



This work is protected by copyright and other intellectual property rights and duplication or sale of all or part is not permitted, except that material may be duplicated by you for research, private study, criticism/review or educational purposes. Electronic or print copies are for your own personal, non-commercial use and shall not be passed to any other individual. No quotation may be published without proper acknowledgement. For any other use, or to quote extensively from the work, permission must be obtained from the copyright holder/s.

Image source identification and characterisation for forensic analysis

Ahmad Ryad Soobhany

Doctor of Philosophy in Computer Science

December 2013

Keele University

Abstract

Digital imaging devices, such as digital cameras or mobile phones, are prevalent in society. The images created by these devices can be used in the commission of crime. Source device identification is an emerging research area and involves the identification of artefacts that are left behind in an image by the camera pipeline. These artefacts can be used as digital signatures to identify the source device forensically. The type of digital signature considered in this thesis is the Sensor Pattern Noise (SPN), which consists mainly of the PRNU (Photo Response Non-Uniformity) of the imaging device. The PRNU is unique to each individual sensor, which can be extracted traditionally with a wavelet denoising filter and enhanced to attenuate unwanted artefacts.

This thesis proposes a novel method to extract the PRNU of a digital image by using Singular Value Decomposition (SVD) to extract the digital signature. The extraction of the PRNU is performed using the homomorphic filtering technique, where the inherently nonlinear PRNU is transformed into an additive noise. The range of the energy of the PRNU is estimated, which makes it easier to separate from other polluting components to obtain a cleaner signature, as compared to extracting all the high frequency signals from an image. The image is decomposed by using SVD, which separates the image into ranks of descending order of energies. The estimated energy range of the PRNU is used to obtain the interesting ranks that are utilised to form part of the digital signature. A case study of an existing image analyser platform was performed by investigating its identification and classification results. The SVD based extraction method was tested by extracting image signatures from camera phones. The results of the experiments show that it is possible to determine the source device of digital images.

Table of Contents

ABSTRACT.....	III
TABLE OF CONTENTS.....	IV
LIST OF FIGURES.....	X
LIST OF TABLES.....	XIV
LIST OF ACRONYMS	XV
ACKNOWLEDGEMENTS.....	XVI
CHAPTER 1 INTRODUCTION.....	17
1.1 Digital Image Forensics	17
1.1.1 Source Identification.....	18
1.1.2 Sensor Noise	19
1.2 Existing Image Analyser Platform	20
1.3 Motivation.....	21
1.4 Contributions.....	23
1.5 Organisation of Thesis	24
CHAPTER 2 DIGITAL IMAGE FORENSICS	26
2.1 Introduction	26

2.2	Human Vision System	27
2.3	Image Acquisition	28
2.3.1	JPEG Compression	33
2.4	Devices Signatures.....	35
2.4.1	Types of Signatures.....	35
2.4.2	Lens Aberration	37
2.4.3	Colour Interpolated Algorithms	39
2.4.4	Quantization Table & CRF.....	40
2.4.5	Sensor Pattern Noise	42
2.5	Device Identification.....	45
2.5.1	Device Linkage	46
CHAPTER 3 SENSOR NOISE AND SIGNATURE EXTRACTION.....		47
3.1	Introduction	47
3.2	Sensor Noise.....	48
3.2.1	PRNU.....	52
3.3	Characteristics of PRNU	54
3.3.1	Multiplicative noise	55
3.3.2	Energy of PRNU.....	57
3.4	Signature Extraction of SPN/PRNU	59
3.4.1	Gaussian Filter	62
3.4.2	Wavelet Based Filter	64
3.4.3	Enhancer.....	70
CHAPTER 4 PRACTICAL CASE STUDY.....		73

4.1	Introduction	73
4.2	Clustering Platforms	74
4.3	Existing Platform	76
4.3.1	Identifier	78
4.3.2	Classifier.....	83
4.3.2.1	Similarity Matrix.....	86
4.3.2.2	Training Phase.....	88
4.3.2.3	Classification Phase.....	89
4.4	Testing Experiments on Existing Platform.....	90
4.4.1	Testing of Image Analyser Identifier.....	92
4.4.2	Classifier Grand Tour Visualisation.....	96
4.4.3	Cross Validation	100
4.5	Limitations.....	105
4.6	Plan to Address Limitations	107
CHAPTER 5 SVD BASED SIGNATURE EXTRACTION MODEL.....		109
5.1	Introduction	109
5.2	Signal Decomposition	110
5.2.1	One-Dimensional Transform Coding	112
5.3	Singular Value Decomposition.....	118
5.3.1	2D Block Transform Coding	119
5.1.1	Interpretation of SVD components.....	121
5.4	Homomorphic filtering	122
5.5	Signature Extraction Model	123

CHAPTER 6	EXPERIMENTS FOR EXISTING PLATFORM.....	128
6.1	Introduction	128
6.2	Experimentation on Existing Platform	129
6.2.1	Image Cropping Position.....	130
6.2.2	Training Size Selection	132
6.3	Results for Existing Platform.....	134
6.3.1	Image Cropping Position.....	134
6.3.2	Training Size Selection	135
6.4	Discussion of Results	138
6.4.1	Image Cropping Position.....	138
6.4.2	Training Size Selection	139
CHAPTER 7	EXPERIMENTS FOR SVD EXTRACTION METHOD.....	141
7.1	Introduction	141
7.2	Experimentation for SVD Based Signature Extraction	141
7.2.1	Estimation of PRNU ranks.....	141
7.2.2	Source Identification using PRNU ranks	142
7.3	Results for SVD Based Signature Extraction	145
7.3.1	Estimation of PRNU Ranks	147
7.3.2	Source Identification of Camera Phones	148
7.4	Discussion.....	151
7.4.1	Estimation of PRNU Ranks	152
7.4.2	Source Identification of Camera Phones	153
7.4.2.1	Correlation Coefficient Identification	153
7.4.2.2	Q-function and p-values.....	156

CHAPTER 8 SUMMARY AND CONCLUSION	158
8.1 Summary	158
8.2 Conclusion	160
8.3 Further Work.....	162
REFERENCES	164
APPENDIX A	170
APPENDIX B	171
Nokia_C2_01_A (cam_1)	171
Nokia_C2_01_B (cam_2)	172
Nokia_E72_A (cam_3)	173
Nokia_E72_B (cam_4)	174
Nokia_N95_A (cam_5)	175
Nokia_N95_B (cam_6).....	176
Samsung_galaxy_S2_A (cam_7)	177
Samsung_galaxy_S2_B (cam_8)	178
Zte_orange_sanfrancisco_A (cam_9)	179
Zte_orange_sanfrancisco_B (cam_10).....	180
APPENDIX C	181

APPENDIX D..... 183

Nokia_C2_01_A (cam_1) 183

Nokia_C2_01_B (cam_2) 183

Nokia_E72_A (cam_3) 184

Nokia_E72_B (cam_4) 184

Nokia_N95_A (cam_5) 185

Nokia_N95_B (cam_6)..... 185

Samsung_galaxy_S2_A (cam_7) 186

Samsung_galaxy_S2_B (cam_8) 186

Zte_orange_sanfrancisco_A (cam_9) 187

Zte_orange_sanfrancisco_B (cam_10)..... 187

List of Figures

Figure 2.1. Image acquisition process inside a digital camera.....	29
Figure 3.1. Sensor Pattern Noise of imaging sensors (Lukas et al, 2006)	52
Figure 3.2. Original image showing a tree (high frequency details) and sky (low frequency details).....	61
Figure 3.3. Extracted SPN from image in figure 3.2 with high frequency details clearly visible and low frequency details absent (sky).	61
Figure 3.4. (a) Two-dimensional image; (b) 1 st level wavelet decomposition with the four sub-images I_A , I_H , I_V and I_D ; (c) 2 nd level wavelet decomposition	67
Figure 3.5. Graphical enhancer model, where the magnitude of strong components in the signature is attenuated and the weak components are not attenuated.	71
Figure 4.1. Signature extraction pipeline in existing platform	76
Figure 4.2. Camera reference signature pipeline in existing platform	79
Figure 4.3. Identifier execution pipeline in existing platform.....	81
Figure 4.4. Stages of the unsupervised classification of images.	84
Figure 4.5. Similarity matrix pipeline in existing platform	86
Figure 4.6. Plot of the average time to create similarity matrix, in minutes, based on different image crop size.....	96
Figure 4.7. Snap shot of the Guided Tour of 50 signatures and 3D plot on the right side shows the four separate clusters.	99
Figure 4.8. k partitions blocks for cross validation.....	102
Figure 5.1. SVD of matrix A with each rank as a separate matrix	119
Figure 6.1. Three cropping positions (red squares) on a picture; top-left, centre and lower-left cropping positions.	131
Figure 6.1. Variance and error classification rate with respect to number of folds when cross validation was performed on 1000 images.	136
Figure 6.2. Effect of number of folds on percentage error rate for different sample sizes (250, 500, 1000) of images.	136

Figure 7.1. Plot of log-scaled singular values of a natural image with 512 ranks after SVD decomposition.	146
Figure 7.2. Nokia_N95_A (cam_5) camera reference signature created from blue sky images and correlation with 100 images. Images 51 to 60 originate from this camera and other 90 images from the 9 other cameras. Red line is the acceptance threshold.	148
Figure 7.3. Nokia_N95_A (cam_5) camera reference signature created from natural images and correlation with 100 images. Images 51 to 60 originate from this camera and other 90 images from the 9 other cameras. Red line is the acceptance threshold.	149
Figure 7.4. zte_orange_sanfrancisco_A (cam_9) camera reference signature and correlation with 100 images. Images 81 to 90 come from this camera, images 91 to 100 from cam_10 and rest of images from the other 9 cameras. Red line is the acceptance threshold.	150
Figure 7.5. zte_orange_sanfrancisco_B (cam_10) camera reference signature and correlation with 100 images. Images 91 to 100 come from this camera, images 81 to 90 from cam_9 and rest of images from the other 9 cameras. Red line is the acceptance threshold.	151
Figure B.1. nokia_C2_01_A (cam_1) SVD camera signature and correlation with 100 images. Images 1 to 10 come from this camera, images 11 to 20 from cam_2 and rest of images from the other 9 cameras.	171
Figure B.2. nokia_C2_01_A (cam_1) wavelet camera signature and correlation with 100 images. Images 1 to 10 come from this camera, images 11 to 20 from cam_2 and rest of images from the other 9 cameras.	171
Figure B.3. nokia_C2_01_B (cam_2) SVD camera signature and correlation with 100 images. Images 11 to 20 come from this camera, images 1 to 10 from cam_1 and rest of images from the other 9 cameras.	172
Figure B.4. nokia_C2_01_B (cam_2) wavelet camera signature and correlation with 100 images. Images 11 to 20 come from this camera, images 1 to 10 from cam_1 and rest of images from the other 9 cameras.	172
Figure B.5. nokia_E72_A (cam_3) SVD camera signature and correlation with 100 images. Images 21 to 30 come from this camera, images 31 to 40 from cam_4 and rest of images from the other 9 cameras.	173
Figure B.6. nokia_E72_A (cam_3) wavelet camera signature and correlation with 100 images. Images 21 to 30 come from this camera, images 31 to 40 from cam_4 and rest of images from the other 9 cameras.	173
Figure B.7. nokia_E72_B (cam_4) SVD camera signature and correlation with 100 images. Images 31 to 40 come from this camera, images 21 to 30 from cam_3 and rest of images from the other 9 cameras.	174
Figure B.8. nokia_E72_B (cam_4) wavelet camera signature and correlation with 100 images. Images 31 to 40 come from this camera, images 21 to 30 from cam_3 and rest of images from the other 9 cameras.	174

Figure B.9. <i>nokia_N95_A (cam_5) SVD camera signature and correlation with 100 images. Images 41 to 50 come from this camera, images 51 to 60 from cam_6 and rest of images from the other 9 cameras.</i>	175
Figure B.10. <i>nokia_N95_A (cam_5) wavelet camera signature and correlation with 100 images. Images 41 to 50 come from this camera, images 51 to 60 from cam_6 and rest of images from the other 9 cameras.</i>	175
Figure B.11. <i>nokia_N95_B (cam_6) SVD camera signature and correlation with 100 images. Images 51 to 60 come from this camera, images 41 to 50 from cam_5 and rest of images from the other 9 cameras.</i>	176
Figure B.12. <i>nokia_N95_B (cam_6) wavelet camera signature and correlation with 100 images. Images 51 to 60 come from this camera, images 41 to 50 from cam_5 and rest of images from the other 9 cameras.</i>	176
Figure B.13. <i>samsung_galaxy_S2_A (cam_7) SVD camera signature and correlation with 100 images. Images 61 to 70 come from this camera, images 71 to 80 from cam_8 and rest of images from the other 9 cameras.</i>	177
Figure B.14. <i>samsung_galaxy_S2_A (cam_7) wavelet camera signature and correlation with 100 images. Images 61 to 70 come from this camera, images 71 to 80 from cam_8 and rest of images from the other 9 cameras.</i>	177
Figure B.15. <i>samsung_galaxy_S2_B (cam_8) SVD camera signature and correlation with 100 images. Images 71 to 80 come from this camera, images 61 to 70 from cam_7 and rest of images from the other 9 cameras.</i>	178
Figure B.16. <i>samsung_galaxy_S2_B (cam_8) wavelet camera signature and correlation with 100 images. Images 71 to 80 come from this camera, images 61 to 70 from cam_7 and rest of images from the other 9 cameras.</i>	178
Figure B.17. <i>zte_orange_sanfrancisco_A (cam_9) SVD camera signature and correlation with 100 images. Images 81 to 90 come from this camera, images 91 to 100 from cam_10 and rest of images from the other 9 cameras.</i>	179
Figure B.18. <i>zte_orange_sanfrancisco_A (cam_9) wavelet camera signature and correlation with 100 images. Images 81 to 90 come from this camera, images 91 to 100 from cam_10 and rest of images from the other 9 cameras.</i>	179
Figure B.19. <i>zte_orange_sanfrancisco_B (cam_10) SVD camera signature and correlation with 100 images. Images 91 to 100 come from this camera, images 81 to 90 from cam_9 and rest of images from the other 9 cameras.</i>	180

Figure B.20. <i>zte_orange_sanfrancisco_B (cam_10) wavelet camera signature and correlation with 100 images. Images 91 to 100 come from this camera, images 81 to 90 from cam_9 and rest of images from the other 9 cameras.</i>	180
Figure D.1. <i>nokia_C2_01_A (cam_1) SVD camera signature and p-values with 100 images. Images 1 to 10 come from this camera, images 11 to 20 from cam_2 and rest of images from the other 9 cameras.</i>	183
Figure D.2. <i>nokia_C2_01_B (cam_2) SVD camera signature and p-values with 100 images. Images 11 to 20 come from this camera, images 1 to 10 from cam_1 and rest of images from the other 9 cameras.</i>	183
Figure D.3. <i>nokia_E72_A (cam_3) SVD camera signature and p-values with 100 images. Images 21 to 30 come from this camera, images 31 to 40 from cam_4 and rest of images from the other 9 cameras.</i>	184
Figure D.4. <i>nokia_E72_B (cam_4) SVD camera signature and p-values with 100 images. Images 31 to 40 come from this camera, images 21 to 30 from cam_3 and rest of images from the other 9 cameras.</i>	184
Figure D.5. <i>nokia_N95_A (cam_5) SVD camera signature and p-values with 100 images. Images 41 to 50 come from this camera, images 51 to 60 from cam_6 and rest of images from the other 9 cameras.</i>	185
Figure D.6. <i>nokia_N95_B (cam_6) SVD camera signature and p-values with 100 images. Images 51 to 60 come from this camera, images 41 to 50 from cam_5 and rest of images from the other 9 cameras.</i>	185
Figure D.7. <i>samsung_galaxy_S2_A (cam_7) SVD camera signature and p-values with 100 images. Images 61 to 70 come from this camera, images 71 to 80 from cam_8 and rest of images from the other 9 camera. .</i>	186
Figure D.8. <i>samsung_galaxy_S2_B (cam_8) SVD camera signature and p-values with 100 images. Images 71 to 80 come from this camera, images 61 to 70 from cam_7 and rest of images from the other 9 camera. .</i>	186
Figure D.9. <i>zte_orange_sanfrancisco_A (cam_9) SVD camera signature and p-values with 100 images. Images 81 to 90 come from this camera, images 91 to 100 from cam_10 and rest of images from the other 9 cameras.</i>	187
Figure D.10. <i>zte_orange_sanfrancisco_B (cam_10) SVD camera signature and p-values with 100 images. Images 91 to 100 come from this camera, images 81 to 90 from cam_9 and rest of images from the other 9 cameras.</i>	187

List of Tables

Table 4.1. Time taken to calculate the camera reference signature, of size 512x512, in relation to quality and original size of images before cropping	93
Table 4.2. The correlation between using combination of high (superfine) and low (normal) quality images for creating the camera reference signature and suspect image.	94
Table 6.1. The average percentage classification error for three different image cropping positions.	135
Table 6.2. Percentage error rates for different sample sizes and varying classifier trainer size when $k = 10$. (E.g. 0.83 %)	137
Table 7.1. Names of mobile phones and their aliases used to represent them in the experimentation. The maximum image resolution of each camera and the number of images taken indoor and outdoor are also listed.	143
Table 7.2. Clustering results for 15 images from 3 cameras (Nokia E71, Nokia N95, Nikon E5200 Coolpix) based on different rank combinations used to create the signature. The cells shaded with pink are the ranks that provided the best c results.	147
Table C.1. Mean of correlation coefficient values when the ten camera reference signatures, extracted using SVD method and wavelet method, are compared with the test images from the same camera and rest of the cameras.	181
Table C.2. Standard deviation of correlation coefficient values when the ten camera reference signatures, extracted using SVD method and wavelet method, are compared with the test images from the same camera and rest of the cameras.	182

List of Acronyms

2D – two dimensional

ADC – Analogue to Digital Conversion

CCD – Charge-Coupled Device

CFA – Colour Filter Array

CMOS – Complementary Metal-Oxide-Semiconductor

CRF – Camera Response Function

DCT – Discrete Cosine Transform

DWT – Discrete Wavelet Transform

EXIF – Exchangeable Image File Format

FAR – False Acceptance Ratio

FPN – Fixed Pattern Noise

JPEG – Joint Photographic Experts Group

NUA – Non-Unique Artefacts

PRNU – Photo Response Non-Uniformity

RGB – Red Green Blue

SNR – Signal to Noise Ratio

SPN – Sensor Pattern Noise

SVD – Singular Value Decomposition

Acknowledgements

It is a pleasure to thank the many people who helped me in one way or another during my studies and made this thesis possible. I would like to thank my supervisor Dr K.P. Lam and Dr Richard Leary from Forensic Pathways Ltd for giving me the opportunity to embark on this research project. I would also like to thank Keele University (Acorn Fund) and Forensic Pathways Ltd who have provided me the funding for this research project. I am deeply grateful to Dr Peter Fletcher, my second supervisor, and Mr David Collins who have been extremely helpful.

Forensic Pathways Ltd provided the existing platform for experiments and technical support. I have to thank the staff at Forensic Pathways Ltd, especially Mr John Thornton, who has always been supportive and understanding.

I would also like to thank my friends and fellow researchers for their insightful comments, help and companionship, John Butcher, Rob Emery, Clive Jefferies, Siffat-Ullah Khan, Louis Major, Usman Nasir and James Rooney. I would like to thank my friends and anybody who provided me with cameras or who helped me in the image gathering process. I am grateful for the assistance that the administrative and technical staff of the Department of Computer Science kindly provided over the years.

My thanks to my sister Farah and her two beautiful children Salim and Aneesah who brought so much fun and happiness in my life. My most special thanks to Sofia Shah who is the light at the end of the tunnel, who has always provided support and guidance in the right direction whenever I was embroiled in doubts.

Most of all, I would like to express my sincerest gratitude to my parents without whose wonderful love, dedication to my success and their multiple sacrifices, I would never have been able to be where I am now. I remain indebted to them and words are not enough to articulate my love and affection for them. To them I dedicate this thesis.

Chapter 1 Introduction

Digital imaging devices, like digital cameras or mobile phones, are widespread everywhere in society. The images created by the digital cameras found in bespoke cameras, mobile phones, tablets or video camcorders can be used for illicit purposes and in the commission of crime. The ability to identify the source device that created a suspect image can be a valuable tool for a forensic investigator. The work presented in this thesis seeks to facilitate the source identification of digital images for forensic analysis.

1.1 Digital Image Forensics

Digital images are usually created by computer systems or digital cameras. The digital images created by computer systems are performed using image processing software (graphics) and drawing software (drawings). Digital images can also be created by the scanning of paper photographs using scanners or by taking pictures using digital cameras found in bespoke cameras, mobile phones, tablets or video camcorders. Moreover, digital pictures can be stored online using social networking sites and exchanged or downloaded through online communication software. It is estimated that an average of 2.7 million photographs are uploaded to the Facebook social network every 20 minutes and that the average person has 350 pictures stored on the site (Greengard 2012). The majority of these pictures come directly from a mobile phone or digital camera without any processing applied.

These images could be used for illicit purposes. When a forensic investigator recovers images from a suspect source, for example a hard disk, a mobile phone or a database, they might want to gather information associated with the images. The investigator may choose to identify the source device that created the image in order to link the images to a suspect or find out if the content of the images have been tampered. Digital image forensics can help the investigator obtain the information and knowledge required to solve a case.

1.1.1 Source Identification

During the image acquisition process inside a camera, the light photons go through several hardware and software processes before being converted to a digital image. These hardware and software processes leave artefacts in the image that have been created by the camera. Source device identification is an emerging research area, where the different artefacts are used as digital signatures to identify the source device.

These digital signatures can allow the identification of the make or model of the source device analogous to the use of fingerprints to identify humans. Most image processing artefacts enable the identification of the camera make or model and some artefacts that are linked to the hardware of the camera can be used to identify the specific camera that took the photograph. In a forensic case it might be useful to pinpoint the exact device that created a digital picture, in order to link the device to the location or suspect of a crime. Some examples of cases where source device identification may be valuable are:

- Child pornography cases
- Gang trophy picture cases
- Identify network of image possessors (e.g. terrorist cells, child abuse)
- Image intellectual property (IP) violation cases

One type of device signature that allows identification of specific devices is part of the sensor noise.

1.1.2 Sensor Noise

The sensor at the heart of a camera is its most expensive part. The sensor is where the light photons are converted to electrical signals. There are two types of sensors that are generally used in consumer cameras, namely the complementary metal–oxide–semiconductor (CMOS) and the charge-coupled device (CCD). These two sensors are made from silicon, which is a natural material. A pattern is formed in each image created by the sensor due to the variations in absorption of light illumination of the silicon photo elements in that sensor. This pattern, also called Sensor Pattern Noise (SPN), can be thought of as unique to each sensor. This also extends to each camera, due to the random nature of the patterns and the number of pixels that exist in each sensor.

The SPN consists mainly of the PRNU (photo-response non-uniformity) noise and other non-unique artefacts (NUA) (Lukas, Fridrich et al. 2006). It is the PRNU that creates the unique pattern that is used in the identification of cameras. The PRNU, which appears as a medium to high frequency signal, can be extracted from digital images by using a denoising high pass filter.

1.2 Existing Image Analyser Platform

There is an existing image analyser platform, which has been developed by Forensic Pathways Ltd¹, which can be used to identify the source of an image or to cluster together images that come from the same device. It consists of an identifier and an image classifier (Forensic Pathways Limited 2009).

The identifier is used when the source device that created the suspect image is present. The camera reference signature of the device can be created in order to identify whether the suspect image originates from that device. The second scenario is more complex and occurs when a large group of images are recovered and the source devices that created these images cannot be located. An unsupervised image classifier is used to group together images that come from the same source device. The platform allows forensic technology analysts to link suspect images to imaging or storage devices recovered from crime scenes.

The image analyser uses a wavelet based filter method to extract the SPN from the images and the uniqueness of this application is the use of a signature enhancer (Li 2010). The weaker SPN components are corrupted by high magnitude scene details which increase the mis-identification rate. The enhancer aims to attenuate the strong scene details while keeping the weaker pattern noise components. The enhanced SPN was

¹ Forensic Pathways Ltd is a company based in Tamworth, UK which owns the patent (Patent No. GB2467767) for the image enhancer. Forensic Pathways Ltd partly funded background work for this research project and provided technical support. Their image analyser platform was used during the experimentation processes.

shown to increase the identification rate of images from digital cameras and allows the use of smaller image crop size.

1.3 Motivation

Not all pixels are the same since the sensors of different cameras differ in size, for example the size of the sensor of a Samsung Galaxy S camera phone is 2.5 x 2.5 mm and that of a mid-range digital camera like the Pentax Optio RW18 is 6.1 x 4.6 mm (Grotta, Grotta 2012). Both cameras produce image sizes of comparable size but the quality of the images are different; the one from the camera phone is of lower quality, which renders source identification more challenging. The SPN enhancer of the existing platform is not effective for the identification of highly compressed images mainly from low to medium end camera phones. The motivations for performing this research project are:

- The existing state-of-the-art used by the existing platform does not perform well for the identification of camera phones.
- The denoising algorithms used for extracting the digital signature of images are somewhat complex and largely empirical by nature. They extract the high frequency scene details along with the high frequency PRNU component.
- The denoising does not consider the characteristics of the PRNU. The choice of the levels of decomposition of the image in the frequency domain is based mainly on empirical techniques.
- The enhancement procedures used to minimise the scene details is heuristic and affects the quality of the weak PRNU component. In the case of camera phones

the PRNU component is already attenuated by the compression process. Thus enhancing the digital signature can result in damaging the weak PRNU component.

The main aim of the research project presented in this thesis is to perform source identification of camera phones. In order to achieve this aim, a set of objectives are defined:

- The performance of the existing platform has to be assessed. The cropping process needs to be automated and the classification of images for varying sample sets studied.
- The characteristics of the PRNU need to be studied. The nature of the PRNU can be analysed and the range of its energy in relation to the image can be estimated.
- The design of a signature extraction method that can distinguish between the PRNU and scene component during the signature extraction process. By decomposing the image, its scene and noise components can be separated. The estimated PRNU can be extracted more efficiently.

1.4 Contributions

The main contributions of this research project presented in this thesis are:

- The performance of the existing platform has been assessed. The optimum cropping position of images has been identified and it has to be located in an area where there are less saturated or dark pixels in order to maximise the identification of images. In addition, cross validation technique was used to assist in finding the optimum training size that provides the best classification rate for varying sample sizes.
- The energy range of the PRNU for CMOS and CCD sensors was studied based on research performed in developing sensor noise models. The estimation of the energy range of the PRNU in natural images has been proposed, which should allow separation of the PRNU from other polluting components and obtain a cleaner signature as compared to extracting all the high frequency signals from an image.
- The design of a novel SVD based signature extraction method is proposed. The extraction of the PRNU is performed in the logarithmic domain, as per the homomorphic filtering technique, where the inherently nonlinear PRNU is transformed to an additive noise. The image is decomposed by using SVD, which separates the image into unit ranked images of descending order of energies. The estimated energy range of the PRNU is used to obtain the interesting unit ranked images that are used to form part of the digital signature. The results of the experiments performed in this research project show that it is possible to

determine the source device of digital images using the SVD based extraction method.

1.5 Organisation of Thesis

The thesis is structured as follows:

- Chapter 2 – An introduction to digital image forensics is provided. The human visual system is described and the image acquisition process in a digital camera is explained. This is followed by a review of different types of device signatures and a description of state of the art in the area of source device identification.
- Chapter 3 – A description of the different types of noise that form part of the sensor noise as well as an investigation of the characteristics of the PRNU. Two signature extraction methods are also detailed.
- Chapter 4 – Two clustering techniques that do not require presence of source device are described. The existing image analyser platform is described. The experiments that were developed on the current platform is discussed in conjunction with the limitations that were observed from the results of these experiments. Finally, further proposal to address the limitations is described.
- Chapter 5 – The concept of signal decomposition is presented and the Singular Value Decomposition (SVD), which can separate an image into ranks of descending order of energies, is investigated. The homomorphic filtering approach is described, which is followed by the presentation of a new signature extraction

model that can better locate the PRNU within an image for the purpose of subsequent extraction.

- Chapter 6 – The experiments performed on the existing platform are described, which include the cross validation experiments and cropping position for images in order to obtain stronger digital signatures. The results of the experimentation process are presented beginning with the results on identifying the most appropriate image cropping position. The results of the classification experiments performed on the existing classifier are elicited. A discussion of the results obtained is performed.
- Chapter 7 – The experimentation performed on the proposed SVD based signature extraction model is described. A description of the results obtained from the tests carried out on the SVD based signature extraction method for ranks estimation and source device identification. Finally a discussion of the results is presented.
- Chapter 8 - The thesis is summarised and the conclusions of the research project are elicited followed by some suggestions for further works.

Chapter 2 Digital Image Forensics

2.1 Introduction

Digital image forensics is a relatively new area of research that stems from the existing Multimedia Security field which involves watermarking and steganography (Redi, Taktak et al. 2011). It makes use of image processing techniques and statistical analysis tools to recover information about digital images. The information gathered can be in terms of the image creation process or any manipulations that the image has undergone after being created. There are two main sub-categories of the digital image forensics research field that are being investigated by researchers to extract the required information from images.

The first sub-area investigates whether an image has undergone any post-processing, after it was created, by studying its statistical properties. The post-processing is usually performed for malicious purposes, also called tampering, in order to hide or add erroneous information about the image source. The second area looks into the creation of the image itself in the camera pipeline, by trying to identify the source device that created the image. The artefacts left behind in the digital image by the camera can be from the characteristics of the imaging device itself or the processing inside the device (Gloe, Kirchner et al. 2007). Most of the time, forensic investigators do not have any previous knowledge about the images they recover and digital image forensics usually works in a 'blind' approach without needing à priori knowledge about the images.

A brief discussion about human vision system will be performed followed by a detailed description of the image acquisition process in digital cameras which will be followed by a review of how artefacts left in images created by digital cameras can be used as device signatures. The different types of digital signatures that have been studied will be described followed by the concept of device identification and linkage.

2.2 Human Vision System

The human vision system (HVS) is responsive to the intensity and colour of light incident on the human eye. The HVS has a non-linear perception of intensity, because of the complex rapport between intensity and perceived brightness that depends on several factors, including the level of surrounding light (Stone 2003). The human eye captures light in the rods and cones of the retina. The rods are highly light sensitive and allow vision in dim lighting conditions. There are three types of cones namely short, medium and long which are sensitive to blue, green and yellow-green light respectively and these sensitivities form the basis of the Red, Blue and Green (RGB) colour primaries in digital imaging (Irie 2009). The short cones are more scarce compared to the medium and long cones and thus the human eye is more sensitive to the green colour.

Digital cameras try to mimic the HVS. They have an automatic white-balance mode which adjusts captured images according to a white reference so that images are displayed to a uniform lighting. This white balancing is performed in the HVS to allow us to see objects with the same colour under different light intensity and colours. The image acquisition process in digital cameras is described in the following section 2.3.

2.3 Image Acquisition

The image acquisition process of a digital imaging device, such as a digital camera or phone camera, takes place in sequential stages starting from the point the light passes through the aperture of the camera optical system to the final stage of outputting the digital picture for storage. Cameras use hardware or software at each stage that corresponds to the quality of the final image to be produced. High-end digital cameras will use better quality hardware, e.g. lens or sensor, and provide more options for altering the software parameters and quality of the image. Furthermore high-end cameras can produce digital images in raw (uncompressed) format, which allows users to store the images at the highest fidelity possible without any loss in quality. Most of the mid-end to low-end digital cameras have to find a trade-off between the quality and price of the hardware or software used in the camera pipeline in order to optimise the quality of the image produced at the end. Most of these cameras store the digital image in a compressed format in order to save on storage space. Mobile/cell phones with digital cameras have the ability to perform all the standard functions of a normal phone with the added functionality of having a digital camera as part of the phone. Phone cameras have to incorporate the camera pipeline in a casing that includes all the other hardware and software of the phone, hence the quality of the components of that camera pipeline will be less than digital cameras. All camera phones store their pictures in a compressed format.

The various image processing stages or components used in a digital camera leave traces in the resultant images produced and these artefacts can be used to identify the source

device. Figure 2.1 depicts these different stages in the image acquisition process in the camera pipeline.

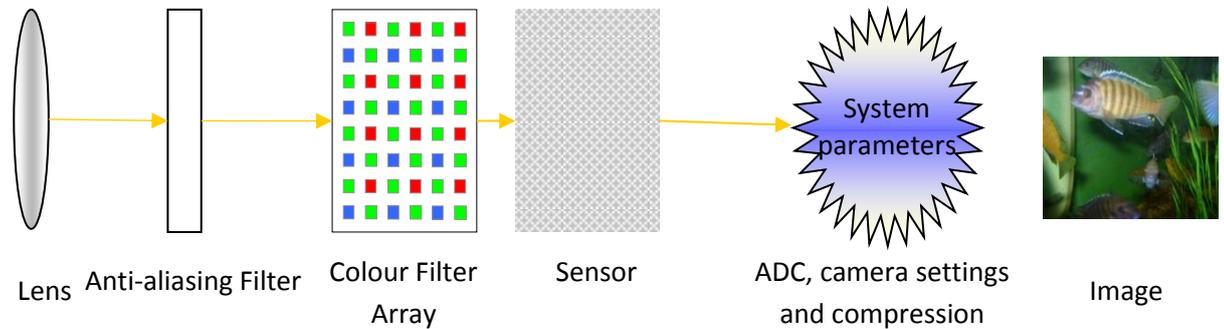


Figure 2.1. Image acquisition process inside a digital camera.

The light passes through the aperture in front of the optical lens and enters the camera through the lens or assembly of lenses, where the light is then focused on the image plane. The purpose of the lens, which is similar to the lens in the human eye, is to project the scene being photographed onto the image sensor. The size and quality of the image formed depends on the size of the aperture or focal length of the lens. Most digital cameras use lenses made of glass and low-end digital cameras or camera phones use plastic lenses, which are cheaper and easier to manufacture. A glass lens is produced by polishing a piece of glass whereas a plastic lens is pressed in a mould. The size of a plastic lens is limited since a larger plastic surface area will expand or shrink too much with varying temperatures (Bartlett 2012).

The anti-aliasing filter acts as a low-pass filter to prevent spatial frequencies higher than that of the individual pixel in the sensor from passing through, which can otherwise create aliasing (Moiré effect). The Moiré waves make a set of closely laid lines appear

curvy or interlaced and the patterns appear in an image when the spatial frequency of the scene is higher than the resolution (Nyquist frequency) of the camera. The Nyquist frequency is half the sampling frequency of a discrete signal system. The filter can be placed anywhere in the optical path between the lens and the imaging sensor. The anti-aliasing filter, placed along the imaging path, will 'blur' the incoming light by eliminating all frequencies above the Nyquist frequency of the resolution of the imaging sensor (Davies, Fennessy 2001).

The colour filter array (CFA) captures the colour components of the light stream and is used because the sensor of a camera is monochromatic (captures only the light intensity and is grayscale). There are different types of CFAs such as the RGB (Red, Green and Blue) filter, the RGBE (Red, Green, Blue and Emerald) filter and the CYGM (Cyan, Yellow, Green and Magenta) filter. The most commonly used filter in digital cameras is the RGB filter known as the Bayer filter, shown in Figure 2.1, which stores the red, green and blue (RGB) colours. Each pixel will store one colour and interpolates the other two colours from the neighbouring pixels. In the Bayer filter, there are twice as many green pixels as red and blue because the human eye is more responsive to the middle frequency in the colour range, which is green. Camera manufacturers use different methods, known as CFA interpolation and demosaicing, to calculate the remaining colours that the pixel does not store. Demosaicing is the most computationally intensive stage in the processing pipeline (Ramanath, Snyder et al. 2005).

The sensor is where the light photons are converted to electrical signals. The sensor contains light sensitive photodiodes which will absorb the energy of the photons and that energy is also the intensity of the light stream. The photodiodes are monochrome and do

not differentiate between the different wavelengths of colours in the light stream. The energy of the photons is then converted to analogue electrical charge (e.g. electrons charge). The signal created by a sensor is stored as a grayscale image. The sensor can be seen as a two-dimensional (2D) array containing a photodiode at each pixel location. The sensor is the most expensive component of the camera and is usually of two main types, the complementary metal–oxide–semiconductor (CMOS) and the charge-coupled device (CCD). There is a third type of sensor used in digital cameras, the Foveon sensor, which makes use of the property of silicon such that different wavelengths of light penetrates to different depths in the silicon (Hytti 2006). The Foveon sensors use three layers of pixels embedded in silicon, which leads to each layer recording the red, green and blue colour respectively. The resolution and colour accuracy from this type of sensor is better than sensors using normal Bayer CFA. The PRNU extraction from the Foveon sensor is beyond the scope of this study.

The main difference between these former two technologies is that in the CCD the charge is shifted out of the array, converted into voltage using one amplifier and then serially read out. That is, when the exposure is complete each pixel's charge packet is transferred sequentially to a common output structure, which converts the charge to a voltage, buffers it and sends it off-chip. On the other hand, in a CMOS sensor, an amplifier is attached to each photodiode and the voltage signals are read out one row at a time (El Gamal, Eltoukhy 2005). The charge-to-voltage conversion takes place in each pixel. Each read-out method had its own advantages and disadvantages. CCD sensors can be designed to contain very small pixels and since charge transfer is passive, they do not introduce temporal noise and fixed pattern noise (FPN). These noises are the pixel to pixel

variations due to device mismatches. On the other hand, the CMOS read-out has several active devices that introduce temporal noise and FPN. The serial read out of charge in CCD makes it slow and more expensive to produce due to the high cost of the specialised technologies. Another advantage of CMOS is that it can be manufactured on similar processing lines as normal processors as compared to specialised lines for CCD. Traditionally, CCD has been more commonly used but CMOS is being utilised more often, mainly in mobile (cell) phones cameras, because it consumes less power and cheap to produce.

