Temporal Image Forensics for Picture Dating based on Machine Learning

FARAH NAFEES AHMED

A thesis submitted in partial fulfilment of the requirements of the University of Northumbria at Newcastle for the degree of Doctor of Philosophy

Research undertaken in the
Faculty of Engineering and Environment

March 2022

Declaration

I declare that no outputs submitted for this degree have been submitted for a research degree of any other institution. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others.

Any ethical clearance for the research presented in this commentary has been approved. Approval has been sought and granted by the Faculty Ethics Committee.

I declare that the Word Count of this Thesis is 42784 words

Name: Farah Nafees Ahmed

Signature: Farah Nafees Ahmed

Date: 02/03/2022

Acknowledgements

First of all, I am grateful to **Almighty Allah** for His grace and wisdom bestowed upon me, giving me the strength, ability to learn and good health in order to complete this research.

I would like to express the deepest appreciation and sincere thanks to my principal supervisor Dr. Fouad Khelifi for providing me with all help, consistent support and constructive direction during my PhD. This thesis would not have been possible without his tireless guidance, patience, motivation, kind supervision and immense knowledge. I am extremely grateful and indebted to him for his expert, sincere and valuable support during my PhD. It is a great honour to work under his supervision. I am truly grateful that he is my mentor. His support was essential to my success here.

I also would like to express my sincere gratitude to Dr. Ahmed Bouridane, who always provide me with a good advice, good examples, encouragement and support. My sincere thanks also go to Dr. Nauman Aslam and Dr. Ammar Belatreche for offering me the teaching opportunities and leading me working on diverse modules.

I would like to express my extreme thanks to Dr. Ashref Lawgaly for sharing his knowledge, time and expertise. His kind endless help, generous advice, suggestions, meaningful assistance and relentless support in every aspect of my research was imperative to my completion of this thesis. Thank you so much for always being there when I needed you.

In addition, I acknowledge research assistant Al-Ani Mustafa and PhD student Ashref Lawgaly, who were significantly involved in taking pictures for the NTIF (Northumbria Temporal Image Forensic) database over a period of time.

Last but not the least, I would like to thank my family, my parents and my friends to support and encourage me throughout my life.

Publications

The following papers have been published.

- Ahmed, F., Khelifi, F., Lawgaly, A. and Bouridane, A., 2019, January. Comparative analysis of a deep convolutional neural network for source camera identification. In *2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3)* (pp. 1-6). IEEE.

- Ahmed, F., Khelifi, F., Lawgaly, A. and Bouridane, A., 2020, June. The 'Northumbria Temporal Image Forensics' Database: Description and Analysis. In *2020 7th International Conference on Control, Decision and Information Technologies (CoDIT)* (Vol. 1, pp. 982-987). IEEE.

- Ahmed, F., Khelifi, F., Lawgaly, A. and Bouridane, A., 2020, August. Temporal Image Forensic Analysis for Picture Dating with Deep Learning. In *2020 International Conference on Computing, Electronics & Communications Engineering (iCCECE)* (pp. 109-114). IEEE.

- Ahmed, F.N., Khelifi, F., Lawgaly, A. and Bouridane, A., 2021. A Machine learning-based approach for Picture Acquisition Timeslot Prediction using Defective Pixels. *Forensic Science International: Digital Investigation, 39.*

ABSTRACT

Temporal image forensics involves the investigation of multi-media digital forensic material related to crime with the goal of obtaining accurate evidence concerning activity and timing to be presented in a court of law. Because of the ever-increasing complexity of crime in the digital age, forensic investigations are increasingly dependent on timing information. The simplest way to extract such forensic information would be the use of the EXIF header of picture files as it contains most of the information. However, these header data can be easily removed or manipulated and hence cannot be evidential, and so estimating the acquisition time of digital photographs has become more challenging.

This PhD research proposes to use image contents instead of file headers to solve this problem. In this thesis, a number of contributions are presented in the area of temporal image forensics to predict picture dating. Firstly, the present research introduces the unique Northumbria Temporal Image Forensics (NTIF) database of pictures for the purpose of temporal image forensic purposes. As digital sensors age, the changes in Photo Response Non-Uniformity (PRNU) over time have been highlighted using the NTIF database, and it is concluded that PRNU cannot be useful feature for picture dating application. Apart from the PRNU, defective pixels also constitute another sensor imperfection of forensic relevance. Secondly, this thesis highlights the fact that the filter-based PRNU technique is useful for source camera identification application as compared to deep convolutional neural networks when limited amounts of images under investigation are available to the forensic analyst. The results concluded that due to sensor pattern noise feature which is location-sensitive, the performance of CNN-based approach declines because sensor pattern noise image blocks are fed at different locations into CNN for the same category. Thirdly, the deep learning technique is applied for picture dating, which has shown promising results with performance levels up to 80% to 88% depending on the digital camera used. The key findings indicate that a deep learning approach can successfully learn the temporal changes in image contents, rather than the sensor pattern noise.

Finally, this thesis proposes a technique to estimate the acquisition time slots of digital pictures using a set of candidate defective pixel locations in non-overlapping image blocks. The temporal behaviour of camera sensor defects in digital pictures are analyzed using a machine learning technique in which potential candidate defective pixels are determined according to the related pixel neighbourhood and two proposed features called local variation features. The idea of virtual timescales using halves of real

time slots and a combination of prediction scores for image blocks has been proposed to enhance performance. When assessed using the NTIF image dataset, the proposed system has been shown to achieve very promising results with an estimated accuracy of the acquisition times of digital pictures between 88% and 93%, exhibiting clear superiority over relevant state-of-the-art systems.

Table of Contents

LIST OF ABBREVIATIONS

PRNU:	Photo Response Non-Uniformity
EXIF:	Exchangeable Image File.
SPN:	Sensor Pattern Noise.
CCD:	Charged Coupled Device.
CMOS:	Complementary Metal Oxide Semiconductor.
CFA:	Colour Filter Array.
R-G-B:	Red Green Blue.
DWT:	Discrete Wavelet Transform.
DCT:	Discrete Cosine Transform.
DIF:	Digital Image Forensics.
SCI:	Source Camera Identification.
JPEG:	Joint Photographic Experts Group.
TIFF:	Tagged Image File Format.
NTIF:	Northumbria Temporal Image Forensic.
DL:	Deep Learning.
SVM:	Support Vector Machine.
NB:	Naives Bayes.
KNN:	k-Nearest Neighbour.
LV:	Local Variation.
FPN:	Fixed Pattern Noise.
NC:	Normalized Correlation.
PCE:	Peak to Correlation Energy.
PCA:	Principal Component Analysis.
CCN:	Circular Cross-Correlation Norm.
FPN:	Fixed Pattern Noise.
FPR:	False Positive Rate.

FNR:	False Negative Rate.
TPR:	True Positive Rate.
TNR:	True Negative Rate.
MAP:	Maximum a Posteriori Probability.
MLE:	Maximum Likelihood Estimator.
CAI:	Context Adaptive Interpolation.
BM3D:	Block-matching 3D filtering.
CNN:	Convolutional Neural Network.

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 MOTIVATION

In this new era of digital technology and with the usage of countless devices and applications, it is possible to keep track of pattern-of-life data points by the year, month, day, hour or even second. Meanwhile, data used in contemporary forensic investigations linked to crime events can be physical, digital or relational. The production of the timelines of events helps to resolve many forensic investigations, which are often now highly dependent on digital traces when physical traces are not present or tell only part of a story. Many such case studies are based on visualizing and constructing the sequences of events or activities captured by digital devices. The triple murder of Cody, Chad and Margaret Amato in Florida in January 2019 is a recent example of a temporal digital case which attracted much attention when forensic investigators handled the case by focusing on establishing the timeline of events captured by 35 different digital devices (Focus, 2020). Similarly, in 2015, a Connecticut man was arrested and charged with the killing of his wife after forensic staff produced a minute-by-minute timeline down of all of her movements during the relevant day which did not match his alibi. Nowadays, massive intensive forensic searches are based on the analysis of crime events to complete investigations involving murder, kidnap or rape. The temporal analysis of a specific device or set of devices helps the forensic analyst to frame the link of any crime incident or other event, and it mainly involves the collection of information within a particular timeframe so that the investigator can determine exactly when a particular crime or associated events happened (Ryser, Spichiger and Casey, 2020). Temporal digital evidence can provide details of event sequences, activity levels and timing. Based on such evidences, the forensic analyst can answer questions about victims or suspects (‘who’), activities (‘what’), places (‘where’) and times (‘when’). Images and other visual recordings are important means to document at any instant in time the condition of specific subjects, which might include scenes related to a crime or accident, victims, suspects or any items of evidence. The forensic analyst can face challenges if there is any discrepancies or unreliability in digital evidences (Casey, 2020). Such thorough and comprehensive photographic images of a scene, suspect or evidence from recordings made by a camera or any digital imaging technology can be helpful in identifying meaningful information which can lead to substantial or circumstantial evidence. If such images are presented in the court for investigation purposes, it is crucial that every possible measure should be taken to ensure

the reliability and accuracy of picture dating. To correlate digital artifacts, the forensic analyst needs to process the correct ordering timeline of any relevant images to allow an understanding of the sequence of events. Helpful clues might be found in the file's header (EXIF) when establishing a timeline as part of a digital forensic investigation. But, since this information could easily have been altered, it may be a significant challenge for the analyst to estimate the picture dating. Motivated by this, this thesis will develop and investigate new methods and techniques to estimate the precise times of acquisition of digital images using image contents. In this regard, in temporal forensic analysis it is very useful to place the digital images of interest in temporally ordered clusters. Classifying images in this way is helpful for the analyst in plotting the significant events relevant to the timespan in which they occurred, and in general assisting in keeping track of the findings.

1.2 RESEARCH BACKGROUND

1.2.1 Multimedia Forensics

As more images and videos continue to flood the internet, the use of high-class digital technologies has changed the type of evidence that is now processed in forensic investigations. There may be many more audio, video and still images available of an actual crime than ever before. Therefore, multimedia evidence has become an important part of forensic investigations. The basic purpose of multimedia forensics is to develop efficient instruments, tools or algorithms that can deal with the disparate digital devices which can be used to manipulate an image. Multimedia forensics can be defined as the branch of digital forensic science that involves the analysis of multimedia signals (audio, videos, and images) and the extraction of electronic information from multimedia devices using a set of scientific techniques to give an assessment of such content during the investigation process and to provide interpretations for use in a court of law.

It aims to:

- reveal the history of digital content,
- identify the acquisition device that produced the data,
- validate the integrity of the contents, and
- retrieve information from multimedia signals.

The basic idea behind multimedia forensics relies on the observation that both the acquisition process and any post-processing operations leave distinctive imprints on the data, as a sort of digital fingerprint. The analysis of such fingerprints may permit the origin of an image or video to be determined and the authenticity of digital content to be established. The examination of such fingerprints in effect suggests how the digital clue should be evaluated, turning it into actual evidence.

1.2.2 Digital Image Forensics

Digital devices are transforming every aspect of our daily lives. With the advent of high-resolution devices available to consumers for the capture of images, an enormous variety of digital images now circulate and user-friendly image editing tools have become commonplace. Unfortunately, these technological advances have also become new weapons of choice for use in criminal acts. Due to the growing frequency of cyber-crime using electronic-assisted devices, it is necessary for forensic scientists and researchers to link images to source devices when the origin and integrity of such images becomes crucial in evidence in a court of law.

Digital image forensics (DIF) aims to validate the authenticity of images by recovering information about their history. It involves the processes and tools utilized in copying a physical storage device in order to conduct investigations and obtaining evidences. DIF can be divided into two categories: active forensics and passive forensics. Active forensic requires additional information to be added to digital pictures such as fragile digital watermarking or cryptographic digital signatures. For instance, it mainly entail creating various types of watermarks or digital signatures of the image content and inserting them into the digital images. The former embedded watermarks are retrieved and examined in the authentication stage to detect whether the original image has been tampered with and, if so, where the altered location is located. A key drawback of active approaches is that they require manipulation of the original image either during capture or storage. To overcome the problem, passive forensic can merely use the image's information to determine the validity or source of the image. In general, passive DIF detects image history or forgeries by looking for certain intrinsic patterns in the original images that were left during capture or storage, or specific patterns in image forgeries that were left during storage or editing (Lin *et al.*, 2018). The principles and motivations of digital image forensics (i.e.

concerning digital images and videos), are to describe the main approaches proposed so far for facing the three basic questions:

a) What is the source of digital content?

b) Is the digital content authentic or not, in the sense of determining if it is genuine and no traces of forgery can be detected?

c) Is the timing information of digital content reliable for forensic investigations?

Answering those questions is relatively easier if the original image is known. Unlike signature- or watermark-based methods, digital image forensics does not require any signature to be generated on an image or a watermark to be embedded in advance. In practice, no such information or reference is usually available for the original image and, therefore, forensic analysts need to authenticate the image's history in a passive and blind way. According to Chen *et al.* (2008), the task of digital forensics can be broadly divided into six areas, as shown in Figure 1-1, where the aim of source classification is to classify images or videos based on their origin, such as, for example, Fujifilm versus Samsung. The objective of device identification is to prove that a given image or video belongs to a particular device. The process of linking content to a device aims to group objects according to their common source. The process of history recovery aims to identify and recover the processing chain which has been performed with the image or video. In the integrity verification process, any malicious processing is discovered that has been added to or removed from the image or video. Finally, the investigation of anomalies explains features observed in an image or video that could be a result of digital processing.

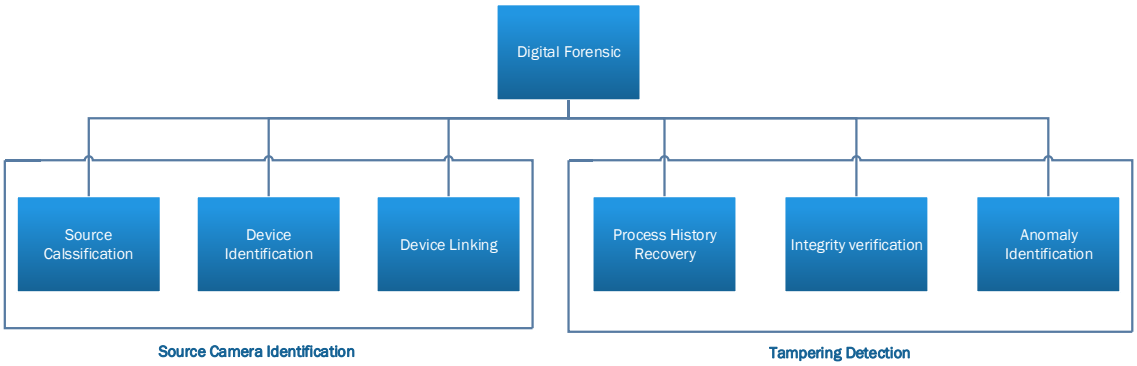


Figure 1-1 Digital Forensic Classification Source: Chen *et al.* (2008)

Source camera identification (SCI) is one of the major strategies used in digital image forensics. It is the process by which an unknown image or suspected images is linked to its source device (Gloe and Böhme, 2010). In digital image forensic investigations, it is the process used to decide which camera has been used to capture a particular image. In 2009, a photograph which became known in court as the 'Hogmanay image', which shows a man assaulting an infant, was among the pieces of evidence used to convict a suspect in a high-profile case. An investigator raids a location and discovers a male who is suspected of downloading child pornography on the Internet. A forensic check of the suspect's computer reveals a number of child pornographic photographs. The culprit was apprehended and stated that he had no knowledge of the photographs' existence. He did admit, though, that he is the one who uses that computer. Is the evidence sufficient to establish the case in court? Can the digital evidences provide more information on: What has the computer user done? When were the images downloaded? Have the photographs been viewed by the user of the computer? Do the images exist and are the images known to the computer user? To answer those questions, the sub-field of image forensics has made significant advances due to efficient exploitation of the imperfections of digital imaging sensors. Figure 1-2 shows the three standard elements (known as artifacts) named as acquisition artifacts, processing artifacts and storage artifacts of a typical digital camera that are used for source camera identification. These three artifacts used for SCI purposes are explained in the following sub-sections.

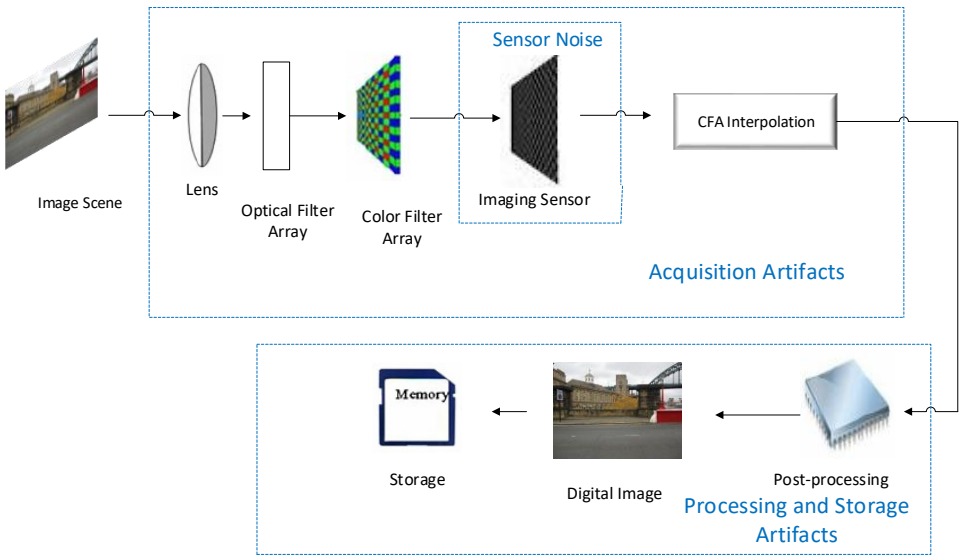


Figure 1-2 The standard components (acquisition, processing and storage artifacts) of the typical digital image acquisition process

1.2.3 Source Camera Identification using Acquisition Artifacts

Three main approaches have been proposed in recent decades to solve source camera identification problems using the acquisition phase of a digital camera. The first approach based on the analysis of artefacts produced in the acquisition phase is lens distortion. A lens aberration is an optical property occurring when light passes through the lens, whereupon it diverges over some region of space instead of focusing towards a point. The effect of subsequent radial distortions affect the straight trajectory of light so that it appears as curved on the output image. This lens radial distortion is then used as a fingerprint to identify the source camera (San Choi, Lam and Wong, 2006). However, this radial distortion cannot be measured if no straight line occurs in the image, since it is computed using the straight-line method. Meanwhile, in addition to intrinsic lens aberrations, chromatic aberration is also produced due to imperfections in the lens (Van, Emmanuel and Kankanhalli, 2007). So, the same distinctive pattern is formed which can be used to identify a camera of the same model, but different brands of camera cannot be distinguished.

The second approach is to use sensor pattern noise (SPN), which can allow different camera brands as well as cameras with the same make and model to be distinguished. Sensor noise is the result of sensor imperfections which can result from three main causes which are pixel defects, fixed pattern noise (FPN) and photo response non-uniformity (PRNU). In the sensor chips, the pixels can be defective. These defects can be of different types, such as point defects, hot pixel defects, cluster defects and stuck pixel defects. These defective pixels cannot capture light information in the image scene and thus they are fixed and formed a pixel defect pattern. One such pattern for a digital camera device is shown in Figure 1-3. The pixel defect pattern can be extracted in order to identify the source camera. However, the pattern can be affected by the sensor's sensitivity to temperature, which may render it less visible. In addition, pixel defect patterns changes over time as the sensor ages and the number of defective pixels increases. Therefore, it is not possible to define a distinctive pattern so as to reliably identify the source camera device (Geradts *et al.*, 2001). Moreover, if a number of devices are used to link an image with a source device, it could be possible that some devices would produce similar pixel defect patterns. Thus, this approach may be impractical for use in identifying the source device.

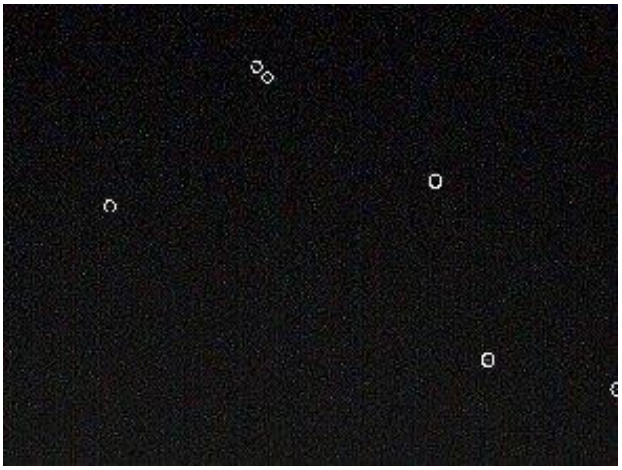


Figure 1-3 Pixel Defect Pattern for Source Camera Identification Source: (Geradts and Bijhold, 2000)

Sensor pattern noise such as PRNU is a unique pattern caused by imperfections in the silicon wafer originating during the sensor manufacturing process. Such imperfections cause pixels to have different sensitivities to light and can be used to identify a source device. It is widely accepted in the use of filtered-based machine learning approaches that PRNU as a type of camera sensor imperfection is an inherent, dominant and unique fingerprint of imaging devices which provides an effective method of source camera identification (Lukáš, Fridrich and Goljan, 2006). Sensor pattern noise is like a fingerprint for a camera and distinctions can be made between different individual cameras as it is unique to each sensor in the device. However, this approach is useful only if the reference pattern noise and test image are of the same size. It is not effective if the images are cropped or subject to video compression, which has a significant impact on sensor pattern noise. In addition, most images will have been cropped to capture the main scene.

The third approach is to use colour filter array (CFA) interpolation, which involves the examination of traces left behind on the image so that a distinctive pattern can be extracted. The CFA is a mosaic of tiny colour filters placed over the image pixel sensors to adopt colour information (Van, Emmanuel and Kankanhalli, 2007; Redi, Taktak and Dugelay, 2011). However, different digital cameras follow the same CFA interpolation, and therefore, this method can only identify the camera make and model instead of linking unique images to their source devices.

1.2.4 Source Camera identification using Post-Processing and Storage Artifacts

Apart from aspects of the acquisition phase, post-processing and storage artifacts can also provide clues to identify the source of an image. The statistical features of images were studied by Kharrazi, Sencar and Memon (2004) acting on the assumption that an image can be affected by colour processing and transformation operations before a final image is stored. The authors analyzed the properties of an image using three approaches: colour-related features, wavelet domain statistics and image quality features. The algorithm developed is only effective if the images considered depict similar contents and are captured by different camera models. Tsai and Wu (2006) subsequently employed similar image features but with different camera sets using multi-class classifiers. However, this method was not effective for similar CCD camera sensors, and also was unable to identify source cameras of the same model.

After the processing phase, most digital cameras generally encode images in JPEG format before saving them to final storage. Some manufacturers use different compression levels and parameters. In research by Farid (2006), differences in compression levels were exploited by analyzing the JPEG quantization table and comparing it against a database of digital cameras, which provides useful clues for source camera identification purpose. The author tested this method on 204 digital camera models, but found that the use of the quantization table is not sufficient to distinguish between all camera models.

When a digital image is saved in a storage format, a thumbnail image which is a reduced version of the full resolution image is also stored together so that users can conduct a quick review. This image thumbnail can also provide some fingerprint information such as relating to filtering operations, contrast adjustment and JPEG compression. The fine details of each implementation process vary according to the manufacturer of the device and the model concerned. Kee and Farid (2010) exploited the steps in the formation of the thumbnail image in order to identify the source camera. They estimated the number of parameters used by a specific device to compute different thumbnail formation steps that include cropping, pre-filtering and post-filtering operations, contrast adjustment and JPEG compression. Such parameters are useful in characterizing a single device and to distinguish between camera models. However, the authors emphasized that a database of different parameters should be available for a number of digital devices in order to facilitate this method. In work by Julliand, Nozick and Talbot (2015), it was concluded that different type of noise affected the images. The pixels inside an image

can be affected when the image is formed due to JPEG lossy compression that creates noise. They showed that a RAW image can have a different histogram than the resulting JPEG image. However, the authors did not propose any specific model for camera identification.

With the rise of deep learning approaches, the convolutional neural network (CNN) has also been exploited to solve source camera and camera model identification problems. Here, large numbers of training images are needed to train the convolutional network so as to achieve high accuracy.

