SOURCE CAMERA IDENTIFICATION FOR ONLINE SOCIAL NETWORK IMAGES USING TEXTURE FEATURE

NORDIANA BINTI RAHIM

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

SOURCE CAMERA IDENTIFICATION FOR ONLINE SOCIAL NETWORK IMAGES USING TEXTURE FEATURE

NORDIANA BINTI RAHIM

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

UNIVERSITY OF MALAYA
ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Nordiana Binti Rahim

Matric No: WHA120011

Name of Degree: PhD Computer Science (Digital Forensics)

Title of Thesis ("this Work"): **SOURCE CAMERA IDENTIFICATION FOR ONLINE SOCIAL NETWORK IMAGES USING TEXTURE FEATURE**

Field of Study: Computer Science (Digital Forensics)

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every right in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

UNIVERSITI MALAYA

PERAKUAN KEASLIAN PENULISAN

Nama: Nordiana Binti Rahim

No. Matrik: WHA120011

Nama Ijazah: PhD Komputer Sains (Forensik Digital)

Tajuk Tesis ("Hasil Kerja ini"): **PENGENALPASTIAN SUMBER KAMERA UNTUK IMEJ RANGKAIAN SOSIAL ATAS TALIAN MENGGUNAKAN CIRI TEKSTUR**

Bidang Penyelidikan: Komputer Sains (Forensik Digital)

Saya dengan sesungguhnya dan sebenarnya mengaku bahawa:

(1) Saya adalah satu-satunya pengarang/penulis Hasil Kerja ini;
(2) Hasil Kerja ini adalah asli;
(3) Apa-apa penggunaan dan mana-mana hasil kerja yang mengandungi hakcipta, telah dilakukan secara urusan yang wajar dan bagi maksud yang dibenarkan dan apa-apa petikan, ekstrak, rujukan atau pengeluaran semula daripada atau kepada mana-mana hasil kerja yang mengandungi hak cipta telah dinyatakan dengan sejelasnya dan secukupnya dan satu pengiktirafan tajuk hasil kerja tersebut dan pengarang/penulisnya telah dilakukan di dalam Hasil Kerja ini;
(4) Saya tidak mempunyai apa-apa pengetahuan sebenar atau patut semunasabahnya tahu bahawa penghasilan Hasil Kerja ini melanggar suatu hak cipta hasil kerja yang lain;
(5) Saya dengan ini menyerahkan kesemua dan tiap-tiap hak yang terkandung di dalam hakcipta Hasil Kerja ini kepada Universiti Malaya ("UM") yang seterusnya mula dari sekarang adalah tuan punya kepada hakcipta di dalam Hasil Kerja ini dan apa-apa pengeluaran semula atau penggunaan dalam apa jua bentuk atau dengan apa juga cara sekalipun adalah dilarang tanpa terlebih dahulu mendapat kebenaran bertulis dari UM;
(6) Saya sedar sepenuhnya sekiranya dalam masa penghasilan Hasil Kerja ini saya telah melanggar suatu hak cipta hasil kerja yang lain sama ada dengan niat atau sebaliknya, saya boleh dikenakan tindakan undang-undang atau apa-apa tindakan lain sebagaimana yang diputuskan oleh UM.

Tandatangan Calon Tarikh:

Diperbuat dan sesungguhnya diakui di hadapan,

Tandatangan Saksi Tarikh:

Nama:

Jawatan:

SOURCE CAMERA IDENTIFICATION FOR ONLINE SOCIAL NETWORK IMAGES USING TEXTURE FEATURE

ABSTRACT

The numbers of Online Social Network (OSN) users have grown extensively in the recent decades due to the production of various affordable high technology devices for example smartphones with high-end featured camera, and free online social network apps. However, the rapid growth of this technology has also increased the risk of cybercrime, which exposed users to identity theft, scamming and fraud risk. Therefore, digital image from OSNs may provide authorities with crucial evidence to probe further into the crimes. This highlights the importance of digital image forensic in aiding the authorities to curb the cybercrime issues. Digital Image Forensic is an area of study that mainly focuses on validating the authenticity of digital images by extracting detailed information in those images; including resolution, type of devices, location, times and dates. There are two main methods under digital image forensic, namely source identification which focusing on extracting details of the device used to take digital images, and forgeries detections which focusing on detecting changes made to digital images. This research proposed a technique to identify the source camera of digital image, particularly for OSN images. Images obtained from OSNs web have been processed and modified to meet the OSNs service provider's requirement prior to publication. The process among others includes reducing its resolutions and size. This process also caused some important information in those images are missing or completely erased, making it difficult or impossible to identify the camera source. In response to this limitation, a new technique was proposed for source camera identification using Gray Level Co-Occurrence Matrix (GLCM) and Gray Level Run Length Matrix (GLRLM) texture feature. Image texture feature is referring to a set of

metrics that provides information about the colour arrangement, intensities or selected region of the image which were derived from image processing. Therefore, in this research, image texture features were utilised to propose the new technique which was evaluated based on the percentage of detection accuracy. The results from this research proved that using GLCM and GLRLM features the proposed technique able to identify the source of camera from both original and OSN images with a high detection accuracy of 99.30% and 99.67% respectively.
Keywords: Digital Forensics, Image Forensics, Source Camera Identification, Texture Features, Social Media

PENGENALPASTIAN SUMBER KAMERA UNTUK IMEJ RANGKAIAN SOCIAL ATAS TALIAN MENGGUNAKAN CIRI TEKSTUR

ABSTRAK

Bilangan pengguna Rangkaian Sosial Atas Talian (OSN) telah berkembang dengan begitu pesat dalam beberapa dekad kebelakangan ini disebabkan oleh pengeluaran pelbagai jenis peranti berteknologi tinggi dan mampu milik seperti telefon pintar dengan kamera berprestasi tinggi dan applikasi media sosial yang ditawarkan percuma secara atas talian. Namun, kepesatan teknologi ini juga telah menyebabkan peningkatan risiko berlakunya jenayah siber, yang meletakkan pengguna terdedah kepada risiko jenayah kecurian maklumat peribadi, penipuan atas talian dan penipuan dalam transaksi perundingan. Oleh hal yang demikian, imej digital daripada OSN boleh menjadi sumber bahan bukti penting bagi pihak berkuasa untuk penyiasatan yang lebih terperinci dalam kes-kes jenayah tersebut. Situasi ini menjelaskan akan kepentingan forensik imej digital dalam membantu pihak berkuasa menangani isu jenayah siber berkenaan. Forensik Imej Digital adalah bidang pengajian yang lebih tertumpu kepada pengesahan keaslian imej digital dengan cara mengekstrak maklumat terperinci daripada imej-imej tersebut; termasuk maklumat resolusi, jenis peranti, lokasi, masa dan tarikh. Terdapat dua kaedah utama di bawah foreksik imej digital, iaitu pengenalpastian sumber yang tertumpu pada pengekstrakan maklumat-maklumat peranti yang digunakan untuk mangambil imej digital, dan pengesanan pemalsuan yang tertumpu kepada mengesan perubahan yang dilakukan pada imej digital. Penyelidikan ini mencadangkan satu teknik untuk mengenal pasti sumber peranti daripada imej digital terutamanya bagi imej OSN. Imej yang diperolehi daripada web OSN telah diproses dan diubah suai bagi memenuhi keperluan pembekal perkhidmatan OSN sebelum ia di paparkan. Proses tersebut antara lain termasuk pengurangan resolusi dan saiz. Proses ini juga menyebabkan sebahagian maklumat penting di dalam imej tersebut samada hilang atau terpadam sepenuhnya, menjadikannya sukar atau hampir mustahil untuk mengenal pasti sumber peranti yang digunakan. Sebagai langkah mengatasi permasaalahan ini, satu teknik baru untuk pengenalpastian sumber kamera telah di cadangkan dengan menggunakan ciri tekstur Gray Level Co-Occurrence Matrix (GLCM) dan Gray Level Run Length Matrix (GLRLM). Ciri imej tekstur merujuk kepada set metrik yang mengandungi maklumat-maklumat berkaitan susunan warna, intensiti atau kawasan terpilih dalam imej tersebut yang di perolehi melalui pemprosesan imej. Penyelidikan ini telah menggunakan ciri tekstur tersebut bagi teknik baru yang di cadangkan, yang telah dinilai bedasarkan peratus ketepatan pengenalpastian. Keputusan hasil dari penyelidikan ini membuktikan bahawa dengan penggunaan ciri GLCM dan GLRLM, teknik yang di cadanganka n berupaya untuk mengenal pasti sumber peranti bagi kedua dua jenis imej asli dan OSN dengan peratusan ketepatan pengenalpastian yang tinggi iaitu masing-masing dengan 99.30% dan 99.67%.

Keywords: Forensik Digital, Forensik Imej, Pengenalpastian sumber kamera, ciri-ciri tekstur, Sosial Media

ACKNOWLEDGEMENTS

First and foremost, I would like to extend my thanks and immense gratitude to Allah S.W.T for giving the good life in completing my PhD successfully. I would also like to express my sincere gratitude to my supervisors Prof. Miss Laiha Mat Kiah and Associate Prof. Dr. Ainuddin Wahid Abdul Wahab for their guidance, support, patience and tolerance to me throughout the completion of this research. Their advice and guidance have been much more important to this PhD successful journey.

Furthermore, my sincere gratitude is dedicated to my parents Rahim Bin Garib and Ramlah Binti Marjuki who have sacrificed their time, efforts and resources to train me both physically, mentally and intellectually to become an important and a valuable person towards the society to the society. My gratitude also is dedicated to my beloved husband Noreffendi Bin Eskan and lovely daughter Auni Adlina Binti Noreffendi for their unconditional love, support and patience throughout my PhD journey. Thank you for their inspiration and goodwill.

A special acknowledgement also goes to the Ministry of Higher Education and Universiti Tun Hussein Onn Malaysia for the financial support towards the success of this PhD study.

This project is dedicated to my father Rahim Bin Garib, my mother Ramlah Binti Marjuki, my husband Noreffendi Bin Eskan, my daughter Auni Adlina Binti Noreffendi, the family of Rahim Bin Garib and the family in-laws of Eskan Bin Sitam.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

AR	:	Auto Regressive
ARMA	:	Auto Regressive Moving Average
BPNN	:	Back-Propagation Neural Network
CCD	:	Charged Couple Device
CCN	:	Correlation to Circular Correlation Norm
CFA	:	Color Filter Array
CMOS	:	Complementary Metal Oxide Semiconductor
CP	:	Conditional Probabilities
CS-LBP	:	Center-Symmetric Local Binary Pattern
DCT	:	Discrete Cosine Transform
DSC		Digital Still Camera
DSLR	:	Digital Single Lens Reflex
DSNU	:	Dark Signal Non-Uniformity
EM	:	Expectation-Maximization
EXIF	:	Exchangeable Image File Format
FCSIT	:	Faculty of Computer Science and Information Technology
FPN	:	Fixed Pattern Noise
GLCM	:	Gray Level Co-Occurrence Matrix
GLNU	:	Gray Level Non-Uniformity
GLRLM	:	Gray Level Run Length Matrix
HGLRE	:	High Gray Level Run Emphasis
IQMs	:	Image Quality Metrics
JPEG	:	Joint Photographic Expert Group
K-NN	:	K-Nearest Neighbour

LBP : Local Binary Pattern

LCA : Lens Chromatic Aberration

LGLRE : Low Gray Level Run Emphasis

Lib-SVM : Library Support Vector Machine

LoG : Laplacian of Gaussian

LPQ : Local Phase Quantization

LRE : Long Run Emphasis

LRHGE : Long Run High Gray Emphasis

LRLGE : Long Run Low Gray Emphasis

MA : Moving Average

MLP : Multi-Layer Perceptron

MMS : Multimedia Messaging Service

NB : Naïve Bayers

OSN : Online Social Network

PCE : Peak to Correlation Energy

PR : Precision Rate

PRNU : Photo-Response Non-Uniformity

RLNU : Run Length Non-Uniformity

ROC : Receiver Operating Characteristics

ROI : Region of Interest

RP : Run Percentage

SGLDM : Spatial Gray Level Dependency Matrix

SIFT : Scale-Invariant Feature Transform

SMS : Short Message Service

SPN : Sensor Pattern Noise

SRE : Short Run Emphasis

SRHGE	:	Short-Run High Gray Emphasis
SRLGE	:	Short Run Low Gray Emphasis
SVM	:	Support Vector Machine
TPR	:	True Positive Rate
WDS	:	Wavelet Domain Statistics

CHAPTER 1: INTRODUCTION

This chapter provides the overview of this study which includes research definition, research background, statement of problem and research objectives. It also covers research questions, research scope and structure of each chapter in this thesis.

1.1 An Overview of Proposed Research

Online Social Networks (OSNs) is defined as web-based services that allow individuals to create, communicate, share and view their public profile and personal information with other OSNs users (Ellison, 2007; Narayanan et al., 2009; Subrahmanyam et al., 2008).

The first OSNs, named as Usenet was developed by Tom Truscott and Jim Ellis in 1979 (Suber, 2009). It is a worldwide discussion system that allows the Internet user to communicate and post public messages (Kaplan et al., 2010). The number of OSNs users has significantly increased year by year. OSNs provide significant advantages to connect peoples around the globe and ease of accessibility via various OSNs platforms such as Facebook, Twitter, Instagram and LinkedIn.

According to the latest study done by an online organization that consolidates information on social networks, (Statista Inc., 2017) as many as 2.4 billion OSNs users are being recognised in 2017 and this number is expected to grow further by 12.60% to 2.77 billion in 2019. The fast-growing number of OSNs users from socio-economic point of view is significantly good as people around the globe are having a more convenient way to communicate across the countries. However, most OSNs users have put full trust on OSNs application (Dwyer et al., 2007). Studies from Boshmaf et al. (2011) and Acquisti et al. (2006) discovered that most OSNs users are freely exposed their personal details about themselves, their friends and their relationships. They tend to share their personal photos, friends, family members and even their home addresses and phone numbers via the OSNs platform. Thus, the information was available for the public and accessible by other OSNs users around the globe (H. Gao et al., 2011; Joe et al., 2014).

The fact that, when personal information was uploaded into the OSNs platform, most of OSNs service provider has taken proactive measure to mitigate the risk. Various security features were developed including the option to set all information as private and can only be seen by limited friends and users. Unfortunately, despite having such security features, there are some users that still ignoring the privacy setting on OSNs website, which indirectly giving chances to strangers to commit offense (Erlandsson et al., 2012; Gross et al., 2005).

From a cyber risk point of view, the above scenario is considered unhealthy as users are exposed to digital crime. Various security attacks have been reported since OSNs was first introduced to the people, which includes identity theft (Bilge et al., 2009), malware (Baltazar et al., 2009; W. Xu et al., 2010) , privacy risk (Boshmaf et al., 2011; Mislove et al., 2010) creating fake profiles (some cases known as sybils) (Q. Cao et al., 2012; Stringhini et al., 2013) or socialbots (Boshmaf et al., 2011; Elyashar et al., 2013)) and sexual harassment (Wolak et al., 2008; Ybarra et al., 2008).

While OSNs service provider continues its effort to educate users on the importance of their security features, an improve mitigation and enforcement is required to fight the cybercrime. As such, in line with the authority's concern on cybercrime law enforcement and cybercrime case investigation, digital image forensics practitioners will further facilitate the process of finding evidence. An improved technique in digital image forensics and source camera identification is required in order to increase the possibility of prudent evidence to be used in fighting cybercrime especially on OSN images (Jang and Kwak 2015; Xu et al., 2016).

1.2 Background of the Research

In digital image forensics, source camera identification technique is applied to determine which camera was used to capture a particular image. It is also classified as the passive method as this method does not require any extra information such as a watermark or signature of the given image (Luo et al., 2007; Meng et al., 2008; Swaminathan et al., 2008).

Various techniques can be used to identify the image source such as metadata, lens aberration, sensor imperfection, Color Filter Array (CFA) interpolation and image statistical feature. However, these existing techniques are only focusing on original images and very limited studies on OSN images. Further details on these techniques will be discussed in Chapter 2.

With the increasing numbers of global OSNs user, the safety and confidentiality of digital images uploaded into OSNs platforms remain as a big concern. A number of publications highlight OSNs security issues including security and privacy concern involving user's data (Bilge et al., 2009; H. Gao et al., 2010; Huber et al., 2011; Jagatic et al., 2007). This fast growing of both OSNs platform and its users will directly increase the risk of cybercrime involving OSN images, which at the moment remain as one of the digital forensic areas that needs improvements to cope with the increasing cybercrime cases and play a bigger and more effective role in assisting the authorities to fight cybercrime.

Therefore, it is clear that digital forensics plays a major role to fight cybercrime. Information retrieved from OSN images could be crucial case evidence that helps authorities in taking closer steps to the right cybercrime culprit. To realise this, an improved digital forensics approach in source camera identification is paramount.

1.3 Problem Statement

Nowadays, thousands of advanced image editing tools are easily available either online or via mobile applications. There are all users friendly and easy to use. Even user with very minimal knowledge of image processing is able to create and modify the digital image content (Nance et al., 2009). Unfortunately, those image editing tools advantages causing more challenges to the authorities to get case evidence form OSN images in the event of cybercrime investigations. Images uploaded and downloaded from OSNs platform will undergone a pre-upload and download process which will change the original features and details of the image. Therefore, the authenticity of the images has become untrustworthy and as mentioned by Farid (2006); Zhu et al. (2004) "Seeing is no longer believing".

In digital image forensics, the main area of concern is deriving the most accurate information about an image. Various techniques were proposed to provide the most accurate answer to several questions including the originality of an image, the type of device manufactured and the source of the image taken from (Kharrazi et al., 2004). These questions are just a few examples of issues faced by the digital investigation and law enforcement agencies in their continuous effort to fight crimes, particularly cybercrimes. However, there are still very limited techniques that could help them in finding the most convincing answers for all the concern raised (Blythe et al., 2004; Luo et al., 2007).

The above also signify the same challenges in source camera identification where several techniques were used to identify the source camera of an image. However, these techniques were only focusing on information extraction from original images. Original images have not undergone any modification like reduced resolution and size compared to OSN images. OSN images, on the other hand, will be modified and processed by

Web 2.0 service tools. All its original characteristic and important information were either changed or totally removed by Web 2.0 during pre-processing into OSN platform (Castiglione et al., 2011). These remain as an area of concern in digital image forensic since the current technique is only focusing on original images characteristic and no techniques were proven accurate and suitable for OSN images.

For example, a study by Kulkarni et al.,(2015) is using texture feature to identify the image source from original images and has been proven capable to provide good results. However, this technique has a limitation which it does not address modified image, including OSN images. Therefore, to cope with the vast development of image editing tools and thousands of new OSN platforms increased year by year, source camera identification requires a significant improvement to address the limitation in existing techniques. This is to ensure that it remains relevant in supporting authorities to fight cybercrime. Hence, this research aims to propose a technique that improves the detection accuracy of source camera identification for OSN images. In addition, since texture features were proven to be able to provide good detection accuracy in original images based on previous research, the same method will be used in this research for OSN images. Further details of the method will be described in Chapter 3.

1.4 Objective of the Research

The main objective of this research is to propose a technique for source camera identification on OSN images. The details are listed as follows:

a. To study the current source camera identification techniques

b. To propose a suitable technique for source camera identification targeting on OSN images.

c. To evaluate the proposed technique in terms of its accuracy.

1.5 Research Question

Most of OSN platform has limitation, particularly in space and storage. To maintain its maximum performance over the time and to cope with fast-growing users, they have to make sure optimum space management but at the same time making sure a very minimal image quality reduction. As such, all original images will undergo a pre-uploading process which will modify or even erase some characteristic from the original images. For example, an original image with resolution 4032 x 3024 and size 1.5MB will be pre-processed before successfully uploaded into OSN platform and the resolution will be reduced to 1008 x 756 and size 95.5KB. This proved the challenges in source camera identification on OSN images because some of its crucial original characteristics were altered or erased during the pre-uploading process into the OSN platform. Hence, it signifies the needs of an improved technique for source camera identification on OSN images. Details of these research objectives and its research questions to be addressed are listed as follows:

Objective 1: To study the current source camera identification techniques

RQ 1: What is the state-of-art for source camera identification?

RQ 2: What are the current detection techniques available for source camera identification?

Objective 2: To propose a suitable technique for source camera identification targeting on OSN images

RQ 3: How to improve the existing techniques to get higher detection accuracy on OSN images?

RQ 4: What are the fundamental elements required to increase the detection accuracy on OSN images?

Objective 3: To evaluate the proposed technique in terms of its accuracy

RQ 5: What is the performance level of the proposed techniques against existing techniques?

1.6 Scope of the Research

The main scope for this research is to propose a technique for source camera identification on OSN images by using a combination of Gray Level Co-Occurrence Matrix (GLCM) and Gray Level Run Length Matrix (GLRLM) features, named as Fusion feature. Both features were selected due to its effectiveness and ability in identifying and classifying source images as proven in previous research.

Meanwhile, Faculty of Computer Science and Information Technology (FSCIT) dataset and Dresden dataset were chosen for this research experiment. Further details on these datasets will be deliberated further in Chapter 3. For the experiment tools, Matlab 2014b was used to execute the main program and Weka 3.6 was used in the classification process.

1.7 Organization of the Thesis

To make sure the end to end process in achieving the research main objectives is clearly explained, this thesis was segregated into six chapters which details are listed as follows:

The **"Introduction"** chapter describes the idea, problem and the objective of this research.

"Literature Review" provides the background information about online social network, digital forensics, image forensics techniques and texture feature extraction techniques. This chapter also presents the studies on source camera identification techniques.

Meanwhile in "**Research Methodology"** it describes the requirements in designing the proposed technique using OSN images. The design of the proposed technique is presented followed by the discussion on the requirement of the research and listing the tools used in the experiment.

"The Proposed Technique" chapter describes the proposed techniques for source camera identification using texture feature on OSN images. There were two texture features selected to extract the image features which are GLCM and GLRLM feature. The GLCM used 22 sets of features properties while GLRLM used seven sets of features properties during the extraction process. This technique was also tested using original images to verify that the prose technique is also capable in source camera detection for both original and OSN images.

The "**Evaluation Experiment and Result Discussion**" chapter describe the evaluation and discussion of the experimental results. This chapter explains how the results of the proposed technique fulfilled the objectives mentioned in Chapter 1.

In **"Conclusion"** chapter it concludes and summarises the research contribution. The achievements and objectives of the research with respect to the experimental results are highlighted along with the findings and significance of the research. This chapter also discusses the future research direction, its impact and benefits to the society in general.

CHAPTER 2: LITERATURE REVIEW

This chapter provides an overview of the current literature pertaining to the relation on OSNs and digital forensics. According to Ellison (2007), a social network is a web service that allows users to construct a profile within various abound systems, articulate list of other users among the shared connections and view the list of user connection within the system. Meanwhile, the definition of digital forensics by Palmer (2001) is the use of scientifically and proven methods towards the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence that derives from the digital devices for the purpose of furthering the reconstruction of event that has been found to be faulty

This chapter is organized into four main ideas which include: i) basic idea of Online Social Network ii) knowledge of forensics science ii) basic idea of digital forensics iii) techniques available in digital image processing and iv) explanation of texture features extraction.

2.1 Online Social Network Background

Nowadays, technology continues to thrive day by day in order to provide a superior, quicker and effective platform to assist individuals by connecting and disseminating ideas, emotions and information to other individuals. The emerging of technology such as social media has become popular not only with the users but it has been the main topic among the researchers to identify the implication of this new technology.

OSNs is a web-based service that allows a user to create a profile within a bounded system that creates a list of others users with whom they are connected and enables the user to cross over other connections made by some other users in order to see their profiles and contents (Dwyer et al., 2007; Ellison et al., 2007b; Fogel et al., 2009;

Gagan Mehra, 2015). It is a platform that allows users to create a social relation among the OSN users that share similar interests, activities and real-life background. OSNs are able to provide users with various mechanisms such as manage, construct and making the content and profile visible. It also allowed users to organize the type of connection with other users and interact by sharing their content and information or even altering their own profiles (Adamic et al., 2016; P. Wang et al., 2015).

There are about more than 200 OSNs applications that are active and their popularity depends on the certain geographical regions (Vincenzo Cosenza, 2012). Figure 2.1 represents the world map of the social network, which was recorded in January 2017. The *QQ* and *QZone* are the most popular social networks in China with the number of users that has over than 1400 million per month. Facebook in the other way still remains as the leading social network in 119 out of 149 countries that used the application as their main social media.

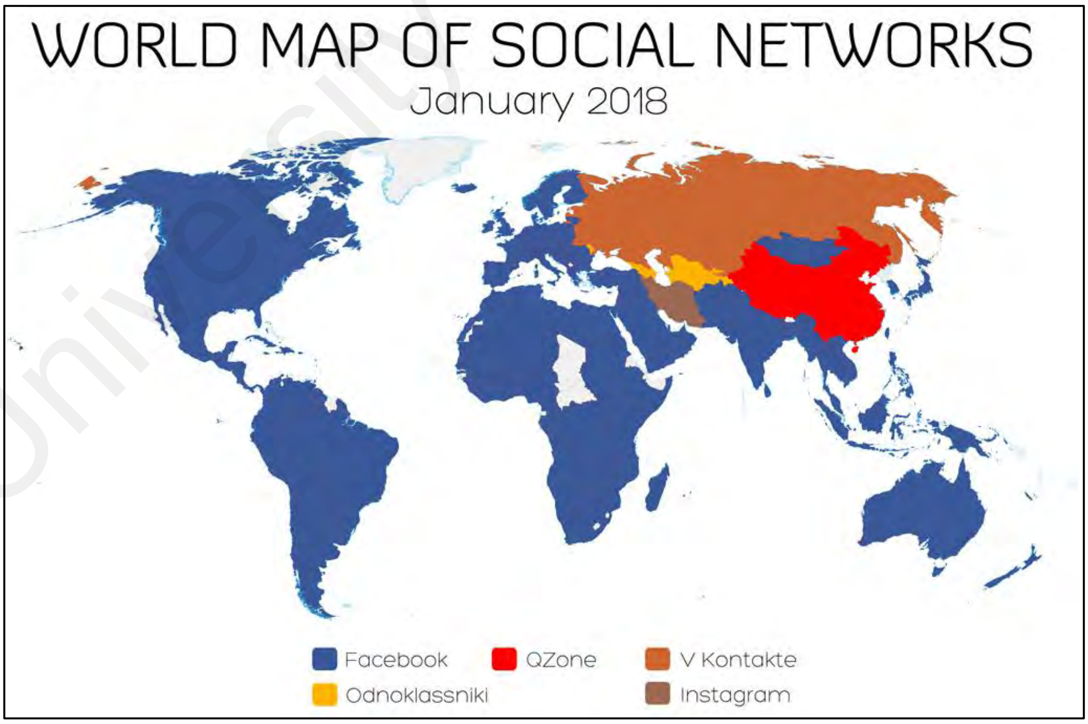


Figure 2.1: World Map of Social Networks on January 2018*

OSNs can be differentiated into several categories based on the basis of the content and target group. For example, Facebook and LinkedIn are targeting to engage a

*Vincenzo Cosenza (2012) – Vincos Blog; World Map of Social Network personal and professional user connection with video or voice chat. Twitter on the other way around provides a service for users to voice out their opinions and sharing breaking news and events. Furthermore, it plays the role as the fastest platform to initiate relief efforts during disasters. The two-way interaction enables users to broadcast the message or information immediately and directly to the target users.