After the light sensing stage, the analogue charges are converted to digital signals by using ADC (analogue to digital converters) and various camera settings are then applied. These settings can be demosaicing, colour processing, gamma correction and ISO settings. In high-end digital cameras, the raw digital image formed after the ADC conversion can be stored along with a separate metadata file containing all the information about the settings and software processing. In most low-end and medium-end digital cameras, the raw image is compressed and stored in memory in order to save storage space. Most digital cameras use JPEG (Joint Photographic Experts Group) compression to store the digital image. JPEG is a lossy compression technique that eliminates the low and high frequencies (which are mostly outside the human visual range) of an image depending on the intended quality of the output image. The higher the compression ratio is, the smaller the size, and poorer the quality, of the resulting image file will be.

2.3.1 JPEG Compression

The JPEG (Joint Photographic Experts Group) is the short name for the image file format JFIF (JPEG File Interchange Format). JPEG is a lossy compression method based on the Discrete Cosine Transform (DCT), which allows for substantial compression to be achieved and generate a reconstructed image with high visual fidelity (ISO/IEC JTC 1991, Wallace 1991). Compression is usually performed on each colour channel separately. The amount of compression is dependent on the characteristics of the image, the final image quality and the speed of compression process needed.

In the lossy baseline coding compression method, the image is converted from the RGB (Red, Green and Blue) colour space to the YCbCr (Luma, Blue and Red difference Chroma) and divided into 8 x 8 pixel blocks where each block is transformed by the forward DCT into a set of 64 values known as the *DCT coefficients*. The compression is performed in three sequential steps (Gonzalez, Woods 2002):

1. DCT computation
2. Quantization
3. Variable-length code assignment

The pixels are processed from top to bottom and left to right. The first element of the block, top left, is called the *DC coefficient* and the other elements are called the *AC coefficients*. Each element is level shifted by subtracting the quantity 2^{n-1} , where 2^n is the maximum number of grey levels. For example, if the image consists of 256 possible values (from 0 to 255) then the number to be subtracted from each pixel will be $2^{8-1} = 2^7$ (128). Each of the 64 coefficients is then quantized using one of the 64 corresponding values from a quantization table. These tables differ between camera manufacturers. A set of at least four quantization tables are used by any JPEG compression method. Following this, the coefficients are prepared for variable-length coding by recording the difference between the DC coefficient of the current block and the DC coefficient of the previous block. The AC coefficients are encoded by using a zigzag sequence, which will create a 1-D sequence of quantized coefficients. The data is compressed further by passing the quantized coefficients to an entropy coding process. An example of such a coding process is the *Huffman encoding*, which is performed by providing the *Huffman table* specifications to the encoder. Once the encoding is performed, the compressed data is still in the frequency domain and has to be converted back to the spatial domain. The reverse procedures are performed by applying an entropy decoder that decodes the quantized DCT coefficients followed by the de-quantization process. The DCT coefficients are transformed to an 8 x 8 pixel block by applying the inverse DCT. During the quantization process some AC coefficient values are rounded off and are eliminated from the data, thus leading to information loss.

2.4 Devices Signatures

Each stage in the camera processing pipeline produces some artefacts that become part of the resulting digital image. These artefacts can be used to identify the make or model of a camera. In some instances they can also help to identify the specific device that created the image. When these artefacts are detected and extracted, they can be used as digital signatures for the camera. Each section of the camera pipeline will produce a different type of signature.

2.4.1 Types of Signatures

Some of the artefacts in the camera pipeline are common to cameras of the same make or model. The 'make' of a camera is the brand of that camera and usually will be linked to the name of the company that produces this type of camera, for example, Sony, Nikon, Canon or Samsung and camera phone makers like Nokia or Apple (iPhone). Each camera manufacturer offers different 'models' of camera that target different sections of the market or to represent updates to the cameras. For example, most digital camera makers have a low, mid and high range camera model, which reflect on the quality and size of the cameras as well as the price of the camera. A low-end camera will cost less than high-end counterparts but the options on the camera will be more limited and a lower quality of hardware is used during the manufacturing process. The mid-range models will contain superior hardware, and an increased number of options, but the price will also be higher. The same applies to high-end cameras, which are mainly used by professional photographers and experienced amateurs. In these cameras, the options offered to the user are more varied and most often feature interchangeable lenses. The camera can also

store images in the raw format as opposed to the low and mid-end cameras, which most often store images in the JPEG format (as explained in section 2.2.1).

When a picture is taken, most digital cameras store information about the settings on the camera at that time. This attached information is usually called tags, EXIF (exchangeable image file format) or metadata of the image and contains details about the date and time the picture was taken, camera firmware, flash fired, image dimension, resolution or geotagging (geographical location of phone when image taken) amongst other some examples. The source of digital images can be identified using the metadata or EXIF headers that are attached to each image taken. However, this method of identification is not reliable since the metadata can be easily removed or modified by image editing software and when uploaded to a website (e.g. Facebook social networking site). A small number of camera models have a watermark inserted into each picture and, because of the limited amount of cameras having this facility, this signature cannot be used for normal cameras. In addition, it would be difficult to get all the camera manufacturers to insert a watermark facility in all their cameras, and this would not be immune to tampering nonetheless. Therefore, other types of identification methods which are more robust to editing have to be used for camera identification. There are several techniques that can be applied to the artefacts left by the stages in the camera pipeline for generating a digital signature, such as the following:

- Lens aberration
- Colour filter array (CFA) interpolation and demosaicing
- Camera response function (CRF)
- JPEG quantization tables
- Sensor pattern noise (SPN)
- Higher order wavelet statistics.

The following subsections describe the different device signatures.

2.4.2 Lens Aberration

Aberrations and distortions in lenses are unwanted artefacts of the design and manufacturing processes of lenses. Some of the distortions and aberrations are:

- Lens radial distortion
- Chromatic distortion
- Spherical aberration
- Coma
- Astigmatism
- Field curvature

More research in the field of device identification has been undergone on the first two types of distortions and therefore they will be explored in greater details. Lens radial

distortion occurs when straight lines from the object are rendered as curved lines on the sensor of the camera, and the difference between the distorted line and the straight line can be measured and used to identify the camera (San Choi, Lam et al. 2006). The lens has various focal lengths and magnifications in different areas across its spherical surface. It is the degree of the radial distortion between a distorted line and a corresponding straight line that can be used as a signature. Chromatic distortion occurs when light of different wavelengths converges at different positions on the camera sensor, which causes misalignment of the RGB channels. There are two kinds of chromatic distortion namely, longitudinal aberration and lateral aberration. The former occurs when different wavelengths focus at different distances from the lens and the second occurs when the wavelengths focus at different positions on the focal plane. The distorted parameters between the RGB colour channels for lateral aberrations can be estimated and used to identify source devices.

Lens radial distortion has been used to classify images from three cameras. However there is a major limitation to this method, the lens distortion cannot be used effectively as a signature when there is manual zooming. Given that most digital cameras nowadays provide the facility to perform manual zooming, therefore this method cannot be used to identify cameras effectively. Lateral aberration, used on its own, is not a good method of identifying cameras from the same model (Van, Emmanuel et al. 2007) and the authors have not used both lateral and longitudinal aberration together as feature sets. Although they used digital images from cell phones, none of the images had been taken with zoom activated and only two cameras came from the same make and model. These latter cameras did not provide a good identification result to differentiate between them.

2.4.3 Colour Interpolated Algorithms

The CFA does not store all the colours at the pixels locations. Instead only one colour channel is stored at any pixel. Camera manufacturers use different demosaicing or interpolation algorithms by using the pixel neighbourhood information to estimate the values of the pixel colours that were not measured. Different camera manufacturers use algorithms with kernels of different sizes and shapes, which are usually unique to each camera 'make'. Sometimes, however, the algorithms can be different between models of the same camera. The interpolation algorithms inserted within each digital image created by the camera results in an artefact that relates to the camera make or model.

These algorithms can be detected in an image and used as a digital signature to identify the model or make of the camera. The identification of the CFA interpolation and demosaicing algorithms present in digital images can be performed by calculating the correlation between the different colour channels in a colour image and estimating the demosaicing algorithm used to produce it (Bayram, Sencar et al. 2005, Gunturk, Glotzbach et al. 2005). The CFA pattern of the image is estimated by creating a set of search patterns. For each pattern, the interpolation coefficients in different types of texture regions of the image are estimated by fitting linear filtering models. These coefficients are then used to re-estimate the output image and find the interpolation error (Swaminathan, Wu et al. 2007).

The identification of the CFA interpolation algorithm is heavily affected by compression (JPEG) of images, where the spatial correlation between the pixels due to the CFA interpolation are suppressed and removed. Hence this method will not provide reliable identification of images from camera phones or heavily compressed images. Another

limitation of this method is that many of the interpolation techniques used are non-linear and dependent on scene details. After the CFA interpolation, non-linear (e.g. gamma correction) and lossy (e.g. JPEG compression) operations are performed to produce the final image and usually investigators do not have access to the raw images.

2.4.4 Quantization Table & CRF

The camera response function (CRF) maps the photon energy captured by the sensor (irradiance) to the light intensity in the image (image intensity) produced by the camera. The mapping can be estimated by fitting a function to corresponding values of intensity and radiance. Using a set of images, the CRF can be fitted to their intensity values and radiance ratios. The CRF can also be estimated by finding the mapping algorithm using a single image, and the generic imaging device can be identified as the source of that image (Lin, Jinwei Gu et al. 2004, Ng, Chang et al. 2007). The CRF estimation method has also been applied to the detection of image splicing (cut and paste). Automatic segmentation is performed on the image and the CRF is estimated followed by a boundary segment classification to detect if splicing occurred (Hsu, Chang 2007).

Celiktutan et al (Celiktutan, Avcibas 2008) have used higher order wavelets statistics together with binary similarity measures and image quality measures as digital signatures for digital images. They applied a support vector machine (SVM) classifier to the signatures to aid in the identification of the camera makes and models.

The last image processing stage in most digital camera pipelines involves compressing the raw images before storage in memory. The JPEG compression algorithm is used to compress the image, and quantization tables are used to determine the quantized effect

(rounding off) on the high and low spatial frequencies of the image in the frequency domain (JPEG compression was explained in section 2.2.1). The main source of variation between the different encoders is the quantization methods that control the compression rates and artefacts. Quantization tables vary between camera manufacturers, and different camera models from the same manufacturer, however the variation is more distinct across manufacturers (Farid 2006). Digital images are usually recompressed for storage or transmission and in these cases the generic device can be still identified (Sorell 2008).

The JPEG quantization tables can also be used to detect whether an image has been processed by an image editing software like Adobe Photoshop (Kornblum 2008). By identifying these images an investigator can eliminate from the investigation the images that have been tampered or generated by image editing software. A large scale test was performed by Farid to examine the effectiveness of using quantization tables to identify the generic source device make or model from a large sample of images downloaded from the Flickr website (Farid 2008). It was found that using the quantization tables can be reasonably effective in narrowing the source of an image to a single camera make and model or to a small set of possible cameras.

The CRF estimation has been used mainly to identify image splicing and has not been shown to work effectively in camera identification. There has not been any further works in CRF estimation either from the authors. Quantization tables can be used to identify camera makes or model, but it is easily suppressed or changed by simply re-saving the digital image in an image editing software. Furthermore, many digital cameras use

different quantization tables for different image light levels or scene details, which makes it very difficult to identify the quantization table of the cameras.

2.4.5 Sensor Pattern Noise

The sensor pattern noise (SPN) consists mainly of the PRNU (photo-response non-uniformity) noise and occurs in the sensor of the camera (Lukas, Fridrich et al. 2006). The properties of PRNU make it unique to each camera sensor. The SPN can be used to identify source devices and to determine whether an image has been tampered with (Chen, Fridrich et al. 2008). Sensor noise also consists of two types of noise; temporal and spatial (Vrhel, Saber et al. 2005). Temporal noise consist of dark noise, shot noise, noise from mechanical vibrations and noise from illumination fluctuations. Spatial noise examples are variations of light sensitivity of sensor elements and dark current non-uniformity. The temporal noise sources can be reduced by averaging and the spatial noise can be minimised by frame subtraction or gain/offset correction methods.

The PRNU is due to imperfections arising from the manufacturing process of the sensor and due to slight variations in conversion of light to electrical energy by individual pixel sensors (Fridrich 2009). A combination of the uniqueness of the imperfections in the silicon material and the different sensitivity of the pixels makes the sensor noise ideal for differentiating between sensors, even if they are made from the same silicon wafer, and hence the respective cameras into which they are embedded. The SPN, n , which appears as a high frequency signal, can be extracted from an image, I , based on the model proposed in (Lukas, Fridrich et al. 2006) as a high pass filter:

$$n = I - f(I)$$

where f is a denoising function, which acts as a low pass filter to extract the noise from the image.

There are several denoising filters that can be used and two of the methods implemented for the signature extraction are a Gaussian filter in the spatial domain and a wavelet domain based approach. The Gaussian filter is two dimensional where the variance of the filter can be adjusted to choose the cut-off frequency that will determine the level of scene content and sensor noise (Alles, Geradts et al. 2008). The second approach applies wavelet decomposition to represent the image in separate detail levels. Then a noise filter, which is a Wiener filter, is applied to the details levels and an image reconstruction is performed to obtain the noise free image. The wavelet domain filtering approach is claimed to give better results than others (Mihcak, Kozintsev et al. 1999, Lukas, Fridrich et al. 2006). The SPN is not the dominant component of the noise residuals that are extracted from the image, thus a smoother image (with less scene details) will provide better magnitude of SPN. However, as reported in (Sencar, Memon 2007), there are some limitations to using SPN as a fingerprint; it is easily contaminated by details of the scene (which are also high frequency signals with higher magnitude), by saturation due to light sources (flash, sun, light bulb) and rotation. This leads to a high misidentification and misclassification rate. A whole image has to be used for the extraction of the fingerprint in order to get a reasonable identification rate. Furthermore, the extraction process can take a long time and it is reported in Hoglund (Hoglund 2009) that it takes about 30 hours to calculate the reference noise using 200 images of sizes 3072 x 2304 pixels. Instead of using a whole image, the computational cost of the extraction process can be greatly

improved by using only part (crop) of the image. A trade-off must be found between the speed and accuracy of identification and classification rate.

The SPN can also be contaminated by the blockiness (row/column noise) created by the JPEG compression and other processing operations performed in the camera pipeline. Consequently, further processing is often applied to facilitate the estimation of the SPN, including the attenuation of non-unique artefacts (NUA) such as the FPN, blockiness and colour interpolation (Chen, Fridrich et al. 2008). The accuracy of SPN can also be improved by attenuating the interference of scene details with the enhancer described in (Li 2010), where the enhanced SPN was shown to increase the identification rate and allow the use of smaller image crop size.

The existing SPN extraction methods extract all the higher frequency details from the digital image which leads to contamination of the signature. Post processing has to be performed on the SPN signature in order to reduce the amount of contamination and the post processing decreases the strength of the signature. For highly compressed images the post processing can prevent source identification due to the already weak SPN signal.

2.5 Device Identification

The source device that created an image can be identified by using one or a combination of the different device signatures described in the previous section. Some of the device signatures allow the identification of the model or make of the source device, whilst other signatures enable the identification of the specific source device. There are two main types of device signatures that can be used for device identification:

- Image processing methods
- Hardware methods

The CFA, CRF, JPEG compression and statistical techniques, which form part of the image processing identification methods, can be used to identify a particular model or make of a camera. The hardware identification methods comprise the lens aberration and SPN techniques, which can be used to ascertain distinct devices of the same model. While some of these methods need specific assumptions to be made before processing images the SPN technique does not. Given that the lenses of higher end digital cameras are exchangeable and lenses can also be swapped with relative ease on lower end ones, the lens aberration component of an image will also change which implies that the camera identification will fail. The sensor, which is relied upon by the SPN technique, is much harder to change as well as being more expensive and hence it is uncommon for a sensor to be swapped.

Another useful capability provided by device signatures for a forensic investigator, is the ability to identify the device type that created the image. For example, being able to determine whether an image was produced by a digital camera, camera phone, scanner

or generated by computer software can help to decrease the number of images to process (McKay, Swaminathan et al. 2008). A fusion of features consisting of the colour interpolation coefficients and the statistical noise features were extracted from the images by McKay et al. Their results based on experiments performed using five cell phone cameras, four scanners, five digital cameras, and computer graphics images show that it is possible to differentiate between generic source device types.

2.5.1 Device Linkage

Extracted device signatures from digital images are very useful to ascertain the source device. When a collection of images are recovered, the signatures can be extracted and compared against device signatures already in the possession of the forensic investigator in order to find a match. Device signatures can be estimated and stored in a database. When suspect images are recovered they can be matched against the device signatures in the database and this can help in showing a network of linked users of the images.

Images that are recovered from secondary storages (e.g. hard disks, memory sticks) or downloaded from online social networks (e.g. Facebook, Flickr) can have their signatures extracted and checked against other signatures of individual images (not device reference cameras). The purpose of these checks is to find one or a set of common source devices that contributed to create these images. This is called device linkage. The signatures can be classified according to their respective source device, for example, the make or model or even specific device. The images can be checked against signatures previously recovered from suspect devices and links can be inferred.

Chapter 3 Sensor Noise and Signature Extraction

3.1 Introduction

The sensor used in a digital imaging device is usually the most expensive hardware part and is at the heart of that device, which makes the changing of sensors in camera a non-trivial process. Forensic investigators can therefore safely link a sensor, for investigative purposes, to a specific imaging device that they suspect created the image. The creation of a digital image will contain artefacts from the camera sensor, which are the sensor noises. Parts of these noises can be used to identify the specific source of the camera. The SPN consists mainly of the PRNU noise that is unique to the individual sensor of a camera and as described in section 2.4 of Chapter 2, using the SPN for device identification does not require any prior assumptions to be made about the type of camera or software used inside the camera and can differentiate between devices of the same model.

This chapter describes the different types of noises that form part of the sensor noise and how they can be used or attenuated in order to assist in device identification. The characteristics of the PRNU will be elaborated where its multiplicative nature and its energy will be explained. Finally, two signature extraction methods will be described with particular attention on the wavelet based denoising method.

3.2 Sensor Noise

The two sensors used in the large majority of digital cameras and camera phones are the complementary metal–oxide–semiconductor (CMOS) and the charge-coupled device (CCD) and they are the two types of sensors whose noise sources will be studied. There are different noise components that form part of what is called the sensor noise (Hytti 2006, Irie, McKinnon et al. 2008):

- PRNU
- Photon shot noise
- Fixed pattern noise or amplifier gain non-uniformity
- Dark current shot noise
- Read-out noise (thermal noise, reset noise)

The photo response non-uniformity occurs due to the different responses to light of individual pixels (photodiodes) on the sensor. The photon shot noise occurs due to the discrete nature of electronic charge. When current flows past any point in a circuit, the arrival rate of electrons will fluctuate slightly and give rise to variation in current flow at that point. The capture of photons is a Poisson process and increases proportionally to the square-root of the sample mean. The fixed pattern noise (FPN) arises from changes in dark currents due to variations in pixel geometry during fabrication of the sensor and it is the pixel to pixel difference when the sensor is not exposed to light. If the average value of the dark currents is subtracted from every pixel the variation remaining is referred to

as FPN, which does not vary much from image to image. Furthermore, FPN depends on the exposure time, temperature and is commonly referred to as amplifier gain non-uniformity. Dark current shot noise arises from the leakage of current in each pixel of the sensor (mainly CCD sensors) and this noise usually doubles with each 8°C rise in temperature of the sensor. Thermal noise arises from equilibrium fluctuations of an electric current inside an electrical conductor due to the random thermal motion of the charge carriers. Read-out noise is generally defined as the combination of the remaining circuitry noise sources between the photoreceptor and the ADC circuitry. Read-out includes thermal noise and reset noise among other minor noises.

The various noise sources that originate from the sensor can be classified in different ways (Hytti 2006). The first classification separates the noise components in three types:

- Signal dependent
- Temperature dependent
- Time dependent

The signal dependent noise components comprise of photon shot noise, PRNU and amplifier gain non-uniformity (aka FPN). Temperature dependent noise components include thermal noise and dark current noise and finally time dependent noise components consist of photon shot noise, dark current shot noise, reset noise and thermal noise.

The second classification method of the noise components is to separate them into random and pattern components. Random components include photon shot noise, dark current shot noise, reset noise and thermal noise. Pattern components are amplifier gain non-uniformity, PRNU, dark current non-uniformity and column amplification offset. As can be observed from the two classifications, the PRNU is a signal dependent noise reliant on the amount of light falling on the sensor and which will have the same pattern for all the images taken from the same sensor.

There are several noise models that have been developed to show the relation between the different sources of noise occurring due to the sensor and how they affect the resulting image produced. The noise models also show how the PRNU is distributed among the other source of noises and the light intensity. All the noise models were developed using RAW images, i.e. not compressed. Three models are showed below and the energy of the PRNU of each model will be described in more details in section 3.3.2.

Irie et al (Irie, 2008) performed a study to present a measurement of CCD sensor noise, where the noisy image capture model, I_{cap} is represented as:

$$I_{cap} = (I + I \times PRNU + SN_{ph}(I) + FPN + SN_{dark} + N_{read}) \times N_D \times N_{filt} + N_Q(I_{filt})$$

where I is the clean image with light intensity, SN_{ph} is the photon shot noise, SN_{dark} is the dark current shot noise, N_{read} is the read noise, N_D is the demosaicing effect, N_{filt} is the digital filtering, and $N_Q(I_{filt})$ is an additive noise source dependent on the image content after digital filtering. This model shows that the PRNU is multiplicative in relation to the light intensity.

In another study, El Gamal & ElToukhy (El Gamal, Eltoukhy 2005) presented a simplified noise model for a CMOS sensor:

$$\text{Average noise power} = q(i_{\text{ph}} + i_{\text{dc}})t_{\text{int}} + q^2(\sigma_{\text{read}}^2 + \sigma_{\text{DSNU}}^2) + (\sigma_{\text{PRNU}} i_{\text{ph}} t_{\text{int}})^2$$

where i_{ph} is the photo current, t_{int} is the time interval, q is the electron charge, σ_{read}^2 is the quantization noise with average power, σ_{DSNU}^2 is the DSNU (dark signal non-uniformity) with average power, $\sigma_{\text{PRNU}} i_{\text{ph}} t_{\text{int}}$ is the PRNU with average power. This model also showed that the PRNU is dependent on the light intensity falling on the sensor.

Chen et al (Chen, Fridrich et al. 2008) presented a sensor output model mainly for source device identification purposes as:

$$I = g^\gamma[(1 + K)Y + \Lambda]^\gamma + \Theta_q$$

where I is the noisy image, g is the colour channel gain, γ is the gamma correction factor, K is the PRNU signature, Y is the incident light intensity, Λ is a combination of the other noise sources including the dark current, shot noise, and read-out noise and Θ_q is the quantization noise. In this model, it can be seen that the PRNU is again dependent on the light intensity.

The noise models show how the different noises contribute to the overall artefacts in an image associated with the sensor. More specifically, the models show how the PRNU is related to the light intensity, a clean (noiseless) image and the other noises affecting the resultant image produced.

The pattern components and signal dependent components of the sensor noises are the parts that are used as the Sensor Pattern Noise (SPN) for source device identification. The

SPN consist of two main components the FPN and the PRNU as shown in Figure 3.1 (Lukas, Fridrich et al. 2006). The FPN can be removed by subtracting a dark frame from an image that was taken by the camera and some mid to high-end digital cameras automatically subtract a dark frame from every image. The main component of the SPN in natural images is the PRNU, which is composed of the pixel non-uniformity (PNU) is the actual difference in sensitivity of pixels to light intensity. The PNU is property that makes the PRNU unique and contributes to the pattern-like nature of the SPN. The PRNU also contains low frequency defects which are due to light refraction on dust particles and optical surfaces as well as camera zoom settings. These components are known as “doughnut” patterns and vignetting (Janesick 2001).

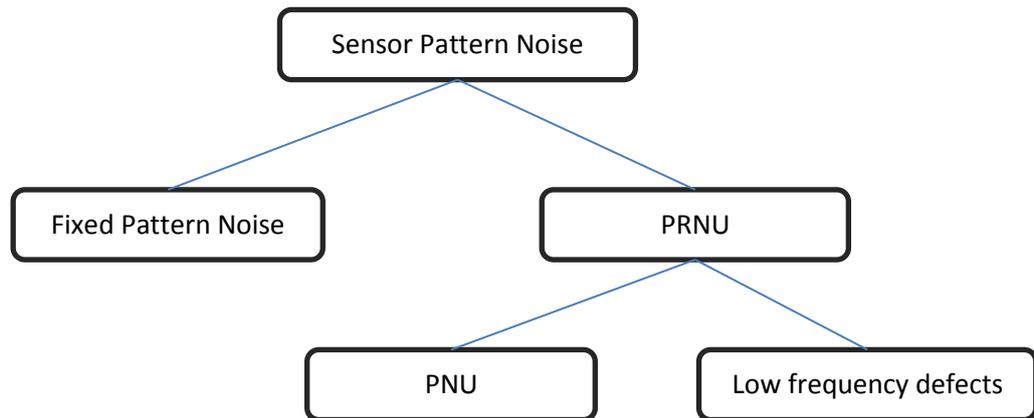


Figure 3.1. Sensor Pattern Noise of imaging sensors (Lukas et al, 2006)

3.2.1 PRNU

The photo response non-uniformity (PRNU) occurs due to the different responses to light of individual pixels (photodiodes) on the sensor. When photons reach the sensor, the photodiodes absorbs the energy of the photons and converts the energy to electrical charge. In theory all the pixels should store the same amount of electrical charge for the same amount of photon energy. In uniform illumination conditions, each photodetector

cell of an image sensor should exhibit the same output voltage. However, this is not the case due to variations in pixel geometry, substrate material, and microlenses. For example, if 1000 units of photon energy are absorbed by a pixel it should store $100e$ (elementary charge) of electrical charge, thus all the pixels present on the sensor that obtains 1000 units of photon energy should store $100e$. In practice, different pixels will store different amount of electrical charge, some might store $90e$ while others might store $75e$.

This difference in charge stored will create a pattern across the sensor and it is this pattern that forms the part of the Sensor Pattern Noise (SPN) which consists mainly of the PRNU noise. The difference in charge stored is linearly proportional to the light intensity falling on the photodiode, which means that stronger intensity of light falling on the photodiode will result in higher magnitudes of the PRNU for that resultant pixel in the image. On the other hand if the intensity of light is very low, i.e. dark image, the magnitude of the PRNU is very low. When a pixel is saturated (too much illumination), the PRNU cannot be extracted. Thus, PRNU is stronger at high illumination and cannot be extracted in dark or saturated pixels. The SPN will consist mainly of the PRNU at normal and higher illumination and the mainly the FPN at low level of light.

The PRNU is due to imperfections arising from the manufacturing process of the sensor. A combination of the uniqueness of the imperfections in the silicon material and the different sensitivity of the pixels makes the sensor noise ideal for differentiating between sensors, even if they are made from the same silicon wafer. The natural imperfections allow the pattern to be unique, due to the randomness of the imperfections, if sufficient numbers of pixels are picked for creating the digital signature. An image of size 128×128

pixels will provide a total of 16384 different values of pixels, which due to the randomness of the imperfections should provide enough pixel differences to produce a unique pattern for its respective sensor. An image of larger size (3264 x 2448 pixels) will provide a stronger signature with more unique patterns but the downside is that the computation complexity increases due to the increase in number of pixels (dimensions).

3.3 Characteristics of PRNU

The sensor at the heart of every camera determines the quality of the digital picture that can be taken by that camera (Grotta, Grotta 2012). The usual metric used by most manufacturers about the pixel count (in megapixels) merely indicates the quantity of pixels in the image but not the quality of these pixels. Consider the £100 Nikon Coolpix S3300 digital camera and the £4000 Nikon D4 DSLR (digital single lens reflex) camera, where both cameras produce images of 16-megapixel. The quality of pictures from the Nikon D4 camera is of a much higher quality than the Coolpix S3300 and this is down mainly to the quality and size of the sensor in the Nikon D4 camera, together with the more precise settings and exchangeable lenses and other higher quality hardware and software (Grotta, Grotta 2012).

A larger sensor will result in larger wells that can collect more electrons for each pixel. Thus, the point where the pixel flips between a '0' and '1' can be much more precise, which means less software processing of the image created by the sensor and a higher signal to noise ratio. Larger sensors produce better noise pattern due to more precise measurement of the PRNU difference as well as less image processing, which degrades the quality of the PRNU. Furthermore, the characteristics of the PRNU, such as its

multiplicative nature and its energy in relation to the image energy, are dependent on the quality of the sensor used. Digital cameras with better and larger sensors should produce stronger digital signatures based on SPN extraction and camera phones with very small sensors should produce weaker SPN that will reduce the ability to identify source devices. Since PRNU is caused by the physical properties of a sensor, it is nearly impossible to eliminate and is therefore usually considered a normal characteristic of the sensor array used in any CCD or CMOS camera. Therefore the characteristics of the PRNU are interesting too.

3.3.1 Multiplicative noise

A noise is said to be multiplicative if the noise amplitude depends on the state variables themselves. By contrast the amplitude of an additive noise is independent of the state variables and will change according to other outer variables. The model for a purely additive noise can be given as (Song, Uhm 1991):

$$Y = \theta X + W$$

where Y is the observed signal component, θ is a signal strength parameter, X is the signal component and W is the additive noise component.

The purely additive noise component W is usually taken to be independent and identically distributed random variables with a common PDF (probability density function), zero centred mean and variance σ^2 . The random signal and purely additive noise from the model are assumed to be statistically independent.

The model for a multiplicative noise can be represented as:

$$Y = ZX$$

Where Y is the observed signal component, X is the signal component and Z is the multiplicative noise component.

The multiplicative noise component Z is a random variable following a Gamma distribution or sometimes normal distribution with unit mean and variance $1/n$ (average of number of samples). The random signal and the multiplicative noise are assumed to be dependent.

The PRNU is proportional to illumination falling on the sensor and is prominent under high illumination levels. Since the PRNU is dependent on the light intensity, it is a multiplicative noise as shown in the sensor noise models in section 3.2. PRNU is assumed to be a zero mean noise-like signal and it is dependent on the light intensity incident at a pixel.

$$PRNU \propto \text{Light intensity}$$

The PRNU is proportional to the light intensity and can be extracted more effectively by either using a multiplicative denoising filter or converting the PRNU noise to an additive noise and applying an additive denoising filter. The multiplicative nature of the PRNU will make it non-existent in very dark or saturated sections of an image. When the PRNU is converted from the spatial domain to the frequency domain, it is found to be a medium to high frequency signal outside the human visual range (Lukas, Fridrich et al. 2006).

3.3.2 Energy of PRNU

The energy of the PRNU in an image depends on the type of device that produced the image. There are two types of sensors that are primarily used in digital cameras, the CCD (Charge-Coupled Device) and the CMOS (Complementary metal–oxide–semiconductor). The CCD is used in most digital cameras both at the low and high-end of the market, whereas the CMOS is used primarily in mobile phones and webcams. The CCD produces less noise but requires more power when compared to the CMOS, hence the reason CMOS used most often in mobile phones where space and battery life is crucial. CMOS sensors are also cheaper to manufacture. The energy of the PRNU in a CMOS will be affected by other sources of noises and the magnitude of the PRNU can be reduced. Dark current (dark noise) in CMOS sensors can be much higher than in CCD sensors and the dark current increases exponentially to rise in temperature (roughly doubling for every 8°C rise). Most of the experiments performed on the sensors in order to calculate the energy of the PRNU were in the research field of Optics or Radiometry and were related to creating a sensor model for the specific type of sensor used (CMOS or CCD).

In Irie et al (Irie, McKinnon et al. 2008), the PRNU was calculated by illuminating a Gretag Macbeth ColorChecker Color Rendition Chart (the chart consists of 24 squares of different colour paint applied to paper then mounted to a cardboard backing with a black frame around all the patches) with controlled lighting to provide a range of reflectances for measurement of the PRNU for a CCD sensor. The captured images were low-pass filtered by defocusing the lens thereby reducing the effect of any high frequency content present in the scene. The PRNU in CCD was measured by calculating the variance (σ^2) of the noise for sets of 100 images. The percentage energy values for the red, green and blue (RGB) responses of the camera were:

$$\sigma_R = 0.010R - 0.008$$

$$\sigma_G = 0.006G + 0.122$$

$$\sigma_B = 0.013B + 0.024$$

The statistical variation of each component is added in the three values. The percentage energy of the PRNU in relation to the total energy of the image is shown above.

The average power of PRNU for a CMOS sensor is given by (El Gamal, Eltoukhy 2005):

$$\frac{1}{q^2} (\sigma_{PRNU} i_{ph} t_{int})^2$$

where q is the electron charge, σ_{PRNU}^2 is the variance of PRNU, i_{ph} is the photocurrent and t_{int} is the integration time. The average power of the PRNU forms part of the sensor noise model for the average noise power in section 3.2. The standard deviation of the PRNU in the example used in el Gamal & Eltoukhy was around 0.6%. The electron charge is a constant and the integration time and the effect of photocurrent is similar for most CMOS sensors and is beyond the scope of this project. Hence the energy of the PRNU can be simplified to the variance of PRNU again. In this case, the energy of the PRNU is around 0.36%.

In Fricker et al (Fricker, Rainer et al. 1999), the lighting levels were varied in order to obtain different PRNU values for an airborne digital CCD sensor and found that up to a PRNU value of 0.02 % of the total energy the SNR (Signal to Noise Ratio) is determined exclusively by the photon noise of the signal, the RMS (root mean square) noise of the CCD and the noise of the analogue channel. At 0.1 %, the PRNU influence becomes

dominant. Hence the SNR decreases for PRNU values above 0.1%, which is to be expected.

From these three experiments it can be assumed that the energy value of the PRNU can vary between 0.02 to 0.4 % for the CMOS and CCD sensors. These experiments were performed in laboratory controlled lighting conditions, where the images were taken at maximum resolutions in a protected environment. In normal conditions for normal digital cameras or mobile phones the energy of the PRNU will represent an approximation of the energy value represented by the results of these experiments. Moreover, other external factors like illumination levels, the camera internal software settings, resolution of the picture and zooming will affect the value of the PRNU. The deterioration of the PRNU will depend on the magnitude of the noise in the image as shown in section 3.2, where the sensor noise models were described.

3.4 Signature Extraction of SPN/PRNU

Two of the most commonly used denoising filters for signature extraction are the Gaussian filter in the spatial domain and the wavelet domain based approach (Alles, Geradts et al. 2008). The estimated PRNU, n , which appears as a high frequency signal, can be extracted from an image, I , based on the model proposed in (Lukas, Fridrich et al. 2006) as a high pass filter:

$$n = I - f(I)$$

where f is a denoising function, which acts as a low pass filter to aid the extraction of the desired spectrum of noise from the image. Most signature extraction methods usually extract the SPN which contains both the FPN and PRNU.

Lines and sharp edges are examples of high frequency scene details in digital images. By subtracting the low frequency details from the image to obtain the SPN, the medium to high frequency scene details are still present in the image as shown in Figure 3.2 and Figure 3.3.



Figure 3.2. Original image showing a tree (high frequency details) and sky (low frequency details)



Figure 3.3. Extracted SPN from image in figure 3.2 with high frequency details clearly visible and low frequency details absent (sky).

The picture in Figure 3.2 shows the original natural image with a tree, some houses and a road with the sky in the background. The tree and pavements contain strong lines and edges which are high frequency details whereas the sky and road surface are smooth areas of the images which are low frequency details. When the SPN of the picture is extracted, Figure 3.3, the trees outline and the pavements are clearly visible in the signature because they are high frequency details. On the other hand, the smooth areas of the sky and road surface are plain with no details because they are low frequency details which have been filtered out. The SPN is invisible in the signature since it is in the high frequency region and is of very small magnitude. Hence scene content of the image and other stochastic noises can interfere and deteriorate the SPN. The SPN can also be contaminated by the blockiness (row/column noise) created by the JPEG compression and other processing operations performed in the camera pipeline.

3.4.1 Gaussian Filter

The Gaussian filter is two dimensional where the variance of the filter can be varied to choose the cut-off frequency that will determine the level of scene content and sensor noise. It is claimed that the Gaussian filter is faster than the wavelet method and obtains comparable identification results (Alles, Geradts et al. 2008). The Gaussian filter approach has 3 parameters that can be altered to allow the extraction of the sensor noise without too much degradation:

- The variance of the Gaussian filter
- The threshold level to suppress scene content
- The macro element size

The variance σ of the kernel of the Gaussian filter determines the cut-off frequency of the filter in order to separate the PRNU from the scene details and stochastic noises. The reference signature of a camera is estimated from 300 images using varying values of σ (Alles, Geradts et al. 2008). The same filter is used to extract the PRNU from another two sets of images with one set originating from the same camera and the second set originating from other cameras. The correlation coefficient values between the camera reference signature and the two sets of signatures are calculated and plotted and the optimal value of variance is chosen ($\sigma = 0.6$).

A similar set of experiments as the ones performed for finding the variance ($\sigma = 0.6$) are carried out to determine the threshold t to suppress the scene content (Alles, Geradts et al. 2008). The camera reference signature is calculated and the two sets of signatures extracted from the same camera and different cameras with varying levels of t . All pixel values above the threshold t are masked out. The optimal value for the threshold was found to be 4. The macro element size is performed to remove the blockiness artefacts, where a group of 2 x 2, 4 x 4 or 8 x 8 pixels are averaged into one macro element. From the experiments, the 4 x 4 averaging was found to provide the best results.

This filter is not able to distinguish between noise and signal features, this method will also distort (blur) the edge integrity (Geradts, Gloe 2009). From the results, it can be seen that the Gaussian filter is more suited for a closed set problem, where the most likely camera is chosen from a fixed group of cameras, as opposed to the open set problem where single images were used and the false rejection rates were unsatisfactory (Alles, Geradts et al. 2008).

