels [16].

This method is useful for identifying the model of a camera taking a particular image. However, for identifying individual cameras within the same model, the identification methods based on CFA interpolations may not work. In fact, for cameras with same models or even same brands, as long as they used the same CFA interpolation algorithms, the camera identification process would fail. Some other unique characteristics have to be used in order to identify individual cameras.

2.4 Sensor Pattern Noise

There are two types of pattern noises in image sensor, one is the dark signal non-uniformity (DSNU) and the other one is the photo response non-uniformity (PRNU) noises.

The DSNU is caused by dark current and corresponds to pixel to pixel difference without illumination of light. The initial work in using sensor pattern noise for camera identification was done by Kurosawa who used the DSNU for camera identification [11]. In the testing, only blank images with low intensity level were tested as DSNU only survives in dark scene. Despite some success, DSNU is a weak signal. It is not easy to accurately extract the DSNU from natural images which hinders the practical application of DSNU in camera identification.

A reliable camera identification method using sensor pattern noise was first introduced by Lukas [1] who used PRNU instead of DSNU for identification. The PRNU corresponds to pixel-to-pixel difference under illumination of light. It is caused by the inhomogeneity of silicon wafers and manufacturing imperfection [26]. These kinds of imperfections make the images taken by a particular camera containing the same kind of pattern noise which acts like a fingerprint of camera. A further study by Chen [3] used the maximum likelihood approach to estimate the camera fingerprint. Recently a phase pattern noise was proposed by Kang [4] who

used the phase component of the pattern noise for camera identification. For simplicity, in this thesis, the original method by Lucas is called the Basic Approach. The modified versions by Chen and Kang are called the MLE (maximum likelihood estimation) approach and the Phase Approach respectively.

In the next section, the sensor model would first be introduced. Then, the Basic Approach for source camera identification would be followed. Afterwards, the idea of MLE approach and Phase Approach would be discussed. Finally, a chapter summary is given.

2.4.1 The Sensor model

The key component of the individual source camera identification is the sensor pattern noise. Therefore, knowing the sensor model is important to understand the idea and the operation of the identification method. In this section, the sensor output model will first be introduced and then the assumptions of generating a simplified camera output model for identification will be given.

2.4.1.1 Sensor noise and the sensor output model

There are different sources of noise in camera sensor during image capture. Let $Y(i,j)$ and $Z(i,j)$ be the ideal sensor output and actual sensor output from a particular

color channel respectively. Figure 2.4 shows different types of sensor noise in affecting the ideal output signal *Y(i,j)*.

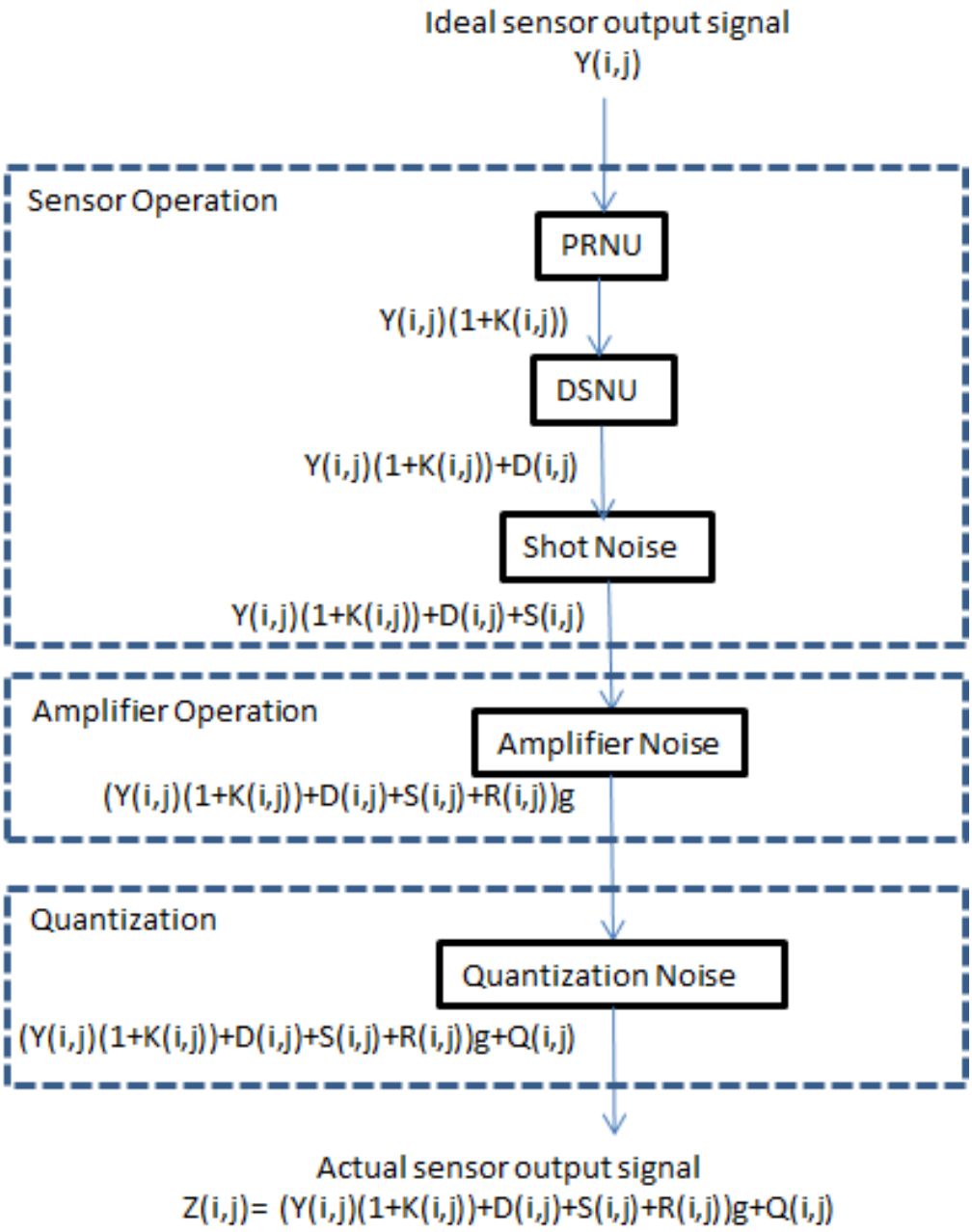


Figure 2. 4 Different types of sensor noise and the sensor output model.

In the image sensor model, the pixel non-uniformities exist in camera output because of the inhomogeneity of silicon wafers and manufacturing imperfection. This kind of imperfection shows up a pixel-to-pixel variation in sensor when responds to the light. The sensor characteristics can be represented by a constant term called PRNU *(K(i,j))* which is fixed but different for each pixel. Hence, the sensor output signal after considering the PRNU becomes,

$$(1+K(i,j))Y(i,j). \tag{2.6}$$

The component of the pattern noise caused by dark currents is called the DSNU. Even in the absence of light, dark currents cause the electrons to accumulate in pixel wells. In general, the charge caused by the dark current consists of a random part and a constant part. The constant part signal is called the DSNU while the random part is called the shot noise. In fact, the dark current varies with temperature and exposure time, but it is independent of the photoelectrons generated. Therefore, the dark current noise can be treated as an additive noise to the sensor output signal. Let *D(i,j)* and *S(i,j)* be the DSNU and the shot noise respectively, then the sensor output signal becomes,

$$(1+K(i,j))Y(i,j)+D(i,j)+S(j,j). \tag{2.7}$$

After the charges accumulated in sensor pixels, they will be transferred to the output amplifier. The amplifier transforms the charge collected at each pixel into a

measurable voltage first and this process will generate an additive noise with zero mean. After that, the amplifier will apply a gain, g, to the signal and this gain factor will also amplify the signal noise. Let the amplifier noise be *R(i,j)*, the sensor output signal becomes,

$$((1+K(i,j))Y(i,j)+D(i,j)+S(i,j)+R(i,j))g. \quad (2.8)$$

The signal passing through the amplifier is subsequently sampled and digitized by the analog to digital converter. The quantization noise *Q(i,j)* will be introduced as an additive noise as follows,

$$((1+K(i,j)\ Y(i,j))+D(i,j)+S(i,j)+R(i,j))g+Q(i,j). \quad (2.9)$$

Equation (2.9) models the output signal from the camera pipeline. We can see that the PRNU is a multiplicative noise while other noises are additive in nature.

2.4.1.2 Simplified camera output model

Since signal processing chain in digital cameras is very complex and varies with camera brands and models, only white balance and gamma correction are considered in the camera model output. A gain factor *G* is used to adjust the sensitivity of intensity in different color channels to obtain a correct white balance and γ is the gamma correction factor. Finally $\theta_q(i,j)$ is added to the output model to represent the quantization noise at camera output. Let *I(i,j)* be the signal of a particular color

channel. Then mathematically, with the consideration of gamma correction and white balance to the model in equation (2.9), the camera output becomes,

$$I(i,j) = [(((1 + K(i,j))Y(i,j) + D(i,j) + S(i,j) + R(i,j))g + Q(i,j))G]^{\gamma} + \Theta_q(i,j) \quad (2.10)$$

Grouping all the additive noises (dark current noise, shot noise and read out noise) together, the camera output model in equation (2.10) can be written as,

$$I(i,j) = G^{\gamma}[\big(1 + K(i,j)\big)Y(i,j) + \Lambda(i,j)]^{\gamma} + \Theta_q(i,j). \quad (2.11)$$

Dropping the pixel indices for better readability, the camera output model in equation (2.11) is given as follows:

$$I = G^{\gamma}[(1 + K)Y + \Lambda]^{\gamma} + \Theta_q. \quad (2.12)$$

To further simplify equation (2.12), the authors in [1] factorize Y out. As $(K + \frac{\Lambda}{Y})$ is a very small positive number and $\gamma \cong 0.45$, Taylor expansion approximation can be used to keep the first two terms only. Hence equation (2.12) can be simplified as follows:

$$\begin{aligned} I &= (GY)^{\gamma}[(1 + K) + \tfrac{\Lambda}{Y}]^{\gamma} + \Theta_q \\ &= (GY)^{\gamma}\left[(1 + \gamma K) + \frac{\gamma\Lambda}{Y}\right] + \Theta_q \\ &= I^{(0)} + I^{(0)}K' + \Theta, \end{aligned} \quad (2.13)$$

where $I^{(0)} = (GY)^{\gamma},\ K' = \gamma K$ *and* $\Theta = (GY)^{\gamma}\left(\frac{\gamma\Lambda}{Y}\right) + \Theta_q$.

$\boldsymbol{K'}$ is the PRNU feature that is often used as a fingerprint for camera identification. Given the model in equation (2.13), one wants to extract $\boldsymbol{K'}$ from camera output I without knowing the actual input $I^{(0)}$. Different methods have been proposed to extract $\boldsymbol{K'}$, for example the Basic Approach and the MLE Approach which will be described in the next section.

2.4.2 Basic Approach

The Basic Approach is a simple but effective way to make use of the pattern noise for individual source camera identification. It was the first reliable approach proposed for extracting the sensor pattern noise. The general framework developed for camera identification is still used by subsequent camera identification methods. In this section, the general flow of the identification method will first be given. After that, each step in the identification process will be discussed in details.

2.4.2.1 General flow of the identification procedure

Figure 2.5 shows the general flow of the identification method of the Basic Approach. To decide whether a photo is taken by a particular camera C, the method requires a set of photos taken from C and another set of photos taken by some other cameras. The set of photos from C is divided into two subsets, i.e., subset A and subset B. Images first undergo image denoising so that the PRNU feature can be

extracted by calculating the difference between the original image and the denoised image. The PRNU feature from images in subset A is used to estimate the reference PRNU for C and that from subset B is used to obtain the probability density function (PDF) of the correlation between the PRNU feature extracted from images in subset B and the reference PRNU. Similarity, the PRNU extracted from other cameras would be used to obtain another probability density function. Using these two PDFs and Neyman Pearson theorem, a threshold is obtained for a pre-defined false rejection rate (FRR) which minimizes the false acceptance rate (FAR). Hence this threshold can be used to judge whether the photo is taken by a particular camera C or not. Besides, the FRR estimated can be used for performance evaluation.

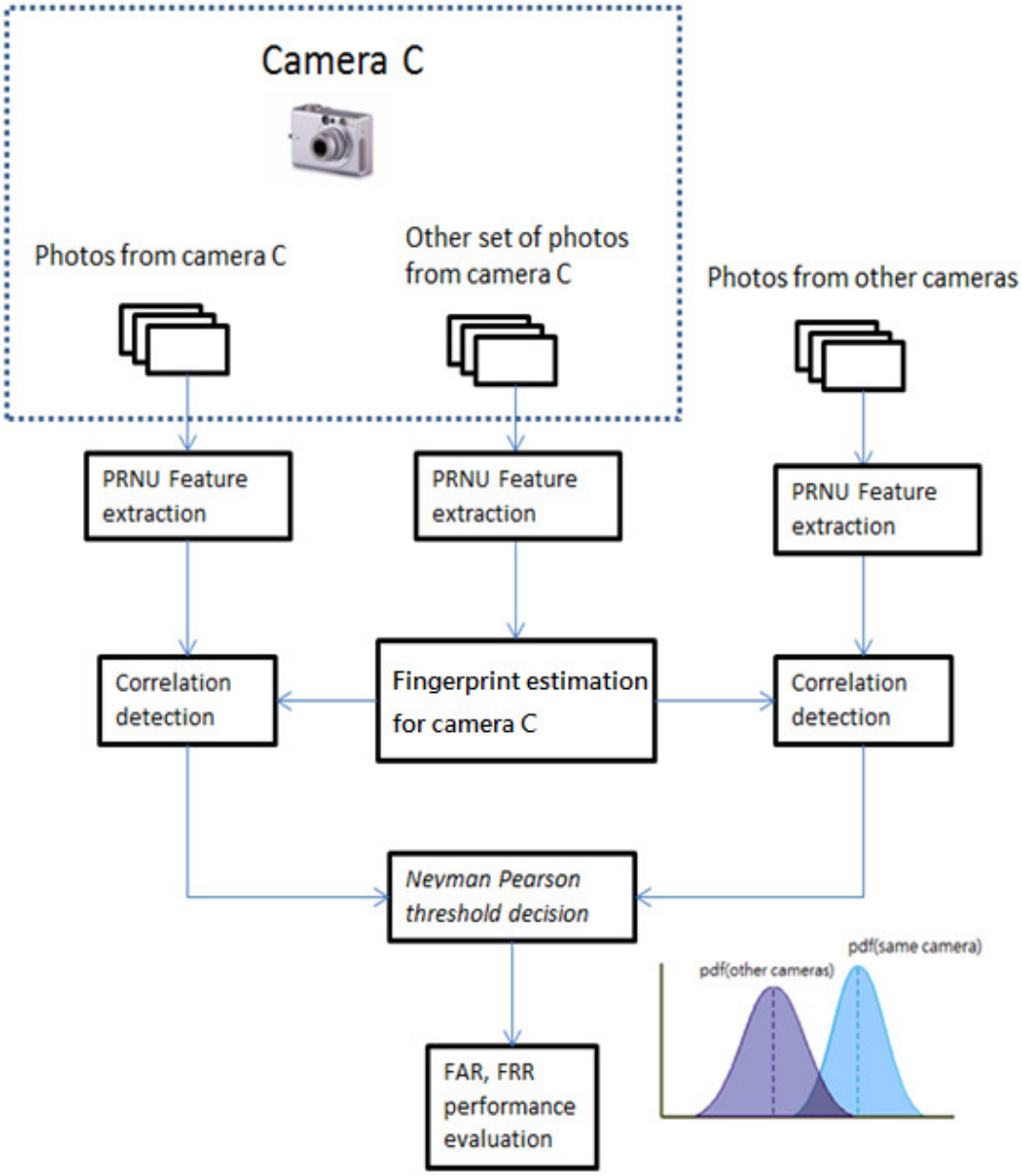


Figure 2. 5 Flow of the identification method of the Basic Approach.

2.4.2.2 Feature extraction by denoising

To extract the PRNU feature of the photos, a denoising filter [34] is used. The noise residue W is obtained by subtracting the original image I from the denoised image $\hat{\boldsymbol{I}}^{(\mathbf{0})}$, i.e.,

$$\begin{aligned} \boldsymbol{W} &= \boldsymbol{I} - \hat{\boldsymbol{I}}^{(\mathbf{0})} \\ &= \boldsymbol{I}\boldsymbol{K}' + \boldsymbol{I}^{(\mathbf{0})} - \hat{\boldsymbol{I}}^{(\mathbf{0})} + (\boldsymbol{I}^{(\mathbf{0})} - \boldsymbol{I})\boldsymbol{K}' + \boldsymbol{\theta} \\ &= \boldsymbol{I}\boldsymbol{K}' + \varepsilon, \end{aligned} \tag{2.14}$$

where ε is the sum of θ and two additional terms introduced by the denoising filter. Assume that ε can be modeled by white Gaussian noise (WGN) with variance σ^2 and is independent of $\boldsymbol{IK'}$. This noise residue can be used to estimate the reference PRNU of a particular camera and for feature comparison.

2.4.2.3 Camera Fingerprint and Correlation detection

To generate a reference fingerprint for a particular camera, simple averaging technique can be used. Let $\widehat{\boldsymbol{W}}_{\boldsymbol{c}}$ be the reference fingerprint for a particular camera C, it can be obtained as the average of the noise residue $\boldsymbol{W}_{\boldsymbol{k}}$ from N images taken by the camera C, i.e.,

$$\widehat{\boldsymbol{W}}_{\boldsymbol{c}} = \frac{\sum_{k=1}^{N} \boldsymbol{W}_{\boldsymbol{k}}}{N}, \tag{2.15}$$

The larger number of N in generating the reference fingerprint, the better the estimation is. Despite that, the estimation of reference fingerprint converges with increasing number of N, the fingerprint estimation with 100 images and 200 images are in fact very similar. Based on the study by Lukas[1], it would be good enough for $N \geq 50$.

After generating the reference fingerprint for the camera C, it is necessary to have a way to measure the similarity of PRNU feature between the reference camera and the testing images. Lukas [1] proposed to use Pearson's correlation to decide whether a photo is taken by a particular camera C, and this technique is still a common way for PRNU comparison.

To decide whether a particular photo P is taken by a specific camera C, the reference fingerprint $\widehat{\boldsymbol{W}}_{\boldsymbol{c}}$ of camera C should first be calculated using equation (2.15). Then the same denoising filter [34] should be used to extract the noise residual $\boldsymbol{W}_{\boldsymbol{p}}$ from a testing photo as equation (2.14): $\boldsymbol{W}_{\boldsymbol{p}} = \boldsymbol{I}_{\boldsymbol{p}}\boldsymbol{K'}_{\boldsymbol{p}} + \varepsilon_p$. The correlation detector calculates the correlation between $\boldsymbol{I}_{\boldsymbol{p}}\widehat{\boldsymbol{K}}_{\boldsymbol{c}}$ and $\boldsymbol{W}_{\boldsymbol{p}}$ as follows:

$$\rho_C(p) = corr\left(\boldsymbol{W}_{\boldsymbol{p}}, \boldsymbol{W}_{\boldsymbol{c}}\right) = \frac{\left(\boldsymbol{W}_{\boldsymbol{p}} - \overline{\boldsymbol{W}_{\boldsymbol{p}}}\right) \odot \left(\boldsymbol{W}_{\boldsymbol{c}} - \overline{\boldsymbol{W}_{\boldsymbol{c}}}\right)}{\left\|\boldsymbol{W}_{\boldsymbol{p}} - \overline{\boldsymbol{W}_{\boldsymbol{p}}}\right\| \left\|\boldsymbol{W}_{\boldsymbol{c}} - \overline{\boldsymbol{W}_{\boldsymbol{c}}}\right\|}. \tag{2.16}$$

where $\overline{\boldsymbol{W}_{\boldsymbol{p}}}$ denotes mean of $\boldsymbol{W}_{\boldsymbol{p}}$, $\odot$ denotes dot product, and $\|.\|$ denotes the L_2 norm. The correlation measures the similarity between the reference PRNU and the PRNU extracted from a particular photo P. If the photo is taken by that camera,

the correlation value calculated should be large. If the photo is not taken by that camera, the value should be small. A threshold is thus required to decide whether the photo is taken by the camera C. The setting of this threshold and system performance evaluation method will be introduced in next section.

2.4.2.4 Neyman Pearson theorem for threshold decision

Assume that there are some photos taken by camera C, and some other photos not taken by camera C. For a particular photo P, one wants to know whether it comes from C or not. The problem is then formulated as a binary hypothesis test as,

$$\begin{cases} H_0\colon\ image\ not\ from\ C \\ H_1\colon\ image\ \ from\ C \end{cases} \tag{2.17}$$

By using the correlation detector in equation (2.16), a set of correlation values can be obtained which is used to obtain the probability density function (PDF) of the correlation between the camera fingerprint ($\widehat{W_c}$) and the noise residual of the testing photo (W_p). In other words, two PDFs are generated: PDF of p($\rho|H_0$) and that of p($\rho|H_1$). The PDF of p($\rho|H_0$) is the distribution of the correlation values for testing photos not coming from camera C while the PDF of p($\rho|H_1$) is that coming from camera C. Figure 2.15 shows a sample plot of the two PDFs. We can see that the correlation values p($\rho|H_0$) are generally smaller than p($\rho|H_0$). By setting the FAR tolerance at a particular small value, e.g., 10^{-3}, a threshold t_o can be set to decide

whether a photo is taken by the camera C. Besides, with the pre-set FAR value, FRR can be used to evaluate the system performance.

The digital camera identification method using PRNU feature has an edge on identifying individual cameras even though they might be of the same brand and model. The accuracy of identification greatly depends on the number of pixels used for PRNU feature extraction. For typical digital cameras, the resolution of digital images is usually very high with more than millions of pixels. This makes the identification accuracy often larger than 90 percent as reported by many researches [1-14].

2.4.3 MLE (Maximum likelihood estimation) Approach

The MLE Approach makes certain modification on the Basic Approach to try to enhance the identification result. For example, a PRNU processing was adopted to suppress the similar artifacts appeared in the same model camera. Besides, the feature used for camera identification was changed to be the multiplicative PRNU factor within the noise residue. Furthermore, a maximum likelihood estimation framework was used for the fingerprint estimation. Finally, a modified statistical detection method which involves a weighting mechanism was also introduced for camera identification.

2.4.3.1 Maximum Likelihood (ML) estimation

Assuming a set of digital images from a particular camera is available. The PRNU $\boldsymbol{K'}$ can be determined from N given images, $I_1, \ldots, I_N$ under a maximum likelihood framework. From equation (2.14), with the assumption that the noise term ε_k can be modeled by white Gaussian noise (WGN) with variance σ^2 and is independent of $\boldsymbol{IK'}$. Then, for each image k=1,…, N, we have:

$$\frac{\boldsymbol{W_k}}{\boldsymbol{I_k}} = \boldsymbol{K'} + \frac{\boldsymbol{\varepsilon_k}}{\boldsymbol{I_k}}. \tag{2.18}$$

The log-likelihood of observing $\frac{W_k}{I_k}$ given $\boldsymbol{K'}$ is

$$L(\boldsymbol{K'}) = -\frac{N}{2}\sum_{k=1}^{N} log\left(\frac{2\pi\boldsymbol{\sigma^2}}{\boldsymbol{I_k}^2}\right) - \sum_{k=1}^{N} \frac{\left(\frac{\boldsymbol{W_k}}{\boldsymbol{I_k}} - \boldsymbol{K'}\right)^2}{\frac{2\boldsymbol{\sigma^2}}{\boldsymbol{I_k}^2}}. \tag{2.19}$$

By taking partial derivatives of equation (2.19), the PRNU can be estimated as follows:

$$\frac{\partial \mathrm{L}(\mathbf{K\prime})}{\partial \mathbf{K\prime}} = \frac{1}{\sigma^2}\sum_{k=1}^{N}(\boldsymbol{W_k} - \boldsymbol{I_k K'})\boldsymbol{I_k} = 0$$

$$=> \widehat{\boldsymbol{K}} = \frac{\sum_{k=1}^{N} \boldsymbol{W_k I_k}}{\sum_{k=1}^{N} (\boldsymbol{I_k})^2}. \tag{2.20}$$

Images used for estimating the PRNU should be those with high luminance and small variance. For example, out-of-focus images of bright blue sky would be the best. The number of photos used, i.e., N, should be large enough to provide a reliable fingerprint estimation. It was suggested that N should be larger than 50 for

those blue sky photos. If natural images are used, N should be larger in order to have a reliable estimation.