Studies found that OSNs users are unaware and do not have sufficient knowledge about the malicious threats and do not even notice that their data have been used or exposed by the other person (Buccafurri et al., 2015; Karavaras et al., 2016). Thus, it has raised concerns about privacy among the users, academicians, law enforcement as well as policymakers. Since that, OSNs have become an important research study in the area of privacy research and a considerable amount of the literature review on privacy issues has been raised (Abdalla et al., 2014; Krishnamurthy et al., 2009; Valsesia et al., 2015b).

2.1.1 Security and Privacy Issue in Online Social Networks

Security and privacy are always the main concerns and issues for users and digital forensics practitioners. In practice, OSNs are applying a data security technology to ensure data or information of the users is stored, secured and unreadable to the unauthorized OSNs users. The objectives of security and privacy in OSNs platform are integrity, availability and privacy as shown in Table 2.1.

Table 2.1: OSNs security and privacy objective

Objective	Description
Integrity	The identities and data of OSNs users have to be protected and secured against unauthorized users' modification.
Availability	OSNs users' data have to be available whenever they are needed.
Privacy	OSNs users' information has to be kept secret from the third party either internal or external to the system.

Integrity is referred to the services provided where users are given complete trust to make sure their data are safe and have not been modified and will be retained as its original data (Hajli et al., 2016). Accessibility in OSNs provider is to ensure that authorized users can access and modify the data at any time and place. OSNs provider has to make sure that the level of privacy is high in order to protect and make sure that the data are appropriately used for information. In OSNs, users' information is also known as information privacy. It applies the technology where the individual data in OSNs can be shared with the third party for legal purposes. Since the popularity of Online Social Networks is continuously growing, users, consumers and researchers are starting to express their concern on security and privacy-related issues. Several researchers have been made in order to identify the behavior and trend of security and privacy in OSNs web. List of paper discussing the security and privacy in OSNs are listed in Appendix 1.

Throughout the years of 2007 until 2017, the diversity and numerous research studies have been done on how the crime networks operate, how the social network attacks are designed, what are the damages caused by the attacks and what are the solutions to improve the security and privacy in OSNs.

2.1.2 Online Social Network and Forensics

In the early twenty-first century, digital forensics was easy to conduct because of the advances and sophisticated technology devices in today's world. Because of the convenient accessibility of the smart devices, almost all of the data are now being shared and stored in a digital form. Pictures, diaries, calendars, video or even daily schedules are stored in a digital form. The demands for the use of technologies are to make all of the data to be stored and shared freely on OSNs web.

This exposure of data and information can be the potential evidence for a crime or incident. Raghavan (2013) studied that social media data is one of the recent developments that attracts everyone in any types and range of social status and it was the platform that produced a large amount of digital data used for analysis. According to Nelson et al. (2015), any type of information that are stored or transferred in a digital form can be identified as digital evidence. The vast used of OSNs and availability of supporting devices have attracted significant forensics analysis on OSNs. It has also increased the numbers of related digital crime involving OSNs. Thus, this is where digital forensics came into the picture. A recent study by Karabiyik et al. (2016) has shown that digital forensics plays the main contribution in investigating the misused and misbehaviour in OSNs.

In digital forensics investigation, social media data may contains an invaluable evidence for the process of investigation. The main goal of investigating the social media data is to understand relations between the actors for the variety of purposes such as solving the criminal activities, preventing terrorist attacks, detecting deceptions, categorizing and matching social networks accounts (Karabiyik et al., 2016). Hence, a need for digital investigators and researchers to understand and be familiar with all the network activities is to confront the related issues and challenges.

2.2 The Basic Knowledge of Forensics Science

Forensics science is referring to a broad range of scientific techniques that are used to analyse the physical, biology and digital data for legal purposes. Commonly, forensic science is the application of science that is implemented in the law within the criminal justice system.

The word forensics is from Latin word, Forensis which means debate or public discussion. In Roman times, the word Forensics presents the case in a group of a forum.

Generally, forensic science is a public presentation of providing fact using a scientific method in a court of justice. Figure 2.2 illustrates the existing branches under forensic science.

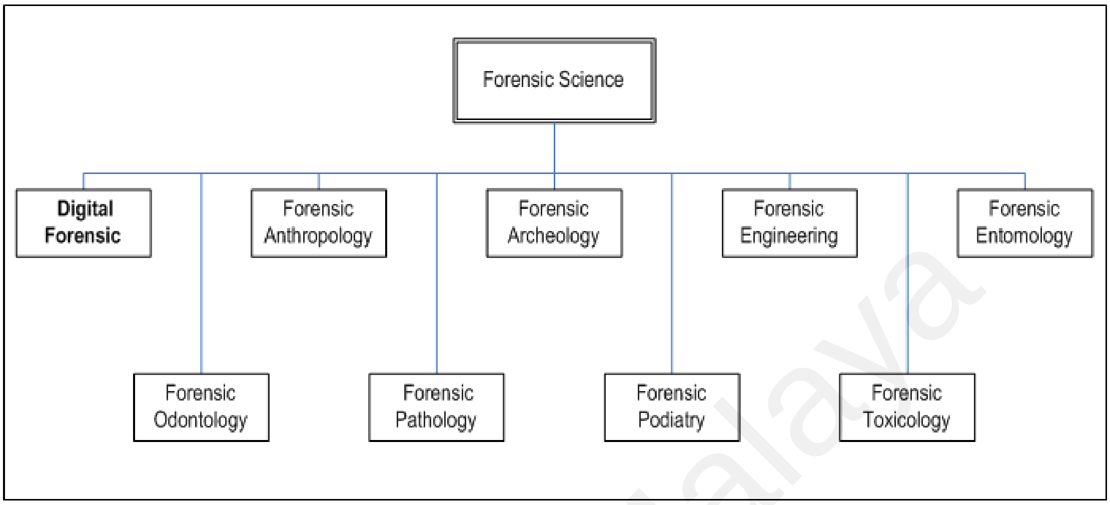


Figure 2.2.: The Hierarchy of Forensics Science

According to this hierarchy, digital forensics is one of the branches listed under forensic science. Details on digital forensics branch were explained in the next section where this thesis focuses on digital forensics that involves the use of scientific techniques and process to validate the authenticity of the given digital evidence.

2.2.1 Introduction to Digital Forensics

Digital forensics is a field of computer science. It related with computer disciplines for presenting the facts in a court of laws. People always think that digital forensics is a cybercrime. Cybercrime is a crime activitie in which a computer and tools that used to commit the crime. It is including phishing, scamming, hacking and child pornography. (Halder et al., 2012). These activities can use as a digital evidence to support or refute the theory of how the offence occurred and also address the critical elements of the offence. (Davis et al., 2004; Morgan, 2016).

The goal of forensic discipline is to have an appropriate procedure in conducting the investigation. But, there is a less proper procedure or consistent method for digital investigation. The current digital forensics investigation procedure is based on the combination of experience from law enforcement, information technology specialists and hackers that has evolved over the time (Reith et al., 2002). The basic process of forensics investigation consists of four phases such as collection, acquisition, analysis and reporting (Ballou, 2010; Carrier et al., 2004; Garfinkel, 2010). These four phases are implemented as guidance for law enforcement in order to protect digital evidence for the recognition. Figure 2.3 illustrates the process involved in a digital forensics investigation.

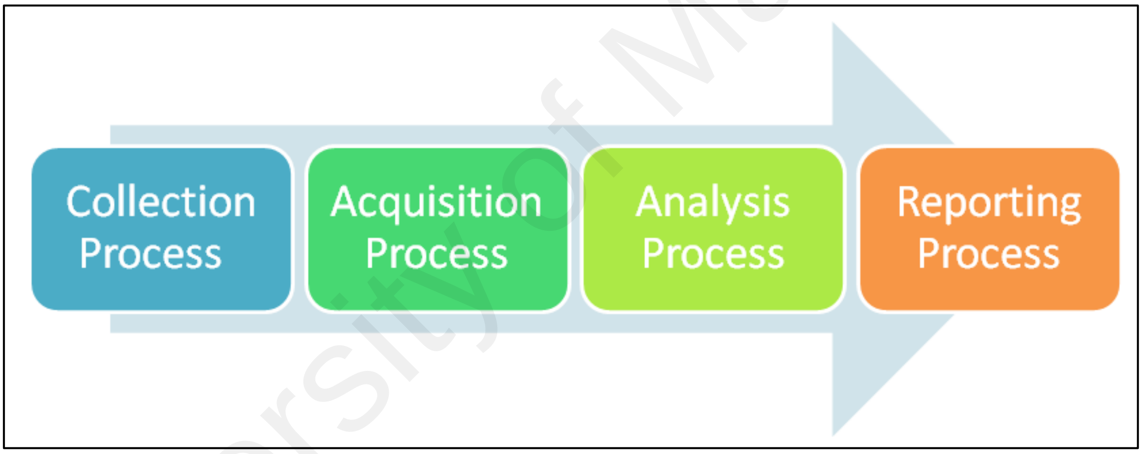


Figure 2.3: Digital Forensics Investigation Process

The procedure starts with the collection of evidence from the scene during the investigation. It is an actual examination where digital media is being seized. Once the evidence is been seized, there is a process of duplication referring to acquisition process where the original evidence is acquired using software imaging tools such as EnCase, FTK Imager or Ditto Forensic FieldStation. The original drives or evidence are returned to secure the storage to prevent a tampering. Then the image files are analysed in order to identify the evidence that used to support the hypothesis in the crime scene (Carrier, 2003). There are various types of techniques used to recover the evidence. For example in keyword searching, the image files can be either identify the relevant evidence or to filter out the unknown files types (Casey, 2011). Once the evidence is recovered, all the related information is analysed in order to reconstruct the events or action that reach to the conclusions (Quick et al., 2014). The process continued with the completed information reported in a suitable form. It includes audit information and other data documentation. The reports are then passed to the commissioning investigation such as law enforcement that is used in court.

In spite of being in use for over a decade, there are significant problems facing by the digital industry with regards to digital forensics. The growing needs to secure the data and information that reside digitally, there are issues regarding how digital evidence is been carried out and accepted in courts. Thus, the digital forensics field has emerged into several branches. According to Battiato et al. (2010), the emerging of digital forensics is because of the huge amount and types of digital data that are being produced every day.

Multimedia forensics is one of the newly emerging fields and is starting to grow over the past few years. This field provides techniques to test the authenticity and integrity of multimedia content such as digital images, audios and videos (Jatinder Kaur, 2012). Figure 2.4 illustrates the emerging of digital forensics field.

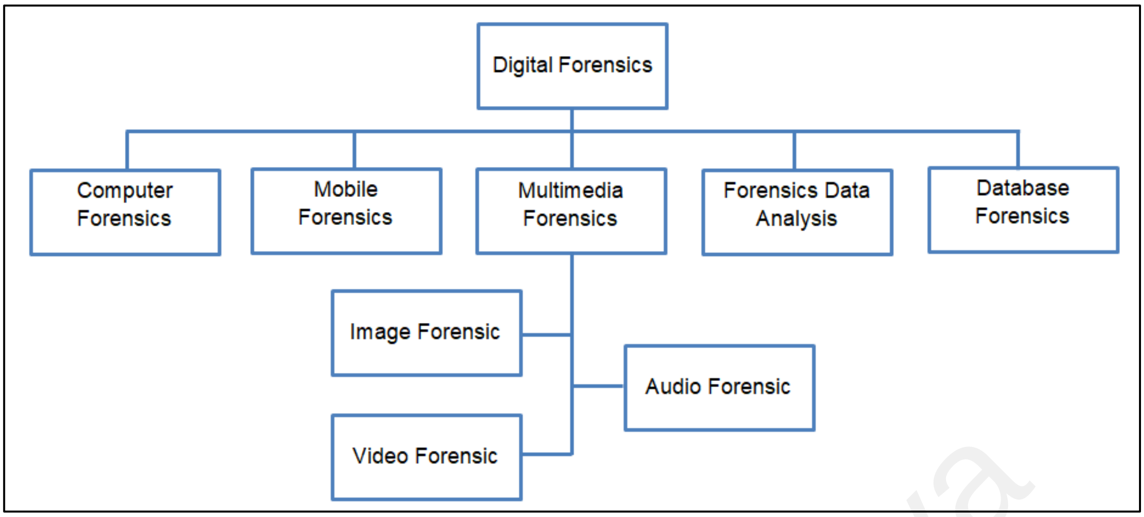


Figure 2.4: Branch of Digital Forensics

The work presented in this thesis puts forward the field of image forensic as a result of the vast emerging of the digital evidence which is also known as a digital image that continuously increases their usage in OSNs nowadays.

2.2.2 Image Forensics in Practice

Digital image and video play a big role in most of the applications such as video surveillance, e-news, social network and others. The availability of sophisticated image processing tools that allows the editing of the digital images and videos has led to the multiplication of fake images (Farid, 2009b). Therefore, the originality of multimedia data such as digital images and videos are doubtful. In order to restore the trust towards this data, image forensics or digital image forensics has become the most suitable branch to overcome the issue.

There are two main research domains under digital image forensics. It is source identification (Ahmet Emir Dirik et al., 2008; San Choi et al., 2006) which the objective is to prove the given digital images or videos that are taken by a specific camera model. Second is semantic forgery detection (Avcibas et al., 2004) which the objective is to discover malicious manipulation given by digital image or videos.

The main objective is to resolve various issues regarding integrity assessment, digital image audio and video authentication. According to Piva (2013), several approaches have been proposed in digital image forensics. They are classified into active and passive approaches. The term ‘active’ approach refers to the assessment of trustworthiness where some of the information has been exploited during the acquisition process. Figure 2.5 illustrates the active and passive approaches based on the reliability of the camera (Blythe & Fridrich, 2004; Friedman, 1993) where it is the way to get the authenticity of the digital images.

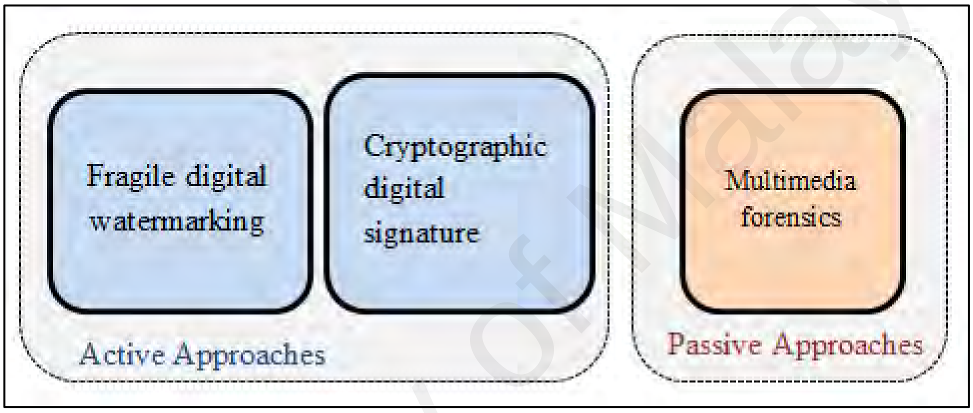


Figure 2.5: Possible Approach in Digital Image Forensics

The modifications or changes of the digital images can be detected by comparing the value of digital watermarking (Barni et al., 2004; I. Cox et al., 2007; I. J. Cox et al., 2002) or digital signature (Menezes et al., 1996; Rivest et al., 1978) which is presented in the digital images.

Passive approach relies on the observation of the digital image formation from the acquisition process till the storing processes (Farid, 2009a; Mahdian et al., 2010; Van Lanh et al., 2007). Each of the processing operations leaves races or digital fingerprints. It is then used to identify the source of the digital image or determine whether it has been undergone a process of modification to the digital image content.

Digital image forensics analyses an image by using processing science technique to identify the source of the images or to detect the manipulation of the image. Source identification is focusing on identifying the source digital devices while forgery detection is attempting to discover the modification to the digital media. As for writing purposes of this thesis, source identification on digital image is the main interest to be focused.

2.2.3 The Formation of Digital Image

Formation of digital image can be presented in three main phases; acquisition, coding and editing. The formation of digital image is illustrated in Figure 2.6.

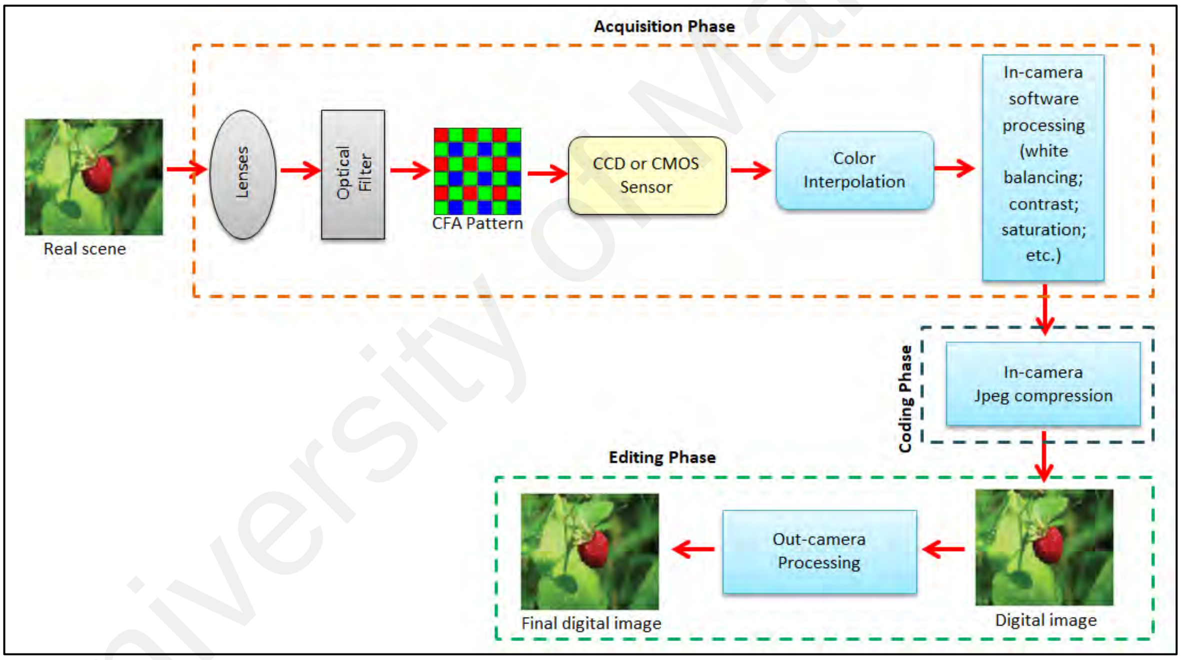


Figure 2.6: Digital image generating process
(Piva, 2013)

In acquisition phase, the light intensity is measured at each pixel and performing a raw image that contains information which is recorded by the camera sensor Charged Couple Device (CCD) or Complementary Metal Oxide Semiconductor (CMOS)) generating the digital image signal. However, before reaching the camera sensor, the light is filtered by the CFA (Color Filter Array) pattern which contains a thin film that selectively a certain component of light to pass through to the sensor. Each pixel contains one main color that consists of red, blue, and green. Through the process called demosaicing, the output from the sensor are interpolated in order to obtain the three main colors for each pixel in the digital image. The obtained digital camera signal, are undergoes an additional pre-processing in-camera that includes white balancing, color processing, image sharpening, contrast enhancement and gamma correction.

In the coding phase, the processed signal is then stored into the camera memory in the form of JPEG format. Finally, the generated image can be pre-processed, for the modification and enhancement of the image contents.

The understanding of these three main phases helped the investigator to trace the evidence or digital fingerprints left during the processing digital image. A study has found that each of the phases (acquisition footprints, coding footprints and editing footprints) in image formation left traces or footprints to the digital image (Piva, 2013).

In acquisition footprints, each of the acquisition components left an intrinsic fingerprint at the final stage of the image output. The patterns of fingerprints depend on the use of specific optical systems, image sensor and camera software. As a result, these footprints were used to differentiate between the kind of devices, the brand or even the model of the digital devices. Moreover, these footprints can also be used to detect forgeries to the digital image by looking at the consistencies of the acquisition feature in a different region and compared it with the same digital image. There are multiple aberrations existing and each of it has a unique feature. For instance, chromatic aberration is responsible for colouring the edges along boundaries by separating the dark and the bright part of the images. With these aberrations, it has derived several techniques to leverage the artefacts for source identification (Deng et al., 2011; A Emir Dirik et al., 2007; Ahmet Emir Dirik et al., 2008; S. Gao, Xu, et al., 2012; Huang et al., 2015; San Choi et al., 2006; Van et al., 2007) and it is also used for tampering detection (Chang et al., 2013; W. Chen et al., 2007; Johnson et al., 2006; Mayer et al., 2016).

Due to the usage of JPEG compression format for efficient storage in digital camera, the research community has been dedicated to discuss and study the compression history of the digital image. JPEG compression includes blocking artefacts, mosquito noise (around edges) and colour degradation. Blocking artefact, for example, it is caused by the underlying block-wise approach and quantisation artefacts. With this artefact, it can estimate the compression parameters by quality factor (W. Fan et al., 2013) and quantisation tables (Z. Fan et al., 2000; Luo et al., 2010). A study by Farid (2008) found that, quantisation table can be used for source identification (Alles et al., 2009; Choi et al., 2006; Liu et al., 2013) and forgery detection(Bianchi et al., 2012; Kee et al., 2011; Lin et al., 2011).

Currently, there are a lot of editing tools that are freely available in the markets or even in the online website. With the numerous usages of editing tools, it can be used by any age and any type of user's social background. It can be used in any appropriate away for the purpose of enhancing the quality of the image or it can be used in any inappropriate way to alter the content of the digital image. One of the most common editing processes is resampling, which it performs a geometric transformation like rotation and resizing. The process is transforming the image from one coordinate system to the other image. These two coordinate systems were related to each other by mapping the function of the spatial information (Kirchner, 2008; Mahdian et al., 2008; Popescu et al., 2005). There are a lot of studies and research on editing footprint which is used either for source identification or forgery detection. For example, contrast enhancement which is a very common manipulating technique is used to increase the quality of the images. The image can be manipulated by darkening the scenery in order to convey a sensation of danger. According to Stamm et al. (2008), global contrast enhancement can be detected by seeking a unique artefacts is that is introduced into an image's histogram.

2.3 Source Camera Identification Techniques

Digital image forensics has been growing in the past few years. A lot of studies and researches have given a concern on the evolution of the digital image forensics area. There are three main types of work that the researcher has focused on; image source identification which aims to determine the sources devices that are used to generate the digital images, discrimination of synthetic image from real image to computer generated image and image forgery detection is to determine whether a given image has undergone any manipulation or modification process (Sencar et al., 2008).

Source camera identification techniques consist of three categories; metadata-based, watermark-based and feature-based. Metadata-based relies on investigating the image source that is related to the information that is embedded in the digital image. The information that is available in metadata includes camera brand, model, date and time when the image was captured. As this information can be easily obtained from the image, this metadata can be easily manipulated.

The watermark-based approach embedded the related information in the digital image. However, these watermarks need to be inserted during the image processing and it involved complicated processing and increases the production cost of digital cameras.

Thus, this leads the researchers to develop the feature-based approach which extracts features on hardware artefacts or software artefacts. This approach is divided into two groups; hardware-related artefacts and software-related artefacts. Hardware-related artefacts includes sensor noise (A Emir Dirik et al., 2007; Geradts et al., 2001; C.-T. Li et al., 2012; Lukas et al., 2006; Maini et al., 2009; Sutcu et al., 2007; Valsesia et al., 2015b; Wu et al., 2012), lens chromatic aberration (Conners et al., 1980; Hall et al., 2009; Van et al., 2007) and color filter array interpolation (Bayram et al., 2005; Bayram et al., 2006; Long et al., 2006). Meanwhile, software-related fingerprints consist of image-related features proposed by Kharrazi et al. (2004). Details of feature-based approach are elaborated in the following sections.

2.3.1 Feature Based Techniques on Sensor Pattern Noise (SPN)

Sensor Pattern Noise (SPN) is caused by the imperfections during the manufacturing process and non-uniformity of photo-electronic conversion due to the inhomogeneity of silicon wafers. It consists of two main components; Fixed Pattern Noise (FPN) and Photo-Response Non-Uniformity noise (PRNU). FPN is produced during the pre-processing stage due to the dark current compensation. FPN is not a robust fingerprint and it is no longer used in any future work.

PRNU is the main component in SPN. It comprises two parts; Pixel Non-Uniformity (PNU) and low-frequency defects. Low-frequency defects are generated through light refractions on dust particles, the optical surface and zooming setting. It does not get affected by the characteristics of the sensor. Meanwhile, PNU is a more dominant component and it is generated based on the sensitivity of pixels. The sensitivity pixel is measured by determining the intensity light, the effect of inhomogeneity of silicon wafer and imperfection of the sensor manufacturing process. Hence, it is the only component in SPN that perfectly contributes to the source camera identification technique.

The initial work by Kurosawa et. al (1999) proposed FPN for source camera identification. In their work, the blank images with low intensity are tested and it can only survive in a dark scene. Although the work is successful, it is still a weak signal and it is not robust for natural image

Geradts et al. (2001) techniques were examining the CCD pixel defect that includes points defect, hot points, dead pixel, pixel traps and cluster defects. The result found that each camera had a different defect pattern. It also notes that, the numbers of defect patterns in the pixels for a camera is different between pictures and varies depending on the content of the images and different level of temperatures. This study has found that cameras with high-end CCD do not face this problem, thus it shows that not all cameras suffered from this problem. It concludes that, most digital cameras contain additional mechanism to reduce this kind of problems.