3.4.2 Wavelet Based Filter

The Gaussian filter described in the previous section 3.4.1 operates in the spatial domain and a filter that performs the denoising in the frequency domain might produce better results since the sensor noise is understood to be in the mid to high frequency spectra region. The sensor noise in digital images can be seen as non-periodic signals with sharp discontinuities (Geradts, Gloe 2009). A Fourier based filter could be applied to extract the SPN, given that Fourier Transform is in the frequency domain and that any function expressed in the Fourier Transform can be recovered fully, with no loss of information, back in the spatial domain through an inverse process.

The Fourier Series for any function that periodically repeats itself, can be expressed as the sum of sines and/or cosines of different frequencies, each multiplied by a different coefficient. The Fourier Transform states that periodic or even stationary functions (e.g. images), with a finite area under the curve, can be expressed as the integral of sines and/or cosines multiplied by a weighing function (Gonzalez, Woods 2002). The Fourier Transform, $F(u)$, of a single variable continuous function, $f(x)$, is defined as:

$$F(u) = \int_{-\infty}^{\infty} f(x)e^{-j2\pi ux} dx$$

where $j = \sqrt{-1}$.

More specifically in the time domain the Fourier Transform can be shown as:

$$F(u) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt$$

which performs the conversion of the time signal, $f(t)$, into a frequency signal $F(u)$. The localisation of the pixel deviations is analogous to localising the time of the frequency change.

When the expression is integrated over the range of $-\infty$ to $+\infty$, the Fourier Transform is invariant to where in time a frequency change occurred. If the signals used to perform the conversion are stationary then it is not a limitation. When the image is considered as a whole it can be viewed as a stationary signal, but in order to extract the signature each discontinuity (each deviating pixel) in the image needs to be localised. The Fourier filtering proves to be a limitation in this instance and is not suitable since the frequencies are not localised.

The short time Fourier Transform (STFT) (windowed Fourier Transform) could be used to cater for the limitation of Fourier Transform by using a small time window to localise the frequency in some time interval. By minimising the window size, the localisation of the pixel variations can be estimated but the minimisation of the window size greatly reduces the precision of the frequency estimation. To solve these issues a wavelet based approach is used to denoise the image.

The basis functions of the Fourier Transform are sinusoids whereas the wavelet transforms are based on small waves, known as *wavelets*, of varying frequency and of limited duration (Gonzalez and Woods 2002). Therefore unlike the Fourier Transform which provides only frequency information, the wavelet transforms provide both the frequency and temporal information of the signal. Wavelets transformations can be the generalised *wavelet series expansion*, the *discrete wavelet transform* and the *continuous wavelet transform* depending on the type of signal being converted to the wavelet

domain. For the scope of image conversion to the wavelet domain, only the discrete wavelet transform will be reviewed.

The *discrete wavelet transform* (DWT) is the set of coefficients obtained when a function being expanded is a sequence of numbers instead of being a continuous signal and the detail coefficients, $W_\psi(j,k)$, can be defined as:

$$W_\psi(j, k) = \frac{1}{\sqrt{M}} \sum f(x)\psi_{j,k}(x)$$

where $f(x)$ is the signal and $\psi_{j,k}(x)$ is the wavelet function (also known as mother wavelet) and M is the upper limit for variable x . The mother wavelet can be modified, by scaling and translation of its parameters, which will result in the size of the time-window changing allowing localisation of where (or when) there is a change in frequency and different functions obtained (daughter wavelets). A slow varying daughter wavelet will be caused by large scale parameters and conversely a fast varying daughter wavelet will be the results of small scale parameters.

Wavelet transformations for images are performed in the two-dimensional space, where the one-dimensional wavelet transform is applied to the rows of the image followed by the one-dimensional wavelet transform of the resulting columns. In the two-dimensional wavelet transform three sets of detail coefficients are produced – the horizontal, vertical and diagonal details. The rows of the image are convolved with a scale and down-sampling its columns will produce two sub-images whose horizontal resolutions are reduced by a factor of 2. A fine scale detects high frequencies (details coefficients) and large (coarse) scale detects the low frequencies. Both sub-images are then filtered along the columns and down-sampled to produce four sub-images (quarter of the size of

original image) – I_A , I_H , I_V and I_D as shown in Figure 3.4(b). A second iteration of the filtering process produces a two-scale decomposition of I_A shown in Figure 3.4(c).

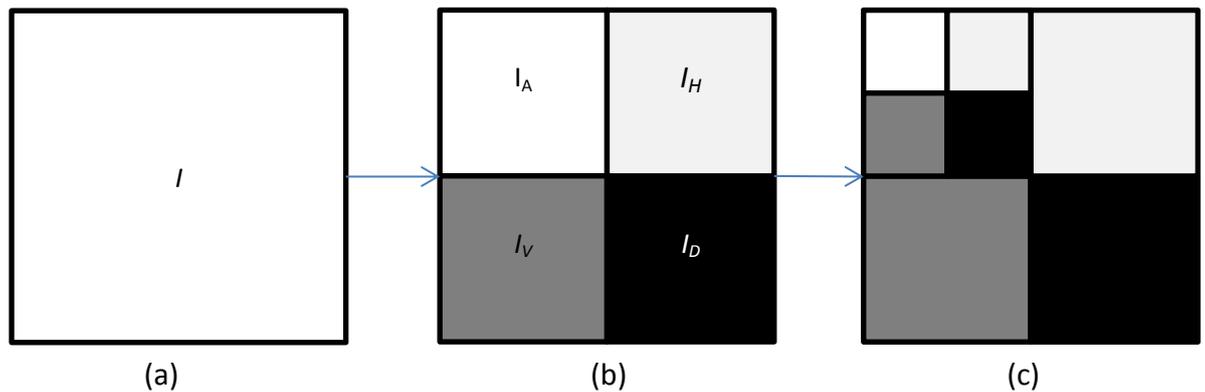


Figure 3.4. (a) Two-dimensional image; (b) 1st level wavelet decomposition with the four sub-images I_A , I_H , I_V and I_D ; (c) 2nd level wavelet decomposition

The decomposition of the signal in two dimensions will also separate the high and low frequencies of the image at each decomposition level. The DWT of an image is both localised in space and in frequency, thus accommodating for the pixel variations. There are different types of discrete wavelets and some of the most common ones are Haar, Coiflet, Daubechies. All the wavelets have at least one vanishing moment, where the wavelet coefficients are zero for polynomials of degree vanishing moments minus one ($p - 1$), where p is the number of vanishing moments. For example the Daubechies wavelets can have 2, 4, 6, 8, 10 up to 20 vanishing moments and these types of wavelets are commonly used in image processing (image denoising).

The second SPN extraction approach applies wavelet decomposition to represent the image at different levels of details and to perform the denoising process. The extraction filter was first presented by Lukas et al (Lukas, Fridrich et al. 2006), where it is described in details, and it is based on a wavelet image denoising method proposed in Mihcak et al (Mihcak, Kozintsev et al. 1999). The method provides a more effective sensor noise extraction when the size of the image is dyadic (powers of 2). A similar extraction method is used by Li (Li 2010) to extract the SPN.

The image is decomposed by performing a four level 2-D (two-dimensional) discrete wavelet transform using the Daubechies 8-tap wavelet QMF (Quadrature Mirror Filters). When a filter is applied to a signal and the reverse filtering process is applied to the results obtained, the same original signal can be recovered, this type of filter is a QMF. The horizontal, vertical and diagonal sub-bands (detail coefficients) are obtained at each level, which represents the high and low frequency components (fine and coarse details) of the image. The local variance is estimated for each coefficient, in each sub-band, with a local square neighbourhood of size 3, 5, 7 and 9 pixels. The minimum variance for each pixel is taken as the final estimate among the varying size neighbourhoods. The image is assumed to be distorted by a zero mean white Gaussian noise (WGN) with variance σ^2 . The σ^2 parameter determines the strength of the noise suppression and is dependent on the image and on the size of the noise (Geradts, Gloe 2009).

The denoised wavelet coefficients are obtained by applying a Wiener filter to the horizontal, vertical and diagonal detail coefficients at each level. The Wiener filter, also called minimum mean square error filter, considers both the degradation function and the statistical characteristics of the noise affecting the image. The aim of the Wiener filter

is to find an estimate of the uncorrupted image such that the mean square error between them is minimised (hence the name, Wiener filter). The Wiener filter is represented as:

$$C_{Hden}(i, j) = C_H(i, j) \frac{\hat{\sigma}^2(i, j)}{\hat{\sigma}^2(i, j) + \sigma_0^2}$$

where $C_{Hden}(i, j)$ is the denoised horizontal detail coefficient, $C_H(i, j)$ is the horizontal detail coefficient, $\hat{\sigma}^2$ and σ_0^2 are the estimate uncorrupted variance and variance respectively. The denoising is performed for the vertical and diagonal detail coefficients similarly. The process is repeated at each decomposition level. The inverse DWT is applied to all the detail coefficients that have been denoised in order to obtain a denoised image. The sensor noise is obtained by subtracting the denoised image from the original image. The sensor noise can also be obtained in the wavelet domain by subtracting the denoised detail coefficients from the original decomposed detail coefficient. If a colour image is used, the denoising can be performed for each colour (RGB) channels or the colour image can be converted to a grayscale image.

The wavelet domain filtering approach is claimed to give better results than others (Mihcak, Kozintsev et al. 1999, Lukas, Fridrich et al. 2006). The Wiener filter acts as a low pass filter and removes all the medium and high frequency details from the image, which when subtracted from the original image leaves some of the medium to high frequency scene details in the signature. The SPN is not the dominant component of the noise residuals that are extracted from the image, thus a smoother image (with less scene details) will provide better magnitude of SPN. However, as reported in (Sencar, Memon 2007), there are some limitations to using SPN as a fingerprint; it is easily contaminated by details of the scene (which are also high frequency signals with higher magnitude), by

saturation due to light sources (flash, sun, light bulb) and by rotation. This leads to a high misidentification and misclassification rate. A whole image has to be used for the extraction of the fingerprint in order to get a reasonable identification rate. Care must be taken in applying those additive-noise-model-based wavelet approaches directly to multiplicative noise such as PRNU (Xie, Pierce et al. 2002).

The SPN can also be contaminated by the blockiness (row and column noise) created by the JPEG compression and other processing operations performed in the camera pipeline. Consequently, further processing is often applied to facilitate the estimation of the SPN, including the attenuation of non-unique artefacts (NUA) such as the FPN, blockiness and colour interpolation (Chen, Fridrich et al. 2008). The rows and columns of the pattern noise can be zero-meant by subtracting the column mean from each pixel in the column followed by the subtraction of the row mean value from all the pixels in the rows. Further processing can be performed by converting the signature in the Fourier domain and applying Wiener filtering to remove the NUA. The accuracy of SPN can also be improved by attenuating the interference of scene details (Li 2010).

3.4.3 Enhancer

The scene details in the resulting signature tend to have higher magnitudes than the PRNU magnitudes. The enhancer developed by Li (Li 2010) aims to attenuate the strong scene details by using a hypothesis where, the stronger signal components in the signature are assumed to be associated to the scene details and are thus less trustworthy components. By assigning less significant weighting factors to strong components in the signature, a better enhanced signature can be obtained with attenuated scene

interference. The enhancer works in the Discrete Wavelet Domain (DWT), which was used to extract the signature.

Five mathematical models were proposed to perform the enhancement of the signature and they are chosen based on the amount and type of scene details in the image. If the magnitude of a component is above a threshold, α , it is considered a strong component and needs to be attenuated monotonically. For the weaker components of the signature they are weighted more than stronger components (which are attenuated). One of the models that were found to perform with most types of scene details is:

$$n_e(i, j) = \begin{cases} e^{-\frac{0.5n^2(i, j)}{\alpha^2}} & , \text{if } n(i, j) \geq 0 \\ -e^{-\frac{0.5n^2(i, j)}{\alpha^2}} & , \text{otherwise} \end{cases}$$

where $n_e(i, j)$ is the enhanced component, $n(i, j)$ is the un-enhanced component and α is the cut-off magnitude threshold. The model can be represented graphically as shown in Figure 3.5, where the magnitude of strong components in the signature is attenuated and the weak components are not attenuated.

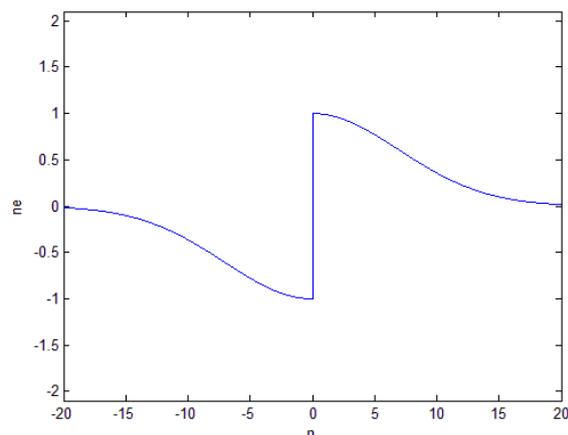


Figure 3.5. Graphical enhancer model, where the magnitude of strong components in the signature is attenuated and the weak components are not attenuated.

The extraction process can take a long time and it is reported in Hoglund (Hoglund 2009) that it takes about 30 hours to calculate the reference noise using 200 images of sizes 3072 x 2304 pixels. Instead of using a whole image, the computational cost of the extraction process can be greatly improved by using only part (crop) of the image. A trade-off must be found between speed and accuracy of identification and classification rate. The enhanced SPN was shown to increase the identification rate and allows the use of smaller image crop size (Li 2010). The development of the enhancer has led to a patent for Forensic Pathways Ltd (Forensic Pathways Limited 2009).

Chapter 4 Practical Case Study

4.1 Introduction

One of the motivations that inspired this research project was to assess the existing image analyser platform and detect any of its limitations. There has been a fair amount of research performed in recent years on the identification of source devices using their digital signatures, but the classification of images for forensic investigation purposes has not been investigated to the same depth. Identification of source devices is performed when the source device is present and the camera reference signature can be extracted. However, in the case when the source imaging device is not present, the signature of the images can be clustered according to their source device. Some of the device identification methods mentioned in chapter 2, Digital Image Forensics, performs classification of images, for example by using a support vector machine (SVM). But classification, in these methods, is mainly employed for the purpose of using or combining sets of different features, extracted from digital images known to come from the source device, to obtain the fingerprint of that source device. In the majority of cases when forensic investigators recover digital images from storage devices, they do not possess the source device that created the image. Hence a classification technique that does not require any prior knowledge of the source has to be employed to link the images coming from the same source.

Two classification techniques that do not need the presence of the source devices are described followed by a description of the existing image analyser platform. The different

components of the current platform will be investigated in details, such as the signature extraction process, the identifier and classification stages. Some experiments that were performed on the current platform will be detailed followed by the limitations that were observed from the results of these experiments. Finally a plan to address these limitations identified will be described.

4.2 Clustering Platforms

There are two clustering techniques that use the SPN as the digital signature and that do not require the source devices that created the images and they are Bloy (Bloy 2008) and Li (Li 2010). In Bloy, images are classified from a mixed set of images without any previous information about the source devices. The extraction of the signatures is simplified in order to decrease the computing complexity. A median filter is used to perform the denoising instead of using a wavelet based one and only the green (G) channel, of the RGB colour channels, of the image is used. A threshold value is calculated by correlating a selection of signatures, created using different amount of images, from a given camera with single images from a different camera. The stages of the classifier are summarised below:

1. Iterate randomly through pairs of images in the dataset until the correlation between two images (one pair) is greater than the threshold and average this pair to form the signature.

2. Correlate the rest of the images with the fingerprint, each time assign the signature to the fingerprint until 50 images have been averaged (clustered) or all images have been used.
3. Correlate the remaining unclustered images with the signatures obtained in step 2 and checking against the threshold (a maximum of 50 images are used to form the fingerprint but more than 50 images can be associated with the cluster).
4. Repeat step 1 until the stopping condition is reached, i.e. tried enough pairs or all of them without success.

This classifier needs a level of knowledge about the different cameras, or using some other camera sources, at the initial stage of the classification process to be able to calculate the threshold. Furthermore this technique needs fifty (50) signatures in order to form the fingerprint cluster, which suggests the need to know prior information about the source devices. The aim of the unsupervised image classifier in Li (Li 2010) is to cluster a large set of images taken from an unknown number of cameras into groups of images corresponding to their source cameras. This classifier does not assume any prior from the images or signatures. The latter will be described in further details in the following section (section 4.3).

4.3 Existing Platform

The existing platform is an image analyser, which is used to extract the signature of the images to identify the source device that created the image or to group together images that comes from the same source device. The platform can be used by forensic technology analysts as a forensic analysis tool to link suspect images to imaging or storage devices recovered from crime scenes. The platform uses SPN as the device signature, which is extracted by using the wavelet extraction technique. The signature extraction pipeline is shown in Figure 4.1.

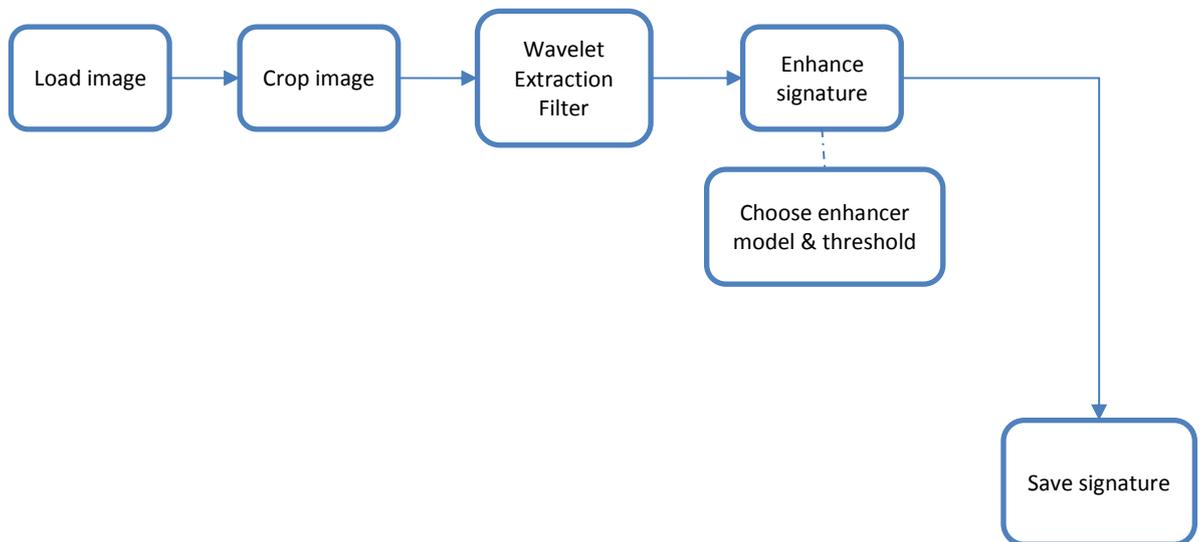


Figure 4.1. Signature extraction pipeline in existing platform

The signature extraction is performed in stages, starting when the image (in JPEG format) is loaded. The image can be cropped to a smaller size in order to increase the speed of signature extraction. The usual size of the cropped image is 512 x 512 pixels, which has been shown to provide reliable identification results (Li 2010). An image cropped to 256 x 256 pixels can also be used, albeit with a lower identification rate. The bigger the size of the image, the more of the sensor pattern noise will be present in the signature. This will

make the signature stronger and increase the identification rate but the downside of using images with increased sizes is the time taken to extract the signature and more scene details present in the resulting signature. Once the image is cropped, it is passed to the wavelet extraction filter.

The wavelet extraction filter makes use of the Daubechies 8-tap wavelets and the decomposition levels goes down to 4 levels, which is chosen because level 4 is the maximum number of wavelet decomposition level for an image of size 256 x 256 pixels. The determinant coefficients at each level are denoised by using a Wiener filter, which acts as a low pass filter. The Wiener filter works by trying to minimise the mean square error between an uncorrupted image and its estimated image. All the coefficients obtained from the 4 levels of decomposition are summed and subtracted from the main wavelet coefficient. The subtraction result is the sensor pattern noise that represents the digital signature of the image. The denoising filter is described in further details in Chapter 3, section 3.4.2.

The resultant signature extracted contains a lot of scene details, which corrupts the signature and reduces the identification rate of the signatures. The main innovation of the existing platform is the enhancing of the signature extracted in order to reduce the amount of scene details present in the signature extracted, which results in a cleaner SPN. There are five different models of the enhancer that are used by the image analyser. The threshold is the magnitude value above which the enhancing will be performed. When the magnitude of the contaminated fingerprint is less than the threshold, the "enhanced" fingerprint should grow monotonically in accordance with contaminated fingerprint. Otherwise the "enhanced" fingerprint should decrease monotonically in

accordance with contaminated fingerprint. The enhancer model and threshold can be chosen according to the type of scene details in the image, but for most types of scene details only one model and a static threshold is used. After the enhancing step the signature is saved.

The image analyser consists of:

1. Identifier
2. Classifier

There are two distinct circumstances when either the identifier or classifier will be used. When a number of images are recovered and the source device that created these images is present, the camera reference signature of the source device can be created and used to identify the images that came from the device. The second scenario occurs when a set of images are recovered, but no source device is present. This second case is more complex since the camera reference signature cannot be created and matched against the images. An unsupervised image classifier is used to group together images that come from the same source device.

4.3.1 Identifier

The identifier is used when the source device that created the images is present. For example, a set of images from a laptop or a storage device is recovered along with a digital camera or mobile phone. The camera reference signature of the recovered camera is created and stored to be used in order to find a match from suspect images. Figure 4.2 shows the steps required to create a camera reference signature.

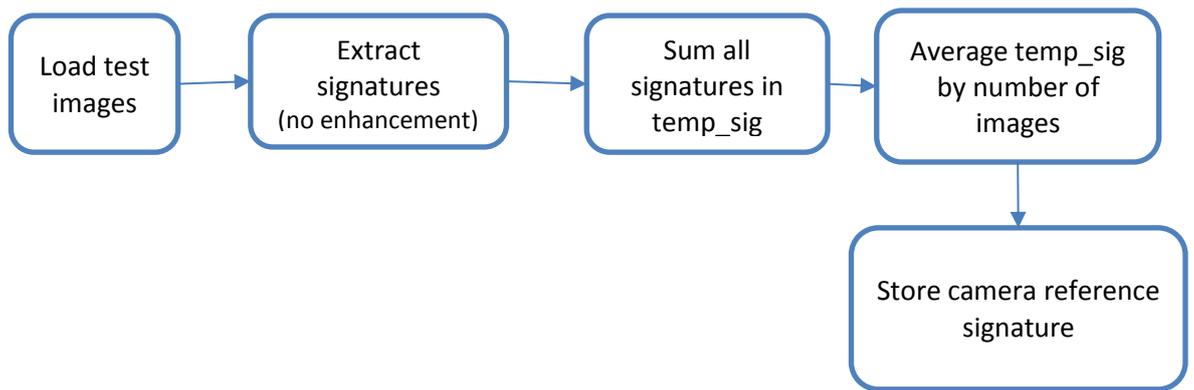


Figure 4.2. Camera reference signature pipeline in existing platform

The camera reference signature of a camera can be created if the imaging device is present. A selection of test pictures are taken from the camera of scenes with uniformly lit backgrounds, also called flat field pictures because they do not contain any scene details or contain minimal scene details. The test images should contain minimum scene details so that the reference signature will contain a cleaner representation of the camera digital fingerprint. Good examples of test images are blue sky pictures or images of a blank white (or light coloured) background (e.g. a wall). The flash on the camera should not be fired (flash off) and if taking images of blank background the pictures should be preferably out of focus so that the features of the wall are not standing out in the pictures. The test images can be images of the same spot on the wall or sky and if the pictures come from different areas, they should have minimal scene details. By using more images to create the camera reference signature, the random noises and low frequency noises in the images are eliminated. Thus the SPN, which is a medium to high frequency noise, will be more prominent in the camera fingerprint.

The number of test images used in the image analyser has been calculated previously in Hoglund (Hoglund 2009). The purpose of Hoglund's experiment was to determine the

number of images that are needed to create the camera reference signature. Hoglund increased the number of test images used to create the reference signature up to 200 per reference signature. A set of images coming from the same camera was matched against the reference signatures and the correlation value recorded. Hoglund found that using at least 50 test images provided a correlation value of nearly 1 (about 0.99) and after 75 images the reference signature did not change much. Using about 25 images provided a good enough reference signature to give a correlation of about 0.98. Thus 50 test images are used to create the camera reference signature.

The test images are loaded and cropped according to the required size and position, the signature of each test image is extracted. There is no need to enhance the signatures of the test images, since there are not prevalent scene details to pollute the signatures. Furthermore, enhancing the signatures of blank images can lead to a weaker signature overall since the magnitude across the signature is constant. The signatures of all the test images are summed and averaged by the number of test images used. The averaging of the signatures provides a stronger final SPN, which is the camera reference signature.

Once the camera reference signature is created, it can be used to match against the signatures of suspect images in the identifier. The identifier execution pipeline is shown in Figure 4.3. The camera reference signature and the suspect images are loaded in the identifier. The suspect images are cropped to correspond to the size and crop position of the camera reference signature, since both the suspect signature and reference signature need to be the same size in order to be able to find the correlation coefficient. The signatures of the suspect images are extracted and enhanced to improve the quality of the signatures. Due to the nature of the SPN being a pattern in relation to the sensor of

the camera, it is located spatially in every image. Hence if the image is rotated geometrically the SPN will change (de-synchronisation of the SPN occurs)) and a match against the reference signature will not be positive. To circumvent this problem, the signature of the suspect images are rotated and correlated against the reference signature and the highest correlation value is used. Some digital cameras rotate a picture that was taken automatically to match the orientation of all pictures taken by that camera. Most users of image editing software and all cameras editing software will only rotate images in 90 degrees angle steps. For example, an image will be rotated by 90 degrees from portrait to landscape or by 180 degrees if the camera was upside down when the image was taken.

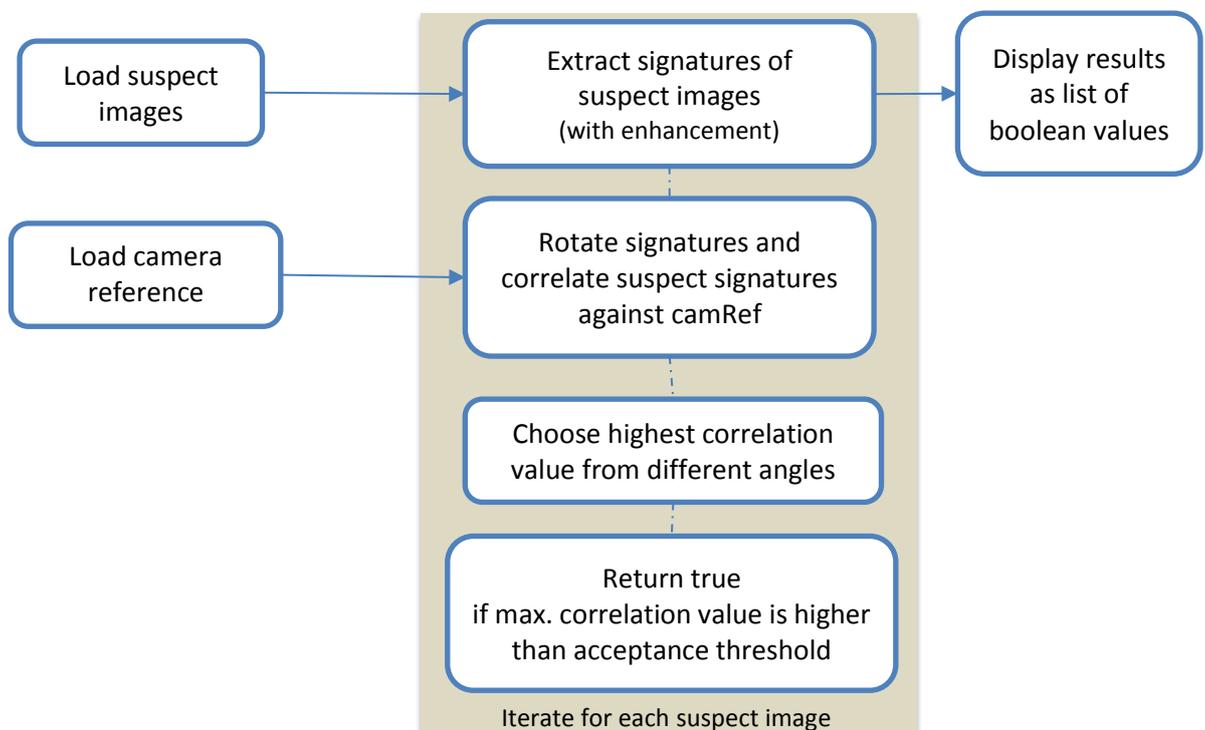


Figure 4.3. Identifier execution pipeline in existing platform

After the suspect image has been correlated against the camera reference signature for all four angles of 0, 90, 180 and 270 degrees, the highest correlation value is chosen and if that value is above the acceptance threshold, a positive match has been found. The identifier will iterate through all the suspect images and display the results of the matching as a list of Boolean (true or false) values indicating a positive or negative match against the camera reference value.

There is a need to determine the acceptance threshold value above which the correlation value will be considered a match between the reference signature and the suspect signature. Since each image provides a correlation value, a large number of images from the same camera can be used to estimate the distribution of the correlation value from that camera. Similarly, a distribution can be obtained by correlating images coming from other cameras against the reference signature of that camera. By deciding to which distribution the correlation value belongs, the suspect image can be linked to the camera. Previous research uses the two distributions obtained by using the reference camera and other cameras to find a point between them (Chen, Fridrich et al. 2007, Fridrich 2009). They model the distribution for each camera as a Generalised Gaussian distribution describing a probability density model. The Generalised Gaussian is a parametric model. The threshold is set to where the False Acceptance Ratio (FAR) is 0.01 (Li 2010). The identifier acceptance threshold for the image analyser is set to a correlation value of 0.01.

4.3.2 Classifier

The classifier of the image analyser is used when a set of images are recovered but no imaging device is present. The camera reference signature cannot be created, which makes the problem of identifying the images non-trivial. These images can be grouped together according to the source device that created them. The classification of images on the existing platform is performed by using an unsupervised clustering of the signatures, since the number of classes (groups) is unknown at the start of the classification process. The classifier finds the hidden pattern amongst the unlabelled image signatures provided. The classifier consists of a training phase that makes use of a similarity matrix to find the number of clusters on a select pool of signatures and a classification phase that groups the rest of the signatures into the formed clusters.

The different stages of the classifier are shown in Figure 4.4 below. These stages involve the classification process starting with the images and ending with all the signatures being placed in its respective cluster.

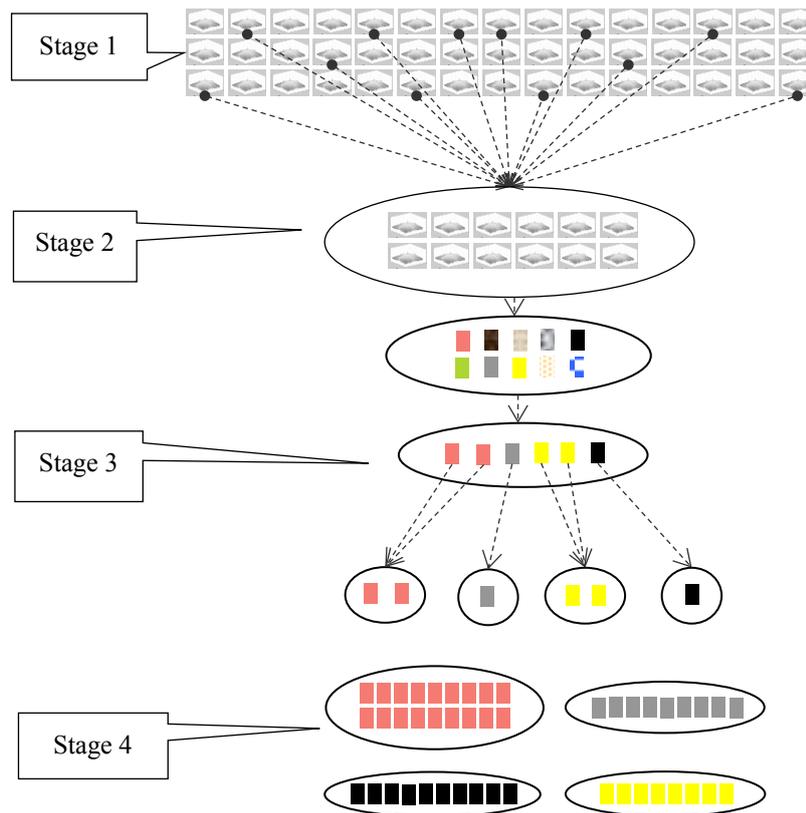


Figure 4.4. Stages of the unsupervised classification of images.

In Stage 1, all the images are cropped to the same size. Usually the size of the cropped images is 512 x 512 pixels but in some instances when some of the images are smaller than 512 x 512 pixels, they all have to be cropped according to the smaller size images. Furthermore, all the signatures have to be of the same size in order to be able to perform the cross correlation. The SPN of each image is extracted and enhanced by choosing the same model and threshold of the enhancer.

For Stage 2, the training set (M) is randomly selected from the dataset and the similarity matrix (S), of size $M \times M$, is created by calculating the correlation between all the SPNs. In Stage 3, the classifier trainer, uses conditional Markov random fields (MRF), in the training phase of the classifier. To determine the class label for each SPN i , the classifier establishes a membership committee C_i with c SPN members from the training set that

are most similar to i . In so doing, the local characteristic of Markov random fields (also known as Markovianity) can be ensured. In terms of Markov random fields, when a SPN i is being visited, the probability (or suitability), $p(\bullet)$, of assigning each class label f_j currently attached to members of C_i and its current class label f_i (if not attached to any members of C_i), are calculated. It is executed based on S and for each fingerprint a membership pool is created and the class labels are updated iteratively until a predefined stopping condition (no change in membership throughout an entire iteration) is reached. A typical strategy for a stopping condition may limit the amount of class label changes. At this stage clusters of class labels are formed and in the following Stage 4, the rest of the images are compared to the centroids of the clusters and are subsequently assigned to the respective cluster of which the centroid is closest to the images. A description of the classifier is provided in more details below, preceded by a description of the similarity matrix creation shown in Figure 4.5.

4.3.2.1 Similarity Matrix

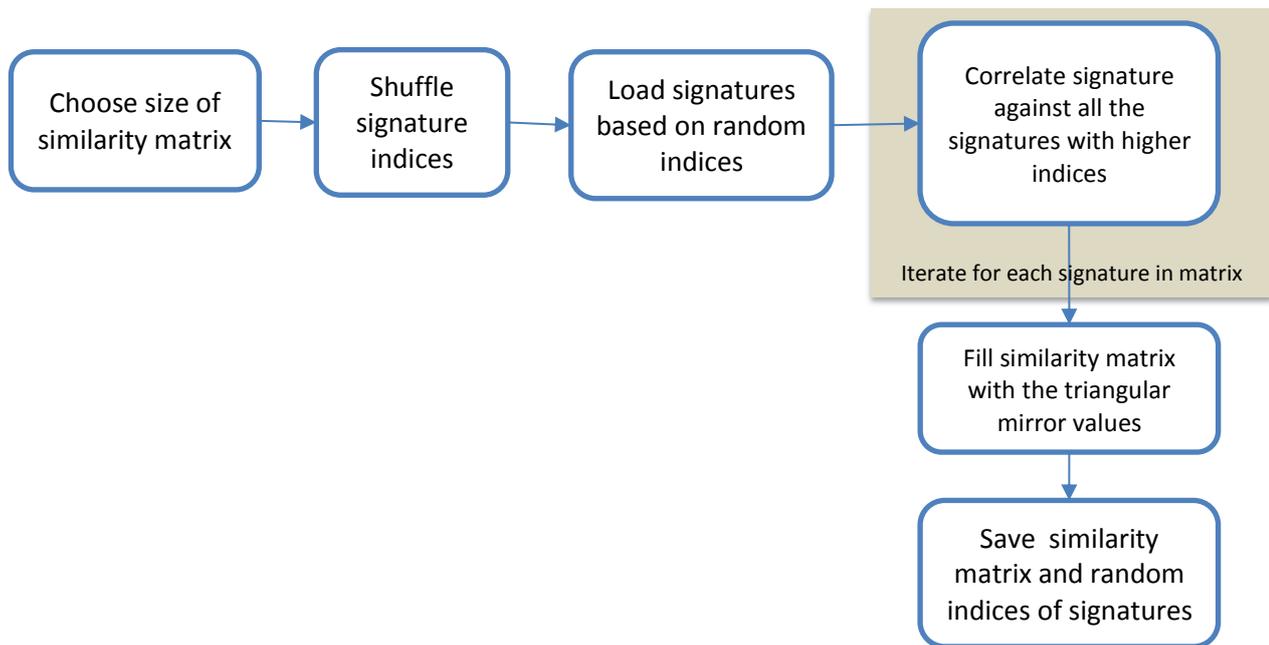


Figure 4.5. Similarity matrix pipeline in existing platform

The size of the similarity matrix is chosen according to the number of images to be classified in the dataset (the choice for optimum size of the matrix is explained in detail in section 4.4). The index of each signature in the dataset is recorded and shuffled. The number of required signatures is picked at random based on the shuffled indices. The first signature that is picked is cross correlated against all the signatures with higher indices in the list and the cross correlation coefficient value is saved as a matrix row. The rest of the signatures are iterated through until the end of the list is reached. Therefore the higher the number of signatures in the similarity matrix, the longer it takes to create the matrix. The correlation values are calculated only for half of the matrix (upper triangular half) since the values are mirrored in a triangular shape. Thus, the whole matrix is filled when the higher triangular half is copied into the lower half. The similarity matrix is saved together with the indices of the signatures that were used to create the matrix.