2.4.3.2 PRNU Processing

The noise residue W in equation (2.14) is not only containing the artifact unique for each sensor, but also other artifacts shared among cameras with same brand/model. Such artifacts can make the PRNU extracted from different cameras of the same model slightly correlated with each other and hence reduce the identification accuracy. These common artifacts include those from 1) Color interpolation, 2) Row and column noise of the sensor and 3) Blocking artifact of JPEG image.

Color interpolation: Most cameras nowadays are equipped with Color Filter Array (CFA). Each pixel of the image sensor captures only a single color component. The other missing color components are generated by color interpolation from neighboring pixels. The CFA and its corresponding interpolation methodology generate a periodic pattern in the extracted PRNU. Hence, cameras equipped with the same interpolation methodology would have the same periodic pattern.

Row and column noise: Due to the row-wise and column-wise operation of the image sensor and the processing circuit, the output images contain bias for each row

and column. Hence cameras of the same brand/model carry the same bias.

Blocking artifact of JPEG images: JPEG is one of the common formats in storing digital images. Due to the block-based processing and quantization, blocking artifact is often left in the JPEG-compressed images.

To solve the problem, some processing operations can be used to suppress the unwanted artifacts. Zero mean operation for every row and column can be done to remove the row and column artifact. Each pixel in the column is subtracted by its column average. Similarly, each pixel in the row is subtracted by its row average. To remove the structural artifact from color interpolation and blocking artifact, the zero-meaned PRNU is transformed into Fourier domain and then filtered with the Wiener Filter. Let the zero mean operation of the PRNU be $ZM(\boldsymbol{K}')$ where K' is the PRNU multiplicative factor, then the Wiener Filtering in Fourier domain can be expressed as,

$$WF(ZM(\boldsymbol{K}'))=F^{-1}\{F(ZM(K')-W(F(ZM(\boldsymbol{K}'))\}, \quad (2.21)$$

where W is the 3x3 Wiener filter, F and F^{-1} are the Fourier transform and inverse Fourier transform respectively.

2.4.3.3 A Weighting mechanism

Equation (2.14) is valid for images with smooth content only. For natural images, the noise term ε should be modeled as a colored Gaussian noise ε_c with unequal variance σ_c^2 as the noise is strongly influenced by image texture. Besides, there should have an attenuation factor T to the PRNU term. Therefore, the noise residue model in equation (2.14) should be modified as,

$$W = TIK + \varepsilon_c. \tag{2.22}$$

Note that T is a pixel-wise multiplicative attenuation factor. As it is difficult to estimate the two non-stationary quantities T and ε_c at every pixel due to insufficient data, a simplified approach is to estimate these quantities in a block based manner. The testing images are divided into M disjoint blocks and T and ε are assumed to be constant within each block. The binary hypothesis test problem becomes,

$$\begin{cases} H_0: W = \varepsilon \\ H_1: W = TX + \varepsilon \end{cases} \tag{2.23}$$

where $\varepsilon[i]$, $i \in B_b$, is white Gaussian noise (WGN) with zero mean and known variance $\hat{\sigma}_b^{\,2}$ and X

The optimal detector for equation (2.23) is the generalized matched filter, i.e.,

$$\rho = \sum_{b=1}^{M} \beta_b \rho_b \tag{2.24}$$

where

$$\beta_b = \frac{\frac{\hat{\boldsymbol{T}}_b}{\hat{\sigma}_b^2}||\boldsymbol{X}_b||\,||\boldsymbol{W}_b||}{\sqrt{\sum_{b=1}^{M}\frac{\hat{\boldsymbol{T}}_b}{\hat{\sigma}_b^2}||\boldsymbol{X}_b||^2}\sqrt{\sum_{b=1}^{M}\frac{1}{\hat{\sigma}_b^2}||\boldsymbol{W}_b||^2}} \tag{2.25}$$

and $\rho_b = corr(X_b, W_b)$. Shaping factor T_b and $\hat{\sigma}_b$ can be estimated under H_1 as,

$$\rho_b = \frac{\boldsymbol{T}_b\|\boldsymbol{X}_b\|^2+\boldsymbol{X}_b\odot\varepsilon_b}{\|\boldsymbol{X}_b\|\sqrt{\boldsymbol{T}_b{}^2\|\boldsymbol{X}_b\|^2+2\boldsymbol{T}_b\boldsymbol{X}_b\odot\varepsilon_b+\|\varepsilon_b\|^2}} \tag{2.26}$$

As ε_b is zero mean and independent of X_b, the mixed term is small as compared to other terms in equation (2.26). This implies,

$$\rho_b \approx \frac{\boldsymbol{T}_b\|\boldsymbol{X}_b\|}{\|\boldsymbol{X}_b\|^2\sqrt{\boldsymbol{T}_b{}^2\|\boldsymbol{X}_b\|^2+\|\varepsilon_b\|^2}} = \frac{1}{\sqrt{1+\frac{\|\varepsilon_b\|^2}{\boldsymbol{T}_b{}^2\|\boldsymbol{X}_b\|^2}}} \tag{2.27}$$

$$\frac{\boldsymbol{T}_b{}^2\|\boldsymbol{X}_b\|^2}{\|\varepsilon_b\|^2} = \frac{\rho_b}{1-\rho_b{}^2} \tag{2.28}$$

From equations (2.22) and (2.28), $\hat{\boldsymbol{T}}_b$ and $\hat{\sigma}_b$ can be expressed as

$$\hat{\sigma}_b{}^2 = \frac{1-\rho_b{}^2}{|B_b|}\|\boldsymbol{W}_b\|^2,\ \hat{\boldsymbol{T}}_b = \frac{\rho_b}{\|\boldsymbol{X}_b\|}\|\boldsymbol{W}_b\|. \tag{2.29}$$

The estimates, $\boldsymbol{T}$ and ε, depend on ρ_b, which is only known under hypothesis H_1 instead of H_0. This problem can be solved by constructing a correlation predictor to predict the normalized correlation based on the image features.

2.4.3.4 Correlation Predictor

The correlation predictor is constructed on small blocks for images from the same cameras as the reference camera. It is constructed as a mapping from the image feature vector to the predicted correlation value. In other words, one tries to find the predicted correlation value by using a set of image features. The authors from [3] found out three factors which affect the correlation the most. They are image intensity, texture, and signal flatting.

As shown in equation (2.13), the PRNU is a multiplicative term to the image intensity as $I^{(0)}K'$. The correlation is directly proportional to the image intensity. However, due to the finite numbers of bits used in digital image representation, the PRNU is absent in the saturated area, i.e. when the intensity is 255 for an eight bit intensity representation. The correlation is also attenuated at pixels near saturated area, as the PRNU near the saturated area affects the PRNU extraction from de-noising. The intensity feature is defined as:

$$f_I = \frac{1}{|B_b|}\sum_{i\in B_b} att(I[i]) \tag{2.30}$$

where $att(I[i])$ is the attenuation function given by,

$$att(I[i]) = \begin{cases} e^{-(I[i]-I_{crit})^2/\tau}, & I[i] > I_{crit} \\ \frac{I[i]}{I_{crit}}, & I[i] < I_{crit} \end{cases} \tag{2.31}$$

where I_{crit} and τ are chosen experimentally.

Due to the use of wavelet-based de-noising filter, the texture feature is extracted in the wavelet domain. The three wavelet high frequency bands at the first scale, i.e., LH HH and HL are combined to generate a high pass image *F*. The texture is then calculated as,

$$f_T = \frac{1}{|B_b|}\sum_{i\in B_b}\frac{1}{1+var_5(F[i])}, \tag{2.32}$$

where $var_5(F[i])$ is the variance of F in the 5x5 neighborhood of pixel i.

For flattened area in an image, the predictor will always overestimate the correlation. Therefore, the third feature f_S is used to model this phenomenon. This feature is defined as the ratio of the number of pixels in a block with an average local variance below certain threshold, i.e.,

$$f_S = \frac{1}{|B_b|}|\{i \in B_b | \sigma_I[i] < cI[i]\}| \tag{2.33}$$

where *c* is the threshold which is chosen experimentally, $\sigma_I[i]$ is the local variance of image intensity $I[i]$ at pixel i estimated from a local 5x5 neighborhood.

Sometimes, the intensity and the texture cannot be treated separately. For example, a highly textured region can also be a high-intensity region. Therefore, a combined texture-intensity feature is used which is defined as,

$$f_{TI} = \frac{1}{|B_b|}\sum_{i\in B_b}\frac{att(I[i])}{1+var_5(F[i])} \tag{2.34}$$

The four features are then modeled in second order to estimate the correlation $\rho(k)$

as:

$$\rho(k) = \alpha_0 + \alpha_1 f_I + \alpha_2 f_T + \alpha_3 f_S + \alpha_4 f_{TI} + \alpha_5 f_I f_I + \alpha_6 f_I f_T + \cdots + \Psi[k] \quad (2.35)$$

where $\Psi[k]$ is the noise term and α is the feature coefficient. In matrix form, equation (2.35) can be written as,

$$\boldsymbol{\rho} = \boldsymbol{F}\,\boldsymbol{\alpha} + \boldsymbol{\Psi} \quad (2.36)$$

where **F** is the matrix of features, $\boldsymbol{\alpha}$ is the vector of unknown feature coefficient and $\boldsymbol{\Psi}$ is the noise term. By using least square estimator (LSE), the feature coefficients can be solved, i.e.,

$$\boldsymbol{\alpha} = \left(\boldsymbol{F}^T \boldsymbol{F}\right)^{-1} \boldsymbol{F}^T \boldsymbol{\rho} \quad (2.37)$$

The image feature can then be used to predict the correlation and thus achieve the weighting purpose as stated in Section 2.4.3.3.

2.4.4 Phase Approach

Phase pattern noise can be used to suppress the image content in camera fingerprint. Studies have found that the large magnitude values in the extracted pattern noise often come from the image content. Removing the magnitude of the pattern noise in frequency domain can help reducing the influence of the image content effect in camera identification.

2.4.4.1 Phase pattern noise

For phase PRNU approach [4], the noise residue W_k first undergoes Discrete Fourier Transform (DFT), i.e.,

$$W_{fk} = DFT(W_k), \tag{2.38}$$

where DFT() denotes the discrete Fourier transform operation. Then, the phase information is kept while the magnitude information is changed to be a constant as follows,

$$W_{\emptyset k} = \frac{W_{fk}}{|W_{fk}|}, \tag{2.39}$$

where $|W_{fk}|$ is the Fourier magnitude of W_{fk}. Afterwards, it is transformed back into spatial domain using Inverse DFT, i.e.,

$$W_{pk} = real(IDFT\Big((W_{\emptyset k})\Big)) \tag{2.40}$$

where *IDFT()* denotes the inverse discrete Fourier transform operation and real() means only the real part is kept because the imaginary part is very small and can be ignored. This frequency whitened noise residue is called the phase pattern noise. The whitening process can remove the contaminations in the frequency domain from the image details as well as periodic noise component and linear pattern. The reference phase pattern noise can be obtained by the following equation,

$$W_{pc} = real(IDFT\left(\frac{\sum_{k=1}^{N} W_{\emptyset k}}{N}\right)), \tag{2.41}$$

where N is the total number of images used in the camera fingerprint extraction. Similar to the Basic Approach, the signal detection can be done by using correlation but with the phase pattern noised instead of the PRNU feature in equation (2.20).

2.5 Chapter Summary

This chapter gives a general review of different methods for camera identification. Digital camera identification can be solved using characteristics from different components in the camera. Lens distortion, CFA interpolation and image sensor noise for source camera identification have been discussed. Among these methods, only image sensor noise can identify individual camera with the same brand and model. For most digital forensic applications, identifying exactly the source camera is much more important than knowing only the camera brand/model. Due to this reason, the sensor pattern noise is often used in camera identification nowadays.

Details of using PRNU feature for source camera identification has been discussed. In particular, three common approaches, namely the Baisc Approach, the MLE Approach and the Phase Approach have been discussed. Table 2.1 shows the similarities and differences for these three approaches in camera identification. The Basic Approach gives the overall framework for the identification process. Later,

some common artifacts which share among for cameras of the same model were found affecting the accuracy of camera identification. PRNU processing was then used to suppress these artifacts. Besides, the use of maximum likelihood estimation (MLE) for the reference camera fingerprint was proposed. The MLE Approach also provides a weighting framework to deal with the scene content effect in the pattern noise, i.e. a weighting is applied in a block-based manner. This block-based processing is, however, not flexible enough. Until now, the scene content effect on the pattern noise is still a big problem. Recently, the Phase Approach was proposed. It examines the phase of the pattern noise in frequency domain. Studies [4] found that the phase only pattern noise can suppress the image content and remove some structural pattern such as JPEG artifact and the demosaicing artifact.

These three approaches are state of the art techniques for camera identification using PRNU. This thesis is a further study on the individual source camera identification. The scene content is the problem this thesis targeted to solve. In the next chapter, a way that formulates the image content effect on the correlation will be introduced and combined into the identification procedure.

Table 2.1 Similarities and differences among three different approaches for camera identification.

	Basic Approach	MLE Approach	Phase Approach
Feature extraction	Wavelets denoising		
Feature used	Pattern noise	Multiplicative factor K from the pattern noise	Phase only pattern noise
PRNU processing	No	Yes	Yes
Fingerprint estimation method	Averaging	Maximum likelihood estimation	Averaging
Signal Detection method	Normalized correlation	Normalized correlation /Generalized matched filter	Normalized correlation
Threshold Decision	Neyman Pearson theorem under binary hypothesis test		

Chapter 3 Image content effect on the camera identification performance

3.1 Introduction

Traditional camera identification using PRNU mainly consists of three components: 1) PRNU extraction, 2) PRNU comparison and 3) probability density function estimation and threshold setting. The reference PRNU for a particular camera is extracted from a number of images taken by that camera using approaches such as the Basic or the MLE approaches as discussed in Section 2.4.2 and Section 2.4.3. Two probability density functions (PDFs) are generated, one for hypothesis H_0 ($p(x|H_0)$) and the other for H_1 ($p(x|H_1)$) as described in Section 2.4.2.4. Using these two PDFs, the false acceptance rate (FAR) is set for the false tolerance of the system. At a given FAR, a threshold can then be set to minimize the false rejection rate (FRR). In fact, FRR can be used for performance evaluation. With this threshold, source camera identification can be achieved by calculating the correlation between the reference PRNU and the pattern noise of a testing photo.

The estimation of the two PDFs is very important. If the two PDFs are not estimated properly, the accuracy of camera identification will be seriously affected. It is found that $p(x|H_1)$ is highly dependent on the image content of the testing images. It is because the image content can be left in the extracted PRNU through denoising.

If the set of images used to generate the $p(x|H_1)$ is "good", i.e., plain images with smooth content, the separation between $p(x|H_1)$ and $p(x|H_0)$ would be large. If the images used to generate $p(x|H_1)$ are highly textured images, then separation between $p(x|H_1)$ and $p(x|H_0)$ would be small. As a result, if the image content effect is not properly accounted for, the decision threshold would be over conservative for the "good" images but too optimistic for the "bad" images.

To address this problem, correlation predictor can be used. The idea is to extract features such as image intensity and texture information from an image. These features are then used to predict the correlation between the PRNU of the testing image and the reference PRNU from a particular camera. A detailed study on the relation between the predicted correlation and the identification accuracy is carried out. With the predicted correlation, the testing images can be classified according to their contents. Using this result, a 2D classifier is proposed so that the predicted correlation can be used to set the decision threshold according to the image content.

This chapter is organized as follows. Section 3.2 studies the image content effect on the two PDFs. Section 3.3 introduces the image feature formulation by the correlation predictor. Section 3.4 discusses the way of evaluating the identification performance according to the image feature. The proposed two-dimensional

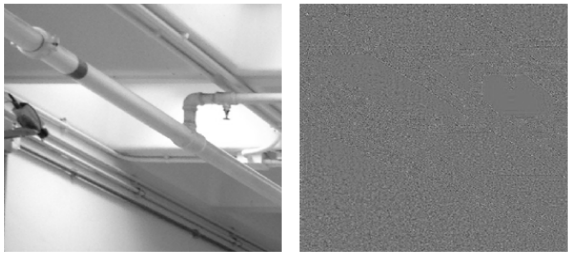
classification method will be discussed in Section 3.5. Finally, Section 3.6 concludes the chapter.

3.2 Effect of Image content on Camera Identification

The PRNU feature is a key to the camera identification. However, the estimated PRNU through denoising can easily be contaminated by the image content. Figure 3.1(a) shows an example of an indoor image and Figure 3.1(b) shows its corresponding noise residue. Scene structure such as the edge of the water pile and ceiling can be seen clearly in Figure 3.1(b). In fact, the noise residue always contains the image content, irrespective of the kind of image used. The question is how serious the image content is left in the extracted PRNU feature.

Traditional camera identification methods [1-14] do not consider the kind of image content used in camera identification. In fact, the content of the images used in the experiment contributes to a major impact in the identification accuracy. As described in Section 2.4.2.4, the threshold decision of the identification system depends on two PDFs: PDF for hypothesis H_0 $(p(x|H_0))$ and that for hypothesis H_1 $(p(x|H_1))$, where H_0 refers to images which are not from the same camera as the reference PRNU and H_1 refers to image from the same camera as the reference PRNU. The PDF $p(x|H_1)$ in fact is dependent of the image content. Figure 3.2 shows an example of the effect of image content on histogram of correlation of H_0 and H_1 for a particular camera. The

histogram of H_1 was generated by two set of images. One set is indoor images with smooth content while the other set is outdoor hill images with textured content. Figure 3.2(e) shows that the distance between histograms of H_0 and H_1 are different for different sets of testing images. Histograms of H_0 and H_1 are more separated when images for H_1 have smooth content. Histograms of H_0 and H_1 are much closer to each other when images for H_1 have texture content. If the system uses H_0 and H_1 with smooth content for testing and sets the threshold at 0.03, the identification accuracy is 100%. However, if one ignores the image content and still uses the same threshold, the result could be bad for other cases. For example, H_1 with textured content is used for testing, the identification accuracy will drop a lot as evident in Figure 3.2 (e) if the threshold is set at 0.03. There will be a lot of wrong identification cases. This threshold is thus too optimistic for the "bad" images. Alternatively, if someone sets the threshold for H_1 with textured images at 0.01, this threshold will be too conservative for those "good" images. Hence, the scene content effect needs to be properly considered.



(a) (b)

Figure 3. 1(a) the ceiling image and (b) its noise residue.

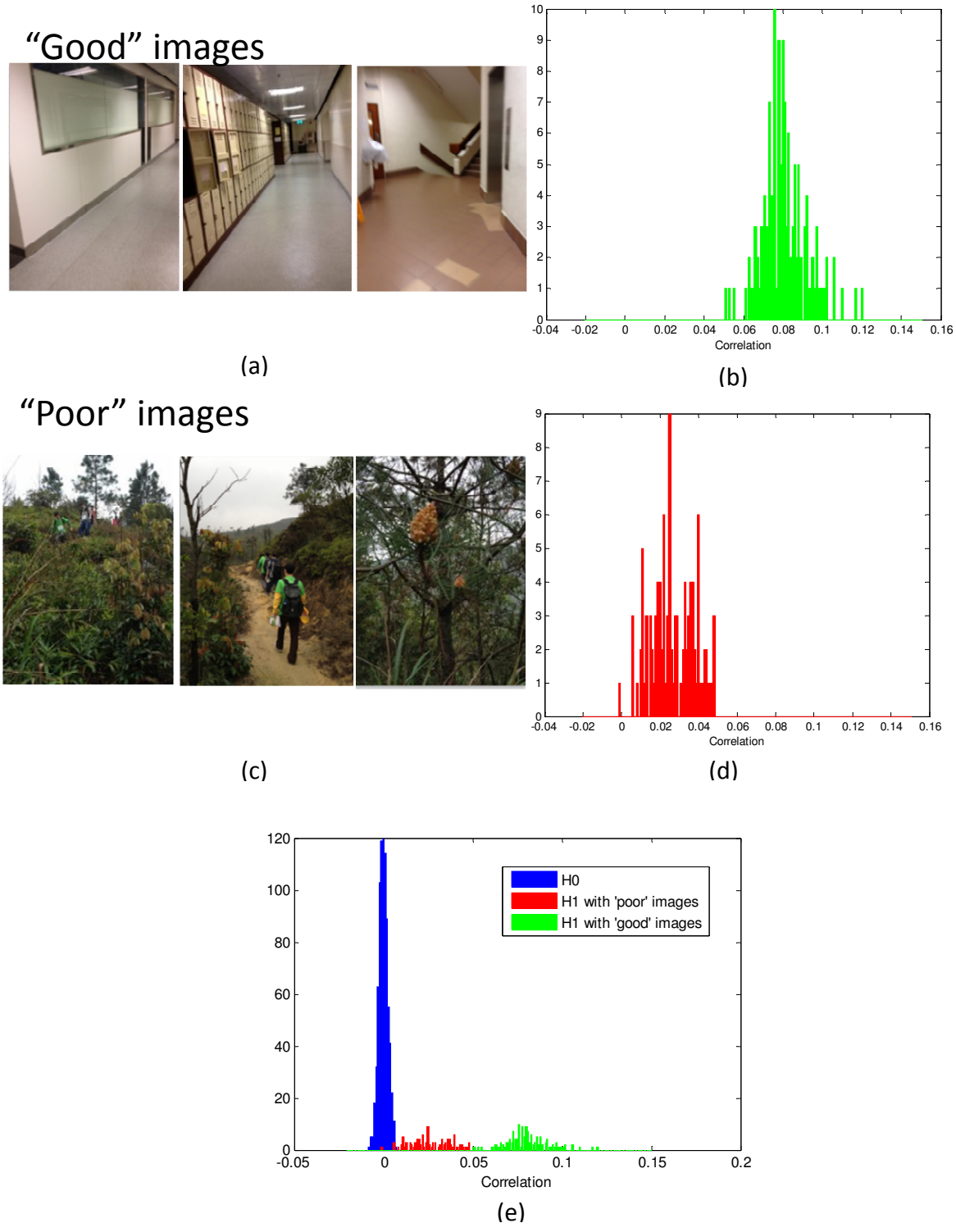


Figure 3. 2 Histogram of H_0, H_1 with 'poor' images and H_1 with 'good' images for camera Minolta DiMAGE X. (a) smooth content images, (b) histogram of H_1 from images in (a), (c) textured images, (d) histogram of H_1 from images in (c), (e) histograms of H_0, H_1 in (b), H_1 in (d) for Minolta DiMAGE X.