Lukas et al. (2006) proposed a method to extract PRNU by denoising the original image with a wavelet denoising filter. The PRNU from a tested camera was obtained by averaging the PRNU image. Then it normalizes the correction coefficient between PRNU of an image and PRNU from the camera reference. This process is to determine whether the image is from the tested camera. An improvement method from Lukas's method has been proposed by Sutcu et al. (2007). By incorporating the camera's demosaicing characteristics into the decision process where it can enhance the reliability of the decision and increase the accuracy of identification. According to Li et al. (2012), color filter array may lead to inaccurate extraction of PRNU. Thus, they proposed a new extraction method that decomposes each of the color channels into four sub-images. Then extracts the PRNU from each of the sub-image and then assembles them. There is another researcher that proposed source camera identification techniques based on SPN.

In the frequency domain, SPN can be largely contaminated by image content and non-unique artefacts of JPEG compression, on-sensor signal transfer, sensor design and colour interpolation. Hence, source camera identification performance that is based on SPN has to be improved for the small size of images and in resisting JPEG compression. Motivated by Goljan et al. (2009) work on test statistics Peak to

Correlation Energy (PCE), Kang et al. (2012) proposed a work that used correlation to Circular Correlation Norm (CCN) as the test statistics which can lower the false positive rate to be half of the former. According to Wu et al. (2012), assuming that SPN is a white signal, they extracted SPN directly from the spatial domain with a pixel-wise adaptive Wiener filter. The result shows that the method is satisfactorily achieved the Receiver Operating Characteristics (ROC) performance among all the state-of-the-art camera source identification techniques. It is also resistance to JPEG image compression with quality factor of 90%. The summary of the techniques are listed in Table 2.2.

Table 2.2: Techniques based on Sensor Noise

Author	Contribution	Limitation
Kurosawa et al. (1999)	- Detecting fixed pattern noise that caused by dark current.	- Noise changes with temperature and time. - FPN fixed for a specific sensor - Not robust
Geradts et al. (2001)	- Examining CCD pixel defect as a unique fingerprint for source camera identification Considered hot / dead	- High-end digital camera rarely produces pixel defects
Lukas et al. (2006)	- Extracting the PRNU by calculating the correlation between noise residual of specific image and camera reference pattern	- Sensitive to appropriate synchronization - Not suitable for modified images
Sutcu et al. (2007)	- Enhanced the SPN by verifying the image processed by the demosaicing algorithm to avoid FPRs.	- Unique for different models and brands but non- unique for cameras with the same model - Less accurate in identifying the same camera model

2.3.2 Feature Based Techniques on Lens Chromatic Aberration (LCA)

Lens Chromatic Aberration (LCA) can occur either when a lens is not capable to bring all the wavelengths colour to the same focal plane or when the wavelengths of colour are focused on different positions in the focal plane. It is also caused by lens dispersion where different colours of light are travelling at different speed whilst passing through the lens. The image looks blurry with the appearance of round objects (with coloured red, green, blue, yellow, purple and magenta) especially with high-contrast situation. LCA can also be applied for forgery detection which has been used by Johnson and Farid (2006) , Yerushalmy et al. (2011) and Mayer and Stamm (2016).

Van et al. (2007) was using an iterative brute-force search where it estimated that three parameters (α x, y) among R and G, also B and G channels through maximising the mutual information. This method causes the reduction of computation time but it has increased the error rates. This is because of the aberrations that are more vising along the image edges.

Gloe et al. (2010) has proposed a technique that can detect a suitable block from the corner points and local estimation for LCA analysis. It works by assuming that each block was a sub-sampled by each factor of u in order to allow the sub-pixel estimation. The factor was estimated carefully in order to increase the accuracy of the estimated displacement vector. It is used to determine the maximum displacement vector for both x and y direction and cropped the reference colour channel for each of the sub-sampled block. The model parameter that was based on the locally estimated displacement vector, were fitted by iterative Gauss-Newton scheme (Mallon et al., 2007). Therefore, this proposed method has successfully reduced the computation time and increased the error rates. On the others hand, the authors have found that there is limitation due to high inter-model similarity of LCA parameter and large variations with different focal length in large scale settings.

Yu et al. (2011) have developed the stable LCA parameter in order to identify lens of the identical in DSLR camera with the inter-changeable lens. It works by using the white paper printed version as a shooting target to eliminate the misalignment due to the small camera movement and corner detection algorithm. This work showed that the reliability of the practical lens forensic depends on maintaining the granularity of focal length and focal distance against increasing the number of lenses to be distinguished. The term granularity is defined as a specific quantisation scale where the lens LCA patterns can be considered as unchanged. The evaluation method is performed by using a mismatch plot of LCA pattern corresponding to all available images by assigning the LCA pattern from one specific image as a reference pattern. The summary of the LCA based techniques are listed in Table 2.3.

Table 2.3: List of LCA based Techniques

Author	Contribution	Limitation
Van et al. (2007)	- Estimate three parameter among *R,G,B* channels through maximizing the mutual information - reduce computation time	- Increase error rates
Gloe et al. (2010)	- Detecting the block from the points and local estimation for LCA analysis	- It is limited to high inter-model due to the similarity of LCA parameter - Less accurate to estimate the optical center - Increase the error rates in identifying the source camera

Yu et al. (2011)	- Identifying the identical lens of DSLR with inter-changeable using LCA parameter (DSLR) with interchangeable lens.	- Depends on the maintaining the granularity of focal length and focal distance. - Large variation of LCA parameters with different focal length negatively influences camera model attribution in large scale settings parameter

2.3.3 Feature Based Techniques on CFA configuration

This method is depending on the fact where most of digital cameras have a Charged Couple Device (CCD) sensor in order to render colour and light where it should be filtered by CFA before it reaches to the sensor. CFA is a pattern arrangement that depends on the manufacturer of digital camera device. It produces a mosaic colour consisting of red, green and blue pixels arranged on a single layer. In order to obtain these three-channels, the signals need to be interpolated. This interpolation model leads to periodic characteristics among image pixels. Since the interpolation models are used differently by digital camera devices, this model can be used for source camera identification techniques. There are two types of CFA interpolation, first is inter-pixel correlation pattern and second is inter-channel correlation pattern.

Bayram et al. (2005) work, used Expectation –Maximization (EM) algorithm to estimate the interpolation coefficients which is designed the amount of contribution from each pixel in the CFA interpolation kernel. A set of the interpolation coefficient that was obtained from the image and the peak location also magnitudes in frequency spectrum were used to design the classifier in order to differentiate the source of camera. The result showed that this method was not working well with heavily compressed images. Therefore, Bayram et al. (2006) extended this work and had improved the method by capturing the periodicity in the second order derivatives on smooth and non-smooth parts of the images separately.

In another techniques, Long and Huang (2006) presented a quadratic pixel correlation model where each of the colour channel obtains the coefficient matrix and extracted the principal component and fed to feed-forward back propagation neural network for source camera identification. The result showed that the method was valid and practical to be used since it does not depend on technical details of the demosaicing algorithms. It was also useful when demosaicing algorithm were complicated to detect.

A method that has identified the CFA pattern and estimated the interpolation parameters was based on minimum mean-square estimation which has been resented by Swaminathan et al. (2007). They used the estimated population coefficients that provided them with useful features that can differentiate the model and brand of digital camera from the given image. Recent work by Hu et al. (2012) have presented an improved work by Ho et al. (2010), where variance map was extracted by estimating the variances of each component of green-to-red and green-to-blue spectrum differences. Meanwhile, the shape and texture feature of the images were obtained for the camera model identification. Table 2.4 represents the contribution and limitation of the available techniques.

Table 2.4: CFA Interpolation techniques

Author	Contribution	Limitation
Bayram et al. (2005)	- Estimate the interpolation coefficients using Expectation Maximization (EM) that can differentiate the source of the camera	- Less detection in smooth and non-smooth parts of the image - no information regarding on the size of the kernels

Long et al. (2006)	- Used quadratic equation and reduce the coefficient matrix to perform demosaicing algorithm	- Large dimensionality - Characteristics of interpolation algorithm need to follow the original group
Swaminathan et al. (2007)	- Provide the useful features that can differentiate the model and brand of digital camera	- Only detect intra-channel correlation - Increases the cost of computation due to the search space. - Content dependency, location dependency and effect of post-processing operations
H. Cao et al. (2009)	- Estimate the intra-channel and cross-channel correlation	- Computation cost - Post-processing is affected due to content dependency and location dependency
Fogel and Nehmad (2009)	- Reduce the negative impact of JPEG compression in linear CFA interpolation model for double JPEG image	- Computation cost - Cropping increases the probability of the block contain texture region or edge and make the technique confusing - Causes the information of CFA pattern damaged by the lossy JPEG compression

There are two limitations for this kind of method. First is the high computational complexity due to the unknown CFA pattern of an image where it needs to be tried for its coefficient estimation during the detection process for various CFA models. Secondly, both of the inter-pixel and inter-channel correlation patterns are sensitive to JPEG compression. Since the compression can be considered as a local homogenization and attenuates characteristics of local correlation pattern, therefore this method is not suitable for detection accuracy particularly on JPEG compression images.

2.3.4 Feature Based Techniques on Image Statistical Features

Each camera model and brand have differences in CFA interpolation algorithm and parameter design. Hence, images taken by different camera have different kinds of statistical characteristics. Therefore, image statistical feature can be used to identify the source of the image.

Preliminary work based on image statistical feature was undertaken by Kharrazi et al. (2004) introduced 34-dimensional features including colour characteristics, Image Quality Metrics (IQMs) and Wavelet Domain Statistics (WDS). Feature of colour characteristics was determined through a colour production with respect to each colour band. Meanwhile, IQMs measured the quality of the scene reproduction by optical systems and wavelet domain statistics quantify a sensor noise which was used to process using wavelet transformation in order to extract the singular frequency component of the image. The three main features were then fed to Support Vector Machine (SVM) classifier in order to identify the source camera.

However, B. Wang et al. (2009) have proposed an effective approach by extracting the 216-dimensional higher-order wavelet features and 135-dimensional wavelet coefficient co-occurrence feature from tested images and applied the Sequential Forward Feature Selection (SFFS) and correlation. Then, it is fed to the SVM classifier to identify the image source. The experimental result shows that the accuracy of this method is able to achieve as high as 98%.

A combination feature from Kharrazi et al. (2004) and Gou et al. (2009) method by Hu et al. (2010) and Y.-C. Chen et al. (2011) has formed the 102-dimensional feature vectors. The identification worked by feeding the features to SVM classifiers. A new approach has been developed by Wahab et al. (2012) using Conditional Probabilities (CP) features. CP features from the selected block-wise Discrete Cosine Transform

(DCT) coefficients were exploited. The method was tested with 400 images captured by four different models of cameras and the result shows the accuracy was able to reach 99.50%. In addition they also tested with cropped and compressed images and result is slightly dropped to 97.75%. Method based on multi-step transition has been presented by S. Gao, Hu, et al. (2012). It worked by using multi-step transition probability matrices to model a different JPEG 2-D arrays along seven directions. The combination of these matrices was used to build a detection features in order to catch the artefacts which introduced by the whole imaging units. This method was able to get the detection accuracy with average 99.27% and it worked well on seven camera models. Summarised of the available techniques on image statistical are shown in Table 2.5.

Table 2.5: Techniques based on Image Statistical

Author	Contribution	Limitation
Kharrazi et al. (2004)	- Consists of 34-dimensional features including of color characteristics, IQMs and WDS	- Increased the computation time - Time consuming
B. Wang et al. (2009)	- Extracting the 216-dimensional higher-order wavelet features and 135-dimensional wavelet coefficient co-occurrence feature from tested images and applied the SFFS and correlation.	- Higher inter-model similarity in compare to another statistical feature
Wahab et al. (2012)	- Using CP from the selected block-wise DCT coefficients	- Not robustness of the features under different post processing operations camera setting and content

S. Gao, Hu, et al. (2012)	- Present multi –step transition probability to model a difference JPEG2-D arrays along seven directions	- Small intra-model similarity interferes with the inter-model similarity when the database includes several devices of the same model

2.4 Texture Feature Extraction Methods

Texture is a repeated pattern of information or arrangement of the structure with regular interval (Larroza et al., 2016; Materka et al., 1998) which has a certain scale, regularity and directionality. It is prominent in natural image and many important properties for image description and revealed through feature observation and extraction.

Feature extraction is a process of capturing a visual content of images for indexing and retrieving (Kumar et al., 2014). The idea is to obtain the most important information from original data and present the information in lower dimensionality space (Racoviteanu et al, 2012). It is a very critical process to measures the texture reasonably and effectively which can solve the problems of spectral heterogeneity and complex spatial distribution in the same category (Myint et al., 2004). Moreover, texture extraction has been employed in image segmentation, classification and pattern recognition.

There are four major methods used to extract texture features (R. Li, Li, et al., 2015) such as using structural based method, model-based method, transform based method and statistical based method as illustrated in Figure 2.7.

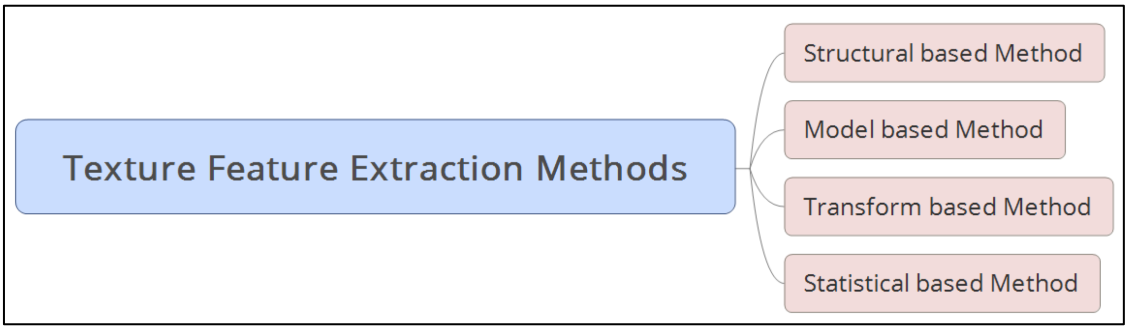


Figure 2.7: List of Texture Feature Extraction Method

As shown in Figure 2.7, there is four listed method in texture feature extraction. The following explanations are the details of the structural based, model-based and transform based method.

Structural based method (Levine, 1985; Weszka et al., 1976) describes texture as the composition of well-defined texture elements such as regularly spaced parallel lines. It is based on topological and geometric properties. The advantage of this is based on that it provides a good symbolic description of the images but it is more useful for image synthesis than image analysis.

Model based method (Cross et al., 1983; Pentland, 1984; Strzelecki et al., 1997) is the structure of an image that can be used for describing the texture and synthesizing it. The examples of this based method are the Fractal model and Markov model. This method describes the image as a probability model or as a linear combination of a set of basic functions.

The fractal model is useful for the modelling of certain natural texture which has a statistical quality of roughness at different scales and self-similarity. It is also useful for texture analysis and texture discrimination.

Different types of models based feature extraction technique are depending on the neighborhood system and noise sources. The different types consist of one-dimensional time series model, Auto Regressive (AR), Moving Average (MA) and Auto Regressive Moving Average (ARMA). A Markov model is a class of local random field models and it is widely used. This is because the conditional probability of the intensity of a given pixel is depends on the intensities of the neighborhood pixels.

Fourier, Gabor and Wavelet transform is a type of transform based method (Bovik et al., 1990; Daugman, 1985; Mallat, 1989; Rosenfeld et al., 1980). It represents an image in a space that coordinates the system which is closely related to the characteristics of a texture. The weakness of Fourier transform is that it could not perform well in spatial localization. Therefore, Gabor provides a better solution for spatial localization. These methods transform the original images by filters and calculate the energy of the transformed images.

2.4.1 Statistical based Method

In statistical based method (Julesz, 1975), it recognises the texture indirectly according to the non-deterministic properties. It manages the relationship between the gray-level on an image. This method is used to analyse the spatial distribution of gray value. It computes the local features at each of the points in the image and derives a set of statistics from the distribution of the local features. This method can be classified into first order statistic (one pixel), second order (pair of pixels) and higher order statistics (three or more pixels) as illustrated in Figure 2.8.

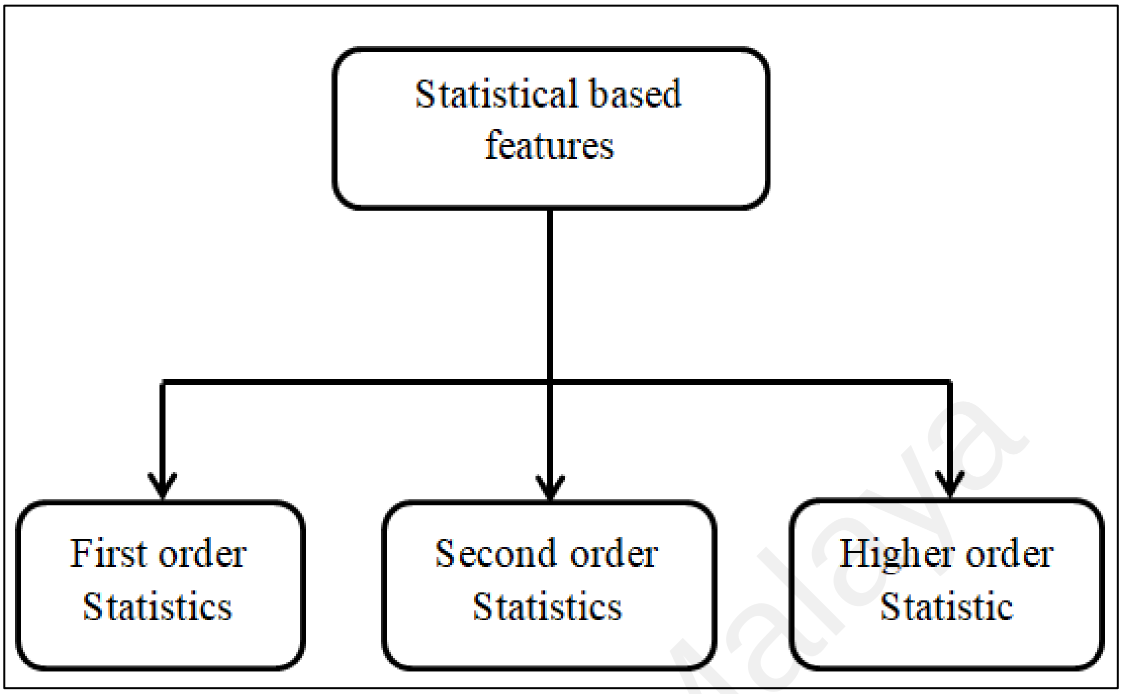


Figure 2.8: Illustration of Statistical based feature

First order statistics or first order histogram provides a different type of statistical properties such as four statistical moments of the intensity histogram of an image. These properties (mean, variance, skewness and kurtosis) depend on the individual pixel values and not on the interaction or co-occurrence of neighbouring pixel values.

The calculation of the first order histogram is described as a histogram h for a gray scale image I with the intensity in the range $I(x, y) \in [0, K-1]$ would contain exactly K entries, where for a typical 8-bit grayscale image, $K = 2^8 = 256$. Therefore, each individual histogram entry is defined as, $h(i)$ = the number of pixels in I with the intensity value I for all $0 \leq I < K$. The equation of first order histogram is shown in Equation (2.1),

$$h(i) = cardinality \{x, y | I(x, y) - i\} \tag{2.1}$$

where *cardinality* denotes the number of elements in a set. The standard deviation σ and intensity of histogram are shown in Equation (2.2) and (2.3).

$$\sigma = \sqrt{\frac{\sum(I(x,y) - m)^2}{N}} \tag{2.2}$$

$$skewness = \frac{\sum(I(x,y) - m)^3}{N\sigma^3} \tag{2.3}$$

In the second and higher order statistics, it estimates the properties with two or more-pixel values occurring at the specific locations that are relative to each other's. The most popular texture feature extraction is the second order statistical feature that is derived from the co-occurrence matrix. Figure 2.9 illustrates the types of second order statistics.

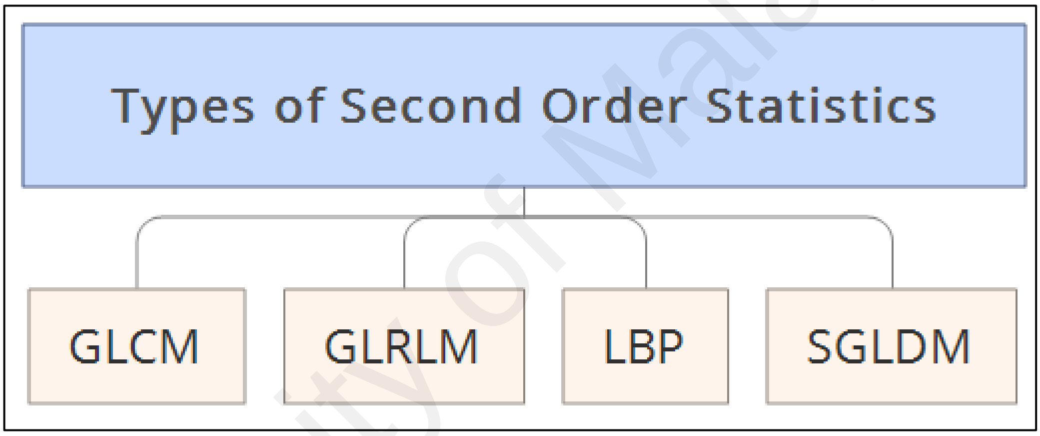


Figure 2.9 Types of Second Order Statistics

(a) ***Gray Level Co-Occurrence Matrix***

Gray Level Co-Occurrence Matrix (GLCM) is a method that extracts the second order statistical texture features and it was proposed by Haralick et al. (1973). It is a matrix where the number of rows and columns is equal to the number of gray level or pixel values in the surface of the images.

GLCM is calculated to give a measurement of variation in intensity at the pixel of interest. Two parameters from co-occurrence matrix are computed between the pixel pair d that measured in pixel number and their relative orientation θ. The orientation θ is quantized in four directions (0º, 45º, 90º and 135º) and it can be various combination of direction. GLCM (Haralick et al., 1973) has 14 features from the co-occurrence matrix, and between them there are more statistical features such as angular second moment, contrast, inverse diff. moment, contrast, and correlation. The complete lists of GLCM are present in Table 2.6.

Table 2.6 : List of GLCM features set

Moment	Equation
Energy	$f_{1}= \sum_{i}\sum_{j} p\ (i,j)^2$
Contrast	$f_{2}= \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j) \middle\| \|i-j\| = n \right\}$
Correlation	$f_{3}= \frac{\sum_i \sum_j (ij)\ p(i,j) -\ \mu_x\mu_y}{\sigma_x\sigma_y}$
Homogeneity	$f_{4}= \sum_{i}\sum_{j} \frac{1}{1+\ (i-j)^2} p(i,j)$
Entropy	$f_{5}= \sum_{i}\sum_{j} p(i,j) \log\big(p(i,j)\big)$
Autocorrelation	$f_{6}= \sum_{i}\sum_{j} (ij)\big(p(i,j)\big)$
Dissimilarity	$f_{7}= \sum_{i}\sum_{j} \|i-j\|, p(i,j)$
Cluster Shade	$f_{8}= \sum_{i}\sum_{j}\ \big(i, i+j-\mu_x-\mu_y\big)^3 p(i,j)$
Cluster Prominence	$f_{9}= \sum_{i}\sum_{j}\ \big(i, i+j-\mu_x-\mu_y\big)^4 p(i,j)$
Maximum Probability	$f_{10}= \underset{i,j}{MAX}\ p(i,j)$

(b) ***Gray Level Run Length Matrix Features***

Gray-Level Run-Length Matrix (GLRLM) that proposed by Galloway (1974) is the number of runs with pixels of gray level i and run length j for a given direction (R. Li, Kotropoulos, et al., 2015). The texture is recognized as a pattern of gray intensity pixel in a particular direction from the reference pixels. Run length is the number of adjacent pixels that have the same gray intensity in a particular direction and can be computed for any direction.

Mostly there are five basic features derived from GLRLM that are commonly used (Tang, 1998). These features are Short Run Emphasis (SRE), Long Run Emphasis (LRE), Gray Level Non-Uniformity (GLNU), Run Length Non-Uniformity (RLNU) and Run Percentage (RPEC). According to Chu et al. (1990), there are two more GLRLM features called Low Gray-Level Run Emphasis (LGLRE) and High Gray-Level Run Emphasis (HGLRE). Table 2.7 shows the list of GLRLM feature set that is commonly used.

Table 2.7: GLRLM features set

Moment	Equation
Short Run Emphasis (SRE)	$\frac{1}{n}\sum_{i.j}\frac{p(i.j)}{j^2}$
Long Run Emphasis (LRE)	$\sum_{i.j} j^2\, p(i,j)$
Gray Level Non-Uniformity (GLN)	$\frac{1}{n}\sum_{i}\left(\sum_{j} p\,(i,j)\right)^2$
Run Length Non-Uniformity (RLN)	$\frac{1}{n}\sum_{i}\left(\sum_{i} p\,(i,j)\right)^2$
Run Percentage (RP)	$\sum_{i,j}\frac{n}{p(i,j)j}$
Low Gray Level Run Emphasis (LGLE)	$\frac{1}{n}\sum_{i,j}\frac{p(i,j)}{i^2}$
High Gray Level Run Emphasis (HGRE)	$\frac{1}{n}\sum_{i,j} i^2\, p(i,j)$

(c) ***Local Binary Patterns***

Local binary pattern (LBP) is a complementary measure for local image contrast (Ojala et al., 2002). It can also be known as texture spectrum approach as described by D.-C. He et al. (1990) and L. Wang et al. (1990). LBP texture operators label the pixels of an image by thresholding the 3x3 neighbourhood of each pixel and consider the result as a binary number. The basic operators of LBP are illustrated in Figure 2.10.