The classifier is divided into two sections, the training phase and the classification phase respectively and is shown in Appendix A. The training phase uses the similarity matrix (SM) to create the clusters. In the classification phase the rest of the signatures, not forming part of the SM, are placed into the clusters created in the training phase.

For each signature in the SM, i.e. each row of the matrix, the similarity rankings of that signature to the rest of the signatures are stored in descending order. The number of classes is not known beforehand and consequently for each signature, a reference similarity is calculated by finding the boundary between the intra-class and inter-class for that signature. Most signatures that belong to the same class as the current signature will belong to the same class (intra-class) and most signatures that belong to different classes as the current signature will belong to separate classes (inter-class).

A kMeans clustering algorithm is performed on the current row of matrix (minus the self-correlation of the current signature) by dividing the rest of the training set into 2 groups. This will provide two clusters consisting of intra-class and inter-class values for that signature respectively. The centroids of the 2 clusters are found and the mean of these two centroids is then calculated, which represents the boundary between intra-class and inter-class for this signature. Hence when the correlation value of other signatures will be compared to the current signature later, if the correlation value between them is greater than the boundary value, it is more likely that the two signatures belong to the same class. Conversely, if the correlation value between the two signatures is less than the boundary value, it is more likely that the signatures belong to different class.

4.3.2.2 Training Phase

Once the values for the similarity ranking and boundary for all the signatures in the SM are calculated, the training phase starts, which is an iterative process that will stop when there is no change in class membership for two consecutive iterations. The training phase starts by assigning a set of singleton clusters corresponding to each signature in the similarity matrix.

For the first iteration, the size of the voting pool can be varied to reduce the time complexity at the expense of accuracy. The size of the voting pool should be chosen so that there are enough signatures selected from the similarity matrix in order to identify the correct number of clusters. For large training sets (similarity matrices), reducing the size of the voting pool will not have a big impact on the accuracy because more signatures are involved in forming the clusters.

For each signature, i , its voting pool is formed and the class ids are allocated according to the similarity ranking order. Each voter is assigned a different (temporary) class id. The cost of the class ids are calculated by calculating the difference between the correlation coefficient of signature i and signature j , $\rho(i, j)$, and the threshold boundary value for signature i (b_i). If the class id for the lowest cost is different from the class id of i , then the new class id is assigned to signature i .

After all the signatures in the voting pool have been assigned the corresponding lowest cost class id, the vector is stored.

The second and successive iterations are similar to the first one, except that the full size of the training set is used. Moreover, the successive iterations are faster since the class

ids that formed the singleton clusters have been assigned to more relevant clusters. Also, after the temporary class ids are assigned, the duplicates are removed from that list based on the vector that was formed in the previous iteration. When the class ids do not change in two successive iterations or a set maximum number of iterations have been reached, the training phase will end.

4.3.2.3 Classification Phase

The centroids of the clusters formed in the training phase are calculated by averaging all the signatures that form part of their respective cluster. The rest of the signatures that did not form part of the correlation matrix are loaded.

The correlation coefficients between a signature and the centroids of the clusters are calculated. The signature is placed in the cluster with the highest correlation value, which indicates that this cluster is the closest in similarity to that signature. The process is repeated until all the signatures have been assigned to its respective cluster.

A Graphical User Interface (GUI) was developed to enhance the usability of the current image analyser platform and to facilitate the testing of the platform. The signature extraction process, camera reference signature creation, identifier, similarity matrix creation and classifier process each had GUIs created and several scenarios were created to test the platform.

For example, the classifier GUI accepts the following parameters:

- similarity matrix of signatures
- directory of signatures to be classified
- size of voting pool
- directory where the results (information about which signature belongs to which cluster) are saved.

Once the signatures are placed in their respective clusters, the cluster contents are saved for inspection by an analyst or for visualisation purposes. The results can also be saved in a database so that if at a later date more images are recovered, the latter can be checked against the database.

4.4 Testing Experiments on Existing Platform

Once the GUIs had been developed for the image analyser, the testing could be performed to check the performance of the platform under different conditions. The algorithms for identifying and classifying imaging source devices had to be improved in order to (Soobhany 2009):

- Work with images of varying sizes and resolutions.
- Automate the image cropping process.
- Identify images that had been rotated.

Different camera models will produce images of varying sizes and resolutions due to the quality of the cameras. For example a low-end digital camera will create images at its highest resolution that are smaller than a high-end digital camera. Furthermore, the quality of the images between the two cameras will vary greatly. The cross correlation coefficient is calculated for the two signatures, which signifies that both signatures need to be of the same size and this can be achieved by cropping all the images to the same size. The image cropping process should be automated in order to allow the processing of large numbers of images from different camera types.

Some cameras usually rotate the digital image they produce automatically according software settings or preferences. Furthermore it is trivial to rotate a digital picture using normal image processing software and almost all of these rotations are performed so that a picture that was taken in portrait can be viewed better in landscape or vice versa. Hence all these rotations are performed by ninety degree steps, for example by 90, 180 or 270 degrees. Given that the SPN is not invariant to image rotation, the image analyser platform had to be modified in order to identify rotated images. The SPN occurs across the sensor in a set pattern according to the position of the photodiodes and the pattern will change if the image is rotated. Whenever the source of a suspect image needs to be identified against a camera reference signature, correlation is calculated for all four possible 90 degree positions of the suspect image and the highest correlation coefficient value is taken as the result to check against the acceptance threshold. In the case of the camera reference signature, there is no need to check for rotations because the test images are taken under controlled situations.

4.4.1 Testing of Image Analyser Identifier

The testing for the existing platform was performed in two stages, by using photographs taken from several digital cameras as well as different mobile phones. The first testing stage was performed on the *identifier* and the second stage performed on the *classifier* part. For the purpose of calculating the camera reference signature a set of uniformly illuminated pictures of a flat surface was taken. The size of the set consisted of about 50 images. The pictures were neutral grey images with minimum variation in light or of a blue sky. There was a mixture of test images, either blue sky or uniformly lit background was used for the different cameras, and the type of images was not mixed within the same test image set for a camera.

The time taken to calculate the camera reference signature is related to the size of the pictures in the set (50 images). A set of crop size, 512x512 pixels, will take approximately 3 minutes and a crop size of 2592x1944 pixels will take on average 36 minutes. Therefore the smaller the size of the crop images, the time complexity for calculation will be reduced. Table 4.1 shows the time taken to calculate the camera reference signature when the quality (compression) and size of the images are altered. The crop size remains constant at 512x512 pixels and the reference camera is a Nikon Coolpix 5200 digital camera. As can be seen from table 4.1, the quality and size of the image does not have a significant effect on the time taken to calculate the reference signature. The size of the crop has a major impact on the time taken for the calculation of the reference signature as opposed to the resolution of the images before the cropping.

Quality of images	Original size of images before cropping	Time taken to calculate reference snp (mins) for crop size of 512x512
Basic	2048x1536	3
Normal	2048x1536	3
Fine	2048x1536	2.5
Basic	1024x768	2
Normal	1024x768	2
Fine	1024x768	1.8

Table 4.1. Time taken to calculate the camera reference signature, of size 512x512, in relation to quality and original size of images before cropping

Experiments were performed on a Blackberry curve 8320 (PDA) using a combination of high (superfine) and low (normal) quality JPEG images to create the camera reference signature. When the quality of the test images for creating the camera reference signature are of different quality to that of the suspect image, the correlation matching is less than the threshold as shown in table 4.2. In the case when the quality of the images is different, there can be false negatives in the identification process.

Quality of images for camera reference signature	Quality of a suspect image	Correlation > threshold
High	High	Yes
High	Low	No
Low	High	No
Low	Low	Yes

Table 4.2. The correlation between using combination of high (superfine) and low (normal) quality images for creating the camera reference signature and suspect image.

The same experiments, as the ones performed on the Blackberry curve 8320, were carried out on the Nikon Coolpix 5200 digital camera with the different qualities and sizes of both the reference (test) images and suspect images. When the size of the images was kept constant and the quality of the images was varied, there was no effect on the correlation. All the test images used for identification gave a correlation above the acceptance threshold, hence giving a positive result that the picture was taken from the camera. On the other hand, when the quality was kept constant and the size of the images was altered, the correlation value was less than the acceptance threshold.

The results indicate that the quality of the images does not affect the identification process when all the images are of the same sizes, i.e. same resolution. Conversely, when the size of the images differs the matching process will provide more false negatives. Cameras create images of different sizes by performing stronger JPEG compression on the raw images. The results obtained show that identification of imaging source devices using sensor pattern noise (SPN) is more difficult when images of different resolutions (sizes)

are used. The suspect images should be of similar resolutions to the test images used to create the camera reference signature.

Several images taken from the Nikon Coolpix 5200 were rotated and then tested using the *identifier*. The *identifier* cannot detect if the rotated picture had been taken by the camera. The correlation value is much less than the acceptance threshold. Hence, if the image is rotated, the SNP of the picture is altered. The improvement to the *identifier* has been explained in section 4.3 above.

The *identifier* cannot identify images that were taken from camera phones with low quality cameras. The quality and size of the images were altered as well as different background images were used as reference images to create the camera reference signature. The *identifier* did not manage to recognise if any of the suspect images were taken from the corresponding mobile phone camera. The mobile phones are performing some post-processing of the photographs taken and hence inserting a higher level of noise in the photograph, for example stronger compression of the images. As a result of the increased noise, the Wiener filter cannot perform the denoising to extract the relevant SPN.

The testing performed mainly concentrated on the signature extraction and identification processes of the image analyser. For the image classification process the next two sections, section 4.4.2 and section 4.4.3, describe some of the testing that was performed on the classifier.

4.4.2 Classifier Grand Tour Visualisation

The preliminary testing on the classification process was performed in two stages by first creating the similarity matrix on some randomly chosen signatures and using the resulting matrix to train the classifier to identify the clusters. The creation of the similarity is the most computationally intensive process in the classification process. Figure 4.6 shows the plot of the average time taken to create the matrix for 231 images, i.e. the size of the matrix is 231 rows by 231 columns. The experiments were run on a laptop with a single core processor with 2 GB of RAM. The extracted signatures of the images are in the high dimensional space, for example a crop size of 512x512 can produce a signature of nearly 900K dimensions for all the three primary colour channels. The difference in dimensions of the signature between the 256x256 crop and the 512x512 crop is nearly 4 times as much. Thus the time taken to create the similarity matrix for signatures of size 512x512 will be expected to be four times as long as the 256x256 crops, but due to other loads (operating system and other software) on the processor and memory, the time taken is more than four fold.

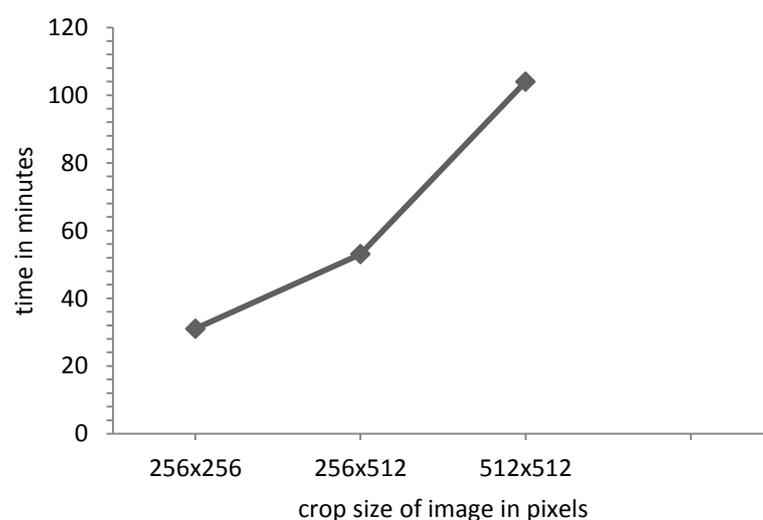


Figure 4.6. Plot of the average time to create similarity matrix, in minutes, based on different image crop size

In order to understand the meaning of the values in the similarity matrix, another method of clustering the signatures according to their source device was applied to the matrix which would help the visualisation of the clusters. The clustering process was geared towards computational methods that ‘tour’ the data systematically looking for interesting structure – clusters, outliers and holes. These are variants of Asimov’s Grand Tour (Buja, Asimov 1986, Buja, Cook et al. 1996, Asimov 1985) and, more generally, projection pursuits that try to look at data set in many 2/3D views with the goal of discovering something interesting and information (Cook, Buja et al. 1995, Lam, Emery 2009).

The Grand Tour procedures are dynamic but are not generally interactive in the sense that the user cannot guide the tour path in the N-dimensional space, by selecting or specifying the starting and end planes (or subspaces) onto which the data are projected. Such interactive tours then proceed from one plane to the other by geodesic interpolation paths between the two subspaces, thus presenting to the user a different view of the data at each step of the interpolation sequence. The latter is also known in some literature as an interpolation tour, whose underlying mathematical foundation was presented in detail in (Asimov & Buja 1994, Hurley & Buja 1990).

A major consideration of such tours concerns the specification of the target spaces. The commonly suggested sub-spaces include those spanned by subsets of the eigenvectors in Principle Component Analysis (PCA) (Lam & Emery 2009). In pattern classification, a number of statistical methods exist including the projection pursuit. Whilst the original technique of projection pursuit is not a visual tour, its use in exploratory analysis, and other similar applications, is well documented (Huber 1985, Friedman & Stuetzle 1981, Jones, Sibson 1987). This tour seeks subspaces that have maximal structure as defined by

some measure, commonly known as the projection index, on which a direct search is based. In this sense, the tour is guided by the data with such indices that indicate certain types of structure that the user is looking for. It keeps touring until possible structures are found.

Standard statistical techniques attempt to test a hypothesis (confirmatory analysis), they tell us how well a dataset fits to a hypothesis. Exploratory analysis is a complementary activity; it looks for underlying patterns within a dataset in a hypothesis free approach, which is similar in some respects to the unsupervised classification method that the current platform uses. The *Grand Tour* allows the analyst to view the data from different angles in the high-dimensional signature space and to determine different patterns. The similarity matrix that was created from the signatures was loaded in the visualisation platform developed by Emery & Lam (Emery, Lam 2008). Functions like *record*, *stop*, *save*, etc are used to stop the movie to view that particular pattern.

A selection 50 images in total from four different cameras was used. The Grand tour platform was found to be limited to visualise 50 dimensions. Beyond that amount of dimensions the computation complexity made it difficult to visualise the 'tours'. The images were sourced as follows:

- 10 images from an Olympus Mju Stylus 1030sw
- 15 images from a Canon Digital IXUS 500
- 7 images from a Blackberry Curve 8310
- 18 images from a Nikon Coolpix E5200

The similarity matrix was created using all the 50 signatures extracted from the images with a centre crop size of 512x512 pixels. The system accepts the similarity matrix of the SPNs and the user starts the animation. Figure 4.7 shows a snapshot of a guided tour with four distinct clusters formed suggesting that the images come from four devices (Soobhany, Lam et al. 2009). Once the animation is started, all the planes are visited in turn and they are displayed as a sequence in the 3D plot on the right side of the window. The 'tour' displays each plane like the image frames in a video sequence (movie). The user can alter the speed of the animation and can also switch the guided mode off.

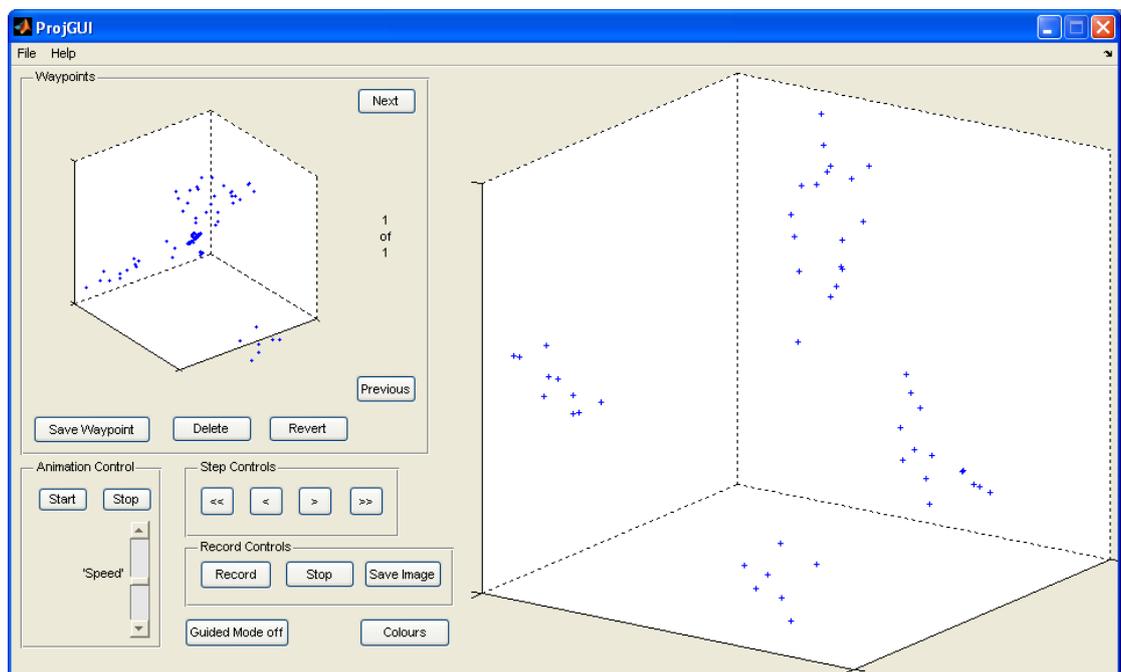


Figure 4.7. Snap shot of the Guided Tour of 50 signatures and 3D plot on the right side shows the four separate clusters.

The user watches the animation until s/he finds an interesting projection plane, similar to the one displayed above, where all the clusters are clearly defined. If the user is looking to

find some outliers or specific plane, then the *step control* buttons can be used to look at each plane in turn. In the example used, the four clusters are shown in Figure 4.7 and they are clearly defined. Practical implementation requires that the sequence of planes is uniformly distributed, so that the movie does not stay too long in one part of the high dimensional space. Computationally, the sequence of projections is dense in the space of all planes and the tour converges to form a pattern.

4.4.3 Cross Validation

The accuracy of a classifier can be estimated by finding its associated prediction error. The main techniques that can predict the errors of classifiers are Holdout (Webb 1999), Bootstrap (Efron, Tibshirani 1997) and cross validation, which is probably the simplest and most often used (Hastie, Tibshirani et al. 2009). Cross validation can be used to compare between classifiers or, in our case, implemented to predict the performance of a single classifier with different sample sizes. This technique is most commonly used in supervised learning, by dividing the dataset into 3 parts, namely training set, validation set and test set and can be used to predict the performance of the supervised learning classifier. The training is performed on the training set, then cross validation is applied to calculate the training error, and finally the test error is calculated based on the test set. The training error can be compared against previous values that were learnt by the classifier. For unsupervised learning classifier, since there is no memory of previous datasets or features, the application of cross validation is not straightforward. Only a training set and a test set were used, since the clusters formed after the classifier training (stage 3 in

Figure 4.4) are already available. The prediction error is calculated by finding the number of class labels that have been misclassified from our dataset.

K -fold cross validation is a method where the dataset δ of size N , is randomly partitioned into k mutually exclusive sets (folds) of roughly the same size. The classification is performed on $k - 1$ partitions (training set) and the remaining partition is applied to the classifier (test set) as shown in Figure 4.8, where the grey boxes correspond to the test set and the white boxes to the training set. The number of labels that are falsely classified from the test set will provide the prediction error $\xi(x_i)$ of the i_{th} partition x_i . The process is repeated K times, where each partition is used as the test set in turn, and the overall prediction error CV_{err} is given by

$$CV_{err} = \frac{1}{K} \sum_{i=1}^k \delta(x_i, \xi(x_i))$$

Take the example when the value of k is 5 and the size of the dataset is 100, with 20 acquired by each camera. There will be 5 partitions of about 20 items in each and cross validation will be executed 5 times. The value of k can be 10 or 20 and if the value of k is 2, k -fold becomes a variation of the holdout method. In the case when the value of k is 1 (also known as leave-one-out), the cross validation will be run 100 times, which is computationally expensive as well as giving a high variance. Although there is a high variance the bias is zero, because all the readings are used for the cross validation. The other values of k will give more bias than leave-one-out but their variance is lower. The typical value of $k = 10$ gives a good trade-off for bias and variance as well as being less computationally intensive.

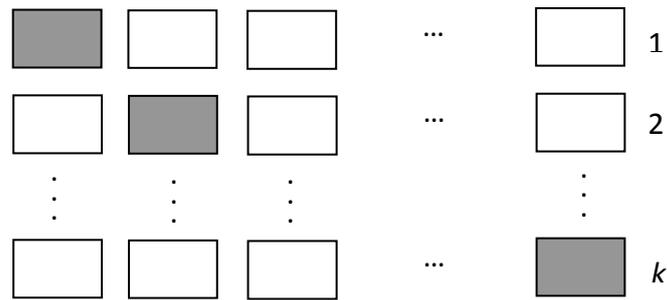


Figure 4.8. k partitions blocks for cross validation

The folds (partitions) can also be stratified, where they will each contain approximately the same amount of labels as compared to the dataset. This ensures that the folds are representative of the distribution of labels in the dataset. Repeating the cross validation multiple times by using different labels for the partitions after each complete run, generally provides a better Monte-Carlo estimate at an added cost (Kohavi 1995). E.g., running the 10-fold cross validation ten times, hence 100 times in total, will provide a better estimation of the error with less variance. The repeated cross validation was not implemented in this experiment due to the additional computational cost it involves and the single cross validation provided a good Monte-Carlo estimate.

For the purpose of the experiments 1000 pictures were used from 5 source devices comprising of 4 digital cameras and one camera phone. The devices were BlackBerry Curve 8310, Nikon Coolpix E5200, Canon Digital IXUS 500, Olympus C-730 UZ, and Olympus Mju Stylus 1030SW. Since images in real forensic situations come from different cameras with different specifications and settings, I did not want to set a specific size for

the photos from the cameras, therefore the sizes ranged from 1600 x 1200 pixels to 3648 x 2736 pixels, which still give a wide margin for cropping of the images. Most of the images were taken at the highest resolution possible for the specific camera. All the images were in JPEG format with a compression quality ranging from about 75 % to 97 % quality. The pictures contain a wide variety of indoor and outdoor sceneries of urban and rural settings, night and day lighting as well as offices and various buildings and some holiday pictures; overall images that would be as close to natural real world situations as possible were captured.

The cropping was performed on the images and the signatures (SPN) were extracted by using a Discrete Wavelet Transform (DWT) followed by a low pass filter and the SPN were enhanced by using one of the models developed in Li (Li 2010). The model chosen was the:

$$n_e(i,j) = \begin{cases} e^{-0.5n^2(i,j)/\alpha^2} & , \text{ if } 0 \leq n(i,j) \\ -e^{-0.5n^2(i,j)/\alpha^2} & , \text{ otherwise} \end{cases}$$

this has been shown to work in (Li 2010) for natural images and the optimal value of α was 7.

To perform the cross validation on the dataset, the folds were randomly selected before the training stage of the classifier. The next step was to create the similarity matrix which, being the most computationally intensive stage of the classification process was no longer a trivial process since for each fold the matrix had to be recalculated. This had the prospect of making the cross validation prohibitively expensive computationally and arises due to the change of size of the matrix by changing the size of the fold and the

different images present in the training phase for each partition. To overcome this problem, the similarity matrix was created at the start of the cross validation process for all the SPNs. Row and column deletions were performed in the matrix as the population of the partitions changed and the size of the folds were altered. This method increased the time complexity to create the similarity matrix, but it had to be created only once, which decreased the time complexity during the cross validation process.

The sizes of the folds were chosen as 20, 10, 5, and 2, which provided a wide range of prediction errors. The size of the dataset was decreased from 1000 to 500 and to 250 so as to provide a learning curve that will indicate what effect the sample size has on the classifier. Once the value of k is chosen for the cross validation and one partition set aside for testing, the other partitions are combined to provide data for the training phase and classification phase. In Li (Li 2010), the selected sizes of the classifier trainer set were 120 and 300, where there was little variation of classification performance between the two sizes. The size of the trainer set was altered with different dataset sizes. Moreover, the partitions were stratified, so that each partition contained approximately the same amount of labels.

4.5 Limitations

The testing performed on the *Image Analyser* platform provided some interesting results for both the *Identifier* and *Classifier*. Some limitations of the software platform were already identified before the testing procedure started. Therefore these improvements were performed on the current platform prior to starting the testing experiments (in section 4.4), such as:

- Processing of images of varying sizes and resolutions
- Automation of the image cropping process
- Identification of images that had been rotated

The cropping of the images has been shown to greatly improve the signature extraction time. For example signature extraction from an image of crop size of 512x512 pixels will take approximately 3 minutes and a crop size of 2592x1944 pixels will take on average 36 minutes. The image cropping process is performed in an ad hoc way, where most of the image cropping is performed from the centre of the image. Thus a limitation of the software platform is the ad hoc nature of the cropping process.

The testing on the *Identifier* with pictures from digital cameras provided positive identification matching results when the respective reference camera signatures were tested against suspect images. The purposes of the tests were mainly to test the findings of Li (Li 2010), and the results for digital cameras and some mobile phones did provide positive matching. Images from the digital cameras were successfully identified when matched against the respective camera reference signatures. One limitation of the identifier was found to be the identification of cameras that produced low resolution

images and highly compressed JPEG images. Test images were taken from Nokia N95 and Nokia C2 camera phones, where the camera reference signatures were created. When some sample images were tested against the reference signatures, the matching values were much lower than the acceptance threshold which indicated false negatives. The inability to identify highly compressed images extends to the clustering of images in the classifier. The images from the camera phones were not placed in the clusters corresponding to the source devices.

The software tool based on the guided tours did provide a way to visualise the formation of the clusters. The stopping condition for the formation of the clusters is dependent on the human analyst/user seeing an interesting pattern or outliers. When the amount of signatures to be classified is low, it is trivial to observe the formation of the clusters, but when the number of signatures increases the number of dimensions will increase too. Thus the computing complexity to project the different planes increases and makes it difficult to visualise the formation of the clusters. Moreover, it can take a long time to 'tour' through the planes which will render the task of the analyst almost impossible to sit and go through all the 'tours'. The limit of the dimensions that could be visualised in the experiments was found to be about 50 dimensions.

4.6 Plan to Address Limitations

The cropping position of the images prior to extracting the signatures can be chosen in a systematic way based on some statistical properties (e.g. mean and variance of scene interference) or empirical methods. An experiment will have to be performed based on images from different cameras taken at different times of the day and night as well as different scenes. The images will have to be cropped from different positions and the signatures extracted, which can be processed by the *image analyser*.

The most significant limitation of the current *image analyser* platform is the lack of effectiveness of identification or classification of images from low-end to medium end camera phones. These images are highly compressed JPEGs, which have low resolutions and are smaller in size too. The compression of digital images in the JPEG format is performed by transforming the image in the frequency domain where the high frequency details of the image is removed, since the human visual system is more attuned to the medium frequencies. The lossy nature of the compression method makes the image smaller and easier to store. Moreover, the compression process introduces blockiness in the image when the pixels are grouped together in blocks of 8 by 8 to be quantized. The JPEG compression is explained in more details in section 2.2.1.

The Sensor Pattern Noise comprising of the PRNU lies in the medium to high frequency region of the image details and is attenuated when the images are highly compressed and processed, which makes it more difficult to be extracted. The signature extracted by using the wavelet method usually contains residual high frequency details from the image, which contaminates the signature. The enhancing process of the signature usually decreases the magnitude of the overall SPN which attenuates the high frequency

contaminating details in the signature but does not make it an ideal solution for the low resolution images on its own. After the SPN has been enhanced, the rows and columns of the SPN can be zero-meant and a Wiener filter applied after a Fourier transform was performed in order to reduce the blocking effect of JPEG compression. It will not allow the identification of low resolution images and highly compressed JPEG images. The reason for this is because; when the enhancer is applied to a signature, the magnitude of the overall signature decreases and the already weak PRNU can be attenuated by a small proportion. Therefore when the zero-meaning and the wiener filtering are performed, the SPN is too weak to act as a reliable signature. A different method would be to explore the nature of the PRNU noise and try to estimate where it lies within the image. Once the PRNU is located, it can be isolated and extracted to produce a stronger signature for highly compressed images.

Chapter 5 SVD Based Signature Extraction Model

5.1 Introduction

The image analyser platform investigated in Chapter 4 uses the wavelet based extraction method in conjunction with the signature enhancer. One of the limitations of the platform is the identification of images that originates from low to mid-end camera phones, which are heavily compressed images. The characteristics of the SPN, which contains the PRNU, were explored and if the PRNU can be converted from a multiplicative noise to an additive noise, it will therefore be much easier to extract the signature from the image. Furthermore, if the range of the energy of the PRNU within an image can be estimated, it will be easier to separate the PRNU from other polluting noises and get a cleaner signature.

The concept of signal decomposition, by extension the 2D image decomposition, is presented and the Singular Value Decomposition (SVD), which can separate an image into ranks of descending order of energies, is investigated. The homomorphic filtering approach is explored and finally a new signature extraction model is presented that can estimate where the PRNU lies within the image. Once the PRNU is located, it can be isolated and extracted to produce a stronger signature for highly compressed images.

5.2 Signal Decomposition

Signal decomposition is an important practical problem as the energy in most real-world signals has unevenly distributed frequency spectra. The uneven distribution of signal energy in the frequency spectra has made signal decomposition an important problem whose solution provides the practical foundation for signal compression techniques under the classical rate-distortion formalism in source coding applications (Berger 1971). The basic principle is to divide a spectrum into signal subspectra (or subspaces) in order that those with more energy content will be given a significantly higher priority for further processing. By analysing and discarding signal subspaces with lower priority, one can thus expect a negligible (synthesis) error in the reconstructed signal to occur subsequent to the *decomposition-synthesis* procedure widely adopted in many real-world applications. In addition to signal compression/coding, such procedure also forms the mathematical basis of modern or non-classical time-frequency based techniques for investigating signal subspace/subbands, providing an expansive means of spectral analysis that naturally leads to the transform coding methods representative of the non-parametric or *eigen decomposition* approach of spectral estimations. These methods are frequently described in the literature as having higher resolution and better frequency estimation characteristics particularly at high noise to signal ratios (or low SNRs). For example, they are particularly effective in identifying narrowband processes in white noise.

Broadly speaking, most eigenvector approaches work by separating the information contained in a signal into two subspaces, which are commonly referred to the signal and noise subspace respectively. The related decomposition/transform generates eigenvalues

of decreasing order and, importantly, eigenvectors that are orthonormal. This latter property is most crucial in ensuring eigenvectors that are deemed part of the noise subspace can be identified, thus allowing influence of that noise to be eliminated effectively. The most challenging aspect of applying eigenvector spectral analysis is the selection of the appropriate dimension of the signal or noise subspace; e.g. if the number of narrowband processes is known, then the signal subspace can be dimensioned on this basis. In general, however, the determination of the signal subspace often relies on trial-and-error approach.

Historically, the purpose of transform coding is to decompose a set of correlated signal samples into a set of uncorrelated spectral coefficients, with energy concentrated in as few of them as possible. Indeed, the orthogonal expansion of a continuous variable function is a subject of extensive studies documented in classic literature (examples are Fourier series and wavelets), which gained in both depth and intensity in the 70s, providing the theoretical underpinnings for many modern applications. From a practical signal processing viewpoint, the decorrelation and energy compaction properties of these transforms constitute the central issues in their applicability for signal coding, laying the important foundations for signal classification and identification particularly for speech and images. Using established concepts from linear vector spaces, a vector-matrix formulation providing a succinct format for block transform manipulation/interpretation is given below.

For the rest of the thesis the following notations will be used:

- Scalar sequence: normal font, lowercase
- Vector: bold font, lowercase
- Matrix: bold font, uppercase

5.2.1 One-Dimensional Transform Coding

Let $f(h)$ be the one dimensional (1D), discrete time signal, sequence representing a the continuous time signal $f(t)$, defined over the interval $[0, N-1]$. Accordingly, $f(h)$ can be expressed as an N -dimensional vector \mathbf{f} by means of the principle of superposition (c.f. linear systems theory):

$$\mathbf{f} = \begin{bmatrix} f(0) \\ f(1) \\ \vdots \\ f(N-1) \end{bmatrix} = f(0) \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + f(1) \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \dots + f(N-1) \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

$$\mathbf{f} = f(0)\mathbf{e}_0 + f(1)\mathbf{e}_1 + \dots + f(N-1)\mathbf{e}_{N-1}$$

This formulation allows \mathbf{f} to be viewed as a point in the N -dimensional Euclidean space spanned by the basis vector set: $\{\mathbf{e}_j, \text{ where } j = 0, 1, \dots, N-1\}$ and these vectors are also linearly independent.

By definition, two sequences $f(h)$ and $g(h)$ with the same support are orthogonal if and only if the inner (or dot) product vanishes, viz,

$$\sum_{j=0}^{N-1} f(j)g(j) = \mathbf{f}^T \mathbf{g} = 0$$

Thus, the basis vector \mathbf{e}_j as described earlier are orthogonal, given that

$$\sum_{h=0}^{N-1} \mathbf{e}_i(h) \mathbf{e}_j(h) = \mathbf{e}_i^T \mathbf{e}_j = 0, \quad \forall i \neq j$$

with $\text{norm}(\|\mathbf{e}_j\|)$ of each basis vector equal to unity:

$$\sum_{h=0}^{N-1} |\mathbf{e}_j(h)|^2 = 1, \quad \forall j = 0, 1, \dots, N-1$$

\mathbf{e}_j provides the simplest set of basis vectors for which the orthogonal expansion of $f(t)$ evaluates to $f(h)$.

In general, f can be characterised using a broad class of orthogonal expansions defined as follows.

Let $\mathbf{X}_n(h)$, $0 \leq n$, $h \leq N-1$ be a family of N linearly independent sequences which are defined on the interval $[0, N-1]$. The orthogonality property requires that,

$$\sum_{h=0}^{N-1} \mathbf{X}_n(h) \mathbf{X}_m(h) = l_n^2 \delta_{n,m} = \begin{cases} l_n^2, & n = m \\ 0, & \text{otherwise} \end{cases}$$

where $l_n = \|\mathbf{X}_n(h)\|$ and \mathbf{X}_m^* denotes the complex conjugate of \mathbf{X}_m . In practice, the orthogonal family is normalised with $\text{norm}(\mathbf{X}_n) = 1$ to form the orthonormal family; that is,

$$\mathbf{W}_n(h) = \frac{1}{l_n} \mathbf{X}_n(h), \quad 0 \leq n \leq N-1$$

giving $\sum_{h=0}^{N-1} \mathbf{W}_n(h) \mathbf{W}_m^*(h) = \delta_{n,m}$

where $\delta_{n,m}$ is the Kronecker-delta sequence as given by

$$\delta_{n,m} = \begin{cases} 1, & n = m \\ 0, & \text{otherwise} \end{cases}$$

A familiar example of the orthonormal basis, in the time vector space for the orthogonal

expansion, is the Discrete Fourier Transform (DFT) of f , with $W_i(h) = \frac{1}{\sqrt{N}} e^{-j2\pi \frac{if(h)}{\alpha}}$

where $j = \sqrt{-1}$ and $f(h)$ uniquely characterised as

$$f(h) = \sum_{n=0}^{N-1} \theta_n W_n(h), \quad 0 \leq h \leq N-1$$

where the spectral coefficients ϑ_i is given as

$$\theta_i = \sum_{h=0}^{N-1} f(h) W_i^*(h), \quad 0 \leq i \leq N-1$$

(which can be computed by multiplying $f(h)$ by $W_m^*(h)$ and summing over h). That is,

$$\sum_h f(h) W_m^*(h) = \sum_n \theta_n \sum_h W_n(h) W_m^*(h) = \sum_h \theta_n \delta_{n+i} = \theta_i$$

In the most general case, the set of coefficients ϑ_i , $0 \leq i \leq N-1$ specifies the spectral characterisation of f relative to the given orthonormal basis function W , for which the classical DFT is a special case where $W_n(h)$ is sinusoidal. In particular, the energy of a signal sequence, computed as the sequence of norm $|f(h)|$

$$\sum_{h=0}^{N-1} \|f(h)\|^2$$

can be computed equivalently as the sum of square of norm $|\vartheta_n(h)|$ as per the Parseval Theorem.

That is,

$$\sum_h^{N-1} f(h)f^*(h) = \sum_h^{N-1} \left(\sum_n^{N-1} \theta_n \mathbf{W}_n(h) \right) \left(\sum_n^{N-1} \theta_n \mathbf{W}_n(h) \right)^* = \sum_n^{N-1} \theta_n \theta_n^*$$

or

$$\sum_{h=0}^{N-1} \|f(h)\|^2 = \sum_{n=0}^{N-1} |\theta_n|^2$$

which asserts that the signal energy is preserved under an orthonormal transformation and can be measured by the square of the norm of either the signal samples, $f(h)$, or the spectral coefficients obtained via the orthogonal expansion of $f(h)$.

The mathematical development leading to unitary matrix $\mathbf{X}^{-1} = \mathbf{X}^{*T}$, subsequently reinstating the Parseval Theorem; i.e. energy preserving relation of \mathbf{f} and $f(h)$; i.e. $\mathbf{W}^T \mathbf{W} = \mathbf{f}^T \mathbf{f}^*$, and, ideally, derivation of the resulting least squares estimator.