1.3 PROBLEM STATEMENT WITH REGARDS TO TEMPORAL IMAGE FORENSICS

As part of digital image forensics, it is essential to define additional information such as temporal parameters to estimate the time of acquisition of digital pictures taken by a camera and to know the conditions under which each picture was taken, such as exposure, date and time. In this field, the concept behind the temporal forensic analysis of an image is that the forensic investigator tries to estimate the acquisition time of any suspect images using prior information such as the camera device used or a set of images captured by the same device at different times. However, very little research has been devoted to the extraction of temporal information from digital pictures. Consequently, there is no established technique that has been deployed in practical forensic investigations which allows the reliable extraction of such information. In these investigations, the visualization of multiple timelines in an incident or case is important to provide a complete summary of a crime. A most important aspect of a forensic investigation is to define the dating and timing of evidence. From this, the forensic analyst can establish an accurate sequence of events and timespans during the investigation and in a court case. When it comes to proof, knowing when an action was committed is often crucial, and a timeline is required to connect the chain of events contained on a digital medium to the chain of events that occurred in the physical world. The analysis and inspection of digital evidence such as CCTV footage, audio sample, image and metadata analysis, makes up about 60% of the evidence. For example, on January 26th 2021, riots in Delhi were seen as a national threat when violent protests by farmers to repeal three farm acts shocked the entire country. When the protesters began to vandalize public property and caused havoc and crisis in parts of Delhi in the guise of a peaceful demonstration, the protests became violent. Dr. Surbhi Mathur, a member of the inquiry team and a leading multimedia

Forensic scientist from National Forensic Sciences University, described how the investigation was difficult because the evidence included over ten thousand CCTV films gathered from public surveillance cameras. After a thorough review of the evidence, the main perpetrators were captured. In another case in 2021, the timeline of key events in the disappearance of Sarah Everard was eventually revealed through the analysis of data from several CCTV cameras which represented vital digital evidence in the forensic investigation and led police to the door of the suspect. Overall, the major challenge for the forensic analyst is to obtain precise dates and times from digital images, which can be accomplished by accessing EXIF headers. However, if this information is deleted or manipulated, finding the date of digital evidence in order to reach conclusions becomes a crucial task.

To the best of the author knowledge, only two research papers on temporal forensics of digital pictures have been reported in the literature. The first framework was presented by (Mao *et al.*, 2009) where the time-dependent camera parameters were assumed to be included in the PRNU. The results obtained through the correlation coefficients of PRNU estimates of different time-based image clusters showed the possibility to rank the clusters according to their acquisition date. The second research framework by (Fridrich and Goljan, 2011) used defective pixels for establishing the temporal relationship. Although, the reported results were not as good as in (Mao *et al.*, 2009) due to some implementation issues, it was clearly shown that camera sensors manifest changes with age. It is worth mentioning here that the distinction between temporal ordering or placement of an individual picture within a set of trusted pictures whose acquisition time is known, as described in (Fridrich and Goljan, 2011) and (Mao *et al.*, 2009), and picture dating, which entails estimating an acquisition date for a given picture is regardless of the time span of the trusted pictures available. In fact, the previously mentioned systems presume that the image under examination was taken during a specified time period, as determined by the known acquisition dates of a collection of trustworthy images, and attempt to locate it in relation to the trusted images. Despite the fact that such systems deliver forensically relevant information, however their applicability in real investigations is limited. In research study of (Fridrich and Goljan, 2011), the authors relied on defective pixels only and acknowledged that the maximum likelihood technique used to estimate the acquisition time of pictures was unable to perform accurately between two consecutive defect onsets because the likelihood is constant. That is, the technique returns a constant acquisition time, which is the average time of the two onsets, for any given picture. The failure to date pictures reliably is also apply to the correlation-based technique used in (Mao *et al.*, 2009). Indeed, the reported

results have clearly shown that the correlation coefficient is not linearly dependent on the time span between the acquisition date of the picture under investigation and the date corresponding to the closest image cluster. Therefore, it would be theoretically incorrect to linearly map the correlation coefficient into a time span for dating purposes. Furthermore, the use of the correlation coefficient on the PRNU noise does not take the type of pixels into account and ignores their temporal evolution. The correlation operation treats all pixels similarly, regardless of their temporal behaviour. In addition to this, the irregular timeslots of a dataset specific to temporal camera analysis in both research works limited the performance of their experimental results.

As observed in section 1.2.3, apart from PRNU, defective pixels also constitute as a sensor imperfection of forensic relevance (Geradts *et al.*, 2001). They usually refer to either dead pixels that always emit the same signal or noisy pixels that have a noise level that exceeds an acceptable threshold. Pixel defects are caused by the environment, such as high altitudes and cosmic radiation exposure (Theuwissen, 2007), or by the manufacturer, such as a high dark current caused by a silicon impurity or a surface flaw on the sensor (Dudas *et al.*, 2007). An interesting characteristic of defective pixels is that they accumulate over time, therefore, defective pixels are exploited in temporal forensics. From this perspective, establishing temporal relationships between digital pictures taken by the same camera for dating purposes is theoretically possible. In this thesis, the analyst is assumed to have the camera device or a set of trusted pictures taken by the same camera and whose acquisition time is known. The goal is then to date one or more images supposedly obtained from the same source and whose acquisition time is unknown. This research can help forensic investigators to analyse incidents and link different events.

1.4 RESEARCH AIM

The aim of this research is to investigate the relationship that exists between digital pictures taken at different times from the same camera device, and to establish a theoretical model of temporal variations in camera sensor imperfections that allows the analyst to estimate or predict the acquisition date of unknown digital pictures.

1.5 RESEARCH OBJECTIVES

The objectives of the present study are as follows:

- To explain the motivation for the research and to explore the importance of temporal digital image forensics in multimedia applications. This objective is covered in Chapter 1.
- To review existing research on camera sensor imperfections in a digital camera, including the investigation of the digital imaging pipeline, source camera identification (SCI) using both the PRNU-based method and deep learning, defective pixels, picture dating and datasets publicly available for image and video forensics. This objective is covered in Chapter 2.
- To propose the 'Northumbria Temporal Image Forensic' (NTIF) database and its relevance for temporal image forensics applications. This objective is covered in Chapter 3.
- With recent findings and the breakthrough of deep learning, this objective attempts to provide a comparative study by assessing the performance of the traditional filter-based approach in comparison with deep learning convolutional neural networks (CNNs) for source camera identification in digital forensic science. This objective is covered in Chapter 4.
- To introduce and assess for the first time the performance deep convolutional neural networks for the application of picture dating. This objective is covered in Chapter 5.
- To design and develop an efficient defective pixel location detection and classification system based on defective pixels, which takes temporal variations into account and to develop novel methods for the automatic dating of digital pictures. This objective is covered in Chapter 6.

1.6 CONTRIBUTIONS OF THE THESIS

The main contributions of this thesis can be summarized as follows:

In the first contribution of the thesis, the usefulness and relevance of the NTIF database for temporal image forensic has been demonstrated in a number of experiments which show changes in the strength of correlations with sensor pattern noise over time. These changes in PRNU correlations indicate that a sensor ages over time. In addition, different patterns of correlation were found for all digital cameras when the images were captured during different time slots as compared to images acquired within the same time slot.

Therefore, NTIF is specifically introduced for temporal digital image forensics which consists of natural images of indoor and outdoor scenes with full high resolution. The images are organized in temporal order with regular acquisition timeslots spanning 94 weeks using ten digital camera devices. 41684 images were captured from 10 digital cameras belonging to different models and brands. Subsets of the images have been annotated with labels covering categories based on a temporal factor of one to two weeks. The NTIF database has been made freely accessible to benefit all researchers in image forensics in academia and industry.

In the second contribution, this thesis sheds light on the application of source camera identification in digital image forensics via a comparative analysis of state-of-the art filter-based techniques and deep learning CNNs. To this end, an estimation technique based on the fundamental patterns of sensor noise using the Wiener filter in the wavelet domain is considered as well as a deep CNN model in order to assess their effectiveness for SCI. Various experiments are carried out using the NTIF database of images with both approaches. From the results in terms of rates of false positives and false negatives and overall accuracy, it has been shown that the filter-based PRNU approach for SCI has given better performance when compared to the CNN approach, especially when a limited number of images are available which is a common scenario in forensic investigations in practice. It is concluded that, since sensor pattern noise is location-sensitive, an approach using machine learning appears to be inefficient. This has particularly been noticed in cases where the PRNU method can distinguish between images taken from different models of the same brand, while the CNN model fails to do so.

In the third contribution of the thesis, the deep learning approach is used to predict the acquisition time of digital images in such a way that the analyst can identify the timeline of unknown digital photographs given a set of pictures from the same source whose time ordering is known. By applying the AlexNet and GoogLeNet convolutional neural network architectures in both feature extraction using CNN and pre-trained CNN model modes, the results show that the CNN models can successfully learn temporal changes in the content of digital pictures acquired from the same source device. Interestingly, although images are divided into non-overlapping blocks in order to increase the number of training samples and feed the CNNs, the estimation accuracy obtained is found to range from 80% to 88%. It is also

concluded that deep learning is good at learning temporal changes for picture dating, while PRNU is a location-sensitive pattern suitable for use in digital image forensics for SCI problems.

In the final contribution of the thesis, temporal variations in the imperfections found in camera sensors are examined so that the acquisition dates of digital pictures can be estimated from the defective pixels within the images and by using a machine learning-based approach. In this context, the technique used to predict the acquisition time slots of digital pictures is based on pixel neighbourhood and local variation features to find the best candidates of defective pixels. Once potential pixel defects are detected, a classifier is trained for each defective pixel location. A number of contributions have been made to enhance the performance of picture dating, such as: combining defective pixel detection and a machine learning technique; using local variation features for the detection of potential defective pixels for picture dating; a re-training-based approach using virtual time slots; and machine learning based on a multi-block model to boost the performance of the system. The results showed very good performance with an estimated accuracy between 88% and 93%.

This research can advance knowledge in the field of forensic science since it offers a new application of great relevance in image forensics. The computer vision and machine learning research communities can benefit from this research, as it widens the application range of machine learning techniques in computer vision for the development of intelligent vision systems. This research investigates the new techniques for efficiently estimating defective pixels in a way that is useful for temporal analysis. Techniques based on machine learning are used to develop a model of the temporal behaviour of camera sensor imperfections that can be used to predict the acquisition time of a given picture. The findings concerning sensor defect detection can be of particular relevance to researchers specializing in image processing for applications in digital forensics and security. Finally, the NTIF dataset of images that has been built up by research staff and PhD students at Northumbria University is uniquely suited for this particular application and has been made accessible to researchers. This will be of great help to the research community in image analysis and forensics across the UK and beyond. This research study is the first in Europe to address the problem of the temporal analysis of camera sensor imperfections and the first ever to address the problem of picture dating in a specific timespan between the acquisition dates of two sets of trusted pictures.

1.7 THESIS OUTLINE

The rest of this thesis is organized as follows: Chapter 2 is based on literature review in which the image acquisition process and the digital camera pipeline are described. Some fundamentals of SCI applications using measures such as PRNU relating to imperfections in digital camera sensors are briefly described. Also, a broader overview is given of image de-noising, PRNU extraction, and estimation and detection methods in order to provide a broader perspective of the significance of PRNU in SCI and temporal forensics applications. The uses of PRNU in digital image forensics applications, and particularly source camera identification problems solved in the context of machine learning and deep learning approaches, are also explained. Other imperfections in digital camera sensors such as defective pixels are detailed to show their usefulness in dating pictures. In addition, a number of image and video datasets used for multimedia forensic techniques are described. In Chapter 3, the NTIF database is introduced which is proposed for use in digital image forensics applications such as the estimation of the acquisition time of images, image classification, authentication and forgery detection, and source camera identification. Also, the usefulness and relevance of the NTIF database for temporal digital image forensics are highlighted by demonstrating changes in sensor patterns over time. Chapter 4 focuses on the comparative analysis of filtered-based techniques such as PRNU and deep learning technique using CNN method in the context of digital image forensic science for SCI problem. In Chapter 5, the time of acquisition of digital pictures is estimated with the help of deep learning, in order to show that a deep learning approach allows temporal changes in digital pictures, rather than the sensor noise pattern, to be learned. Chapter 6 addresses the problem of picture dating through the temporal forensic analysis of imperfections in digital camera sensors using defective pixels. Finally, Chapter 7 presents the conclusions of the thesis, a summary and recommendations for future work.

CHAPTER 2 LITERATURE REVIEW

2.1 THE DIGITAL IMAGE ACQUISITION PIPELINE

Most digital cameras have the same imaging pipeline, as shown below in Figure 2-1. Light from the scene passes through a lens and then through a set of optical filter arrays. Of these, an infrared filter is a reflective filter that only allows light in the visible part of the spectrum to pass, while blocking infrared radiation which can reduce the sharpness of the image formed. Anti-aliasing filters decrease aliasing and support the finer spatial frequency of the target patterns. After passing through optical filter arrays, the light hits an imaging sensor (charge-coupled device (CCD)/Complementary Metal Oxide Semiconductor (CMOS)), which is also referred to as the 'heart' of the digital camera. This sensor is composed of arrays of rows and columns of pixels. When it strikes the pixel arrays, the light is divided into a number of pixels to capture only intensity information. Image sensors are unable to differentiate between colours and can only record the brightness of light and thus provide monochromatic output (Farid, 2019). The output from the image sensor with a Bayer filter is a mosaic of red, green and blue pixels with different intensities which can acquire colour information. The colour filter array (CFA) is a specific colour mosaic (Kirchner and Böhme, 2009) which allows each pixel to detect a single colour, as shown in Figure 2-2. The most commonly preferred CFA pattern is the Bayer filter (Bayram *et al.*, 2005). A colour interpolation algorithm which conducts a process called interpolation or de-mosaicing is applied to achieve a full colour image with three colour components, such as red, green and blue channels for each pixel, by estimating missing pixel colours based on the values of existing neighbouring pixels, as also shown in Figure 2-2.

There are also various sources of noise which are present during the image acquisition process, such as photon shot (Poisson), read-out, fixed pattern (FPN), and PRNU noise. Noise sources such as readout, shot, interpolation and quantization noise that are present in the output of the signal are not dominant, and so they are not useful for forensic purposes. FPN occurs when there is no light and a small amount of current accumulates in each pixel of the sensor, and thus it is also referred to as dark current noise. FPN and PRNU noise are caused by imperfections in the process of manufacturing the sensor. The main cause of PRNU noise is the non-homogeneity of silicon wafers that provide variations in sensitivity of pixels to light. It is widely accepted that PRNU is the dominant characteristics of image sensor pattern

noise for use in SCI since it is stable and unique throughout the lifetime of the camera device. In the post-processing stage, three main stages are used for image enhancement processing, which are white balancing, de-mosaicing and gamma correction, before the final image is saved in the storage memory device in the common JPEG format.

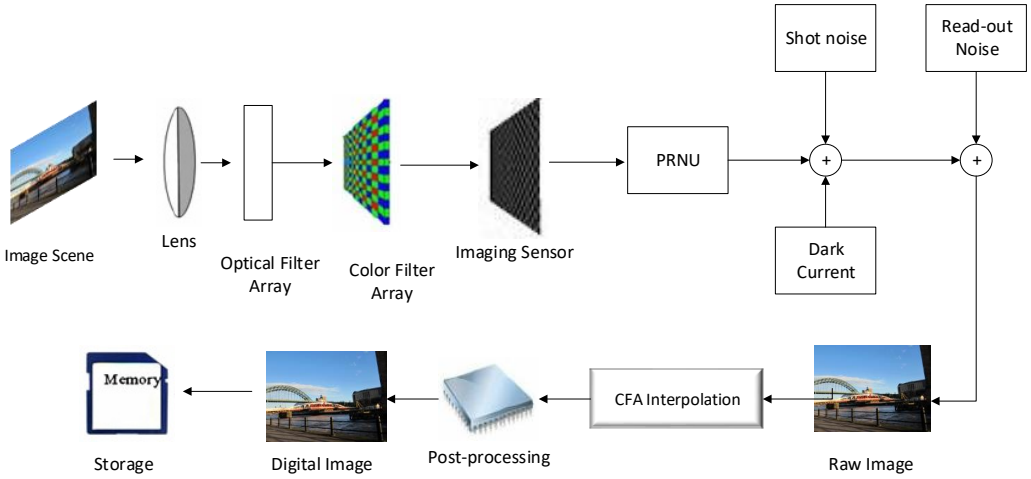


Figure 2-1 Digital Camera Acquisition Pipeline

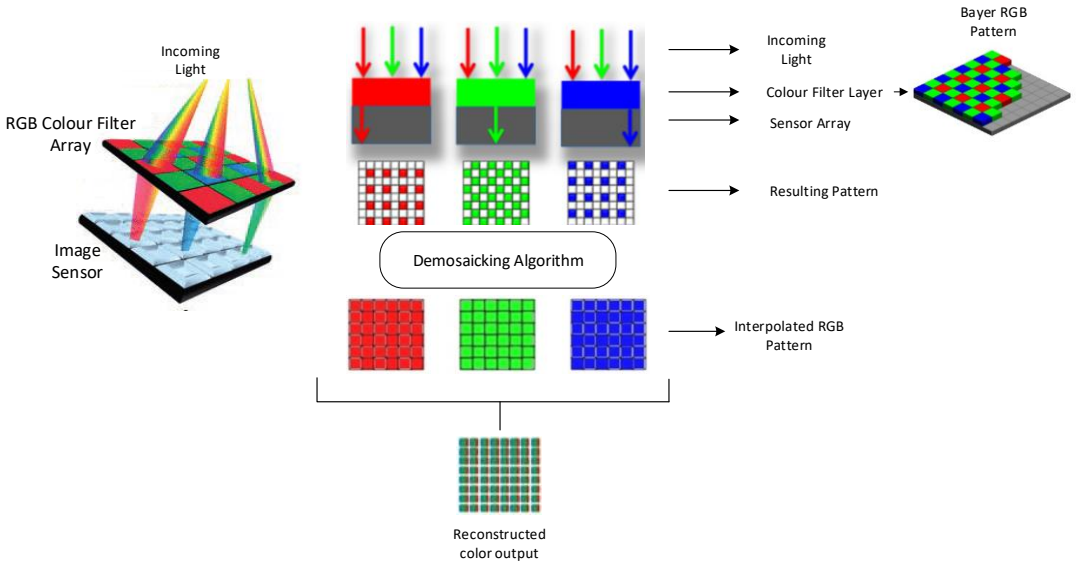


Figure 2-2 Process of Colour Interpolation

2.2 SOURCE CAMERA IDENTIFICATION (SCI)

Due to increasing proliferation of image acquisition devices, social media and multimedia sharing platforms, digital images are becoming a pervasive part of daily life. There are many sources available to capture digital images, such as digital cameras, mobile telephones, and tablets, which in turn can be used for illegal purposes. As manipulation, image forgery, and the downloading and sharing of digital material has become more commonplace over the years, an urgent necessity has arisen to confirm the integrity and authenticity of digital content and provide verification. Detecting the sources of images such as the brand and model of camera used to obtain a photograph is still a main focus, so that the owner of illegitimate and contentious material such as child pornography image and images of terrorist scenes can be determined. It is important for the forensic investigator to identify the source device of images gathered from a suspect source. Source camera identification is one of the major interests in digital image forensics. It can solve a wide series of forensics problems, from copyright infringement to ownership attribution. It is the process by which an unknown image or suspected images is linked to its source device. For this purpose, several algorithms have been proposed in recent decades to scrutinize the intrinsic traces on images. The two main approaches that are effective for source camera identification are feature-based and fingerprint PRNU-based methods (Lukáš, Fridrich and Goljan, 2006); (Li, 2010); (Lawgaly and Khelifi, 2016).

2.2.1 Feature-based Techniques for SCI

Kharrazi, Sencar and Memon (2004) proposed a feature-based approach for source camera identification using a set of numerical features of an image which are computed in both spatial and wavelet domains with the help of image quality metrics (IQM) and high-order wavelet statistics (HOWS) using a supervised learning technique to assign each picture to a digital camera. San Choi, Lam and Wong (2006) proposed a lens distortion effect method and demonstrated that it is possible to measure the intrinsic lens radial distortion of each sensor to enhance accuracy in SCI. They relied on the fact that inherent radial distortions act as a unique fingerprint in pictures. Other researchers have proposed the use of different sets of the attributes of images to identify their legitimate source; for instance, feature sets based on binary similarity metrics (BSM) (Çeliktutan, Sankur and Avcibas (2008), image modelling (Swaminathan, Wu and Liu, 2008), noise characteristics (Filler, Fridrich and Goljan,

2008), and de-mosaicizing strategies (Bayram *et al.*, 2005). Bayram *et al.* (2005) and Milani, Tagliasacchi and Tubaro (2014) exploited the effect of CFA interpolation to identify the source camera. Meanwhile, Chen and Hsu (2007) used gain histogram characteristics to solve camera identification problems, whereas Dirik, Sencar and Memon (2008) analyzed the traces left by dust on sensors and natural image modelling. In Deng, Gijsenij and Zhang (2011) study an auto-white balance algorithm was proposed. Gloe (2012) used an extended colour feature set, Xu *et al.* (2016) employed various texture features such as local binary pattern (LBP) and local phase quantization (LPQ) features, and an ensemble of de-mosaicing features was proposed by Chen and Stamm (2015). In (Borole and Kolhe, 2020) research study, statistical features based on Gray Level Co-occurrence Matrix (GLCM) are used for source camera identification with an average accuracy about 89.8%.

2.2.2 Filtered-based Techniques for SCI

More recently researchers have focused on intrinsic features; i.e. features that represent the imaging device sensor rather than the output data. Indeed, regardless of the imaging sensor type (CCD or CMOS) light-induced photo-charge is accumulated for each pixel. The signal output from the sensor is normally linear with the accumulated charge (Adams and Pillman, 2013). This charge depends on the physical dimensions of the pixel photosensitive area and on the homogeneity of the silicon. Because such homogeneity is never perfect and the pixel's physical dimensions slightly vary due to manufacturing imperfections, a white Gaussian noise, called the PRNU, always characterizes the output of the camera (Chen *et al.*, 2008). Furthermore, even in the absence of light, a small amount of current is accumulated in each pixel of the sensor. This is called dark current noise and is sometimes referred to as FPN. Finally, other noise sources such as readout noise, shot noise, interpolation noise, and quantization noise that can be added to the output do not constitute the dominant part and they are not useful for forensic purposes for the simple reason that they vary from one picture to another. The initial work in this context was reported by Kurosawa, Kuroki and Saitoh (1999) who used the FPN for video-camera identification. Because this technique requires dark frames to estimate the FPN, it fails to identify the imaging device in uncontrolled environments under which forensic investigations take place.

The most robust and accurate method using PRNU is exploited by Lukáš, Fridrich and Goljan (2006). PRNU is a unique fingerprint of a sensor that occurs due to imperfections in the manufacturing process.

The most important property of PRNU is that it can be used for determining the individual brand of digital camera for example to distinguish between images captured by Canon and Samsung digital camera. Also, the PRNU could be used to identify the source camera even if two camera devices are of the same brand and model (i.e., device linking). The abundant information that PRNU carries in terms of the frequency content makes it unique, and consequently suitable for identifying the source camera and detecting image forgeries.

In the past decade, a large number of studies have established the PRNU approach as a well-established forensic tool for matching an image to a source camera (Lukáš, Fridrich and Goljan, 2006) (Li, 2010) (Lawgaly and Khelifi, 2016). To identify the source device, it is now widely accepted in digital image forensics that PRNU is an inherent, dominant and unique fingerprint of imaging devices which provides an effective means of source camera identification.

A simplified camera sensor output model is used to derive the PRNU estimator, as illustrated in (Chen *et al.*, 2008). Regardless of the sensor type, the sensor output model as shown in Equation 2.1 can be expressed as follows:

$$I = g^{\gamma}[(1 + K)Y + \Theta_c + \Theta_s + \Theta_r]^{\gamma} + \Theta_q \qquad (2.1)$$

where I is the observed output image and Y is the original input signal of scene light intensity, g is the colour channel gain (used to adjust the pixel intensity level to achieve the correct white balance), γ is the gamma correction factor (whose value is determined as 0.45), Θ_c represents dark current noise (also referred to as fixed pattern noise) which can be removed by subtracting a dark frame from images captured by the same camera (Lukáš, Fridrich and Goljan, 2006), and Θ_s and Θ_r are shot and read-out noise respectively which are considered as random noise where Θ_q stands for quantization noise and K is PRNU multiplicative noise.

In Equation 2.1, the most dominant term is Y. In order to factor out the most dominant term Y from Equation 2.1, the first two expressions of Taylor expansion is applied. The final output of the image is obtained as follows in Equation 2.2:

$$I = I^{(0)} + \gamma I^{(0)} K + \Psi \qquad (2.2)$$

where $I^{(0)} = (gY)^{\gamma}$ is the sensor output in the absence of noise, K is PRNU sensor pattern noise, and Ψ represents the independent components of the random noise ensemble of Θ_c,Θ_s , Θ_r and Θ_q.

Most research has been devoted to characterizing each stage of the image acquisition process in order to identify the source camera device for an image under examination. The key stages for source camera identification include an image denoising process, the estimation, enhancement, and identification of PRNU, and detection performance measures as explained below.