178	25	123
60	86	102
189	68	243

3x3 neighbourhood

1	0	1
0		1
1	0	1

Thresholding

1	2	4
128		8
64	32	16

Weighting value

1	0	4
0		8
64	0	16

LBP value

Figure 2.10: Basic Operator of LBP

The process of LBP can be mathematically expressed as follow;

$$LBP_{P,R} = \sum_{i=0}^{P-1} s(g_i - g_c).2^i, \qquad s(x) = \begin{Bmatrix} 1, & x \geq 0 \\ 0 & x < 0 \end{Bmatrix} \tag{2.4}$$

Where g_c and g_i ($0 \leq i \leq P\text{-}1$) denote the gray value as a centre pixel and the gray value of neighbour pixel on a circle of radius R, respectively and P is the number of the neighbours.

LBP is one of the most frequently used in practical application since it has the advantages of simple implementation and fast performance. The related features of LBP are Scale-Invariant Feature Transform (SIFT) descriptor, Local Phase Quantization (LPQ) operator, Center-Symmetric LBP (CS-LBP) and Volume-LBP.

2.5 Chapter Summary

This chapter discusses the general information of OSNs including the issues of privacy and security concerns. It also discusses the emerging field of digital forensics which focuses on digital image forensics. There are two main problems in digital image forensics; image forgery detection and source camera identification. Each of the problems has its own specific issues but for this thesis purposes, the focus discussion is solely on source camera identification. The issue faced by the investigators is lack of available technique that can be applied to the OSN images. To address these issues, texture feature-based has been proposed by extracting the texture feature from OSN images in order to create a statistical value for classification purposes. However, texture feature especially using GLCM and GLRLM features has limited application applied to the source camera identification techniques. Furthermore, textural features give very distinctive information of a region or object. Therefore, GLCM and GLRLM texture feature has been considerably proposed to help investigators solve the issues in their entirety.

CHAPTER 3: RESEARCH METHODOLOGY

This chapter presents the methodology used to carry out research of source camera identification technique on OSN images. It covers the end to end process from the evaluation and selection of suitable features extraction technique, followed by the image filtering technique, and the classification stage.

3.1 Overview on Proposed Technique for Source Camera Identification

Source camera identification is a well research area for the state-of-art in image forensics. A few source camera identification techniques were selected and discussed in the previous Chapter 2. However, most of the techniques were focusing on source camera identification on original images and none of them managed to prove that their technique is also can be used for OSN images. Therefore this research is focusing on improving that limitation by proposing a solution based on the image statistical features techniques. The image statistical features technique has also been used by a few other researchers to identify the source camera from original images (Y.-C. Chen et al., 2011; S. Gao, Hu, et al., 2012; Gou et al., 2009; Hu et al., 2010; Kharrazi et al., 2004; Wahab et al., 2012; B. Wang et al., 2009; G. Xu et al., 2012).

In this research, evaluation for the suitable techniques begins with the first selection of six previous related researches by Kharrazi et al. (2004), Filipczuk et al. (2012), Nurtanio et al., (2013) Kulkarni et al. (2015), Singh (2016), and B. Xu et al. (2016) which were subsequently further shortlisted into four final list.

Two techniques by Kharrazi et al. (2004) and Kulkarni et al. (2015) were chosen based on their good result in fulfilling the suitable criteria of source camera identification. Multiple types of digital images from different models and brand were tested and provided good results.

Meanwhile, the other two techniques selected were from Filipczuk et al. (2012) and Singh (2016). Both are based on the performance of the texture feature GLCM and GLRLM feature used in their techniques which were proven capable to provide good detection accuracy for original images.

The above two factors of source camera identification and GLCM and GLRLM texture features are the main components selected for a further experiment in this research. More details of the selection process were explained in Section 3.2.1.

3.2 Research Methodology

This research was conducted in four main phases, starting from understanding the current issues, limitation and area of improvement, followed by setting up the research objective to improve the limitation identified earlier. The third phase covers the methodology used to come out with the proposed solutions and the last phase is the experiment and evaluation of the proposed techniques to prove that the research successfully meets its objective. All four phases involved were illustrated in the following Figure 3.1.

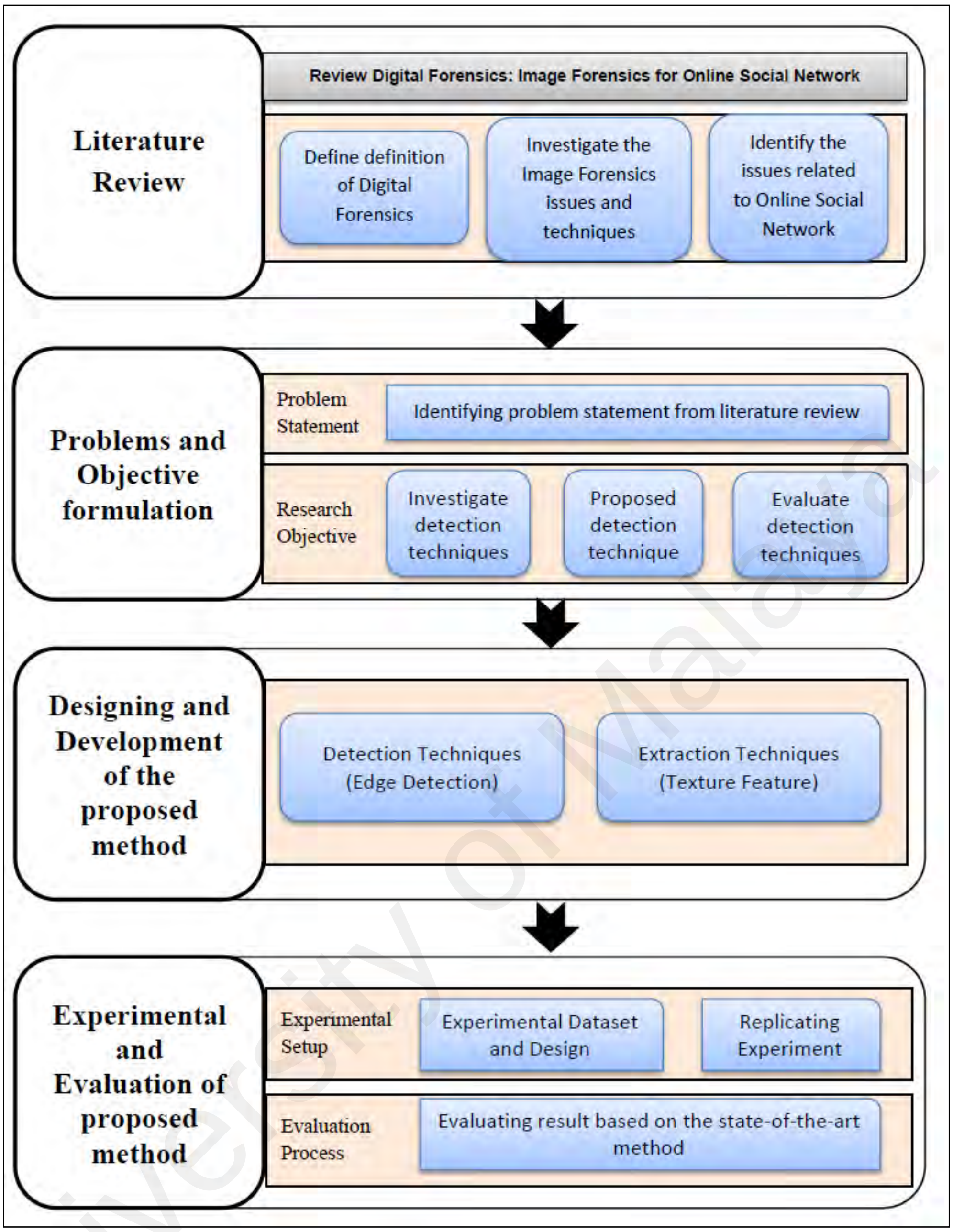


Figure 3.1: The end to end process flow of this research

The first phase is **"Literature Review"** which focuses on developing knowledge and idea in digital forensics, digital image forensics and online social network area. In this phase, a thorough study has been conducted to understand previous techniques and its limitation as well as deliberating its area of improvement. The review of existing techniques was not only limited to digital forensic but also covers other fields. For example, it was found that GLCM and GLRLM were also been used in Bio-Science field to classify the detection of cancers as per research conducted by Filipczuk et al.

(2012) and it was proved that both features are able to provide high detection accuracy. The issues on source camera identification for OSN images was also identified in this phase where it was found that none of the previous studies managed to prove any techniques suitable to detect source camera from OSN images.

The second phase is **“Problems and Objective Formulation”** which focused on identifying the problem statement and specifying the research objective to solve the identified issues. Problem statements were identified based on finding from Phase 1. In this case, the issues of source camera identification on OSN images were identified and three key objectives were set to solve the issues, which has been described in details under Chapter 1 of this thesis.

The third phase, **“Designing and development of the Proposed Method”** involved two main processes which are image filtering and texture feature extraction. Image filtering for OSN images was conducted using Laplacian of Gaussian (LoG) algorithm with default filter size, d=0.3. Value for the filter size was selected from 0.1 to 0.3. The result shows that larger filter size will provide smoother edges and lesser image noise. The second process, texture feature extraction was conducted using GLCM and GLRLM feature on the filtered OSN images. Further details on consideration made to choose GCLM and GLRLM texture feature extraction will be discussed in details under Section 3.2.1.

The fourth and final phase is **“Experimental and Evaluation of Proposed Method”** which involved two main processes. The first is experimental setup and followed by the evaluation process. The experimental setup included the use of FSCIT and Dresden dataset and covering experiment for both original and OSN images. The experiment results were then evaluated and compared further against the existing techniques by Kharrazi et al. (2004), Kulkarni et al. (2015), Filipczuk et al. (2012) and Singh (2016).

These four techniques were selected earlier as a benchmark to evaluate the proposed technique performance.

3.2.1 Selection Techniques used in the Proposed Solution

The selection process covers two main categories. The first selection is for source camera identification technique while the second selection is for the suitable feature extraction technique.

For the source camera identification, three techniques were shortlisted and further evaluated. The techniques are from Kulkarni et al. (2015), Kharrazi et al. (2004) and B. Xu et al. (2016). These three techniques were proven effective in giving good results for source camera identification on original images in their respective research. Among these three, a technique by Kulkarni et al. (2015) has given the highest detection accuracy with 97.59% followed by the technique by Kharrazi et al. (2004) in second place with 88.02%. Technique by B. Xu et al. (2016) however has given slightly lower detection accuracy from the earlier two techniques. In addition, there are challenges in getting the source code for techniques from B. Xu et al. (2016) where few attempts to contact and communicate with them did not receive any response. Hence, further evaluation of techniques by B. Xu et al. (2016) could not be conducted. Kharrazi et al. (2004) on the other hand has given a good cooperation in sharing their source code and provide comments on the area of improvement to be looked at based on their previous research limitation. The above are the reasons that both techniques from Kulkarni et al. (2015) and Kharrazi et al. (2004) were selected as a benchmark for the proposed technique in this research.

The second selection process is focusing on choosing the most suitable texture features for the proposed technique. Texture features were used in this research due to its high discrimination accuracy and require less computation time (Mohanaiah et al.,

2013) which will effectively expedite the experiment process. There were total of three existing techniques that used texture features in this research which are Filipczuk et al. (2012), Singh (2016) and Nurtanio et al. (2013). Research by Filipczuk et al. (2012) was used in Biomedical field which focusing on differentiating benign and malignant tumour in biopsy microscopic images for breast cancer diagnosis. In their research, the combination of GLCM and GLRLM texture feature under K-Nearest Neighbour (K-NN) classifier has given a significantly high detection accuracy of 90.00%. Another research in the Biomedical field by Nurtanio et al. (2013) was also using the combination of GLCM, GLRLM and First-Order statistic features. Under SVM classifier, the combined texture features provide slightly lower detection accuracy at 75.56%. Research by Singh (2016) on the other hand was used in the agricultural field which focusing on classifying the cereal grain for agriculture products. In their research, GLCM and GLRLM texture features were tested separately under Back Propagation Neural Network (BPNN) classifier with GLRLM provide 99.75% detection accuracy and GLCM provides slightly lower at 86.50% detection accuracy.

However, there was some limitation for the technique by Nurtanio et al. (2013) where it provides very low detection accuracy when the algorithm was re-executed for this research as part of the evaluation and selection process. In addition, the technique used in Nurtanio et al., (2013) used a small Region of Interest (ROI) for image features extraction which is not suitable to be adopted in source camera identification.

Hence based on the above evaluation, GLCM and GLRLM texture features used in technique by Filipczuk et al. (2012) and Singh (2016) were selected. Summary of the features for all four selected techniques are illustrated in the following Table 3.1.

Table 3.1: Description of features used in selected techniques

<table>
<tr><th>Author</th><th>Features</th><th>Feature Details</th></tr>
<tr><td>Kharrazi et al. (2004)</td><td>IQMs
Color Characteristics
Wavelet Domain Statistics</td><td>• Green color features have more contributions in achieving high accuracy because of Bayer-Pattern CFA with alternating row of R-G and G-B filters that are widely used in digital cameras.
• Computing the wavelet statistics for each color band has increased the accuracy of identification.</td></tr>
<tr><td>Kulkarni et al. (2015)</td><td>4 properties of GLCM feature
• Entropy
• Contrast
• Homogeneity
• Correlation</td><td rowspan="2">GLCM Feature
• GLCM is designed as two-dimensional histogram of gray levels for a pair of pixels. It is separated by a fixed spatial relationship.
• It able to reveal certain properties of spatial distribution of gray level in the texture image.
GLRLM Feature
• Provides information on the connected length of particular pixel in a definite direction and help in getting better classification accuracy.</td></tr>
<tr><td>Filipczuk et al. (2012)</td><td>4 properties of GLCM feature
• Contrast
• Correlation
• Homogeneity
• Energy
11 properties of GLRLM feature
• Short Run Emphasis (SRE)
• Long Run Emphasis (LRE)
• Gray-Level Non-Uniformity (GLNU)
• Run Length Non-Uniformity (RLNU)
• Run Percentage (RP)
• Low Gray- Level Run Emphasis (LGLRE)
• Short Run Low Gray Emphasis (SRLGE)
• Short-Run High Gray Emphasis (SRHGE)
• Long Run Low Gray Emphasis (LRLGE)
• Long Run High Gray Emphasis (LRHGE)</td></tr>
</table>

Singh (2016)	4 properties of GLCM feature • Contrast • Correlation • Homogeneity • Energy 4 properties of GLRLM feature • Short Run Low Gray Emphasis (SRLGE) • Short-Run High Gray Emphasis (SRHGE) • Long Run Low Gray Emphasis (LRLGE) • Long Run High Gray Emphasis (LRHGE)	

3.3 Proposed Solution

The proposed technique will be tested to identify the source camera from both original and OSN images. Despite the focus is to propose a technique for source camera identification on OSN images, experiment on original images is equally important to compare the performance of the new technique. There are three main stages involved in the proposed technique depicted in the following Figure 3.2.

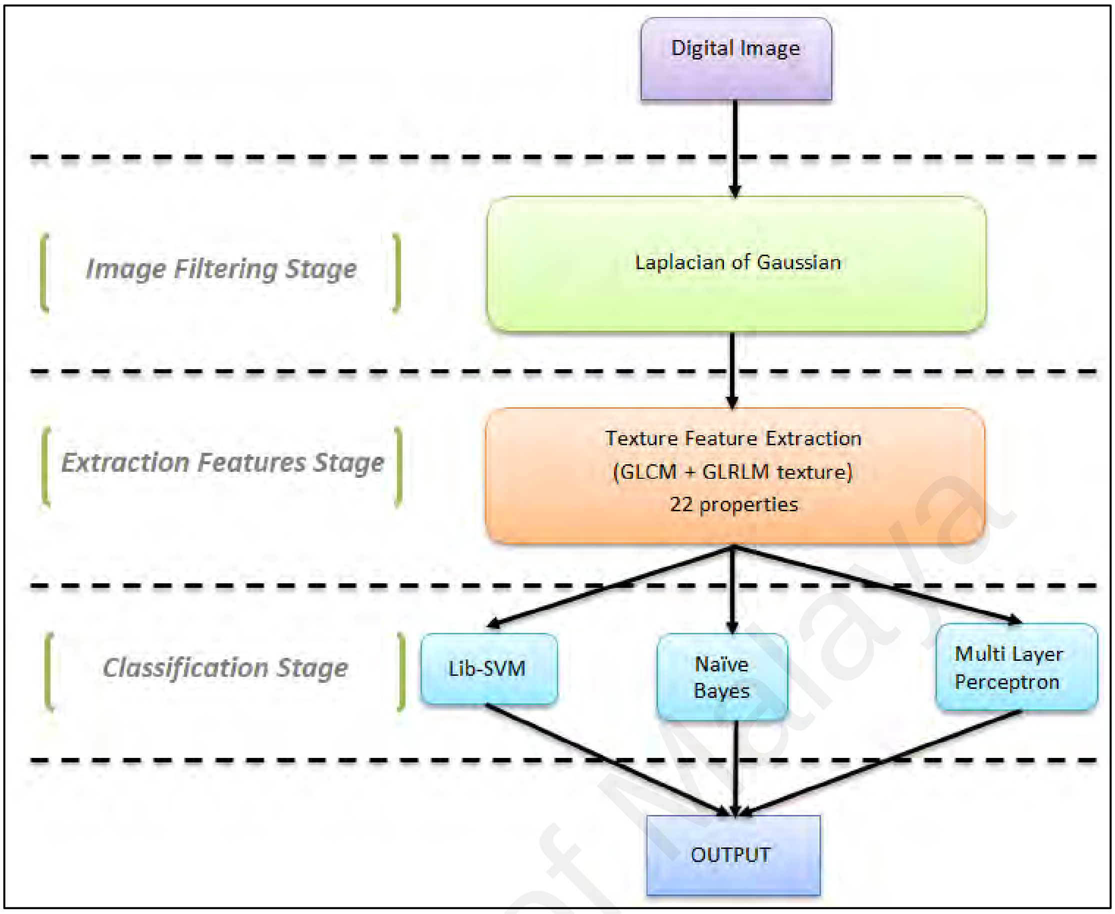


Figure 3.2: The structure of the proposed technique.

The first is image filtering stage that involved image filtration process via Laplacian of Gaussian algorithm. Both techniques were selected due to its good performance in the previous research for image filtering.

The second is extraction features stage which involved texture features extraction technique. Based on the existing research as described in Section 3.2.1 in this thesis, the combination of GLCM and GLRLM texture features were proposed. Both features have proven capable in providing good detection accuracy in previous research under the Biomedical field hence the combination of both would increase the detection accuracy for source camera identification, particularly on OSN images.

The third and final stage is the classification stage. Three classifiers were selected for this research which is LiB-SVM, Naïve Bayes and Multi-Layer Perceptron. These three classifiers is a universal classifier that commonly used in image processing and suitable to analyse a large dataset.

The above three stages also comprise various components and processes including image datasets, detection process, extraction process and classification process. These components will be described further in the next Section 3.4 in this thesis.

3.4 Research Requirement

There were six camera makers consisted of several camera models used in this experiment. Four camera models were collected from FSCIT community members and the others were collected from Dresden Image database. FSCIT dataset is a collection of images that available in the FSCIT community where this it can access or collected from the community members. All images were captured as natural images which were taken under the variety of natural indoor and outdoor scenery. The detailed information of these cameras as listed in Table 3.2 and 3.3.

Table 3.2: List of digital cameras collected from FCSIT community

ID	Model	Resolution	Format
I4	iPhone 4	1936x2592	JPEG
I4S	iPhone 4S	2448x3264	JPEG
I5	iPhone 5	1536x2048	JPEG
I5S	iPhone 5S	2448x3264	JPEG

Table 3.3: List of digital cameras collected from Dresden dataset

ID	Model	Resolution	Format
AGFA_504	Agfa DC504	4032x3024	JPEG
AGFA_733	Agfa DC733s	3072x2304	JPEG
AGFA_830I	Agfa DC830i	3264x2448	JPEG
AGFA_S505	Agfa Sensor505x	2592x1944	JPEG
CanonA640	Canon PowerShot A640	3648x2736	JPEG
NikonCoolPix	Nikon Cool PixS710	4352x3264	JPEG
Nikon D70	Nikon D70	3008x2000	JPEG
SonyW170	Sony DSC W170	3648x2736	JPEG

These images were filtered using LoG algorithm in order to reduce the high frequency of noise component and to increase the detection accuracy (Gonzalez et al., 2009). During the extraction process, the image features were extracted based on GLCM and GLRLM properties such as homogeneity, entropy, contrast and more. These properties were used during the classification processes.

3.4.1 Performance Evaluation

Performance evaluation in the thesis was to measure the proposed technique of source identification for OSN images. The F-Measure score was used to measure the performance of data classification. It measured how separable the dataset is when thresholding is applied to classify the data.

The F-Measure which also known as F_1 *score* is calculated as:

$$F_1 = \frac{2 * true\ positives}{2 * true\ positives + false\ positives + false\ negatives}$$

The classification performance was assessed using two constructed measurements. First, the number of constructed clusters was divided by the number of clusters that test the data. Thus, it gave a ratio of one if the output has too many clusters and below one if the tested group of image is sufficient. Each image in each cluster that was not from the same source as the majority in the respective cluster was counted. It gave an indication on how well the clustering algorithm was managed to avoid clustering images from different sources. This number was divided by the total number of images to come up with an error rate. The details and results of the evaluation technique are explained in Chapter 5.

3.4.2 Experimental Tool

This proposed technique is developed on Matlab programming software version R2014b. It performed using the laptop with Intel(R) Core (TM) i7-4700MQ CPU 8.00 GB RAM.

Weka application (Hall et al., 2009) is a data mining software used for classifying and identifying the camera models. There are three types of classifiers used in this experiment which consists of Lib-SVM Naive Bayes and Multi-Layer Perceptron. Table 3.4 is details of the tools used in the experiment.

Table 3.4: Lists of tools

Experimental Tool	Description
Personal Computer	Intel(R) Core (TM) i7-4700MQ CPU 8.00 GB RAM.
Matlab Software	Matlab version R2014b
Weka Machine Learning	Weka version 3.6.13

3.5 Chapter Summary

This chapter elaborated on the processes involved in this thesis. This experiment used two types of datasets that consists of FSCIT and Dresden dataset. This two dataset were then uploaded and downloaded from the OSN web for creating the OSN images dataset. Details of the process involve and results achieve are explained in Chapter 4 and 5.

CHAPTER 4: THE PROPOSED TECHNIQUE

This chapter described in details the proposed technique for source camera identification using texture features on original and OSN images. The structure of the technique will be explained further followed by the processes of collecting image dataset, image filtering process using LoG algorithm, the image extraction process of GLCM, GLRLM and combination of GLCM and GLRLM feature (Fusion feature) and image classification process.

4.1 Structure of Proposed Technique

The experiment is focusing on identifying source image based on texture features of GLCM and GLRLM. Both features are statistical feature that widely used in image processing and has a unique pattern in each digital image (Kharrazi et al, 2004). That was the main consideration to choose both GLCM and GLRLM features for the proposed technique. Detailed end to end process flow involved in the proposed technique experiment is illustrated in Figure 4.1.

The process begins with the collection of digital images from FSCIT community members and "Dresden Image Database" (Gloe & Böhme, 2010) from *Dresden Image website. Both image datasets comprise of natural images taken under various indoor and outdoor scenery. The experiment process has been executed with 400 images from FSCIT dataset and 600 images from Dresden dataset. Those images were captured using different types of camera makers and camera models. All images were subsequently filtered by LoG algorithm to reduce the high-frequency noise component and increase the accuracy of identification.

* http://forensics.inf.tu-dresden.de/ddimgdb/

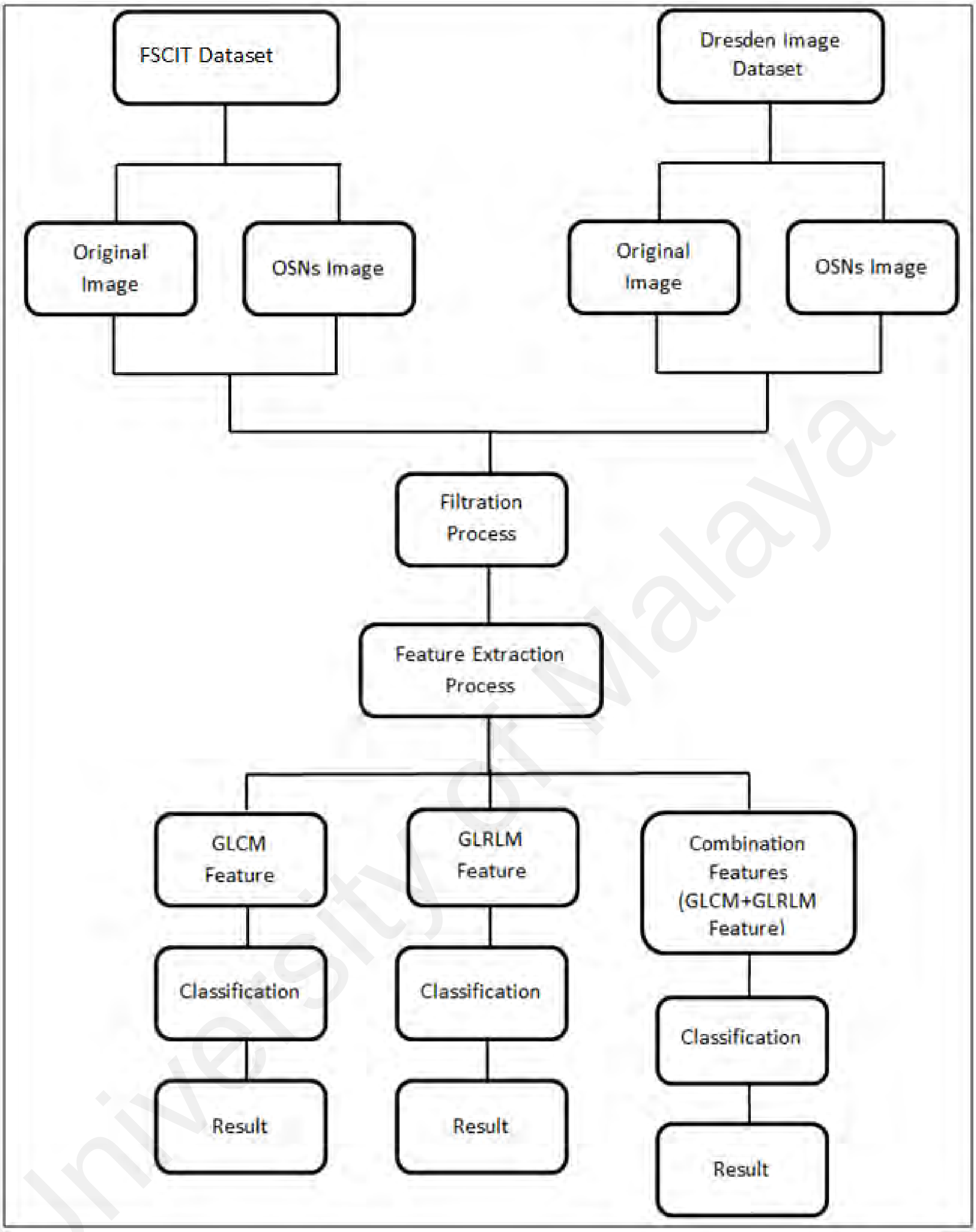


Figure 4.1: The experimental flow process

4.2 Experimental Settings

This section describes the structure of image dataset used in this experiment. Two types of dataset consist of original and OSN images were tested. Original images defined as a digital image that directly derived from digital camera devices without any image processing. The processes of creating original images were discussed in Chapter 2.