Typically the decorrelated components of \mathbf{W} , $\{w\}$, have different variances which simply state that the corresponding signal coefficients sequence is non-stationary. Thus a second important objective of the transformation is to repack the signal energy (see above equation) into a relatively small number of spectral coefficients; that is, the performance/metrics or quality of an orthonormal transformation depends on its signal decorrelation and energy repacking properties. Computationally, the classical technique of zonal sampling is often used for approximating \mathbf{f} wherein only a small subset of the spectral coefficients is used to represent the original signal vector (least squares approximation already addressed in the above derivation). Thus the best zonal sampler is one that packs the maximum energy into the first L coefficients $\{w_L, L = 1, N\}$, where $L =$

$\min\{N\}$. The Karhunen-Loeve transform (KLT), a signal/input dependent block transform best known for its optimality in data compaction (Jolliffe 1986), has this property.² A brief summary of KLT is presented below (Lam, Emery 2009).

Given \mathbf{x} is an N random vector, KLT computes orthogonal transforms that maximise the amount of data variance with the goal of seeking a projection to best represent the data in a least-squares sense. Geometrically, this can be understood by noting that whilst the mean vector \mathbf{m}_x of the set of d -dimensional samples $\{x_1, x_2, \dots, x_n\}$ offers the minimum sum of squared distances between itself and the set of samples that it represents, a more *interesting* representation of the data set can be obtained by projecting the data onto a line running through \mathbf{m}_x . The procedure is followed by finding an optimal set of coefficients $\{a_k, k=1, n\}$ which minimises the squared-error criterion function J concerning the sum of squared differences between \mathbf{x}_k and $(\mathbf{m}_x + a_k \mathbf{e})$, where \mathbf{e} denotes a unit vector in the direction of the line sought. In other words, the least-squares solution is achieved by projecting the vector \mathbf{x}_k onto the line in the direction of \mathbf{e} that passes through the sample mean \mathbf{m}_x . Now let the matrix \mathbf{A} define a linear transformation that generates a new vector \mathbf{y} from \mathbf{x} , as $\mathbf{y} = \mathbf{A} (\mathbf{x} - \mathbf{m}_x)$, where \mathbf{A} is constructed such that its rows are the eigenvectors of \mathbf{C}_x , the covariance matrix as defined by:

$$\mathbf{C}_x = \mathbf{E}\{(\mathbf{x} - \mathbf{m}_x)(\mathbf{x} - \mathbf{m}_x)^t\} = \frac{1}{N} \sum x_n x_n^t - \mathbf{m}_x \mathbf{m}_x^t$$

² In essence, KLT is a continuous signal transformation analogous to that developed by Hotelling (Hotelling 1933) removing the correlation among the elements of a random vector and hence, the widely known method of principal components.

which is real and symmetric, with diagonal elements representing the variances of the individual random variables whilst the off-diagonals are their covariances. For convenience, the row eigenvectors of \mathbf{A} are constructed in order of decreasing magnitudes of the corresponding eigenvalues. The transformed vector \mathbf{y} is a zero-mean random vector whose covariance matrix (\mathbf{C}_y) is related to that of \mathbf{x} by $\mathbf{C}_y = \mathbf{A} \mathbf{C}_x \mathbf{A}^T$; i.e. \mathbf{x} and \mathbf{y} are *similar* under the similarity transformation that preserves eigenvalues (as \mathbf{A} is real and unitary), with

$$\mathbf{C}_y = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{bmatrix}$$

that represents a diagonal matrix having the eigenvalues of \mathbf{C}_x along its diagonal. Thus, the linear orthogonal transformation of \mathbf{A} removes the correlation amongst the original variables of \mathbf{x} , resulting in elements of \mathbf{y} which are uncorrelated (as the off-diagonal elements of \mathbf{C}_y are all zero). Furthermore, each λ_k is the variance of y_k , the k^{th} element of \mathbf{y} , which is ordered in decreasing value of k in accordance with the order of the eigenvectors in \mathbf{A} . As such, the dimensionality of \mathbf{y} can be reduced by ignoring one or more eigenvectors that have small eigenvalues; that is, given $M \ll N$, the transformation matrix \mathbf{B} ($M \times N$) having the first M rows of the matrix \mathbf{A} will produce new transformed vectors of $\tilde{\mathbf{y}}$ which are smaller (M columns), as given by $\tilde{\mathbf{y}} = \mathbf{B} \tilde{\mathbf{x}}$. Here the reconstructed vectors $\tilde{\mathbf{x}}$ can be obtained by means of the inverse relation as $\tilde{\mathbf{x}} = \mathbf{B}^T \tilde{\mathbf{y}}$, whose mean square error (MSE) is simply the sum of the eigenvalues corresponding to the discarded eigenvectors. That is, $\tilde{\mathbf{y}}$ represents the least squares fit of \mathbf{xx} ; viz.,

$$MSE = \sum_{k=M+1}^N \lambda_k$$

More importantly, if \mathbf{C}_x is singular, its rank R is less than N and it will have $N-R$ zero eigenvalues. Thus PCA (*Principal Component Analysis*) or the associated KLT/Hotelling Transform always reduces the dimensionality from N to R , gracefully yielding a computationally more tractable solution.

5.3 Singular Value Decomposition

Using signal compaction techniques, the energy of a signal can often be redistributed into a significantly smaller number of frequency sub-bands, allowing them to be divided into sub-spectra in order that those with more energy content will be given a significantly higher priority for further processing. By analysing and discarding signal subspaces with lower priority, a signal can be reconstructed or approximated by a decomposition-synthesis procedure described in section 5.2.1.

Most eigenvector approaches work by separating a multidimensional signal into two subspaces, which are commonly referred to the signal and noise subspaces. In practice, the most challenging part of eigenvector spectral analysis is to compute the appropriate dimension of the signal or noise subspace, which often resorts to a trial and error procedure. Mathematically, a matrix \mathbf{A} with m rows and n columns with rank r , $r \leq n \leq m$, can be expanded or decomposed into:

$$\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}^T$$

where \mathbf{U} and \mathbf{V}^T are two orthogonal matrices of size $m \times m$ and $n \times n$ respectively (Moler 2004). \mathbf{S} is the diagonal matrix, of size $m \times n$, containing r non-zero singular values. The decomposition of matrix \mathbf{A} is known as Singular Value Decomposition (SVD). When an

image is decomposed using SVD, the ranks of the image can be represented as component matrices with decreasing energy contents as shown in Figure 5.1 (Andrews, Patterson 1976). SVD can be used to separate the spectra of the image and the ranks can be selected in accord with the aggregated total image energy of the individual ranks; that is, sum of $\sqrt{\lambda_i^2}$, where λ_i represents the eigenvalue associated with that eigenvector (\mathbf{e}_i).

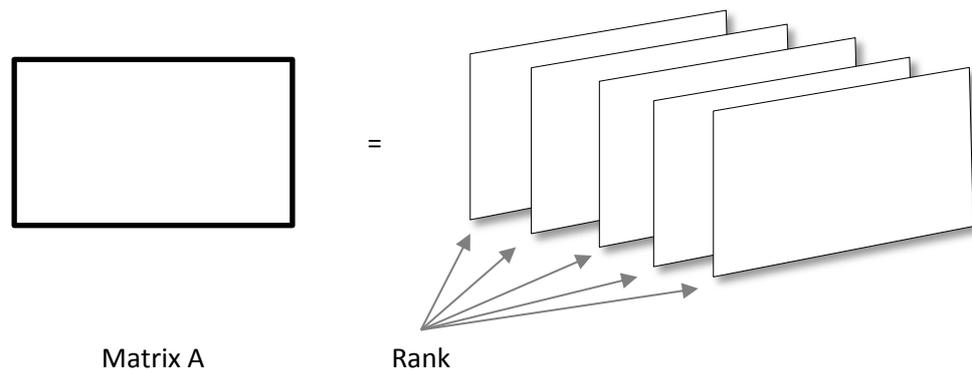


Figure 5.1. SVD of matrix A with each rank as a separate matrix

5.3.1 2D Block Transform Coding

The 2D formulation of transform coding is easily extrapolated from the discussion in section 5.2.1. Generally, a 2D array or image is divided into $(N \times N)$ subblocks, each of which is separately encoded. These blocks are usually square, with 4x4, 8x8 and 16x16 being representative sizes. In a similar vein, the extension of the least square fit procedure of KLT to 2D images leads to an efficient image dependent decomposition known as the *Singular Value Decomposition* (SVD). Here an $N \times N$ image/block \mathbf{F} is computed using the weighted sum of the N^2 basis array $\mathbf{B}_{i,j} = \boldsymbol{\Phi}_i \boldsymbol{\Phi}_j^T$, wherein the $\{\boldsymbol{\Phi}_i\}$ are columns of a preselected unitary matrix. An outer product expansion is sought, wherein the outer products are matched specifically to the particular image so that the

double sum over N^2 basis images reduces to a single sum over R arrays, where $R \leq N$. The matrix expansion required has the form:

$$\mathbf{F} = \sum_{i=1}^R \sigma_i \boldsymbol{\Phi}_i \boldsymbol{\Psi}_i^T = \sum_{i=1}^R \sqrt{\lambda_i} \boldsymbol{\Phi}_i \boldsymbol{\Psi}_i^T$$

where R is the rank of \mathbf{F} (see above); viz. $R = \text{rank}(\mathbf{F})$. The expansion (or decomposition) is most commonly known as the SVD, which can be constructed as follows. Define the $N \times R$ transformation matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ such that the R columns of each of these matrices are the R non-zero eigenvectors of $\mathbf{F}^T \mathbf{F}$ and $\mathbf{F} \mathbf{F}^T$ respectively. Further, since both $\mathbf{F}^T \mathbf{F}$ and $\mathbf{F} \mathbf{F}^T$ have the same eigenvalues, we have

$$(\mathbf{F}^T \mathbf{F}) \boldsymbol{\psi}_k = \lambda_k \boldsymbol{\psi}_k$$

$$(\mathbf{F} \mathbf{F}^T) \boldsymbol{\Phi}_k = \lambda_k \boldsymbol{\Phi}_k, \quad k = 1, 2, \dots, R$$

with non-zero eigenvalues $\{\lambda_k\}$ are the singular values of \mathbf{F} . It can be shown that \mathbf{F} can be written as the well-known expression:

$$\mathbf{F} = \boldsymbol{\Psi} \sqrt{\boldsymbol{\Lambda}} \boldsymbol{\Phi}^T$$

which, upon expansion, produces

$$\mathbf{F} = \sum_{i=1}^R \sqrt{\lambda_i} \boldsymbol{\Phi}_i \boldsymbol{\Psi}_i^T$$

as described above. As with the 1D expansion, for $R \ll N$, the signal compaction achieved by SVD is always optimal in the least squares sense. As such, the N^2 image samples often need only be encoded with $2NR$ samples; i.e. R times of the N samples for each $\boldsymbol{\Phi}_k$ and

ψ_k . In other words, Φ and ψ are determined solely by the particular image being analysed (cf. KLT) and, as such, must be re/calculated for each image block in order that it can be expanded by SVD with least squares error. The familiar outer-product expansion of $F = U\sqrt{\Lambda}V^T$ that we have been applying to F can be viewed as a limiting case where $N = 1$; i.e. 2D transform applied to each image pixel.

5.1.1 Interpretation of SVD components

The major issue associated with SVD is the interpretation of the linearly combined components (planes) that are created. The energy of the ranks in relation to the aggregated total image energy can be represented by the eigenvalue, λ_i , where $i = \{1, 2, \dots, N\}$, where N is the number of ranks. The variance, σ^2 , of the image is proportional to the sum of the eigenvalue squares.

$$\sigma^2 \propto \sum_{i=1}^N \lambda_i$$

A scree plot, line plot showing the fraction of total variance in an image, can be used to display how the components are distributed. The energy of an image is proportional to its variance and the relevant components of the SVD can be computed by using the properties of the variance and eigenvalues.

5.4 Homomorphic filtering

Homomorphic filtering is a process where a multiplicative noise, such as the PRNU, can be transformed to an additive noise. The process works by mapping the signal (or image) to a different domain where linear filtering techniques can be applied to denoise or enhance the signal. Consider a signal $f(x)$ in the spatial domain, which is a combination of two sub-signal components $m(x,y)$ and $n(x,y)$ represented as

$$f(x, y) = m(x, y)n(x, y)$$

The relationship between m and n is multiplicative and it can be difficult to separate these two signal components in the spatial domain using linear techniques. If the frequency components of m and n had to be operated separately, then the above equation cannot be used directly because the Fourier transform of the product of two functions is not separable (Gonzalez, Woods 2002).

$$\mathcal{F}\{f(x, y)\} \neq \mathcal{F}\{m(x, y)\} \mathcal{F}\{n(x, y)\}$$

To circumvent this problem the signal f can be converted to the logarithmic domain where the property of logarithm converts a product to a sum. Thus, transforming f to the logarithmic domain will provide

$$g(x, y) = \log f(x, y) = \log m(x, y) + \log n(x, y)$$

Then the Fourier transform can be applied to f in the logarithmic domain

$$\mathcal{F}\{\log f(x, y)\} = \mathcal{F}\{\log m(x, y)\} + \mathcal{F}\{\log n(x, y)\}$$

where any enhancement or filtering can be performed in the frequency domain and the inverse Fourier transform followed by application of the exponential function (inverse

operation of logarithm) can be applied to convert the signal back to the spatial domain. Homomorphic filtering is used in the field of image processing to enhance (denoised) images that have been corrupted by multiplicative noises such as speckle noise (Chitwong, Thongsila et al. 2005). It has also been used in adaptive image thresholding to separate the illumination and reflectance of the image (Lam 1998).

5.5 Signature Extraction Model

The existing denoising approach uses the wavelet method, and scene details can be included in the extracted signature. Hence the SPN will contain the PRNU and some of the NUA. If the signature of an image can be extracted by separating the PRNU from the scene details which lie in the similar frequency region, this will create a cleaner PRNU that in turn should provide a better SPN. A novel signature extraction model is proposed, which is described in more detail below.

By performing homomorphic filtering of the 2D image, the PRNU can be separated from the scene components. Moreover, if the range of energy of the PRNU can be estimated, it is easier to apply a filter to separate the PRNU from the image content. The Singular Value Decomposition (SVD) can be used to decompose the image into unit ranked images of descending energy. It is difficult to interpret the SVD separated planes in a linear fashion. By using the estimated energy of the PRNU the range of planes where the PRNU is located can be estimated. These ranks (planes) can be grouped together to form the extracted PRNU. Gul & Avcibas have investigated the use of SVD to identify the model of a mobile phone. Their method uses the singular values of SVD to estimate the relative

linear dependency of image rows/columns, which identifies the CFA interpolation algorithm of the camera (Gul, Avcibas 2009).

Let us consider the sensor output model presented in (Chen et al, 2008) for source device identification purposes. They presented a mathematical model relating the clean image and the PRNU together with the other noises that form part of the noisy image as shown in Chapter 3 (section 3.2).

$$I = g^\gamma [(1 + K)Y + \Lambda]^\gamma + \Theta_q$$

where I is the noisy image, g is the colour channel gain, γ is the gamma correction factor, K is the PRNU signature, Y is the incident light intensity, Λ is a combination of the other noise sources including the dark current, shot noise, and read-out noise and Θ_q is the quantization noise. Since in natural images the dominant term is the light intensity Y , it is factored out and only the first two terms of the Taylor expansion

$$(1 + x)^\gamma = 1 + \gamma x + O(x^2) \quad \text{at } x = 0$$

are kept to provide:

$$I = (gY)^\gamma \left[1 + K + \frac{\Lambda}{Y} \right]^\gamma + \Theta_q = (gY)^\gamma \left(1 + \gamma K + \gamma \frac{\Lambda}{Y} \right) + \Theta_q$$

To simplify the notation and avoid using too many symbols, the gamma correction factor is absorbed in the PRNU factor K . $(gY)^\gamma$ is represented by I_0 and $\gamma I_0 \Lambda/Y + \Theta_q$ is represented by η .

Hence the simplified model for a noisy image I can be represented as

$$I = I_0 + I_0 K + \eta$$

where I_0 is the clean image (perfect absorption of light energy by pixels), K is the PRNU and η is the remaining noise, such as shot noise, dark noise and read-out noise, associated with the image. The noisy image is comprised of the clean image to which is added the product of the clean image and PRNU and the random noise components. The model can be represented as

$$I = I_0(1 + K) + \eta$$

The purpose of grouping the PRNU component K using the brackets is to make the multiplicative nature of the PRNU in relation to the clean image (sensor output in the absence of noise).

The image model can be transformed from the spatial domain to the logarithmic domain (Xie, Pierce et al. 2002). Thus, the result is an additive model consisting of the image and PRNU noise, as follows

$$\hat{I} = \hat{I}_0 + \hat{K}$$

where $\hat{I} = \log(I)$, $\hat{I}_0 = \log(I_0)$, $\hat{K} = \log(1+K)$ and the noise components η are suppressed from the equation by averaging many images created by the sensor. The $\log(1+K)$ representation ensures that the PRNU in logarithm is never an undefined value (due to possibility of $\log(0)$) and the value of K is always positive.

The next stage is to determine the energy of the PRNU in an image, which depends on the type of device that produced the image and is a fraction of the total energy of the image. There are two types of sensors that are primarily used in digital cameras, the CCD (Charge-Coupled Device) and the CMOS (Complementary metal–oxide–semiconductor). The exact energy of the PRNU in each image is computationally intractable to calculate,

thus the range of energy where the PRNU lies can be estimated instead based on the type of sensor. The type of scene content might have an impact on the energy of the PRNU, hence the reason of estimating the range where the PRNU is located. The CCD produces less noise but requires more power when compared to the CMOS, hence the reason CMOS is used most often in camera phones where space and battery life are crucial. The energy of the PRNU in a CMOS will be affected by other sources of noises and the approximate energy (variance) of the PRNU can be reduced (El Gamal, Eltoukhy 2005). The PRNU in CCD was measured by calculating the variance (σ^2) of the noise in 100 image sets (Irie, McKinnon et al. 2008). Further details about the PRNU energy experiments can be found in chapter 3 (section 3.3.2). The energy of the PRNU can be estimated to be in the range of 0.01% to 1.5%, depending on the type of image and sensor.

The range has to be increased from the findings in the experiments because these experiments were performed in highly controlled environments where the images used were carefully selected based on the scene contents and lighting of the sensor. On the other hand, the range of energy to be used in the extraction model will be based on real world images of natural images taken from a wide range of cameras, where the lighting and scene contents will vary greatly.

Once the energy range can be estimated, the next stage is to identify the location of the PRNU in the energy spectrum of the image. The logarithmic image model is then decomposed into ranks by using SVD. The energy of the PRNU can be used as a guide to select the relevant ranks using the property of the relationship (as described in section 5.3.2) between the variance of an image and the singular values of the SVD components. The estimation of the PRNU in the signal subspace can be performed empirically by using

a trial and error approach in relation to the total energy (variance) of the image; more details on the approach used will be explained in chapter 6. If the PRNU is being extracted from a colour image, the SVD process should be applied to each colour (RGB) channels separately or the image converted to a grayscale image. Figure 5.2 shows the process stages of the novel signature extraction model.

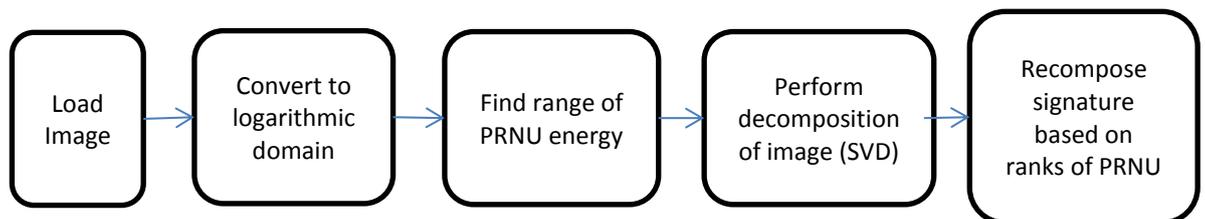


Figure 5.2. Process stages of the novel signature extraction model.

When the PRNU ranks have been estimated, they are grouped together and the logarithmic signature is reconstructed using the selected range of ranks in accordance with their associated energy. The latter should be chosen to contain the PRNU of the camera that created the image. The digital signature can be converted back from the logarithmic domain to the spatial domain, using the exponential function. In this case, the original image cannot be recovered, which signifies that the signature can be stored or transferred securely. The signature obtained can be stored and processed in the logarithmic domain itself. The signatures extracted by this method can be used to create the camera reference signatures of the source device for identification purposes or compared against the reference signatures of other cameras for linkage purposes.

Chapter 6 Experiments for Existing Platform

6.1 Introduction

The experiments were divided into two main categories where firstly the existing platform was assessed and secondly the SVD based extraction method was tested. The experiments on the existing platform, presented in section 6.2, were designed in order to find:

- The best cropping position of an image prior to extracting its digital signature in order to obtain a stronger signature, which will help reduce identification and classification error rates.
- A good training size to sample size ratio of the classification process of images in order to get better clusters that will provide lower classification error rates.

The tests for the existing platform were performed using 5 cameras, and images of varying level of scene details as well as different times of day and night were taken.

The second experimentation process, described in section 6.3, was designed to evaluate the proposed SVD based signature extraction method. This test is performed in two parts to better achieve the objectives of the research:

1. The range of energy of the PRNU has been estimated and the test is designed to locate the PRNU in the SVD separated planes for the images. The most appropriate ranks that contain the PRNU for most types of cameras are located empirically.

2. The range of ranks that was located in test 1 is used to validate the SVD based extraction method. The camera reference signatures from a set of images from 10 mobile phones are calculated and the source identification of unknown images is performed. The results obtained from the SVD based method will be compared against the wavelet based method.

6.2 Experimentation on Existing Platform

For the purpose of the experimentation on the existing platform, 1000 pictures were used from 5 source devices comprising of 4 digital cameras and one camera phone with device contributing 200 images each. The devices were:

- BlackBerry Curve 8310 (camera phone)
- Nikon Coolpix E5200
- Canon Digital IXUS 500
- Olympus C-730 UZ
- Olympus Mju Stylus 1030SW

Since images in real forensic situations come from different cameras with different specifications and settings, a specific size for the photos from the cameras were not set, therefore the sizes ranged from 1600 x 1200 pixels to 3648 x 2736 pixels, which still give a wide margin for cropping of the images. Most of the images were taken at the highest resolution possible for the specific camera. All the images were in JPEG format with a compression quality ranging from about 75 % to 97 % quality. The pictures contain a wide

variety of indoor and outdoor sceneries of urban and rural settings, night and day lighting as well as offices and various buildings and some holiday pictures; overall images that would be as close to natural real world situations as possible were captured.

6.2.1 Image Cropping Position

The cropping position of the images, prior to performing the extraction of the signatures, can be chosen in a systematic way. Different areas of the image will be affected by varying scene details and camera software processing. For example when a picture is taken at night, the flash fires and the centre of the image can be saturated. A trivial solution would involve cropping the images in the dataset at different positions in order to obtain the least scene details or better illumination levels. But the sensor pattern is spatially distributed in the image, which indicates that the cropping position will have to be the same for all the images in the dataset. Images of different resolutions and sizes will have different numbers of pixels in the cropping area, but overall a majority of pixels will be overlapping for most of the cropped images. The experimentation process will have to address the following points:

- Select an area where pixel saturation is less for most images
- Identify the area of images where different scene details varies most
- Most importantly, the cropping area will need to provide a strong signature

The experiment was designed using images from different cameras taken at different times of the day and night as well as different scenes. The experiment was performed on real world natural images. The combination of images from day and night provides

different illumination levels and areas of illumination in the image. Moreover the set of images contains images with indoor and outdoor scenes with some close ups and panoramic pictures. The images were cropped from different positions and the signatures extracted, which was processed by the current *image analyser* platform.

The images were cropped to a size of 512 x 512 pixels, which provide a trade-off between the computational complexity and the accuracy of the signature. Furthermore, the 512x512 size provide reasonable processing time given the limited computing resources and as reported in Hoglund (Hoglund 2009), it takes about 30 hours to calculate the reference noise using 200 images of sizes 3072 x 2304 pixels, whereas 512x512 takes less than one hour. Hence cropping the images from their normal sizes greatly reduces the computational complexity. The magnitude and accuracy of the SPN does not deteriorate greatly when the image is cropped to 512 x512 and preliminary tests performed in Li (Li 2010) indicated the size of the cropping can be as low as 256 x 512.

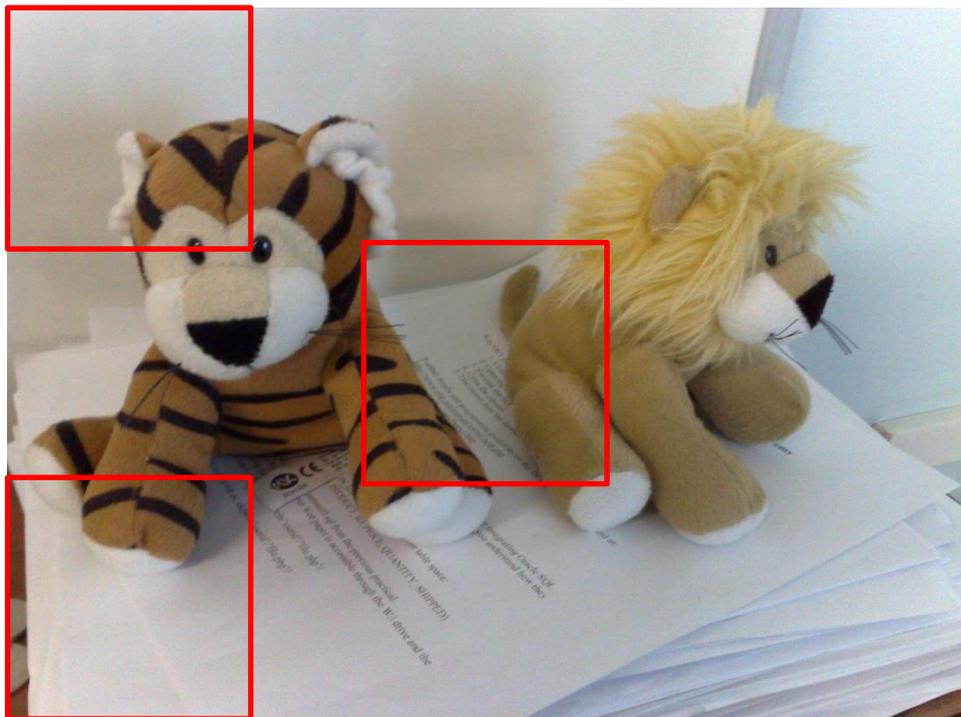


Figure 6.1. Three cropping positions (red squares) on a picture; top-left, centre and lower-left cropping positions.

The images were cropped from three positions; the centre, the upper left corner and lower left corner as shown in Figure 6.1. The preference for choosing these positions was based on the fact that there can be large differences in light levels and scene details between the lower and upper parts of the images as well as the centre. Moreover, the right side and left side of a picture present approximately the same amount of lighting and same average scene details, hence the decision to use only one side of the images to perform the cropping. The average entropy (measure of light intensity and scene contents) of the left side and right side of 1000 images were calculated and the results for both sides were found to be between 7.3 and 7.4, which shows that the light intensity for both sides are the equivalent. The choice of left side cropping over the right side is due to the pixel numbering starting from the left side of the picture. These three chosen positions provide a good distribution of light levels, zooming effect and brightness when the flash is fired. The optimum cropping position will be the one that provides the lowest classification error rates. The results are presented in section 6.3.

6.2.2 Training Size Selection

The setup for performing the cross validation experiments on the existing image analyser platform is described in this section. The cropping was performed on the images and the signatures (SPN) were extracted by using a Discrete Wavelet Transform (DWT) followed by a low pass filter and the SPN were enhanced by using one of the models developed in Li (Li, 2010). The enhancing model chosen was the:

$$n_e(i, j) = \begin{cases} e^{\frac{-0.5n^2(i, j)}{\alpha^2}} & , \text{if } n(i, j) \geq 0 \\ -e^{\frac{-0.5n^2(i, j)}{\alpha^2}} & , \text{otherwise} \end{cases}$$

where $n_e(i,j)$ is the enhanced version of the $n(i,j)$ component and this model has been shown to work for natural images in Li (Li 2010) and the optimal value of α was 7.

To perform the cross validation on the dataset, the folds (explained in Chapter 4, section 4.43) were randomly selected before the training stage of the classifier. The next step was to create the similarity matrix which, being the most computationally intensive stage of the classification process was no longer a trivial process since for each fold the matrix had to be recalculated. This had the prospect of making the cross validation prohibitively expensive computationally and arises due to the change of size of the matrix by changing the size of the fold and the different images present in the training phase for each partition. To overcome this problem, the similarity matrix was created at the start of the cross validation process for all the SPNs. Row and column deletions were performed in the matrix as the population of the partitions changed and the size of the folds were altered. This method increased the time complexity to create the similarity matrix, but it had to be created only once, which decreased the time complexity during the cross validation process.

The sizes of the folds were chosen as 20, 10, 5, and 2, which provided a wide range of prediction errors. The size of the dataset was decreased from 1000 to 500 and to 250 so as to provide a learning curve that will indicate what effect the sample size has on the classifier. The same number of images was chosen from each device for each sample size. Once the value of k is chosen for the cross validation and one partition set aside for testing, the other partitions are combined to provide data for the training phase and classification phase. In Li (Li 2010), the selected sizes of the classifier trainer set were 120 and 300, where there was little variation of classification performance between the two

sizes. The size of the trainer set was altered with different dataset sizes. Moreover, the partitions were stratified, so that each partition contained approximately the same amount of labels.

6.3 Results for Existing Platform

The results for the experiments performed on the existing platform are separated in two sections. The first sets of results are described for the image cropping position and the second sets are for the training size selection results.

6.3.1 Image Cropping Position

The first set of results obtained when performing cross validation (for $k=10$), was giving a higher error rate than expected (table 6.1) when cropping from the centre of the images. The position of the cropping was then changed from the centre to the upper left corner and subsequently to the lower left corner of the image and the experiment repeated each time. All three different positional cropping of the SPNs were used to classify the same set of images to determine the position with the lowest classification errors. The average percentage classification errors for the centre, upper-left corner and lower-left corner are shown in table 6.1.

Cropping Position	Percentage error (%)
Centre	2.2
Upper left	2.5
Lower left	0.81

Table 6.1. The average percentage classification error for three different image cropping positions.

The results show that, for the dataset used for the experiments, the best position to perform cropping of an image for classification is the lower left corner (Soobhany, Leary et al. 2011).

6.3.2 Training Size Selection

Cross validation was performed with sizes of folds of 2, 5, 10, and 20 and it was found that the variance of the error rates increases when the number of folds decreases. Figure 6.1 shows the error and variance for different folds when cross validation was performed on 1000 images.

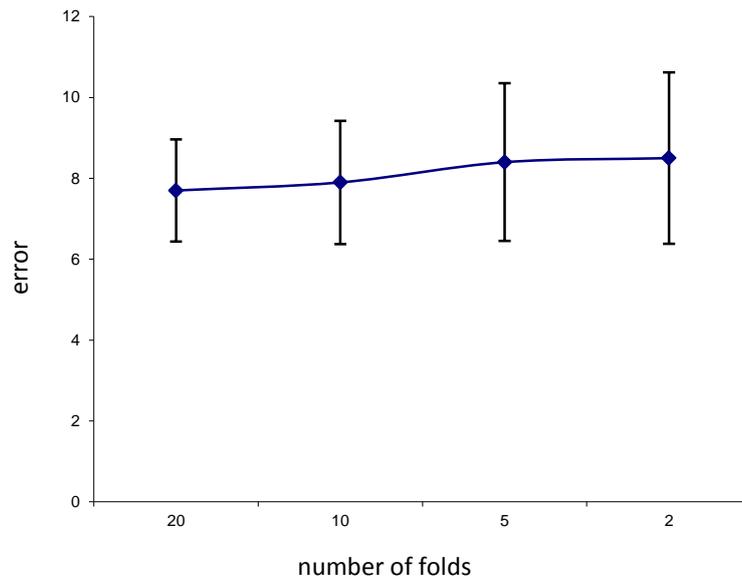


Figure 6.1. Variance and error classification rate with respect to number of folds when cross validation was performed on 1000 images.

Two fold and five fold cross validation produced higher values of variance whereas twenty fold produced the least variance of errors, but the variance of ten and twenty fold are close to each other.

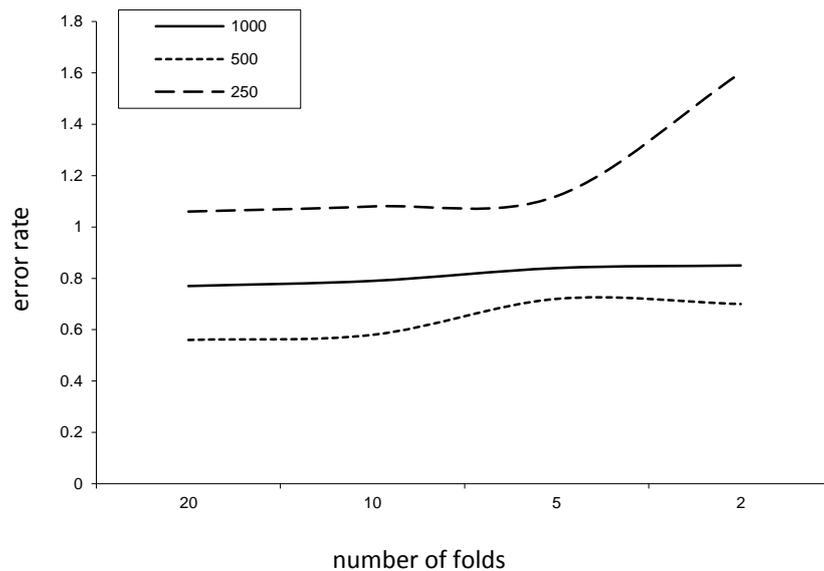


Figure 6.2. Effect of number of folds on percentage error rate for different sample sizes (250, 500, 1000) of images.

In Figure 6.2, the error rates for all three sample sizes (1000, 500, 250) are shown with different sizes of folds. The error rates for $k = 10$ and $k = 20$ are quite similar for all three curves and there is large difference when the number of folds goes towards $k = 2$. It is interesting to note that when the sample size is 500, the error rates are at their lowest values. For a smaller sample size, 250, the error rate increases drastically when the value of k is less than 5.

Sample Size	Classifier Trainer Size			
	900	500	250	125
1000	0.83	0.79	0.74	0.91
500		0.54	0.58	0.88
250			1.84	1.12

Table 6.2. Percentage error rates for different sample sizes and varying classifier trainer size when $k = 10$. (E.g. 0.83 %)

The error rates in Table 6.2 shows that the unsupervised classifier performs better when the size of the training set is less than 50% of the size of the sample space but more than approximately 125 images. When the size of the sample space is 1000 and 250, the error rates obtained were higher when the training set was nearly as large as the size of the sample space.

Furthermore, an additional test was performed to check if the classifier will perform with a small dataset of 50 images comprising of 7, 18, 15 and 10 pictures from the BlackBerry Curve 8310, the Nikon Coolpix E5200, the Canon Digital IXUS 500, and the Olympus Mju Stylus 1030SW respectively. The signatures were clustered in their respective groups according to their camera source together with an outlier group that consisted images that had high level of saturation and darkness. The purpose of this last was to check the suitability of the classifier to cluster small image sets.

6.4 Discussion of Results

The discussion will be structured into two parts, namely image cropping position and the training size selection.

6.4.1 Image Cropping Position

The image cropping from the lower left corner provided the best classification accuracy. The centre of the image usually has a high light intensity (saturation) in the event the flash is activated when the picture is taken at night or in low light, similarly the upper left corner of the image can have high intensity of light since light sources from light bulbs or sunlight, are most often present in that space of the picture. Outdoor pictures taken at night will affect the top of the image too, due to the low light intensity in this area of the image which will adversely affect the multiplicative nature of the PRNU. The saturation of the pixels in these areas of the images usually corrupts the signature of the image.

The lower left corner of the image usually has normal light intensity and less scene details than the centre of the image. Hence cropping the images from the lower left corner provide better results for this image set. These findings are different from Li (2010) because this image dataset had a proportion of images taken during the night and indoors, when the flash was fired as well as in bright sunlight where the top part of the image was saturated. The edges of the images can be providing some further help to enhance the sensor pattern signature due to the edge effect. Therefore the optimum cropping position for images that have been recovered from a crime scene or suspect device should be performed in the lower left corner if the images contain night scenes or highly saturated sunlit images.

6.4.2 Training Size Selection

It can be observed from Figure 6.1 that simply adding more folds beyond $k = 10$ does not necessarily yield drastically better results. The performance of the clustering process will plateau after $k = 10$ folds are used, meaning computing time can be improved by choosing $k = 10$, while keeping error rates low. It was also observed from the results that when the size of the sample space was increased the variance tends to decrease, because the classifier has more images during the training stage of classification which increases the ability to create better clusters. Both outcomes were expected and are in line with previous studies in the area.

When the sample size is 500, the error rates are at their lowest values when it should have been lowest when the sample size is 1000. For a smaller sample size, 250, the error rate increases drastically because insufficient training of the limited number of signatures

are being picked during the training phase of the clustering process. It was stated in (Li, 2010) that the unsupervised classifier needed a large sample of unknown images to work best. The error rates were found to be smaller when the training sample size was less than half the sample space.

When the training sample size is too close to the sample space over training might occur in stage 3 of the unsupervised classifier as depicted in chapter 4 (section 4.3.2), where the intra class and extra class boundaries of the clusters appear to overlap and are thus indistinguishable. The error rates included in table 6.2 are the estimated prediction error rates (P_{est}). They are the maximum likelihood estimate of the true error rate (P_{true}) of the classifier. The relationship between P_{est} and P_{true} with a 95% confidence interval (95% being considered high enough to be accepted by a forensic investigator, obtained during interviews with forensic investigators) for different sample size has been calculated. The average P_{est} including all the tests was below 2%, giving P_{true} between 0.5% to 6% and 1.6% to 3% for sample size of 125 and 1000 respectively.