2.2.2.1 Image Denoising

Image denoising is an integral part of image processing to reduce the amount of noise in images. Many denoising algorithms have been proposed in both the spatial and frequency domains. In the former, the techniques used are the hybrid median, Wiener, and bilateral filters, and histogram equalization; whereas the most popular technique used in the frequency domain is the wavelet domain. The favoured choice among most of researchers is to lower the noise using the wavelet domain technique, due to its locality, compression, and multi-resolution properties. To de-noise the image in the wavelet domain, many state-of-the-art methods have been introduced to estimate noise-free coefficients. In studies of source camera identification, Lukáš, Fridrich and Goljan (2006) proposed a wavelet-based denoising Wiener filter to extract noise patterns. They were the first to propose the extraction of SPN by filtering a series of pictures taken with the same camera, and Wavelet-based denoising filters perform very well in the extraction of PRNU. This filter is divided in two phases in which the variance in local image is computed and then the de-noised images are estimated in the wavelet domain by applying a local Wiener filter. The operation of wavelet de-noising filter proceeds as follows. Firstly, the noisy image is decomposed into fourth-level wavelet decomposition and it is observed into 8-tap Daubechies mirror filters. Different sub-bands such as those for the diagonal, horizontal and vertical are denoted as $d(i,j)$, $h(i,j)$ and $v(i,j)$, where (i,j) belongs to M which is the set of wavelet coefficients that depends on the level of decomposition. The high-frequency wavelet coefficients of a noisy image are modeled as an additive mixture of i.i.d (independent and identical distributed) signal with zero mean (a noise-free image) and white Gaussian noise $N(0,{\sigma_0}^2)$ (the noise component). The denoised wavelet coefficients can be obtained by using the Wiener filter, as shown in Equation 2.3 below:

$$h_{den}(i,j)=h(i,j)\frac{\sigma^2(i,j)}{\sigma^2(i,j)+\sigma_0{}^2} \quad (2.3)$$

where $\sigma^2(i,j)$ refers to the estimated local variance of the wavelet coefficients of the original image that can be estimated by using a MAP (maximum *a posteriori* probability) method for each sub-band, as shown in Equation 2.4:

$$\sigma_q^2(i,j) = \max(0,\frac{1}{q^2}\sum_{(i,j)\epsilon M} h^2(i.j) - \sigma^2) \quad (2.4)$$

where $q \times q$ is the size of the window that can be set with values of 3, 5, 7, and 9.

In addition, Dabov *et al.* (2007) proposed the BM3D algorithm as a denoising filter and found that better performance in color image denoising could be achieved. Wu, Kang and Liu (2012) employed edge-directed interpolation using their proposed context adaptive interpolation (PCAI), in which edges and smooth regions are taken into account. They estimated the smooth region by using a mean filter that accurately predicts the centre pixel value while the other edged regions, such as horizontally-edged, vertically-edged and other regions, are estimated through the median filter. To obtain more accurate results, they performed pixel-wise adaptive Wiener filtering and used maximum *a posteriori* probability (MAP) estimation to predict local variance. The results showed that their proposed PCAI method outperforms other denoising filters such as the Mihcack wavelet-based filter and the BM3D filter. Li, Guan and Li (2014) then introduced the principal component analysis (PCA)-based denoising of SPN for source camera identification, in which the content of the scene in an image is suppressed by reducing the dimensionality of PRNU noise and by transforming PRNU into an eigenspace spanned by the principal components. Kang *et al.* (2014) attenuated most of the contents of image scenes by using a filter based on eight-neighbour context-adaptive interpolation (CAI). Al-Ani *et al.* (2015) proposed a new 2-pixel image de-noising method for the estimation of white PRNU noise in which one adjacent pixel is used in spatial domain filtering to lower the pixel-to-pixel correlation. The results indicated that their proposed technique outperforms the basic sensor pattern noise (SPN), maximum likelihood estimation (MLE), phase SPN, and CAI methods. It is also worth highlighting the comprehensive study

undertaken by (Al-Ani and Khelifi, 2016) which includes an evaluation of SCI at different stages such as pre-processing, the filtering stage, and PRNU enhancement techniques. Lin and Li (2016) proposed that removing denoising distortions in the filtering stage can enhance the camera reference pattern noise. Moreover, Lawgaly and Khelifi (2016) proposed the locally adaptive discrete cosine transform (LADCT) filter to denoise the wavelet coefficients of the noisy image, in which the filter is operated on sliding blocks to derive sufficient information about the local effect of noise. The results showed that LADCT achieved better performance as compared to existing state-of-the-art systems such as BM3D filter, a predictive filter based on eight-neighbour CAI, and wavelet-based Mihcack. As PRNU is abundant in its frequency content and contains most of the high frequency details such as edges and textures, Gupta and Tiwari (2018) extracted it separately in the low- and high-frequency components of the image using discrete Fourier transform (DCT) for decomposition to improve performance. To increase the quality of the results, a weighting function was proposed which assigns weights for pixel-based correlation. Higher weights were given to image regions which provide reliable PRNU, whereas smaller weights were given to the regions providing lower reliability. In (Al-Ani and Khelifi, 2016) evaluation study, it has been shown that Lukáš, Fridrich and Goljan (2006) de-noising method of PRNU extraction is the most computationally efficient method among other de-noising methods.

2.2.2.2 Estimation of PRNU

In studies of the sensor pattern noise (SPN)-based identification of source camera devices, it has now become widely accepted in the image forensics community that PRNU is the most efficient pattern that characterizes a camera sensor and it is unique to each digital camera. Lukas *et al.*'s (2006) study laid down the fundamental pillars in the identification of source camera devices by indicating that PRNU is the most efficient and unique pattern that characteristics a camera. Lukáš, Fridrich and Goljan (2006) pointed out that PRNU is abundant in the high-frequency components of an image and can be extracted from the raw output of the sensor using a flat fielding process. However, the raw sensor output cannot be accessed by most consumer cameras. Therefore, the only way to extract the PRNU is to average the number of images taken by the same camera, as shown in below Figure 2-3.

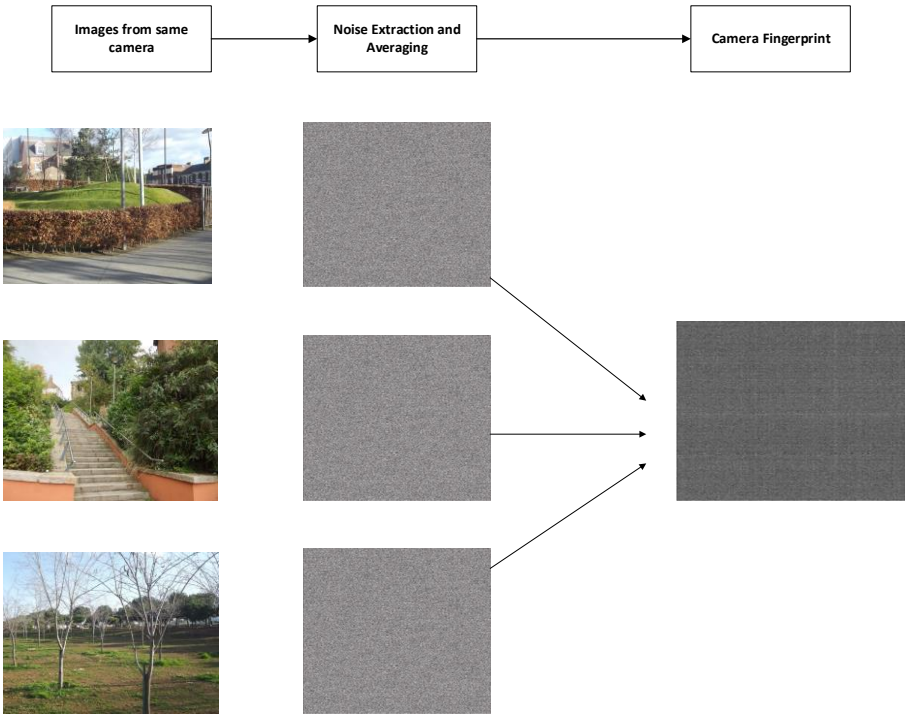


Figure 2-3 Extraction of PRNU from Noise Residual of Images

The image contents are suppressed through the denoising filtering process. So, the noise residual W is obtained through the denoising of the original image I, where F is the de-noising filter as shown in Equation 2.5:

$$W = I - F(I) \tag{2.5}$$

The PRNU of the sensor camera can then be estimated by averaging the noise residuals W_i with $i = 1 \ldots . L$, as shown in Equation 2.6:

$$K = \frac{\sum_{i=0}^{L} W_i}{L} \tag{2.6}$$

where L is the number of images used in the extraction of PRNU noise.

The PRNU estimation process and the decision concerning the test image is shown in Figure 2-4 below.

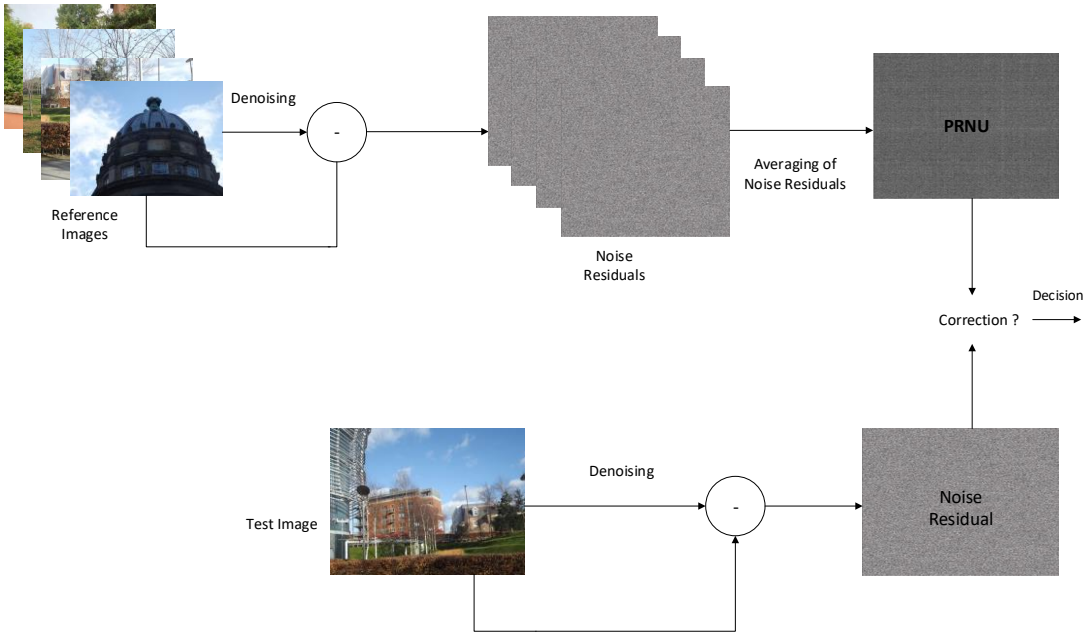


Figure 2-4 Estimation process of the PRNU for SCI

It is known that PRNU is multiplicative noise, and therefore Chen *et al.* (2008) proposed maximum likelihood estimation (MLE) to estimate the PRNU multiplicative factor from reference images to determine source camera identification, as shown in Equation 2.7:

$$K = \frac{\sum_{i=0}^{L} W_i I_i}{\sum_{i=0}^{L} I_i I_i} \tag{2.7}$$

Lawgaly, Khelifi and Bouridane (2014) questioned the suitability of the averaging method, emphasizing that it is only optimal if there is constant variance in all observations. However, as the noise present depends on the image content, its variance differs in each observation. Therefore, they proposed a weighted averaging method for sensor pattern noise estimation, as shown in Equation 2.8:

$$K = \sum_{i=1}^{L} W_i r_i \tag{2.8}$$

where r_i is the optimal weight corresponding to the i^{th} noise residual W_i and it depends on the variance of the undesirable noise which is present in each observation.

Mehrish, Subramanyam and Emmanuel (2018) derived an algorithm for the estimation of PRNU from probabilistic estimated raw data by using a joint Poisson probability process, and used MLE for the estimation of PRNU specifically for images with small patch dimensions.

2.2.2.3 Enhancement of PRNU

Many researchers have been devoted to improving the estimation of PRNU for source camera identification in order to remove unwanted artifacts such as CFA interpolation, which introduces periodic bias patterns in the column and row averages of the estimated PRNU and the row-wise and column-wise operation of sensors and processing units. Chen *et al.* (2008) proposed zero-mean (ZM) operation in every row and column of the estimated PRNU to suppress unwanted periodic patterns and biases. Goljan (2008) introduced a peak-to-correlation energy (PCE) measure to suppress periodic noise in the enhancement of camera source identification. Li (2010) proposed five models to attenuate the interference effect of scene details on PRNU and enhanced SPN by assigning higher weights to reliable components. Danielyan, Katkovnik and Egiazarian (2011) and Katkovnik, Danielyan and Egiazarian (2011) expanded the concept and generalized BM3D as a restoration tool, including de-blurring, after BM3D had initially been proposed as a denoising filter by Dabov *et al.* (2007). Moreover, Li and Li (2011) reduced the interpolation noise caused by the colour filter array (CFA) to obtain better estimation of the actual components of PRNU. A study by Kang *et al.* (2011b) indicated that actual PRNU is white Gaussian noise and the phase-only component of estimated PRNU would better represent the camera sensor noise. Therefore, they proposed a reference-phase PRNU to remove non-white contamination and to store the phase-only component of PRNU. Moreover, Wu, Kang and Liu (2012) adopted four-neighbour CAI to improve the estimation of PRNU and to determine the noise-free image and suppress most of the scene content prior to the extraction of SPN, and proposed an edge-adaptive SPN predictor. Lawgaly, Khelifi and Bouridane (2013) proposed an image sharpening approach to amplify the high frequency content of PRNU noise in images before PRNU estimation. Li, Guan and Li (2014) introduced an unsupervised learning method using principal component analysis (PCA) in the de-noising of SPN for source camera identification. Later, Li, Li and Guan (2018) proposed a supervised

learning method using linear discriminant analysis (LDA) to further boost performance in source camera identification by enhancing the extracted PRNU feature. Lin and Li (2016) lowered the periodic artifacts and implemented pre-processing on enhanced SPN through spectrum equalization. They subsequently proposed the removal of filtering distortion to enhance SPN and to improve filtering error (Lin *et al.* 2016). Meanwhile Lawgaly and Khelifi (2016) enhanced PRNU by concatenating the PRNUs estimated from three different colour channels to exploit the strong presence of physical PRNU components. In (Tiwari and Gupta, 2018) research study, the accuracy of SCI has been improved by using weighting function for correlation. Their idea is motivated by the fact that intensity based features and high-frequency details of the image can affect the quality of the extracted PRNU, therefore the weighting function can enhance the quality of PRNU by providing higher weights to reliable PRNU image regions and less weights to the image regions that do not give reliable PRNU.

2.2.2.4 Detection of PRNU Reference Pattern

The Normalized Correlation statistic test was used by Lukáš, Fridrich and Goljan (2006), Chen *et al.* (2008) and Li (2010) to determine the correlation ρ between the noise residual W and the camera reference pattern such as PRNU K, in order to make a decision about whether or not a specific image I was taken by camera C. As shown in Equation 2.9, the normalized correlation coefficient between the noise residual W and the PRNU of camera C such as K_C is defined as:

$$\rho_C(I) = corr(W, K_C) = \frac{(W - \bar{W}).(K_C - \bar{K}_C)}{\|W - \bar{W}\|\|K_C - \bar{K}_C\|} \tag{2.9}$$

where the bar above symbols represents the mean value, "." is the usual dot product, and ||.|| is the Least square norm.

Another correlation detection statistic proposed by Goljan (2008) is called the peak-to-correlation energy (PCE) measure, which is a replacement for normalized correlation that detects similarity between the noise residual and the PRNU pattern of a camera while reducing the false acceptance rate. Periodic noise contamination was also suppressed by using PCE to enhance the accuracy of camera source identification. However, the false positive rate was doubled when PCE detection was used.

PCE ratio is defined as the squared correlation divided by sample variance of the circular cross-correlations. Equation 2.10 for the determination of PCE is defined as:

$$PCE(x,y) = \frac{r_{xy}^2\ (0,0)}{\frac{1}{u \times t - |A\ |} \sum_{m_1,m_2 \notin A} r_{xy}^2\ (m_1 m_2)} \tag{2.10}$$

where $r_{xy}(m_1, m_2)$ represents the circular cross-correlation between x and y expressed as in Equation 2.11:

$$r_{xy}(m_1, m_2) = \frac{\sum_{i=0}^{u-1} \sum_{j=0}^{t-1} x(i,j) y(i \oplus m_1, j \oplus m_2)}{u \times t} \tag{2.11}$$

Here x and y are the two-centered vectors in $\mathbb{R}^n$ where $\mathbb{R}^n$ represents one-dimensional vectors of real numbers and $\oplus$ is the modulo addition. In Equation 2.10, A is a small neighbour area typically of size $u \times t$ around the central point at $(0,0)$, and $|A\ |$ is the cardinality of the area.

Kang *et al.* (2011a) proposed a further detection statistic called correlation over the circular correlation norm (CCN). This is to reduce false positive rates and suppress periodic noise contamination. The authors pointed out that the squaring operation r_{xy}^2 may increase the false positive rate by converting negative correlations into a positive PCE values. Therefore, the authors suggested the use of CCN, which can lower false positive rates approximately by half compared to PCE. The correlation over circular correlation norm (CCN) is derived using Equation 2.12:

$$CCN = \frac{r_{xy}\ (0.0)}{\frac{1}{u \times t - |A\ |} \sum_{m \notin A} r_{xy}^2\ (m_1 m_2)} \tag{2.12}$$

As seen above, machine or statistical learning techniques have been applied by forensic researchers in SCI. Examining digital evidence regarding a crime which has been committed in order to be utilised as legal proof in a court of law is part of a digital forensic investigation. Machine learning is seen as an effective technique which can solve the challenges that arise in the field of digital forensics. In the

process of retrieving and analysing digital evidence, a variety of machine learning methods and approaches are beneficial. So, machine learning improves the process by allowing it to cope with massive amounts of data in a short amount of time while maintaining a high degree of accuracy and producing high-quality results. Investigators are motivated to employ these approaches throughout forensic analysis as they provide the ability to combat a variety of crimes (Qadir and Varol, 2020).

2.3 Deep Learning Approach

Apart from machine learning approach, recently a models of deep learning (DL), which is also known as feature learning or representation learning, have achieved a high recognition rates in source camera identification. In deep learning approaches, no pre-processing is required to extract the features of the image. These methods are based on automatic learning and the ability to gain knowledge from the inherent characteristics of training data. Features are learned directly from photographs in order to generalize the camera model. This is a type of machine learning in which a model is used to perform the classification task by automatic learning directly from images, texts or videos. Different patterns are tuned to capture patterns and to extract features automatically, which are usually created using a neural network architecture. Traditionally, a neural network has only two or three layers but deep neural networks may have hundreds of layers: the more layers, the deeper the network. The difference between traditional neural network and deep learning neural network is shown in Figure 2-5.

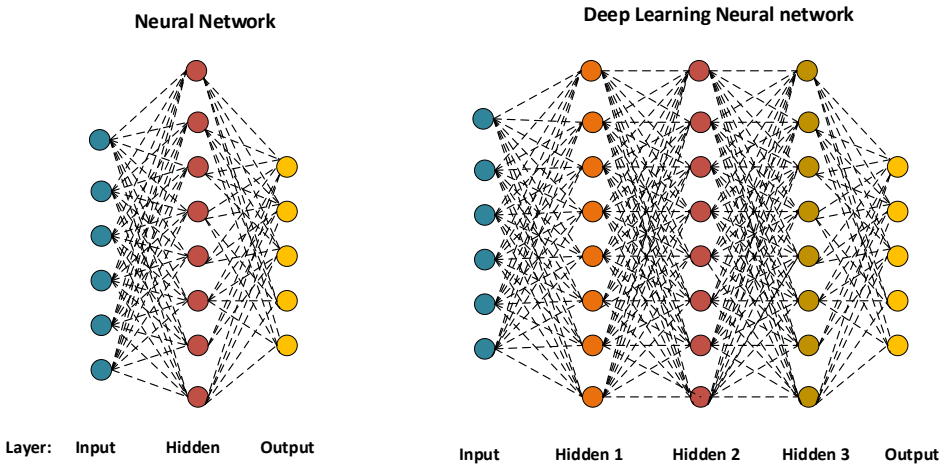


Figure 2-5 Difference between traditional neural network and deep learning neural network

The degree of accuracy of deep learning heavily relies on technology enablers such as datasets, which are useful for training on many different types of objects. Meanwhile high-performance GPUs reduce the time required for training from days to hours. In deep learning, there are multiple non-linear processing layers which operate in parallel. Inspired by biological nervous systems, the neural network consists of an input layer, a number of hidden layers, and an output layer. All layers are interconnected with others through neurons, and each layer uses the output of the previous layer as its input. By using training data, deep learning can begin to understand the residual noise patterns or any features of raw images, which can eventually be linked with corresponding classes. Each layer of the network operates by obtaining data from the previous layer, transforming it, and then forwarding it to the next layer. Since the network learns directly from the data, no information is needed concerning what kind of features are being learned.

2.3.1 Convolutional Neural Network (CNN) Architecture and Operation

The CNN is one of the most popular deep learning models. It has shown remarkable performance in many computer vision applications, including face recognition, handwriting detection, image classification and pedestrian detection. Inspired by the human neural system, the CNN is constructed as a complex computational model with many interconnected nodes. A numerical weighting parameter is used to define connections between neurons which are tuned so that the neural network has the ability to learn complex functions. The CNN is the most popular algorithm in use for deep learning with images and videos. It is also constructed in a similar way as other neural networks in consisting of an input layer, several hidden layers and then a final output layer, with all layers interconnected via neurons and using the output of the previous layer as input. As multiple non-linear processing layers operate in parallel, the network is therefore is defined as a sequence of layers where each changes a volume of inputs to a new volume of output through differentiable functions. Feature learning, detection and classification elements are unified in one model and are learned jointly as shown in below Figure 2-6. Multiple layers of different processing units perform operations such as convolution, pooling, sigmoid/hyperbolic tangent squashing, rectification and normalization, and are responsible for learning (representing) a hierarchy of features from low-level to high-level.

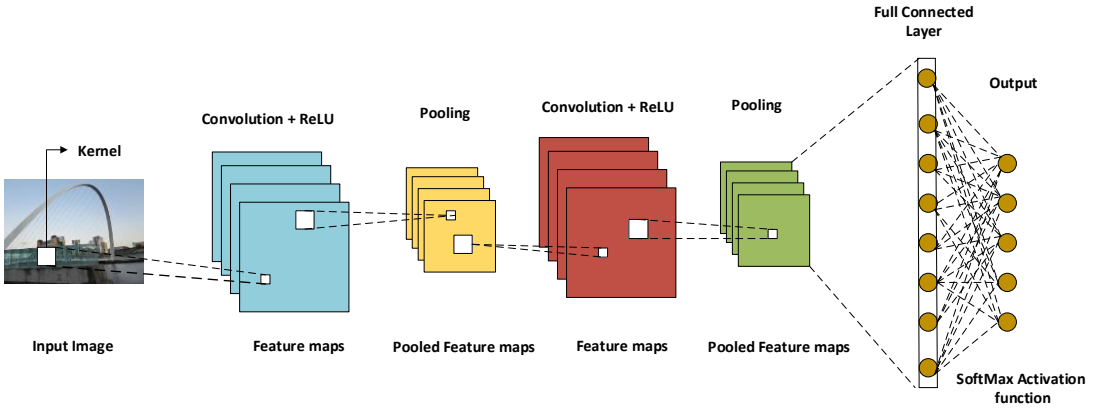


Figure 2- 6 A basic convolutional neural network structure

To do this, the first layer of the CNN is the input layer, in which the data is in the form of blocks or patches of images fed into the network. Low level features are usually extracted here, such as edges, shapes and color contrast, whereas deeper layers identify more complex visual patterns After the first input layer, there are many hidden layers consisting of different processing units responsible for performing various operations in a feature detection task, such as convolution, normalization, pooling and the rectified linear unit (ReLU). Convolution is the major operation, which convolves around the image using a number of kernels which act as filters. Each filter is specified with a small width and height and the same depth as given to the input volume, as shown in Figure 2-7.