3.3 Image Content effect formulated by Correlation predictor

As discussed in Section 3.2, distribution of H_1 can easily be affected by the image content such as image intensity and texture. To take into account the image content effect, researcher [2-3] applied attenuation factors to the PRNU. In particular, a correlation predictor is used to assign different weightings to different image regions as discussed in Section 2.4.3.3. An image is first divided into a number of non-overlapping blocks. For each block, three types of features are extracted. They are intensity in equation (2.30), texture in equation (2.32) and signal flatting in equation (2.33). Besides, the correlation between the PRNU feature in each block and the reference PRNU is calculated. Then, a linear prediction model is built to take into account the effect of the image feature to the correlation, i.e.,

$$\boldsymbol{\rho} = \boldsymbol{H}\boldsymbol{z}, \tag{3.1}$$

where $\boldsymbol{\rho}$ is a column vector consisting of correlation terms from different image data, $\boldsymbol{z}$ is a column vector consisting of coefficients associated with each feature and H is a matrix whose row and column contain respectively the image features and the number of image data. By using least square estimator, the coefficients term $\boldsymbol{z}$ for each feature can be solved. With the trained coefficients, the predicted correlation can be calculated by eq. (3.1) for any testing image by using the same type of image features.

The predicted correlation is a way to quantify the reliability of the correlation obtained. If the predicted correlation is large in magnitude, it implies that the correlation obtained for that block is reliable. Hence, a large weighting will be given to the block. Similarly, a small weighting is assigned to the block with a low predicted value. Therefore, the weighted correlation, which is the sum of the normalized weighted non-overlapping block correlation for the image, can be used to characterize the similarity between two pattern noises.

3.4 Relation Between Image Content and Identification Accuracy

As discussed in Section 3.2, the camera identification accuracy can vary a lot with different image content used in testing. If the identification system is trained with "good" images but tested with "poor" images, the identification result can be very misleading. This section introduces a way to relate the identification accuracy and the image feature. The method will be described in Section 3.4.1 and Section 3.4.2 will show the experimental result.

3.4.1 Methodology

The general framework of the identification process follows the Basic Approach described in Figure 2.5 in Section 2.4.2.1. The reference PRNU for a particular camera is first generated by a set of images from that camera. Afterwards, another set of images from that camera is used to train the correlation predictor as described in Section 3.3. This correlation predictor aims to formulate the seriousness of image content effect on the extracted PRNU.

For each image, the PRNU feature is first extracted by the wavelet based filter [34]. This PRNU is then divided into M non-overlapping blocks. For each block, the correlation is calculated by equation (2.16). Let the correlation for each block be $\rho(b)$ where b =1,2,…M. Then, the actual correlation for an image is obtained as,

$$\rho = \sum_{b=1}^{M} \frac{\rho(b)}{M} \tag{3.2}$$

The actual correlation represents the average of the similarity between the pattern noise of a testing image and the reference PRNU over all the blocks in an image. Using the actual correlation and the image features in each block, θ in eq. (3.1) is trained for a particular camera. Using the trained coefficients in θ, predicted correlation can be obtained for any testing image by analyzing its image features. Hence, the correlation predictor can then be used to estimate the seriousness of image content effect. Let the predicted correlation for each block be $\hat{\rho}(b)$ where b

=1,2,…M. Then, the predicted correlation for each image can be obtained as the average over all the blocks, i.e.,

$$\hat{\rho} = \sum_{b=1}^{M} \frac{\hat{\rho}(b)}{M} \tag{3.3}$$

Figure 3.3 shows a plot of the actual correlation against the predicted correlation for two sets of images. The block size used to train the predictor was 128x128 and the image size used for testing was 512x512. The first set of the images consisted of 400 images (represent by blue dot) taken from the same camera (Minolta DiMAGE X) that generated the reference PRNU. The other set of images (represent by green dot) consisted of 400 photos taken by each of the following cameras: Digital Canon IXUS 65 and Sony DSC T-500. For same camera case, the actual correlation is linearly proportional to the predicted correlation. However, for different camera case, the predicted correlation is not related to the actual correlation. In the Basic Approach, only the actual correlation is used for identification. At low actual correlation values, it is difficult to make a concrete judgment regarding the origin of the testing image as the distance between blue and green dots is very small. In this case, the histogram shown in Figure 3.4, which is generated with actual correlation only, will be used to generate the PDFs of $p(x|H_0)$ and $p(x|H_1)$. We can see that the separation of the correlation values of H_0 and H_1 is small. If the identification system uses a threshold at 0.008, FAR and FRR are equal to 0.01 and 0.05 respectively for data in Figure 3.4.

In fact, the FRR should be varied with the image content as we can clearly see from Figure 3.3 that the overlapping between blue and green dots happens when predicted correlation is smaller than 0.03. Hence, the FRR should be set adaptively depending on the predicted correlation. In other words, FRR should be 0 for predicted correlation larger than 0.03 and FRR should increase when predicted correlation is smaller 0.03. As the performance of the identification depends on the image content, this section studies the relation between the scene content effect and identification accuracy by making use of the correlation predictor.

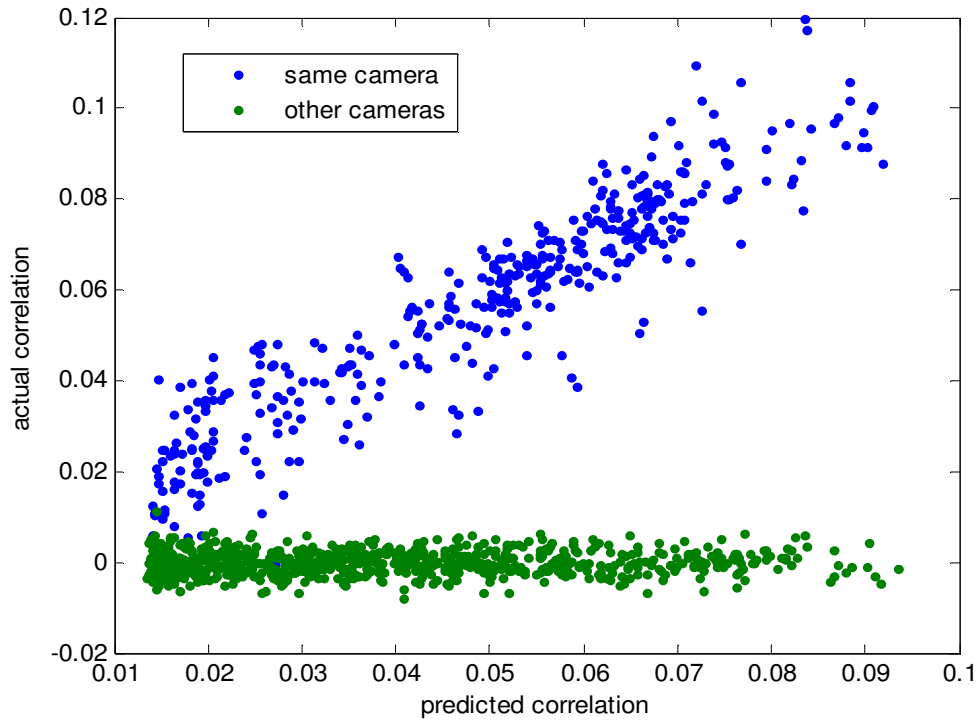


Figure 3. 3 Plot of ρ vs $\hat{\rho}$ for 400 images from Minolta Minolta DiMAGE X (blue dot) and 800 images from other cameras (green dot) using predictor trained from another 50 images from Minolta DiMAGE X.

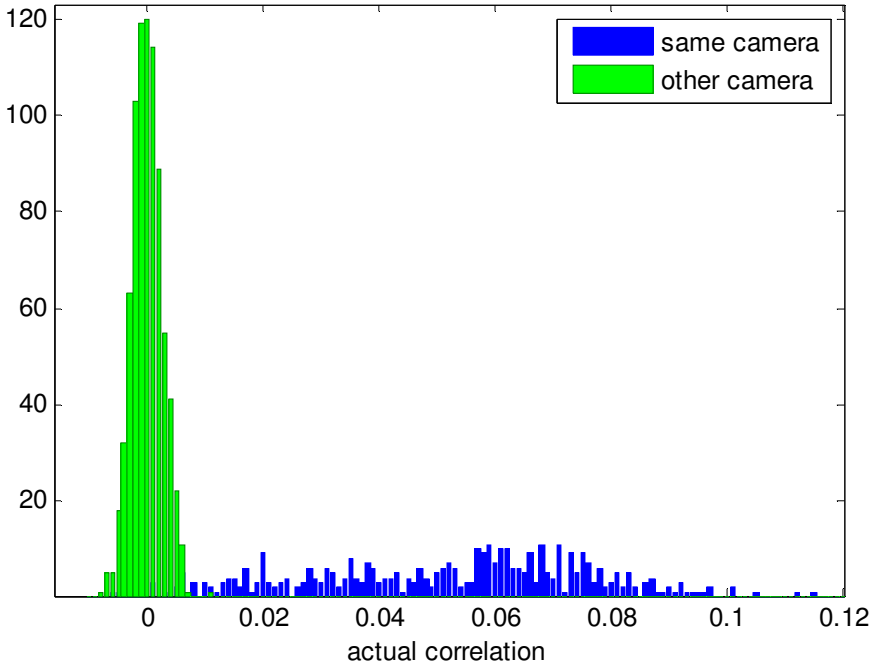


Figure 3. 4 Histogram of actual correlation using data in Figure 3.3

As the actual correlation is independent of predicted correlation for the H_0 hypothesis (images from other camera), the probability density function of H_0 hypothesis can simply be modeled by one dimensional Gaussian distribution. Using the simple one dimensional Gaussian model fitting, the following equation can be obtained:

$$p(x|H0) = \frac{1}{\sqrt{2\pi}s} e^{-\frac{(x-u)^2}{2s^2}}, \tag{3.4}$$

where

$$\boldsymbol{u} = \frac{1}{N}\sum_{x\in Xi} \boldsymbol{x}, \tag{3.5}$$

$$s^2 = \frac{1}{N}\sum_{x\in Xi}(\boldsymbol{x} - \boldsymbol{u})^2. \tag{3.6}$$

x is actual correlation, u and s^2 are respectively the mean and variance of *x*.

For the alternative hypothesis H_1, its probability density function is modeled by a two dimensional Gaussian model. The following equations can be used to solve for the model parameters,

$$p(\boldsymbol{x}|\boldsymbol{w_i}) = \frac{1}{2\pi|\sum|^{1/2}}\, exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{u})^T\sum^{-1}(\boldsymbol{x}-\boldsymbol{u})\}, \tag{3.7}$$

where

$$\boldsymbol{x} = \begin{bmatrix}\boldsymbol{x_1}\\ \boldsymbol{x_2}\end{bmatrix}, \tag{3.8}$$

$$\boldsymbol{u} = \frac{1}{N}\sum_{x\in Xi}\boldsymbol{x}, \tag{3.9}$$

$$\sum = E\{(\boldsymbol{x}-\boldsymbol{u})(\boldsymbol{x}-\boldsymbol{u})^T\}. \tag{3.10}$$

$\boldsymbol{x_1}$is the predicted correlation, $\boldsymbol{x_2}$ is the actual correlation, u is the mean vector and $\sum$ is the covariance matrix. Let

$$\boldsymbol{u} = \begin{bmatrix}\boldsymbol{u_1}\\ \boldsymbol{u_2}\end{bmatrix} \tag{3.11}$$

$$\sum = \begin{bmatrix}\sum_{11} & \sum_{12}\\ \sum_{21} & \sum_{22}\end{bmatrix} \tag{3.12}$$

Then, the conditional parameters for *p(x2|x1)* are given by

$$\boldsymbol{u}_{2|1} = \boldsymbol{u}_2 + \sum_{21}\sum_{11}{}^{-1}(\boldsymbol{x}_1 - \boldsymbol{u}_1) \tag{3.13}$$

$$\sum_{2|1} = \sum_{22} - \sum_{21}\sum_{11}{}^{-1}\sum_{12} \tag{3.14}$$

These two values are the mean and variance of the PDF of H_1at a particular predicted correlation.

Figure 3.5 shows the 2D Gaussian model fitting for the H_1 hypothesis based on the data in Figure 3.3. By using the concept of generating the 1D Gaussian from the 2D Gaussian [55], we can find the relation between scene content and identification accuracy at different predicted correlations. Figure 3.6 shows an example of the conditional PDF of $p(x2|H_1,x1=0.03)$ and the PDF of H_0 $p(x|H_0)$. In such case, the relation can be obtained as $p(x|H_0)$ and $p(x|H_1)$ can be generated for different image content.

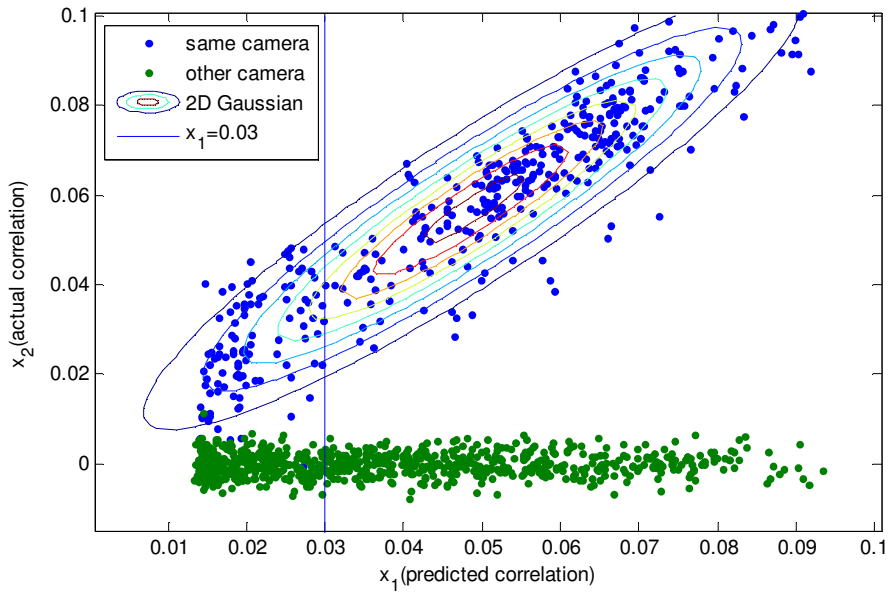


Figure 3. 5 Two dimension Gaussaian modeling on H_1 for data in Figure 3.3

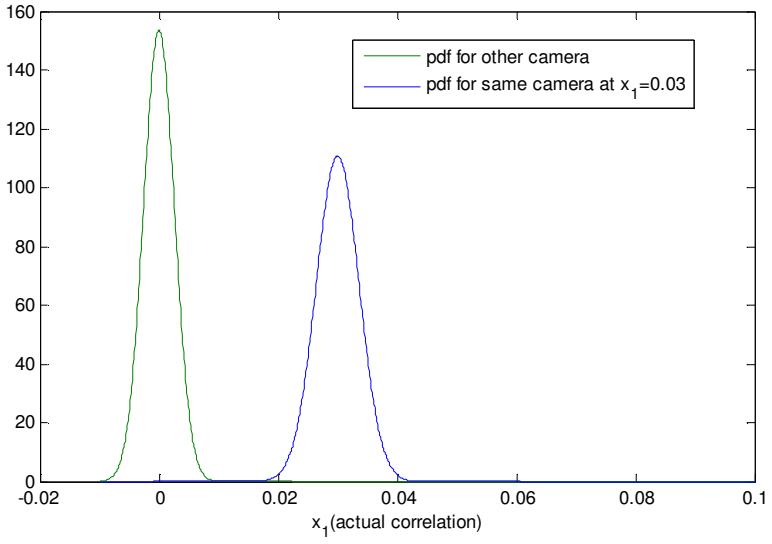


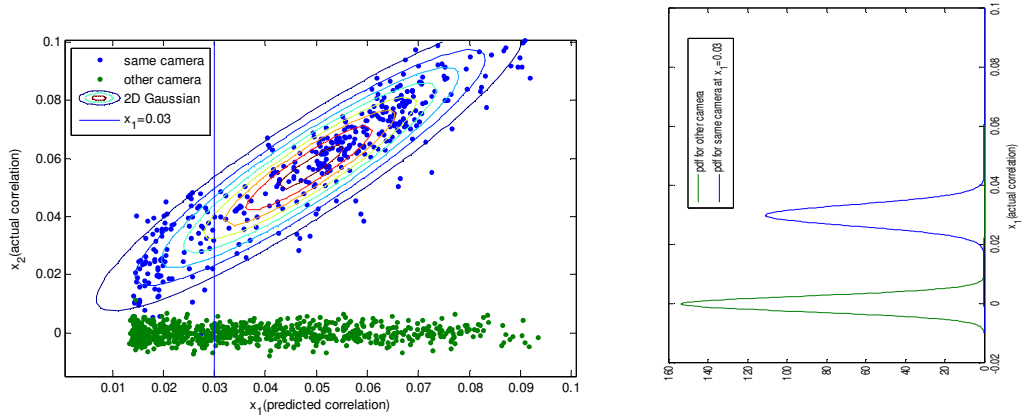
Figure 3. 6 PDF of other camera and PDF of same camera at predicted correlation =0.03.

3.4.2 Experimental Results

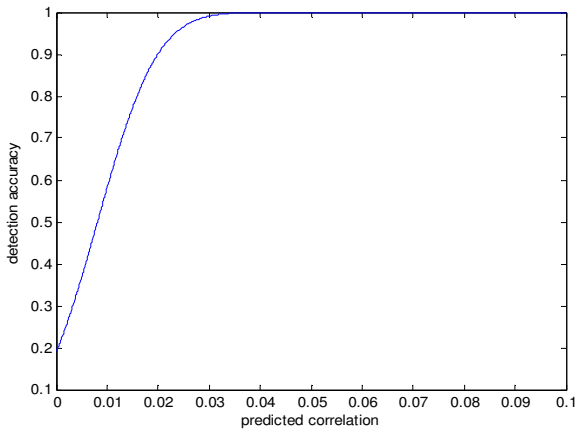
The cameras used in the experiment were Minolta DiMAGE X, Digital Cannon IXUS 65 and Sony DSC T-500. For each camera, 50 images were used to generate the reference PRNU. Afterwards, another 50 images were used to train the correlation predictor by using the method described in Section 3.3. Finally, another 400 images for each camera were used for performance evaluation. Hence, there were 400 images from the same camera and 800 images (400 for each other camera) to generate the PDFs under the binary hypothesis testing framework.

For the H_0 case, the Gaussian model is only based on the actual correlation while for the H_1 case, the Gaussian model is based on both actual correlation and predicted

correlation. Then, for each predicted correlation, the Gaussian PDF for H_0 and H_1 can be obtained as described in Figure 3.5 and Figure 3.6. By using this information, identification accuracy can be obtained. The FAR for H_0 case is set to 0.001. By examining the error rate, the detection accuracy (accuracy in correctly identify images from same camera) can be obtained at each predicted correlation. Figure 3.7 shows the flow of obtaining the detection accuracy at different predicted correlation.



(a) (b)



(c)

Figure 3. 7 The flow to obtain the detection accuracy at different predicted correlation. (a) Step 1: generating the Gaussian model for the data; (b) Step 2: At each predicted correlation, obtain the PDF for both H_0 and H_1 and (c) Step 3: By setting a

pre-defined FAR value, obtain the detection rate for each predicted correlation.

Figure 3.8 to Figure 3.10 show the experimental results for three different cameras. These figures show the changes of performance of the identification system with respect to different types of image content. For example, if the testing image is highly textured which leads to a low predicted correlation, the detection rate will be low. If the image content is smooth, the predicted correlation will be large and hence, the detection rate will be high. The detection rate is in fact related to the image content in which traditional identification method simply ignores. Our study clearly indicates the relation between the scene content (as indicated by the predicted correlation) and the detection accuracy. For example, if the testing image with a predicted correlation larger than 0.02 is found for camera Minolta DiMAGE-Xt, the identification result is reliable as the detection rate is nearly 100%. Otherwise, the identification is not reliable as the detection accuracy drops sharply when predicted correlation is less than 0.02 which indicates the image content contaminates the PRNU feature seriously.

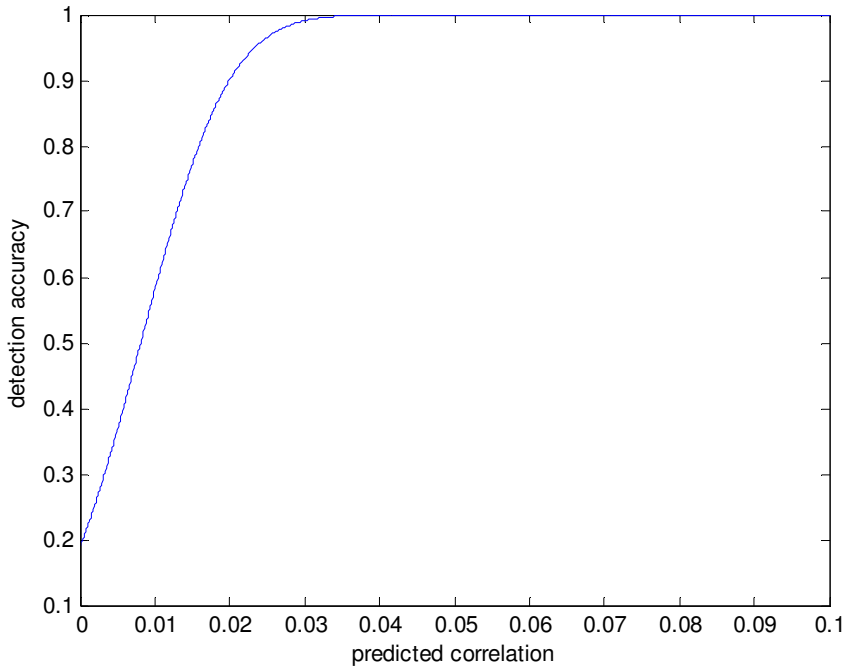


Figure 3. 8 Detection accuracy vs predicted correlation with FAR = 0.001 for camera Minolta DiMAGE X.

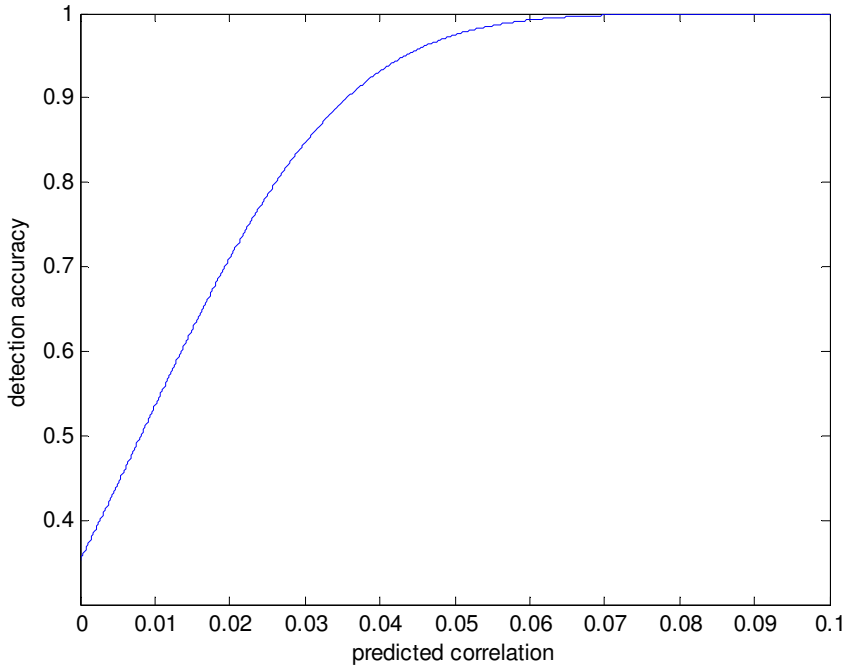


Figure 3. 9 Detection accuracy vs predicted correlation with FAR = 0.001 for camera Digital Cannon IXUS 65.