Chapter 7 Experiments for SVD Extraction Method

7.1 Introduction

This chapter presents the results of the experimentation process beginning with the results on identifying the most appropriate image cropping position. The results of the classification experiments performed on the existing classifier are described followed by the results obtained from the tests carried out on the SVD based signature extraction method for ranks estimation and source device identification. Finally a discussion of the results is presented.

7.2 Experimentation for SVD Based Signature Extraction

The experiments for the SVD based signature extraction were performed in two stages. Firstly the ranks corresponding to the PRNU had to be determined and, secondly, the associated single rank images were combined (linearly) to produce the signature that would facilitate identification of the respective camera phones.

7.2.1 Estimation of PRNU ranks

The range of energy of the PRNU has to be located based on the energy of the images. Given the mathematical property of ($\sigma^2 = \lambda$), SVD was performed on a set of images (cropped to 512x512 pixels) and single rank images with different energy levels were recombined to form the signature. The process was performed empirically by estimating

the ranks where the PRNU might be located and refining the range of energy by using the classifier of the existing image analyser platform.

A set of 15 images were used for the experiments originating from 3 cameras. The cameras were:

- Nikon E5200 Coolpix digital camera
- Nokia N95 camera phone
- Nokia E71 camera phone

Only 5 images were chosen from each camera in order to keep the processing time to a minimum while keeping in mind that the accuracy of the classifier might be affected. Some preliminary tests were performed on the wavelet based extraction method to determine whether the choice of these images and cameras will produce reliable results. It was found that most of these images could be clustered according to their source device.

7.2.2 Source Identification using PRNU ranks

For the purpose of the source device identification experiments, a total of 1000 images were chosen evenly from 10 mobile phones; i.e., each device contributed 100 images. Most of these phones were older models of the respective make and, as such, offer a significantly lower image quality, particularly when compared with images taken from digital cameras. To demonstrate that the proposed extraction method can differentiate between devices of the same make and model, there were five different makes/models, each of which has two phones.

Mobile Phone	Alias	Max Image resolution	Number of images	
			Indoor	Outdoor
nokia_C2_01_A	cam_1	1536x2048	70	30
nokia_C2_01_B	cam_2	1536x2048	77	23
nokia_E72_A	cam_3	2592x1944	42	58
nokia_E72_B	cam_4	2592x1944	15	85
nokia_N95_A	cam_5	2592x1944	45	55
nokia_N95_B	cam_6	2592x1944	34	66
samsung_galaxy_S2_A	cam_7	3264x2448	20	80
samsung_galaxy_S2_B	cam_8	3264x2448	73	27
zte_orange_sanfrancisco_A	cam_9	1536x2048	19	81
zte_orange_sanfrancisco_B	cam_10	1536x2048	14	86

Table 7.1. Names of mobile phones and their aliases used to represent them in the experimentation. The maximum image resolution of each camera and the number of images taken indoor and outdoor are also listed.

Table 7.1 shows that, for each model, the two phones of the same make share the same prefix but are distinguished by different labels ‘A’ and ‘B’ appended as suffix. Most of the phones are products from Nokia, since it is one of the most popular makes in the low-to-medium end of the camera phone market. Camera (or phone) manufacturers use the same colour interpolation algorithms or quantization tables for JPEG compression for several camera (or phone) models, which provide NUA across camera models from the same manufacturer. In addition, the inclusion of different phone models from the same make was also expected to better demonstrate the identification performance of our method. In all cases, the pictures were taken at the highest native resolution of the cameras and stored in the JPEG format, which is the de facto compression format for still images from camera phones. To ensure generality, the pictures were natural images consisting of a mixture of outdoor and indoor scenes, captured during the day and at

night as showed in table 7.1. The images captured at night were all of indoor scenes. Further, given the expectedly different sizes of the captured images, they were all cropped to the same size of 512x512 pixels, consisting of the lower left corner of the image as described in (Soobhany, Leary et al. 2011). The mixture of different types of scenes for the image dataset and cameras with different native resolutions will make identification of source devices more challenging but more realistic. Furthermore, some of the images were taken in low light levels and the flash of the camera has been fired, which usually distorts the synchronisation of the sensor pattern noise. This represents a more realistic scenario where the external/environmental factors vary and out of the control of an investigator.

The SVD based signature extraction procedure as described in Chapter 5 (section 5.5) was applied to these cropped images, allowing the creation of reference signatures for the individual cameras. The experiments are based on real world natural images, which could have been recovered from suspect devices or downloaded from the web. The ability to take uniformly illuminated pictures from the cameras is therefore restricted. Hence the camera reference signatures are created by extracting the signature of 50 images randomly out of the 100 sample, which are natural images with varying scene details, and the extracted signatures are averaged. The creation of the camera reference signature process is described in chapter 4, section 4.3.1. The remaining 50 signatures from each camera dataset were then compared against the ten camera reference signature computed for the individual cameras, similar to the identification of cameras described in chapter 4, section 4.3.1.

To determine whether a specific image was taken by a camera, the cross correlation between the signature of the suspect image and the camera reference signature is calculated. The 2D cross correlation coefficient, r , is given by

$$r = \frac{\sum_m \sum_n (X_{mn} - \bar{X})(Y_{mn} - \bar{Y})}{\sqrt{(\sum_m \sum_n (X_{mn} - \bar{X})^2) (\sum_m \sum_n (Y_{mn} - \bar{Y})^2)}}$$

where X and Y are 2D matrices of size $m \times n$, \bar{X} is the mean of X and \bar{Y} is the mean of Y .

It is assumed that there was no geometrical transformation, such as rotation or rescaling, applied to the images and the tests were mainly to check the effectiveness of the proposed extraction method. It has been shown in previous research in source device identification that an acceptance threshold of 0.01 for the cross correlation coefficient was reasonable (Li, 2010). Hence, the same threshold will be used to determine a suspect image originates from the same camera as the camera reference signature.

7.3 Results for SVD Based Signature Extraction

The required energy range for the extraction of the PRNU was found to vary slightly (across the range described in Chapter 3 (section 3.3.2) depending on the amount of scene detail in the image. Generally, images with high scene detail content will have their energy spread out more widely across the top ranks after performing SVD, whereas less 'busy' images had most of their energy concentrated among the first couple of high energy ranks.

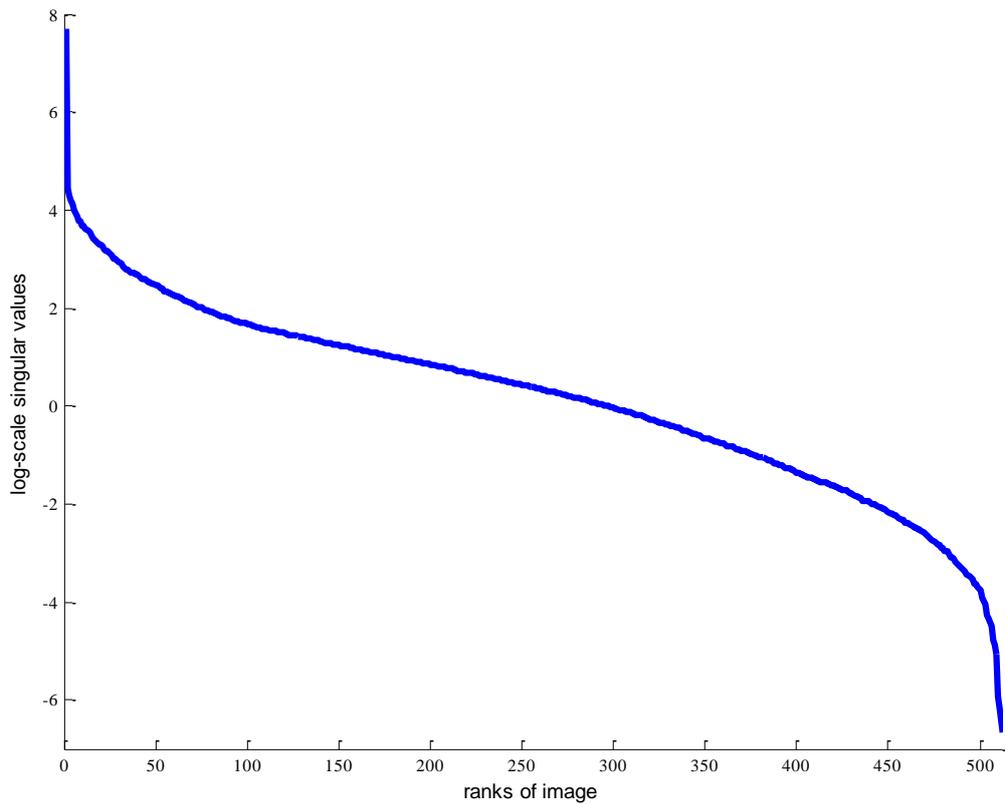


Figure 7.1. Plot of *log*-scaled singular values of a natural image with 512 ranks after SVD decomposition.

Figure 7.1 shows the plot of the *log*-scaled singular values of a natural image with 512 ranks. There is a sharp drop after the first rank (from 7.6 to 4.4). For a blue sky image, the drop is from about 7.8 to 0.4. The result indicates that most of the scene detail energy is concentrated in the first (topmost) rank. The difference between the blue sky image and the natural image further suggest that most of the smooth scene details will be concentrated among the first few higher ranks and the stronger details is located further down the ranks.

7.3.1 Estimation of PRNU Ranks

Preliminary results obtained from experiments to choose the ranks based on PRNU energy suggested that the location of the PRNU is widely spread among ranks in images from different types of cameras. Table 7.2 shows the clustering percentage error based on different selection of ranks for images of 512 ranks from two mobile phones and one digital camera (Nokia E71, Nokia N95, Nikon E5200 Coolpix).

Ranks	Average Energy	Percentage error (%) for each camera		
		Nokia E71 (C1)	Nokia n95 (C2)	Nikon E5200 (C3)
4-20	0.2128	40	60	60
4-100	0.1819	80	60	40
4-200	0.1770	60	80	60
25-150	0.0366	80	80	20
30-150	0.0315	40	60	60
35-150	0.0281	20	20	0
40-150	0.0252	20	20	20
45-150	0.0228	0	20	20
50-100	0.0242	40	40	20
50-115	0.0230	40	40	20
50-125	0.0224	20	20	20
50-130	0.0221	0	40	20
50-135	0.0219	0	0	20
50-150	0.0212	0	0	20
50-200	0.0197	0	0	20
50-250	0.0187	0	0	20
75-150	0.0153	20	60	20
100-150	0.0123	40	40	0
100-200	0.0108	20	80	100
150-200	0.0077	60	80	20

Table 7.2. Clustering results for 15 images from 3 cameras (Nokia E71, Nokia N95, Nikon E5200 Coolpix) based on different rank combinations used to create the signature. The cells shaded with pink are the ranks that provided the best c results.

The three cameras produced three clusters where Nokia E71, Nokia N95 and Nikon E5200 Coolpix are the clusters C1, C2 and C3 respectively. When the ranks combinations contain high energy ranks, for example ranks 4 – 100, it can be observed that the results produce

mixed clusters with high percentage error. The energy of the starting ranks in the signature was decreased each time by 5 ranks and the clustering results can be seen to improve. The best clustering results were observed consistently when the starting energy ranks started around 50. The ranks combinations of 50 up to 250 provided the best results with lowest percentage error and the ranks combination were narrowed down to use the ranks between 50 and 150 for extracting the signature of the images. The average energy of the combined ranks can be found to agree in broad terms with the range of energy in Chapter 3 (section 3.3.2).

7.3.2 Source Identification of Camera Phones

The Figure 7.2 shows the result of using blue sky images to create the camera reference signature for *cam_5*. There were 100 images used where each camera contributed 10 images each and images 51 to 60 originate from *cam_5*. The red line indicates the acceptance threshold of the correlation coefficient as discussed in section 6.3.2.

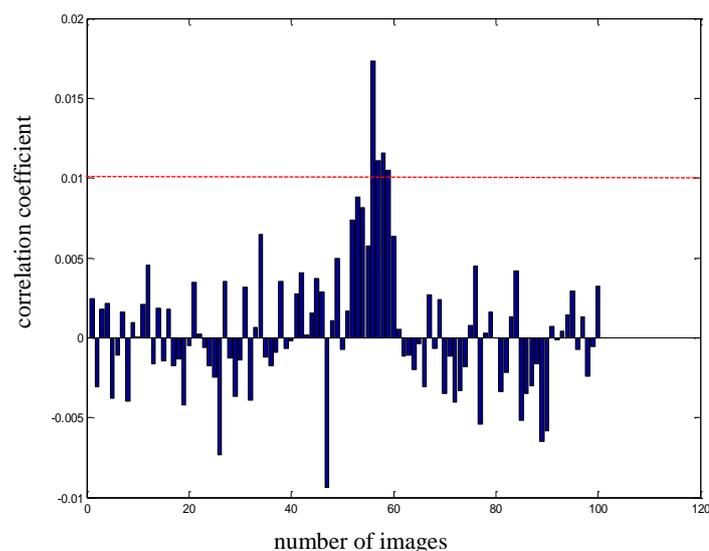


Figure 7.2. Nokia_N95_A (*cam_5*) camera reference signature created from blue sky images and correlation with 100 images. Images 51 to 60 originate from this camera and other 90 images from the 9 other cameras. Red line is the acceptance threshold. 148

The same camera phone was used as source device to create its camera reference signature from natural scene images. Figure 7.3 shows the plot of the correlation was performed between the 100 suspect images and the camera reference signature. The red line indicates the acceptance threshold of the correlation coefficient, similar to Figure 7.2. The rest of the experiments were performed with the camera reference signature created using natural scene images.

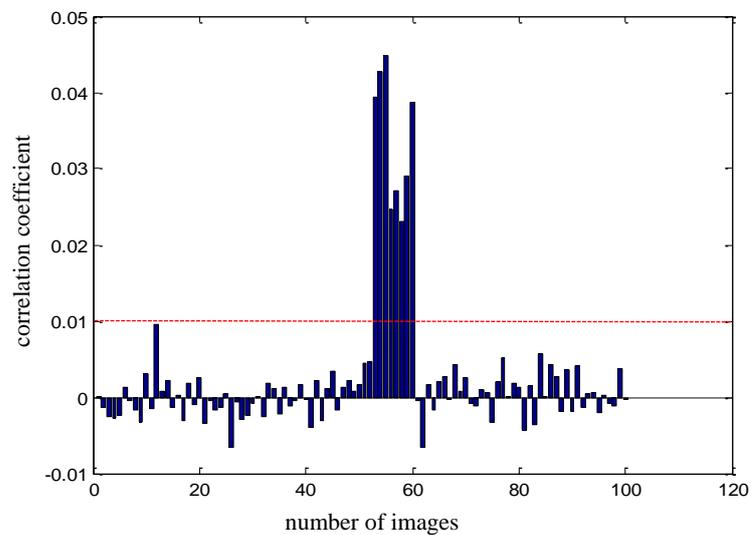


Figure 7.3. Nokia_N95_A (*cam_5*) camera reference signature created from natural images and correlation with 100 images. Images 51 to 60 originate from this camera and other 90 images from the 9 other cameras. Red line is the acceptance threshold.

The graph in Figure 7.4 shows the result of comparing the camera signature of the ZTE Orange San-Francisco A (*cam_9*) with 100 signatures consisting of ten images from each camera. The images 81 to 90 come from this camera, images 91 to 100 from ZTE Orange San-Francisco B (*cam_10*) and the rest of the images originate from the other 9 cameras. The red line indicates the acceptance threshold of the correlation coefficient. The *cam_9*

provided the best results in terms of the separation in values between the true positives of identification and true negatives.

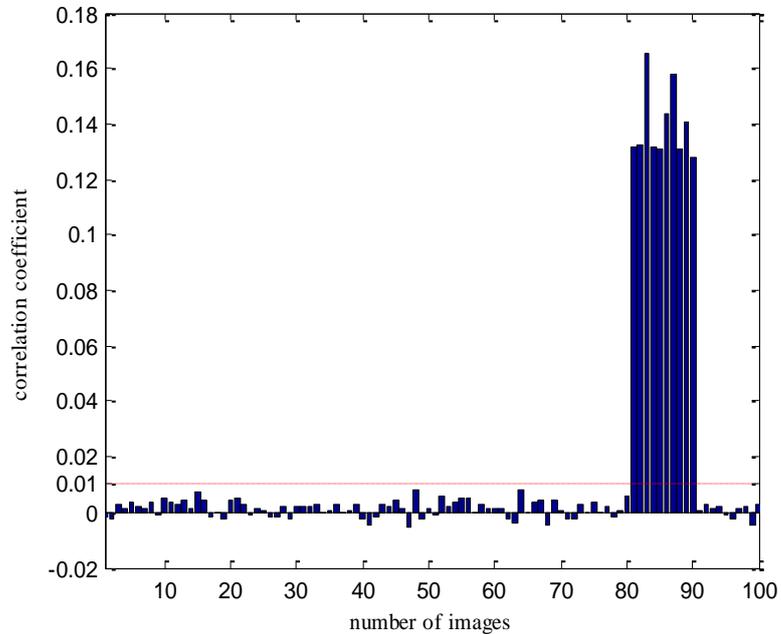


Figure 7.4. zte_orange_sanfrancisco_A (*cam_9*) camera reference signature and correlation with 100 images. Images 81 to 90 come from this camera, images 91 to 100 from *cam_10* and rest of images from the other 9 cameras. Red line is the acceptance threshold.

The graph in Figure 7.5 shows the result of comparing the camera reference signature of the ZTE Orange San-Francisco B (*cam_10*) with the same test signatures as the one in Figure 7.4. On this plot the overall correlation values can be seen to be lower than the *cam_9* results above, but the positive identification of images from *cam_10* is still above the threshold.

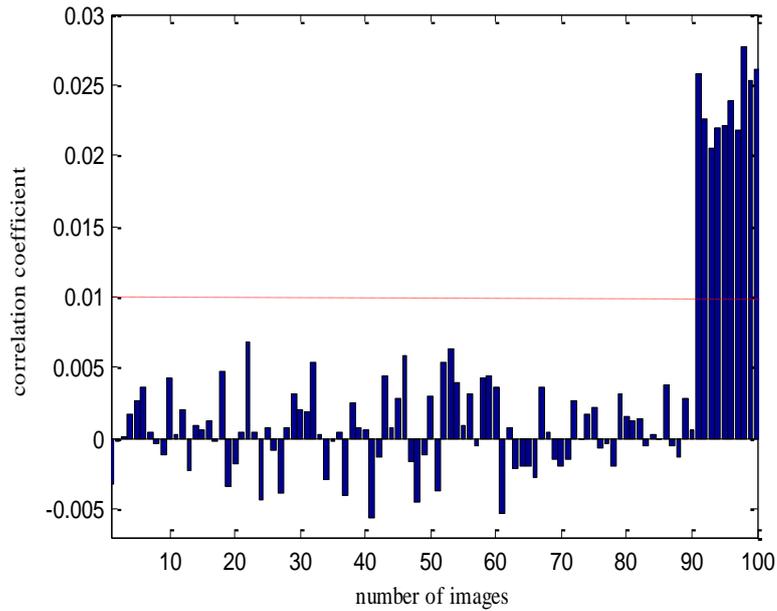


Figure 7.5. zte_orange_sanfrancisco_B (*cam_10*) camera reference signature and correlation with 100 images. Images 91 to 100 come from this camera, images 81 to 90 from *cam_9* and rest of images from the other 9 cameras. Red line is the acceptance threshold.

Table C.1 and C.2 (in Appendix C) shows the mean and standard deviation of correlation coefficient values, respectively. The ten camera reference signatures, extracted using SVD method and wavelet method, are compared with the test images from the same camera and rest of the cameras. The same images were used to extract the sets of signatures using the wavelet and the SVD extraction method and the images were cropped from the same position. The wavelet signatures were enhanced.

7.4 Discussion

The SVD based signature extraction method tends to work better with natural images when generating the camera reference signature. Figure 7.2 illustrates that the majority of false negatives were attributed to *cam_5*, suggesting that the extracted PRNU from

those 'blue sky' images was relatively weak. On the other hand, Figure 7.3 shows no false negatives whilst a near false positive has been identified (image 12) having a computed correlation value very close to the acceptance threshold. The discussion will be described in the following two sections, the estimation of the PRNU ranks and the source identification of camera phones.

7.4.1 Estimation of PRNU Ranks

When the high energy ranks (e.g. rank 4) were used to extract the signature, the classifier produced mixed clusters. This indicates that the PRNU was not the only component that was extracted, if at all, and the signature was being highly corrupted by other scene components. Another source of pollution in the signature at high energy ranks may be the NUA (non-unique artefacts) that are not suppressed in these high energy ranks. The signatures extracted using ranks from 50 to 150 were found to produce clusters, that grouped signatures to their respective source devices, which indicate the PRNU lies in that region and can be effectively extracted to form part of the signature.

The average energy of the majority of the rank combinations was found to be within the range of PRNU, around 0.01% to 1.5%. The results are in the range similar to the findings (range) in the literature of sensor noise model (Chapter 3, section 3.3.2). However, only the rank combinations of around 50 to 150 were found to produce results that grouped the images according to their source devices. Therefore it can be concluded that the PRNU lies in a narrow range of ranks.

7.4.2 Source Identification of Camera Phones

The experiments for the identification of the camera phones were carried out to see if the results produced will show that the PRNU can be reliably extracted by estimating its energy and applying the SVD based extraction method.

7.4.2.1 Correlation Coefficient Identification

It can be seen in Figure 7.4 that the correlation coefficient for the images from *cam_9* is significantly higher than the acceptance threshold, whereas the correlation coefficient for the other cameras is close to zero confirming the expectedly uncorrelated relationship. Furthermore, the identification results between images from *cam_10* and the reference signature of *cam_9* is similar to the results of other cameras, clearly demonstrating that the SVD based method can differentiate between two cameras of the same model. Figure 7.5 corroborates the results from *cam_9* when using the camera reference signature of *cam_10*, although the correlation values for test signatures from *cam_10* are lower as shown in the Figure 7.4. The lower correlation values were largely due to the quality of the images that were selected arbitrarily to create the reference signature of *cam_10*; in particular, there were more saturated pixels present in these pictures.

Appendix B shows the plot of the correlation values versus the image numbers for all the 10 cameras when their camera reference signatures are matched against the 100 image sample set. The odd numbered figures (e.g. Figure B.1, Figure B.3, etc) show the results for the SVD based method and the even numbered figures (e.g. Figure B.2, Figure B.4, etc) show the results for the wavelet based method. An inspection of the graphs for the

wavelet method shows that the overall correlation values have increased for the matching and non-matching images and there is less separation between them.

Cam_1 (Figure B.1 and B.2) does not provide proper correlation values for both methods. For cam_2 (Figure B.3 and B.4), more images can be identified for the wavelet method, the SVD method could not differentiate between cam_1 and cam_2. Again for cam_3 (Figure B.5 and B.6), only about half the images could be identified by both methods with some false positives in both methods. Cam_4 (Figure B.7 and B.8) provides a good signature in both extraction methods, but the non-matched cases are close to zero in SVD method, indicating a better extraction of the PRNU. Cam_5 (Figure B.9 and B.10) do not provide good reference signature and closer inspection of the images from that camera revealed a high percentage of saturated ones.

Cam_6 (Figure B.11 and B.12) provided a better identification for SVD method, but still some false positives from the other two Nokia phones in the dataset, the wavelet method did not produce any coherent result. The SVD method managed to distinguish between the two Nokia N95 (cam_5 and cam_6). Cam_7 (Figure B.13 and B.14) provided better identification results for SVD method albeit with lower correlation values. The wavelet method could not differentiate between cam_7 and cam_8, which are of the same model.

The SVD method identified all the images from cam_8 (Figure B.15 and B.16), whereas the wavelet method did not manage to identify all the signatures and some false positive were obtained from cam_7. For cam_9 (Figure B.17 and B.18) obtained the best camera signature with the SVD method with a good separation between matched and unmatched case. The wavelet method obtained good identification results with cam_9 too, but with narrower separation between matched and unmatched cases. Cam_10 (Figure B.19 and

B.20) provided a good identification rate for the SVD method, albeit with lower correlation values than those obtained from cam_9. The wavelet method did not manage to identify all the 10 images from cam_10 and also some false positives were included.

The mean values (table C.1) of the correlation coefficient values are about the same level for both the wavelet method and SVD method, with the wavelet method with marginally higher mean values, when the camera reference signatures are compared with signatures from the same cameras. The main difference in the mean values between the two methods occurs when the camera reference signatures are matched against images from other cameras.

The table C.1 shows that the mean values for the SVD method is close to zero and there is a wide margin of difference (order of 100 times) between the matched case and unmatched case means. For the wavelet method, the mean values for unmatched cases are between 0.03 and 0.05, which indicate that the overall non match correlation values have increased substantially. The difference between the matched and unmatched case is small (order less than 10 times). The reason for this rise in values might be due to blurring of the SPN when the enhancer model was applied to the signature. The acceptance threshold can be altered to reflect this change, but the difference between true positive and true negative is still very small.

The standard deviation from the mean (Table C.2) is lower for the wavelet method for the unmatched cases, but the mean values are much closer to the mean for the matched case. The deviations for matched images, for both methods, are similar for cameras 2, 4, 6 and 7. The deviations for cameras 3, 8, 9 and 10 are higher for the wavelet method and given the fact that the mean values for matched wavelet cases are close to the

unmatched cases, it shows that it will be problematic to identify an appropriate acceptance threshold level for the wavelet method.

7.4.2.2 Q-function and p-values

The Q-function is the tail probability of the standard normal distribution. It is the probability that a normal (Gaussian) random variable will obtain a value larger than x standard deviations above the mean. Peak to Correlation Energy (PCE) ratio is the squared correlation divided by sample variance of the circular cross-correlations and it is a suitable detection statistic (Goljan 2009). Using the PCE, the p-value (Q-function) can be obtained to show that the result can be rejected (i.e. image does not belong to the same camera reference signature). The results for the 10 camera phones are displayed in Appendix D (Figures D.1 to D.10). The images that originate from the camera that belong to the correct camera reference signature produce very low p-values close to zero (approximately order of 10^{-15}). Hence the probability that the image does not originate from this camera is very close to zero. On the other hand, if the image comes from a different camera than the camera reference signature the p-value (probability) will be higher which indicates that the image comes from a different camera.

The Figures (D.1 to D.10) in Appendix D, shows that when the images originate from the camera that was used to create the camera reference signature, the p-values (Q-function) are close to zero. Cam_1 and cam_2 (Figure D.1 and D.2) show that the p-values are zero for their respective images in the sequence, but both graphs cam_8 also provide lower p-values than the other cameras. The signatures of the images from cam_8 interfere with

the reference signature of both Nokia_C2_01 camera phones, but the latter phones can be differentiated from cam_8.

The p-values for cam_3 and cam_4 (Figure D.3 and D.4) are nearly zero each of the respective images from these two cameras. For cam_5 (Figure D.5), there are a couple of false negatives (images 41 to 43) due to the quality of the signature of these images. Cam_6 (Figure D.6) did not provide any identification of the images that came from the same camera (cam_6). The reference signature for that camera does not allow the signatures of images 51 to 60 to be identified, which indicates that the reference signature is too weak to use Q-function.

The p-values for cam_7 and cam_8 (Figure D.7 and D.8) are nearly zero for the images from the corresponding source camera. Cam_9 (Figure D.9) got the best results for the p-values, which is consistent with the results obtained when using correlation coefficients (Figure B.15). This camera has the cleanest reference signature and the images that were used for the creation of its reference signature were similar quality (brightness and scene details) to the other cameras. The cam_10 (Figure D.10) has three false negatives.

The discussion has focused on explaining the results obtained by using the SVD based extraction method, which allows the identification of camera phones. The Q-function can be used as a statistical feature to accept or reject an image (signature) when correlated with a camera reference signature. Some of the phones that were not identified were found to contain a high proportion of saturated images and the digital zoom (which affects the synchronisation of the PRNU) had been activated for these images sets too.

Chapter 8 Summary and Conclusion

This chapter provides a summary of the research work presented and concludes the thesis. Some recommendations for further work are also provided.

8.1 Summary

This thesis focused on the source identification of digital images for forensic analysis. A case study was performed on an existing image analyser platform, used by Forensic Pathways Ltd, which consists of an identifier and an image classifier. The identifier is used when the source device that created the image is present and the image classifier is used when a large number of images are collected but no source device is present. The existing platform uses the wavelet signature extraction method together with an enhancer, which attenuates high magnitude details in the signatures. The current method does not perform well for the identification of camera phones. The performance of the existing platform has been assessed in order to find the best cropping position for images before signature extraction and to find the optimum training size of the image classifier for varying image sample size.

The wavelet base method extracts the high frequency details from the image, which comprises both the PRNU and scene details. Hence the signature contains unwanted scene details. The denoising algorithm is empirical and does not consider the characteristics of the PRNU when extracting the signature. The choice of the detail levels for the wavelet decomposition is performed empirically. The enhancement procedures

used to minimise the scene details is heuristic and affects the quality of the weak PRNU component. In the case of camera phones the PRNU component is already attenuated by the compression process and the enhancement procedure of the digital signature can result in damaging the PRNU component further.

To address these issues two characteristics of the PRNU were investigated, namely the multiplicative nature of the PRNU and its energy range. The latter was estimated based on previous research performed on eliciting the sensor noise model of CCD and CMOS sensors. The concept of signal decomposition, where a spectrum is divided into signal subspaces in order that those with more energy content will be given higher priority for further processing. Homomorphic filtering is applied to the multiplicative PRNU noise in order to facilitate the separation of the image and the PRNU.

The SVD based signature extraction method was implemented by applying SVD to images in the logarithmic domain in order to obtain the separated planes of the images as unit ranked images. It is complex to interpret the SVD separated planes in a linear way. This is where the property of the image variance being proportional to the singular values obtained has been used to estimate the location of the PRNU based on its energy. The estimated single ranked planes corresponding to the PRNU energy are extracted to form the signature.

8.2 Conclusion

The outcome of the case study allowed the detection of some limitations pertaining to the existing platform. The image analyser was improved so that it accepts images with varying sizes and resolutions; the cropping process was also automated. Furthermore the platform can identify images that have been rotated.

The cropping position for images, before extracting the SPN, was ascertained in an ad hoc method on the existing platform. An experiment was designed to locate several positions on the image that are most affected by light levels and scene details. A selection of mixed images from different scenes was chosen from 5 cameras. The images were cropped from three positions, the top left, the centre and the lower left and the extracted signatures were processed in the image classifier of the existing platform. The results obtained suggest, when images in a sample set are taken at night and with high illumination levels, the best cropping position is the lower section of the image. Both the right and left side of the image can be chosen.

The training size for the image classifier was varied to assess the impact it had on the unsupervised classification error rate. Cross validation technique, which is mainly used for supervised classification, was used in the context of unsupervised classification. The results of the study showed that the use of cross validation can help ascertain the validity of the clusters formed. Furthermore the best training size for different sample space was found to be more than 50% of that sample space.

The main limitation of the existing platform was the identification of highly compressed images, which occur mainly from mobile phones. The characteristics of the PRNU were studied, such as its energy range and its multiplicative nature, and a novel PRNU

extraction method using SVD was introduced. The energy range of the PRNU is estimated and homomorphic filtering is performed on the image. SVD is then applied to decompose the image into unit ranked images of descending energy contents and the PRNU is located and extracted. The identification results of the test performed on 10 camera phones showed that the method can differentiate between two cameras of the same make and model, suggesting that the signature is highly related to the SPN of the camera.

It was also showed that the PRNU signature could be extracted relatively straightforwardly with most real-world/natural images other than the representative uniformly lit (e.g. blue sky) pictures, which are difficult, if not impractical, to obtain from recovered evidence in most forensic investigations. The performance of the SVD-based method was also compared to the wavelet method used by the existing platform and the former did produce encouraging results by being able to distinguish between camera phones of the same model. The extraction of the signature in the SVD based method is backed by theory and performed in a qualitative way, that is, by finding the energy range where the PRNU lies.

The main contributions that were made in this research project are as follows:

- Determine the best cropping position of an image for SPN signature extraction that will result in better identification rate for images of varying scene details and illumination levels.
- Identify the optimum training size for the unsupervised classifier used by the existing platform.

- The characteristics of the PRNU have been investigated, where the energy range of the PRNU in images can be estimated.
- Design of a novel SVD based extraction method for PRNU signatures where homomorphic filtering is applied to the image and signal decomposition is used to separate the image into unit ranked images of descending energy in order to facilitate the localisation of the PRNU. Once the ranks where the PRNU is located have been identified, they can be combined to form the signature of the image.

8.3 Further Work

Based on the study in this thesis, there are several improvements that can be performed on the proposed method.

Once the energy of the PRNU is estimated, the ranks of the PRNU noise are selected manually at present in the proposed method. Further work can be performed on applying algorithms for automating the rank selection process by making use of the image characteristics. Broken stick algorithm or scree test can be used to automate the ranks selection. Scree test is a graphical technique that can be used to retain the correct number of factors in a factor analysis. In the broken stick model the apportioned resource is the total variance of the data set (the variance is considered a resource shared among the principal components). A component is retained if its associated eigenvalue is larger than the value given by the broken stick distribution (Cangelosi, Goriely 2007).

The scene details have an impact, as shown in the results section, on the energy distribution in images. The identification results could be improved by dividing the

suspect image into small blocks depending on scene contents within the image and applying SVD to the small blocks. The computational complexity might increase but the advantage of this method will be the localisation of areas of high or low scene details and enhancing the PRNU estimation.

The proposed method has been applied to identify camera phones, where it was shown to enable the identification of some camera phones with a high accuracy. The SVD based extraction method could be applied to the identification of videos and online images. The signature extraction in videos will follow the same principles as in image extraction since the videos consist of frames, which are inherently stationary images. A compressed video sequence will contain some frames with the complete stationary image known i-frames and other intermediate frames between 2 i-frames known as p-frames. The latter frames store only the difference between two consecutive i-frames, which might hinder identification rate of the video. A large-scale experimentation from an online image database, such as Flickr, can be applied to the SVD method for identification of a large number of images.

References

- ALLES, E.J., GERADTS, Z.J.M.H. and VEENMAN, C.J., 2008. Source Camera Identification for Low Resolution Heavily Compressed Images, *International Conference on Computational Sciences and Its Applications, 2008. ICCSA '08*. 2008, pp. 557-567.
- ANDREWS, H. and PATTERSON, C., 1976. Singular value decompositions and digital image processing. *IEEE Transactions on Acoustics, Speech and Signal Processing*, **24**(1), pp. 26-53.
- ASIMOV, D., 1985. The grand tour. *SIAM Journal of Scientific and Statistical Computing*, **6**(1), pp. 128-143.
- ASIMOV, D. and BUJA, A., 1994. The grand tour via geodesic interpolation of 2 frames, *Visual Data Exploration and Analysis, Symposium on Electronic Imaging Science and Technology, IS&T/SPIE (Soc. for Imaging Sci. and Technology/Internat. Soc. for Optical Engineering)* 1994.
- BARTLETT, K., 5 March 2012, 2012-last update, Nokia 808 PureView: Carl Zeiss science of making the perfect lens [Homepage of Conversations by Nokia], [Online]. Available: <http://conversations.nokia.com/2012/03/05/nokia-808-pureview-carl-zeiss-science-of-making-the-perfect-lens/> [June/15, 2012].
- BAYRAM, S., SENCAR, H., MEMON, N. and AVCIBAS, I., 2005. Source camera identification based on CFA interpolation, *IEEE International Conference on Image Processing, 2005. ICIP 2005*. pp. III-69-72.
- BERGER, T., 1971. *Rate Distortion Theory: A Mathematical Basis for Data Compression*. NJ: Prentice-Hall.
- BLOY, G.J., 2008. Blind Camera Fingerprinting and Image Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **30**(3).
- BUJA, A. and ASIMOV, D., 1986. Grand tour methods: an outline, *Computer Science and Statistics: Proceedings of the 17th Symposium on the Interface* 1986, pp. 63-67.
- BUJA, A., COOK, D. and SWAYNE, D.F., 1996. Interactive high-dimensional data visualization. *Journal of Computational and Graphical Statistics*, pp. 78-99.
- CANGELOSI, R. and GORIELY, A., 2007. Component retention in principal component analysis with application to cDNA microarray data. *Biology Direct*, **2**(2),.
- CELIKTUTAN, O.S. and AVCIBAS, B., 2008. Blind identification of source cell-phone model. *IEEE Transactions on Information Forensics and Security*, **3**(3), pp. 553-566.

CHEN, M., FRIDRICH, J., GOLJAN, M. and LUKAS, J., 2008. Determining image origin and integrity using sensor noise. *IEEE Transactions on information forensics and security*, **3**(1), pp. 74-90.

CHEN, M., FRIDRICH, J. and GOLJAN, M., 2007. Digital imaging sensor identification (further study), *In Security, Steganography, and Watermarking of Multimedia Contents IX. Edited by Delp, Edward J., III; Wong, Ping Wah. Proceedings of the SPIE, Volume 6505* 2007.

CHITWONG, S., THONGSILA, A., INTAJAG, S., NILAS, P. and CHEEVASUVIT, F., 2005. Speckle noise reduction using adaptive singular value decomposition in logarithmic domain, *ASPRS 2005 Annual Conference, 7th - 11th March 2005, Baltimore, Maryland*.

COOK, D., BUJA, A., CABRERA, J. and HURLEY, C., 1995. Grand tour and projection pursuit. *Journal of Computational and Graphical Statistics*, , pp. 155-172.

DAVIES, A. and FENNESSY, P., 2001. *Digital Imaging for Photographers*. 4th edn. Focal Press.

EFRON, B. and TIBSHIRANI, R.J., 1997. *An introduction to the bootstrap*. Chapman & Hall.

EL GAMAL, A. and ELTOUKHY, H., 2005. CMOS image sensors. *IEEE Circuits and Devices Magazine*, **21**(3), pp. 6-20.