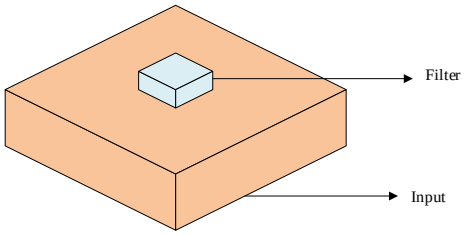
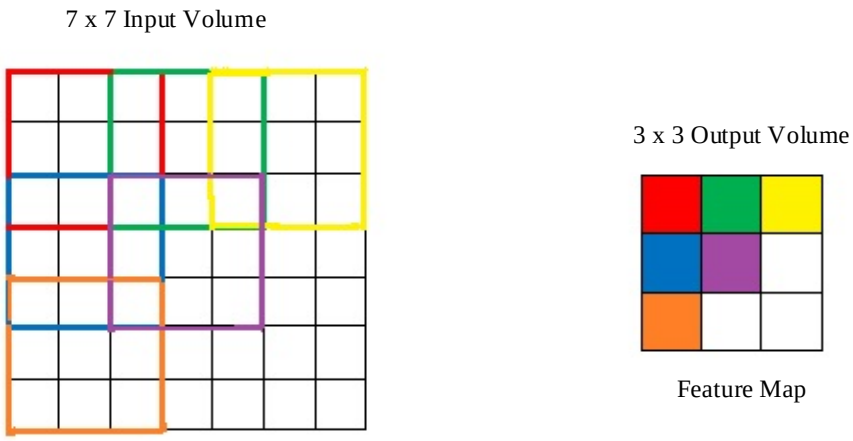


Figure 2-7 Filtering operation with small width and height and the same depth as given to input volume

The convolution layers work on the input images through a set of defined convolutional filters. Every output of a convolution operation is a feature-map. Each convolutional filter activates certain features from the images so as to produce several feature maps. The size padding and stride of each kernel used for convolution need to be specified, as they affect the quality of the output of the layer. Using a smaller

size of kernel, finer features are extracted. For instance, as shown in Figure 2-8a, the 3×3 filter, also referred to as a neuron or a kernel, convolves around the input volume 7×7 by shifting two units at a time. The area that is covered by the filter size during convolution called the receptive field. The amount that determines how much the filter should shift is called the stride. In this example, the stride is set at '2' and padding is set as '0'. As a result, an output volume of 3×3 feature map is obtained as shown in Figure 2-8b using Equation 2.13. Each position of the filter results in an activation of the neuron, and the filter convolves around every location in the input volume, which in turn gives an activation map or a feature map.



(a) (b)

Figure 2-8 The convolution process and feature map

The formula for computing the output size for any given convolutional layer is calculated as shown in Equation 2.13:

$$O = \frac{(H-T+2P)}{S} + 1 \tag{2.13}$$

where O represents the output height/length, H is the input height/length, T represents the size of the filter, P stands for padding and S stands for stride.

In the below Figure 2.9, the input image is 5×5, size of the filter is 3×3, padding is set as '1' and stride is set as '2'. The filters moves to the right with a stride value set as '2' till it parses the complete width. Moving on, it hops down to the left of the image with the same stride value. This process is repeated until the whole image is traversed. Thus, the output volume of the feature map is resulted as 3×3 using Equation 2.13.

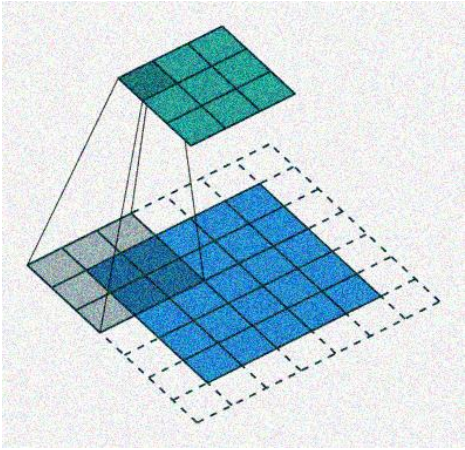


Figure 2-9 Another example showing feature map when padding is set as '1'

After computing the required number of feature maps, the CNN architecture performs an activation function layer for feature extraction and then further enhances the output through a pooling layer. An activation function in CNN is used to provide non-linearity in the data. The most commonly used activation functions are sigmoid and tanh, but the most widely used activation function in deep learning is the rectified linear unit (ReLU), $f(x) = log(1 + ex)$. The benefit of using ReLU over other activation functions such as the sigmoid is that the gradient of the latter reaches nearly zero when there is increase or decrease in the input signal, while in ReLU the gradient does not vary when the input signal changes. The ReLU layer performs faster training by retaining positive values and mapping negative values to zero. After ReLU, a pooling operation is performed which may fall into one of three categories, which are max, average and L2 norm. In deep learning, max pooling is usually employed, and this pooling operation further simplifies the number of parameters for network learning by computing non-linear down- sampling. These three operations are recurrent in different numbers of hidden layers in order to detect different features.

The final output layer is based on the classification layer from which the fully connected layer produces a vector of Q dimensions, where Q is the number of classes that the deep-learning CNN network can predict. The vector contains the probabilities for each category of any picture being classified. The other most essential component involved in the CNN architecture is that which evaluates the performance of the learning task. Therefore, a cost function as an 'optimize error' function is used in deep learning in camera model classification tasks in order to minimize classification errors by finding an optimal set of theta values, where theta is the parameter space of the architecture.

Four kinds of optimizers are commonly used in deep learning: SGD (Stochastic Gradient Descent), RMS (Root Mean Square) propagation, Adam (Adaptive Moment Estimation) and Adadelta (Adaptive Delta) (Ruder, 2016). Image batch size plays a crucial role in the representation learning. If the batch size is larger, it means that higher memory is required. The advantage of using batch size is that there is no need to load the memory all at once, and this helps in speeding up learning by updating the weight of the network at every iteration. Therefore, tuning different parameters such as the block size of the image, the number of convolution layers, batch size, the number of epochs, and the kind of optimizer used helps in the achievement of better accuracy of the deep neural network.

To prevent overfitting, a drop-out layer is also used in the CNN as a regularization technique. It helps in speeding up training, reducing interactions between neurons and learning more robust features. The final classification layer of the output layer uses a softmax function to provide the classification output. This last layer performs a loss function. The Equation 2.14 for the softmax activation function is as follows:

$$Softmax\ (\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{Q} e^{z_j}} \tag{2.14}$$

where $\vec{z}$ is the input vector to the softmax function and all the z_i values are the elements of the input vector to the softmax function and it can take any real value, positive, zero or negative. Q is the number of classes. The exponential act as the non-linear function. The term $\sum_{j=1}^{Q} e^{z_j}$ is referred as the

normalization term which ensure that all the output values of the function will sum to 1 and each be in the range (0,1). The normalization values are then converted into probabilities.

Since there is no known formula to determine the number of layers required, the best approach is to tune a few layers and then assess how the network performs.

2.3.2 Convolutional Neural Network (CNN) for Source Camera Identification

This section briefly reviews recent deep learning contributions to the task of SCI. A few papers have suggested the use of deep learning for source camera identification. For instance, Baroffio *et al.* (2016) proposed a CNN model comprised of a structure with three convolutional layers for camera identification in which 27 camera models were used. The dataset used consisted of 415k patches extracted from 16982 natural images, of which 300k patches were used for training while 15k were used for testing. Three days were devoted to processing the training images for their experimental results in order to achieve an error rate of approximately 6%. In addition, Tuama, Comby and Chaumont (2016) proposed an architecture for a CNN model also with three convolutional layers, and a single pooling layer was applied after the convolutional layers to achieve an error rate of less than 10%. Bondi *et al.* (2016) presented a structure with four convolutional layers followed by a transfer learning strategy using SVM classification for camera model identification. For this, more than 13000 images with a 64 x 64 patch size taken by 18 camera models were used for their dataset to achieve 93% accuracy. To address the SCI problem, Freire-Obregón *et al.* (2019) conducted experiments using CNNs on mobile telephone images to determine the mobile device used. They tuned various hyper-parameters of the CNN to measure the model's accuracy. Their findings indicated that the deep learning approach provides good performance in the case of identifying the source device of an image, but performance declined when the numbers of cameras produced by the same manufacturer increased.

2.4 TEMPORAL IMAGE FORENSIC ANALYSIS

Digital image acquisition devices have become widely used in people's daily lives. With the advent of high-resolution devices and user-friendly image editing tools, an enormous variety of digital images are in circulation. Unfortunately, these technological advances have also become a new weapon of choice

for use in criminal acts. Due to growing volumes of cyber and electronics-assisted crime, it is now crucial for forensic scientists and researchers to be able to link images to their source devices, and the originality of such images may constitute critical evidence in a court of law. Detecting the type of the source of images, such as brand or model of camera used to acquire a photograph, is still the focus of attempts to identify the owners of illegitimate and contentious material. The identification of the camera device used to capture a specific image has attracted great interest and can solve a wide range of forensics problems from copyright infringement to ownership attribution. As a unique reference pattern for every individual digital camera, PRNU is created by small differences in the sensitivity of the camera's light sensors. Camera sensors such as CCD and CMOS components in digital cameras convert photons (light particles) into electrons (electrical charge). The pixel value is measured according to the electrical charge. During the sensor manufacturing process, small variations are caused in the electrical signals which result in the form of a unique noise pattern. The PRNU of the digital imaging device is useful in the digital image forensic analysis. The other noise sources such as readout noise, shot noise, interpolation noise, and quantization noise that can be added to the output do not constitute the dominant part and they are not useful for forensic purposes for the simple reason that they vary from one image to another. With regards to temporal information, not all artifacts of imaging sensor device exhibit changes over time and therefore, they are not useful for estimating the acquisition time of digital images. The artifacts such as defective pixels and sensor dust are the only reasonable candidates for the temporal image forensic analysis as they gathered over time. However, sensor dust can be cleaned from the sensor protective glass and thus, the temporal information cannot be useful for the forensic investigator. Therefore, defective pixels constitute as a sensor imperfection seem to be a more reasonable candidate as they accumulate over time which can be exploited in temporal forensic analysis of images.

2.4.1 Defective Pixels

Digital image technology is widely used in our daily routines such as in devices like tablets, smart digital phones, notebooks, and digital cameras. Nowadays, a small form factor has become a popular trend and therefore there is a tendency to reduce the size of image sensors in contemporary devices with small pixel size and with higher resolution. This increasing trend of using small pixel sizes and the higher sensitivity in the digital sensors enhanced the possibility of defective pixels (Chapman *et al.*,

2010). Even though defective pixels may be low in quantity, they are very noticeable. As with other microelectronic devices, digital images develop defects over their lifetime. Geradts *et al.* (2001) emphasized that defective pixels also constitute another imperfection of sensor of relevance to forensics. Therefore, such pixels can be used to determine the age of pictures given that causes of the defective pixels may be associated with the ageing of the imaging sensor. Theuwissen (2007) indicated that defective pixels inside the camera sensor are caused due to imperfections in their silicon lattices, and this damage is actually mainly due to the impact of cosmic ray radiation. Such infield defects develop and continuously increase in number over the device's lifetime. Defects can be removed by factory calibration, but this process is extremely costly. These in-field defects do not heal over time and, once the pixels become defective, the damage is permanent and their parameters cannot be changed (Leung *et al.,* 2009).

2.4.2 Types of Defective Pixels

Defective pixels located in an image sensor often produce different output from that of neighbouring pixel (Redi, Taktak and Dugelay, 2011). Therefore, they degrade the quality of captured images. There is no standard exact definition of a defective pixel, but usually they are divided into three main groups: hot pixels, which are 'stuck-high'; dead pixels which are also classified as 'stuck-low'; and abnormal-sensitivity pixels classified as partially-stuck. Hot pixels are those which produce output values much higher compared to that of their neighbouring pixels. Stuck-low pixels are those whose values are much lower than adjacent pixels. They usually occur as black pixels or are called dead pixels, and are permanently damaged and do not receive any power. Partially-stuck pixels have values different to a certain degree from adjacent pixels by a small factor and exhibit irregular behaviour (Chapman *et al.*, 2010). The most prominent types of defective pixel which occurs in a sensor over its lifetime are stuck and hot pixels. The former are independent of illumination, and their numbers increase linearly with exposure time. A hot pixel is a pixel with large dark current which is caused by the higher leakage of charge (Fridrich et al., 2011). If the dark current of a pixel is high then it is referred to as hot, whereas if the offset is high it is denoted as a stuck pixel. The appearance of hot pixels usually resembles bright spots with fixed spatial locations in an output image. Partially-stuck hot pixels have an additional offset but this is independent of both illumination and exposure time. A stuck pixel has a constant fixed value,

c, between $0 < c < 255$ in the output of the image. These types of defective pixels mainly appear as factory time defects. Dudas *et al.* (2007) proposed five levels of defect model, of which the first level includes the most commonly found defects in camera sensors. These level-1 defects are stuck-high and stuck-low pixels. The former always output a nearly full-scale value, whereas the latter always return a value close to zero. In addition, Leung *et al.* (2007) claimed that they had never found true stuck pixels in their experimental results. Meanwhile, Chapman *et al.* (2013) emphasized that, in the past, researchers found that hot pixels have an illumination-independent component that increases linearly with exposure time. However, this is not true on most occasions but in fact depends on illumination. They also categorized hot pixels into two types: standard hot pixels that increase linearly with exposure time; and partially-stuck hot pixels that have an extra offset which can be identified at zero exposure. In addition, Fridrich and Goljan (2011) classified abnormally-behaving pixels as an example of PRNU and dark current. Chapman *et al.* (2017) indicated that in-field defects are always stuck hot pixels. Their analysis showed that, when the size of the pixel shrinks, a large amount of hot pixels are produced at higher International Organization for Standardization (ISO) settings. They projected the growth rate in numbers of hot pixel defects in terms of defect density (defects/year/mm^2), where the hot pixel densities grow according to a power law. The power law of hot pixels is described as the inverse of pixel size raised to the power of about 3 and the ISO parameter raised to the power of about 0.5. They developed special procedures to analyze collected dark frame data, and found the the growth rate of hot pixel defects becomes higher when pixel size is reduced to 1 micron. The quality of the images produced by digital sensors are mostly affected by dark current and hot pixels. The latter and stuck pixels are depicted in Figure 2-10.

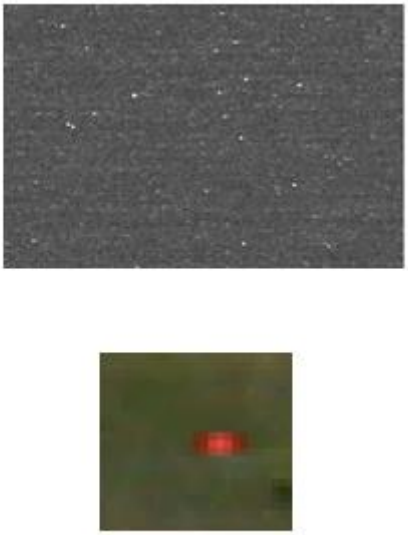


Figure 2-10 Hot pixels captured from a dark frame images (top); close up of a stuck pixel in an image (bottom) Source: (Fridrich and Goljan, 2011)

2.4.3 Detection of Defective Pixels

The detection of defective pixels is still challenging. The central problem lies in the image processing pipeline of the digital imager, because the presence of defective pixels degrades the perceived quality of images. Many theoretical and statistical approaches have been proposed for the detection of defective pixels and to assess the number of them a sensor produces per year. However, the detection of defective pixels can be performed at the factory during the manufacture of digital imagers by conducting dark frame calibration to identify hot pixels and light-field calibration to detect dead pixels. However, such calibration involves complex and expensive processes. (An, Lee and Kim, 2007) proposed a temporal verification method for dead pixels that is based on scene-change detection used to reduce false detection rates and enhance accuracy. Leung *et al.* (2008b) and Chapman *et al.* (2012) found that the rate of detection of defective hot pixels can be increased if the ISO/exposure time is lengthened, especially when there are low light conditions. According to the study by Leung *et al.* (2007), defective pixels are caused by random processes; therefore, they can appear as singlets, doublets and triplets or even clusters of hot, dead and stuck pixels. Leung *et al.* (2008a) analysed the statistical distribution and date of development of defects which develop rapidly over the sensor's lifetime. They presented a hot defect pixel tracing algorithm and concluded that these defects were caused due to cosmic radiation

rather than material failure. They identified hot pixels by utilizing an algorithm based on Bayesian statistics to detect the presence and absence of defects in sets of colour images.

According to Dudas *et al.* (2007) and Leung *et al.* (2008b), defective pixels accumulate over time and space on the sensor independently of each other. In addition, they found that locations of defective pixel follow a normal random distribution across the sensor area and inter-defect times follow an exponential distribution. Chan (2009) proposed a real-time dead pixel detection method highly suitable for embedded devices. However, the method yielded much higher false positive rates. Cho *et al.* (2011) proposed a real-time dead pixel method using reference data from a to-be-tested pixel. The results achieved very high detection rates and image quality compared to Chan's approach.

As the physical dimensions of pixels change slightly due to imperfections introduced during the manufacturing process, Fridrich and Goljan (2011) pointed out that a large value of PRNU can lead to a point defect, which is also referred to as a pixel with abnormal sensitivity. Longer exposure, larger ISO and higher temperature increase the dark current of a pixel, which can lead to the development of hot defective pixels. Looking the impact of all the defects as discussed above, Fridrich and Goljan (2011) derived the raw output model of a sensor as shown in Equation 2.15 to detect defective pixels.

$$I = I^o + I^o K + \tau D + c + \Psi \tag{2.15}$$

Where I is the observed output image, $I^{(0)}$ is the sensor output in the absence of noise and τ is a scalar multiplicative factor which depends on exposure time, temperature and ISO value, D represents a dark current factor, c is the matrix of offset values referring to stuck pixels, K represented as PRNU and Ψ is a collection of noise sources.

Meanwhile, Bergmüller *et al.* (2014) used a statistical approach for the detection of defective pixels and, moreover, Kauba and Uhl (2016) concluded that all images must be aligned pixel-wise according to the sensor in order for hot defective pixels to be detected effectively. El-Yamany (2017) detected defective pixels by exploiting Bayer sensors in the digital camera acquisition pipeline. The locations of defective pixels were identified and the sensor's defect map updated in order to produce high quality

images. Dong *et al.* (2019) explored the digital image pipeline to detect defective pixels by employing multiple edge detection, a detection channel and threshold segmentation. They built a digital weld defect database to classify various defects using an SVM classifier. Wang *et al.* (2019) proposed a blind pixel detection method based on visual characteristics to locate dead pixels based on pixel values lower and higher than others and to obtain information about a pixel's position to fix the dead pixels. Forcina and Carbone (2020) formulated a statistical model called the Gaussian mixture model to determine the effect of temperature and duration of exposure on dark current, hot pixels and thermal noise in the analysis of astronomical image data under conditions of darkness. Tchendjou and Simeu (2020) used pixel neighbourhood analysis by conducting only simple arithmetic operations on the neighbourhood to detect defective pixels. They estimated the distance between a pixel under examination and surrounding pixels. They also computed the mean, median, variance and other statistical dispersion parameters of neighbourhood pixels.

An interesting characteristic of defective pixels is that they accumulate over time and space on the sensor independently of each other. Thus, once they have occurred in an imaging sensor, defective pixels proliferate over time. They become a permanent part of the sensor, which does not heal itself. From this perspective, establishing temporal relationships between digital pictures taken by the same camera for dating purposes is theoretically possible by detecting the defective pixels in a temporal series of images which were taken over a larger timespan. It is worth mentioning that the studies reviewed above concerning the detection of defective pixels were conducted with the aim to enhance the quality of images, whereas the present study intends to develop an application which is able to use information from such pixels to estimate the acquisition date of images.

2.5 PICTURE DATING

Methods to determine the dates at which images were taken have been introduced in many computer vision applications. The first work that deals with dating historical images is coined by Schindler, Dellaert and Kang (2007). They presented an approach to order a collection of cityscape pictures temporally. For this, the authors reconstructed the 3D world from which many overlapping images of the same location were acquired. Palermo, Hays and Efros (2012) proposed an approach to

automatically predict the age of historical colour images in which they designed combinations of various colour descriptors to model the historical colour photographs. They then analyzed variations in colour distributions statistically in order to classify the photographs, while splitting the database into a chronological time-order as well as taking into consideration the capture of temporally discriminative information based on time. Fernando *et al.* (2014) further improved the work of Palermo, Hays and Efros (2012) by adding colour derivatives and angles. They noticed that extracting the features of colour distribution was not in itself efficient in discriminating between different decades. Therefore, they proposed colour features which are device-dependent (such as colour-films) for dating historical colour images in which two colour descriptors such as RGB and colour angle are obtained for image dating purposes. Some research works, such as that by Jae Lee, Efros and Hebert (2013) and Vittayakorn, Berg and Berg (2017), have looked at the visual style of objects so as to discover visual connection of images in space and time, while (Lee *et al.*, 2015) used architecture styles as a cue in dating pictures. Moreover, Martin, Doucet and Jurie (2014) applied a binary classification algorithm to classify image by making decisions in terms of an 'older' or 'newer' image when compared against a reference image. Jae Lee, Efros and Hebert (2013) evaluated and modelled stylistic differences across time and space by identifying the street and car object-type view images with different groups of patches. Some authors, such as He *et al.* (2016) and Li *et al.* (2015), have worked on estimation methods to age historical documents by exploring contour and stroke fragments, and the CNN along with optical character recognition have been applied. In addition, Ginosar *et al.* (2015) and Salem *et al.* (2016) estimated the dates of pictures by learning a dynamic map of visual attributes through the process of finding differences in the patterns of human appearance, fashion and clothing styles. Furthermore, Müller, Springstein and Ewerth (2017) predicted the time of acquisition of images which were captured between 1930 to 1999. In looking for timestamp manipulation, authors such as Kakar and Sudha (2012) and Li *et al.* (2017) have worked on verifying the timestamps of images by estimating the sub-azimuth angle with the help of shadow angles and the appearance of the sky, and then assessing it according to the position of sun computed from image metadata. Whereas Padilha *et al.* (2021b) verified the date of image capture by examining the consistency of their timestamps with its scene content, geographic location and satellite imagery. Based on the appearance of the scene, the images from the real-world events are temporally sorted by estimating the image's temporal position within the duration of the event occurred (Padilha *et al.*, 2021a).

The above research has provided very impressive results; however, the methods used require the presence of particular visual elements in a scene and geographic location as cues to reliably predict the date of image capture. Only a very few studies such as those by Fridrich and Goljan (2011) and Mao *et al.* (2009) have been devoted to the extraction of temporal localisation information from digital pictures. For image forensics, Fridrich and Goljan (2011) research used natural images with different indoor and outdoor scenes to estimation the acquisition time of digital images. They assumed that the accuracy of their proposed method would depend on the appearance of new defects on the sensor. Therefore, a higher frequency of the occurrence of new defects in an image sensor would lead to higher accuracy in determining the age of digital pictures. However, the results were not promising, and they acknowledged that the maximum likelihood technique used to estimate the acquisition time of pictures was unable to perform accurately between two consecutive defect onsets. This is because the likelihood is constant since the technique returns a constant acquisition time which is the average time of the two onsets for any given picture. Another research framework was presented by Mao *et al.* (2009) where time-dependent camera parameters were assumed to be included in the PRNU. The results obtained through the correlation coefficients of PRNU estimates for different time-based image clusters showed that it is possible to rank the clusters according to their acquisition date. The results clearly showed that theses correlation coefficients were not linearly dependent on the timespan between the acquisition date of the picture under investigation and the date corresponding to the closest image cluster. The lack of a dataset specific to temporal camera analysis can be clearly observed in the experimental part of the two proposed frameworks in this area. Indeed, Fridrich and Goljan (2011) used only three cameras and acknowledged that irregular time slots limited the performance evaluation of their estimation technique. Also, the pictures downloaded from the photo sharing website Flickr (Flickr, 2018) and used in the other framework by (Mao *et al.*, 2009) were acquired in non-regular time slots and represented three digital cameras only. Moreover, the authors pointed out that some pictures, for which the temporal ordering was inaccurate, were manipulated before being uploaded on Flickr.

Also, the dataset of images used in their experiments is consisted of irregular timeslots. Furthermore, the use of the correlation coefficient on the PRNU noise does not take the type of pixels into account and ignores their temporal evolution. That is, the correlation operation treats all the pixels equally

without any prior knowledge on their temporal behaviour. Therefore, neither of these research frameworks have not yet reached conclusive findings with regards to accuracy and applicability for dating purposes. In fact, the previously mentioned systems assume that the picture under investigation has been taken in a specific time-span, defined by known acquisition dates of a set of trusted pictures, and attempt to locate it in relation to the trusted pictures. Although such systems provide forensically relevant information, their applicability in practical investigations remains limited.

2.6 IMAGE AND VIDEO DATABASES FOR MULTIMEDIA FORENSICS

In the field of multimedia forensics, image device fingerprinting is used by analysts to establish the history of an image, identify the source device and verify the content. Therefore, source camera identification applications have attracted the attention of many researchers who use different kinds of datasets to facilitate their work. Over the years, plenty of such datasets have been constructed which are available for use to gain more understanding of multimedia forensics applications. Nowadays, different digital devices can capture and record video images. In addition, collections of covert videos and illegal seen contents on a large scale on the worldwide web. Even unskilled users can manipulate and tamper with the content of images with only minimal effort. Meanwhile not all digital content can be protected with the help of digital watermarks or signatures. In fact, passive techniques can provide protection and the authentication of digital content. As described by (Chen *et al.*, 2008), the field of digital forensics can be divided into six main tasks: source classification, device identification, the linking of groups of images to the same source device, the recovery of processing history, integrity verification, and anomaly investigation. Each task is associated with specific problems such as copyright protection, forgery detection or the verification of images or videos against a specific camera device. With the advances in novel techniques such as generative adversarial networks and computer generated imagery (CGI), it has become more challenging for forensic analysts to identify the integrity, trustworthiness and authenticity of visual content.