OSN images are defined as a digital image that has been uploaded and downloaded from the OSN web and has gone through a pre-processing stage that reduces the image size and improves the image appearance. Figure 4.2 illustrates the process of creating original and OSN images.

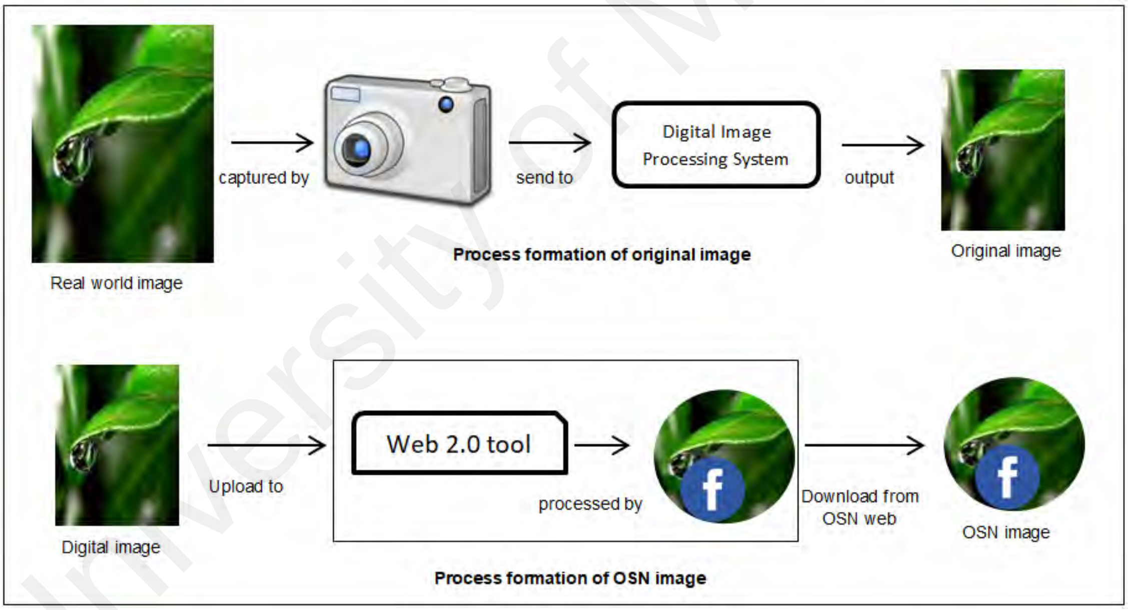
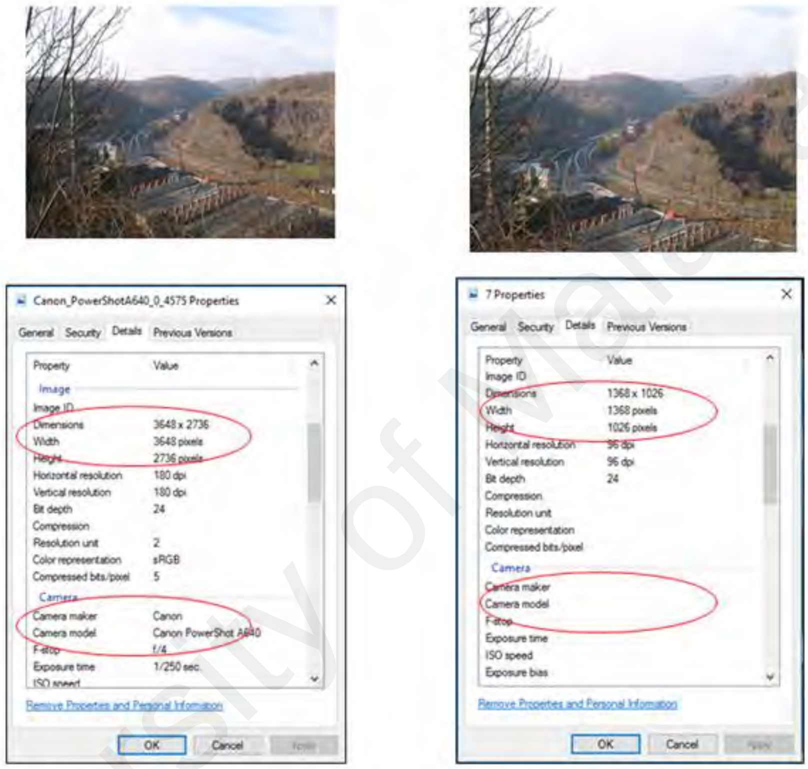


Figure 4.2: Process Formation of Original and OSN image

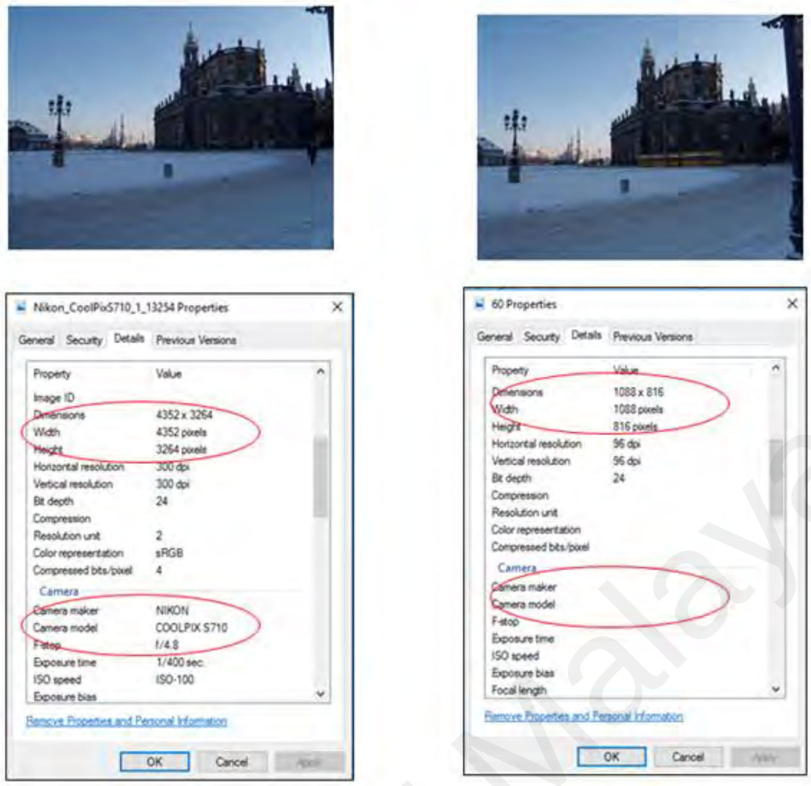
Original image derived by capturing the real image and processed through the image processing system in the digital camera. This original image subsequently uploaded into OSN website in order to create OSN image.

Each digital image contains metadata, detailed information of the image including time taken, date taken, camera maker and camera model. Extracting image metadata is the basic technique to identify the image source. However, this technique is limited for original images only. The same information will not be able to be retrieved from images that have gone through pre-processing including cropping, scaling and resizing. Images that have gone through modification on Web 2.0 tool system will also have the same limitation. The comparison of information available between original images and OSN images are illustrated in Figure 4.3 and Figure 4.4.



i. Information from Original Image ii. Information from OSN Image

Figure 4.3: Comparison of image metadata in an image taken by camera Canon Power Shot A640



i. Information from Original Image   ii. Information from OSN Image

Figure 4.4: Comparison of image metadata in an image taken by camera Nikon Coolpix S710

Based on the example above, the original image metadata contains more detailed information including camera maker and the camera model used to take the images. Meanwhile, in OSN image, the metadata does not include camera maker and model details.

In addition, it was also shown that the image dimension in OSN image significantly deteriorated compared to original images. This was due to the pre-processing image by the web server where it will reduce the image size and improve image appearance before uploading to OSN web.

To show in further details the impact of web server pre-processing into an original images metadata, each original image from each dataset was uploaded and subsequently downloaded from the OSN web. The detailed list of image dataset and the modification made to its resolution by the web service are recorded in Table 4.1.

Table 4.1: Image modification performed by Facebook web application

Image Dataset	Model ID	Image Resolution		Image Size	
		Original	OSN	Original (mb)	OSN (kb)
FSCIT Dataset	I4	1936x2592	717x960	2.47	207
	I4S	2448x3264	720x960	2.03	84.0
	I5	1536x2048	720x960	1.22	64.4
	I5S	2448x3264	720x960	3.39	53.6
Dresden Dataset	AGFA_504	4032x3024	1008x756	1.05	95.5
	AGFA_733	3072x2304	1152x864	2.72	139
	AGFA_830I	3264x2448	1224x918	1.94	66.7
	AGFA_S505	2592x1944	972x729	1.21	106
	CanonA640	3648x2736	1368x1026	4.89	215
	NikonCoolPix	4352x3264	1088x816	5.24	210
	Nikon D70	3008x2000	1128x750	0.71	116
	SonyW170	3648x2736	1368x1026	3.37	233

Based on the information above, it was clear that images being published into OSN platform have undergone a significant change in its original characteristic including image resolution and image size.

4.2.1 OSN Dataset

A Facebook web account was created to facilitate the preparation of OSN dataset. All function in the Facebook page was set to default settings. Facebook was selected as the OSN platform to be used on in this research due to their big number of the user (Malleson et al., 2015). All images collected from both FSCIT and Dresden dataset been uploaded into the Facebook page as shown in Figure 4.5 to 4.8.

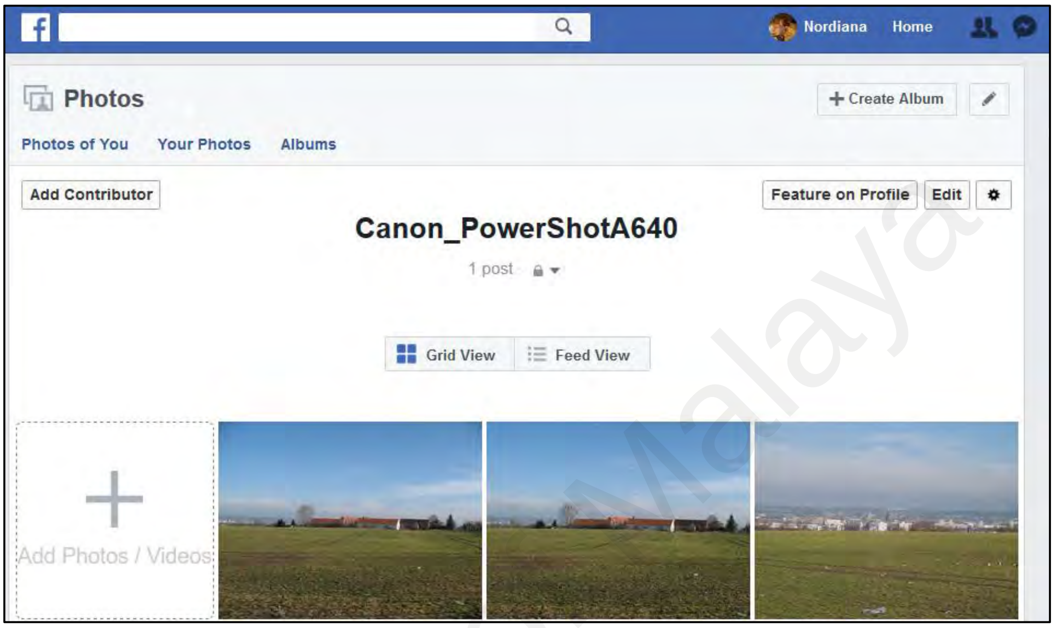


Figure 4.5: Digital Images from Canon Power Shot A640 in Facebook album

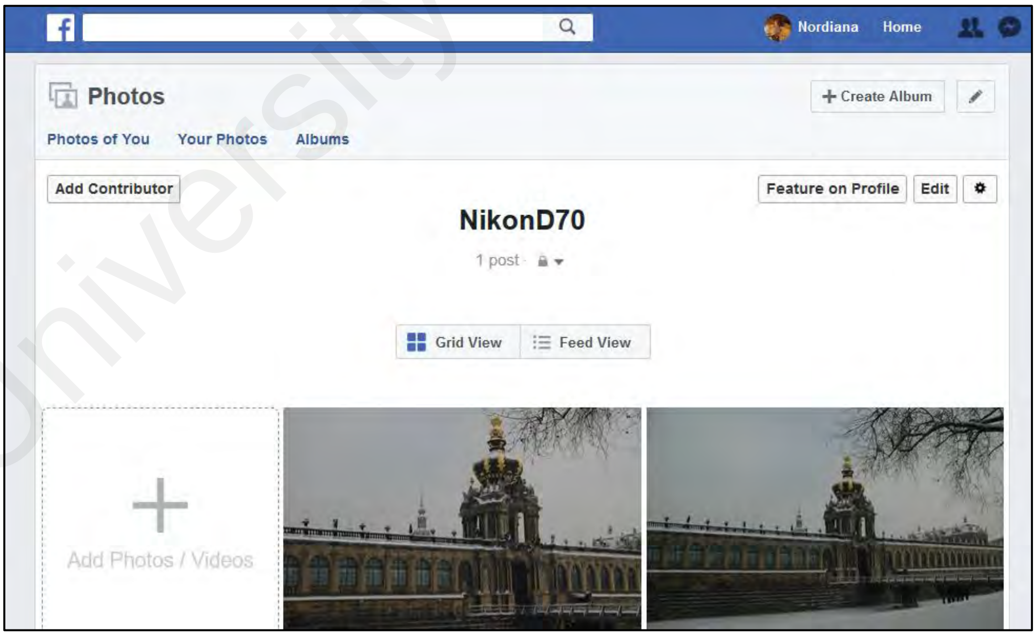


Figure 4.6: Digital Image from Nikon D70 in Facebook album

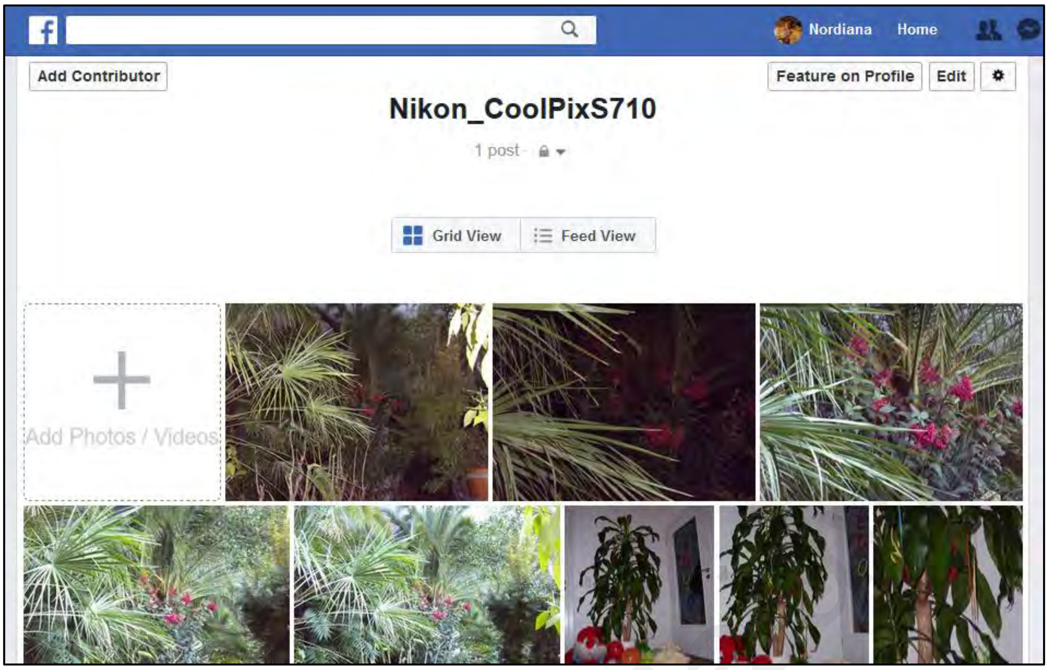


Figure 4.7: Digital Images from Nikon Cool Pix S700 in Facebook album

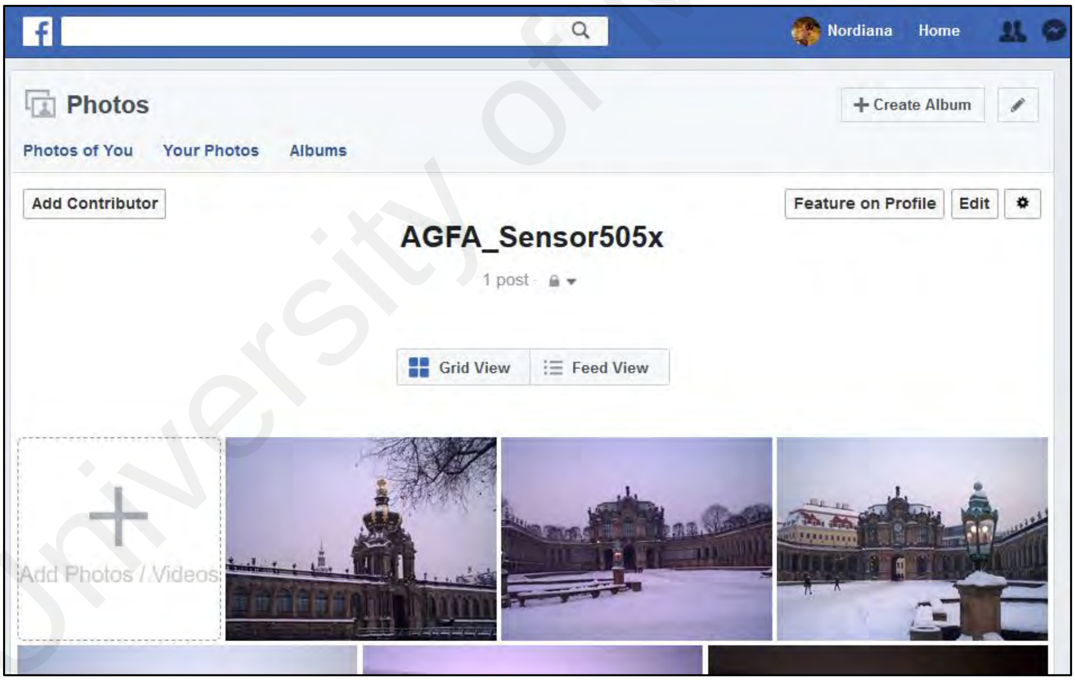


Figure 4.8: Digital Images from Agfa Sensor 505x in Facebook album

4.3 Image Filtering Process

All digital images including original and OSN images are affected by noise. LoG filter was applied in this thesis is to filter and detect the edges in order to reduce the noise level in each image used in this experiment. The process of filtration using LoG filter is shown in Figure 4.9.

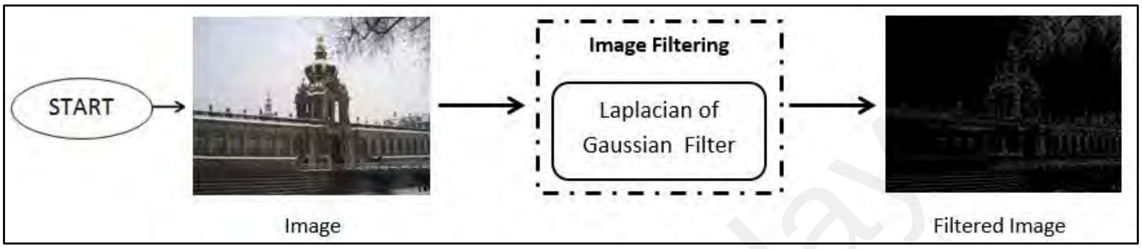


Figure 4.9: Image filtering using Laplacian of Gaussian filter.

This filtration process is using filter size, *d*=0.3 as explained in Section 3.2, Para 3. The illustration of the filtering process in each dataset is shown in Figure 4.10.

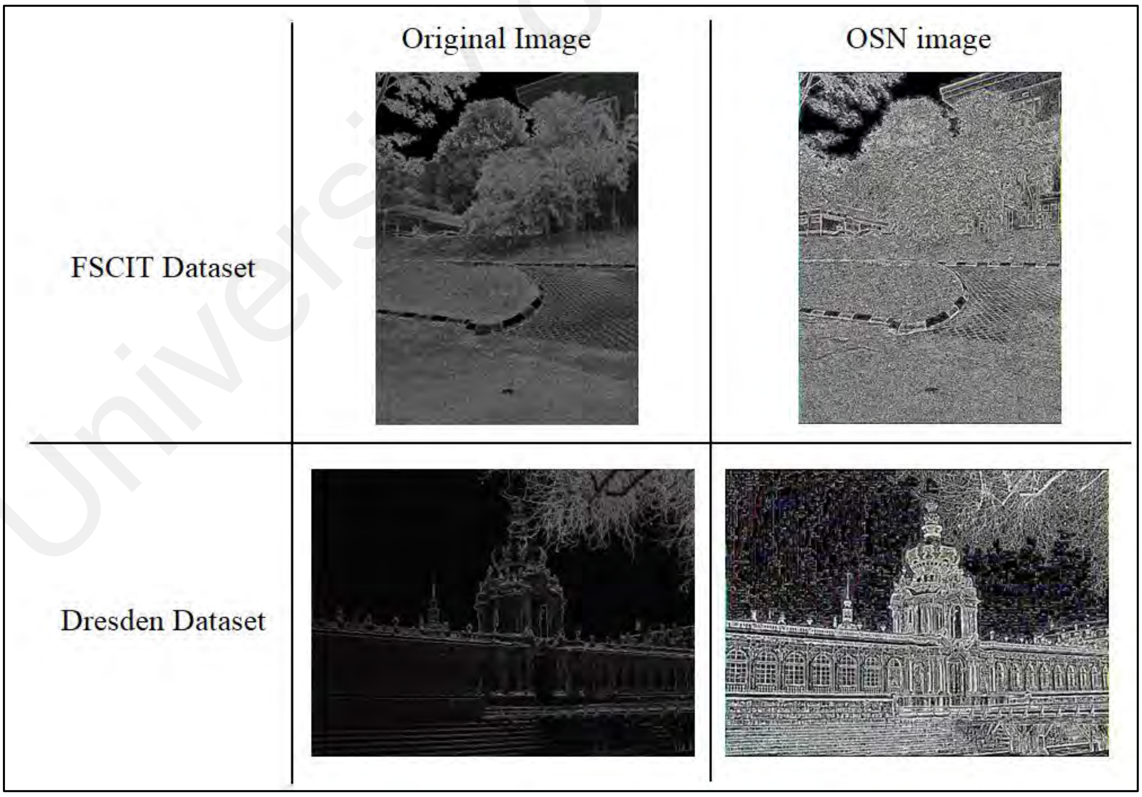


Figure 4.10: Image filtration using filter size, d=0.3 for both Original and OSN images

The figure above shows the result of filtration in both original and OSN images using LoG filter. The image features subsequently were extracted in the extraction process.

4.4 Extraction Process

In the extraction process, GLCM and GLRLM texture feature were extracted from the filtered images. Both GLCM and GLRLM feature contains of 22 and seven set feature properties respectively, including image contrast, correlation, homogeneity, entropy, short run emphasis and long run emphasis as illustrated in Figure 4.10.

GLCM Properties

Autocorrelation
Contrast
Correlation
Correlation
Cluster Prominence
Cluster Shade
Dissimilarity
Energy
Entropy
Homogeneity
Homogeneity
Maximum probability
Sum of squares: Variance
Sum average
Sum variance
Sum entropy
Difference variance
Difference entropy
Information measure of correlation1
Information measure of correlation2
Inverse difference
Inverse difference normalized
Inverse difference moment normalized

GLRLM Properties

Short Run Emphasis
Long Run Emphasis
Gray Level Non-Uniformity
Run Percentage
Run Length Non-Uniformity
Low Gray Level Run Emphasis
High Gray Level Run Emphasis

Figure 4.11: Properties of GLCM and GLRLM features

The above extraction process was applied to all images from both FSCIT and Dresden dataset. These properties will be used for performance evaluation which will be explained further in Chapter 5 of this thesis. Once the features extraction process completed, the next process is to classify the images via selected classifiers.

4.5 Classification Process

There are three classifiers selected for this research comprises of Lib-SVM, Naïve Bayes and Multi-Layer Perceptron. All these three classifiers are universal classifiers and commonly used in image processing application (X. He et al., 2010). The classification process started with image properties extracted from GLCM feature and followed by image properties extracted from GLRLM features. The training and testing for all three classifiers were selected using a 10-fold cross-validation method. Both training and testing datasets for each classifier comprises of 400 and 600 images from FSCIT and Dresden dataset respectively.

4.6 Chapter Summary

This chapter discusses the end to end experiment process of the proposed technique for source camera identification using image filtering algorithm (Laplacian of Gaussian) and texture feature GLCM and GLRLM. The dataset used in these experiments are collected from FSCIT research community and Dresden Image database.

CHAPTER 5: EVALUATION EXPERIMENTS AND DISCUSSIONS

This chapter explained the details of the experimental process and result based on the proposed technique. The result of original and OSN images was obtained from FSCIT and Dresden Dataset (Gloe et. al, 2010). All the steps and processes involved in the experiment were detailed in Chapter 4. In general, the experiment flows for both datasets depicted in Figure 5.1 below.