EMERY, R. and LAM, K.P., 2008. *Visual Detectives for Structures in High Dimensional Space*. Presentation edn. 3ME/EPSRC Sandpit.

FARID, H., 2008. *Digital ballistics from jpeg quantization: A followup study*. TR2008-638. Dartmouth College: Department of Computer Science.

FARID, H., 2006. *Digital image ballistics from JPEG quantization*. TR2006-583. Dept. Comput. Sci., Dartmouth College, Hanover, NH.

FORENSIC PATHWAYS LIMITED, 2009. *Methods for identifying image devices and classifying images acquired by unknown imaging devices*. Patent No. GB2467767, UK.

FRICKER, P., RAINER, S.A., STEWART WALKER, A. and DIEGO, S., 1999. Digital Photogrammetric Cameras: Possibilities and Problems.

FRIDRICH, J., 2009. Digital Image Forensic Using Sensor Noise. *IEEE Signal Processing Magazine*, **26**(2), pp. 26-37.

FRIEDMAN, J.H. and STUETZLE, W., 1981. Projection pursuit regression. *Journal of the American statistical Association*, , pp. 817-823.

GERADTS, Z. and GLOE, T., 2009. *Identification of images*. D6.8b. Future of Identity in the Information Society (FIDIS).

- GLOE, T., KIRCHNER, M., WINKLER, A. and BÖHME, R., 2007. Can we trust digital image forensics? *Proceedings of the 15th international conference on Multimedia 2007*, ACM, pp. 78-86.
- GOLJAN, M., 2009. Digital Camera Identification from Images – Estimating False Acceptance Probability, *Digital Watermarking, Lecture Notes in Computer Science*, **5450**, pp. 454-468.
- GONZALEZ, R.C. and WOODS, R.E., 2002. *Digital Image Processing*. Second edn. New Jersey: Prentice-Hall.
- GONZALEZ, R.C. and WOODS, R.E., 2002. Homomorphic Filtering. *Digital Image Processing*. Second edn. New Jersey: Prentice-Hall, pp. 191.
- GREENGARD, S., 2012. Digitally possessed. *Communications of the ACM*, **55**(5), pp. 14-16.
- GROTTA, S.W. and GROTTA, D., 2012. Not all pixels are created equal [Tools & Toys]. *Spectrum, IEEE*, **49**(5), pp. 22-24.
- GUL, G. and AVCIBAS, I., 2009. Source cell phone camera identification based on singular value decomposition, *First IEEE International Workshop on Information Forensics and Security, 2009. WIFS 2009*. 2009, pp. 171-175.s
- GUNTURK, B.K., GLOTZBACH, J., ALTUNBASAK, Y., SCHAFER, R.W. and MERSEREAU, R.M., 2005. Demosaicking (sic): Color filter array interpolation in single chip digital cameras. *IEEE Signal Processing Magazine*, **22**(1), pp. 44–54.
- HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J., 2009. *The elements of statistical learning: data mining, inference, and prediction*. 2nd edn. Springer.
- HOGLUND, T., 2009. Digital Camera Identification: A Brief Test of a Method Based on the Sensor Noise. *Journal of Forensic Identification*, **59**(5), pp. 27.
- HOTELLING, H., 1933. An analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, **24**(6), pp. 417-441.
- HSU, Y.F. and CHANG, S.F., 2007. Image splicing detection using camera response function consistency and automatic segmentation, *IEEE International Conference on Multimedia and Expo 2007*, pp. 28-31.
- HUBER, P.J., 1985. Projection pursuit. *The annals of Statistics*, , pp. 435-475.
- HURLEY, C. and BUJA, A., 1990. Analyzing High-Dimensional Data with Motion Graphics. *SIAM Journal on Scientific and Statistical Computing*, **11**(6), pp. 1193-1211.
- HYTTI, H.T., 2006. Characterization of digital image noise properties based on RAW data. *Proceeding SPIE-IS&T Electronic Imaging, Image Quality and System Performance III*, **6059**, pp. 60590A.

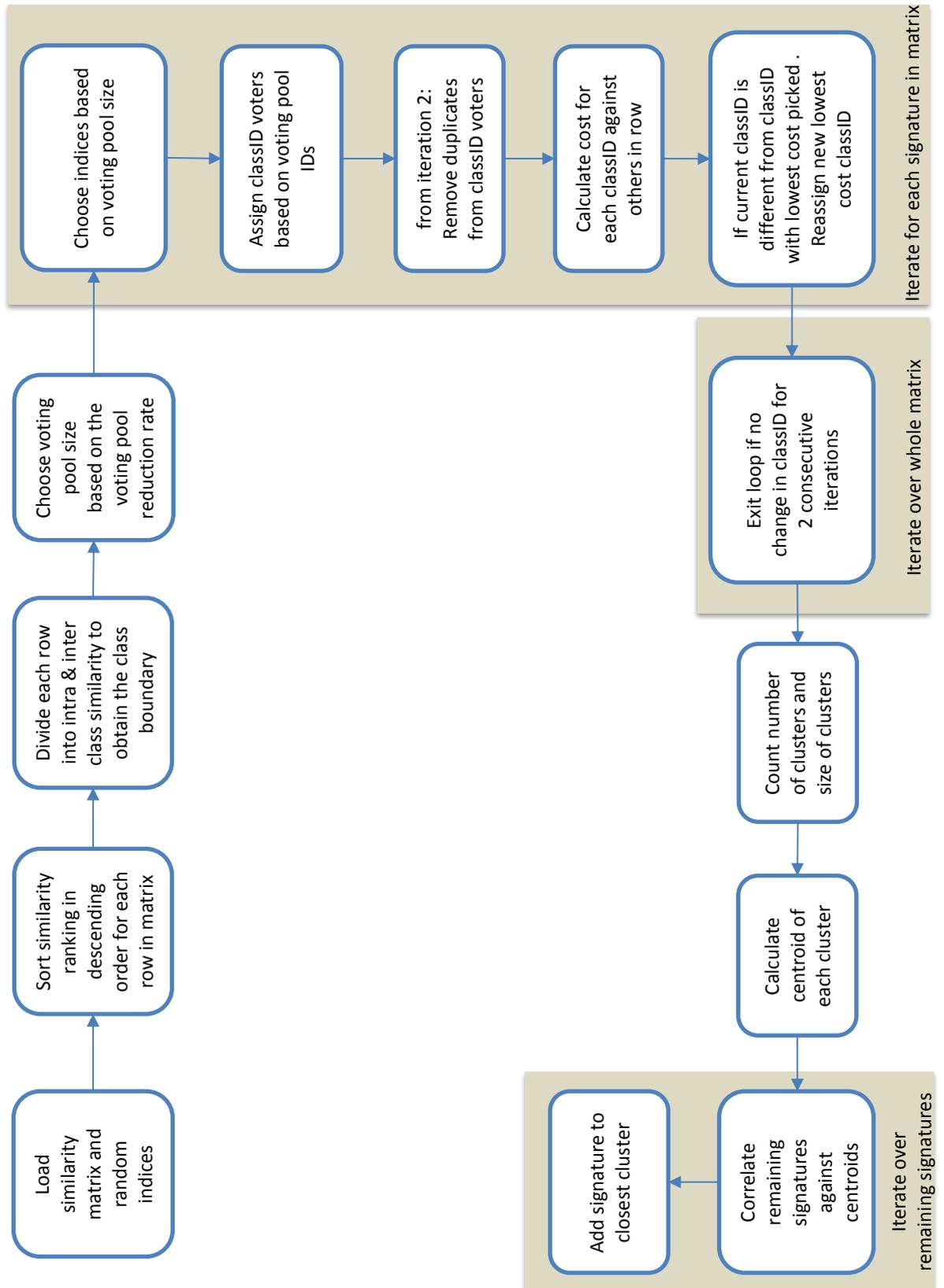
- IRIE, K., MCKINNON, A.E., UNSWORTH, K. and WOODHEAD, I.M., 2008. A model for measurement of noise in CCD digital-video cameras. *Measurement Science and Technology*, **19**(4), pp. 045207.
- IRIE, K., 2009. Noise-limited scene-change detection in images. *Ph.D. Thesis*, Lincoln University, UK.
- ISO/IEC JTC, 1991. *Digital compression and coding of continuous tone still images, Part 1 - Requirements and guidelines*. ISO/IEC JTC Committee Recommendation 10918_1.
- JANESICK, J.R., 2001. *Scientific Charge-Coupled Devices*. SPIE Press Book.
- JOLLIFFE, I.T., 1986. *Principal Component Analysis*. New York: Springer-Verlag.
- JONES, M.C. and SIBSON, R., 1987. What is projection pursuit? *Journal of the Royal Statistical Society. Series A (General)*, , pp. 1-37.
- KOHAVI, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection, *International joint Conference on artificial intelligence 1995*, Citeseer, pp. 1137-1145.
- KORNBLUM, J.D., 2008. Using JPEG Quantization Tables to Identify Imagery Processed by Software, *Proceedings of the Digital Forensic Workshop*, August 2008, pp. 21-25.
- LAM, K.P., 1998. High-performance thresholding with adaptive equalization, *Proceedings of SPIE 1998*, SPIE, pp. 148-333.
- LAM, K.P. and EMERY, R., 2009. Image pixel guided tours: a software platform for non-destructive x-ray imaging, *Proceedings of SPIE 2009*, pp. 72450N.
- LI, C.T., 2010. Source Camera Identification Using Enhanced Sensor Pattern Noise. *IEEE Transactions on Information Forensics and Security*, **5**(2), pp. 280-287.
- LI, C.T., 2010. Unsupervised Classification of Digital Images Using Enhanced Sensor Pattern Noise, *IEEE International Symposium on Circuits and Systems*, May 30th - June 2nd 2010 2010.
- LIN, S., JINWEI GU, YAMAZAKI, S. and HEUNG-YEUNG SHUM, 2004. Radiometric calibration from a single image, *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004*. 2004, pp. II-938-II-945 Vol.2.
- LUKAS, J., FRIDRICH, J. and GOLJAN, M., 2006. Digital camera identification from sensor pattern noise. *IEEE Transactions on Information Forensics and Security*, **1**(2), pp. 205-214.
- MCKAY, C., SWAMINATHAN, A., HONGMEI GOU and MIN WU, 2008. Image acquisition forensics: Forensic analysis to identify imaging source, *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008*. 2008, pp. 1657-1660.

- MIHCAK, M.K., KOZINTSEV, I. and RAMCHANDRAN, K., 1999. Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising, *Proceedings of the Acoustics, Speech, and Signal Processing, 1999. on 1999 IEEE International Conference-Volume 06* 1999, IEEE Computer Society, pp. 3253-3256.
- MOLER, C., 2004. Eigenvalues and Singular Values. *Numerical Computing with MATLAB*. <http://www.mathworks.co.uk/moler>: The MathWorks, Inc, pp. 269.
- NG, T., CHANG, S. and TSUI, M., 2007. Using Geometry Invariants for Camera Response Function Estimation, *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR '07*. 2007, pp. 1-8.
- RAMANATH, R., SNYDER, W.E., YOO, Y. and DREW, M.S., 2005. Color image processing pipeline. *Signal Processing Magazine, IEEE*, **22**(1), pp. 34-43.
- REDI, J.A., TAKTAK, W. and DUGELAY, J., 2011. Digital image forensics: a booklet for beginners. *Multimedia Tools Appl.*, **51**(1), pp. 133-162.
- SAN CHOI, K., LAM, E.Y. and WONG, K.K.Y., 2006. Source camera identification using footprints from lens aberration. *Digital Photography II SPIE*, **6069**(1), pp. 172-179.
- SENCAR, H.T. and MEMON, N., 2007. Overview of State-of-the-art in Digital Image Forensics. *Indian Statistical Institute Platinum Jubilee Monograph series titled Statistical Science and Interdisciplinary Research*, .
- SONG, I. and UHM, T.S., 1991. Multiplicative noise model and composite signal detection. *IEE Proceedings for Radar and Signal Processing*, **138**(6),.
- SOOBHANY, A.R., 2009. *Forensic Identification & Classification of Digital Images*. poster edn. Keele Graduate Symposium.
- SOOBHANY, A.R., LAM, K.P. and FLETCHER, P., 2009. Exploratory Visual Search of Image Sensor Noise in High-Dimensional Space, *13th IEEE Information Visualisation (IV09) conference*, July 2009.
- SOOBHANY, A.R., LEARY, R. and LAM, K.P., 2011. On the Performance of Li's Unsupervised Image Classifier and the Optimal Cropping Position of Images for Forensic Investigations. *International Journal of Digital Crime and Forensics (IJDCF)*, **3**(1), pp. 1-13.
- SORELL, M.J., 2008. Conditions for effective detection and identification of primary quantization of re-quantized JPEG images, *e-Forensics '08: Proceedings of the 1st international conference on Forensic applications and techniques in telecommunications, information, and multimedia and workshop* 2008, ICST, pp. 1-6.
- STONE, M. C., 2003. *A Field Guide to Digital Color*, A K Peters Ltd.
- SWAMINATHAN, A., WU, M. and LIU, K.J.R., 2007. Nonintrusive component forensics of visual sensors using output images. *IEEE Transactions on Information Forensics and Security*, **2**(1), pp. 91-106.

- VAN, L.T., EMMANUEL, S. and KANKANHALLI, M.S., 2007. Identifying Source Cell Phone Using Chromatic Aberration, *2007 IEEE International Conference on Multimedia and Expo 2007*, pp. 883-886.
- VRHEL, M., SABER, E. and TRUSSELL, H.J., 2005. Color image generation and display technologies. *Signal Processing Magazine, IEEE*, **22**(1), pp. 23-33.
- WALLACE, G.K., 1991. The JPEG still picture compression standard. *Communication of the ACM*, **34**(4), pp. 30-44.
- WEBB, A., 1999. *Statistical pattern recognition*. A Hodder Arnold Publication.
- XIE, H., PIERCE, L.E. and ULABY, F.T., 2002. Statistical properties of logarithmically transformed speckle. *IEEE Transactions on Geoscience and Remote Sensing*, **40**(3), pp. 721-727.

Appendix A

Classification pipeline in existing platform.



Appendix B

Nokia_C2_01_A (cam_1)

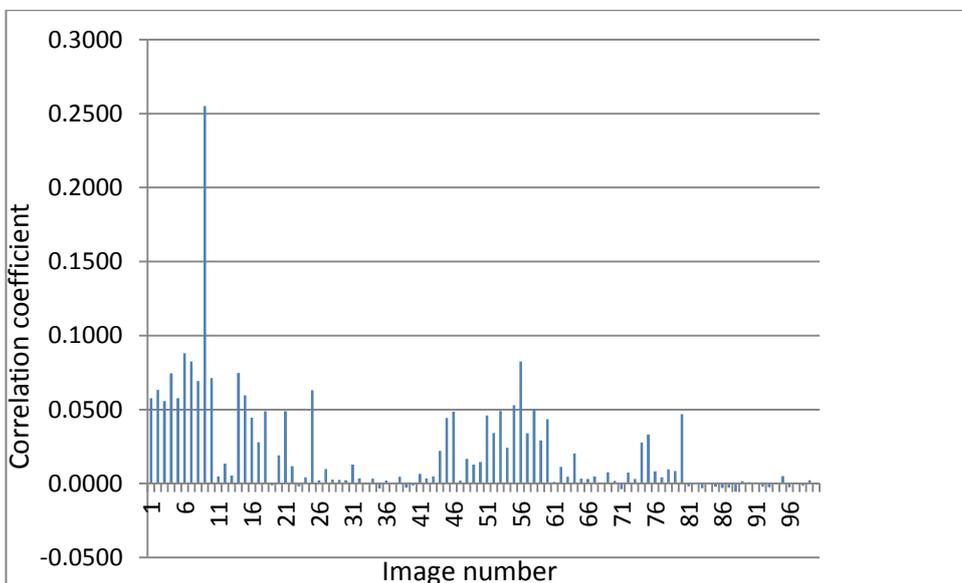


Figure B.1. nokia_C2_01_A (*cam_1*) SVD camera signature and correlation with 100 images. Images 1 to 10 come from this camera, images 11 to 20 from *cam_2* and rest of images from the other 9 cameras.

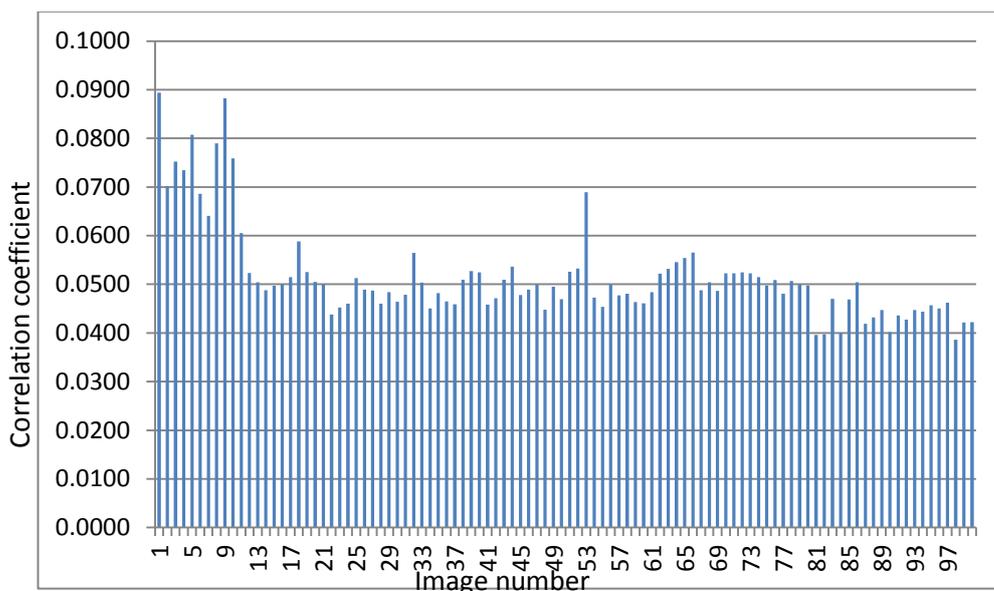


Figure B.2. nokia_C2_01_A (*cam_1*) wavelet camera signature and correlation with 100 images. Images 1 to 10 come from this camera, images 11 to 20 from *cam_2* and rest of images from the other 9 cameras.

Nokia_C2_01_B (cam_2)

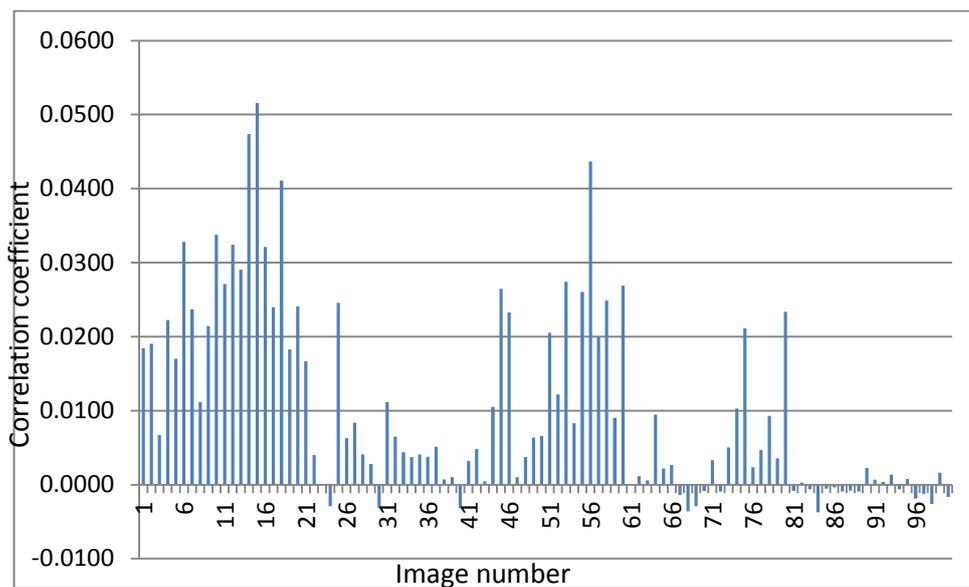


Figure B.3. nokia_C2_01_B (*cam_2*) SVD camera signature and correlation with 100 images. Images 11 to 20 come from this camera, images 1 to 10 from *cam_1* and rest of images from the other 9 cameras.

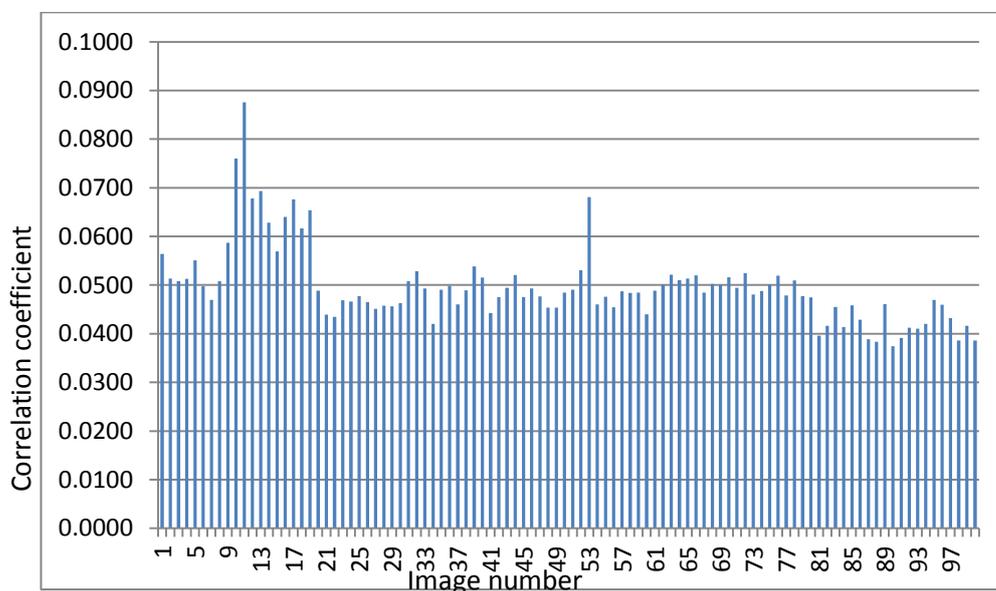


Figure B.4. nokia_C2_01_B (*cam_2*) wavelet camera signature and correlation with 100 images. Images 11 to 20 come from this camera, images 1 to 10 from *cam_1* and rest of images from the other 9 cameras.

Nokia_E72_A (cam_3)

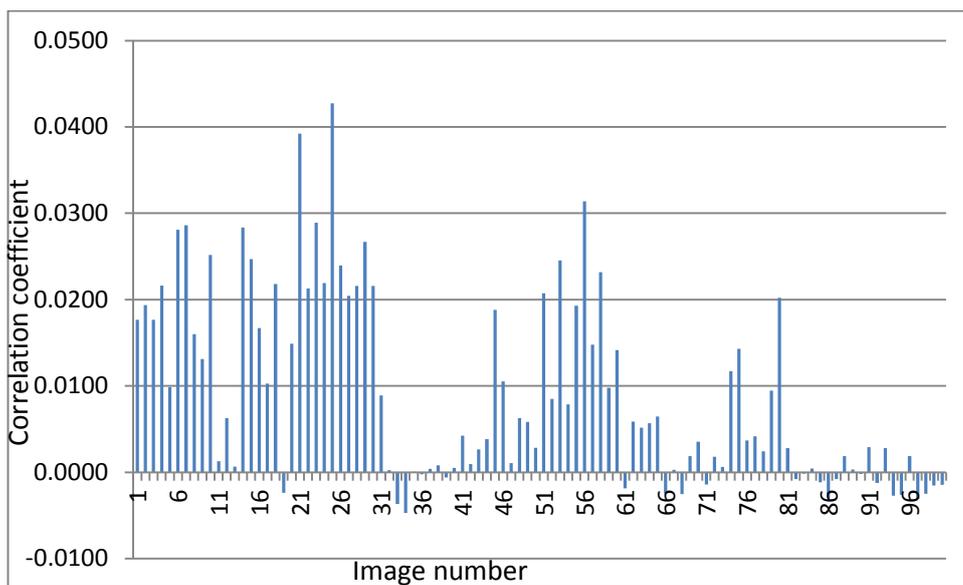


Figure B.5. nokia_E72_A (*cam_3*) SVD camera signature and correlation with 100 images. Images 21 to 30 come from this camera, images 31 to 40 from *cam_4* and rest of images from the other 9 cameras.

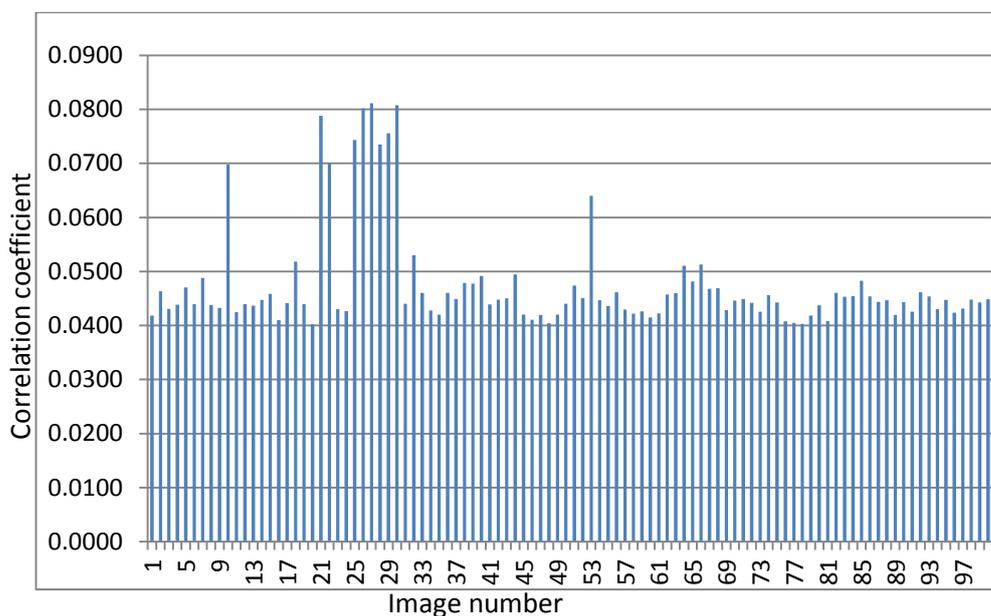


Figure B.6. nokia_E72_A (*cam_3*) wavelet camera signature and correlation with 100 images. Images 21 to 30 come from this camera, images 31 to 40 from *cam_4* and rest of images from the other 9 cameras.

Nokia_E72_B (cam_4)

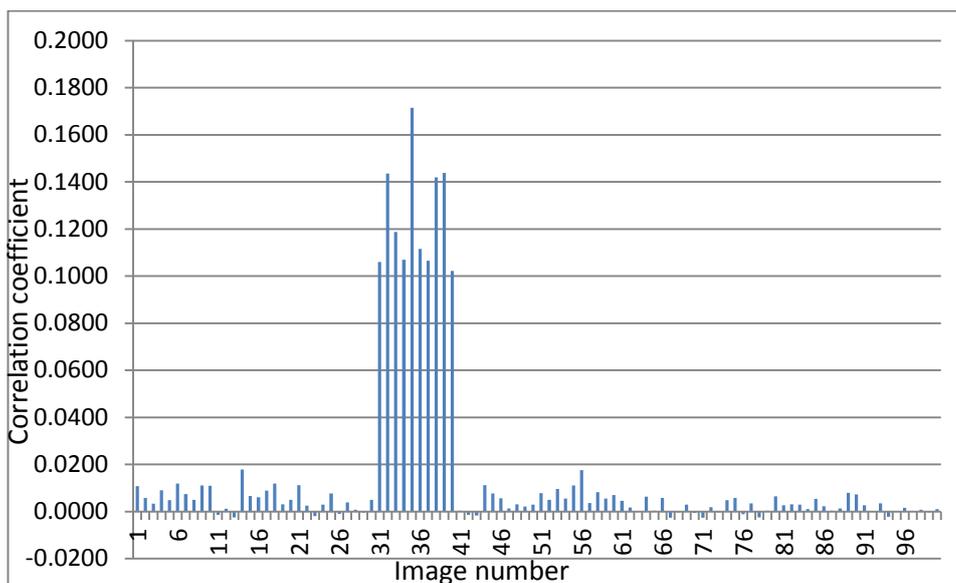


Figure B.7. nokia_E72_B (*cam_4*) SVD camera signature and correlation with 100 images. Images 31 to 40 come from this camera, images 21 to 30 from *cam_3* and rest of images from the other 9 cameras.

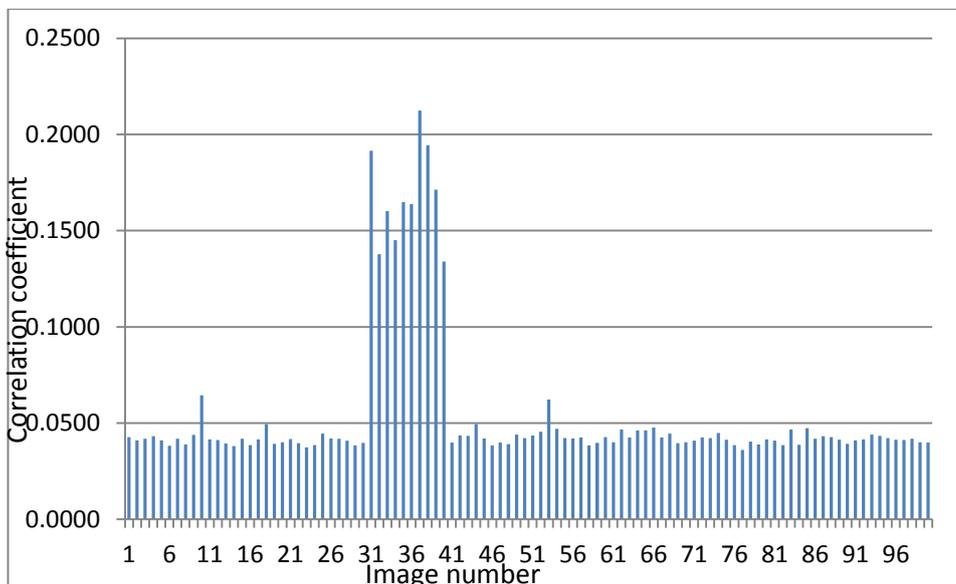


Figure B.8. nokia_E72_B (*cam_4*) wavelet camera signature and correlation with 100 images. Images 31 to 40 come from this camera, images 21 to 30 from *cam_3* and rest of images from the other 9 cameras.

Nokia_N95_A (cam_5)

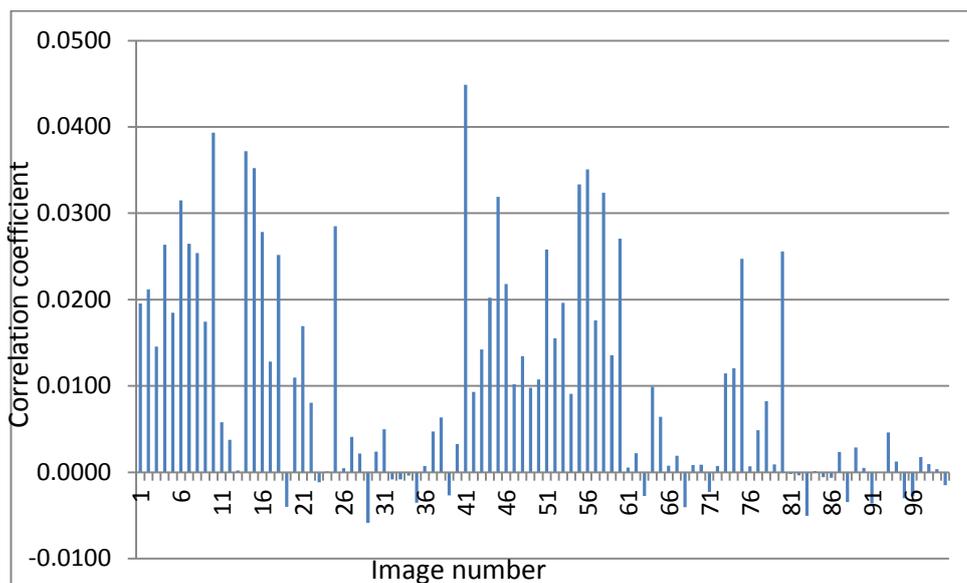


Figure B.9. nokia_N95_A (*cam_5*) SVD camera signature and correlation with 100 images. Images 41 to 50 come from this camera, images 51 to 60 from *cam_6* and rest of images from the other 9 cameras.

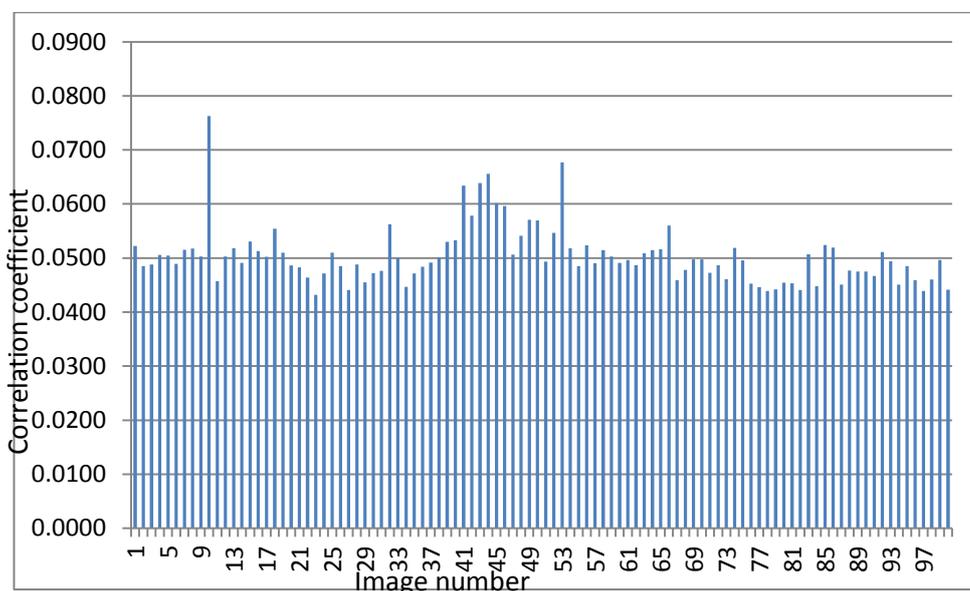


Figure B.10. nokia_N95_A (*cam_5*) wavelet camera signature and correlation with 100 images. Images 41 to 50 come from this camera, images 51 to 60 from *cam_6* and rest of images from the other 9 cameras.

Nokia_N95_B (cam_6)

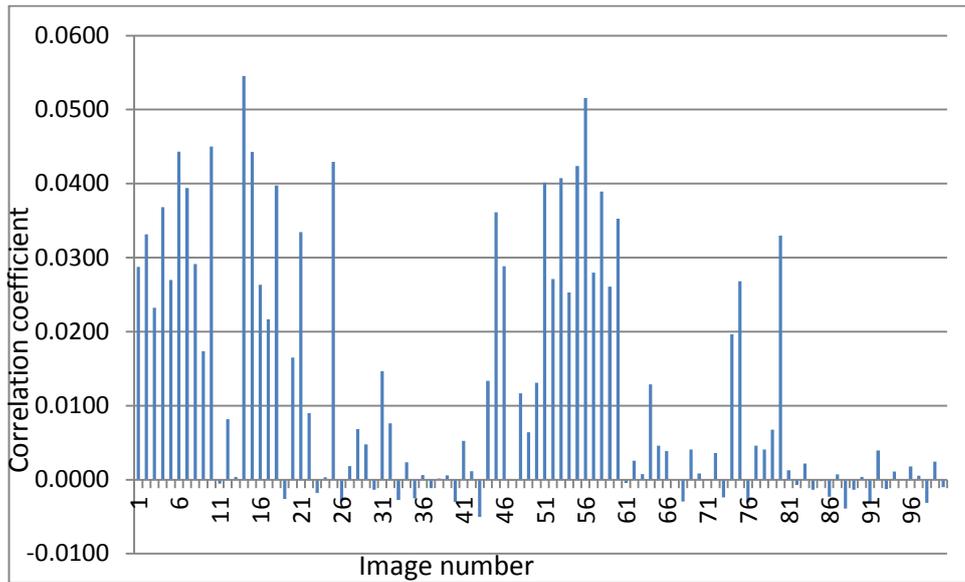


Figure B.11. nokia_N95_B (*cam_6*) SVD camera signature and correlation with 100 images. Images 51 to 60 come from this camera, images 41 to 50 from *cam_5* and rest of images from the other 9 cameras.

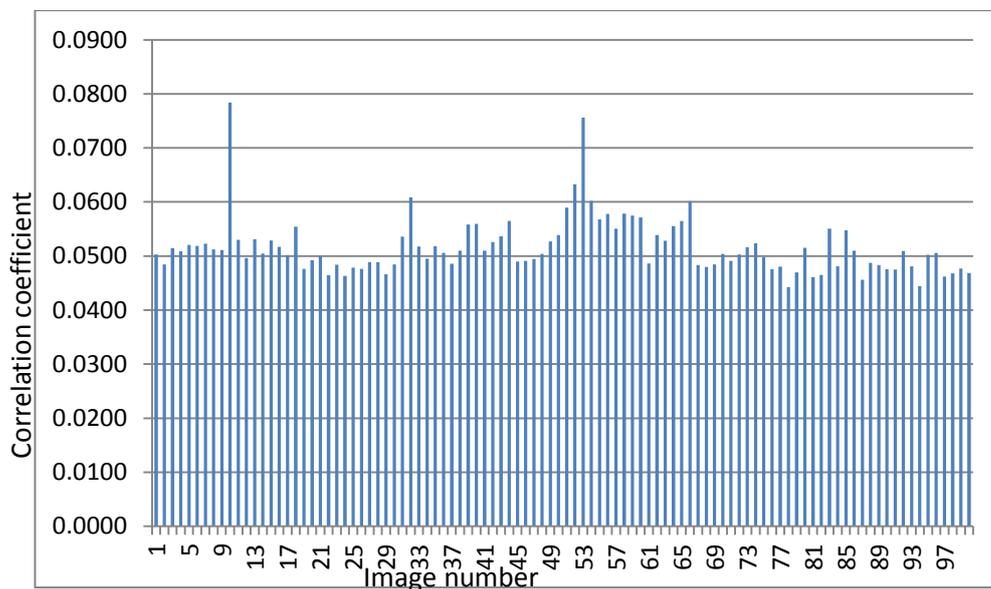


Figure B.12. nokia_N95_B (*cam_6*) wavelet camera signature and correlation with 100 images. Images 51 to 60 come from this camera, images 41 to 50 from *cam_5* and rest of images from the other 9 cameras.