With the popularity of low-cost imaging devices such as camcorders, digital cameras, smartphones, IoT devices, tablets, and scanners, still images and videos have become popular means to transmit persuasive messages. Similarly, the worldwide web also facilitates the sharing and uploading of digital

content by multimedia users through social media platforms. Huge volumes of digital pictures and videos are captured, shared and distributed on websites and social networks every second. Such images and videos available on different platforms, including cloud storage, smartphones and social media, are also used as a primary source of information for many people and institutions. Huge amounts of digital media content are spread widely in different ways. If such pictures or videos are related to a crime, then multimedia forensics staff aim to investigate the history, authentication and timings and any correlations of this digital content. Many forensic techniques have been developed to identify the history of digital images, perform forgery detection and identify the source device and camera model used. To address these issues, the availability of many high-quality forensic databases collected from different devices such as digital cameras, smartphones, camcorders and social networks have helped in the development of forensic algorithms. However, several databases have been built and presented for the forensic research community to benchmark forensic techniques.

In the following overview, several popular datasets used for image and video source identification are analyzed and described in detail such as UCID (Uncompressed Colour Image Dataset) which is the first and oldest image dataset available to the multimedia forensic research community was originally constructed for the evaluation of image retrieval techniques. The images are very small in size at either 512 x 384 or 384 x 512 pixels, and they may have suffered some degree of out-camera post-processing (Schaefer and Stich, 2003). The Dresden Image database was the first large and publicly available image database built for the research community, and was specifically designed for forensic applications such as source camera identification and later for forgery detection. This dataset contains more than 14000 images, including many indoor and outdoor scenes acquired from 73 camera devices selected from among 25 camera models with the highest quality in JPEG settings. This dataset has proven very popular since 2010 in the forensic research community dealing with benchmarking for SCI problems, and the database is intended for the development and benchmarking of camera model identification solutions (Gloe and Böhme, 2010). The RAw ImageS datasEt (RAISE) portray a large variety of semantic contents and technical parameters and are classified into seven different types of scenes: ‘outdoor’, ‘indoor’, ‘landscape’, ‘nature’, ‘people’, ’objects’ and ‘buildings’. The RAISE dataset is specifically used for camera verification tasks only (Dang-Nguyen *et al.*, 2015). Shullani *et al.* (2017) proposed the Video and Image Dataset for Source Identification (VISION), which is a dataset

containing both images and videos, as a collection of media content captured by portable devices mainly for the purpose of video and image source identification. The images and videos are split into two categories of native and social exchange content from Facebook, YouTube and WhatsApp. To test the relevance of the VISION dataset for multimedia forensic, the authors tested the dataset using image and video forensic tools such as sensor noise patterns for the source identification of both the native and social media content. In addition, they highlighted the importance of such a dataset for the forensics research community interested in exploring the comparison of reference fingerprints of cameras estimated from still images and video. Although this dataset is efficient and useful for the development of video forensic algorithms, the VISION database is not sufficiently sizeable to provide the large number of training video clips needed for video source and camera model identification, as it only contains between 10 and 30 videos from each device (Shullani *et al.*, 2017).

The Surrey University Library for Forensic Analysis (SULFA) consists of a database of original and forged video files which was constructed for the purpose of video forensics to be used in applications for video camera identification and integrity verification (Qadir, Yahaya and Ho, 2012). In this dataset, special consideration was given to the development of the spatial and temporal characteristics of natural videos. For the construction of forged videos, Adobe Photoshop and AfterEffects software was used to alter the original videos. To further verify the SULFA dataset for video forensic algorithms, experiments were conducted with PRNU-based camera identification techniques to detect forgeries and for camera identification. The results have shown very good detection rates in forgery detection and camera identification. The authors also aimed in future work to design and develop more videos from different sources such as camcorders and smart phones. Hence, this dataset is specifically focused on the purposes of video identification and forgery detection. The dataset was also constructed to support research on the cloning that can be implemented by means of Adobe Photoshop CS3 and Adobe AfterEffect CS5 (Qadir, Yahaya and Ho, 2012). The Digital Media Forensic Challenges (MFC) dataset is one of the largest-scale benchmark datasets constructed specifically for media forensics evaluation tasks (Guan *et al.*, 2019). The authors constructed a series of datasets for different purposes in media forensic system evaluations, such as evaluating images and video device identification, forgery detection, and event verification. Five evaluation tasks were performed: image manipulation detection and the localization of images and videos (MDL); splice manipulation detection and localization (SDL);

event verification (EV); provenance filtering (PF); and provenance graph building (PGB) for media forensics. In image and video MDL, the problem is to detect if an image or video has been tampered and, if so, to then locate the region that was manipulated. In video MDL, temporal localization is used to identify the manipulated frame temporally. The goal of the SDL task is to detect whether or not a region of a given image has been spliced into an examined image and, if it has, then the task is to provide mask regions for both images. For the EV task, a number of images or videos were taken during an event and then the test image can be assessed as to having been captured during the event. The goal here is to verify that test images were taken during that event. For this reason, collections of different events were developed, such as hurricanes, parades, marathon races and air shows, to verify that a tested image belongs to the given event. In the PF task, the test HP images and manipulated images were analyzed against the given large-scale 'world' dataset. The task of PGB is to retrieve relevant images from the world dataset and then to rebuild a provenance phylogeny graph that shows the relationships of images with manipulation sequences. The two additional tasks of camera verification (CV) and generative adversarial networks (GAN) have also been considered. The CV task is to perform sensor fingerprint verification employing cross-modality evaluation. For instance, the system was trained with images but tested on videos sourced from the same set of cameras. In the GAN challenge task, the authors created a full MFC18 GAN dataset consisting of 1340 images and a GAN crop set comprised of 1000 images, whereas the GAN video set has a collection of 118 videos. For all evaluation tasks, the results showed that the MFC evaluation dataset has good capability to support the six tasks of media forensics. For all such evaluation tasks, the authors released specific datasets, such as the NC16 kickoff dataset, NC17 (development and evaluation dataset) and the MFC18 dataset which includes MFC18 development datasets 1 and 2, MFC18 evaluation, the MFC18 GAN image and video challenge dataset, a camera ID verification dataset, and MFC18 event verification development and evaluation datasets. The authors indicated that the development of MFC19 and MFC20 evaluation datasets will continue in the future. However, this database lacks information concerning camera model identification (Guan *et al.*, 2019).

The Video Authentication and Camera Identification (video-ACID) database is presented specifically for problems in identifying the model of camera used to take video images. The videos represent a diverse range of content, including clips with different lighting conditions and motion sequences. No

manipulation of the videos had been performed and all were unedited, representing the original format directly output by the camera. This dataset has been used in the development of systems used for the identification of video camera models, and the database also contains videos taken by 19 different smartphones and tablets. To reduce noise distortion, the videos were captured from rear-facing cameras and at the default zoom level. A wide range of content was collected when capturing videos useful for forensic purposes. The full dataset has been evaluated using a CNN algorithm in benchmarking state-of-the-art camera model identification problems (Hosler *et al.*, 2019). The Source Camera Recognition on Smartphones database named SOCRateS is comprised of a dataset of both still images and video, and was specifically constructed for the identification of source digital cameras on smartphones. The database is publicly available for the research community. SOCRateS was also designed for the benchmarking of image forensics algorithms for smartphones as well as source camera identification. The authors such as Galdi, Hartung and Dugelay (2019) highlighted the benefits of this database, which includes images from the highest number of different sensors which is useful for the benchmarking of methods for the identification of source smartphone cameras. The database is very useful in other digital image forensic techniques, such as the user authentication process where identity can be verified using iris, biometric or face recognition from images taken by a smartphone, and SOCRateS has been used in the proposal of a method combining smartphone identification with face recognition (Galdi, Hartung and Dugelay, 2019). The Daxing Smartphone Identification Dataset is the largest dataset that has used multiple smartphones per model for image and video source identification, and it is freely available for research purposes. The images were directly copied from smartphones without any post-processing. The resolution of the images and videos captured are the same as the resolution size of the sensor. The authors intend to further extend this dataset by including audio samples from the same devices for smartphone identification (Tian *et al.*, 2019). The Warwick Image Forensics dataset was presented in 2020 for use by researchers in image device fingerprinting. It consists of flat-field images using seven different exposure (ISO) settings. Their dataset of images provides good diversity in scenes and exposure parameter settings. The main aim of this database is to give researchers deeper insights concerning the impact of using a diversity of exposure settings and parameters as well as HDR imaging in the identification of device fingerprints. For this purpose, Quan *et al.* (2020) developed a set of time-sequential HDR images with different exposure settings. For each camera, they generated about 100 flat field-images to be used for reference in SPN extraction. Whilst capturing the flat-field images, they

also ensured that they should not be too dark or saturated, and therefore they kept the exposure settings as normal. They captured the SDR images in single-shot mode taken with different exposure settings ranging from ISO 100 to 6400. For each ISO speed, they generated 30 images with various different scenes and conditions for each camera. The HR images of 20 different scenes were captured using three different modes: AEB (auto exposure bracketing), in which continuous shots of images are captured with different exposure settings; normal exposure burst shots and under exposure burst shots; and fast continuous shots selected so as to obtain seven continuous shots of images with the same exposure settings. From a number of experiments, they showed that different ISO speeds have different levels of impact on SPN quality. Hence, special attention was drawn to the diverse range of exposure setting parameters to be used in SDR and HDR images for the purpose of device fingerprinting (Quan *et al.*, 2020). Below Table 2-1 highlighted the key differences between the existing image forensic datasets.

Table 2-1 Characteristics of Datasets used for Digital Image and Video Forensics

Datasets	**No. of Digital camera devices**	**No. of Images/Videos**	**Purpose**
Uncompressed Colour Image Dataset (UCID) (Schaefer and Stich, 2003)	Single Minolta Dimage 5 camera	1338 uncompressed images in tag image file format (TIFF) format. The images are very small in size at either 512 x 384 or 384 x 512 pixels	Evaluation of image retrieval techniques.
Dresden Image database (Gloe and Böhme, 2010)	73 camera devices	This dataset contains more than 14000 images. The image resolution ranges from 3072 x 2304 to 4352 x 3264 pixels.	Source camera identification and Camera model identification.
RAw ImageS datasEt (RAISE) (Dang-Nguyen *et al.*, 2015)	3 Nikon devices: the Nikon D40, Nikon D90 and Nikon D7000	8156 high-resolution raw images. All of the images were captured in high resolution at 3008 x 2000, 4288 x 2848, and 4928 x 3264 pixels stored in an uncompressed	Camera verification tasks.

		format of either Compress Raw12-bit or Lossless Compress Raw14-bit.	
Video and Image Dataset for Source Identification (VISION) (Shullani *et al.*, 2017)	35 portable smartphone and tablet devices belonging to 11 different brands including Apple, Asus, Huawei, Lenovo, Lucky Goldstar (LG) electronics, Microsoft, OnePlus, Samsung, Sony, Wiko and Xiaomi.	The dataset consists of 34427 images and 1914 videos. The video resolution ranges from 640 x 480 to 1920 x 1080 and the image resolution ranges from 960 x720 to 5312 x 2988 pixels.	Video Forensics: Video Source and Camera model identification
Surrey University Library for Forensic Analysis (SULFA). (Qadir, Yahaya and Ho, 2012)	3 different camera brands: the Canon SX220, Nikon S3000 and Fujifilm S2800HD	150 videos compressed in H.264/AVC and Motion JPEG (MJPEG) format from three different camera brands. The resolution size of each video is 320 x 240 pixels with a duration of 10 seconds and 30 fps (frames per second)	Video forensics to be used in applications for video camera identification and integrity verification
Digital Media Forensic Challenges (MFC). (Guan *et al.*, 2019)	About 574 distinct HP cameras were enrolled to build this database.	176000 high provenance (HP) images and 11000 HP videos. In addition to this, more than 100000 manipulated images, 4000 manipulated videos, 35 million internet images and 300000 video clips are included.	Media forensics evaluation tasks such as evaluating images and video device identification, forgery detection, and event verification.
Video Authentication and Camera Identification (video-ACID). (Hosler *et al.*, 2019).	The video clips are captured from 46 physical devices belonging to 36 different models and 18 different brands of camera.	The database consists of more than 12000 video clips. Over 250 of the videos were captured by each device. The highest quality of videos are represented in	Identification the model of camera used to take video images.

		this database, with 1080P at 30fps. The duration of each video is longer than 5 seconds.	
Source Camera Recognition on Smartphones database (SOCRateS). (Galdi, Hartung and Dugelay, 2019).	103 different smartphones belonging to 15 different brands and about 60 different models.	9700 images and 1000 videos. One hundred pictures were captured from each device, of which 50 depicted blue-sky images and the other 50 portray different kinds of scenes. Approximately 10 video clips were collected from each smartphone device.	Smartphones Identification.
Daxing Smartphone Identification Dataset. (Tian *et al.*, 2019).	90 smartphones, including 22 models belonging to 5 brands	43400 original images and 1400 original videos. From each smartphone device, more than 482 images were captured with 8 different types of scene content.	Smartphones Identification.
Warwick Image Forensics Dataset. (Quan *et al.*, 2020).	14 digital cameras	58600 flat-field images, standard dynamic range (SDR) images and high dynamic range (HDR) images.	Image device fingerprinting

2.7 CONCLUSION

The motivations for the development of digital forensics techniques have been considered in this chapter. Photo Response Non-Uniformity (PRNU) can be considered a strong unique pattern of images produced by an imaging sensors and its use has been a key element in many digital applications. A large number of studies have proved that PRNU is a well-established forensic tool for matching an image to a source camera. Approaches to PRNU estimation have traditionally relied on wavelet-based de-noising techniques. Many researchers have made significant contributions to source camera identification (SCI)

applications. Many other critical issues with SPN application such as improving their robustness have also been addressed. While a variety of forensic tools and techniques have been developed for SCI, little research has been conducted in the area of the temporal forensic analysis of digital images. The increasing importance of temporal information in forensics applications has created a need to develop techniques for temporal forensic image analysis. The idea is motivated by the fact that all semiconductors and electronic components degrade with time gradually. Therefore, defective pixels also constitute another sensor imperfection of forensic relevance which can be useful for picture dating applications. As advanced research in multimedia forensics continues to develop, there is a need for more datasets to be available in the field of digital image and video applications. The highest quality of results depends on datasets which should have sufficient images for training and validation using machine learning or deep learning techniques. However, very little research has been conducted in the area of the forensic analysis of videos. Only a few digital video datasets are available for use in forensics scenarios. Research into video forensics algorithms is proceeding at a slower pace compared to work in digital image forensic methods. This is due to the lack of availability of large databases of videos for use in the benchmarking of video forensics algorithms. The research community is still working on forensic algorithms to identify video manipulation and forgery, the detection of video frame deletion, and the identification of source devices and camera models. However, among the image and video databases currently used for benchmarking, there is as yet no publicly available database suitable for the specific purpose of temporal image forensics analysis.

CHAPTER 3 THE 'NORTHUMBRIA TEMPORAL IMAGE FORENSICS' DATABASE: DESCRIPTION AND ANALYSIS

3.1 MOTIVATION

Imaging technologies now play a crucial role in forensics and criminology when there is an encouraging rise in digital images. There are many situations in which imaging technologies play a vital role in forensic work. Images and video recordings are important means to document the condition of a specific subject at any instant in time. The content of such images might include scenes related to crimes or accidents, victims, suspects or any items of evidence in a laboratory. Thorough and complete photographic images from recordings made by a camera or any digital imaging technology showing a scene, suspect or items of evidence can be helpful in gathering meaningful information which can lead to the identification of direct or circumstantial evidence. To make inferences and establish correlations, temporal image analysis in digital image forensics gives analysts the perfect idea to predict the connections of crimes and its whereabouts. Also, forensic researchers can examine, model and analyze the patterns of images within a time series to make links between events. To derive insights from temporal image analysis, it is important to have available temporally ordered sequence of images taken at regular intervals of time which could provide opportunities for analysts or forensic investigators to analyze incidents or link different events by monitoring temporal parameters from digital images. Several algorithms have been proposed over the last two decades to find the source of an image for source camera identification (SCI) purposes. Blind image forensics takes advantage of and exploits the traces left behind by the stages of processing during image acquisition and storage to determine the history of an image. Among those traces, the most useful camera trace found so far is photo response non-uniformity (PRNU), which is a unique sensor 'fingerprint' which can therefore be used to identify a specific camera (Lukáš, Fridrich and Goljan, 2006). Extensive research has been done conducted to identify the sources of images, and the database most commonly used to fulfil the purpose of SCI in digital forensic academics is the Dresden Image Database (Gloe and Böhme, 2010).

However, only two research papers have been published in the academic literature which focus specifically on the temporal forensic analysis of digital pictures. Mao *et al.* (2009) research was based on time-dependent camera parameters that were assumed to be included in PRNU, while Fridrich and

Goljan (2011) employed defective pixels to establish temporal relationships. The authors of both studies acknowledged that the lack of regular acquisition time slots for the capture of digital pictures limited their analysis and the performance of the methods proposed.

To the best of our knowledge, there is no database available for use by forensic researches to analyze a regular series of images over time. To overcome some of the problems in research in temporal image forensics, the Northumbria Temporal Image Forensics (NTIF) database is introduced which will be helpful in the wider fields of digital image forensics and criminology (Ahmed *et al.*, 2020a). By using this database, researchers can conduct analysis to deduce the approximate timeline of pictures taken by the same camera device. The database can also be used for source camera identification and forgery detection (Lawgaly and Khelifi, 2016; Lawgaly, Khelifi and Bouridane, 2014; Lawgaly, Khelifi and Bouridane, 2013; Al-Ani and Khelifi, 2016; Al-Ani *et al.*, 2015; Ahmed *et al.*, 2019).

3.2 CHARACTERISITCS OF IMAGES

In the application of forensic analysis techniques, a number of parameters may be involved. In fact, studying the behaviour of a set of pixels over time, using different camera brands and models, different exposure setting parameters (ISO, exposure time, aperture), as well as different environmental conditions and different image contents need to be taken into account. Therefore, substantial effort is devoted to the consideration of all relevant parameters when constructing a database of images to be useful for digital image forensic applications, including the temporal analysis of digital images. For instance, a large number of images taken from different cameras of the same make and model is necessary for source camera identification. In this context, only one suitable image forensics database has been available for researchers, which was developed by Gloe and Böhme (2010). This has been used for source camera identification and camera model classification, but the Dresden database does not have structured images taken at regular intervals of time to be appropriate for use in temporal image forensics. In fact, to the best of our knowledge, there is no publicly accessible dataset which allows the behaviour of pixels over time to be analyzed and the acquisition time of digital pictures to be estimated accordingly.

The main feature of the NTIF database is that a large number of images have been taken at different intervals of time on a regular basis using a number of cameras of various makes and models. All the images depict natural scenes and blue-sky images are included, and great care has been taken to avoid any copyright violations or privacy issues.

A total of 10 digital cameras were selected in constructing the database of digital images on a weekly/bi-weekly basis. In total, 41684 images were captured for the NTIF database. Among them are 980 blue sky images taken by different cameras. Out of 10 digital cameras, 5 are different models from the same manufacturer. All of the digital cameras are available in the highest resolution and JPEG quality settings. Most of the images captured by the cameras involve similar indoor and outdoor scenes. The database also consists of images taken by different individual cameras of the same make and model. This allows researchers to assess source camera identification techniques in practical and realistic scenarios. Each digital camera has been used to produce digital images taken at regular intervals spanning a 94-week period of time. The image set constructed for each camera is designed in such a manner that several images are grouped and categorized sequentially in weekly or bi-weekly time slot subsets up to 94 weeks. For example, Figure 3-1 illustrates the number of images that were categorized on a weekly and bi-weekly basis from time slots numbered from 1 to 71 for the Canon IXUS115HS-1. Samples of natural images captured by the Canon IXUS115HS-1 are depicted in Figure 3-2.

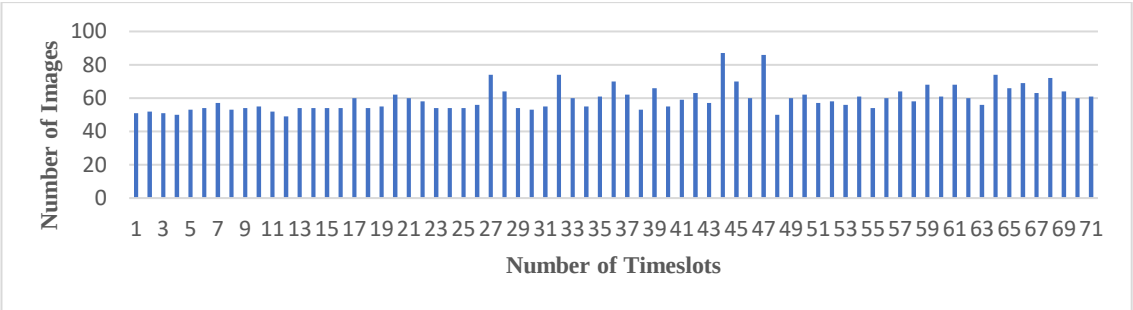


Figure 3-1 The numbers of images categorized from 1 to 71 time slots covering a 94-week period for the Canon IXUS115HS-1 camera

Figure 3-2 Sample of images with different indoor and outdoor scenes captured from the Canon IXUS115HS-1 digital camera

Details relating to each digital camera and its respective images, including make, model, the whole acquisition time-period of images, camera resolution size, image format, type of sensors, sensor size, number of natural and blue sky images, and range of exposure settings have been recorded, as depicted in Table 3-1.

Table 3-1 Characteristics of NTIF Database

Camera model	**No. of devices**	**Resolution size**	**Sensor type**	**Sensor size (inches)**	**Total time-period of image capture**	**ISO (min-max)**	**Focal length (mm) (min-max)**	**Shutter speed (min-max)**	**No. of images**	**No. of blue-sky images**
Canon IXUS115HS-1	1	4000 × 3000	CMOS	1/2.3”	21st Oct 2014-30th Dec 2016	100-1600	5-20	6.6667e-4-0.1250	4234	191
Canon IXUS115HS-2	1	4000 × 3000	CMOS	1/2.3”	21st Oct 2014-30th Dec 2016	100-1600	5-20	6.6667e-4-0.1250	4269	159
Fujifilm S2950-1	1	4288 × 3216	CCD	1/2.3”	22nd Oct 2014-28th Dec 2016	64-800	5-75	6.2500e-04-0.2500	4057	122

Fujifilm S2950-2	1	4288 × 3216	CCD	1/2.3”	22nd Oct 2014-27th Dec 2016	64-1600	5-75	6.25e-04-0.25	4064	-
Nikon Coolpix L330-1	1	5152 × 3864	CCD	1/2.3”	21st Oct 2014-28th Dec 2016	80-1600	4-104	6.6667e-04-1	4117	-
Nikon Coolpix L330-2	1	5152 × 3864	CCD	1/2.3”	1st Jan 2014-25th Dec 2016	80-1600	4-71.5	6.6667e-04-0.7692	4239	137
Panasonic DMC TZ20-1	1	4320 × 3240	CMOS	1/2.33”	21st Oct 2014-29th Dec 2016	100-1600	4.3-68.8	4e-04-0.25	4358	143
Panasonic DMC TZ20-2	1	4320 × 3240	CMOS	1/2.33”	21st Oct 2014-30th Dec 2016	100-1600	4.3-41.5	5e-04-0.5	4193	104
Samsung pl120-1	1	4320 × 3240	CCD	1/2.33”	21st Oct 2014-30th Dec 2016	80-800	4.9-24.5	5e-04-0.125	4089	124
Samsung pl120-2	1	4320 × 3240	CCD	1/2.33”	21st Oct 2014-30th Dec 2016	80-800	4.9-24.5	5e-04-0.7	4064	-
∑	**10**								**41684**	**980**

3.3 RELEVANCE OF DATABASE TO TEMPORAL DIGITAL IMAGE FORENSICS

This section demonstrates the relevance of the database to temporal image forensics. It mainly highlights the changes detected in PRNU over time.

3.3.1 Sample of Images used in Experiments

NTIF database has been used to highlight the changes in PRNU over time. A total of 10 digital cameras have been used, from each of which 3200 indoor and outdoor scene images which are uniformly spanned over 64 time slots were considered. The list of cameras used in experiments is given in Table 3-2. To ensure that reasonable time should be allowed for the imaging sensor to age over time, the images captured over an 8-week period are grouped in separate classes, where class 1 consists of images taken in weeks 1-8, class 2 represents images taken in weeks 9-16, and so on. As a result, each class consists of 400 images covering an 8-week period as shown in Table 3-3 are used to show the intra-correlations and inter-correlations.

Table 3-2 Characteristics of Digital Cameras used in Experimental study

Name of digital camera	**Size of resolution**	**Type of sensors**
Canon IXUS115HS-1	4000 × 3000	CMOS
Fujifilm S2950-1	4288 × 3216	CCD
Nikon Coolpix L330-1	5152 × 3864	CCD
Panasonic DMC TZ20-1	4320 × 3240	CMOS
Samsung pl120-1	4320 × 3240	CCD
Canon IXUS115HS-2	4000 × 3000	CMOS
Fujifilm S2950-2	4288 × 3216	CCD
Nikon Coolpix L330-2	5152 × 3864	CCD
Panasonic DMC TZ20-2	4320 × 3240	CMOS
Samsung pl120-2	4320 × 3240	CCD

Table 3-3 Number of Classes and 512×512 Block Images used for each Digital Camera to show Intra-correlations and Inter-correlations

Class	No. of Weeks	No. of Images
1	1-8	400
2	9-16	400
3	17-24	400
4	25-32	400
5	33-40	400
6	41-48	400
7	49-56	400
8	56-64	400
Total	**64**	**3200**

3.4 EXPERIMENTS AND RESULTS

PRNU has been proven to be an effective method for the identification of a source device. However, if only a single image is used to estimate PRNU, the entity to use as an approximation of PRNU would be the noise residual.