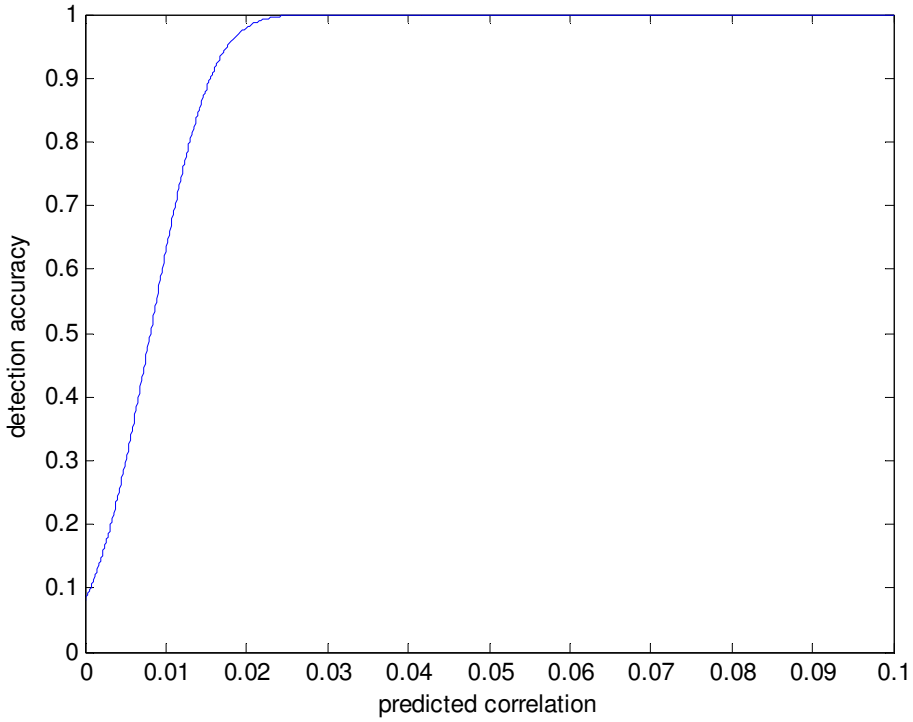


Figure 3. 10 Detection accuracy vs predicted correlation with FAR = 0.001 for camera Sony DSC T-500.

3.5 A two dimensional camera identification method using PRNU feature

In Section 3.4, we can see that the identification performance depends on the image contents. The predicted correlation can be used as an indicator to show the reliability of the correlation and determine the detection accuracy. In this part, the predicted correlation will be used for the source camera identification. In traditional identification method, the threshold is set by the tolerance of false accepting rate (FAR) of the H_1 hypothesis. This makes the same threshold to be used irrespective of the image content. In this part, the predicted correlation will be used to formulate the

image content effect so that a flexible threshold setting scheme can be obtained.

3.5.1 Methodology

As discussed in Section 2.4.3.4, the predicted correlation is used to decide the weights for the correlation in different blocks [2-3]. Indeed, it can be used to classify images according to their features. In this section, the usage of the predicted correlation is extended in the classification procedure. The predicted correlation is considered as another feature used together with the actual correlation to give a two-dimensional classification.

Figure 3.3 shows an important feature of the correlation predictor, i.e., the predicted value is close to the actual correlation if that image comes from the same camera as the reference PRNU. Otherwise, the predicted value is independent of the actual correlation. This suggests that the strategy used to classify images from the same camera and different cameras should vary with the predicted correlation value.

If using solely the actual correlation, the identification process can only have a single value for threshold decision. In other words, if the actual correlation is greater than this threshold, we consider that the testing image comes from the same source. However, Figure 3.3 shows that the distance between the blue dots and the green dots increases with the predicted correlation value. Therefore, the optimal threshold

decision should also vary with the predicted correlation value. To have different thresholds for different predicted correlation values, the prediction correlation which takes image content effect into account should also be considered in the identification process. Hence, we extend the traditional 1D classification method to a 2D classification. The traditional classifier only uses the actual correlation for decision. In our 2D classifier, we used both actual and predicted correlation for decision. For both classifiers, the support vector machine (SVM) is used. It was implemented with Matlab default statistical learning toolbox. The kernel function used was the radial basis function.

To show the advantage in using the proposed 2D classifier as compared to the traditional classifier, the 2D classifier was trained from half of the data in Figure 3.3. To give a more reasonable range for the training threshold, some artificial data were generated in SVM training to give an upper and lower bound for the threshold. The artificial data at upper bound was set to have the actual correlation equals to 1.5 times the maximum of the actual correlation from different cameras while that at lower bound was set to 0. Figure 3.11 shows the training results. The traditional method has a fixed threshold irrespective of the predicted correlation value (as shown by a straight horizontal line at actual correlation equals to 0.006). However, the thresholds in the 2D classifier vary with the predicted correlation. The new

threshold is larger than the traditional threshold at mid predicted correlation values while it is smaller than the traditional threshold at low predicted correlation value. Hence, the threshold setting can be made adaptive to the scene content.

After training, the second half of the data in Figure 3.3 was used to test the trained classifier. Both of the traditional and the 2D classifiers were set to have the same false rejection rate (FRR). Figure 3.12 shows a plot of the threshold values of the traditional and the 2D classifiers. For the traditional classifier, the same threshold value is used for all the predicted correlation values. Data at high predicted correlation were easily falsely accepted by the traditional classifier. In contrast, our 2D classifier adopts different threshold values according to the predicted correlation and the actual correlation. Those falsely accepted data were correctly identified by our proposed 2D classifier. This shows the advantage of our proposed 2D classifier which can identify images not solely based on the actual correlation but also based on the prior knowledge of the image content. Figure 3.12 considered the FAR by fixing the FRR. In fact, similar improvement on FRR can be found by fixing the FAR.

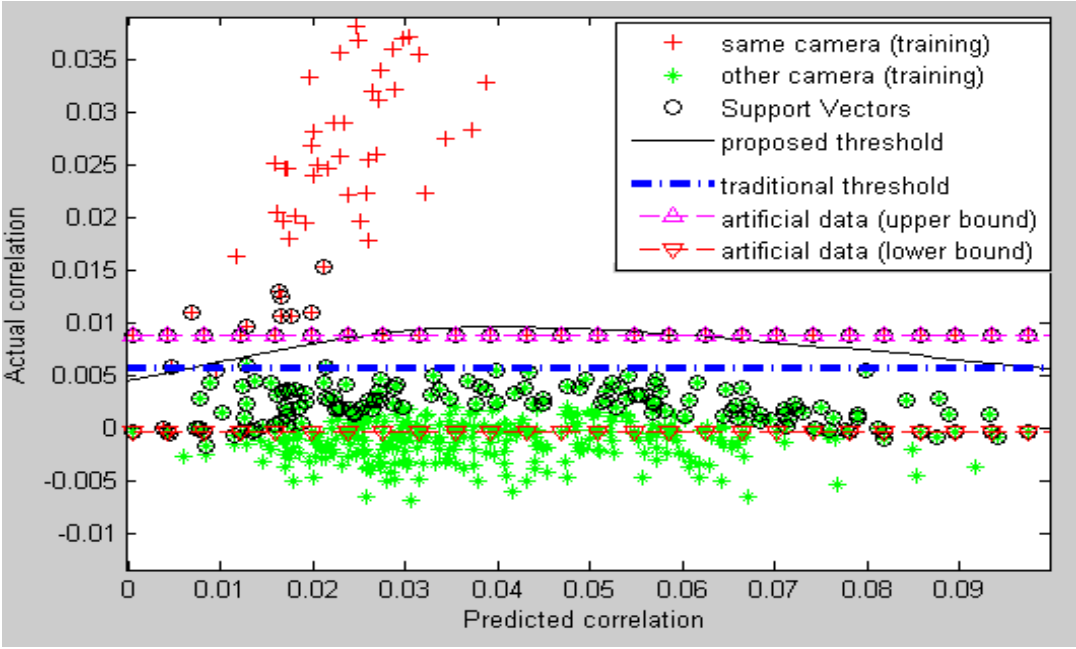


Figure 3. 11 A close up of the proposed threshold and traditional threshold using half of the data in Figure 3.3 for training.

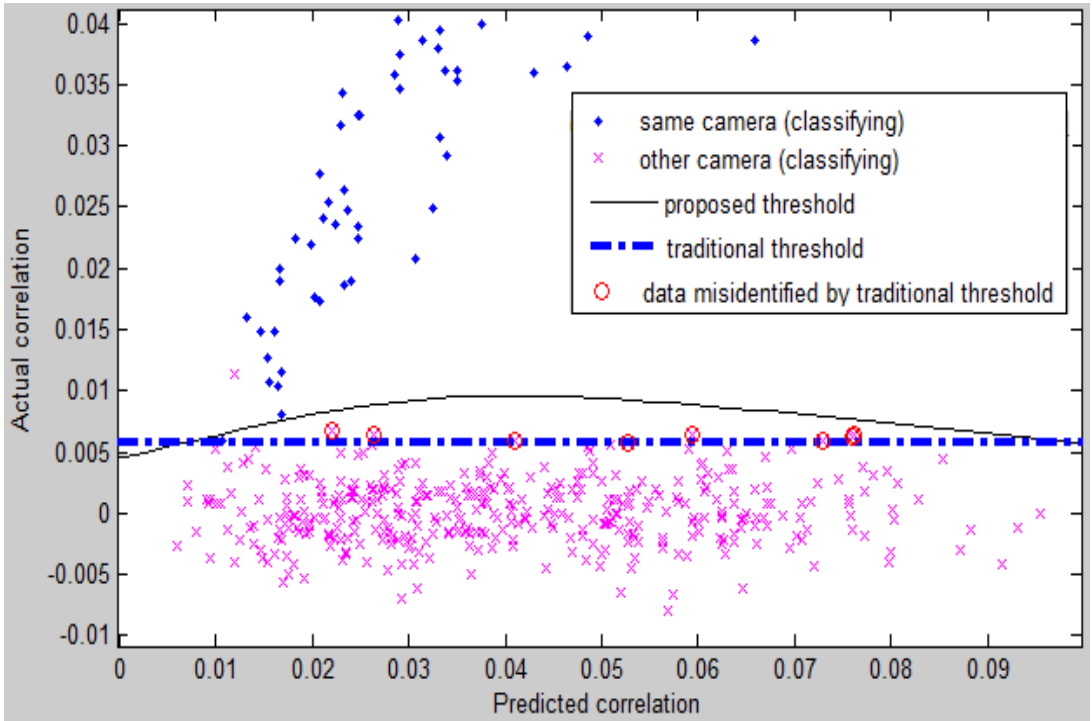


Figure 3. 12 Classifying result using the trained results in Figure 3.11.

3.5.2 Experimental Result

The experimental settings adopted and the data used were the same as that in Section 3.3. As discussed in Section 2.4.2.4, the classification problem is a binary hypothesis problem. Note that H_0 represents the hypothesis that the testing image and the reference PRNU are from different cameras while H_1 represents the hypothesis that they are from the same camera. For each camera, there were 800 images for hypothesis H_0 and 400 for hypothesis H_1. For images in hypothesis H_0, there are 400 images from each of the other cameras in the experiment. To train the 2D classifier, half of the data from hypothesis H_0 and hypothesis H_1 were randomly picked for training. Some artificial data were generated as mentioned in Section 3.5.1 to give the upper and lower bound for the training threshold. Afterwards, the other half of the data from hypothesis H_0 and hypothesis H_1 were used for classification. The average time for running the whole identification process was 30 minutes, 45 minutes and 60 minutes respectively for the Basic approach, the MLE approach and our proposed approach. As the result of SVM classifier varies with different training sets and classifying sets, the SVM training and classifying processes were repeated 50 times with the same pool of data but with random combination in training set and classifying set in order to achieve a reliable result.

To compare the performance of our proposed 2D classifier with the traditional classifier using either Basic Approach [1] or MLE approach [2,3], the FAR in the 2D classifier was used to set the threshold in the traditional classifier in the same training set. The FAR, FRR and accuracy are then compared in the classifying set.

Table 3.1 shows the average performance of our proposed 2D classifier and the traditional classifier over 50 trials for different training and classifying sets. In terms of accuracy and FRR, our proposed method shows improvement in camera 2 but has comparable performance in camera 1 and camera 3. The reason is that, under the case of high accuracy, the overlapping area for null hypothesis and alternative hypothesis data is low. Hence, the performance of using the same threshold and the varying threshold values would be similar. For the case of low accuracy as in camera 2, the advantage of using flexible threshold of the 2D classifier is obvious. The accuracy increases by about 1% to 2% as compared to the Basic or MLE approach.

We have also compared the performance for a smaller block size. By using a smaller block size, the overlapping area between null hypothesis and alternative hypothesis data will increase. The experiment was repeated at 128x128 block size and the result is shown in Table 3.2. From Table 3.2, we can see the improvement in camera 2 is significant in which the accuracy improves by 6% and 4% as compared to the Basic and the MLE Approaches respectively.

Table 3. 1 Average performance in terms of FAR, FRR and accuracy of the proposed 2D classifier and the traditional classifier with a block size of 512x512.

		Camera1	Camera2	Camera3
FAR (%)	Basic approach [1]	0.25	0.23	0.26
	MLE [2]	0.23	0.28	0.28
	Proposed Method	0.13	0.11	0
FRR (%)	Basic Approach [1]	1.32	7.08	1.11
	MLE [2]	1.31	5.65	1.51
	Proposed Method	1.42	4.79	1.24
Accuracy (%)	Basic Approach [1]	98.43	92.69	98.63
	MLE [2]	98.46	94.07	98.21
	Proposed Method	98.45	95.10	98.76

Table 3. 2 Average performance in terms of FAR, FRR and accuracy of the proposed 2D classifier and the traditional classifier with a block size of 128x128.

		Camera1	Camera2	Camera3
FAR (%)	Basic Approach [1]	2.13	2.42	4.23
	MLE [2]	2.14	2.44	4.28
	Proposed Method	2.13	2.40	4.16
FRR (%)	Basic Approach [1]	11.82	21.41	20.27
	MLE [2]	11.90	19.36	18.47
	Proposed Method	11.46	15.54	17.95
Accuracy (%)	Basic Approach [1]	86.05	76.17	75.5
	MLE [2]	85.96	78.20	77.25
	Proposed Method	86.41	82.06	77.89

3.6 Chapter Summary

In this chapter, we have investigated the relation between the scene content effect and the identification accuracy. We have shown that the PDF of H_1 can greatly be affected by the image content. Hence, it is important to find a way to characterize the scene content effect. We have used a 2D Gaussian function to model the data. The statistics of H_0 and H_1 can be obtained at different predicted correlation values. Hence, the detection accuracy can be obtained. Using this method, the relation between the scene content effect and the identification accuracy can be obtained. Our experimental results have shown that the detection accuracy will be dropped significantly at low predicted correlation value.

As the predicted correlation is able to characterize the scene content, we have extended the traditional classifier to a 2D classifier. There are two features used in the 2D classifier, they are the actual correlation and the predicted correlation. The actual correlation shows the similarity of the pattern noise of a testing image with the reference PRNU of a particular camera while the predicted correlation shows the seriousness of the scene content effect. Experimental results show that the proposed 2D classifier provides a flexible threshold setting mechanism according to the seriousness of the scene content. Hence, the 2D classifier is able to provide an accurate source camera identification.

Chapter 4 A Confidence Map and Pixel-Based Weighted Correlation for PRNU-based Camera Identification

4.1 Introduction

In Chapter 3, the scene content effect for testing images is formulated by a correlation predictor. Its objective is to determine the reliability of the PRNU feature for each block in an image. However, the scene content effect model from the correlation predictor only gives an overall model for the whole image, it cannot determine precisely which small area in a block of an image is reliable. Therefore, in this chapter, the study of scene content effect is extended to a finer level where the scene content effect can be assessed to the pixel level.

In this chapter, a confidence map and a pixel-based weighted correlation method for digital camera identification using the photo-response non-uniformity (PRNU) is proposed. In traditional camera identification method, a simple denoising technique is used to extract the PRNU feature as the difference between the original image and its denoised version. One of the major problems is that the image content is left behind in the noise residue which affects the correlation calculation for identifying the source camera. In order to solve the problem, this chapter first studies the image content effect by examining the relation between different image features and correlation. Afterwards, the image content effect is formulated by a non-linear regression model.

Then this relationship is used to obtain a confidence map for the testing images. The aim of the confidence map is to show the reliability of each pixel in correlation calculation. It can then be used as a weighting function so as to give a higher weight for pixel which is more reliable and a lower weight for less reliable pixel.

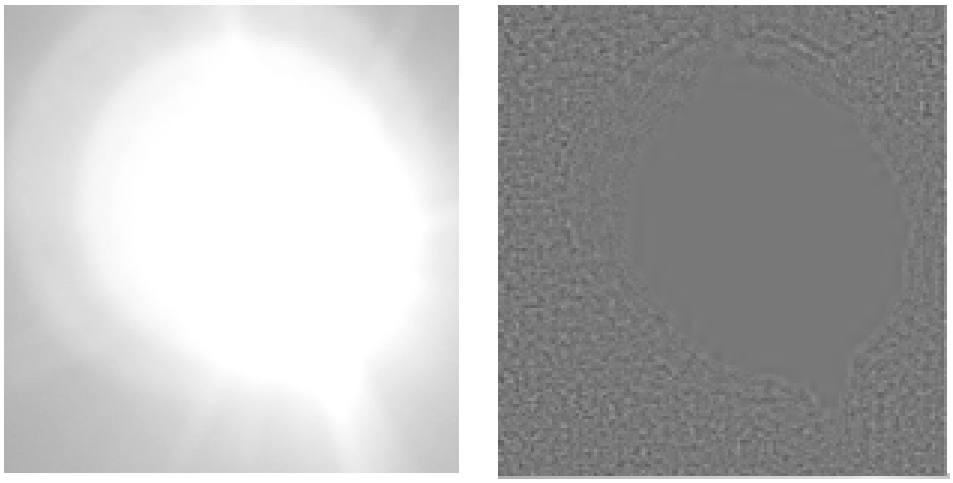
The rest of this chapter is organized as follows. Section 4.2 first describes the image content effect on camera identification using pattern noise. A non-linear regression model will be discussed in Section 4.3. The confidence map and weighted correlation will be given in Section 4.4. Experimental results are shown in Section 4.5. The use of the confidence map in existing PRNU-based camera identification is discussed in Section 4.6 while a comparative study is described in Section 4.7. Finally, Section 4.8 concludes this chapter.

4.2 The Image Content Effect

Image content can seriously affect the accuracy of the camera identification. There are mainly two factors, namely the image intensity and the image texture. The PRNU is a multiplicative factor to the intensity as shown in equation (2.14). The PRNU characteristics can be detected easily with the increase in the image intensity. However, if the image intensity reaches the maximum level, i.e., it is saturated, the PRNU feature would be lost completely in such a case. Figure 4.1(a) and Figure 4.1(b)

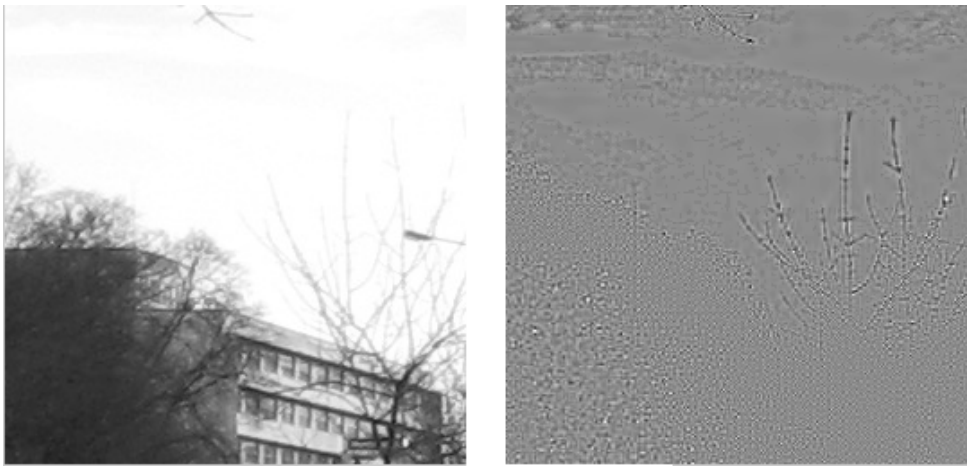
show respectively an image of the sun and its corresponding noise residue. As evidenced in Figure 4.1(b), the PRNU feature is completely missing in the saturated area of the sun.

Besides, image texture is another factor affecting the PRNU extraction. For texture/edge regions, the high frequency components will be left in the noise residue after denoising. These high frequency terms contaminate the extracted PRNU. Figure 4.2(a) and Figure 4.2(b) show respectively an outdoor street scene and its corresponding noise residue. Tree details are clearly left in the noise residue. These two examples show that scene characteristics can be left in the noise residue which in turn affects the extracted PRNU. Thus, the effect of the scene content on the correlation should be studied. In particular, this section investigates the relationship between the correlation and image features. We considered both image intensity and image texture.



(a) (b)

Figure 4. 1 (a) the sun image and (b) its noise residue *W*.



(a) (b)

Figure 4. 2 (a) an outdoor street image and (b) its noise residue *W*.

To obtain the correlation, an image is divided into a number of non-overlapping blocks. The block size was set to be 128x128 in order to have a homogeneous content within each block and at the same time have a statistically stable correlation calculation. The image intensity feature f_{bi} and image texture feature f_{bt} are defined respectively as,

$$f_{bi} = \frac{1}{|B_b|}\sum_{i\in B_b}(I[i]), \tag{4.1}$$

$$f_{bt} = \frac{1}{|B_b|}\sum_{i\in B_b}\frac{1}{1+var_5(I[i])}, \tag{4.2}$$

where B_b is the pixel block used to calculate the correlation, $I[i]$ is the intensity value at pixel i and $var_5(I[i])$ is the variance of intensity I at pixel i in the 5x5 neighborhood. The image intensity feature and texture feature have been defined differently as those in Chapter 2. The definition for intensity and texture features in

equations (4.1) and (4.2) has been simplified as compared to their corresponding definition in equations (2.30) and (2.32). For example, there is no attenuation function defined in the intensity feature in equation (4.1) so that there is no need to determine the parameters I_{crit} and τ. Besides, the texture feature is extracted in the spatial domain rather than the wavelet domain so that the computational complexity is reduced.

Figure 4.3 and Figure 4.4 show respectively the variation of the correlation with respect to the image intensity feature and image texture feature for camera Minolta DiMAGE X. The correlation generally increases with the image intensity, but drops when the intensity is nearly saturated. The correlation also drops sharply when the image texture feature is bigger than about 0.6. Due to the nonlinearity shown in Figure 4.3 and Figure 4.4, the linear prediction model in equation (2.3.5) in Section 2.4.3.4 is not appropriate. Therefore, we propose to use the Kernel principal component analysis (KPCA) to model and formulate the scene content effects. The KPCA is a nonlinear prediction model in which image features can be projected to a high dimensional space to deal with the nonlinear relation between image feature and the correlation. Hence KPCA uses more components for the regression model as compared to the linear model which can help to achieve a better regression result [50].