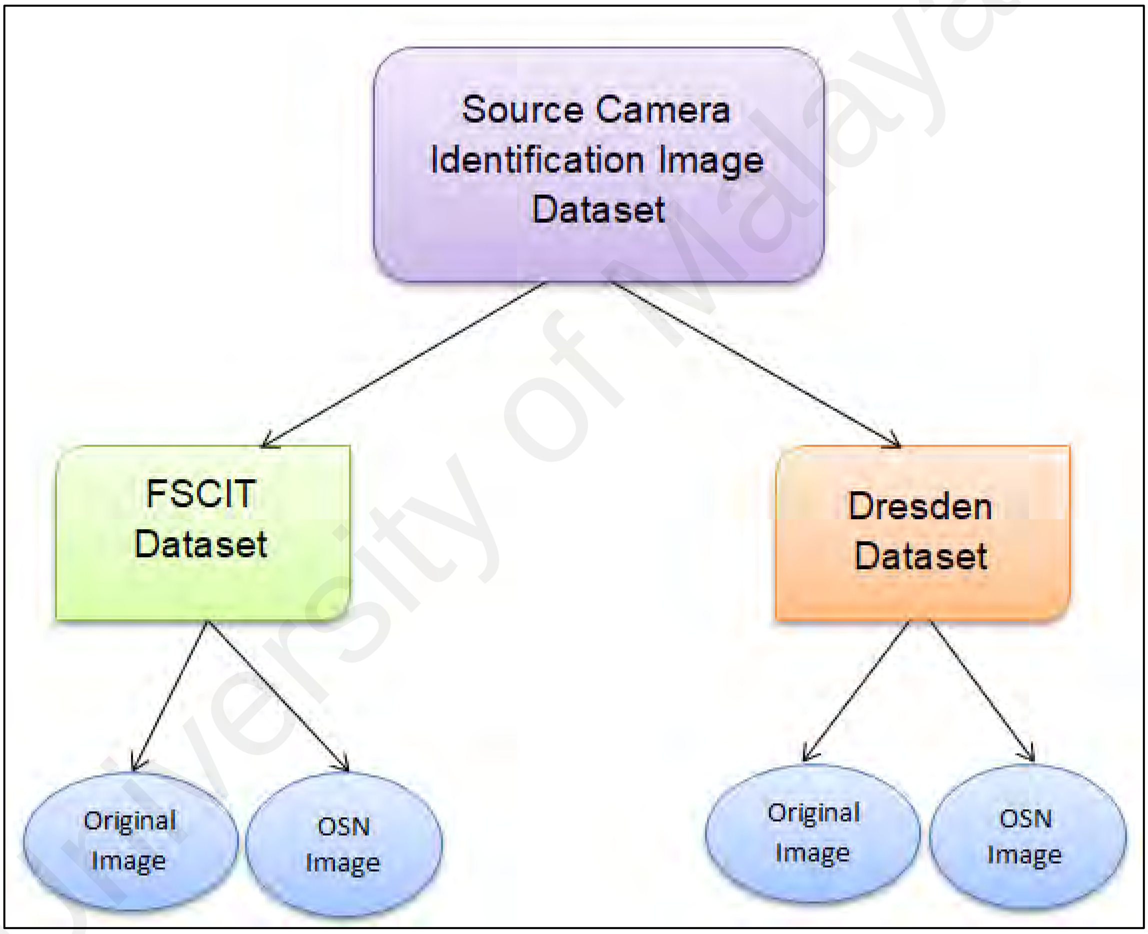


Figure 5.1: Experiment flow for both dataset

Each dataset was further breakdown into two types; the first is original images dataset and second is OSN images dataset. For OSN images, detailed process to create the dataset was described in Chapter 4.

5.1 Experimenting the Original Image

This experiment is to test the detection accuracy of the proposed technique to detect and identify the source of images in its original features without any modification, from different camera maker and model. Original images dataset are divided into three categories for this experiment. The first is from FSCIT dataset which includes four different models of iPhone. The second and third category is from Dresden dataset which was divided into two categories consist of a different model with the same maker denote as *DI-I* dataset and different model with different maker denote as *DI-II* dataset. Figure 5.2 illustrated the structure of the datasets used in this experiment.

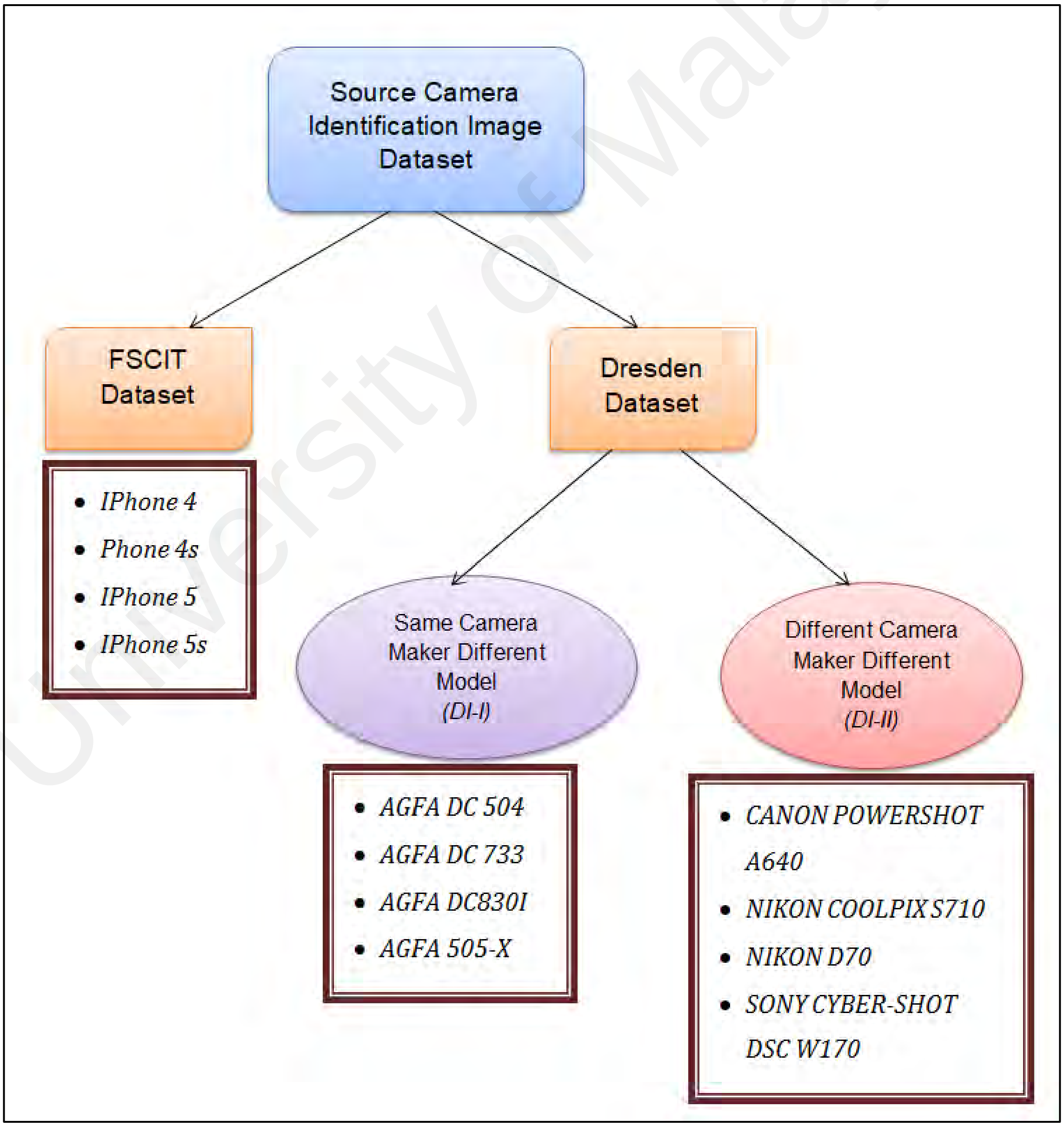


Figure 5.2 : Structure of dataset used in this experiment

Each experiment used 400 to 600 images in which each model represented by 100 to 150 images.

5.1.1 Result Analysis on Source Camera Identification Technique

In this section, the experiment results on original images will be elaborated further. Results are divided into three subsection which the first subsections are the experimental result from FSCIT dataset, second is the experimental result from Dresden dataset *DI-I* (same maker different model) and third is Dresden dataset *DI-II* (different maker different model).

5.1.1.1 Experimental Result from FSCIT dataset

There are 400 original images used for this experiment. Relatively it is not a big dataset but still acceptable for a significant result. It is due to the limitation of getting original images from each iPhone camera model. Details of the iPhone camera model and the original image's resolution and size are listed in Table 5.1.

Table 5.1: List of camera make and model from FSCIT dataset

Faculty Dataset on Original Images			
ID	**Model**	**Resolution**	**Size (mb)**
I4	iPhone4	1936x2592	2.5
I4S	iPhone4S	2448x3264	2.2
I5	iPhone5	1536x2048	2.0
I5S	iPhone5S	2448x3264	3.5

Images are filtered using LoG algorithm in order to reduce the noise level. The process continues by extracting image texture feature based on GLCM and GLRLM. These two features are combined (namely as Fusion feature) in order to improve the detection accuracy derived from the single feature. The extraction results are then fed into three types of classifiers which are Naïve Bayes, Lib-SVM and Multi-Layer Perceptron (MLP) for the final result of the camera source identification. Details of the detection accuracy are illustrated in Figure 5.3.

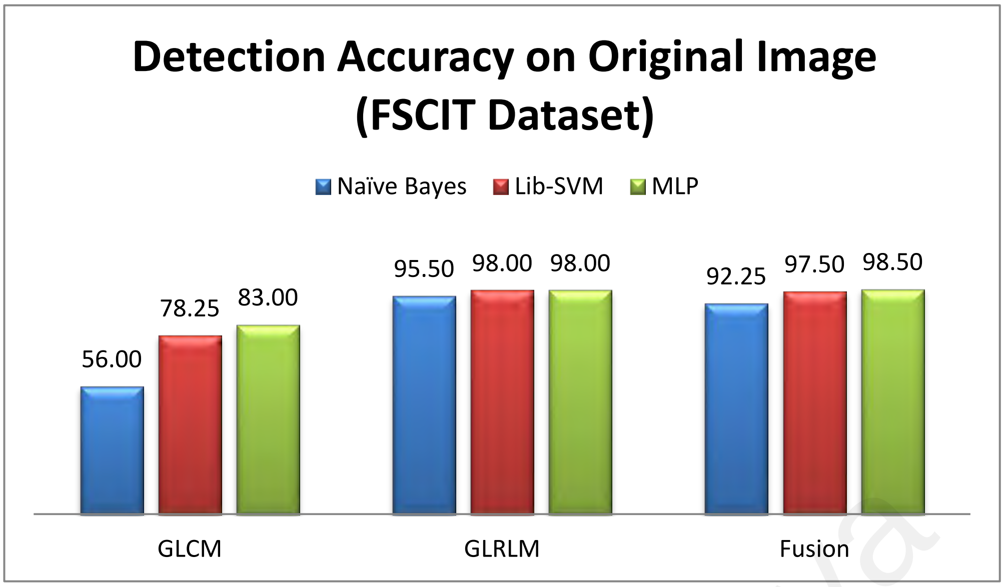


Figure 5.3: Detection Accuracy on Original Images for FSCIT dataset

Based on the results obtained, individual experiment for each GLCM and GLRLM features provided reasonably high detection accuracy in all classifier with above 80.00% accuracy except GLCM under Lib-SVM and Naïve Bayes having accuracy below 80.00%. The highest detection accuracy was derived from GLRLM feature under MLP classifier with 98.00%. However, the combination of GLCM and GLRLM features (Fusion features) provided even better accuracy with 98.50% under MLP classifier. The following Table 5.2 to Table 5.4 are the results of a confusion matrix for Fusion under three classifiers. Each row represents the detection accuracy for each iPhone camera models.

Table 5.2: Confusion Matrix for Fusion feature classified using Naïve Bayes

	Identified			
	I4	**I4S**	**I5**	**I5S**
I4	**97%**	2%	0%	2%
I4S	0%	**91%**	0%	6%
I5	0%	0%	**95%**	3%
I5S	5%	9%	0%	**86%**
Average	**92.25%**			

Table 5.3: Confusion Matrix for Fusion feature classified using Lib-SVM

	Identified			
	I4	**I4S**	**I5**	**I5S**
I4	**100%**	0%	0%	%
I4S	0%	**99%**	0%	1%
I5	0%	0%	**100%**	0%
I5S	0%	7%	0%	**93%**
Average	**97.50%**			

Table 5.4: Confusion Matrix for Fusion feature classified using MLP

	Identified			
	I4	**I4S**	**I5**	**I5S**
I4	**100%**	0%	0%	0%
I4S	0%	**96%**	0%	4%
I5	0%	0%	**100%**	%
I5S	0%	4%	0%	**96%**
Average	**98.50%**			

Based on the results depicted earlier, it was proven that the proposed technique of Fusion feature is capable in identifying the camera source of original digital images with a high average detection accuracy of 98.50% for iPhone mobile camera using MLP classifier.

5.1.1.2 Experimental Result from Dresden dataset from same camera Maker and different Model (*DI-I*)

This section describes the result of the second category of original images which is Dresden dataset (*DI-I*). The process flow for *DI-I* dataset remains the same as per FSCIT dataset. Details of the camera model and the original image's resolution and size used are listed in Table 5.5.

Table 5.5 : List of the same camera maker and different model from Dresden dataset (*DI-I*)

Dresden Dataset *(DI-I)* on Original Images			
ID	**Model**	**Resolution**	**Size (mb)**
AGFA_504	Agfa DC504	4032x3024	1.50
AGFA_733	Agfa DC733s	3072x2304	2.72
AGFA_830I	Agfa DC830i	3264x2448	1.94
AGFA_S505	Agfa 505-x	2592x1944	1.21

There are 600 original images used for this experiment and the results are illustrated in Figure 5.4

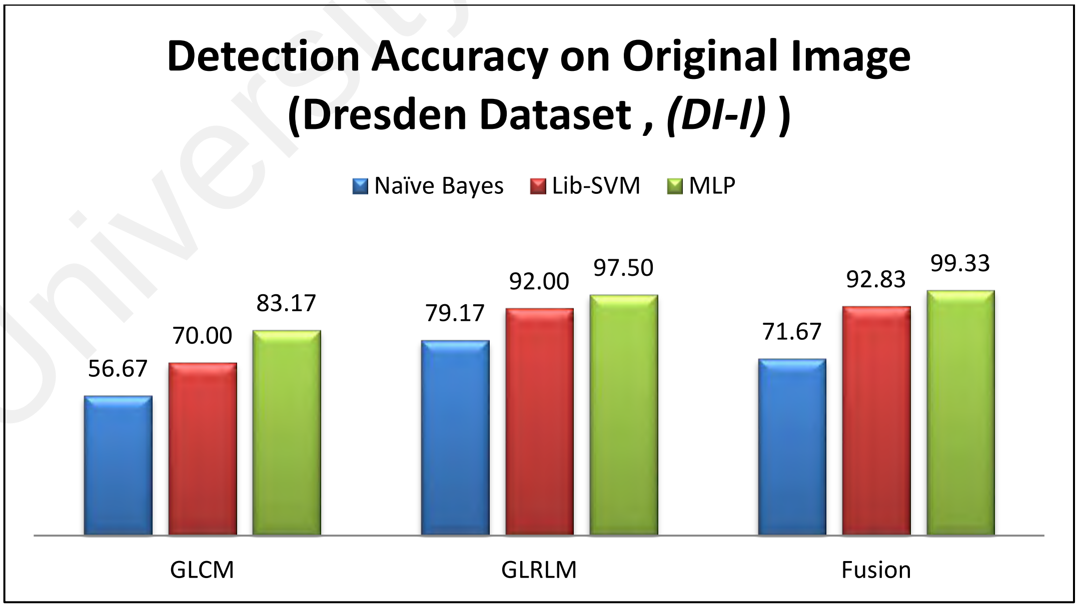


Figure 5.4: Detection Accuracy on Original Image for Dresden dataset (*DI-I*)

Individual experiment results on *DI-I* dataset for both GLCM and GLRLM features showing slightly lower detection accuracy compared to GLCM and GLRLM results for

FSCIT dataset. GLRLM feature is still giving higher detection accuracy in all classifier compared to GLCM feature, which is also the same pattern as per FSCIT dataset experiment.

However, experiment on *DI-I* dataset via Fusion feature has given the best result with 99.33% detection accuracy under MLP classifier. The following Table 5.6 to Table 5.8 are confusion matrix for Fusion feature under three classifiers.

Table 5.6: Confusion Matrix of *DI-I* dataset for Fusion feature classified using Naïve Bayes

	Identified			
	AGDC504	**AGDC733**	**AGDC830**	**AGS505**
AGDC504	**77.30%**	8.0%	4.67%	10.0%
AGDC733	6.0%	**61.30%**	30.67%	2.0%
AGDC830	3.30%	13.33%	**83.30%**	0%
AGS505	20.67%	12.67%	3.0%	**64.70%**
Average	**71.67%**			

Table 5.7: Confusion Matrix of *DI-I* dataset for Fusion feature classified using Lib-SVM

	Identified			
	AGDC504	**AGDC733**	**AGDC830**	**AGS505**
AGDC504	**86.70%**	0%	2%	11.30%
AGDC733	1.30%	**97.30%**	0%	1.40%
AGDC830	2.0%	10.70%	**87.30%**	0%
AGS505	0%	0%	0%	**100%**
Average	**92.83%**			

Table 5.8: Confusion Matrix of *DI-I* dataset for Fusion feature classified using MLP

	Identified			
	AGDC504	**AGDC733**	**AGDC830**	**AGS505**
AGDC504	**98.00%**	0%	1.00%	1.00%
AGDC733	0%	**99.30%**	0%	0.70%
AGDC830	0%	0%	**100%**	0%
AGS505	0%	0%	0%	**100%**
Average	**99.33%**			

Based on the above results, it was proven that the proposed technique of Fusion feature is also capable in identifying the camera source from original images derived from *DI-I* dataset with the best detection accuracy of 99.33% under MLP classifier.

5.1.1.3 Experimental result from Dresden dataset with different camera maker and different model *(DI-II)*

This section describes the result for the third category of original images, Dresden dataset *(DI-II)*. The reason for this experiment is to see whether the proposed technique is also providing good detection accuracy for *DI-II* dataset, consistent with FSCIT and *DI-I* dataset tested earlier. The process flow for *DI-II* dataset remains the same as per FSCIT and *DI-I* dataset.

Images used in this experiment were taken by three entry-level digital still cameras (DSC) and one semi-pro digital single lens reflex (DSLR) camera. Details of the camera model and the original image's resolution and size used are listed in Table 5.9.

Table 5.9: List of the camera maker and model from Dresden dataset (*DI-II*)

Dresden Database *(DI-II)* on Original Images			
ID	**Model**	**Resolution**	**Size (mb)**
CanonA640	Canon PowerShot A640	3648x2736	4.89
Nikon CoolPix	Nikon Cool PixS710	4352x3264	5.24
Nikon D70	Nikon D70	3008x2000	1.20
SonyW170	Sony DSC W170	3648x2736	3.37

In order to maintain the same comparison level, a total of 600 original images used for this experiment which is also the same number of images used in *DI-I* dataset experiment. The results are presented in Figure 5.5.

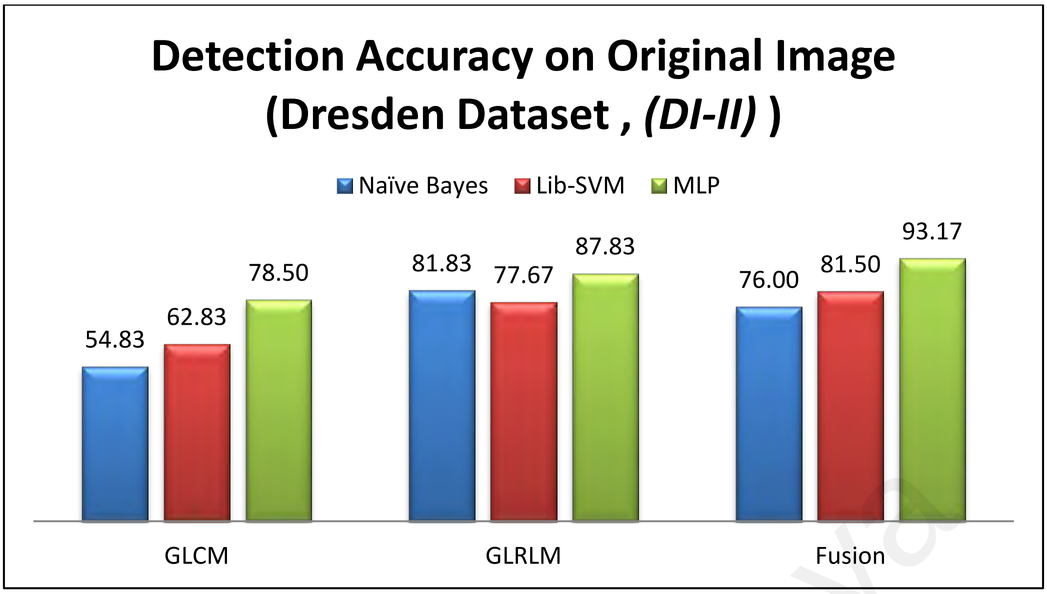


Figure 5.5: Detection Accuracy on Original Image for Dresden dataset *(DI-II)*

Individual experiment results on *DI-II* dataset for both GLCM and GLRLM features showing slightly lower detection accuracy compared to GLCM and GLRLM results from FSCIT and *DI-I* dataset. But GLRLM feature is still providing higher detection accuracy in all classifier compared to GLCM feature, same pattern as FSCIT and *DI-I* dataset.

However, experiment on *DI-II* dataset via Fusion feature showing slightly lower detection accuracy compared to FSCIT and *DI-I* dataset, with the best is 93.17% under MLP classifier. The following Table 5.10 to Table 5.12 are confusion matrix for Fusion feature under three classifiers.

Table 5.10: Confusion Matrix of *DI-II* dataset for Fusion feature classified using Naïve Bayes

	Identified			
	CanonA640	**NikonCoolPix**	**NikonD70**	**SonyW170**
CanonA640	**72.70%**	6.67%	1.33%	19.30%
NikonCoolPix	0.70%	**99.30%**	0%	0%
NikonD70	0%	0.70%	**93.30%**	6%
SonyW170	51.30%	10%	0%	**38.70%**
Average	**76.00%**			

Table 5.11: Confusion Matrix of *DI-II* dataset for Fusion feature classified using Lib-SVM

	Identified			
	CanonA640	**NikonCoolPix**	**NikonD70**	**SonyW170**
CanonA640	**66.00%**	0%	1.30%	32.70%
NikonCoolPix	0%	**100%**	0%	0%
NikonD70	0%	0%	**100%**	0%
SonyW170	40.00%	0%	0%	**60.00%**
Average	**81.50%**			

Table 5.12: Confusion Matrix of *DI-II* dataset for Fusion feature classified using MLP

	Identified			
	CanonA640	**NikonCoolPix**	**NikonD70**	**SonyW170**
CanonA640	**86.70%**	0%	1.30%	12.00%
NikonCoolPix	0%	**100%**	0%	0%
NikonD70	0%	0%	**100%**	0%
SonyW170	14.00%	0%	0%	**86.00%**
Average	**93.17%**			

Based on all three experiments covering FSCIT, *DI-I* and *DI-II* dataset, it was confirmed that the proposed technique of Fusion feature giving higher detection accuracy compared to the existing GLCM and GLRLM features for original images. The highest detection accuracy was derived from *DI-I* dataset under MLP classifier with 99.33%, followed by FSCIT dataset under MLP classifier with 98.50% and lastly *DI-II* dataset under MLP classifier with 93.17%. It was also proven that MLP classifier is best suited the proposed Fusion feature for higher detection accuracy compared to Naïve Bayes and Lib-SVM.

5.1.2 Performance Evaluation Metrics

This section will deliberate further the performance metrics used in original images experiments described in the earlier section of this chapter. Performance metrics is a common method to validate experimental results which in this case, to validate the proposed technique performance for original images. The performance metrics include detection precision rate and false positive rate.

Precision is one of the basic performance measures that used in the evaluation of the identification problems. In this experiment, precision is used as a ratio of true identification of image source with respect to the total number of images. Percentage (%) sign is used as a quantity precision measurement. The precision measurement is defined as the following equation.

$$Precision = \frac{N_D}{N_D + N_{ID}} \tag{5.1}$$

N_D denotes as the number of true detections of the image source and N_{ID} is the numbers of false detection of the image source.

Meanwhile, false positive rate also helps in determining the performance level of the proposed Fusion feature technique. It is referring to the error obtained during evaluation of the scenario in a certain condition that is observed as positive or close to true condition instead of a fully false condition. In this experiment, false positive is represented in a form of the percentage of incorrect detection ratio against the total number of tested digital images. The equation of the false positive rate is defined as follows,

$$false\ positive\ rate = \frac{N_D}{N_D + N_M} \tag{5.2}$$

Where N_D denote as the numbers of the true source image detection and N_M denote as the number of miss detection source images. The result for both precision rate and false positive rate for all techniques used in these original images experiment are listed in Table 5.13.

Table 5.13: Performance Evaluation Metrics for source camera Identification techniques on Original images

FSCIT Dataset						
Texture Feature	**Precision Rate (%)**			**False Positive Rate (%)**		
	Naïve Bayes	Lib-SVM	MLP	Naïve Bayes	Lib-SVM	MLP
GLCM	53.90%	78.40%	83.10%	14.70%	7.30%	5.70%
GLRLM	95.60%	98.00%	98.00%	1.5%	0.7%	0.7%
Fusion	92.30%	97.50%	98.50%	2.60%	0.8%	0.5%
Dresden Dataset (*DI-I*)						
Texture Feature	**Precision Rate (%)**			**False Positive Rate (%)**		
	Naïve Bayes	Lib-SVM	MLP	Naïve Bayes	Lib-SVM	MLP
GLCM	55.70%	70.00%	83.40%	14.40%	10.00%	5.60%
GLRLM	80.40%	92.60%	97.70%	6.90%	2.70%	0.80%
Fusion	72.40%	93.20%	99.33%	9.40%	2.40%	0.20%
Dresden Dataset (*DI-II*)						
Texture Feature	**Precision Rate (%)**			**False Positive Rate (%)**		
	Naïve Bayes	Lib-SVM	MLP	Naïve Bayes	Lib-SVM	MLP
GLCM	54.80%	62.30%	78.40%	15.10%	12.40 %	7.20%
GLRLM	82.10%	77.70%	87.40%	6.10%	7.40%	4.20%
Fusion	75.60%	81.40%	93.10%	8.00%	6.20%	2.30%

Based on the above results, the proposed technique provided the highest precision rate and lowest false positive rate in all dataset via MLP classifier compared to GLCM and GLRLM feature. *DI-I* dataset giving the best result for Fusion feature with 99.33% precision and 0.20% false positive rate, followed by FSCIT dataset with 98.50% precision and 0.5% false positive rate and lastly *DI-II* dataset with 93.10% precision and 2.30% false positive rate.