The noise residual is extracted from a single image by using a de-noising filter. Many state-of-the-art methods have been introduced to estimate noise-free coefficients; for instance, Mihcak's wavelet-based filter (Mihcak *et al.*, 1999) , BM3D (Dabov *et al.*, 2007), LADCT (Lawgaly and Khelifi, 2016), the wavelet-based de-noising Weiner filter (Lukáš, Fridrich and Goljan, 2006) and the PCA-based de-noising filer (Li, Guan and Li, 2014). In this chapter, the most popular Weiner de-noising filter, which was introduced by (Lukáš, Fridrich and Goljan, 2006), is used. To compute the noise residuals of images from the given set of N images, suppose that I^1, I^2, …., I^N are the images generated by camera C. The noise residuals W^1, W^2,….., W^N are obtained by using a wavelet-based de-noising filter F. Equation 3.1 below shows the process of the extraction of noise residuals from the original images:

$$W^N = I^N - F(I^N) \tag{3.1}$$

For the experiment, the noise residuals were estimated from cropped blocks from the central part of images and which are set to be of the size 512×512. Hence, the total number of noise residuals extracted per class is 400, resulting in a total of 3200 corresponding noise residuals extracted from 8 classes of each camera. Once the noise residuals were extracted, the intra-correlations and inter-correlations can be computed. An intra-correlation is defined as the correlation between two distinct noise residuals from the same class (that is, noise residuals corresponding to images taken within the same timeframe), whereas an inter-correlation, on the other hand, is the correlation between two noise residuals of different classes (noise residuals corresponding to images taken in different timeframes).

It's worth noting that the similarity between noise residuals has been assessed using a number of measures in the literature, including the peak-to-correlation energy (PCE) (Goljan, 2008), correlation over circular cross-correlation norm (CCN) (Kang *et al.*, 2011b) and normalized correlation (Lukáš, Fridrich and Goljan, 2006). Without loss of generality, the normalized correlation ρ_{W^1,W^2} is used which is computed between two noise residuals W^1 and W^2 as shown in Equation 3.2.

$$\rho_{W^1,W^2} = \frac{(W^1 - \mu_{W^1}).(W^2 - \mu_{W^2})}{\| W^1 - \mu_{W^1} \|.\| W^2 - \mu_{W^2} \|} \tag{3.2}$$

where μ_{W^1} and μ_{W^2} are the mean values of the W^1 and W^2 vectors.

To compute the intra-class correlation of one class from a single camera, one noise residual image (as a reference noise residual) is correlated against all other noise residual images within the same class. As a result, a total of $400^2 - 400 = 159600$ correlation scores for each class of the camera were obtained. All of the scores are then averaged to obtain the final score of intra-correlation per camera. This process is repeated for each camera.

For inter-class correlation, each noise residual image (reference noise residual) from one class (reference class with 400 noise residuals) was correlated against the noise residuals of all other classes (2800 noise residuals). As a result, a total of $400 \times 2800 = 1120000$ inter-class scores for every

reference class is computed. The scores are then averaged to get the final score of inter-correlation per camera. The results in Table 3-4 show that the intra-correlations for noise residuals generated from images in the same time period are clearly higher than the inter-correlations of noise residuals taken in different timeframes over the 64 weeks timespan. This means that there is a strong correlation between the noise residuals taken from the same class. So, when the noise residuals are correlated with different timeframes of classes, the strength of correlations decreases, which means that the strength of noise residuals becomes weaker over time or less well-correlated with other timeframes of class. For instance, in Table 3-4 the average scores of intra-correlations of camera Canon IXUS115HS-1 and Fujifilm S2950-1 are 0.0340 and 0.0724 which is higher than its average score of inter-correlations which are 0.0310 and 0.0577. So, image pixels started to get defective in image sensors over the time. Therefore, changes in the noise residual fingerprint can be observed over the 64 weeks timespan, conditioned by the ageing of the sensors.

Table 3-4 Results of Intra-correlation and Inter-correlations of Ten Digital Cameras

Digital Cameras	**Mean Score of Intra-correlations**	**Mean Score of Inter-correlations**
Canon IXUS115HS-1	0.0340	0.0310
Fujifilm S2950-1	0.0724	0.0577
Nikon Coolpix L330-1	0.0265	0.0190
Panasonic DMC TZ20-1	0.0857	0.0836
Samsung pl120-1	0.1237	0.0968
Canon IXUS115HS-2	0.0466	0.0422
Fujifilm S2950-2	0.0861	0.0606
Nikon Coolpix L330-2	0.0250	0.0171
Panasonic DMC TZ20-2	0.1131	0.1060
Samsung pl120-2	0.1030	0.0858

Furthermore, PRNU fingerprints of different camera sensors have been analyzed to provide evidence for sensor ageing over time. For this purpose, each camera's dataset of images is divided into five classes (with approximately 15 time slots per class). For each class of camera, approximately 750 images (with approximately 50 images per time slot) are used to generate the PRNU fingerprint, which is simply the average of the noise residuals in each class (time slot). In each time slot, there are images

from a weekly or bi-weekly period. Thus, a total of 5 classes are used to show the changes in PRNU over time with respect to its reference PRNU for every digital camera. That is, class 1 consists of approximately 750 images taken in time slots 1-15, class 2 represents 750 images taken in time slots 16-30, and so on. The PRNU noise residual is then extracted from 750 colour images with a size of resolution of a sensor from each class. A total of 5 PRNU fingerprints (spanning up to 71-73 time slots covering 94 weeks) were generated by extracting the PRNU using a wavelet-based Wiener de-noising filter for each digital camera. From a total of 5 classes, each class act as reference PRNU and correlated with other PRNUs of classes by using normalized correlation. In Table 3-5, the highest correlation value of 1 is obtained when PRNU from the same classes is correlated with itself. For example, in Table 3-5 the PRNU of time slots 1-15 is correlated with the reference PRNU of same class (time slots 1-15). Therefore, all classes have the high correlation value of 1 as the reference PRNU of each class correlated with itself.

From the results for the ten digital cameras as shown in Table 3-5, it can be observed that for most of the classes the correlation scores which are nearer to the respective class reference PRNU (which is 1) have higher correlations, and the scores of classes which are far from the reference PRNU class showing weak correlations over time. Thus, the NTIF database can be a good means to highlight such changes in sensor characteristics on a temporal basis which can be exploited further in image forensics.

Table 3-5 Changes in PRNU Correlations over Time

Classes	**1**	**2**	**3**	**4**	**5**
	Timeslots (1-15)	***Timeslots (16-30)***	***Timeslots (31-45)***	***Timeslots (46-60)***	***Timeslots (61-73)***
	Canon IXUS115HS-1_PRNU_Correlations				
1	1	0.551250	0.571357	0.572748	0.494645
2	0.551250	1	0.736748	0.696113	0.608135
3	0.571357	0.736748	1	0.772543	0.656592
4	0.572748	0.696113	0.772543	1	0.658268
5	0.494645	0.608135	0.656592	0.658268	1
	Fujifilm S2950-1_PRNU_Correlations				
1	1	0.635052	0.550815	0.544425	0.531021
2	0.635052	1	0.789016	0.784475	0.741334

3	0.550815	0.789016	1	0.904002	0.854614
4	0.544425	0.784475	0.904002	1	0.848889
5	0.531021	0.741334	0.854614	0.848889	1
	Nikon Coolpix L330-1_PRNU_Correlations				
1	1	0.48447	0.558866	0.552941	0.537059
2	0.48447	1	0.659799	0.656548	0.623062
3	0.558866	0.659799	1	0.795315	0.76377
4	0.552941	0.656548	0.795315	1	0.758505
5	0.537059	0.623062	0.76377	0.758505	1
	Panasonic DMC TZ20-1_PRNU_Correlations				
1	1	0.692824	0.660519	0.651409	0.602928
2	0.692824	1	0.832423	0.83404	0.80022
3	0.660519	0.832423	1	0.901151	0.826104
4	0.651409	0.83404	0.901151	1	0.838008
5	0.602928	0.80022	0.826104	0.838008	1
	Samsung pl120-1_PRNU_Correlations				
1	1	0.673287	0.606406	0.605888	0.580013
2	0.673287	1	0.901754	0.897895	0.860183
3	0.606406	0.901754	1	0.931408	0.884516
4	0.605888	0.897895	0.931408	1	0.889795
5	0.580013	0.860183	0.884516	0.889795	1
	Canon IXUS115HS-2_PRNU_Correlations				
1	1	0.605448	0.608573	0.591696	0.584656
2	0.605448	1	0.792462	0.789691	0.784672
3	0.608573	0.792462	1	0.827488	0.819144
4	0.591696	0.789691	0.827488	1	0.848829
5	0.584656	0.784672	0.819144	0.848829	1
	Fujifilm S2950-2_PRNU_Correlations				
1	1	0.711028	0.580929	0.567745	0.561392
2	0.711028	1	0.801994	0.797299	0.777057
3	0.580929	0.801994	1	0.915688	0.890757
4	0.567745	0.797299	0.915688	1	0.901678
5	0.561392	0.777057	0.890757	0.901678	1
	Nikon Coolpix L330-2_PRNU_Correlations				
1	1	0.522641	0.552734	0.554840	0.555255
2	0.522641	1	0.666750	0.666844	0.651406

3	0.552734	0.666750	1	0.792599	0.779158
4	0.554840	0.666844	0.792599	1	0.782999
5	0.555255	0.651406	0.779158	0.782999	1
Panasonic DMC TZ20-2_PRNU_Correlations					
1	1	0.709098	0.664563	0.685091	0.675229
2	0.709098	1	0.853287	0.843076	0.844832
3	0.664563	0.853287	1	0.907534	0.899984
4	0.685091	0.843076	0.907534	1	0.904788
5	0.675229	0.844832	0.899984	0.904788	1
Samsung pl120-2_PRNU_Correlations					
1	1	0.695992	0.629976	0.620483	0.608472
2	0.695992	1	0.911120	0.905259	0.884481
3	0.629976	0.911120	1	0.934255	0.912187
4	0.620483	0.905259	0.934255	1	0.913928
5	0.608472	0.884481	0.912187	0.913928	1

In Figures 3-3 to 3-7 below for ten digital cameras, each graph shows five time slot reference values of PRNU for weeks 1-15, 16-30, 31-45, 46-60 and 61-75 when compared the strength of the correlations with other time slots.

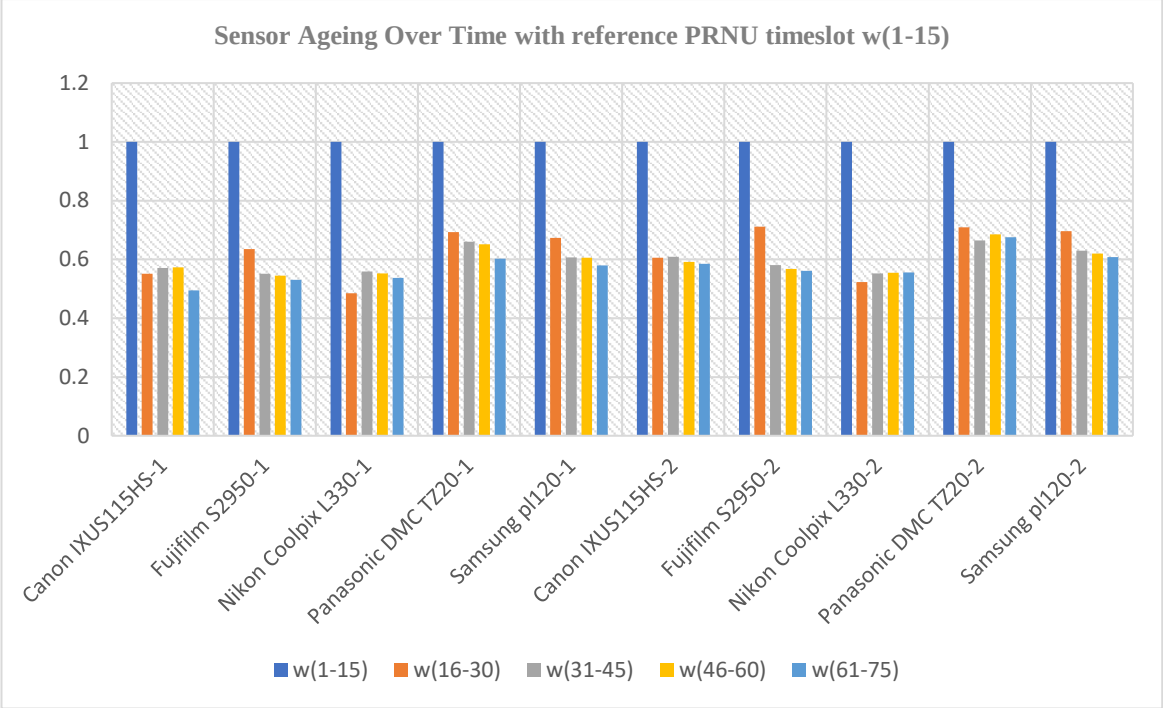


Figure 3-3 Correlation values which are nearer to the reference PRNU (which is equal to 1) of the respective class for weeks 1-15 have higher correlations over time, and values which are far away from the reference PRNU class for weeks 1-15 have weaker correlations

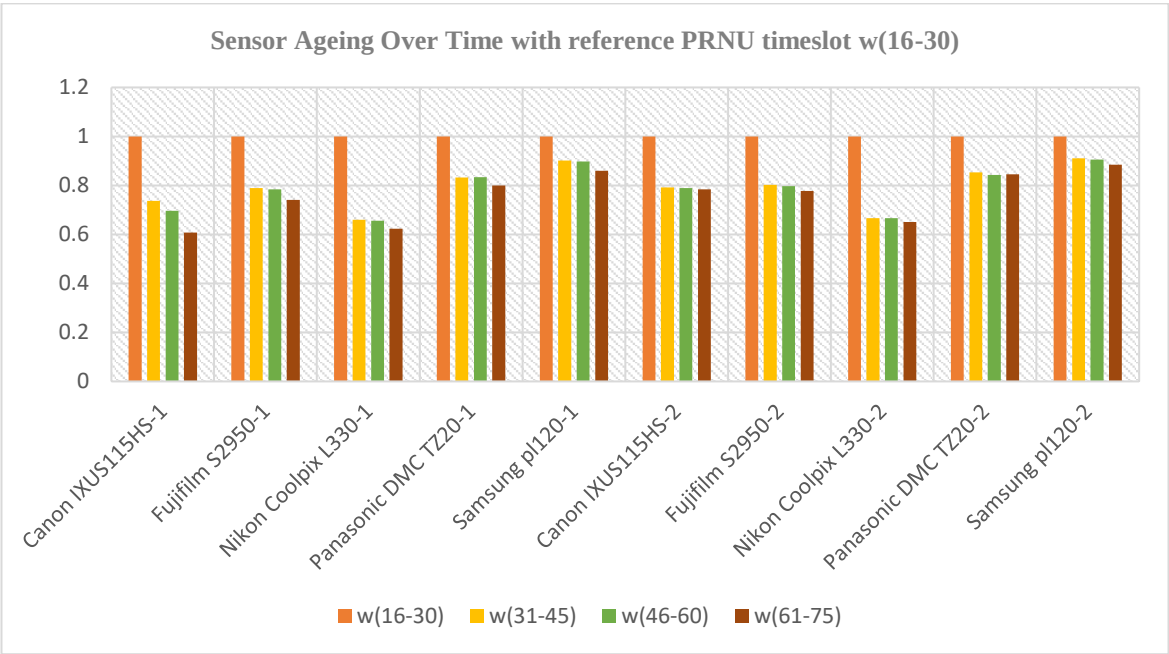


Figure 3-4 Correlation values which are nearer to the reference PRNU (which is equal to 1) of the respective class for weeks 16-30 have higher correlations over time, and values which are far away from the reference PRNU class for weeks 16-30 have weaker correlations

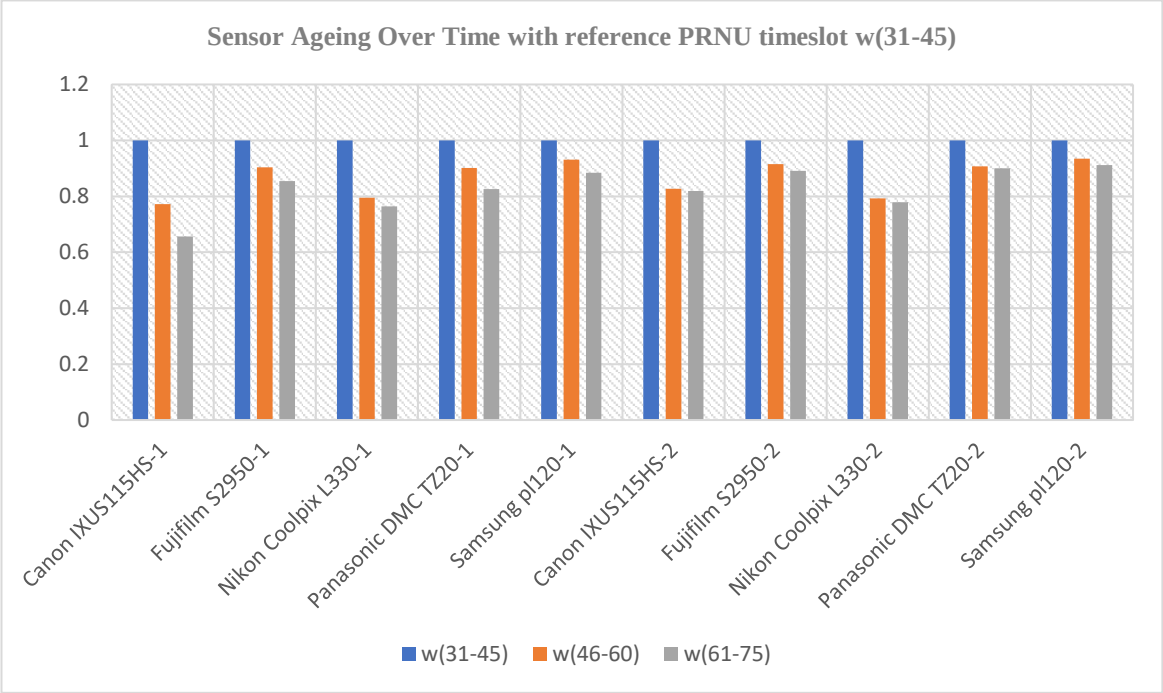


Figure 3-5 Correlation values which are nearer to the reference PRNU (which is equal to 1) of the respective class for weeks 31-45 have higher correlations over time, and values which are far away from the reference PRNU class for weeks 31-45 have weaker correlations

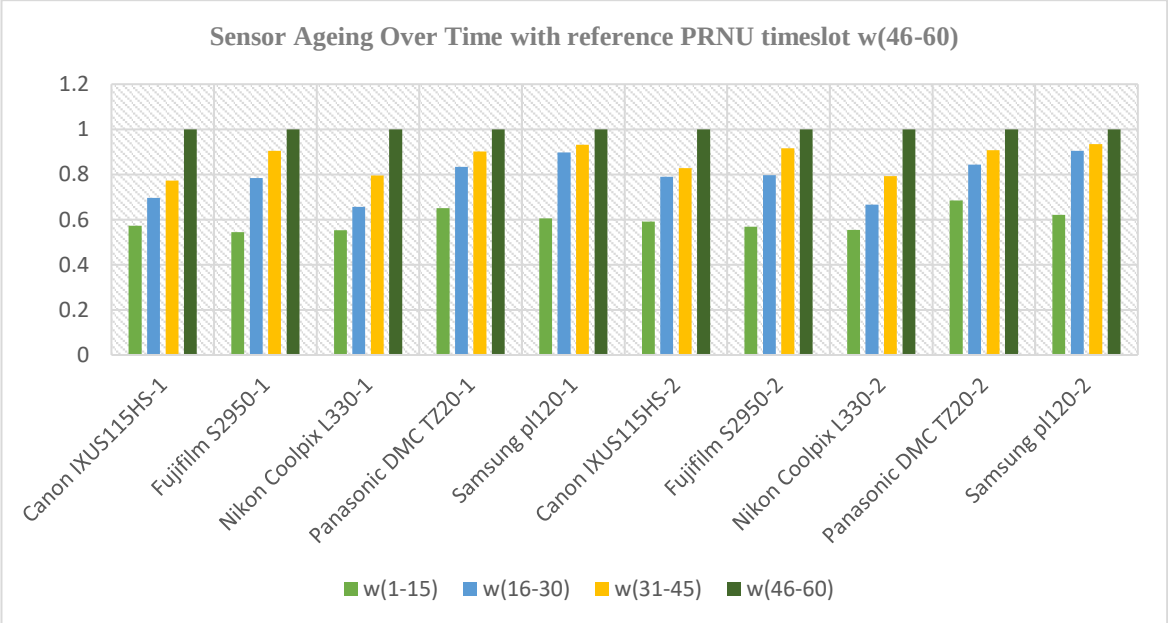


Figure 3-6 Correlation values which are nearer to the reference PRNU (which is equal to 1) of the respective class for weeks 46-60 have higher correlations over time, and values which are far away from the reference PRNU class for weeks 46-60 have weaker correlations

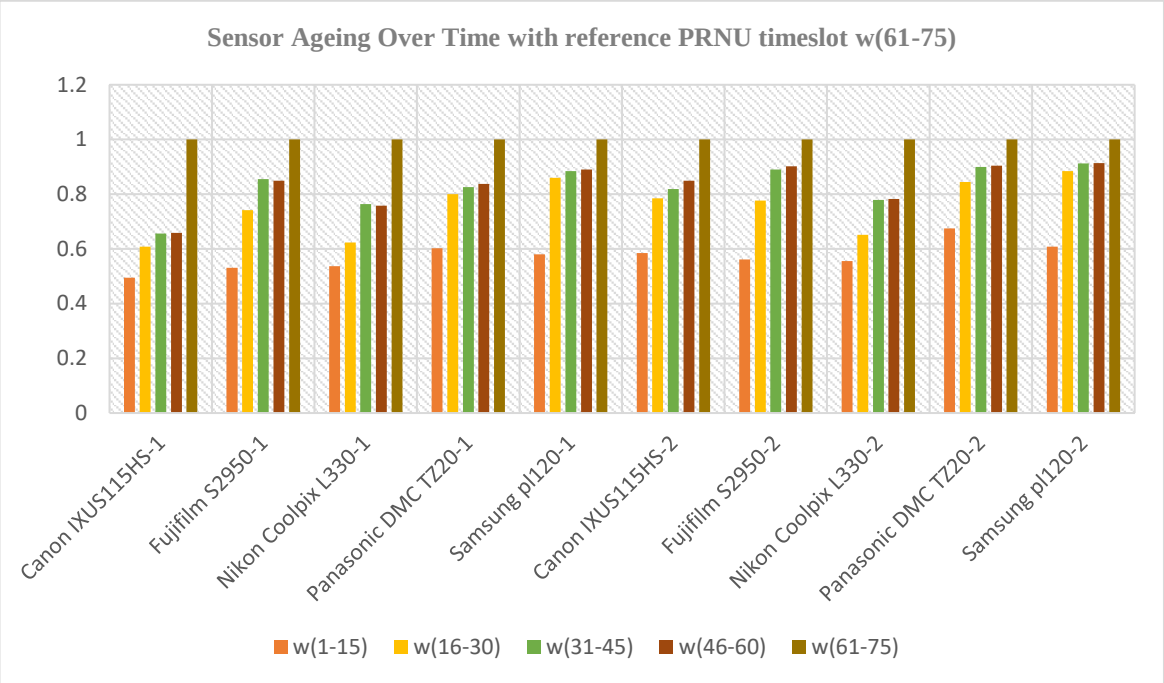


Figure 3-7 Correlation values which are nearer to the reference PRNU (which is equal to 1) of the respective class for weeks 61-75 have higher correlations over time, and values which are far away from the reference PRNU class for weeks 61-75 have weaker correlations

3.5 Database Access

The NTIF database has been made freely available to the research community for scientific and research purposes. Access to the NTIF requires the license agreement file to be signed, which is accessible along with the NTIF at the following URL (Uniform Resource Locator):

https://github.com/Northumbria-CIS/Northumbria-Temporal-Image-Forensics-NTIF-Database

3.6 Conclusion

In this chapter, a temporal image database called the Northumbria Temporal Image Forensics (NTIF) is introduced which is specifically designed for temporal digital image forensic analysis. The current benchmark dataset in image processing, which is the Dresden Image Database, has played a critical role in advancing source camera identification research. The characteristics, quality and relevance of NTIF indicate that it can be widely used for many applications, including source camera identification, brand and model classification, image authentication and image acquisition time estimation. The high-quality

images of the NTIF will enable it to become a new and challenging benchmark dataset for future research that could potentially lead to breakthroughs in image forensics.