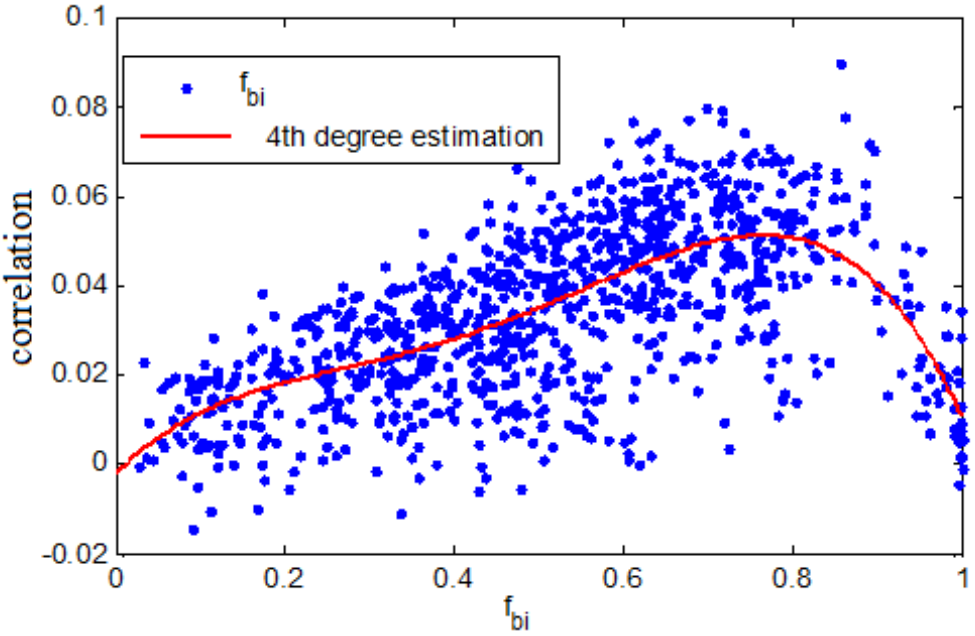


Figure 4. 3 Plot of correlation vs f_{bi} for 100 images with each image containing ten 128x128 blocks in Minolta DiMAGE X and the 4th degree estimation for the relationship.

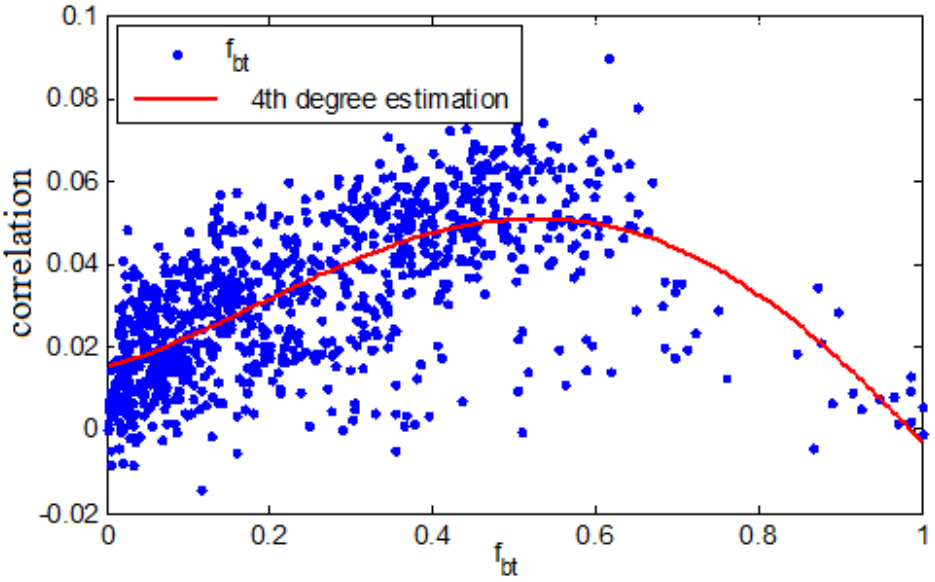


Figure 4. 4 Plot of correlation vs f_{bt} for 100 images with each image containing ten 128x128 blocks in Minolta DiMAGE X and the 4th degree estimation for the relationship.

4.3 Formulation of Image Content Effect Using Kernel Principal Component Analysis (KPCA)

Because of the nonlinear relation between the correlation and the image features including both intensity and texture, KPCA is proposed to study the image content effect. Let $\Phi(\boldsymbol{X})$ be a nonlinear mapping of X from input space Y to the feature space F, where $X \in Y$ and $\Phi(\boldsymbol{X}) \in F$. Then the regression model can be expressed as,

$$G=\Phi(\boldsymbol{X})\boldsymbol{w}+e \tag{4.3}$$

where G is a column matrix consisting of correlation between the PRNU from the reference camera and the pattern noise from the training images, $\Phi(X)$ is a matrix whose column consisting of image features and row corresponding to the training data, w is a column matrix consisting of weighting for each feature and e is a column matrix corresponding to noise. The nonlinear mapping function $\Phi(X)$ is not easy to determine. In KPCA, a kernel matrix K is constructed so that there is no need to estimate $\Phi(\boldsymbol{X})$. The kernel matrix is defined as the inner product of $\Phi(X)$ in the feature space [50], i.e.,

$$\boldsymbol{K}=\Phi^{\mathrm{T}}(\boldsymbol{X})\,\Phi(\boldsymbol{X}). \tag{4.4}$$

where Φ^{T} denotes the transpose of Φ . Performing standard principal component analysis (PCA) on the kernel matrix gives,

$$K=UDU^{\mathrm{T}} \tag{4.5}$$

where columns of U are the eigenvectors of K and D is a diagonal matrix with eigenvalues in the diagonal elements. The principal axes of $\Phi(X)$ can be written as,

$$Y=\Phi(X)\, UD^{-\frac{1}{2}}. \tag{4.6}$$

The principal component $\widehat{\Phi}(X)$ of $\Phi(X)$ can be obtained and expressed in terms of K, U and D as follows,

$$\begin{aligned}\widehat{\Phi}(X)&= Y^{T}\Phi(X)\\ &=D^{-\frac{1}{2}}U^{T}\,\Phi^{T}(X)\,\Phi(X)\\ &= D^{-\frac{1}{2}}U^{T}K\end{aligned} \tag{4.7}$$

From equation (4.7), the principal component $\widehat{\Phi}(X)$ of $\Phi(X)$ can be estimated without knowing $\Phi(X)$. Now the KPCA regression model becomes,

$$G=\widehat{\Phi}(X)w+e. \tag{4.8}$$

Similar to the linear correlation predictor in equation (2.35) in Section 2.4.3.3, the weighting w in each feature can be solved under a least squares framework, i.e., w can be obtained by,

$$w=(\widehat{\Phi}(X)^{T}\,\widehat{\Phi}(X))^{-1}\,\widehat{\Phi}(X)^{T}G. \tag{4.9}$$

Afterwards, the estimated w and the KPCA projected features can be used to predict the correlation between the PRNU from the reference camera and the pattern noise

from the testing image. Mathematically, the predicted correlation can be written as,

$$\hat{G}=\hat{\Phi}(X)w. \tag{4.10}$$

The predicted correlation can be used to determine the severity of image content effect. A low predicted correlation means a large amount of image content is left in the noise residue. Hence, the image content mixes with the PRNU characteristics in that region. On the other hand, a large predicted correlation means that the region does not suffer from image content effect and hence the PRNU characteristics estimated from the region is reliable.

4.4 Confidence map

The KPCA method can be used to quantify the seriousness of the image content effect and the reliability of each pixel in correlation calculation. For example, a saturated region will have a low predicted value which implies that the region is unreliable in correlation calculation. Hence, equation (4.10) can be used to obtain the predicted correlation and determine whether the pixel is suitable for camera identification or not. There are two phases in the proposed estimation process, namely the training phase and the testing phase. In the training phase, an image is divided into a number of non-overlapping blocks. For each block, two image features are obtained. They are the image intensity and image texture features as in equation (4.1) and

equation (4.2) respectively. Besides, the correlation between the PRNU from the reference camera and the pattern noise from the training image can be obtained as in equation (2.16) in Section 2.4.2.3. After that, the weighting w can be obtained from equation (4.9). In order to have a good estimation in correlation, the block size cannot be too small. However, the block size cannot be too large to avoid a heterogeneous content within the block. The block size was chosen to be 128×128. In fact, similar results can be obtained if the block size is 64×64.

In the testing phase, the block size is set to be 1×1 so that a pixel-based reliability analysis can be achieved. In particular, the confidence map is generated using equation (4.10) for a testing image. This confidence map is able to characterize the reliability of each pixel in correlation calculation. In other words, a more reliable pixel in the testing image will have a larger value in the confidence map. This pixel-based weighting mechanism is much more flexible than the block-based weighting mechanism in [2,3] as it is able to quantify the image content effect in each pixel and determine the reliability of each pixel in the testing image in correlation calculation.

Figure 4.5(a) shows an image of a snow mountain. The image contains saturated regions, as well as low and high variance regions. As discussed, these kinds of image content will affect the correlation calculation. Figure 4.5(b) shows the corresponding

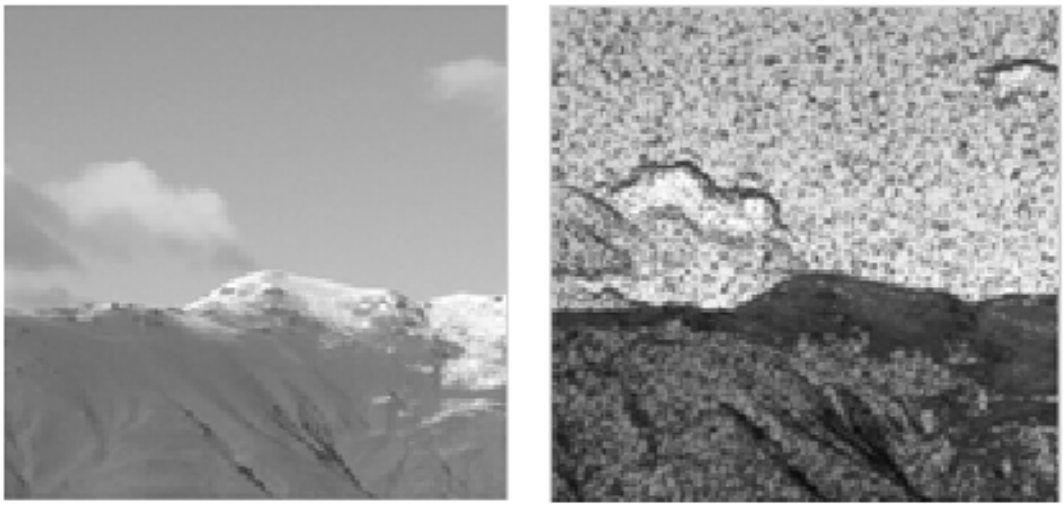
confidence map. The intensity of the confidence map shows the reliability of each pixel in calculating the correlation. The bright area in the confidence map means a reliable location for PRNU estimation and vice versa. From Figure 4.5(b), we can see that pixels in the sky area are more reliable than that in the mountain in PRNU estimation as the intensity level is higher in the sky region. The edge of the mountain shows a dark area in the confidence map because the texture complexity would affect the extracted PRNU. Besides, the saturated area of the snow mountain also shows a dark intensity level in the confidence map since there is no PRNU feature in the saturated area. In this example, the testing image contains different types of image features. The confidence map is able to determine the suitability of different parts of the image for camera identification.

To incorporate the confidence map information into the camera identification framework, a weighted function is applied in calculating the correlation. For each pixel, the reliability is different due to the image feature at that pixel. A weighted function can be used to give weighting according to the reliability of the pixel so as to reduce the effect of image content contamination to the PRNU feature. The confidence map can help to achieve such purpose. The predicted correlation $\hat{G}$ in the confidence map can be treated as a weighting function. Let the noise residue, image intensity of the testing image, and the reference PRNU be W_p, I_p and $\widehat{K}_c$

respectively. Then the weighted correlation by incorporating the confidence map $\hat{G}$ is defined as,

$$corr\left(W_p, I_p\hat{K}_c; \hat{G}\right) = \frac{(\hat{G}(W_p - \overline{W_p}) \odot \left(I_p\hat{K}_c - \overline{I_p\hat{K}_c}\right))}{\left\|\sqrt{\hat{G}}(W_p - \overline{W_p})\right\| \left\|\sqrt{\hat{G}}(I_p\hat{K}_c - \overline{I_p\hat{K}_c})\right\|}. \quad (4.11)$$

$corr\left(W_p, I_p\hat{K}_c; \hat{G}\right)$ can help reducing image content effect as a low weighting (i.e., small $\hat{G}$ value in certain pixels) would be given to saturated regions or highly textured regions. In some cases, $\hat{G}$ could be very small for the whole image. This implies poor reliability caused by serious image content effect over the whole image which would cause numerical instability in equation (4.11). So a threshold l is set such that if the average of $\hat{G}$ is greater than l, the weighted correlation in equation (4.11) is calculated, otherwise correlation in equation (2.16) is used instead.



(a) (b)

Figure 4. 5 (a) A testing image and (b) its confidence map.

4.5 Experimental Results

There were six cameras used in the experiment. Table 4.1 summarizes the camera model, sensor type, resolution and picture format of these cameras. Images were acquired in default camera setting and were taken under different environments, from indoor to outdoor at different times and locations. To evaluate the performance of identifying images from the same camera model, three iphone 4S were included in the experiment.

For each camera, 100 images were used to estimate the PRNU fingerprint. To evaluate the KPCA regression method, another 100 images were used for selecting the kernel coefficients as well as the feature weightings in the regression model. Another 100 images were then used for evaluating source camera identification performance. The classification method was a binary hypothesis test. The classification was done for each camera. In each classification, 100 images from the same camera were treated as the positive samples, and 100 images from each of the other cameras were treated as the negative samples.

Table 4. 1 Camera Details used in the experiment

Camera model	Sensor	Resolution	Format
Konika Minolta DiMAGE X	1/2.7" CCD	2816x2112	JPEG
Cannon IXUS 65	1/2.5" CCD	2048x1536	
Sony DSC T-500	1/2.3" CCD	3048x2736	
Apple iphone 4S (3 iphones)	1/3.2" CCD	3264x2448	

A Performance of regression model

The root mean square error (RMSE) between the predicted correlation and the actual correlation was used to evaluate the performance of the regression model. It is defined as,

$$RMSE=\left(\frac{\sum_{t=1}^{N}\hat{G}(t)-G(t)}{\sum_{t=1}^{N}G(t)}\right)^{\frac{1}{2}}, \quad (4.12)$$

where N is the total number of the testing images, $\hat{G}(t)$ and $G(t)$ are the predicted correlation and the actual correlation of the t-th testing image respectively. The heavy-tailed Radial Basis Function (RBF) was selected as the kernel function in the KPCA regression because of its good performance as compared to the polynomial or Gaussian radial basis kernel [50]. The heavy-tailed RBF is defined as,

$$K(\mathrm{x_i}, \mathrm{x_j}) = exp\left(-\sum_{n=1}^{N} |x(n)_i^a - x(n)_j^a|^b\right). \quad (4.13)$$

where $\mathrm{x(n)_i}$ and $\mathrm{x(n)_j}$ are the *i*th and *j*th input image feature of the *n*th correlation observation. There are two parameters in RBF, namely *a* and *b*. By using different combinations of *a* and *b*, the one that achieves the minimum RMSE would be used in the KPCA regression model.

In the experiment, 100 images were used to train the regression model. For each image, four 128x128 image blocks were used for feature extraction and correlation calculation. Half of the images were randomly selected to train the weighting of the features for different combinations of *a* and *b* in equation (4.13). Then another half of the images were used to test the performance of the regression model for each combination of *a* and *b*.

Table 4.2 shows the average change of RMSE of the proposed regression model among all cameras as compared to the traditional regression model in Section 2.4.3.4. The block size for calculating the predicted correlation $\widehat{\mathrm{G}}(\mathrm{t})$ and the actual correlation $\mathrm{G}(\mathrm{t})$ was 512x512. The negative value in RMSE means that the proposed method has a better performance than the traditional regression model in modeling the image content effect. Table 4.2 shows that the combination *a*=0.5 and *b*=1 has the best performance in the data set. Hence they were chosen for the non-linear model in subsequent experiments.

Table 4. 2 Percentage change in RMSE of the proposed KPCA regression model as compared to the traditional regression method.

	b=2	b=1	b=0.5
a=1	0.026	-2.34	-2.63
a=0.5	-2.32	-2.66	-2.61
a=0.25	-1.31	-2.41	-2.50
a=0.125	-0.62	-1.93	-2.51

B. Performance of the weighted correlation

The kernel parameters having the minimum RMSE is used to project the image features in the pixel domain into a higher dimension in the KPCA model. By using the regression model in equation (4.10), a confidence map for each testing image can be generated. Weighted correlation using the confidence map is calculated according to equation (4.11). Table 4.3 shows the average correlation and the weighted correlation of the testing images with its corresponding source camera using a block size of 512x512. We can see that the correlation of images from same camera increased by using the confidence map. This indicates that the proposed weighting function based on the confidence map can eliminate the image content effect according to the image

features of the testing images. For weighted correlation between the reference PRNU and pattern noise of testing images from different cameras, as the two pattern noises are independent of each other, the weighting function does not have any effect on the weighted correlation calculation. In other words, the correlation and the weighted correlation values are more or less the same.

Table 4. 3 Average correlation and weighted correlation on the same camera cases for 100 images with a block size of 512x512.

Camera	Average correlation	Average weighted correlation
Konika Minolta DiMAGE X	0.0352	0.0369 (4.8%)
Cannon IXUS 65	0.0274	0.0286 (4.4%)
Sony DSC T-500	0.0217	0.0219 (0.9%)
iphone 4S -A	0.0570	0.0576 (1.1%)
iphone 4S -B	0.0760	0.0782 (2.9%)
iphone 4S -C	0.0362	0.0373 (3.0%)

C. Overall Performance

Using the trained results as discussed in subsections *A* and *B*, we tested the performance of the proposed confidence map on source camera identification. The Receiver Operating Characteristics (ROC) curve is obtained by using LIBSVM tool [51] in the implementation. Figure 4.6 and Figure 4.7 show the overall ROC curve for different cameras with a block size of 512x512 and 256x256 respectively. We can see that the true positive rates are improved by using the proposed confidence map at all false positive rates. Hence, the proposed method is able to enhance the existing PRNU-based identification methods and achieves a better performance in terms of ROC curves. This improvement is consistent with different image sizes.

Besides, the compression effect on the proposed method was also examined. The same set of images was used but compressed at the quality factor of 90, 70 and 50 in JPEG format. Results were shown in Figure 4.8 to Figure 4.13. As the PRNU feature is affected by the compression, the accuracy is reduced in general. But still, the proposed method is able to enhance the results. The identification accuracy is improved by using the proposed weighting scheme through the confidence map, at all the compression ratios and for all block sizes. This proves the effectiveness of the proposed scheme in reducing the scene content effect in camera identification.

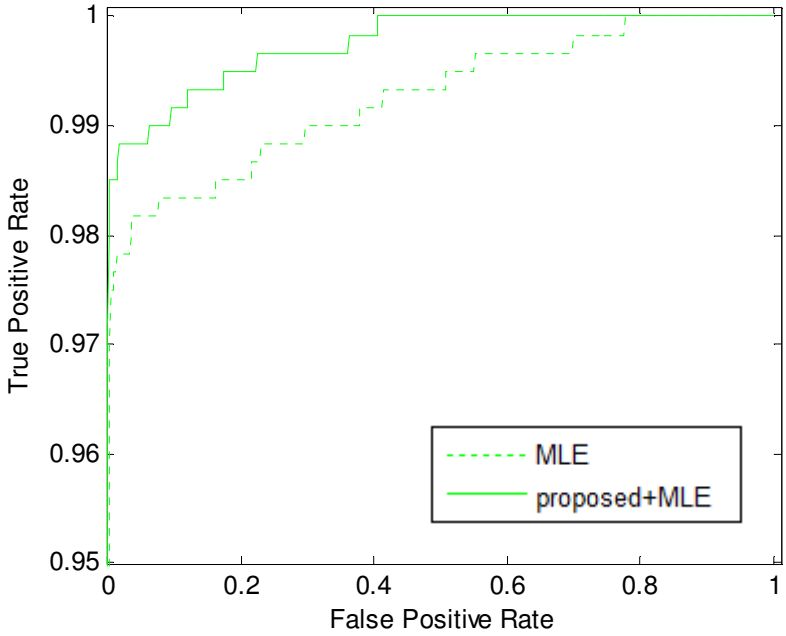


Figure 4. 6 Overall ROC curves using a block size of 512x512.

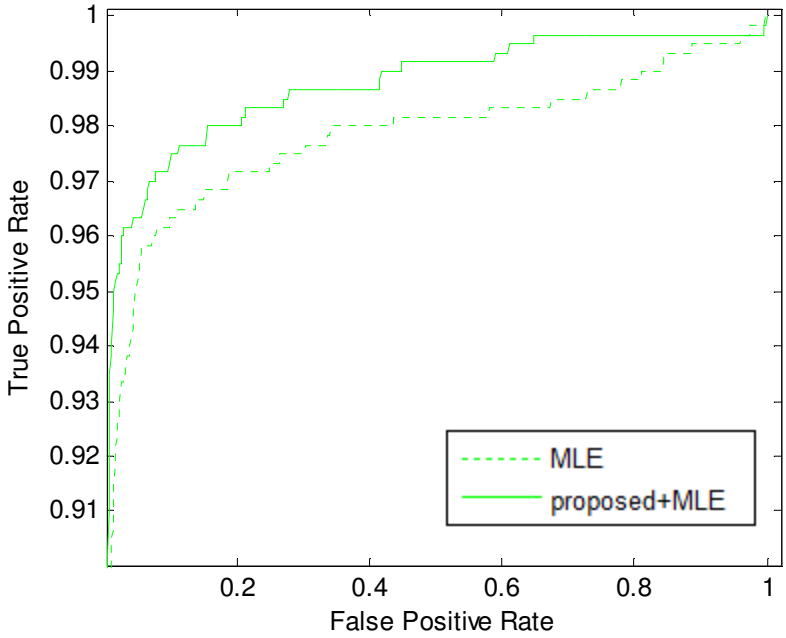


Figure 4. 7 Overall ROC curves using a block size of 256x256.

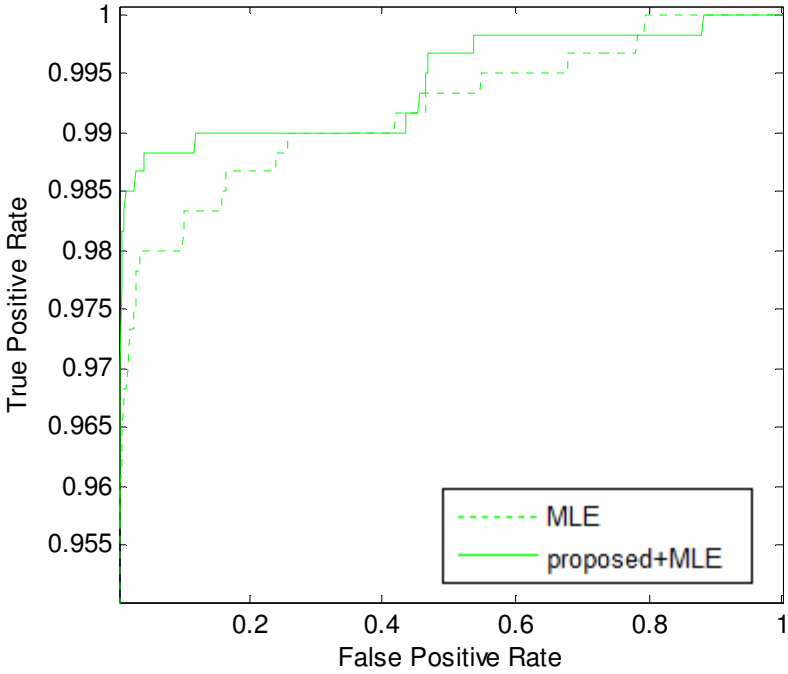


Figure 4. 8 Overall ROC curves using a block size of 512x512 at JPEG quality factor of 90.