5.1.3 Texture Feature Comparison against Proposed Technique

This section compares the performance of the proposed techniques against two other techniques proposed earlier by Filipczuk et al. (2012) and Singh (2016). Both techniques were selected due to its good performance on classification process via GLCM and GLRLM feature. Details of their techniques were explained in Chapter 3 Section 3.2.1. Both techniques were also used the same measurements metric of precision rate and false positive rate.

The comparison is fundamental in determining the performance level of the proposed technique against the existing techniques. All techniques were evaluated using original images from *DI-I* dataset under MLP classifier since it provided the highest detection accuracy than the other two datasets, FSCIT and *DI-II* dataset in the previous experiments. Detailed comparisons were tabulated in Table 5.14 below.

Table 5.14: Comparison of detection accuracy based on Texture Feature

Techniques	**GLCM**		**GLRLM**		**Fusion**	
	Precision Rate	**False Positive Rate**	**Precision Rate**	**False Positive Rate**	**Precision Rate**	**False Positive Rate**
Proposed Technique	83.40%	5.60%	97.70%	0.80%	99.33%	0.20%
(Filipczuk et al., 2012)	67.70%	10.70%	81.60%	6.20%	82.30%	5.90%
Singh (2016)	68.80%	10.30%	66.20%	11.10%	75.90%	8.10%

The above result shows the proposed technique provided the highest precision rate and lowest false positive rate in all three texture features, GLCM, GLRLM and Fusion feature.

The precision rate of the proposed technique using Fusion feature is the highest with 99.33%, an increase by 20.69% from the second highest provided by Filipczuk et al. (2012) technique with 82.30%. Meanwhile, false positive rate of 0.20% provided by the proposed technique under Fusion texture is the lowest among all other techniques. It was a significant reduction by 96.61% from the second highest false positive rate provided by Filipczuk et al. (2012) technique at 5.90%.

The above comparison clearly shows that the proposed technique is having a better performance than the existing technique by Filipczuk et al. (2012) and Singh (2016).

5.1.4 Comparison with Other Source Camera Identification Techniques

This section further compares the performance of proposed technique against Kharrazi et al. (2004) and N. Kulkarni et al. (2015) techniques on source camera identification. The comparison was based on two measurements metric comprises of the average precision rate and the average false positive rate. Detailed comparison recorded in Table 5.15 below.

Table 5.15 : Comparison with other source camera identification techniques

Technique	Average Precision Rate (PR) (%)	Average False Positive Rate (FPR) (%)
Kharrazi et al. (2004)	93.00	2.4
Kulkarni et al. (2015)	70.60	10.4
Proposed Technique	99.33	0.2

All techniques were evaluated using a total of 600 original images derived from *DI-I* dataset. The above results has proven that proposed technique of Fusion feature is performing better than the existing technique in identifying the camera source of original images with the highest average precision rate of 99.33% and lowest average false positive rate of 0.20% using MLP classifier.

5.2 Discussion Result on Original Image

In all experiments conducted to assess the performance of proposed technique, it was confirmed and proven that the proposed technique is capable in giving better detection accuracy for source camera identification from original digital images compared to the existing techniques.

The above confirmation was derived by evaluating and comparing the proposed technique against existing source camera identification techniques from Kharrazi et al.

(2004) and Kulkarni et al. (2015). All technique were tested using the same *DI-I* dataset under MLP classifier. The dataset are selected based on the recommendation by Gloe and Böhme (2010) as a benchmark data for source camera identification.

The performance of proposed technique was evaluated based on two fundamental metrics comprises of precision rate and false positive rate. The precision rate is defined as the percentage of true source camera identification from digital images. A higher percentage represents a better precision rate in determining the camera source. False positive rate on the other hand is defined as the percentage of false or almost true source camera identification from digital images. A lower percentage represents a better performance by the technique used in determining the digital image's camera source.

Based on the result explained in the earlier section of this chapter, it shows that the proposed technique is capable in providing better detection accuracy in identifying camera source from original digital images compared to the existing techniques.

The proposed techniques provided highest precision rate and lowest false positive rate in all dataset used in the experiment for original images covering FSCIT, *DI-I* and *DI-II* under MLP classifier. In addition, comparison against the existing techniques also showed an improvement of precision rate by 20% and significant reduction of false positive rate by 96.61% from the closest existing technique performance by Filipczuk et al. (2012).

5.3 Experimental Results on OSN Images

This experiment is focusing on the performance level of the proposed technique to identify camera source of OSN images. OSN images are selected in this experiment due to the fact that the original information of the images has been removed and altered resulted in difficulties in identifying the original source of the image. The Dresden dataset *DI-I* and *DI-II* were selected to be used in this experiment due to its good result and detection accuracy in the previous experiment for original images. The process flow for these experiments also remains the same as per the previous experiment.

5.3.1 Result analysis on Source Camera Identification Technique

This section will deliberate in details on each experiment results for OSN images from each dataset and classifier. The results will be segregated in two subsections comprises the results for *DI-I* datasets and *DI-II* dataset.

5.3.1.1 Experimental Result on Social Network images for Dresden Dataset *(DI-I)*

This section describes the result of *DI-I* dataset on OSN images. Details of the camera model and OSN image's resolution and size used are listed in Table 5.16.

Table 5.16: List of the camera used for this experiment (*DI-I*)

Dresden Database (*DI-I*) on OSN Images			
ID	**Model**	**Resolution**	**Size (kb)**
AGFA_504	Agfa DC504	1008x756	95.5
AGFA_733	Agfa DC733s	1152x864	139
AGFA_830I	Agfa DC830i	1224x918	66.7
AGFA_S505	Agfa 505-x	972x729	106

There are 600 OSN images used from this experiment and the results were illustrated in Figure 5.6.

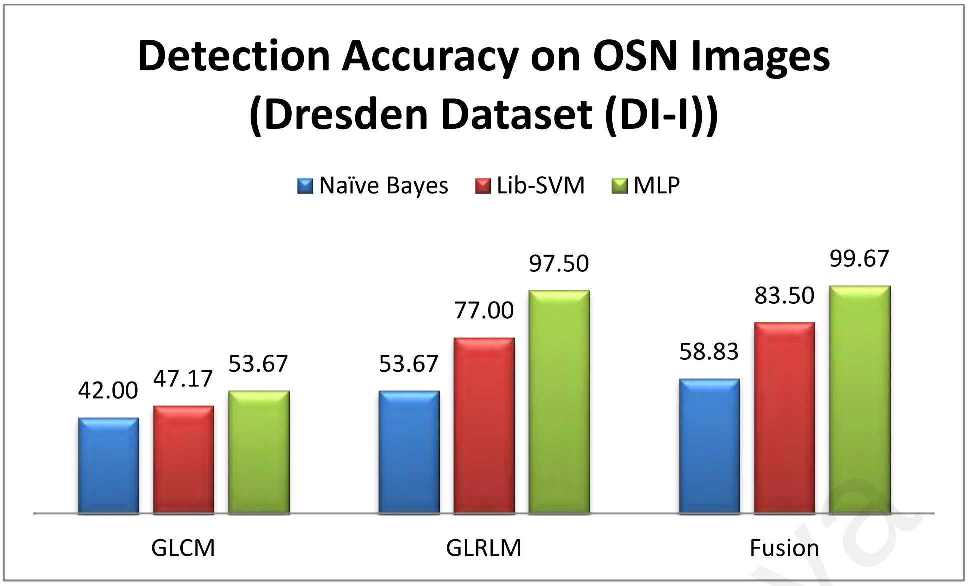


Figure 5.6: Detection Accuracy on OSN Images for Dresden dataset *(DI-I)*

Individual experiment results for each GLCM and GLRLM on OSN images using *DI-I* datasets show lower detection accuracy than the GLCM and GLRLM results on original images except for GLRLM under MLP classifier. GLRLM under MLP classifier provided the same detection accuracy between original and OSN images with 97.50%.

However, Fusion feature for OSN images under MLP classifier shows the highest detection accuracy with 99.67%, which is also slightly better detection accuracy than the same technique on original images with 99.33%. The following Table 5.17 to Table 5.19 are confusion matrix for Fusion feature under three classifiers.

Table 5.17: Confusion Matrix of *DI-I* dataset on OSN images for Fusion feature classified using Naïve Bayes

	Identified			
	AGDC504	**AGDC733**	**AGDC830**	**AGS505**
AGDC504	**44.70%**	16.67	6.63%	32%
AGDC733	8%	**58.70%**	30%	3.30%
AGDC830	2%	16.70%	**81.30%**	0%
AGS505	20%	17.30%	12%	**50.70%**
Average	**58.83%**			

Table 5.18: Confusion Matrix of *DI-I* dataset on OSN images for Fusion feature classified using Lib-SVM

	Identified			
	AGDC504	**AGDC733**	**AGDC830**	**AGS505**
AGDC504	**78.00%**	0%	0%	22.00%
AGDC733	0.70%	**91.30%**	8.30%	0%
AGDC830	0%	8.00%	**92.00%**	0%
AGS505	27.30%	0%	0%	**72.70%**
Average	**83.50%**			

Table 5.19: Confusion Matrix of *DI-I* dataset on OSN images for Fusion feature classified using MLP

	Identified			
	AGDC504	**AGDC733**	**AGDC830**	**AGS505**
AGDC504	**99.30%**	0%	0%	0.70%
AGDC733	0.70%	**99.30%**	0%	0%
AGDC830	0%	0%	**100%**	0%
AGS505	0%	0%	0%	**100%**
Average	**99.67%**			

The above result has proved that the proposed technique is capable to provide high detection accuracy in identifying camera source from OSN images derived from *DI-I* dataset. The highest detection accuracy derived from Fusion feature under MLP classifier with 99.67%.

5.3.1.2 Experimental Result on Social Network Images for Dresden Dataset (*DI-II*)

This section describes the result of OSN images from *DI-II* dataset. The process flow and camera model in this experiment remain the same as per *DI-I* dataset. Details of the camera model and the OSN image's resolution and size used are listed in Table 5.20.

Table 5.20: List of the camera used for this experiment (*DI-II*)

Dresden Database (*DI-II*) on OSN Images			
ID	**Model**	**Resolution**	**Size (kb)**
CanonA640	Canon PowerShot A640	1368x1026	215
Nikon CoolPix	Nikon Cool PixS710	1088x816	210
Nikon D70	Nikon D70	1128x750	116
SonyW170	Sony DSC W170	1369x1026	233

There are 600 OSN images used for this experiment and the detailed results depicted in Figure 5.20.

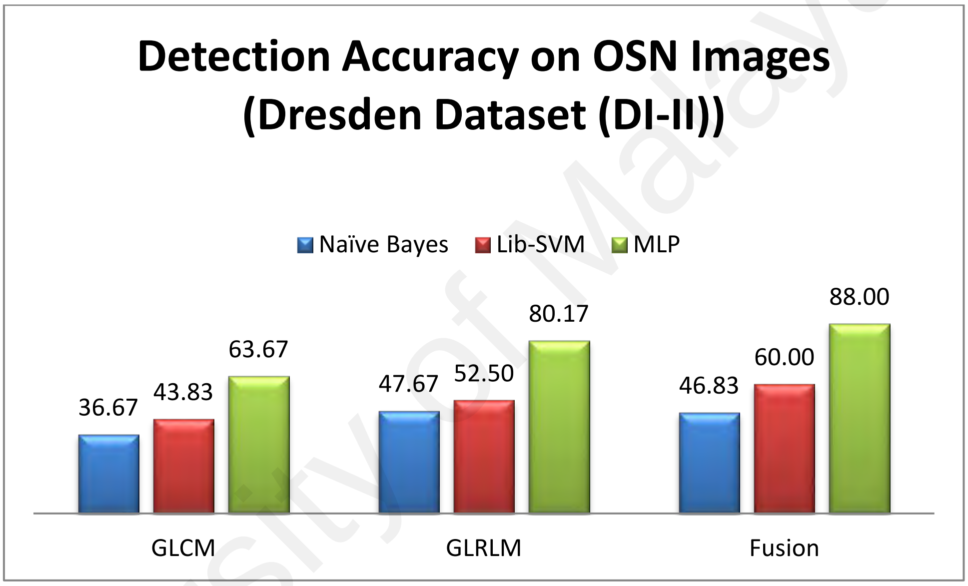


Figure 5.7: Detection Accuracy on OSN Images for Dresden dataset (*DI-II*)

The above result shows that detection accuracy is slightly lower in all features under all classifiers compared to *DI-I* dataset results on OSN images. This is due to the *DI-II* dataset contains OSN images from different camera models compared to *DI-I* dataset. Nevertheless, this result for the proposed technique is considered successful since the false positive rate is lower and the precision rate is higher than the existing features of GLCM & GLRLM. Detailed results illustrated in Table 5.21 below.

Table 5.21: Performance Evaluation Metrics for source camera Identification techniques on OSN images

Dresden Dataset (*DI-I*)						
Texture Feature	**Precision Rate**			**False Positive Rate**		
	Naïve Bayes	Lib-SVM	MLP	Naïve Bayes	Lib-SVM	MLP
GLCM	37.60%	42.10%	49.50%	20.40%	19.50%	17.00 %
GLRLM	55.70%	67.60%	97.50%	15.40%	10.80%	0.80%
Fusion	58.70%	83.60%	99.70%	13.70%	5.50%	0.10%
Dresden Dataset (*DI-II*)						
Texture Feature	**Precision Rate**			**False Positive Rate**		
	Naïve Bayes	Lib-SVM	MLP	Naïve Bayes	Lib-SVM	MLP
GLCM	37.70%	43.10%	63.60%	21.10%	18.70 %	12.10%
GLRLM	47.30%	52.70%	81.20%	17.40%	15.80 %	6.60%
Fusion	47.10%	60.00%	88.00%	18.30%	13.30 %	4.00%

The above results have proven that proposed technique of Fusion feature is capable to provide high detection accuracy on OSN images derived from both *DI-I* and *DI-II* dataset with the highest detection accuracy of 99.70% and 88.00% respectively under MLP classifier.

5.3.2 Texture Feature Comparison against Proposed Technique on the OSN images

Comparison of the proposed technique performance against two other existing techniques proposed by Filipczuk et al. (2012) and Singh (2016) using OSN images will be explained further in this section. These two techniques were selected as a benchmark to the proposed techniques since both have applied texture feature extraction specifically on GLCM and GLRLM feature for their research purposes. The same performance metric used in original images were evaluated in this experiment which comprises of precision rate and false positive rate. The comparison results are summarized in Table 5.22.

Table 5.22: Comparison on Social Media Images with others techniques

Techniques	GLCM		GLRLM		Fusion	
	Precision Rate	False Positive Rate	Precision Rate	False Positive Rate	Precision Rate	False Positive Rate
Proposed Technique	49.40	17.00%	97.50%	0.80%	99.70%	0.10%
Filipczuk et al. (2012)	60.40%	13.10%	73.30%	9.00%	75.20%	8.40%
Singh (2016)	60.10%	13.20%	62.30%	12.30%	71.20%	9.60%

Based on the above results, it shows that the proposed technique of Fusion feature provided the highest precision rate at 99.70%, an increase of 32.58% from the closest existing techniques by Filipczuk et al (2012) at 75.20%. The false positive rate, on the other hand, shows a more significant reduction by 98.81% from 8.40% by Filipczuk et al (2012) to 0.10% by the proposed technique.

5.3.3 Comparison with other Source Camera Identification Techniques on OSN images

This section described a further comparison between the performance of the proposed technique against Kharrazi et al. (2004) and Kulkarni et al. (2015) using OSN images *DI-I* dataset under MLP classifier. The performance is measured via two measurements metric comprises of the average precision rate and the average false positive rate. Detailed results are recorded in Table 5.23.

Table 5.23: Comparison with other source camera identification techniques on OSN images

Technique	Average Precision Rate (PR) (%)	Average False Positive Rate (FPR) (%)
Kharrazi et al. (2004)	88.00	4.00
Kulkarni et al. (2015)	57.10	14.20
Proposed Technique	99.70	0.10

A total of 600 OSN images were used in all techniques above. It shows that the proposed technique of Fusion feature is significantly having a better precision rate with an average of 99.70% and better false positive rate with average 0.1% in identifying the camera source of OSN images.

5.4 Discussion Result on Original images and OSN images

The overall result of proposed technique comparisons and evaluation against existing techniques by Kharrazi et al. (2004), Filipczuk et al. (2012), Kulkarni et al. (2015), and Singh (2016), using 600 original and OSN images from *DI-I* Dresden dataset are tabulated in Table 5.24 and Table 5.25.

Table 5.24: Performance Detection of source camera identification on Original images

	Kharrazi et al. (2004)	Kulkarni et al. (2015)	Filipczuk et al. (2012)	Singh (2016)	Proposed Technique
PR (%)	93	70.60	82.30	75.90	99.33
FPR (%)	2.40	10.40	5.90	8.10	0.20

Table 5.25: Performance Detection of source camera identification on OSN images

	Kharrazi et al. (2004)	Kulkarni et al. (2015)	Filipczuk et al. (2012)	Singh (2016)	Proposed Technique
PR (%)	88.00	57.10	75.20	71.20	99.70
FPR (%)	4.00	14.20	8.40	9.60	0.10

Based on the above comparison, it was clear that all techniques are capable of giving high detection accuracy of source camera identification on original images derived from *DI-I* dataset with the precision rate above 70%. The proposed technique has proven its better performance compare to the other technique with the highest detection accuracy at 99.33% precision rate. The proposed technique has also provided the lowest false positive rate at 0.20% for the original images.

However, experiment results for source camera identification on OSN images highlighted the advantages of the proposed technique compared to the others. While all other techniques performance dropped with lower precision rate and higher false positive rate for OSN images, the proposed technique performed significantly better. The proposed technique provides a higher precision rate for OSN images at 99.70% compared to original images at 99.33%. The proposed technique also gives a lower false positive rate for OSN images at 0.10% compared to original images at 0.20%. The detailed comparison of the performance and movement of the original images against OSN images for each technique depicted in Table 5.26 below.

Table 5.26: Performance movement of source camera identification from Original images against OSN images.

Performance	Precision Rate (%)			False Positive Rate (%)		
Images	**Original**	**OSN**	**Movement**	**Original**	**OSN**	**Movement**
Kharrazi et al. (2004)	93.00	88.00	- 5.38	2.40	4.00	+ 66.67
Filipczuk et al. (2012)	82.30	75.20	- 8.63	5.90	8.40	+ 42.37
Kulkarni et al. (2015)	70.60	57.10	- 19.12	10.40	14.20	+ 36.54
Singh (2016)	75.90	72.10	- 5.01	8.10	9.60	+ 18.52
Proposed Technique	99.33	99.70	+ 0.37	0.20	0.10	- 50.00

5.5 Chapter Summary

This chapter has explained in details the experimental results for both original and OSN images, including dataset used and comparison against existing techniques. The experiments successfully proved that the proposed technique using image filtering and texture feature extraction has significantly improved the detection accuracy in identifying source camera for both original and OSN images.

The Fusion feature, a combination of GLCM and GLRLM feature has proven capable in providing high detection accuracy with high precision rate and low false positive rate. The used of classifiers is also improves the detection accuracy. Three classifiers were selected for this experiment comprises Naïve Bayes, Lib-SVM and MLP based on the research conducted on most common classifier used in existing techniques and research. Those three classifiers are also universal classifiers and have been implemented in the image processing application (X. He et al.,2010). Among those three classifiers, MLP has proven best suited the proposed technique with the highest detection accuracy compared to other classifiers.

CHAPTER 6: CONCLUSION

These chapter deliberate actual findings of the conducted research work and its contribution, followed by achievement in reaching research objectives set earlier. The chapter will end with the deliberation of future work pertaining to this research.

6.1 Contribution of the Research

The main objective of this research is to develop and propose a source camera identification technique that can contribute to high detection accuracy especially on OSN images. The objectives were successfully met with the proposed technique, Fusion feature which is a combination of GLCM and GLRLM texture features. The Fusion feature, under MLP classifier, has provided significantly high detection accuracy with high precision rate and low false positive rate in both original and OSN images, better than the existing source camera identification techniques.

The contributions of this research are as follows:

- A thorough analysis of various existing source camera identification techniques was conducted in this research to assess the limitation and area of improvement. The challenges and limitation in the technique used by Kharrazi et al. (2004) and Kulkarni et al. (2015) to identify the camera source of modified images from other sources (OSNs web) were selected as the base case to be improved further in this research. An improvised technique subsequently was proposed based on texture feature. Texture feature will provide the list of all statistical features in digital images which facilitate source camera identification. A unique value generated via texture feature during the classification process is also fundamental in source camera identification technique.

- Three classifiers were used in this research comprises of Naïve Bayes, Lib-SVM and Multi-Layer Perceptron to classify the source images. The experiment results have proved that these three classifiers are able to identify and detect the source camera for both original and OSN images.
- The benchmarking results indicated that the proposed technique offers better identification and detection performance, compared to the-state-art-of techniques.

6.2 Achievement of the Research Objectives

The achievements of the research objective in this study are further elaborated in the following sections:

1. To study the current source camera identification techniques.

    - A comparative study of existing source camera identification techniques is thoroughly conducted in this research. List of available techniques was recorded and tabulated in Chapter 2. It can be concluded that the existing source camera identification techniques were mainly focusing on original images and intentionally modified images. None of the techniques was proven effective and having good detection accuracy when applied to OSN images. With this finding, the first objective of this research was successfully achieved.

2. To propose a suitable technique for source camera identification targeting on OSN images.

    - A new technique has been proposed in this research. GLCM and GLRLM features were chosen to be extracted from the digital images due to their effective performance in image recognition, text classification and biological information processing. The combination of GLCM and GLRLM features, which is also the proposed technique, has proven capable in giving high detection accuracy for source camera identification on both original and OSN images. The results were proven better than the existing techniques. Therefore, the second objective was successfully achieved.

3. To evaluate the proposed technique in terms of its accuracy.

    - Performance of the proposed technique was tested and compared against the-state-of-art techniques. Results show that the proposed technique performs better with higher detection accuracy compared to the-state-of-art techniques. The proposed technique, Fusion feature has also provided the highest precision rate and lowest false positive rate in both original and OSN images compared to the-state-of-art techniques. Hence the third objective of this research was successfully achieved.

6.3 Future Work

The limitation and area of improvement from previous research on the technique used for source camera identification on OSN images were deliberately explained and addressed with the proposed technique. Not only limited to OSN images, but the proposed technique has also performed significantly well with high detection accuracy for original images. This proves that the proposed technique is capable of giving good detection accuracy in both original and OSN images.

Nevertheless, there is still an area of improvement for future works on this research. It generally includes but not limited to the following:

i.) The proposed technique achieved significantly high detection accuracy over 90% for both original and OSN images, as indicated in Chapter 5. However, this proposed technique need to be tested on different types of images source such as computer-generated image or images generated from scanner devices.

ii.) This research used GLCM and GLRLM features in identifying the camera source for both original and OSN images. This is mainly due to the capability of both GLCM and GLRLM feature in image recognition, text classification and biological information processing used in previous research. However, it not necessarily meant that GLCM and GLRLM features are the best features to provide the most accurate detection rate. Hence, future improvement via other texture analysis techniques can be applied in this research to cover more features and wider comparison scope. The following texture analysis techniques can be considered for future studies since both have also capable of giving good results in classification and identification problem.

- Local binary patterns
- Tamura Texture Feature

Wider comparative studies and evaluation to include other texture features and bigger datasets will be beneficial for future research, with a possibility in getting close to 100% accuracy or at least an improvement from the proposed technique in this research for source camera identification.

REFERENCES

Abdalla, A., & Yayilgan, S. Y. (2014). *A review of using online social networks for investigative activities.* Paper presented at the International Conference on Social Computing and Social Media.

Acquisti, A., & Gross, R. (2006). *Imagined communities: Awareness, information sharing, and privacy on the Facebook.* Paper presented at the Privacy enhancing technologies.

Adamic, L. A., Lento, T. M., Adar, E., & Ng, P. C. (2016). *Information evolution in social networks.* Paper presented at the Proceedings of the ninth ACM international conference on web search and data mining.

Alles, E. J., Geradts, Z. J., & Veenman, C. J. (2009). Source camera identification for heavily jpeg compressed low resolution still images. *Journal of forensic sciences, 54*(3), 628-638.

Avcibas, I., Bayram, S., Memon, N., Ramkumar, M., & Sankur, B. (2004). *A classifier design for detecting image manipulations.* Paper presented at the Image Processing, 2004. ICIP'04. 2004 International Conference on.

Baer, R. (2010). *Resolution Limits in Digital Photography: the Looming End of the Pixel Wars.* Paper presented at the Imaging systems.

Ballou, S. (2010). *Electronic crime scene investigation: A guide for first responders*: Diane Publishing.

Baltazar, J., Costoya, J., & Flores, R. (2009). The real face of koobface: The largest web 2.0 botnet explained. *Trend Micro Research, 5*(9), 10.

Barni, M., & Bartolini, F. (2004). *Watermarking systems engineering: enabling digital assets security and other applications*: CRC Press.

Battiato, S., Bruna, A. R., Messina, G., & Puglisi, G. (2010). *Image processing for embedded devices*: Bentham Science Publishers.

Bayram, S., Sencar, H., Memon, N., & Avcibas, I. (2005). *Source camera identification based on CFA interpolation.* Paper presented at the Image Processing, 2005. ICIP 2005. IEEE International Conference on.