CHAPTER 4 COMPARATIVE ANALYSIS OF A DEEP CONVOLUTIONAL NEURAL NETWORK FOR SOURCE CAMERA IDENTIFICATION

4.1 MOTIVATION

Recently, deep learning and in particular CNN-based approaches have been gaining attention in relation to many computer vision topics such as image classification, biometrics, source camera identification (SCI) and medical image analysis. For SCI, large numbers of training images have been required to train the system. In addition, high computing resources are required, with the training process taking days to complete. In contrast to deep learning, in traditional machine learning based on the PRNU approach for source camera identification, the stages of feature selection, extraction and classification are done separately, and limited numbers of training images are used to extract the PRNU. These factors provided the motivation to discover which approaches are effective in the field of digital image forensics for SCI when only limited numbers of images are available to the analyst. Therefore, the aim of this chapter is to shed light on the suitability of approaches to source camera identification in digital forensic science by comparing state-of-the art technique based on the filtered-based PRNU or deep learning convolutional neural networks (CNNs).

4.2 CNN ARCHITECTURE

4.2.1 Basic CNN Structure

Basically, a CNN is a complex computational model inspired by the human brain's neural system which consists of many interconnected nodes organized in stacked layers to perform simple operations on input data. In this way, a CNN can easily capture patterns in the data. There are many pre-trained CNN networks available that have gained popularity and which accept predefined input dimensions to recognize objects in images. Some authors such as Freire-Obregón *et al.* (2019), Baroffio *et al.* (2016), Bondi *et al.* (2016) and Tuama, Comby and Chaumont (2016) have proposed their own CNN models by tuning a number of convolutional layers interspersed with an activation function using rectified linear units (ReLUs) to add non-linearity, batch normalization and max-pooling layers followed by fully-connected layers and classification to identify the source camera device. Convolution is an operation in which various operations are performed such as smoothing and sharpening as well as edge

and feature detection based on the two functions $f(x)$ and $g(x)$ that give a third function as output interpreted as a filtered version of function f. Therefore, the major advantage of deep learning is that it does not require the manual extraction of features from the image, as the network learns to extract features by using a number of convolutional kernels during the training process. In the convolution equation shown in Equation 4.1, $f(x)$ and $g(x)$ are the two functions of a discrete variable x, where g is interpreted as a filter.

$$f(x) * g(x) = \sum_{m=-\infty}^{\infty} f[m].g[x-m] \tag{4.1}$$

Hence, the large number of filters inside CNN layers mean that it automatically learns the features, and the output of the convolutional operation is in the form of a number of feature maps.

In the proposed CNN model for SCI, the main purpose of the convolutional layers inside the convolutional neural network is to extract features from the input image by using different sizes of matrix, which are also called 'filters' or 'kernels', that slide over the image according to different striding factors and compute the dot product that results in the form of different activation maps for a single input image. Therefore, the number of filters used in a convolutional layer provides different feature or activation maps when convolved over the entire image. If a larger number of filters are used in the convolution layer then a larger number of features maps are extracted and thus makes the network better able to recognize the patterns of validated data. For instance, if three filters are used in the convolution layer, three feature maps will be produced as an output. All of the feature maps are assembled at the output of the convolutional layers and are then forwarded to the next layer. However, the size of the activation maps is controlled by the three parameters in deep learning, which are depth, stride and padding. After the convolution operation, a non-linear activation function used in deep learning such as a ReLU layer is added to give non-linearity, where all the negative values inside the feature map are replaced by zero. In addition, spatial pooling is also applied, which reduces the dimensionality of each feature map but retains the most valuable information. This pooling is available in different types, such as Max, Average and Sum pooling. At the end of the CNN layers, a fully

connected layer as an 'output' layer of the CNN architecture relates to the SoftMax activation function layer that classify various features of images into different classes based on the training dataset.

4.2.2 The CNN Model of investigation

The framework of the CNN model of investigation for source camera identification is shown in Figure 4-1, where the detailed settings of the CNN architecture are described. The CNN model of investigation for SCI is constructed with one input layer, three convolutional layers with batch normalization and a rectified linear unit as an activation function, and a max pooling layer followed by one fully connected layer, as well as a drop-out and classification layer.

To fit the requirements of the model, fewer layers (6 layers) are used in the CNN chosen in order to save the time taken in computation and power required, unlike the standard CNN models such as GoogleNet (Szegedy et al., 2015) and AlexNet (Krizhevsky, Sutskever and Hinton, 2012) that use networks 22 and 8 layers deep respectively. The input layer of the CNN model consists of noise residual images extracted using a method similar to that defined in Lukáš, Fridrich and Goljan (2006). The first CNN layer (Convolution 1) treats the noise residual image with 128 kernels of size 3×3 and with a stride value of 2 followed by a max pooling operation with window size 2×2 and stride 2. The second convolutional layer (Convolution 2) applies convolutions using 96 kernels of size 3×3 again with a stride value of 2 followed by a max pooling operation with window size 2×2 and stride 2. The third convolutional layer (Convolution 3) applies convolutions with 64 kernels of size 3×3 with a stride value of 1 followed by a max pooling operation with window size 3×3 and stride 2.

For instance, an input noise residual image with dimensions of 128×128 is fed into the CNN network. It is convolved in a first convolutional layer with 128 filters of size 3×3 and stride of 2, resulting in dimensions of $64 \times 64 \times 128$. The next layer is a pooling operation with filter size 2×2 and stride of 2. Hence the resulting image dimension is $32 \times 32 \times 128$. Similarly, the second convolution layer involves a convolution operation with 96 filters of size 3×3 and stride of 2, resulting in dimensions of $16 \times 16 \times 96$. The second pooling operation with a similar filter size of 2×2 and stride of 2 results in image dimensions reduced to $8 \times 8 \times 96$. In the third convolutional layer, the convolutional operation with 64 filters of size 3×3 and stride of 1 results in image dimensions of $6 \times 6 \times 64$, followed by a

third pooling layer with filter size of 3×3 and stride of 2. Thus, the resulting image dimensions are reduced to $3 \times 3 \times 64$.

The ReLU layer is applied as an activation function to the output of every convolutional layer followed by a batch normalization layer. The output of the fully connected layer is fed to a softmax function. The values used for the convolution layer as well as the weights for the fully connected layers are computed through backpropagation (Rumelhart, Hinton and Williams, 1986), which is performed by the deep neural network that computed the error in order to see how much the network can be adjusted and changed.

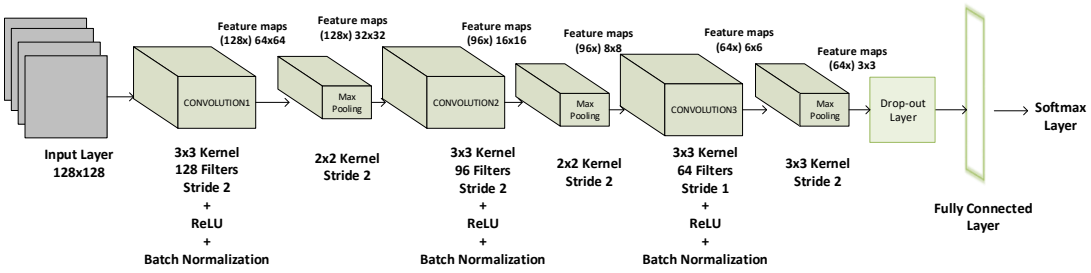


Figure 4-1 CNN Architecture used in Experimental Work

To approximate a task in real forensic investigations, limited numbers of training images are used to train the network to classify the source of an image. Some studies such as those by Baroffio *et al.* (2016) and Tuama, Comby and Chaumont (2016) have used patches to increase the number of images for training. However, the difficulty with patches relates to the location-sensitive pattern in this application. As a result, the noise pattern of an image can change and behave completely differently, which makes it difficult for the CNN to learn its features. Based on the patches of every single image, a majority voting strategy can also be used to classify the number of patches for different cameras. But it will be quite difficult to make a final decision if multiple decisions for different patches in the same image are available.

4.3 EXPERIMENTAL RESULTS

To comparison filtered-based PRNU and deep learning CNN approaches, the proposed CNN model is trained on three different training datasets in which only 50, 100 and 250 training images respectively are used for each camera. Also, these three training datasets are used to extract the reference pattern of a camera for the PRNU-based method to determine the effect on performance. So, the larger the number of images, the more the random noise components and the impact of the scene are suppressed. The two different dimensions of training images are used such as 128×128 and 256×256 images feed into the input layer of the CNN-based model as well as for the PRNU-based method. The test image dataset is consisted of 200 images for each camera. For both methods, the training and test images are in the form of noise residual patterns extracted in the same way as described in Lukáš, Fridrich and Goljan (2006).

To gauge the effectiveness of the different approaches for SCI, the natural images are used from NTIF database which is consisted of ten digital cameras. The one more digital camera such as Samsung L301 is also included in this experiment. Lawgaly and Khelifi (2016) introduced this unique dataset of images captured from number of digital camera devices specifically for the purpose of source camera identification. The digital cameras used for the experiments are listed in Table 4-1 along with dimensions and resolutions. As mentioned, three different training datasets of images are built in order to compare the two approaches for SCI and to judge how it affects performance. Table 4-2 summarizes the NTIF datasets for each camera used in the experiments in the form of training and test images as well as input size blocks for both techniques.

Table 4-1 Characteristics of Eleven Digital Cameras used in Experimental Study

Camera model	**Resolution**	**Type of sensors**
Canon IXUS115HS-1	4000 × 3000	1/2.3", CMOS
Canon IXUS115HS-2	4000 × 3000	1/2.3", CMOS
Fujifilm S2950-1	4288 × 3216	1/2.3", CCD
Fujifilm S2950-2	4288 × 3216	1/2.3", CCD
Nikon Coolpix L330-1	5152 × 3864	1/2.3", CCD
Nikon Coolpix L330-2	5152 × 3864	1/2.3", CCD
Panasonic DMC TZ20-1	4320 × 3240	1/2.33", CMOS
Panasonic DMC TZ20-2	4320 × 3240	1/2.33", CMOS
Samsung p1120-1	4320 × 3240	1/2.33", CCD
Samsung p1120-2	4320 × 3240	1/2.33", CCD
Samsung L301	4000 × 3000	1/2.3", CCD

Table 4-2 Summary of Datasets used in Experimental Study for each Digital Camera

Techniques	**Two sets of input dimensions**	**No. of training images in three sets**	**No. of test images**
PRNU	$128 \times 128/256 \times 256$	50, 100 and 200	200
CNN	$128 \times 128/256 \times 256$	50, 100 and 200	200

For traditional approaches such as machine learning PRNU-based techniques, Lukáš, Fridrich and Goljan (2006) method is applied to estimate the PRNU by using a wavelet-based filter. To accomplish SCI, the estimation of PRNU is calculated by averaging the number of training images taken by the same camera and suppressing the image content through a de-noising filtering process. The noise residual W is obtained by denoising the original image I, as shown in Equation 4.2:

$$W = I - F(I) \tag{4.2}$$

where F is the denoising filter.

And the camera reference pattern K is obtained by averaging the noise residuals W, as shown in Equation 4.3:

$$K = \frac{\sum_{i=0}^{L} W_i}{L} \tag{4.3}$$

where L is the number of training images used in the extraction of the PRNU.

The normalized correlation coefficient between the noise residual test image and the camera reference PRNU, as shown in Equation 2.9, is used to make a decision concerning whether or not the specific image was taken by camera C.

This normalized correlation statistic test (Lukáš, Fridrich and Goljan, 2006) was used to determine the correlation between the noise residual of the test image and the camera reference pattern for filtered-based PRNU in source camera identification. For the CNN-based deep learning approach, noise residual images with two input dimension blocks such as 128×128 and 256×256 are fed into the input layer of the CNN model. To make fair a comparison with the PRNU-based approach, the noise residual images for CNN taken from eleven digital cameras are extracted in the same way as described in Lukáš, Fridrich and Goljan (2006). For CNN experiments, three different training datasets of images were trained one by one on Nvidia GEFORCE GTX (Giga Texel Shader eXtreme) GPU (Graphics Processing Unit).

Many CNN configurations were tested and tuned with different parameters to find the lowest error rate. The best CNN model is then chosen, as illustrated in Section 4.2, for evaluating SCI performance. For the identification of source camera devices, the two well-established performance measures, the false positive rate (FPR) and false negative rate (FNR) are computed in assessing the performance of both approaches. A false positive, which is also referred to as a false alarm, is an incorrect positive prediction which occurs when the image under test belongs to a source camera device that is in fact not the source camera device under investigation. Meanwhile, a false negative is an incorrectly negative prediction which occurs when the image under test is said to belong to a source camera device other than the one under investigation. For instance, the false negative rate (FNR) for camera "A" represents the proportion of test images taken by camera "A" which are, however, incorrectly assigned to a different

camera. On the other hand, the false positive rate (FPR) for camera "A" is the proportion of test images assigned to camera "A" but which were in fact taken by a different camera.

Similarly, the true positive rate (TPR) is defined as the correct prediction of images that belong to the source camera device under investigation, whereas the true negative rate (TNR) is the proportion of images for which correct negative predictions were made.

A total of 200 test images captured by the same sensor were selected to calculate the false negative rate, while 2000 images captured by the other ten cameras (200 images for each) were chosen to compute the false positive rate. Moreover, the overall FPR and FNR are computed by taking the average of FPR and FNR of eleven cameras for both approaches. In addition, the overall accuracy of eleven cameras computed for PRNU and CNN is shown in Equation 4.4.

$$Accuracy\ in\ \% = \frac{TPR + TNR}{2} \tag{4.4}$$

where $TPR = 1 - FNR$ and $TNR = 1 - FPR$.

From the results for FPR, FNR and overall accuracy, this chapter shows that the PRNU-based approach has achieved superior SCI performance as compared to the CNN-based approach for each training dataset used for the identification of the source cameras of test images. The false positive rates and false negative rates of the filtered-based approach for each camera are much lower than those for the CNN-based approach, as depicted in Tables 4-3 and 4-4.

Moreover, the CNN converges on the training images but it is likely that these images do not adequately represent the real world situation because of the low number of images per class (an overfitting problem). The PRNU which is the unique feature for each camera is location-sensitive, therefore, feeding image blocks at different locations into the CNN for the same class is theoretically not sound. The conclusion from these findings is that the CNN could be used in a more efficient way but definitely not in the traditional way, as is well-known in other applications where input training images could be cropped or resized or divided into blocks and fed into the CNN as part of the same class samples.

In addition, a clear improvement is seen with filtered-based PRNU in the form of reduced numbers of false negative and false positive rates for the both input block size of 256×256 block as compared to 128×128 blocks. This is due to the fact that large input block size of images has more signal components for correlation than small input block size. In addition, the FPR and FNR are lowered when the number of training images are increased in the PRNU-based method for both input block sizes of images. This is due to the fact that when there are large numbers of images, more of the scene content of images can be suppressed and, as a result, the PRNU is enhanced.

In the CNN-based approach, the FNR and FPR are much higher for each camera model when compared to other camera models of the same brand of camera. For instance, for the camera brand Fuji model no. S2950-1, the FNRs of image input size block 128×128 with 50, 100 and 200 training images are higher than those of the Fuji S2950-2. For the camera brand Samsung L301, the CNN-based approach shows good performance in FNRs and FPRs as compared to other camera models of different brands. This is due to the fact that the Samsung with one model L301 is used in the system for training. The results indicate that performance of the CNN-based approach declines when different models of the same camera brand are used. The CNN-based approach is less likely to distinguish between images of different models of the same camera brand compared to the PRNU-based technique.

Tables 4-5 and 4-6 show the overall performance for the eleven digital cameras with two different image size blocks using both the PRNU and CNN approaches, where the overall FPR, FNR and average accuracy rate of PRNU are much higher than those of the CNN. These findings indicate that the PRNU-based technique is much more effective for SCI in image forensics applications compared to the CNN-based method when limited numbers of images are available to the analyst. To achieve a higher accuracy for SCI using the CNN, the deep learning technique requires a large number of training images to train the CNN model. However, this is not practical in real investigation situations in the field. This gives an advantage to traditional machine learning approaches such as the PRNU-based technique in identifying a source camera because fewer training images are needed to extract the PRNU. Thus, the results of these experiments show that the PRNU-based technique for SCI has been shown to outperform the CNN-based technique.

Table 4-3 False Negative Rate (%) for Test Images of PRNU-based and CNN-based Approaches

No. of cameras (11)	**Two sets of approaches**	**Input dimensions of training images 128×128**			**Input dimensions of training images 256×256**		
		No. of Training Images			**No. of Training Images**		
		50	***100***	***200***	***50***	***100***	***200***
Canon IXUS115HS-1	PRNU-based	20.0	9.0	5.0	3.0	1.0	0.5
	CNN-based	61.0	39.5	35.0	48.0	54.0	5.0
Canon IXUS115HS-2	PRNU-based	16.5	7.0	3.5	6.5	3.0	2.5
	CNN-based	58.0	60.5	57.5	76.5	43.5	99.5
Fujifilm S2950-1	PRNU-based	21.0	18.5	13.5	11.0	8.0	4.0
	CNN-based	60.5	56.5	52.5	59.0	54.0	22.5
Fujifilm S2950-2	PRNU-based	11.0	9.0	7.5	3.5	2.5	2.5
	CNN-based	39.5	32.5	50.5	39.5	46.5	65.5
Nikon Coolpix L330-1	PRNU-based	57.0	55.0	48.5	43.0	41.0	34.0
	CNN-based	62.0	79.0	60.5	51.5	57.0	100.0
Nikon Coolpix L330-2	PRNU-based	71.5	64.5	60.0	61.0	60.0	51.0
	CNN-based	38.5	35.0	76.5	53.0	56.5	36.5
Panasonic DMC TZ20-1	PRNU-based	4.5	2.0	1.5	0.5	0.5	0.5
	CNN-based	67.5	34.0	30.0	76.0	52.5	80.0
Panasonic DMC TZ20-2	PRNU-based	2.5	1.0	0.0	0.0	0.0	0.0
	CNN-based	39.5	67.5	63.5	33.5	48.5	27.5
Samsung p1120-1	PRNU-based	12.5	9.5	12.5	1.0	2.0	3.5
	CNN-based	75.5	73.5	50.5	48.5	49.0	64.0
Samsung p1120-2	PRNU-based	0.0	1.5	0.0	0.0	0.0	0.0
	CNN-based	35.0	43.5	66.5	47.0	57.5	34.0
Samsung L301	PRNU-based	1.0	0.5	0.5	0.0	0.0	0.0
	CNN-based	17.5	10.5	7.5	12.5	2.5	11.5

Table 4-4 False Positive Rate (%) for Test Images of PRNU-based and CNN-based Approaches

No. of Cameras (11)	**Two sets of approaches**	**Input dimensions of training images 128×128**			**Input dimensions of training images 256×256**		
		No. of Training Images			**No. of Training Images**		
		50	***100***	***200***	***50***	***100***	***200***
Canon IXUS115HS-1	PRNU-based	1.50	1.50	1.10	1.25	1.50	1.10
	CNN-based	5.30	10.0	9.45	6.15	4.65	13.05
Canon IXUS115HS-2	PRNU-based	1.45	1.75	1.60	1.50	1.20	0.85
	CNN-based	4.75	5.95	4.30	3.45	4.90	0.00
Fujifilm S2950-1	PRNU-based	2.20	1.60	1.45	1.45	1.10	0.95
	CNN-based	4.20	3.10	4.70	4.20	3.65	6.80
Fujifilm S2950-2	PRNU-based	2.10	1.65	1.75	0.70	1.00	0.85
	CNN-based	8.75	6.05	4.90	6.55	4.70	2.30
Nikon Coolpix L330-1	PRNU-based	1.45	1.10	0.60	0.80	0.80	0.50
	CNN-based	3.65	1.75	5.00	4.50	5.10	0.00
Nikon Coolpix L330-2	PRNU-based	1.65	1.00	0.90	0.70	0.50	0.20
	CNN-based	2.95	6.30	3.30	2.75	4.50	14.20
Panasonic DMC TZ20-1	PRNU-based	2.70	2.00	1.65	1.85	1.65	1.90
	CNN-based	2.80	7.35	7.95	2.30	4.15	1.50
Panasonic DMC TZ20-2	PRNU-based	3.00	1.70	1.55	1.15	1.20	0.75
	CNN-based	5.75	2.65	2.90	7.20	3.95	7.05
Samsung p1120-1	PRNU-based	2.15	2.15	1.50	1.25	1.30	1.35
	CNN-based	5.85	2.15	7.65	10.65	7.75	4.35
Samsung p1120-2	PRNU-based	1.90	1.85	1.75	1.40	1.05	0.95
	CNN-based	10.45	6.40	3.45	5.15	7.45	5.10
Samsung L301	PRNU-based	1.65	1.40	1.40	0.90	0.5	0.45
	CNN-based	1.00	1.50	1.45	1.60	0.9	0.25

Table 4-5 Overall Performance of PRNU and CNN for Eleven Digital Cameras with Image Size 128×128

	Overall performance (%)								
Technique	***FPR***			***FNR***			***Accuracy***		
	No. of Training Images			***No. of Training Images***			***No. of Training Images***		
	50	***100***	***200***	***50***	***100***	***200***	***50***	***100***	***200***
PRNU-based	1.98	1.61	1.39	19.77	16.14	13.86	89.13	91.13	92.38
CNN-based	5.04	4.84	5.00	50.41	48.36	50.05	72.28	73.40	72.48

Table 4-6 Overall Performance of PRNU and CNN for Eleven Digital Cameras with Image Size 256×256

	Overall performance (%)								
Technique	***FPR***			***FNR***			***Accuracy***		
	No. of Training Images			***No. of Training Images***			***No. of Training Images***		
	50	***100***	***200***	***50***	***100***	***200***	***50***	***100***	***200***
PRNU-based	1.18	1.07	0.90	11.77	10.73	8.95	93.53	94.10	95.08
CNN-based	4.95	4.70	4.96	49.55	47.41	49.64	72.75	73.95	72.70

In addition to this, the CNN model can only identify the images taken by different brands, as it is a brand-specific design for SCI problems. Table 4-7 shows the FNR and FPR computed for the only five digital camera brands with no models used of any camera brand. A clear improvement can be seen in the FPRs and FNRs of the CNN-based approach for five camera brands for different training image datasets as compared to the eleven digital camera models when used to train the network for SCI, as indicated in Table 4-8.

Table 4-7 False Negative Rate (%) and False Positive Rate (%) for 200 Test Images for Five Digital Camera Brand

	False negative rate			**False positive rate**		
Digital camera brand	**Training images of CNN-based method, input size 128 x 128**			**Training images of CNN-based method, input size 128 x 128**		
	No. of Training Images			***No. of Training Images***		
	50	***100***	***200***	***50***	***100***	***200***
Canon IXUS115HS-1	40.1	10.5	6.5	5.38	5.13	6.13
Fujifilm S2950-1	24.44	1.0	0.0	14.25	1.25	2.63
Nikon Coolpix L330-1	52.34	23.0	30.5	2.00	4.25	1.00
Panasonic DMC TZ20-1	36.30	9.0	6.0	7.38	0.50	1.25
Samsung L301	16.71	8.0	7.0	4.75	1.75	1.50

Table 4-8 Overall Performance in (%) of 200 Test images of size 128×128 for Five Digital Cameras and Eleven Digital Camera Brands with Models used for Training

	Overall Performance in %								
Techniques	**FPR**			**FNR**			**Accuracy**		
	No. of Training Images			***No. of Training Images***			***No. of Training Images***		
	50	**100**	**200**	**50**	**100**	**200**	**50**	**100**	**200**
CNN for Five Digital Camera Brands	**6.75**	**2.58**	**2.50**	**27.0**	**10.3**	**10.0**	**73.0**	**89.7**	**90.0**
CNN for Eleven Digital Camera Brands and Models	5.04	4.84	5.00	50.4	48.4	50.1	72.3	73.4	72.5

4.4 CONCLUSION

The findings indicate that the CNN-based approach is not as efficient as the PRNU-based technique for SCI in image forensics as it clearly exhibits lower performance at a higher computational cost. This is due to the fact that, when different models of the same camera brands are used in the training process, the performance of CNN declines. In particular, the CNN-based technique fails to distinguish between images captured by the same models of the same brand of camera. This constitutes a clear drawback in practice, since the forensic analyst normally deals with images from the same model and brand. In addition, the CNN converges on the training images, but due to there being fewer images per class, it is possible that these images fail to accurately represent the real world. Moreover, the low performance of the CNN as compared to the PRNU-based technique is due to the fact that image blocks are fed into the CNN at different locations for the same class, which is theoretically not sound because PRNU is location-sensitive.