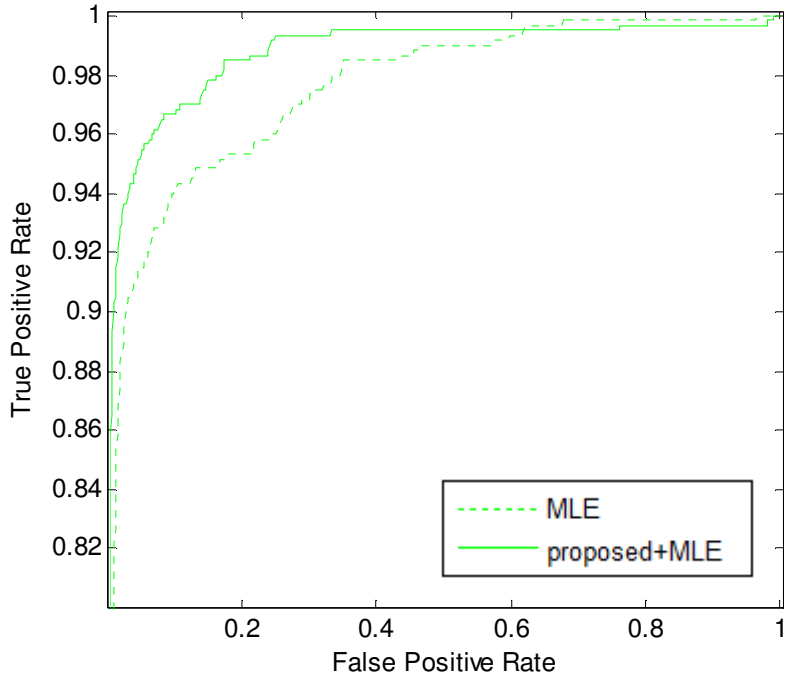


Figure 4. 9 Overall ROC curves using a block size of 256x256 at JPEG quality factor of 90.

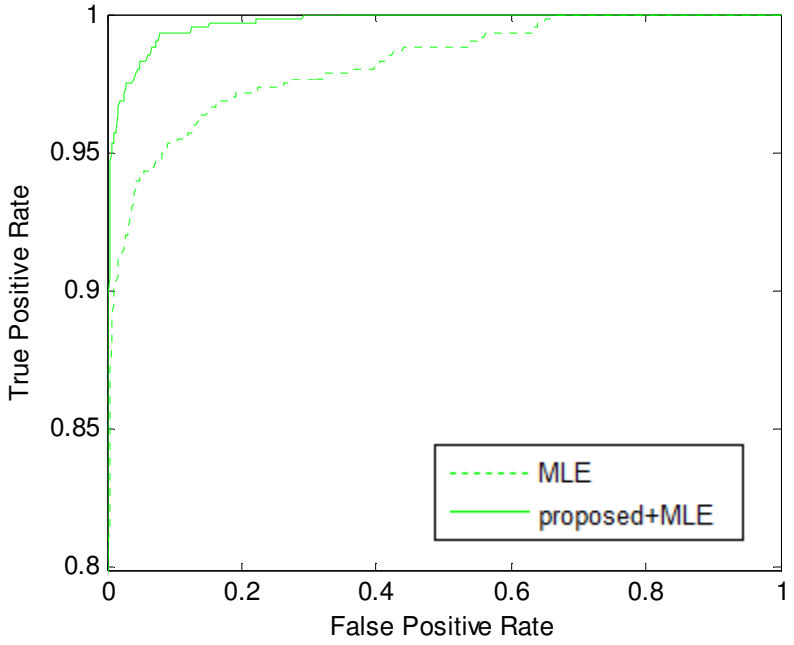


Figure 4. 10 Overall ROC curves using a block size of 512x512 at JPEG quality factor of 70.

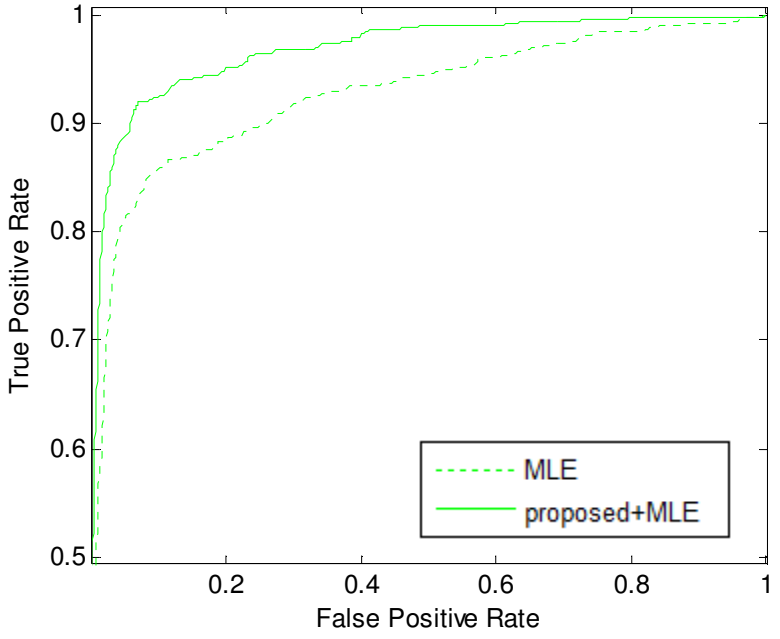


Figure 4. 11 Overall ROC curves using a block size of 256x256 at JPEG quality factor of 70.

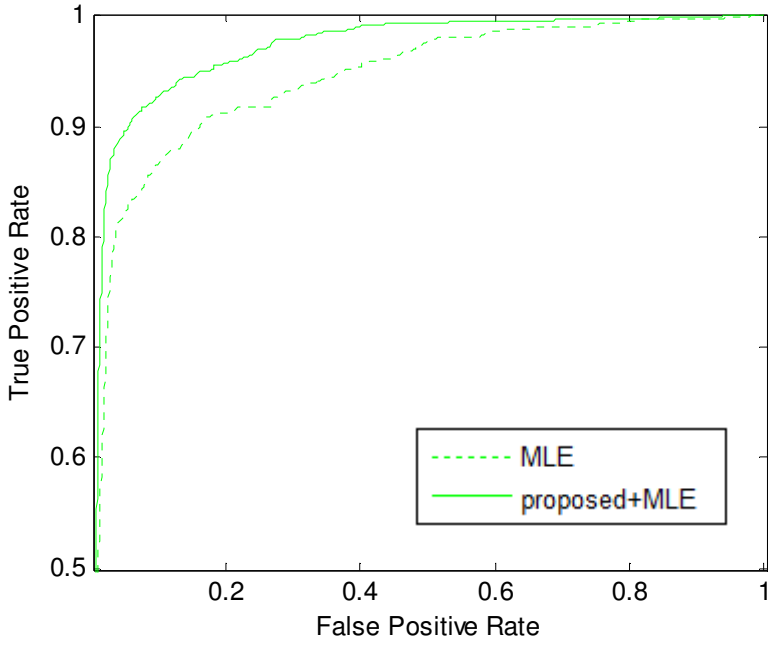


Figure 4. 12 Overall ROC curves using a block size of 512x512 at JPEG quality factor of 50.

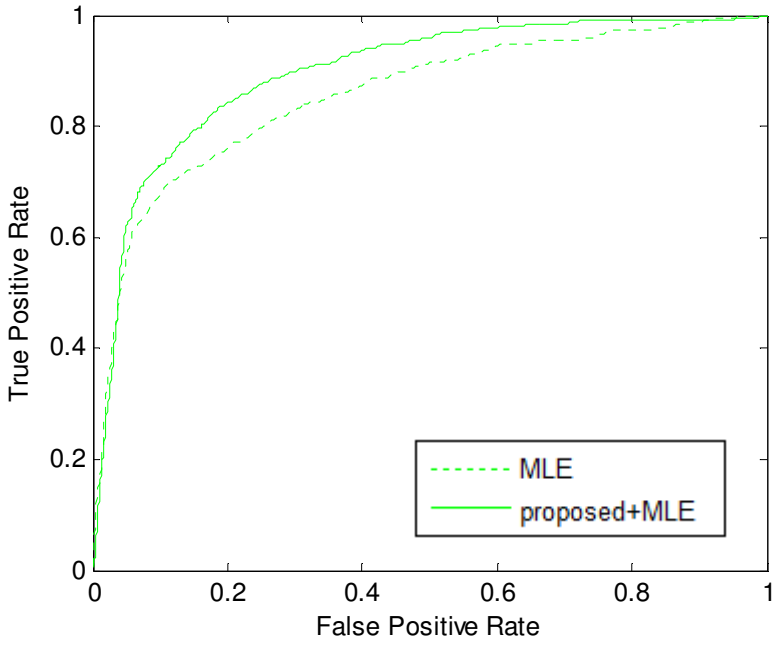


Figure 4. 13 Overall ROC curves using a block size of 256x256 at JPEG quality factor of 50.

4.6 Use of Confidence Map in other PRNU-based Camera Identification Methods

In Section 4.5, the results were based on the maximum likelihood approach (MLE) [2,3]. In fact, the proposed confidence map can work with other existing PRNU-based camera identification methods. In this part, the proposed weighting scheme was applied to three different existing PRNU-based camera identification methods. They are the Basic approach (Section 2.4.2) [1], the phase PRNU approach (Section 2.4.4) [4] and Li's methods [13]. For Li's method, we selected model 3 and model 5 which have the best performance in their study for comparison. To use the confidence map in other state of the art camera identification technology, we replace the correlation in equation (2.16) by the weighted correlation in equation (4.11). The experimental settings remain the same as those in Section 4.5.

Figure 4.14 and Figure 4.17 show the ROC curve for different cameras with a block size of 512x512 and 256x256 respectively. Similar to the results obtained for MLE approach, we can see that the true positive rates are improved by using the proposed confidence map at all false positive rates. Hence, the proposed weighting scheme is able to enhance the existing identification methods and achieve a better performance. We have also examined the compression effect. Figure 4.18 to Figure 4.29 show the results. We can see that the identification accuracy is improved by using the confidence map for all the methods and at all the quality factors.

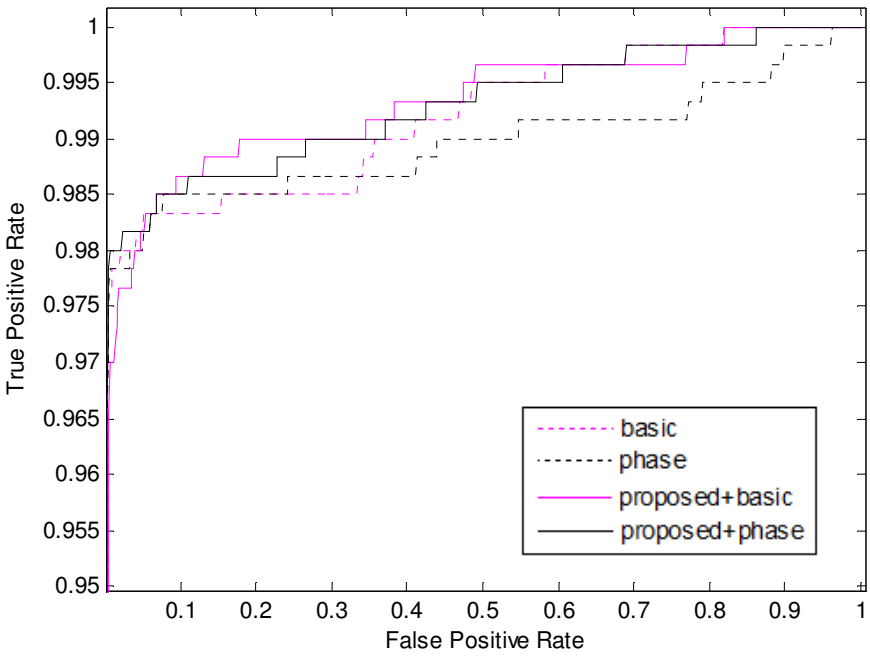


Figure 4. 14 Overall ROC curves for the Basic approach and the phase approach using the proposed confidence map with a block size of 512x512.

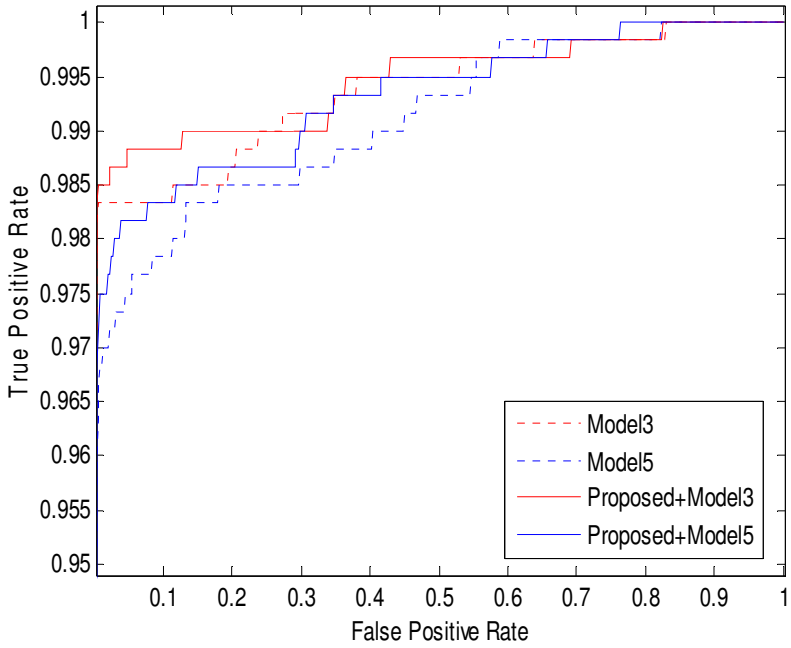


Figure 4. 15 Overall ROC curves for Li's model3 and model 5 with a block size of 512x512.

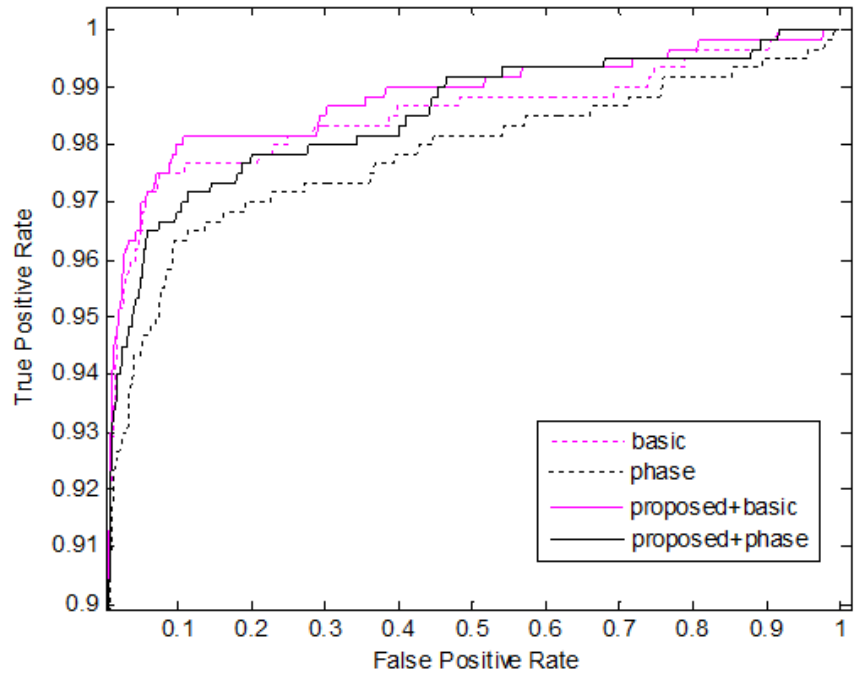


Figure 4. 16 Overall ROC curves for the Basic approach and the phase approach using the proposed confidence map with a block size of 256x256.

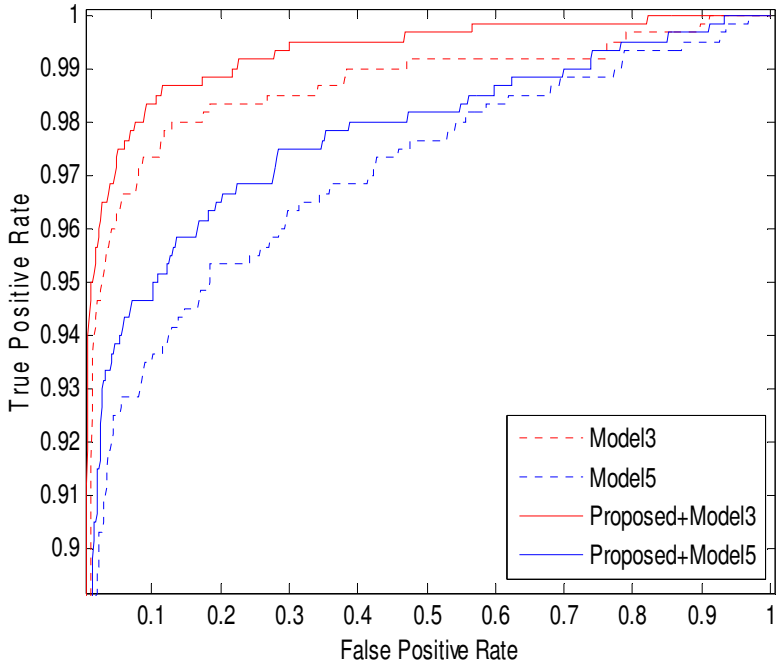


Figure 4. 17 Overall ROC curves for Li's model3 and model 5 with a block size of 256x256.

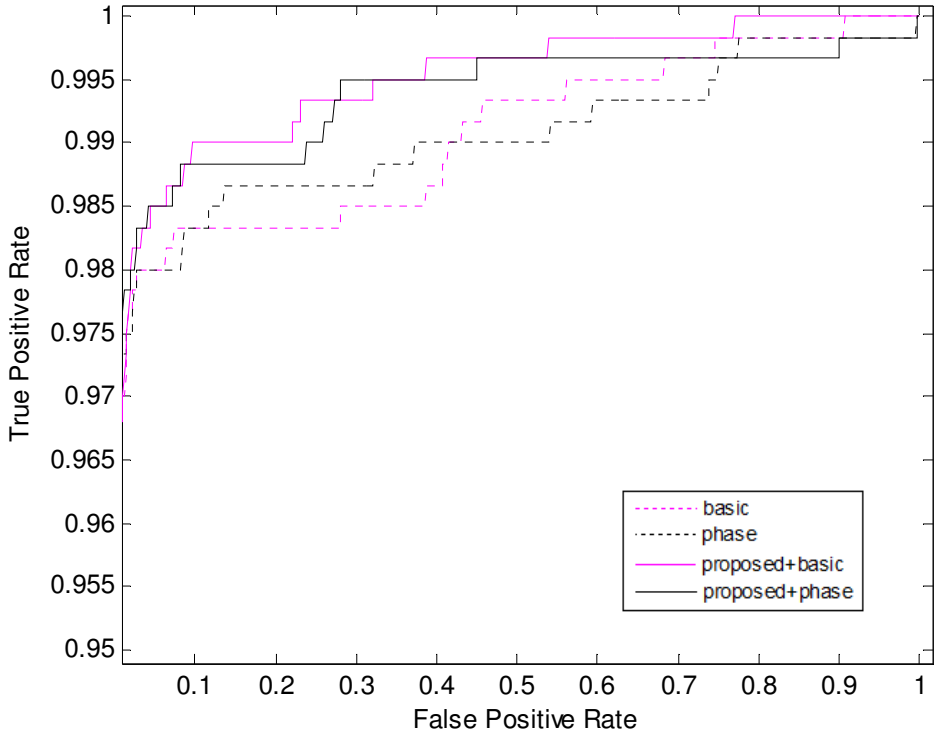


Figure 4. 18 Overall ROC curves for the Basic approach and the phase approach using the proposed confidence map with a block size of 512x512 at JPEG quality factor of 90.

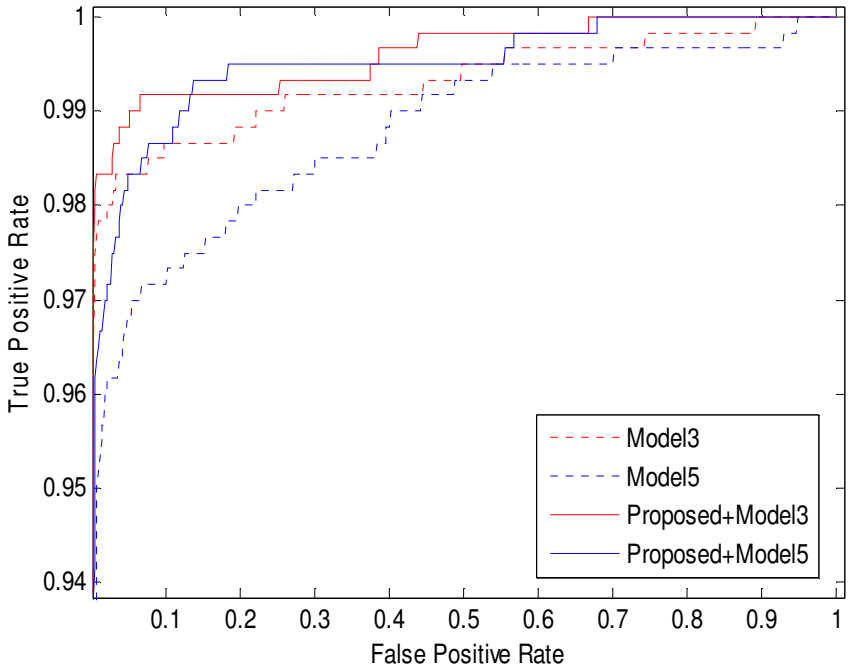


Figure 4. 19 Overall ROC curves for Li's model3 and model 5 with a block size of 512x512 at JPEG quality factor of 90.

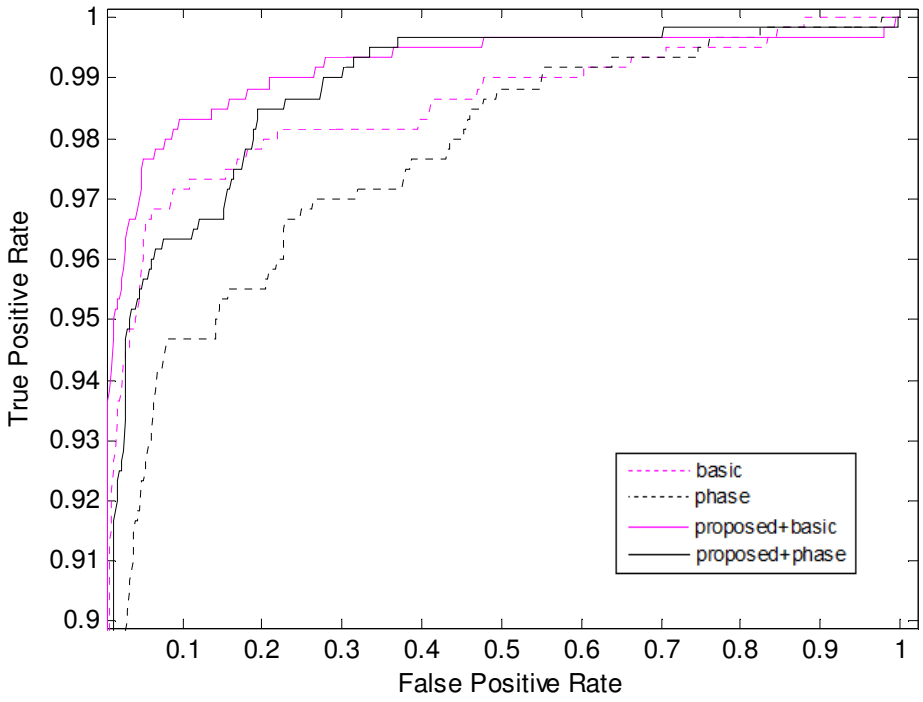


Figure 4. 20 Overall ROC curves for the Basic approach and the phase approach using the proposed confidence map with a block size of 256x256 at JPEG quality factor of 90.