Bayram, S., Sencar, H. T., Memon, N., & Avcibas, I. (2006). Improvements on source camera-model identification based on CFA interpolation. *Proc. of WG, 11*, 24-27.

Bianchi, T., & Piva, A. (2012). Image forgery localization via block-grained analysis of JPEG artifacts. *IEEE Transactions on Information Forensics and Security, 7*(3), 1003-1017.

Bilge, L., Strufe, T., Balzarotti, D., & Kirda, E. (2009). *All your contacts are belong to us: automated identity theft attacks on social networks.* Paper presented at the Proceedings of the 18th international conference on World wide web.

Blythe, P., & Fridrich, J. (2004). *Secure digital camera.* Paper presented at the Digital Forensic Research Workshop.

Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2011). *The socialbot network: when bots socialize for fame and money.* Paper presented at the Proceedings of the 27th Annual Computer Security Applications Conference.

Bovik, A. C., Clark, M., & Geisler, W. S. (1990). Multichannel texture analysis using localized spatial filters. *IEEE Transactions on pattern analysis and machine intelligence, 12*(1), 55-73.

Buccafurri, F., Lax, G., Nicolazzo, S., & Nocera, A. (2015). Comparing Twitter and Facebook user behavior: Privacy and other aspects. *Computers in human behavior, 52*, 87-95.

Cao, H., & Kot, A. C. (2009). Accurate detection of demosaicing regularity for digital image forensics. *IEEE Transactions on Information Forensics and Security, 4*(4), 899-910.

Cao, Q., Sirivianos, M., Yang, X., & Pregueiro, T. (2012). *Aiding the detection of fake accounts in large scale social online services.* Paper presented at the Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation.

Carrier, B. (2003). Defining digital forensic examination and analysis tools using abstraction layers. *International Journal of Digital Evidence, 1*(4), 1-12.

Carrier, B., & Spafford, E. H. (2004). *An event-based digital forensic investigation framework.* Paper presented at the Digital forensic research workshop.

Casey, E. (2011). *Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet*: Academic Press.

Castiglione, A., Cattaneo, G., & De Santis, A. (2011). *A forensic analysis of images on online social networks.* Paper presented at the Intelligent Networking and Collaborative Systems (INCoS), 2011 Third International Conference on.

Chang, I.-C., Yu, J. C., & Chang, C.-C. (2013). A forgery detection algorithm for exemplar-based inpainting images using multi-region relation. *Image and Vision Computing, 31*(1), 57-71.

Chen, W., Shi, Y. Q., & Xuan, G. (2007). *Identifying computer graphics using HSV color model and statistical moments of characteristic functions.* Paper presented at the Multimedia and Expo, 2007 IEEE International Conference on.

Chen, Y.-C., Hu, Y., & Li, C.-T. (2011). *Further studies on forensic features for source camera identification.* Paper presented at the Imaging for Crime Detection and Prevention 2011 (ICDP 2011), 4th International Conference on.

Choi, K. S., Lam, E. Y., & Wong, K. K. (2006). *Source camera identification by JPEG compression statistics for image forensics.* Paper presented at the TENCON 2006. 2006 IEEE Region 10 Conference.

Chu, A., Sehgal, C. M., & Greenleaf, J. F. (1990). Use of gray value distribution of run lengths for texture analysis. *Pattern Recognition Letters, 11*(6), 415-419.

Conners, R. W., & Harlow, C. A. (1980). A theoretical comparison of texture algorithms. *IEEE Transactions on pattern analysis and machine intelligence*(3), 204-222.

Cox, I., Miller, M., Bloom, J., Fridrich, J., & Kalker, T. (2007). *Digital watermarking and steganography*: Morgan Kaufmann.

Cox, I. J., Miller, M. L., Bloom, J. A., & Honsinger, C. (2002). *Digital watermarking* (Vol. 1558607145): Springer.

Cross, G. R., & Jain, A. K. (1983). Markov random field texture models. *IEEE Transactions on pattern analysis and machine intelligence*(1), 25-39.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *JOSA A, 2*(7), 1160-1169.

Davis, C., Philipp, A., & Cowen, D. (2004). *Hacking exposed computer forensics*: McGraw-Hill, Inc.

Deng, Z., Gijsenij, A., & Zhang, J. (2011). *Source camera identification using auto-white balance approximation.* Paper presented at the Computer Vision (ICCV), 2011 IEEE International Conference on.

Dirik, A. E., Sencar, H. T., & Memon, N. (2007). *Source camera identification based on sensor dust characteristics.* Paper presented at the Signal Processing Applications for Public Security and Forensics, 2007. SAFE'07. IEEE Workshop on.

Dirik, A. E., Sencar, H. T., & Memon, N. (2008). Digital single lens reflex camera identification from traces of sensor dust. *IEEE Transactions on Information Forensics and Security, 3*(3), 539-552.

Dwyer, C., Hiltz, S., & Passerini, K. (2007). Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. *AMCIS 2007 Proceedings*, 339.

Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210-230.

Ellison, N. B., Steinfield, C., & Lampe, C. (2007a). The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication, 12*(4), 1143-1168.

Ellison, N. B., Steinfield, C., & Lampe, C. (2007b). The Benefits of Facebook" Friends:" Social Capital and College Students' Use of Online Social Network Sites. *Benefits, 3*, 11.

Elyashar, A., Fire, M., Kagan, D., & Elovici, Y. (2013). *Homing socialbots: intrusion on a specific organization's employee using Socialbots*. Paper presented at the Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.

Erlandsson, F., Boldt, M., & Johnson, H. (2012). *Privacy threats related to user profiling in online social networks.* Paper presented at the Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom).

Fan, W., Wang, K., Cayre, F., & Xiong, Z. (2013). *JPEG anti-forensics using non-parametric DCT quantization noise estimation and natural image statistics.* Paper presented at the Proceedings of the first ACM workshop on Information hiding and multimedia security.

Fan, Z., & De Queiroz, R. (2000). *Maximum likelihood estimation of JPEG quantization table in the identification of bitmap compression history.* Paper presented at the Image Processing, 2000. Proceedings. 2000 International Conference on.

Farid, H. (2006). Digital doctoring: how to tell the real from the fake. *Significance, 3*(4), 162-166.

Farid, H. (2008). Digital image ballistics from JPEG quantization: A followup study. *Department of Computer Science, Dartmouth College, Tech. Rep. TR2008-638, 7*, 1-28.

Farid, H. (2009a). Image forgery detection. *Signal Processing Magazine, IEEE, 26*(2), 16-25.

Farid, H. (2009b). Seeing is not believing. *IEEE Spectrum, 46*(8).

Filipczuk, P., Fevens, T., Krzyżak, A., & Obuchowicz, A. (2012). GLCM and GLRLM based texture features for computer-aided breast cancer diagnosis. *Journal of Medical Informatics & Technologies, 19*.

Fogel, J., & Nehmad, E. (2009). Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in human behavior, 25*(1), 153-160.

Friedman, G. L. (1993). The trustworthy digital camera: Restoring credibility to the photographic image. *IEEE Transactions on consumer electronics, 39*(4), 905-910.

Gagan Mehra. (2015). 91 Leading Social Networks Worldwide Retrieved 19 June, 2017, from http://www.practicalecommerce.com/91-Leading-Social-Networks-Worldwide

Galloway, M. M. (1974). Texture analysis using grey level run lengths. *NASA STI/Recon Technical Report N, 75*.

Gao, H., Hu, J., Huang, T., Wang, J., & Chen, Y. (2011). Security issues in online social networks. *Internet Computing, IEEE, 15*(4), 56-63.

Gao, H., Hu, J., Wilson, C., Li, Z., Chen, Y., & Zhao, B. Y. (2010). *Detecting and characterizing social spam campaigns.* Paper presented at the Proceedings of the 10th ACM SIGCOMM conference on Internet measurement.

Gao, S., Hu, R.-M., & Tian, G. (2012). Using multi-step transition matrices for camera model identification. *International Journal of Hybrid Information Technology, 5*(2), 275-288.

Gao, S., Xu, G., & Hu, R.-M. (2012). Camera model identification based on the characteristic of CFA and interpolation. *Digital Forensics and Watermarking*, 268-280.

Garfinkel, S. L. (2010). Digital forensics research: The next 10 years. *Digital Investigation, 7*, S64-S73.

Geradts, Z. J., Bijhold, J., Kieft, M., Kurosawa, K., Kuroki, K., & Saitoh, N. (2001). *Methods for identification of images acquired with digital cameras.* Paper presented at the Enabling Technologies for Law Enforcement.

Gloe, T., & Böhme, R. (2010). The dresden image database for benchmarking digital image forensics. *Journal of Digital Forensic Practice, 3*(2-4), 150-159.

Goljan, M., Fridrich, J., & Filler, T. (2009). *Large scale test of sensor fingerprint camera identification.* Paper presented at the IS&T/SPIE Electronic Imaging.

Gou, H., Swaminathan, A., & Wu, M. (2009). Intrinsic sensor noise features for forensic analysis on scanners and scanned images. *Information Forensics and Security, IEEE Transactions on, 4*(3), 476-491.

Gross, R., & Acquisti, A. (2005). *Information revelation and privacy in online social networks.* Paper presented at the Proceedings of the 2005 ACM workshop on Privacy in the electronic society.

Hajli, N., & Lin, X. (2016). Exploring the security of information sharing on social networking sites: The role of perceived control of information. *Journal of Business Ethics, 133*(1), 111-123.

Halder, D., Jaishankar, K., & Jaishankar, K. (2012). *Cyber crime and the victimization of women: laws, rights and regulations*: Information Science Reference Hershey, PA.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter, 11*(1), 10-18.

Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*(6), 610-621.

He, D.-C., & Wang, L. (1990). Texture unit, texture spectrum, and texture analysis. *IEEE Transactions on geoscience and remote sensing, 28*(4), 509-512.

He, X., & Xu, S. (2010). Artificial neural networks. *Process Neural Networks: Theory and Applications*, 20-42.

Ho, J. S., Au, O. C., Zhou, J., & Guo, Y. (2010). *Inter-channel demosaicking traces for digital image forensics.* Paper presented at the Multimedia and expo (ICME), 2010 IEEE international conference on.

Hu, Y., Li, C.-T., Lin, X., & Liu, B.-b. (2012). *An improved algorithm for camera model identification using inter-channel demosaicking traces.* Paper presented at the Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on.

Hu, Y., Li, C.-T., & Zhou, C. (2010). *Selecting forensic features for robust source camera identification.* Paper presented at the Computer Symposium (ICS), 2010 International.

Huang, Y., Zhang, J., & Huang, H. (2015). Camera Model Identification With Unknown Models. *IEEE Transactions on Information Forensics and Security, 10*(12), 2692-2704.

Huber, M., Mulazzani, M., Leithner, M., Schrittwieser, S., Wondracek, G., & Weippl, E. (2011). *Social snapshots: Digital forensics for online social networks.* Paper presented at the Proceedings of the 27th annual computer security applications conference.

Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. *Communications of the ACM, 50*(10), 94-100.

Jatinder Kaur, G. S. (2012). COMPREHENSIVE STUDY OF DIGITAL FORENSICS. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 1*(5), pp: 180-184.

Jin, L., Chen, Y., Wang, T., Hui, P., & Vasilakos, A. V. (2013). Understanding user behavior in online social networks: A survey. *IEEE Communications Magazine, 51*(9), 144-150.

Joe, M. M., & Ramakrishnan, D. B. (2014). A Survey of Various Security Issues in Online Social Networks. *International Journal of Computer Networks and Applications, 1*(1), 11-14.

Johnson, M. K., & Farid, H. (2006). *Exposing digital forgeries through chromatic aberration.* Paper presented at the Proceedings of the 8th workshop on Multimedia and security.

Julesz, B. (1975). Experiments in the visual perception of texture. *Scientific American, 232*(4), 34-43.

Kamalakannan, J., & Rajasekhara Babu, M. (2016). Early detection of breast cancer using GLCM feature extraction in mammograms. [Article]. *IIOAB Journal, 7*(5), 170-179.

Kang, X., Li, Y., Qu, Z., & Huang, J. (2012). Enhancing source camera identification performance with a camera reference phase sensor pattern noise. *IEEE Transactions on Information Forensics and Security, 7*(2), 393-402.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons, 53*(1), 59-68.

Karabiyik, U., Canbaz, M. A., Aksoy, A., Tuna, T., Akbas, E., Gonen, B., & Aygun, R. S. (2016). A Survey of Social Network Forensics. *Journal of Digital Forensics, Security and Law, 11*(4), 8.

Karavaras, E., Magkos, E., & Tsohou, A. (2016). *Low User Awareness Against Social Malware: An Empirical Study and Design of a Security Awareness Application.* Paper presented at the 13th Eur. Mediterr. Middle East. Conf. Inf. Syst.(EMCIS 2016).

Kee, E., Johnson, M. K., & Farid, H. (2011). Digital image authentication from JPEG headers. *IEEE Transactions on Information Forensics and Security, 6*(3), 1066-1075.

Kharrazi, M., Sencar, H. T., & Memon, N. (2004). *Blind source camera identification.* Paper presented at the Image Processing, 2004. ICIP'04. 2004 International Conference on.

Kirchner, M. (2008). *Fast and reliable resampling detection by spectral analysis of fixed linear predictor residue.* Paper presented at the Proceedings of the 10th ACM workshop on Multimedia and security.

Krishnamurthy, B., & Wills, C. E. (2009). *On the leakage of personally identifiable information via online social networks.* Paper presented at the Proceedings of the 2nd ACM workshop on Online social networks.

Kulkarni, N., & Mane, V. (2015). *Source camera identification using GLCM.* Paper presented at the Advance Computing Conference (IACC), 2015 IEEE International.

Kulkarni, S., Shirodkar, S., Hegadi, R., & Kulkarni, U. (2013). Iris Recognition using Fusion of Gray Level Co□ occurrence Matrix and Gray Level Run Length Matrix: ICSIP.

Kumar, G., & Bhatia, P. K. (2014). *A detailed review of feature extraction in image processing systems.* Paper presented at the Advanced Computing & Communication Technologies (ACCT), 2014 Fourth International Conference on.

Kurosawa, K., Kuroki, K., & Saitoh, N. (1999). *CCD fingerprint method-identification of a video camera from videotaped images.* Paper presented at the Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on.

Larroza, A., Bodí, V., & Moratal, D. (2016). Texture analysis in magnetic resonance imaging: review and considerations for future applications *Assessment of Cellular and Organ Function and Dysfunction using Direct and Derived MRI Methodologies*: InTech.

Levine, M. D. (1985). *Vision in man and machine*: McGraw-Hill College.

Li, C.-T., & Li, Y. (2012). Color-decoupled photo response non-uniformity for digital image forensics. *IEEE Transactions on Circuits and Systems for Video Technology, 22*(2), 260-271.

Li, R., Kotropoulos, C., Li, C. T., & Guan, Y. (2015). *Random subspace method for source camera identification.* Paper presented at the IEEE International Workshop on Machine Learning for Signal Processing, MLSP.

Li, R., Li, C. T., & Guan, Y. (2015). *Incremental update of feature extractor for camera identification.* Paper presented at the Proceedings - International Conference on Image Processing, ICIP.

Lin, G.-S., Chang, M.-K., & Chen, Y.-L. (2011). A passive-blind forgery detection scheme based on content-adaptive quantization table estimation. *IEEE Transactions on Circuits and Systems for Video Technology, 21*(4), 421-434.

Liu, Q., Cooper, P. A., Chen, L., Cho, H., Chen, Z., Qiao, M., . . . Sung, A. H. (2013). Detection of JPEG double compression and identification of smartphone image source and post-capture manipulation. *Applied Intelligence, 39*(4), 705.

Long, Y., & Huang, Y. (2006). *Image based source camera identification using demosaicking.* Paper presented at the 2006 IEEE Workshop on Multimedia Signal Processing.

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

Luo, W., Huang, J., & Qiu, G. (2010). JPEG error analysis and its applications to digital image forensics. *IEEE Transactions on Information Forensics and Security, 5*(3), 480-491.

Luo, W., Qu, Z., Pan, F., & Huang, J. (2007). A survey of passive technology for digital image forensics. *Frontiers of Computer Science in China, 1*(2), 166-179.

Mahdian, B., & Saic, S. (2008). Blind authentication using periodic properties of interpolation. *IEEE Transactions on Information Forensics and Security, 3*(3), 529-538.

Mahdian, B., & Saic, S. (2010). A bibliography on blind methods for identifying image forgery. *Signal Processing: Image Communication, 25*(6), 389-399.

Maini, R., & Aggarwal, H. (2009). Study and comparison of various image edge detection techniques. *International journal of image processing (IJIP), 3*(1), 1-11.

Mallat, S. G. (1989). Multifrequency channel decompositions of images and wavelet models. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 37*(12), 2091-2110.

Malleson, N., & Andresen, M. A. (2015). The impact of using social media data in crime rate calculations: shifting hot spots and changing spatial patterns. *Cartography and Geographic Information Science, 42*(2), 112-121.

Mallon, J., & Whelan, P. F. (2007). Calibration and removal of lateral chromatic aberration in images. *Pattern Recognition Letters, 28*(1), 125-135.

Materka, A., & Strzelecki, M. (1998). Texture analysis methods–a review. *Technical university of lodz, institute of electronics, COST B11 report, Brussels*, 9-11.

Mayer, O., & Stamm, M. (2016). *Improved forgery detection with lateral chromatic aberration.* Paper presented at the Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on.

Menezes, A. J., Van Oorschot, P. C., & Vanstone, S. A. (1996). *Handbook of applied cryptography*: CRC press.

Meng, F., Kong, X., & You, X. (2008). A new feature-based method for source camera identification *Advances in Digital Forensics IV* (pp. 207-218): Springer.

Mislove, A., Viswanath, B., Gummadi, K. P., & Druschel, P. (2010). *You are who you know: inferring user profiles in online social networks.* Paper presented at the Proceedings of the third ACM international conference on Web search and data mining.

Mohanaiah, P., Sathyanarayana, P., & GuruKumar, L. (2013). Image texture feature extraction using GLCM approach. *International Journal of Scientific and Research Publications, 3*(5), 1.

Morgan, S. (2016). Cyber crime costs projected to reach $2 trillion by 2019. *Forbes. Retrieved September, 22*.

Myint, S. W., Lam, N. S.-N., & Tyler, J. M. (2004). Wavelets for urban spatial feature discrimination. *Photogrammetric Engineering & Remote Sensing, 70*(7), 803-812.

Nance, K., Hay, B., & Bishop, M. (2009). *Digital forensics: defining a research agenda.* Paper presented at the System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on.

Narayanan, A., & Shmatikov, V. (2009). *De-anonymizing social networks.* Paper presented at the Security and Privacy, 2009 30th IEEE Symposium on.

Nelson, B., Phillips, A., & Steuart, C. (2015). Guide to Computer Forensics and Investigations.

Nurtanio, I., Astuti, E. R., Ketut Eddy Purnama, I., Hariadi, M., & Purnomo, M. H. (2013). Classifying cyst and tumor lesion using support vector machine based on dental panoramic images texture features. *IAENG International Journal of Computer Science, 40*(1), 29-37.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence, 24*(7), 971-987.

Palmer, G. (2001). *A road map for digital forensic research.* Paper presented at the First Digital Forensic Research Workshop, Utica, New York.

Pentland, A. P. (1984). Fractal-based description of natural scenes. *IEEE Transactions on pattern analysis and machine intelligence*(6), 661-674.

Piva, A. (2013). An overview on image forensics. *ISRN Signal Processing*.

Popescu, A. C., & Farid, H. (2005). Exposing digital forgeries in color filter array interpolated images. *Signal Processing, IEEE Transactions on, 53*(10), 3948-3959.

Quick, D., & Choo, K.-K. R. (2014). Impacts of increasing volume of digital forensic data: A survey and future research challenges. *digital investigation, 11*(4), 273-294.

Raghavan, S. (2013). Digital forensic research: current state of the art. *CSI Transactions on ICT, 1*(1), 91-114.

Reith, M., Carr, C., & Gunsch, G. (2002). An examination of digital forensic models. *International Journal of Digital Evidence, 1*(3), 1-12.

Rivest, R. L., Shamir, A., & Adleman, L. (1978). A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM, 21*(2), 120-126.

Rosenfeld, A., & Weszka, J. S. (1980). Picture recognition *Digital pattern recognition* (pp. 135-166): Springer.

San Choi, K., Lam, E. Y., & Wong, K. K. (2006). Automatic source camera identification using the intrinsic lens radial distortion. *Optics express, 14*(24), 11551-11565.

Sencar, H., & Memon, N. (2008). Overview of state-of-the-art in digital image forensics. *Algorithms, Architectures and Information Systems Security, 3*, 325-348.

Singh, K. (2016). A Comparison of Gray-Level Run Length Matrix and Gray-Level Co-Occurrence Matrix Towards Cereal Grain Classification. *International Journal of Computer Engineering and Technology, 7*(6), 9-17.

Spiekermann, S., Krasnova, H., Koroleva, K., & Hildebrand, T. (2010). Online social networks: Why we disclose.

Stamm, M., & Liu, K. R. (2008). *Blind forensics of contrast enhancement in digital images.* Paper presented at the Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on.

Statista Inc. (2017). Number of social media users worldwide from 2010 to 2020 (in billions) Retrieved 17 June, 2017, from https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

Stringhini, G., Wang, G., Egele, M., Kruegel, C., Vigna, G., Zheng, H., & Zhao, B. Y. (2013). *Follow the green: growth and dynamics in twitter follower markets.* Paper presented at the Proceedings of the 2013 conference on Internet measurement conference.

Strzelecki, M., & Materka, A. (1997). *Markov random fields as models of textured biomedical images.* Paper presented at the Proc. 20th National Conf. Circuit Theory and Electronic Networks KTOiUE'97.

Suber, P. (2009). Timeline of the open access movement.

Subrahmanyam, K., Reich, S. M., Waechter, N., & Espinoza, G. (2008). Online and offline social networks: Use of social networking sites by emerging adults. *Journal of applied developmental psychology, 29*(6), 420-433.

Sutcu, Y., Bayram, S., Sencar, H. T., & Memon, N. (2007). *Improvements on sensor noise based source camera identification.* Paper presented at the 2007 IEEE International Conference on Multimedia and Expo.

Swaminathan, A., Wu, M., & Liu, K. R. (2007). Nonintrusive component forensics of visual sensors using output images. *IEEE Transactions on Information Forensics and Security, 2*(1), 91-106.

Swaminathan, A., Wu, M., & Liu, K. R. (2008). Digital image forensics via intrinsic fingerprints. *IEEE Transactions on Information Forensics and Security, 3*(1), 101-117.

Tang, X. (1998). Texture information in run-length matrices. *Image Processing, IEEE Transactions on, 7*(11), 1602-1609.

Valsesia, D., Coluccia, G., Bianchi, T., & Magli, E. (2015a). *Image retrieval based on compressed camera sensor fingerprints.* Paper presented at the Proceedings - IEEE International Conference on Multimedia and Expo.

Valsesia, D., Coluccia, G., Bianchi, T., & Magli, E. (2015b). Large-Scale Image Retrieval Based on Compressed Camera Identification. [Article]. *IEEE Transactions on Multimedia, 17*(9), 1439-1449. doi: 10.1109/tmm.2015.2455417

Van Lanh, T., Chong, K.-S., Emmanuel, S., & Kankanhalli, M. S. (2007). *A survey on digital camera image forensic methods.* Paper presented at the Multimedia and Expo, 2007 IEEE International Conference on.

Van, L. T., Emmanuel, S., & Kankanhalli, M. S. (2007). *Identifying source cell phone using chromatic aberration.* Paper presented at the Multimedia and Expo, 2007 IEEE International Conference on.

Vincenzo Cosenza. (2012). World Map of Social Networks Retrieved 21 FEB 2017, from http://vincos.it/world-map-of-social-networks/

Wahab, A. W. A., Ho, A. T., & Li, S. (2012). *Inter-camera model image source identification with conditional probability features.* Paper presented at the Proceedings of IIEEJ 3rd Image Electronics and Visual Computing Workshop (IEVC 2012).

Wang, B., Guo, Y., Kong, X., & Meng, F. (2009). *Source camera identification forensics based on wavelet features.* Paper presented at the Intelligent Information Hiding and Multimedia Signal Processing, 2009. IIH-MSP'09. Fifth International Conference on.

Wang, L., & He, D.-C. (1990). Texture classification using texture spectrum. *Pattern Recognition, 23*(8), 905-910.

Wang, P., Xu, B., Wu, Y., & Zhou, X. (2015). Link prediction in social networks: the state-of-the-art. *Science China Information Sciences, 58*(1), 1-38.

Weszka, J. S., Dyer, C. R., & Rosenfeld, A. (1976). A comparative study of texture measures for terrain classification. *IEEE Transactions on Systems, Man, and Cybernetics*(4), 269-285.

Wolak, J., Finkelhor, D., Mitchell, K. J., & Ybarra, M. L. (2008). Online" predators" and their victims: myths, realities, and implications for prevention and treatment. *American Psychologist, 63*(2), 111.

Wu, G., Kang, X., & Liu, K. R. (2012). *A context adaptive predictor of sensor pattern noise for camera source identification.* Paper presented at the 2012 19th IEEE International Conference on Image Processing.

Xu, B., Wang, X., Zhou, X., Xi, J., & Wang, S. (2016). Source camera identification from image texture features. *Neurocomputing*.

Xu, G., & Shi, Y. Q. (2012, 2012). *Camera model identification using local binary patterns*.

Xu, W., Zhang, F., & Zhu, S. (2010). *Toward worm detection in online social networks.* Paper presented at the Proceedings of the 26th Annual Computer Security Applications Conference.

Ybarra, M. L., & Mitchell, K. J. (2008). How risky are social networking sites? A comparison of places online where youth sexual solicitation and harassment occurs. *Pediatrics, 121*(2), e350-e357.

Yerushalmy, I., & Hel-Or, H. (2011). Digital image forgery detection based on lens and sensor aberration. *International journal of computer vision, 92*(1), 71-91.

Yu, J., Craver, S., & Li, E. (2011). *Toward the identification of DSLR lenses by chromatic aberration.* Paper presented at the IS&T/SPIE Electronic Imaging.

Zhu, B. B., Swanson, M. D., & Tewfik, A. H. (2004). When seeing isn't believing [multimedia authentication technologies]. *IEEE Signal Processing Magazine, 21*(2), 40-49.