Samsung_galaxy_S2_A (cam_7)

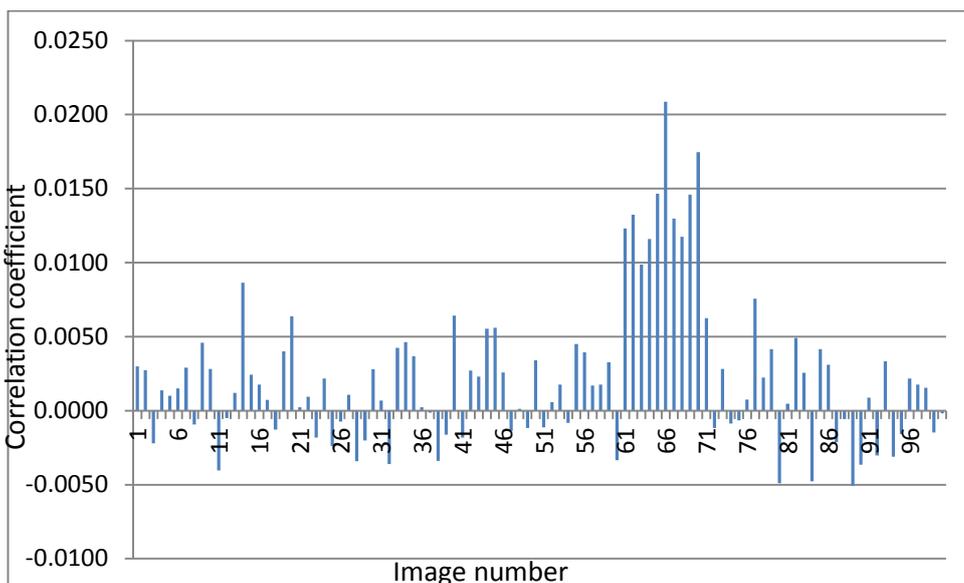


Figure B.13. samsung_galaxy_S2_A (*cam_7*) SVD camera signature and correlation with 100 images. Images 61 to 70 come from this camera, images 71 to 80 from *cam_8* and rest of images from the other 9 cameras.

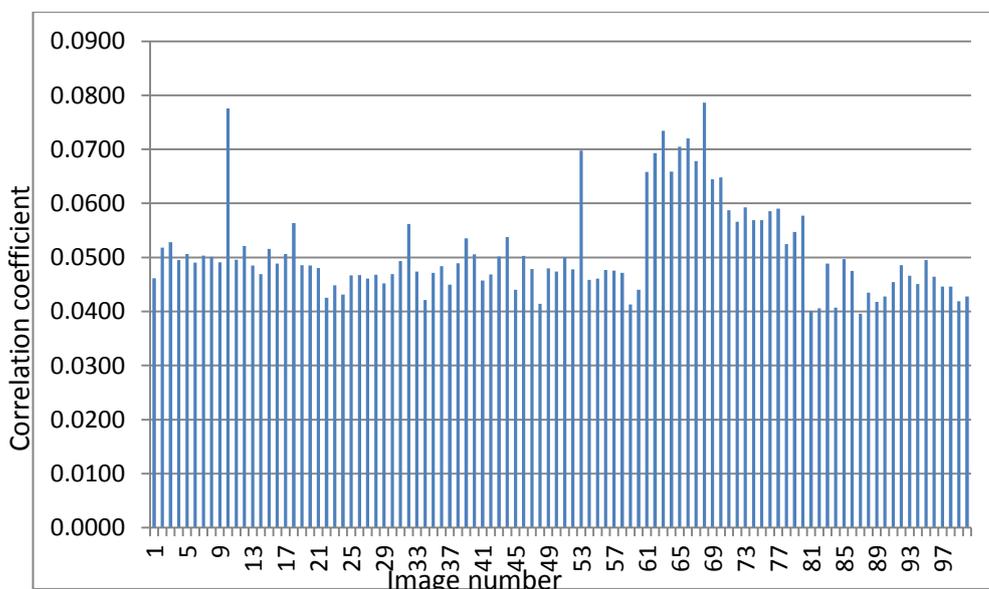


Figure B.14. samsung_galaxy_S2_A (*cam_7*) wavelet camera signature and correlation with 100 images. Images 61 to 70 come from this camera, images 71 to 80 from *cam_8* and rest of images from the other 9 cameras.

Samsung_galaxy_S2_B (cam_8)

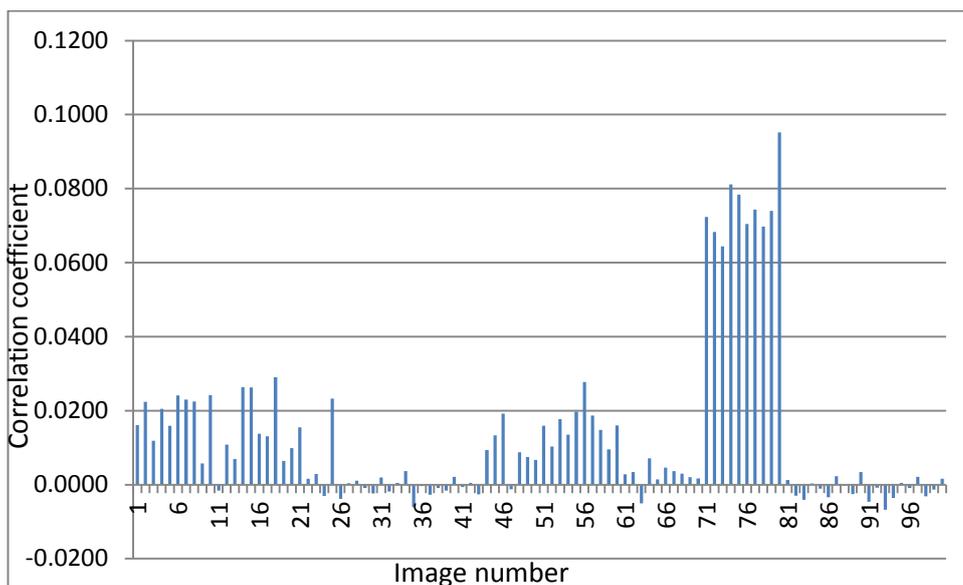


Figure B.15. samsung_galaxy_S2_B (*cam_8*) SVD camera signature and correlation with 100 images. Images 71 to 80 come from this camera, images 61 to 70 from *cam_7* and rest of images from the other 9 cameras.

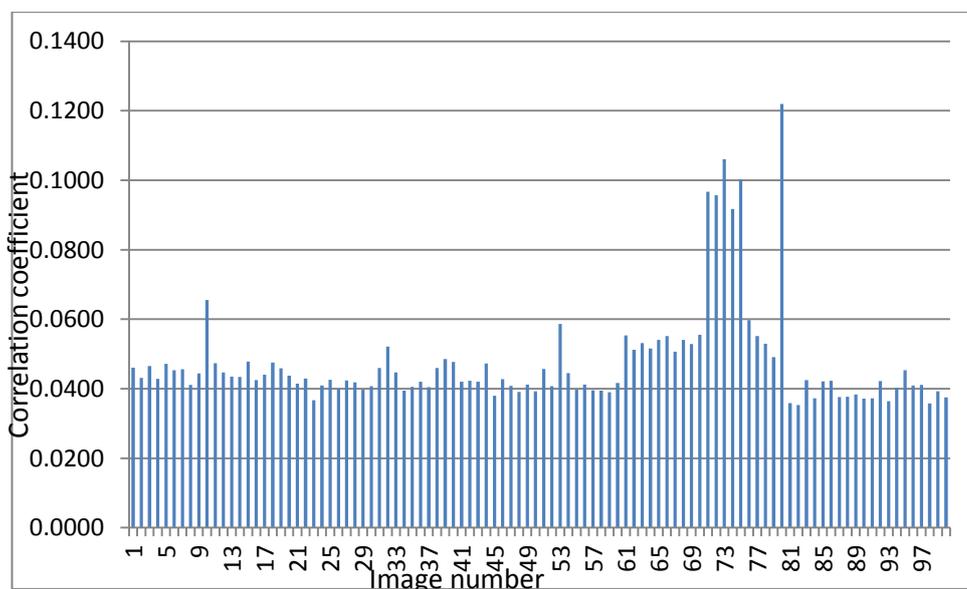


Figure B.16. samsung_galaxy_S2_B (*cam_8*) wavelet camera signature and correlation with 100 images. Images 71 to 80 come from this camera, images 61 to 70 from *cam_7* and rest of images from the other 9 cameras.

Zte_orange_sanfrancisco_A (cam_9)

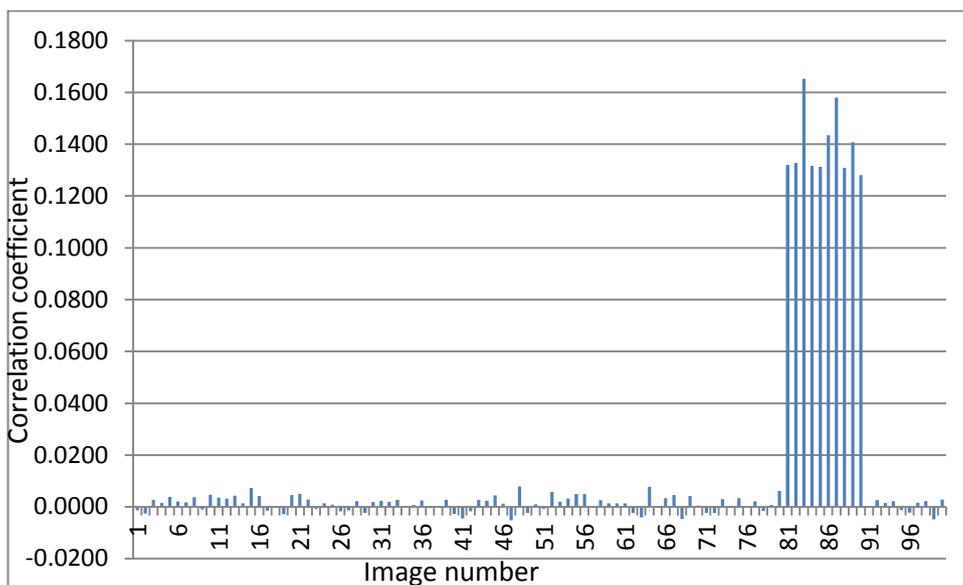


Figure B.17. zte_orange_sanfrancisco_A (*cam_9*) SVD camera signature and correlation with 100 images. Images 81 to 90 come from this camera, images 91 to 100 from *cam_10* and rest of images from the other 9 cameras.

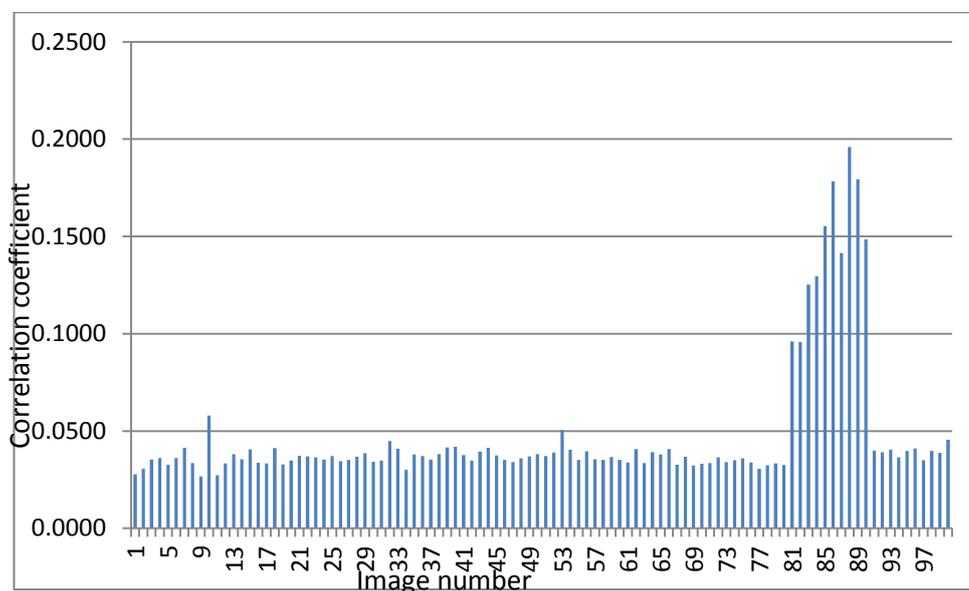


Figure B.18. zte_orange_sanfrancisco_A (*cam_9*) wavelet camera signature and correlation with 100 images. Images 81 to 90 come from this camera, images 91 to 100 from *cam_10* and rest of images from the other 9 cameras.

Zte_orange_sanfrancisco_B (cam_10)

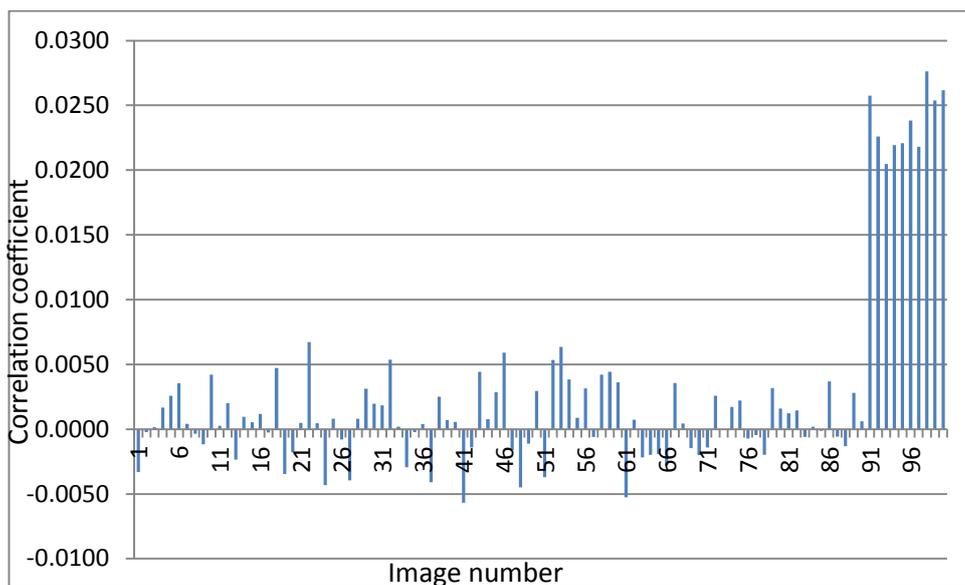


Figure B.19. zte_orange_sanfrancisco_B (cam_10) SVD camera signature and correlation with 100 images. Images 91 to 100 come from this camera, images 81 to 90 from *cam_9* and rest of images from the other 9 cameras.

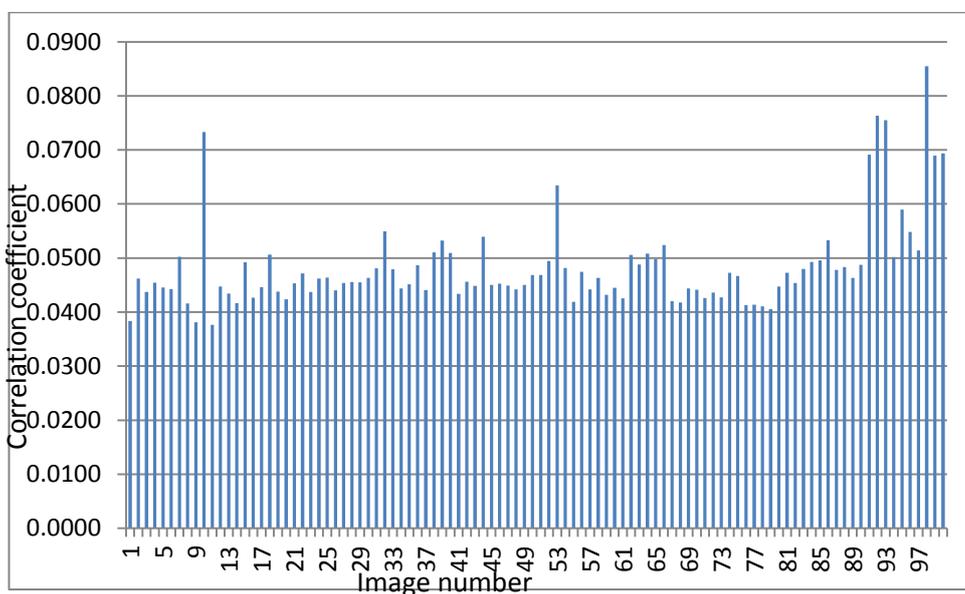


Figure B.20. zte_orange_sanfrancisco_B (cam_10) wavelet camera signature and correlation with 100 images. Images 91 to 100 come from this camera, images 81 to 90 from *cam_9* and rest of images from the other 9 cameras.

Appendix C

Camera	Images from same camera		Images from other cameras	
	<i>SVD</i>	<i>Wavelet</i>	<i>SVD</i>	<i>wavelet</i>
cam_1	0.0238	0.0765	0.0006	0.0487
cam_2	0.0327	0.0652	0.0077	0.0479
cam_3	0.0268	0.0700	0.0070	0.0450
cam_4	0.1253	0.1675	0.0020	0.0422
cam_5	0.0187	0.0589	0.0088	0.0494
cam_6	0.0355	0.0600	0.0096	0.0507
cam_7	0.0139	0.0693	0.0010	0.0489
cam_8	0.0748	0.0829	0.0065	0.0437
cam_9	0.1394	0.1445	0.0013	0.0367
cam_10	0.0238	0.0660	0.0006	0.0463

Table C.1. Mean of correlation coefficient values when the ten camera reference signatures, extracted using SVD method and wavelet method, are compared with the test images from the same camera and rest of the cameras.

Camera	Images from same camera		Images from other cameras	
	<i>SVD</i>	<i>Wavelet</i>	<i>SVD</i>	<i>wavelet</i>
cam_1	0.059845	0.008136	0.020353	0.004719
cam_2	0.010787	0.009912	0.010409	0.005634
cam_3	0.007956	0.01476	0.009313	0.004241
cam_4	0.023346	0.025638	0.004408	0.004171
cam_5	0.011648	0.0046	0.012079	0.004558
cam_6	0.008728	0.005915	0.014862	0.004327
cam_7	0.003206	0.004513	0.00298	0.005976
cam_8	0.008639	0.026104	0.009419	0.005609
cam_9	0.012744	0.034084	0.002854	0.004444
cam_10	0.002339	0.011761	0.002339	0.004862

Table C.2. Standard deviation of correlation coefficient values when the ten camera reference signatures, extracted using SVD method and wavelet method, are compared with the test images from the same camera and rest of the cameras.

Appendix D

Nokia_C2_01_A (cam_1)

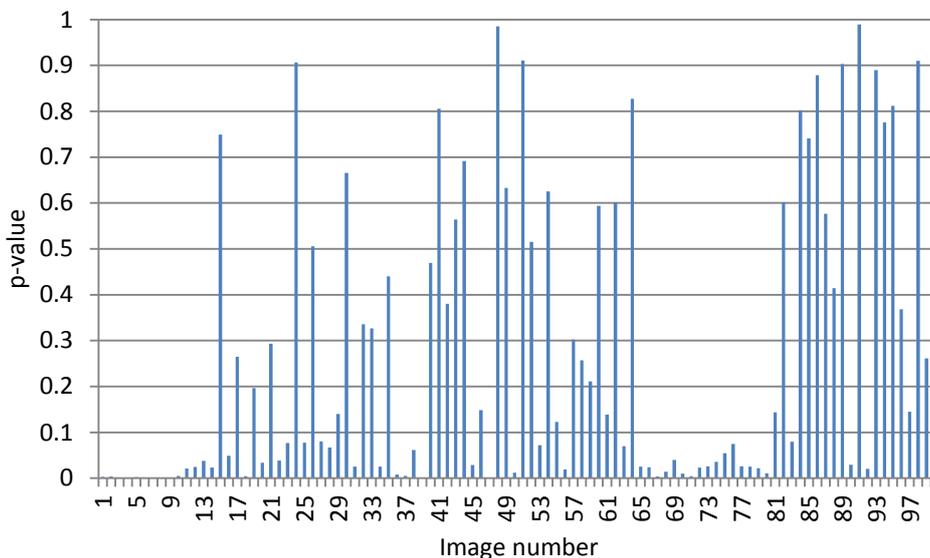


Figure D.1. nokia_C2_01_A (*cam_1*) SVD camera signature and p-values with 100 images. Images 1 to 10 come from this camera, images 11 to 20 from *cam_2* and rest of images from the other 9 cameras.

Nokia_C2_01_B (cam_2)

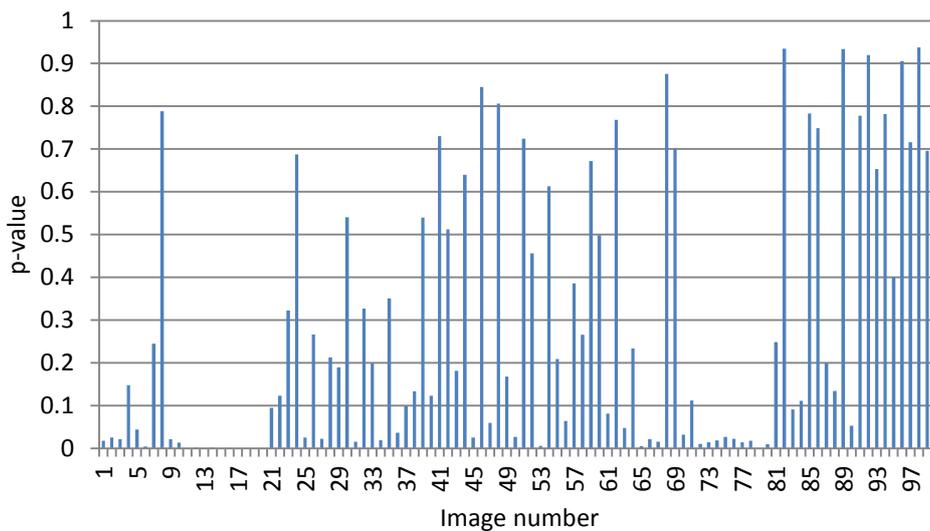


Figure D.2. nokia_C2_01_B (*cam_2*) SVD camera signature and p-values with 100 images. Images 11 to 20 come from this camera, images 1 to 10 from *cam_1* and rest of images from the other 9 cameras.

Nokia_E72_A (cam_3)

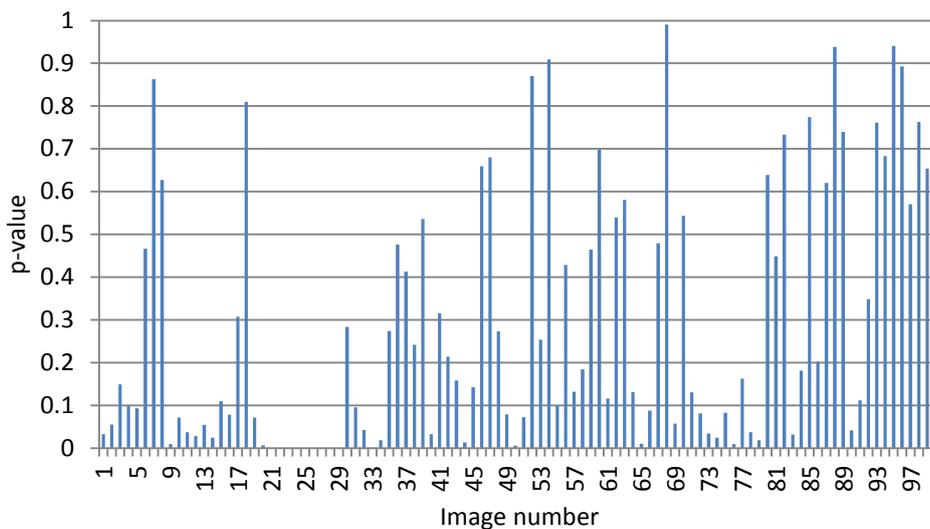


Figure D.3. nokia_E72_A (*cam_3*) SVD camera signature and p-values with 100 images. Images 21 to 30 come from this camera, images 31 to 40 from *cam_4* and rest of images from the other 9 cameras.

Nokia_E72_B (cam_4)

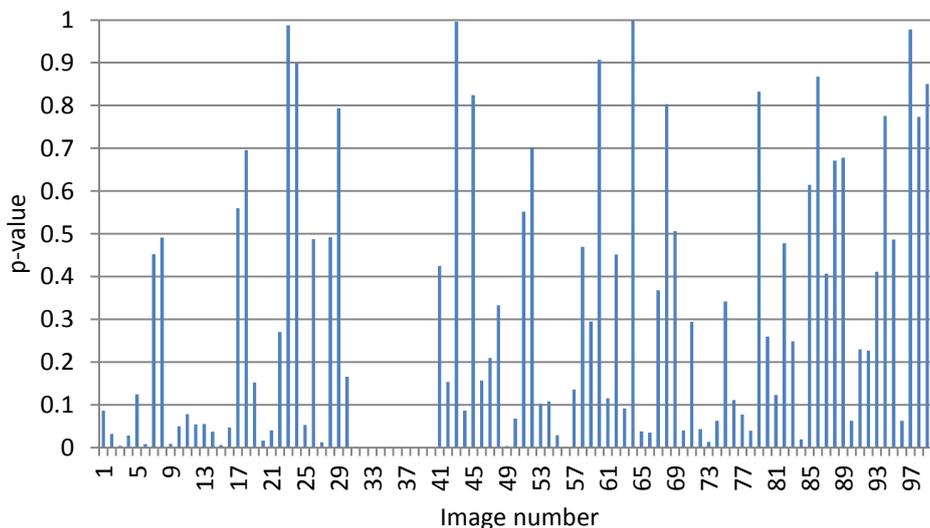


Figure D.4. nokia_E72_B (*cam_4*) SVD camera signature and p-values with 100 images. Images 31 to 40 come from this camera, images 21 to 30 from *cam_3* and rest of images from the other 9 cameras.

Nokia_N95_A (cam_5)

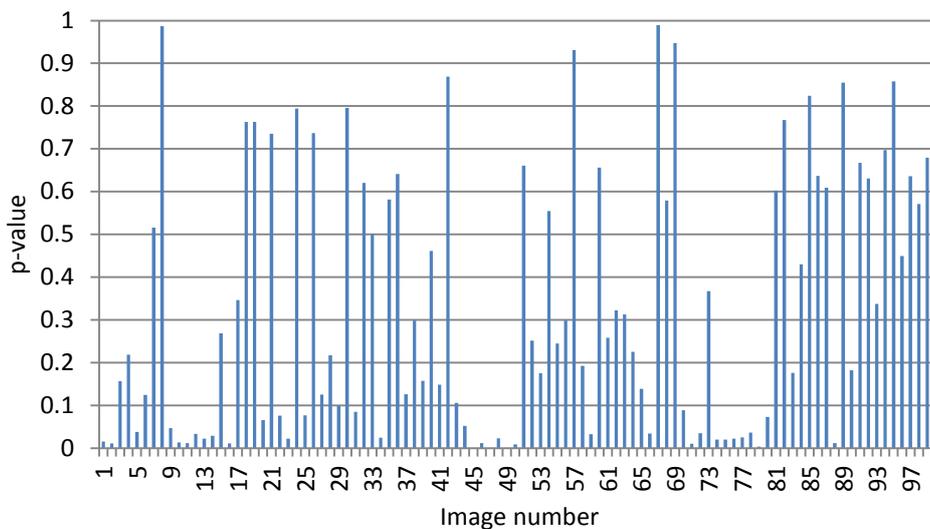


Figure D.5. nokia_N95_A (*cam_5*) SVD camera signature and p-values with 100 images. Images 41 to 50 come from this camera, images 51 to 60 from *cam_6* and rest of images from the other 9 cameras.

Nokia_N95_B (cam_6)

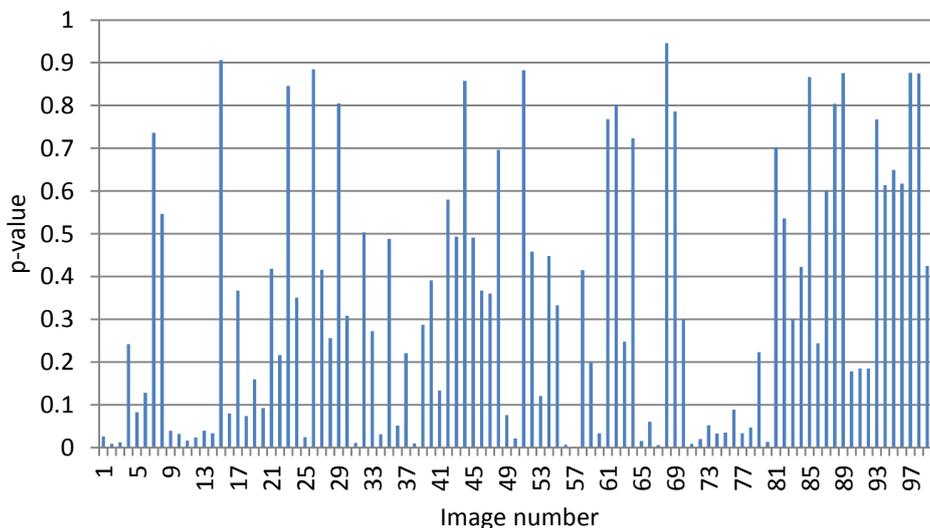


Figure D.6. nokia_N95_B (*cam_6*) SVD camera signature and p-values with 100 images. Images 51 to 60 come from this camera, images 41 to 50 from *cam_5* and rest of images from the other 9 cameras.

Samsung_galaxy_S2_A (cam_7)

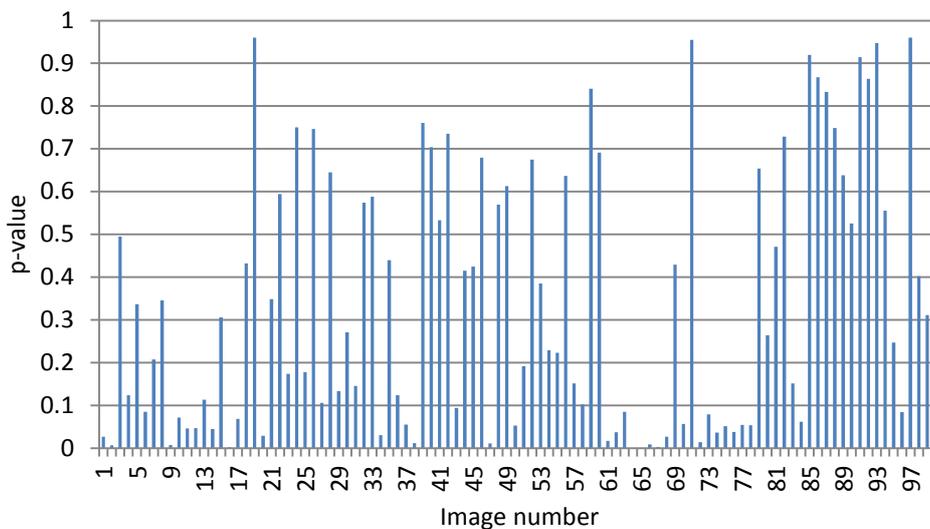


Figure D.7. samsung_galaxy_S2_A (cam_7) SVD camera signature and p-values with 100 images. Images 61 to 70 come from this camera, images 71 to 80 from *cam_8* and rest of images from the other 9 camera.

Samsung_galaxy_S2_B (cam_8)

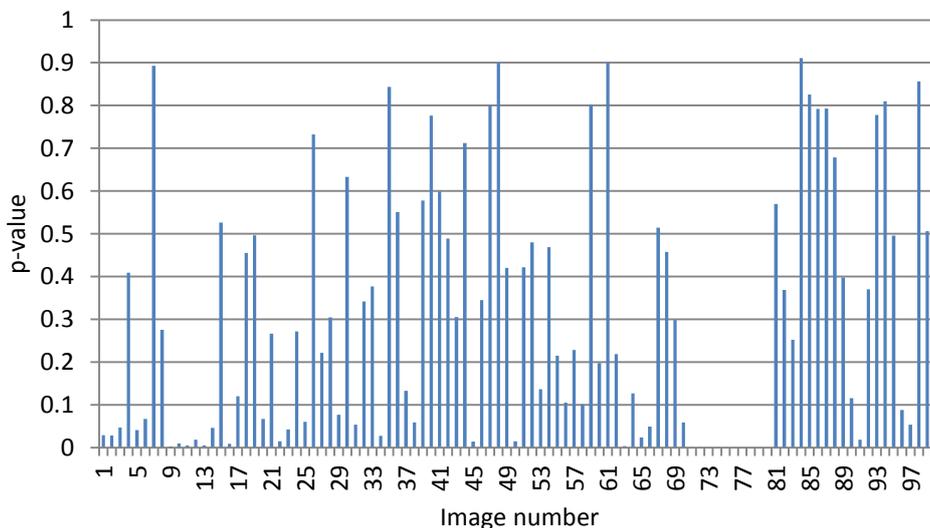


Figure D.8. samsung_galaxy_S2_B (cam_8) SVD camera signature and p-values with 100 images. Images 71 to 80 come from this camera, images 61 to 70 from *cam_7* and rest of images from the other 9 camera.

Zte_orange_sanfrancisco_A (cam_9)

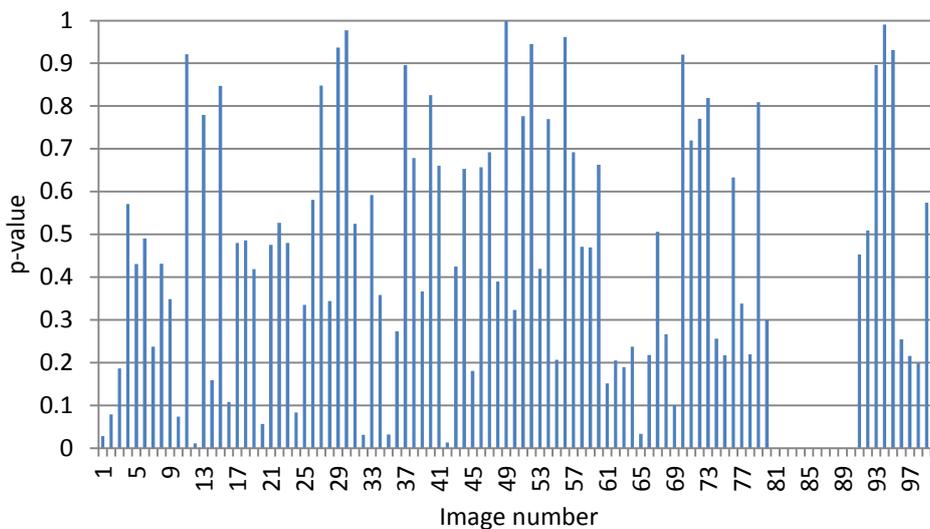


Figure D.9. zte_orange_sanfrancisco_A (*cam_9*) SVD camera signature and p-values with 100 images. Images 81 to 90 come from this camera, images 91 to 100 from *cam_10* and rest of images from the other 9 cameras.

Zte_orange_sanfrancisco_B (cam_10)

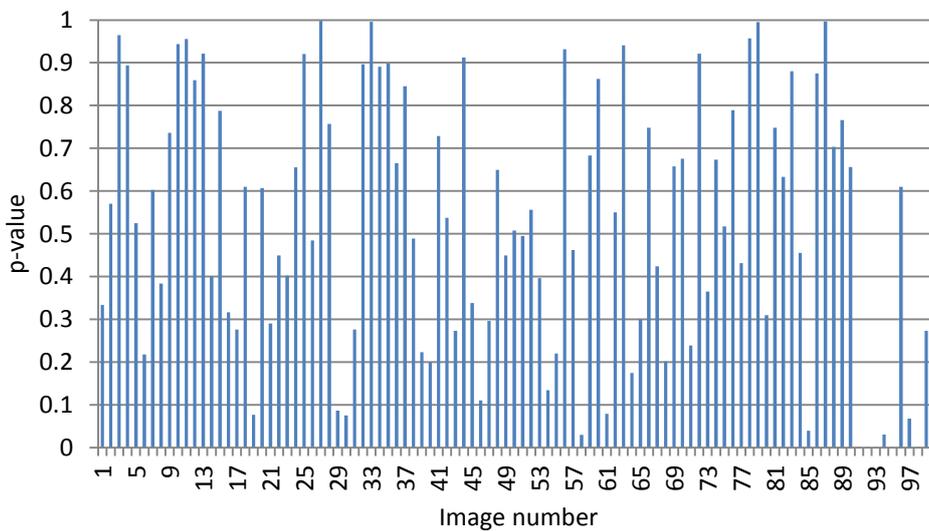


Figure D.10. zte_orange_sanfrancisco_B (*cam_10*) SVD camera signature and p-values with 100 images. Images 91 to 100 come from this camera, images 81 to 90 from *cam_9* and rest of images from the other 9 cameras.