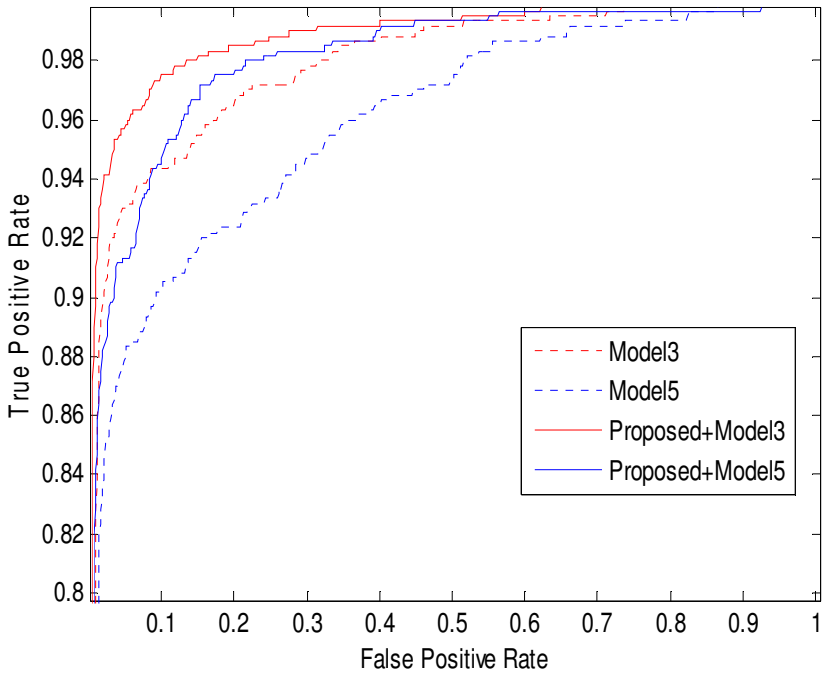


Figure 4. 21 Overall ROC curves for Li's model3 and model 5 with a block size of 256x256 at JPEG quality factor of 90.

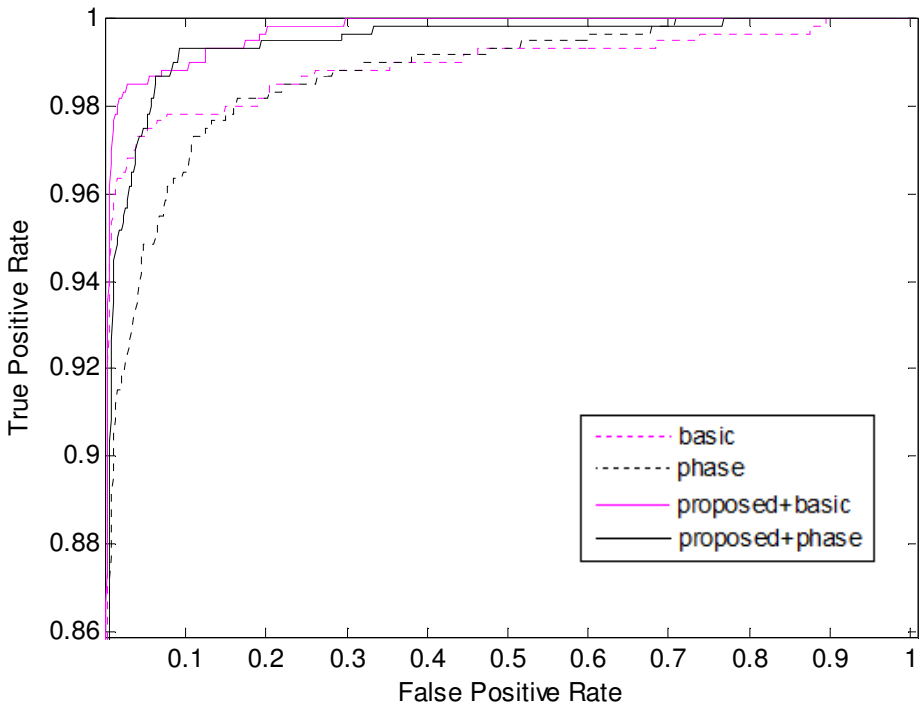


Figure 4. 22 Overall ROC curves for the Basic approach and the phase approach using the proposed confidence map with a block size of 512x512 at JPEG quality factor of 70.

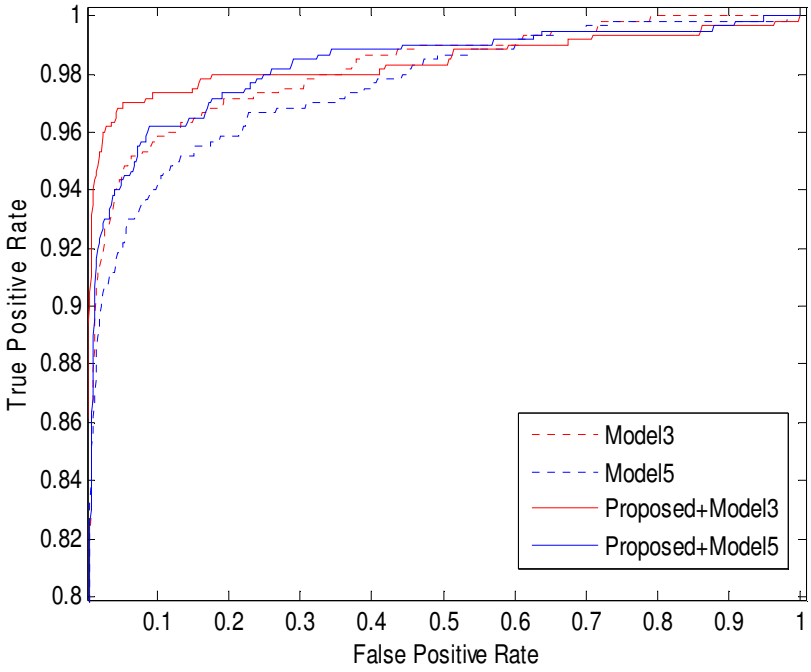


Figure 4. 23 Overall ROC curves for Li's model3 and model 5 with a block size of 512x512 at JPEG quality factor of 70.

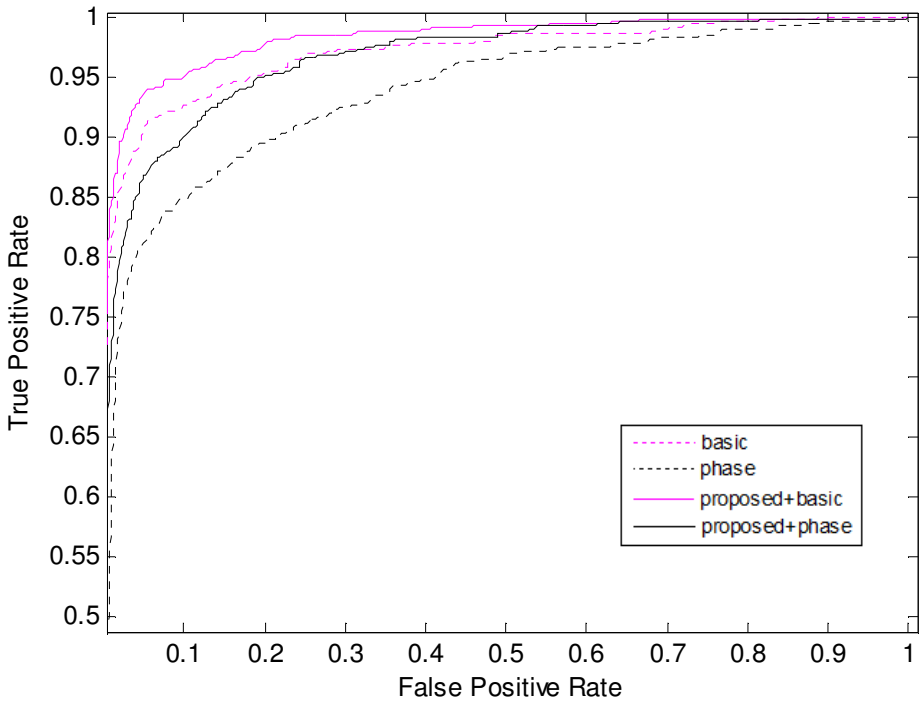


Figure 4. 24 Overall ROC curves for the Basic approach and the phase approach using the proposed confidence map with a block size of 256x256 at JPEG quality factor of 70.

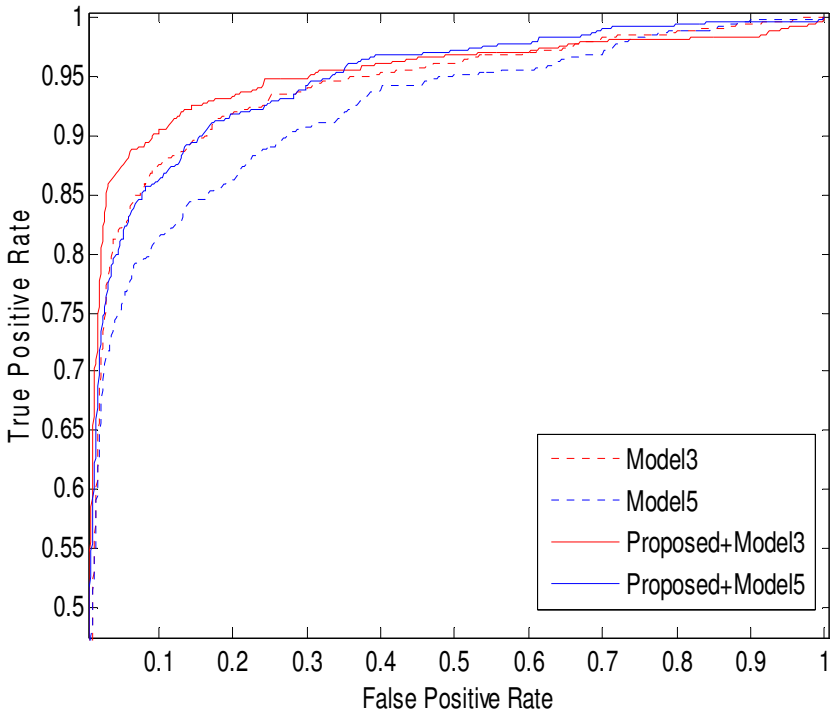


Figure 4. 25 Overall ROC curves for Li's model3 and model 5 with a block size of 256x256 at JPEG quality factor of 70.

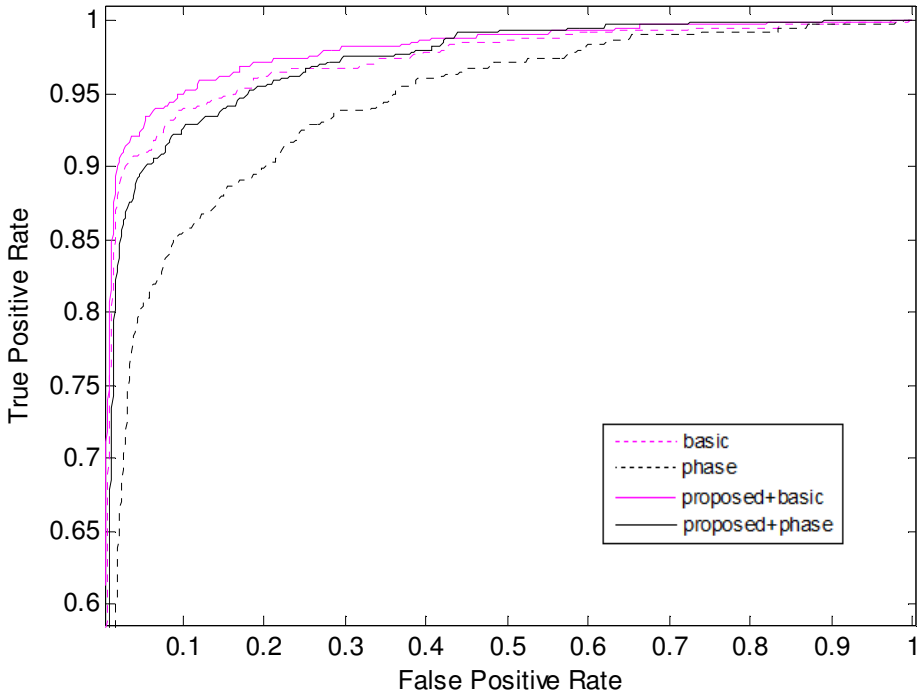


Figure 4.26 Overall ROC curves for the Basic approach and the phase approach using the proposed confidence map with a block size of 512x512 at JPEG quality factor of 50.

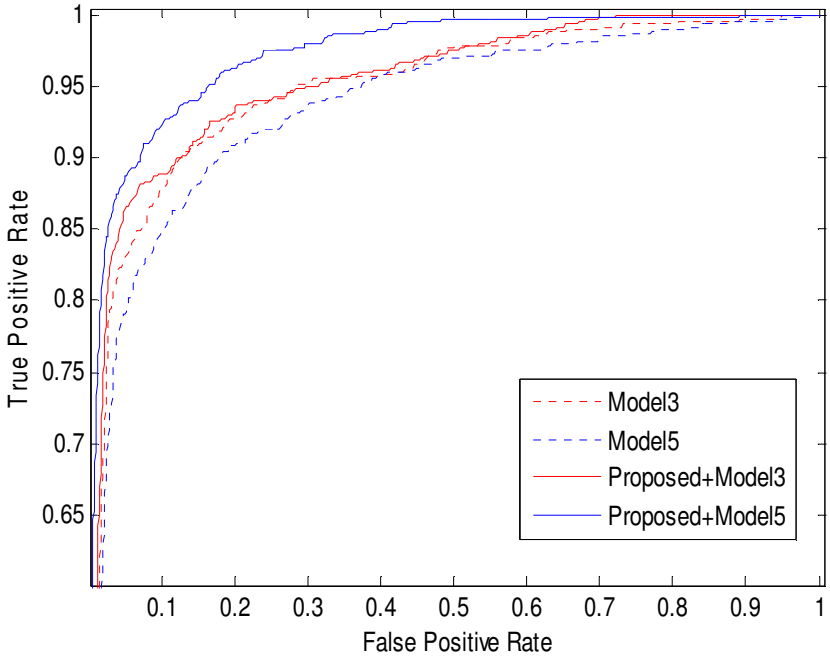


Figure 4.27 Overall ROC curves for Li's model3 and model 5 with a block size of 512x512 at JPEG quality factor of 50.

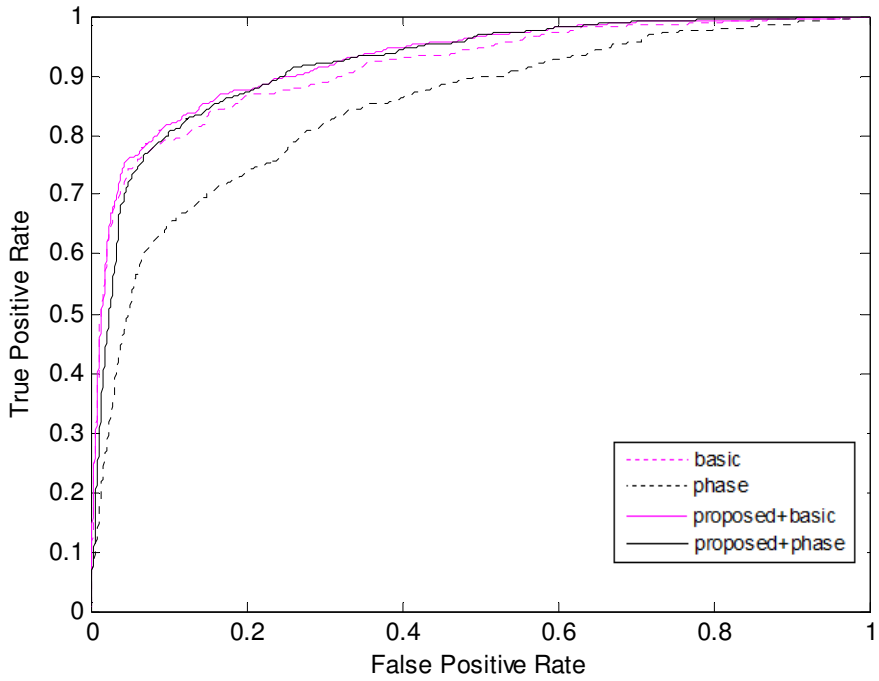


Figure 4.28 Overall ROC curves for the Basic approach and the phase approach using the proposed confidence map with a block size of 256x256 at JPEG quality factor of 50.

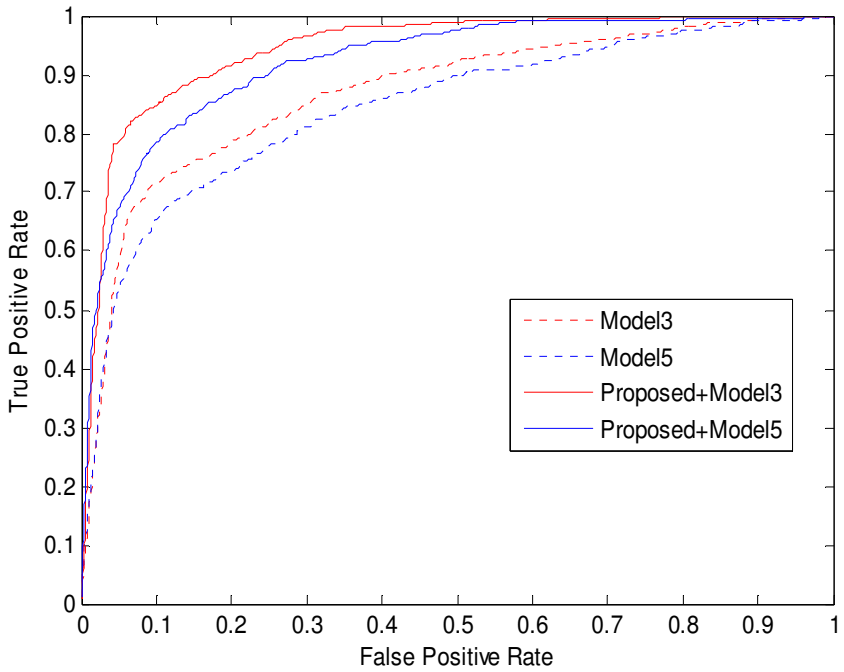


Figure 4.29 Overall ROC curves for Li's model3 and model 5 with a block size of 256x256 at JPEG quality factor of 50.

4.7 Comparative Studies

A detailed comparative study was performed at 0.01 and 0.05 false positive rates (FPRs). Table 4.4 shows the results of the three different methods at 256x256 block size with default camera setting. We can see that the true positive rate (TPR) improves for all the methods at the two FPRs. The improvement at 0.01 FPR is better than that at 0.05 FPR for the MLE, the phase approach and Li's models. However, the improvement at 0.05 FPR is better than that at 0.01 FPR for the Basic approach.

Table 4.5 to Table 4.7 show the TPR rate at FPR=0.01 and FPR=0.05 for the testing images recompressed at JPEG format with different quality factor, i.e. 90, 70 and 50. For a high quality factor such as 90, the improvement of the proposed method is modest for all the three approaches. The percentage improvement in TPR is around 2 to 8%. However, for a medium to low quality factor such as 70 and 50, the improvement of the proposed method is significant. For example, for the recompressed images at a quality factor of 70, Table 4.6 shows that there are 6.7% and 4.4% overall improvement in TPR at FPR equals to 0.01 and 0.05 respectively when combining the proposed confidence map with the Basic approach. The improvements for combining the proposed confidence map with MLE method are 13.6 % and 10.7% in TPR at FPR equals to 0.01 and 0.05 respectively. When combining the proposed confidence map with the phase method, the improvements

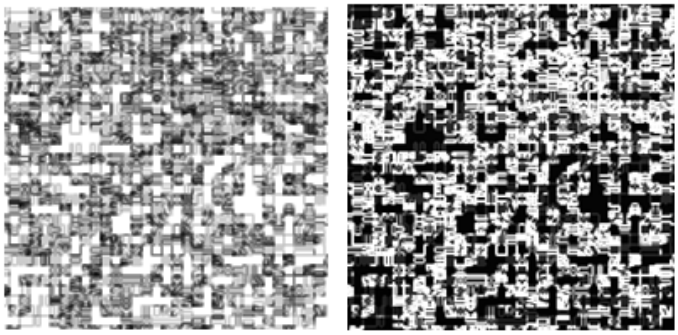
are 17.3% and 6.5% in TPR at FPR equals to 0.01 and 0.05 respectively. Similar improvement can be found with Li's models. From Table 4.5 to Table 4.7, we can see that our proposed confidence map and weighted correlation method is able to achieve a consistent improvement to the existing methods for all different JPEG quality factors.

One interesting observation in the experiment is that the Basic method is having the best result as compared with some newly proposed methods such as the MLE and the phase approaches. It may probably be due to the removal of the linear pattern which also attenuates the PRNU feature for classification. Despite that, if the linear pattern leads to misclassification from the same camera model but not the same camera, it is necessary to use the newly approaches such as MLE and phase approaches to avoid this kind of problem. Irrespective of which methods the experiments were using, experiment results show that the proposed confidence map is able to consistently provide accurate source camera identification. Also, the improvement at high compression ratio is always more significant than that at low compression ratio.

At high compression ratio, images are often suffered from blocking artifacts and the loss of high frequency content. It is caused by the quantization within the 8 by 8 blocks of an image in the JPEG scheme. The quantization causes most of the ac components within the 8 by 8 block to be truncated to zero. As PRNU is a weak

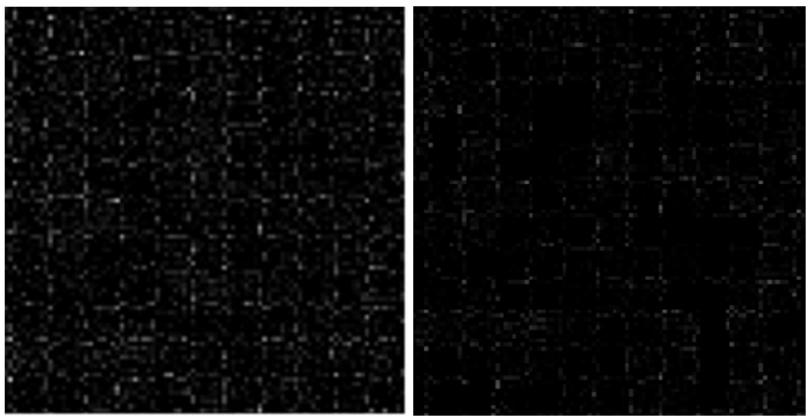
signal, the PRNU feature would be lost after quantization because of the heavy compression. Figure 4.30(a) shows a compressed image with JPEG quality factor of 70. Within the region marked with a red color box, Figure 4.30(b) shows the values of the image texture feature in equation (4.2). The white color pixels in Figure 4.30 (b) indicate the complete lost of the PRNU feature caused by the removal of the ac components in the quantization. We can see that about 30% of pixels do not carry any PRNU characteristics. Figure 4.30(c) shows the corresponding confidence map in which these pixels are marked as unreliable and hence are excluded from the weighted correlation calculation. In this way, our proposed weighting scheme helps to compensate for the lost of PRNU feature caused by heavy compression. In addition, our confidence map is able to compensate for the effect of blocking artifacts. As shown in Figure 4.31(a), blocking artifacts would appear in the estimated PRNU under high compression. After applying the confidence map as a weighting function as in Figure 4.31(b), we can see that the blocking artifacts are compensated by marking these regions as unreliable regions. As a result, the identification accuracy increases for all the PRNU-based methods using our proposed weighting scheme.

(a)



(b) (c)

Figure 4.30 (a) An image recompressed with JPEG quality factor of 70, (b) the image texture feature of the red block region in Figure 4.30(a) and (c) the corresponding confidence map



(a) (b)

Figure 4.31 (a) The estimated PRNU from a set of highly compressed images and (b) the weighted PRNU

Table 4. 4 TPR at FPR=0.01 and FPR=0.05 for the basic, MLE and phase methods with a block size of 256×256.

	FPR=0.01	FPR=0.05
Basic approach	0.925	0.962
Proposed confidence map with the basic approach	0.933(0.86%)	0.975(1.35%)
MLE	0.883	0.938
Proposed confidence map with the MLE approach	0.913(3.39%)	0.945(0.75%)
Phase approach	0.907	0.942
Proposed confidence map with the phase approach	0.932(2.76%)	0.955(1.38%)
Li's model3	0.88	0.95
Proposed confidence map with Li's model3	0.92(5.1%)	0.97(1.6%)
Li's model5	0.86	0.92
Proposed confidence map with Li's model5	0.88(1.5%)	0.93(1.4%)

Table 4. 5 TPR at FPR=0.01 and FPR=0.05 for the basic, MLE and phase methods with a block size of 256×256 at JPEG quality factor=90.

	FPR=0.01	FPR=0.05
Basic approach	0.913	0.950
Proposed confidence map with the basic approach	0.935(2.4%)	0.968(1.89%)
MLE	0.831	0.913
Proposed confidence map with the MLE approach	0.861(3.61%)	0.930(1.86%)
Phase approach	0.846	0.901
Proposed confidence map with the phase approach	0.863(2.00%)	0.945(4.88%)
Li's model3	0.86	0.92
Proposed confidence map with Li's model3	0.93(8.1%)	0.97(5.4%)
Li's model5	0.81	0.88
Proposed confidence map with Li's model5	0.85(5%)	0.91(3.4%)

Table 4. 6 TPR at FPR=0.01 and FPR=0.05 for the basic, MLE and phase methods with a block size of 256×256 at JPEG quality factor=70.

	FPR=0.01	FPR=0.05
Basic approach	0.823	0.900
Proposed confidence map with the basic approach	0.878(6.7%)	0.940(4.4%)
MLE approach	0.641	0.783
Proposed confidence map with the MLE approach	0.728(13.6%)	0.938(19.8%)
Phase approach	0.654	0.812
Proposed confidence map with the phase approach	0.767(17.3%)	0.865(6.5%)
Li's model3	0.63	0.80
Proposed confidence map with Li's model3	0.74(18%)	0.84(5.9%)
Li's model5	0.61	0.74
Proposed confidence map with Li's model5	0.66(8%)	0.81(9.7%)

Table 4. 7 TPR at FPR=0.01 and FPR=0.05 for the basic, MLE and phase methods with a block size of 256×256 at JPEG quality factor=50.

	FPR=0.01	FPR=0.05
Basic approach	0.525	0.711
Proposed confidence map with the basic approach	0.526(0.19%)	0.725(1.96%)
MLE	0.398	0.560
Proposed confidence map with the MLE approach	0.401(0.75%)	0.568(1.42%)
Phase approach	0.388	0.520
Proposed confidence map with the phase approach	0.481(24.0%)	0.690(32.7%)
Li's model3	0.45	0.57
Proposed confidence map with Li's model3	0.55(20%)	0.70(22%)
Li's model5	0.38	0.52
Proposed confidence map with Li's model5	0.47(23%)	0.65(25%)

4.8 Chapter Summary

The accuracy of PRNU-based camera identification method can easily be affected by image content. We have shown that image intensity values and image texture are two main factors that can affect the correlation calculation. Hence, a non-linear regression model was built to formulate the image content effect. In particular, the image content effect was first studied in block based manner. Then with the use of the kernel principal component analysis, a confidence map was generated which can be used to quantify the reliability of each pixel in PRNU estimation. By using the confidence map as a weighting function in correlation calculation, the image content effect can be reduced as weighting was given according to the image feature. Thus, a large weight is applied to reliable pixels while a small weight is applied to unreliable pixels which are believed to be affected seriously by scene image.

The confidence map was applied to state-of-the-art PRNU-based identification methods including the Basic approach, the MLE approach, the phase approach and Li's method [14]. Experimental results show that the proposed confidence map is able to enhance all these existing methods to achieve an accurate source camera identification result. As the camera fingerprint may be affected by the JPEG compression, experiments at a high compression level were also carried out. Based on

the experimental results, the improvement achieved by the proposed method maintains a good performance at a high compression level.

Chapter 5 Conclusions and Future Work

5.1 Conclusions

With the increasing popularity of digital imaging devices such as digital camera and mobile phone, camera identification has become an important topic in digital forensic applications. Existing methods of digital camera identification can be divided into two types: source model identification and individual source camera identification. The former uses proprietary image formation operations such as the color interpolation algorithm to determine the brand and the model of the camera while the latter finds patterns caused by manufacturing imperfections such as pattern noise to uniquely identify each individual camera. Due to the importance of identifying the exact camera in forensic applications, this thesis focused on the individual source camera identification using photo-response non-uniformity noises (PRNU).

PRNU is present in every image irrespective of image content. Existing algorithms extract PRNU through image denoising. A major problem of the PRNU-based identification method is the scene content left in the PRNU after denoising. Examples have been given in this thesis to show that the extracted PRNU feature is contaminated by saturated regions and highly textured regions in an image. This thesis thus studied the scene content effect in camera identification as well as

proposed some solutions in solving the scene content problem.

We have studied the effect of scene content on the accuracy of camera identification. The PRNU-based camera identification method is based on a binary hypothesis test: H_0 (the testing image is considered to be from other cameras) and H_1 (the testing image is from the reference camera). The separation between the probability density functions of H_0 and H_1 influences the accuracy of the identification performance. It is found that the probability density function of H_1 highly depends on the scene content in the images. If images for H_1 contain smooth content, the PRNU is less affected by the scene content which makes a large separation between the PDFs of $p(x|H_0)$ and $p(x|H_1)$. However, if images for H_1 are highly textured, the scene content will be left in the PRNU after denoising. This narrows the separation between $p(x|H_0)$ and $p(x|H_1)$. As the performance of the identification accuracy depends on the overlap area between the two PDFs, a narrow separation implies that a wrong camera classification can easily occur. Hence, the identification method has to consider the scene content of the testing images.

To demonstrate the idea clearly, we have performed a detailed study on the relations between the seriousness of scene content effect and the identification accuracy. In particular, the predicted correlation is used to characterize the scene content effect. Our study indicates clearly that a low detection accuracy is obtained

if the scene content effect is serious. Using this result, a 2D classifier is proposed for individual camera identification. The 2D classifier uses two features for camera identification. The first feature is the correlation between the PRNU of the testing image and the reference PRNU from a particular camera. It represents the similarity of two pattern noises. The second feature is the predicted correlation which characterizes the seriousness of the scene content effect. The predicted correlation helps setting different correlation thresholds for different types of image content. Experimental results show that the proposed 2D classifier can have a more flexible threshold setting mechanism which gives better identification accuracy as compared to the traditional identification methods.

Although the predicted correlation is able to characterize the scene content effect, the formulation is not flexible enough as the characterization is done in a block-based manner. Therefore, we refined the correlation prediction model so that it can describe the scene content effect in pixel level. In other words, the severity of scene content effect on each pixel can be quantified. In our proposed method, kernel principal component analysis is used to model the non-linear relationship between the image features and the correlation. Then a confidence map is generated which indicates the reliability of each pixel in the PRNU estimation. The confidence map can be used as a weighting function in which the weighting is given according to the

seriousness of the scene content effect. In other words, a large weighting is given to reliable pixels in PRNU estimation and verse versa. Experimental results show that the proposed confidence map is able to quantify the severity of scene content effect and hence it achieves accurate camera source identification even at high compression ratios.

In conclusion, we have considered two ways of compensating the scene content effect in individual source camera identification. The first method uses the predicted correlation as one of the features in a 2D classifier which helps setting the threshold according to the scene content for the camera identification. The second method extends the block-based scene content characterization to pixel level so that the reliability of each pixel in correlation estimation can be quantified. Through these two methods, the scene content effect can be properly taken into account into the camera identification problem. Experimental results and comparative studies with existing methods have shown the effectiveness of these two methods.

5.2 Future works

Besides source camera identification, image tampering detection is another important research topic in digital forensics. Image tampering detection aims to determine whether the image content has been modified. The basic idea of image

tampering detection can be very similar to source camera identification. One tries to determine whether there are any statistical inconsistencies among blocks in an image. For example, a tampered region will have different PRNU feature as compared to the same region in the reference PRNU. Thus, if the correlation is smaller than certain value, it is likely that an image tampering has occurred. In order to demonstrate the idea, a simple example is given in Figure 5.1. Figure 5.1(a) and Figure 5.1(b) show an outdoor image and its tampered version respectively. Figure 5.1(b) is tampered with a back view of a man adding into the original image. In order to detect the tampered region, PRNU characteristics of the image in Figure 5.1 (b) is extracted with a block size of 100×100. Then the correlation between the extracted PRNU and the reference pattern noise is calculated. Figure 5.2 shows the resultant correlation values in the image. A bright intensity indicates a high correlation value and vice versa. We can see that the tampered area (i.e., the back view of the man) has a lower intensity in the correlation map than other areas. This shows the statistical inconsistencies between the detected PRNU in the testing image and the reference pattern noise. However, the over-exposed area on the right hand corner also has a low intensity. It is because the PRNU feature is lost in the over-exposed area. Hence the correlation values cannot be used solely to determine the tampered region. To distinguish whether the low correlation value corresponds to a tampered region or due

to the contamination of the image content on the PRNU feature, we need a measure that can indicate the reliability of the obtained correlation. In fact, the confidence map proposed in this thesis is able to quantify the reliability of each pixel in the PRNU estimation. Hence it can be combined with the correlation values in Figure 5.2 to achieve reliable tampered region detection.

Figure 5.3 shows the confidence map of the tampered image in Figure 5.1(b). In order to have a direct comparison between the confidence map and the correlation values in Figure 5,2, the new confidence map is obtained by performing averaging with a block size of 100×100 as shown in Figure 5.4. The averaged confidence map is used to quantify the seriousness of image content effect and hence characterize the reliability of the obtained correlation. The high intensity pixels indicate a reliable correlation value. In other words, the region for the back view of the man is a reliable region while the over-exposed area is an unreliable region. Then, by comparing Figure 5.2 and Figure 5.4, it can be seen that the low correlation value for the over-exposed area in Figure 5.2 is caused by the image content as the corresponding value in the confidence map is low, while the low correlation value for the man is not caused by the image content, thus the low correlation values in this area is likely to be caused by image manipulation.

(a) (b)

Figure 5. 1(a) An outdoor image and (b) its tampered version.

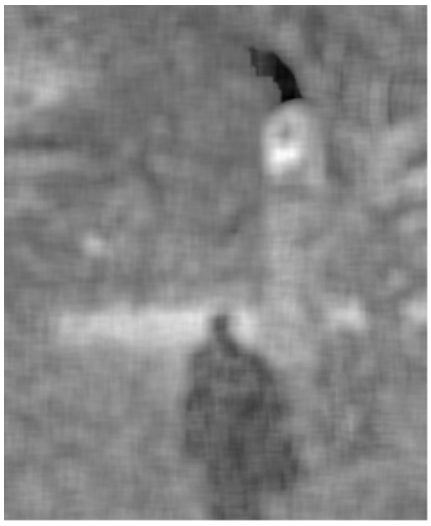


Figure 5. 2 Correlation between the PRNU extracted in image in Figure 5.1(b) and the reference PRNU with a block size of 100 x100.

Figure 5. 3 Confidence Map of Figure 5.1(b).

Figure 5. 4 Averaged Confidence Map in Figure 5.3 with a block size of 100x100.

In summary, we notice that the scene content affects not only the source camera identification, but also image manipulation detection using pattern noise. For example, the textured area or the over-exposed image region is easily classified into tampered region. In order to solve the problem, we will apply our study of scene content effect

on camera identification to image manipulation detection. We plan to use the confidence map discussed in Section 4.4 to decide whether the low value in the correlation map corresponds to a tampered image region or due to the scene content contamination on the pattern noise.

References

1. J.Lukas and J.Fridrich, “Digital camera identification from sensor pattern noise,” *IEEE Trans. Information Forensics and Security*, vol.1, no. 2, pp.205-214, 2006.
2. M. Chen, J. Fridrich, and M. Goljan, and J. Luká , “Determining image origin and integrity using sensor noise,” *IEEE Trans. Inform. Sec. Forensics*, vol. 3, no. 1, pp. 74–90, Mar. 2008.
3. M. Chen, J. Fridrich, and M. Goljan, “Digital Imaging Sensor Identification (Further Study),” *in Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contests IX,* San Jose, CA, Jan. 28-Feb.2,2007, vol.6505, pp. 0P-0Q.
4. X. Kang, Y. Li, Z. Qu, J. Huang, "Enhancing Source Camera Identification Performance with a Camera Reference Phase Sensor Pattern Noise," *IEEE Transactions on Information Forensic and Security,* vol.7, Issue 2, pp.393-402, April, 2012.
5. F. Gharibi, F. Akhlaghian, J. RavanJamjah, B. ZahirAzami, “Using the Local Information of Image to Identify the Source Camera,” Signal Processing and Information Technology, 15-18 Dec. 2010 , Luxor, pp. 515 – 519.
6. B.-B Liu, H.-K Lee, and Y Hu, “Source camera identification from significant noise residual regions,” in *IEEE International Conference on Image Processing*, Hongkong, Sep 2010.
7. Y. Li and C.-T. Li, “Decomposed Photo Response Non-Uniformity for Digital Forensic Analysis,” *Forensics in Telecommunications, Information and Multimedia*, pp. 166–172, 2009.
8. S. McCloskey, “Confidence Weighting for Sensor Fingerprinting”, Computer Vision and Pattern Recognition Workshops, Anchorage, AK, June, 2008.
9. N. Khanna, A. K. Mikkilineni, and E. J. Delp, “Forensic camera classification: Verification of sensor pattern noise approach,” *Forensic Science Communications (FSC)*, vol. 11, no. 1, January 2009.
10. M. Goljan and T. Filler, “Large Scale Test of Sensor Fingerprint Camera Identification”, *Proc. SPIE, Electronic Imaging, Media Forensics and Security XI,* San Jose, CA, January 18-22, 2009, pp. 0I 1-0I 12.
11. K. Kurosawa, K. Kuroki, and N. Saitoh, “CCD fingerprint method — identification of a video camera from videotaped images,” in *Proc ICIP*, Kobe, Japan, Oct. 1999, pp. 537–540.
12. N. Khanna, A. K. Mikkilineni, and E. J. Delp, “Forensic camera classification: Verification of sensor pattern noise approach,” *Forensic Science Communications*

(FSC), vol. 11, no. 1, January 2009.

13. C.-T. Li, "Source Camera Identification Using Enhanced Sensor Pattern Noise," *IEEE Transactions on Information Forensics and Security*, 2010.
14. Amerini I., Caldelli R., Cappellini V., Picchioni F., and Piva, A. 2010. Estimate of PRNU noise based on different noise models for source camera identification. *International Journal of Digital Crime and Forensics*. 2, 2 (Apr.-Jun. 2010), 21-33.
15. S. Bayram , H. T. Sencar , and N. Memon , " Improvements on source cameramodel identification based on CFA interpolation ," in *Proc. WG 11.9 Int. Conf. Digital Forensics*, Orlando , FL , Jan. 2006.
16. A. Swaminathan , M. Wu , and K. J. R. Liu , " Component forensics of digital cameras: A non-intrusive approach ," in *Proc. Conf. Information Sciences and Systems* , Princeton , NJ , Mar. 2006 , pp. 1194 – 1199.
17. S. Bayram , H. T. Sencar , N. Memon , and I. Avcibas , "Source camera identification based on CFA interpolation ," in *Proc. IEEE Int. Conf. Image.*
18. A. Swaminathan , M. Wu , and K. J. R. Liu , " Non-intrusive component forensics of visual sensors using output images ," *IEEE Trans. Inform. Forensics Sec.* , vol. 2 , no. 1 , pp. 91 – 106 , Mar. 2007.
19. A. C. Popescu and H. Farid , " Exposing digital forgeries in color filter array interpolated images ," *IEEE Trans. Signal Processing* , vol. 53 , no. 10 , part 2 , pp. 3948 – 3959 , Oct. 2005.
20. H. Cao and A. C. Kot, "Mobile camera identification using demosaicing features," in *Proc. ISCAS*, 2010, pp. 1683–1686, 2010.
21. S.Bayram, H.T. Sencar, N. Memon, Classification of digital camera-models based on demosaicing artifacts. Digital Investigation 5, 46–59, 2008.
22. M. Kirchner, "Efficient estimation of CFA pattern configuration in digital camera images," in *Proc. SPIE Conf. Media Forensics and Security*, San Jose, CA, 2010.
23. Ng T.-T. and Chang S.-H., "Blind Detection of Digital Photomontages using Higher Order Statistics", *ADVENT Technical Report* #201-2004-1, Columbia University, June 2004.
24. M. Kharrazi, H. T. Sencar, and N. Memon, "Blind source camera identification," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2004.
25. M. J. Tsai, C. S. Wang, J. Liu, and J. S. Yin, "Using Decision Fusion of Feature Selection in Digital Forensics for Camera Source Model Identification," *Computer Standards & Interfaces,* 2011.
26. M.J. Tsai and C.S. Wang, Adaptive Feature Selection for Digital Camera Source Identification, in: IEEE International Symposium on Circuits and Systems, Seattle, WA, pages 412-415, 2008
27. K. S. Choi, E. Y. Lam, and K. K. Y. Wong, "Source camera identification using

footprints from lens aberration*", Proceedings of the SPIE 2006*.

28. K. S. Choi, E. Y. Lam, and K. Y. Wong, "Automatic source identification using the intrinsic lens radial distortion," *Opt. Express*, vol. 14, no. 24, pp. 11551–11565, Nov. 2006.
29. Y. Ido and H.O. Hagit, Digital Image Forgery Detection Based on Lens and Sensor Aberration, in: International Journal of Computer Vision, 92:1(71-91), 2011.
30. J. Yu, A. Scott and E. Li, Toward the Identification of DSLR Lenses by Chromatic Aberration, in: SPIE Conference on Media Watermarking, Security, and Forensics, San Francisco, CA, 2011.
31. F. Devernay and O. Faugeras, "Automatic calibration and removal of distortion from scenes of structured environments," in *Investigative and Trial Image Processing*, *Proc. SPIE* **2567**, pp. 62–67, 1995.
32. J. R. Janesick, "Scientific Charge-Coupled Devices," in *Monograph,* vol. PM83. Bellingham, WA: SPIE, 2001.
33. S. M. Kay, *Fundamentals of Statistical Signal Processing*, vol. II Prentice-Hall, New York, 1998,
34. M. K. Mihcak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Phoenix, AZ, Mar. 1999, vol. 6, pp. 3253–3256.
35. 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 *Proceedings of Digital Signal Processing*, pages 1–7, Santorini, Greece, July 2009.
36. K. Rosenfeld, H.T. Sencar, "A Study of the Robustness of PRNU-Based Camera Identification." *Proc. SPIE, Media Forensics and Security XI*, vol. 7254, San Jose, CA, January 18–22, pp. 0M–0N, 2009**.**
37. A. Cortiana, V. Conotter, G. Boato, and F. G. B. De Natale, "Performance comparison of denoising filters for source camera identification," *Proc. of the SPIE*, vol. 7880, pp. 788007-788007-6, 2011.
38. G. Chierchia, S. Parrilli, G. Poggi, C. Sansone, and L. Verdoliva. "On the influence of denoising in prnu based forgery detection," in *Proc. MiFor*, 2010.
39. S. Mahalakshmi, K. Vijayalakshmi and S. Priyadharsini, "Digital Image Forgery Detection and Estimation by Exploring Basic Image Manipulations", *Digital Investigation*, 8:3-4(215-225), 2012.
40. M. Stamm and K. J. Ray Liu, "Forensic Detection of Image Manipulation Using Statistical Intrinsic Fingerprints", *IEEE Transactions on Information Forensics and*

Security, 5:3(492-506), 2012.

41. G. Gül, I. Avcibas and F. Kurugollu, “SVD Based Image Manipulation Detection”, *International Conference on Image Processing*, pp. 1765-1768 Hong Kong, 2010.
42. H. Farid, “A Survey of Image Forgery Detection”, *in IEEE Signal Processing Magazine*, 2:26(16-25), 2009.
43. P. Zhang and X. Kong, “Detecting Image Tampering Using Feature Fusion, in: International Conference on Availability”, *Reliability and Security*, pp. 335-340, 2009.
44. Y. Fang and J. Yin, “Digital Image Forensics For Photographic Copying”, *in SPIE Conference on Media Waternarking, Security, and Forensics*, San Francisco, CA, pages 8303-12, 2012.
45. B. L. Shivakumar and S. Santhosh Baboo, “Detecting Copy-Move Forgery in Digital Images: A Survey and Analysis of Current Methods”, *in Global Journal of Computer Science and Technology*, 10:7(61-65), 2011.
46. V. Christlein, C. Riess and E. Angelopoulou, “A Study on Features for the Detection of Copy-Move Forgeries”, in *Information Security Solutions Europe*, Berlin, Germany, 2010.
47. J. Zhang, Z. Feng and Y. Su, “A New Approach for Detecting Copy-Move Forgery in Digital Images”, *in IEEE Singapore International Conference on Communication Systems*, Guangzhou, China, pages 362-366, 2008.
48. J. Fridrich, D. Soukal and J. Lukas, “Detection of Copy Move Forgery in Digital Images”, *in Digital Forensic Research Workshop*, Cleveland, OH, 2003.
49. B. Mahdian and S. Saic, “Detection of Copy Move Forgery Using a Method Based on Blur Moment Invariants”, *in Forensic Science International*, 171(180-189), 2007.
50. B. Scholkopf, A. J. Smola, and K.R. Muller, “Nonlinear component analysis as a kernel eigenvalue problem”, *Neural Comput.*, vol. 10, pp.1299.1319, 1998.
51. C. Chang and C. Lin, “LIBSVM: a library for support vector machines,” 2001. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm
52. J. Armando Dominguez-Molina, G. Gonzalez-Farias, and R. M. Rodriguez-Dagnino, A practical procedure to estimate the shape parameter in the generalized Gaussian distribution, CIMAT Tech. Rep. I-01- 18_eng.pdf. [Online]. Available: http://www.cimat.mx/reportes/enlinea/I-01-18_eng.pdf.
53. J. E. Adams and J. F. Hamilton, Jr., "Adaptive color plane interpolation in single color electronic camera," U.S. Patent 5 506 619 (Apr. 1996).
54. T. Sakamoto, C. Nakanishi and T. Hase, “Software pixel interpolation for digital still cameras suitable for a 32-bit MCU,” *IEEE Trans. Consumer Electronics*, vol. 44, no. 4, November 1998.

55. T. Minka, “Inferring a Gaussian distribution”, Technical report,MIT, 2000.
56. H. Lim; S.Y. Park, S.J. Kang, W.H. Cho, “FPGA implementation of image watermarking algorithm for a digital camera”, *IEEE Pacific Rim Conference on Communications, Computers and signal Processing*, vol.2, pp. 1000- 1003, 2003
57. D. Kundur, D. Hatzinakos , “Digital watermarking for telltale tamper proofing and authentication”, *Proceedings of the IEEE* , vol.87, no.7, pp.1167-1180, July,1999