CHAPTER 5 TEMPORAL IMAGE FORENSIC ANALYSIS FOR PICTURE DATING WITH DEEP LEARNING

5.1 MOTIVATION

Estimating the acquisition date of digital photographs is crucial in image forensics. In this context, it is seen that PRNU has abundant high frequency components so that it has been widely used in SCI applications. In addition, PRNU is a location-sensitive noise pattern, which therefore cannot learn the temporal changes in the content of digital pictures. Motivated by this, the task of dating images by processing their contents should be reasonably accurate for use in court to resolve high profile and complex cases. The goal of temporal digital image forensic analysis is to discover the links in time between two or more pieces of evidence. Therefore, the problem of picture dating is addressed here from a machine learning perspective by adopting a deep learning approach for the first time in temporal image forensics. This has been demonstrated on a new database called the Northumbria Temporal Image Forensics (NTIF) database which has been made publicly available for researchers in image forensics. NTIF is the first public database that includes a large number of images at different time slots captured on a regular basis using 10 different digital cameras. This will serve the research community as a solid ground for research on picture dating and other image forensics applications.

5.2 INTRODUCTION

As manipulation, image forgery, downloading, and sharing digital material are widely increasing challenges over the years, the timing information associated with digital images is very susceptible to alteration or modification. If such images are presented in court or used for criminal investigation purposes, it is crucial that every measure be taken to ensure the reliability and accuracy of picture dating. Reconstructing the history of evidence in images concerning crime in order to verify the correct timeline of the events which have occurred is seen as an essential task in digital forensic analysis (Casey, 2009). To correlate the digital artifacts, the forensic analyst needs to identify the correct ordering and timeline of any relevant images so as to understand the sequence of events. It is also necessary to correlate such evidence on the basis of past, current or future events. Every digital image can be evidence of at least one event. Therefore, the forensic investigator develops or exploits many hypotheses to answer questions about the timeline of events. Such questions may include “Did the event occur before this

date or after this date?" or "When did the event occur?" and "how events occurred before, during or after the incident". To answer these questions, it is necessary to develop a correct timeline of the images concerned. Helpful clues might be found in the file header (EXIF), but this information can be easily altered and therefore cannot be deemed reliable for forensics purposes. Besides, retrieving genuine timestamp information is also not possible due to the ease of manipulations of the time settings of a digital device. Therefore, forensic analysts need to test hypotheses about the order of events in the timeline on the basis of the images provided. Establishing a timeline as part of a digital forensic investigation is a significant challenge and a critical part of understanding the events which have occurred.

For this purpose, temporal image forensics leverages a temporal analysis of images by creating the correct ordering of their timeline, where the forensic investigator estimates the acquisition time of suspect images using prior information or knowledge such as the camera device used or maybe a set of images captured by the same devices in different time slots. In this regard, temporal forensic analysis is preferred to place the digital images relative to temporally ordered cluster classes. This chapter contributes to the body of knowledge in this area, where the acquisition timeline of digital pictures is estimated in supporting temporal image forensic analysis with the help of deep learning. However, despite temporal information being crucial in a digital forensics investigation, there is a lack of a suitable image database for the digital image forensics community to exploit. Therefore, the Northumbria Temporal Image Forensics (NTIF) database provides a valuable means for temporal image forensics applications as it contains structured images captured in a regular series over time.

5.3 PICTURE DATING

In the literature, the problem of picture dating has been tackled in different ways from the perspectives of both machine and deep learning. Some studies are available that have predicted the dating of images based solely on the visual contents of and objects in images. For example, Palermo, Hays and Efros (2012) and Fernando *et al.* (2014) predicted the age of historical colour images and Salem *et al.* (2016) estimated the date of pictures based on the appearance of people in the scene. They used convolutional neural networks (CNNs) to explore the relationship between human appearance and time. They concluded that visual contents such as clothing, hairstyles and glasses are useful features to discriminate the timeline of images. A study by Müller, Springstein and Ewerth (2017) attempted to estimate the

acquisition year of photos captured in the period from 1930 to 1999 using a deep learning approach. However, it is believed that the results shown were biased because the quality of most the pictures used was clearly related to the time at which they were acquired (and many of the pictures were originally in a non- digitised form). Moreover, many pictures represented historical events that were directly linked to the timeframes that the authors were looking to predict. Martin, Doucet and Jurie (2014) treated picture dating as a binary task, in deciding whether an image is older or newer than a reference image. Thus, it can be observed that, in the majority of research in the literature so far as discussed in section 2.5, picture dating has been seen as a task based on the appearance of objects in the images. In contrast to previous work, the picture dating task here is not considered to be limited to classifying the timeline of pictures on the basis of objects or scenes. This chapter specifically relates to digital image forensic analysis where limited numbers of images are available to the analyst trying to estimate the date of any suspect images. For the first time in temporal digital image forensics, this chapter addresses the problem of picture dating of natural images as a classification task from the perspective of deep learning using the NTIF database (Ahmed *et al.*, 2020b).

5.4 PROPOSED CNN-BASED SYSTEM FOR PICTURE DATING

The problem of dating natural colour images is approached from a multiclass classification perspective with the help of deep learning using existing CNN models. Firstly, a number of images captured at different times are divided into a number of classes. Each class represents a specific timeline of a group of images. The number of colour images are then divided into two sets called the training and testing sets. The two different stages shown in Figure 5-1 illustrate the proposed CNN-based system for picture dating. In the first stage, the training samples are divided into a number of non-overlapping colour patches. The adopted CNN model is then loaded, the final layer of CNN is replaced according to training dataset for classification. Then, the training patches are directly fed into the adopted CNN model to train the whole network. In the second stage, the test images are divided into a number of non-overlapping patches to feed into the trained network for the prediction of N decisions for the number of patches. The next stage adopts the majority voting method to reach a final decision using multiple patches from the test image.

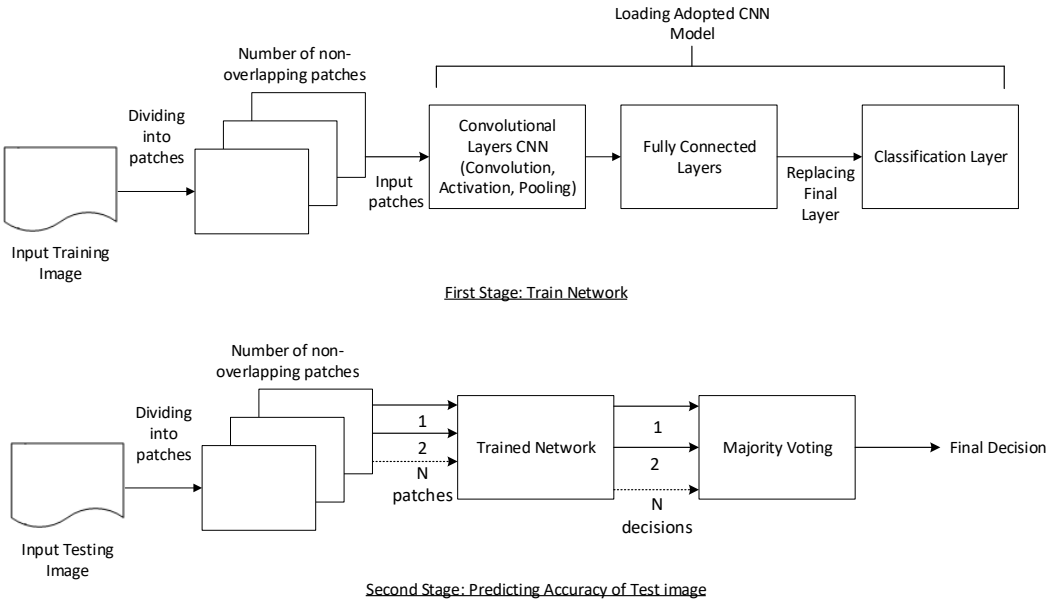


Figure 5-1 The proposed CNN-based System for picture dating

5.4.1 Basic Architecture of Convolutional Neural Network (CNN) and its Operation

In the CNN, data in the form of input images are fed into the network input layer. It usually extracts low-level features such as edges, shapes, and colour contrasts, whereas deeper layers identify more complex visual patterns. The basic architecture of the CNN is shown in Figure 5-2 where the convolution is the major operation that convolves around the image by using several kernels acting as a filter. Every output of the convolution operation is a feature-map.

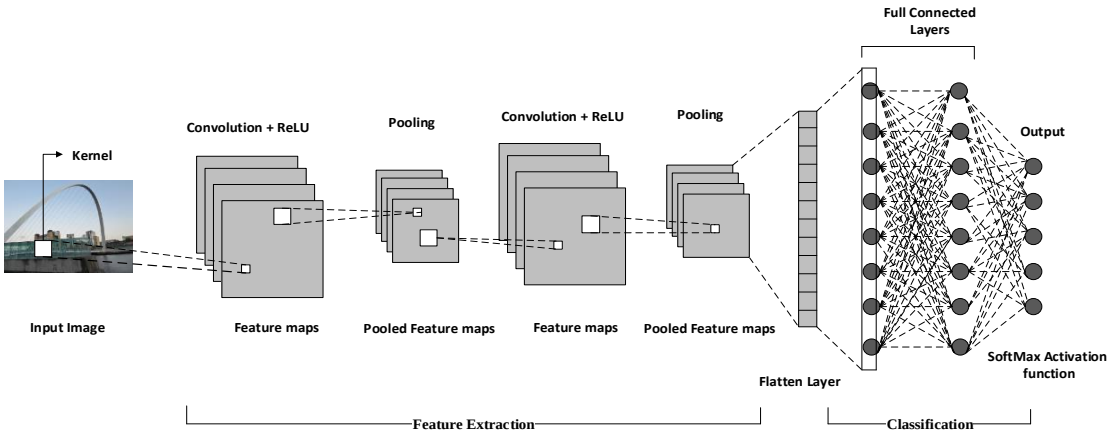


Figure 5-2 Basic architecture of the CNN

After the computation of feature maps, an activation function layer performs feature extraction and then the output is further enhanced through a pooling layer. An activation function in the CNN is used to provide non-linearity in the data. The most commonly used activation function in deep learning is the rectified linear unit (ReLU) (Nair and Hinton, 2010). The ReLU layer performs more rapid training by retaining positive values and mapping negative activations to zero. After the ReLU layer, the pooling operation further simplifies the number of parameters for network learning by computing non-linear down-sampling. The three main operations, convolution, the activation function, and the pooling operation recur in a different number of hidden layers to detect various features. Once, the pooled featured map is obtained, flatten layer transforms the entire pooled feature map into a single column feature vector to feed into the next layer for processing. The final output layer is based on the classification layer from which the fully connected layer produces a vector of k dimensions where k is the number of classes that the deep-learning CNN network can predict. A back-propagation algorithm is used to train the CNN. Most commonly, four kinds of optimizers are used in deep learning: SGD (Stochastic Gradient Descent), RMS prop, Adam, and AdaDelta. To prevent overfitting, a drop-out layer is also used in the CNN as a regularization technique (Srivastava *et al.*, 2014). Thus, tuning different parameters in a deep neural network, such as the block size of the image, number of convolution layers, activation function, pooling layers, batch size, number of epochs, and different kinds of optimizers, helps in the achievement of higher accuracy.

5.4.2 Deep Learning Models adopted for Picture Dating

The present interest is to be able to estimate or place unknown images in different time slots using prior knowledge of the camera device used where the images are taken at different time slots. The pixel behaviour of images is identified over time using a deep learning CNN model which is specifically designed to read visual patterns directly from pixel images. Two well-known and award-winning object recognition CNN architectures such as AlexNet (Krizhevsky, Sutskever and Hinton, 2012) and GoogLeNet (Szegedy *et al.*, 2015) have been shown to achieve remarkable performance in forensic tasks involving camera model identification problems (Tuama, Comby and Chaumont, 2016; Athanasiadou, Geradts and Van Eijk, 2018). Therefore, these two CNN models are assessed along with

some fine-tuning approaches by using two different modes: pre-trained model and feature extraction using pre-trained CNN model.

In pre-trained CNN mode, the training images are trained on the existing CNN models using the deep learning softmax classification layer. Feature extraction using AlexNet CNN network works in such a way that learned image pixel features are trained on the pre-trained architecture of the AlexNet network. The trained features are then fed into the other machine-learning multiclass classifiers to predict the timeline of the images. In the classification process, the most frequently used classifiers are the SVM, k-nearest neighbour (KNN), Random Forest (RF) and naive Bayesian (NB). Combinations of the CNN network with a traditional classifier such as SVM have been used in previous studies (Basly et al., 2020; Shi et al., 2019; Bondi et al., 2016; Huang et al., 2018; Al Banna et al., 2019). For feature extraction using the AlexNet CNN network, only two machine learning multiclass classifiers are used in evaluating the accuracy of identification in this chapter, which are the SVM and KNN. The outputs of the trained SVM and trained KNN were used for the final classification.

In both modes, the true positive rate (TPR) is computed as a performance measure for each class, in which only correctly classified images are determined. The final accuracy of every CNN model for the digital camera is also measured based on the total percentage of images correctly classified. The final accuracy for each digital camera is computed as shown in Equation 5.1:

$$TPR = \frac{M}{T} \times 100 \tag{5.1}$$

where M is the total number of images correctly classified per digital camera and T is the total number of images per digital camera.

5.5 PROPOSED NTIF DATABASE FOR TEMPORAL IMAGE FORENSICS

As described in chapter 3, NTIF is the first database that consists of a large number of natural scene images at different time slots taken on a regular basis using 10 different digital cameras. For each digital camera, there are approximately 1 to 71 time slots. In NTIF database, the number of images is grouped in each weekly and biweekly time slot. The total timespan of images for each digital camera is

approximately within the range of 94 weeks from 21st October 2014 to 30th December 2016. This huge database contains 41684 natural colour images covering a broad range of indoor and outdoor scenes along with 980 blue-sky images captured from different digital cameras which have been made publicly available for the research community in digital image forensics.

This database can be widely used for many digital image forensic applications such as image authentication, brand and model classification, source camera identification (SCI), forgery detection and picture dating. The characteristics of the NTIF database include a number of parameters which can be of benefit to digital image forensic analysis, such as different image contents and formats, different types of sensors and sizes, different exposure parameters (including a vast range of ISO settings, shutter speed and focal length), environmental conditions and different brands and models. The NTIF database has been used for SCI purposes in previous research (Lawgaly and Khelifi, 2016; Al-Ani and Khelifi, 2016; Al-Ani *et al.*, 2015; Lawgaly, Khelifi and Bouridane, 2014; Ahmed *et al.*, 2019), and it has proved highly suitable for employment in SCI since multiple images are included which were taken by different individual cameras of the same make and model. The main advantage of the NTIF database to be used for picture dating purposes is that the forensic analyst can accomplish a temporal analysis by studying the pixel behaviour of a set of images over time, since the images were captured as a sequential series over time. Using this database, picture dating is possible where researchers can estimate the time of acquisition of images taken by the same camera device.

5.6 EXPERIMENTS AND RESULTS

This section is divided into two subsections. The first describes the database and set-up used in the experiment, and the second describes the experiment and presents the results for the estimation of acquisition time of images captured at different time slots.

5.6.1 Experimental Dataset and Set-up

The ten digital cameras used in this experiment are listed in Table 5-1. For each camera, the images from the NTIF database collected in the first 25 time slots are categorized into 5 classes. The number of classes

represents the timeline of the images. This full timeline is grouped into consecutive temporal classes where each class is consisted of 5 timeslots.

Table 5-1 Characteristics of Ten Digital Cameras

Name of digital camera	**Size of resolution**	**Type of sensors**
Canon IXUS115HS-1	4000 × 3000	CMOS
Fujifilm S2950-1	4288 × 3216	CCD
Nikon Coolpix L330-1	5152 × 3864	CCD
Panasonic DMC TZ20-1	4320 × 3240	CMOS
Samsung pl120-1	4320 × 3240	CCD
Canon IXUS115HS-2	4000 × 3000	CMOS
Fujifilm S2950-2	4288 × 3216	CCD
Nikon Coolpix L330-2	5152 × 3864	CCD
Panasonic DMC TZ20-2	4320 × 3240	CMOS
Samsung pl120-2	4320 × 3240	CCD

For example, class 1 (C1) represents the timeline of images taken during the first five acquisition dates range from timeslots 1 to 5 (w1-w5), class 2 (C2) represents the timeline of images taken during the other next 5 time slots range from timeslots 6 to 10 (w6-w10) and so on. The dataset for each class is divided into two sets of images called the training and testing datasets for each camera. The description of numbers of classes, the dates of images, numbers of images in both training and testing datasets and the timespan of each class are shown in Table 5-2.

To make the dataset suitable for the existing CNN model conditions and to increase the number of training and testing images, each image is cropped into a number of patch sizes with a non-overlapping block patch size of 500×500. Hence, a total of $N = 48$ patches were extracted from each image. Thus, a total of 7200 patches were extracted from 150 training images and a total of 4800 patches were extracted from 100 testing images per class for each camera.

Table 5-2 Summary of Datasets used in Experimental Study for each Digital Camera

No. of classes	Timespan/date	No. of training images	No. of testing images
C1(w1_w5)	21-10-2014 to 26-11-2014	150	100
C2(w6_w10)	27-11-2014 to 04-01-2015	150	100
C3(w11_15)	05-01-2015 to 12-02-2015	150	100
C4(w16_20)	13-02-2015 to 19-03-2015	150	100
C5(w21_25)	30-03-2015 to 07-05-2015	150	100
Total		**750**	**500**

Finally, the larger dataset is consisted of 36000 non-overlapping training image patches and 24000 non-overlapping testing image patches from 5 classes for each digital camera. To classify the number of patches that belong to a particular image, the decision is made by a majority voting process. Once the majority voting process has been computed for each image, the final performance is evaluated in terms of the TPR for all CNN models.

5.6.2 Experimental Results and Discussion

Number of eexperiments were conducted for each digital camera to classify the acquisition time of images captured at different time slots with four different approaches: pre-trained model using GoogLeNet, pre-trained model using AlexNet, feature extraction using AlexNet with KNN and feature extraction using AlexNet with SVM multiclass classifiers. Several tuning configurations of different parameters were tested on the existing CNN model to achieve a minimal error rate. The colour image patches from both training and testing datasets were augmented according to the input size of the CNN model. The 'Adam' optimizer was used. The learning rate is defined as 0.0001 in the training options.

The results for the ten digital cameras are depicted in Tables 5-3 to 5-12, which show that all four approaches using existing CNN models are best at dating the overall 500 test images in their correct timeline of classes with a high maximum accuracy of 80% to 88%. In addition, the best overall accuracy for nine of the digital cameras is achieved with the pre-trained model using the AlexNet CNN model, with performance ranging from 82% to 88%. Moreover, the overall accuracy of feature extraction using

AlexNet with the k-NN classifier for all digital cameras is substantially higher as compared to feature extraction using AlexNet with the SVM classifier. It is worth noting that the test images of class 5 dating from 30th March 2015 to 7th May 2015 are accurately predicted with a higher rate of accuracy from 90% to 100% for all digital cameras as compared to other classes. Interestingly, while images are divided into non-overlapping blocks to increase the training samples, the results show that the temporal changes in image content, as modelled by CNNs, are not dependent on block location.

Table 5-3 TPR per class and overall Accuracy of Acquisition Time Estimation of 100 Test Images of Digital Camera Canon IXUS115HS-1

No. of Classes	**Transfer Learning using GoogLeNet (%)**	**Transfer Learning using AlexNet (%)**	**Feature Extraction using AlexNet with k-NN classifier (%)**	**Feature Extraction using AlexNet with SVM classifier (%)**
C1(w1_w5)	94	87	85	78
C2(w6_w10)	68	64	69	66
C3(w11_15)	89	84	86	83
C4(w16_20)	88	84	85	77
C5(w21_25)	90	98	94	99
Overall Accuracy	**85.8**	83.4	83.8	80.6

Table 5-4 TPR per class and overall Accuracy of Acquisition Time Estimation of 100 Test Images of Digital Camera Fujifilm S2950-1

No. of Classes	**Transfer Learning using GoogLeNet (%)**	**Transfer Learning using AlexNet (%)**	**Feature Extraction using AlexNet with k-NN classifier (%)**	**Feature Extraction using AlexNet with SVM classifier (%)**
C1(w1_w5)	97	95	87	86
C2(w6_w10)	70	71	73	76
C3(w11_15)	74	80	81	75
C4(w16_20)	70	87	80	80
C5(w21_25)	99	97	92	94
Overall Accuracy	82.0	**86.0**	82.6	82.2

Table 5-5 TPR per class and overall Accuracy of Acquisition Time Estimation of 100 Test Images of Digital Camera Nikon Coolpix L330-1

No. of Classes	**Transfer Learning using GoogLeNet (%)**	**Transfer Learning using AlexNet (%)**	**Feature Extraction using AlexNet with k-NN classifier (%)**	**Feature Extraction using AlexNet with SVM classifier (%)**
C1(w1_w5)	86	85	91	73
C2(w6_w10)	78	73	69	64
C3(w11_15)	79	86	84	82
C4(w16_20)	78	90	89	88
C5(w21_25)	97	98	96	95
Overall Accuracy	83.6	**86.4**	85.8	80.4

Table 5-6 TPR per class and overall Accuracy of Acquisition Time Estimation of 100 Test Images of Digital Camera Panasonic DMC TZ20-1

No. of Classes	**Transfer Learning using GoogLeNet (%)**	**Transfer Learning using AlexNet (%)**	**Feature Extraction using AlexNet with k-NN classifier (%)**	**Feature Extraction using AlexNet with SVM classifier (%)**
C1(w1_w5)	83	93	85	84
C2(w6_w10)	57	77	65	68
C3(w11_15)	86	79	88	86
C4(w16_20)	85	94	91	90
C5(w21_25)	100	96	98	97
Overall Accuracy	82.2	**87.8**	85.4	85.0

Table 5-7 TPR per class and overall Accuracy of Acquisition Time Estimation of 100 Test Images of Digital Camera Samsung pl120-1

No. of Classes	**Transfer Learning using GoogLeNet (%)**	**Transfer Learning using AlexNet (%)**	**Feature Extraction using AlexNet with k-NN classifier (%)**	**Feature Extraction using AlexNet with SVM classifier (%)**
C1(w1_w5)	90	82	85	80
C2(w6_w10)	86	93	84	89
C3(w11_15)	77	80	79	75
C4(w16_20)	79	89	89	89
C5(w21_25)	97	96	95	96
Overall Accuracy	85.8	**88.0**	86.4	85.8

Table 5-8 TPR per class and overall Accuracy of Acquisition Time Estimation of 100 Test Images of Digital Camera Canon IXUS115HS-2

No. of Classes	**Transfer Learning using GoogLeNet (%)**	**Transfer Learning using AlexNet (%)**	**Feature Extraction using AlexNet with k-NN classifier (%)**	**Feature Extraction using AlexNet with SVM classifier (%)**
C1(w1_w5)	90	89	77	76
C2(w6_w10)	62	64	70	68
C3(w11_15)	86	92	86	82
C4(w16_20)	80	80	85	78
C5(w21_25)	97	97	98	97
Overall Accuracy	83.0	**84.4**	83.2	80.2

Table 5-9 TPR per class and overall Accuracy of Acquisition Time Estimation of 100 Test Images of Digital Camera Fujifilm S2950-2

No. of Classes	**Transfer Learning using GoogLeNet (%)**	**Transfer Learning using AlexNet (%)**	**Feature Extraction using AlexNet with k-NN classifier (%)**	**Feature Extraction using AlexNet with SVM classifier (%)**
C1(w1_w5)	89	96	91	88
C2(w6_w10)	67	66	57	64
C3(w11_15)	79	84	86	74
C4(w16_20)	81	87	82	84
C5(w21_25)	93	95	96	90
Overall Accuracy	81.8	**82.8**	82.4	80.0

Table 5-10 TPR per class and overall Accuracy of Acquisition Time Estimation of 100 Test Images of Digital Camera Nikon Coolpix L330-2

No. of Classes	**Transfer Learning using GoogLeNet (%)**	**Transfer Learning using AlexNet (%)**	**Feature Extraction using AlexNet with k-NN classifier (%)**	**Feature Extraction using AlexNet with SVM classifier (%)**
C1(w1_w5)	83	93	85	79
C2(w6_w10)	70	76	78	68
C3(w11_15)	83	81	77	76
C4(w16_20)	73	84	75	87
C5(w21_25)	96	93	95	94
Overall Accuracy	81.0	**85.4**	82.0	80.8

Table 5-11 TPR per class and overall Accuracy of Acquisition Time Estimation of 100 Test Images of Digital Camera Panasonic DMC TZ20-2

No. of Classes	**Transfer Learning using GoogLeNet (%)**	**Transfer Learning using AlexNet (%)**	**Feature Extraction using AlexNet with k-NN classifier (%)**	**Feature Extraction using AlexNet with SVM classifier (%)**
C1(w1_w5)	84	93	80	83
C2(w6_w10)	66	73	71	65
C3(w11_15)	83	74	86	81
C4(w16_20)	81	77	78	78
C5(w21_25)	95	99	98	94
Overall Accuracy	81.8	**83.2**	82.6	80.2

Table 5-12 TPR per class and overall Accuracy of Acquisition Time Estimation of 100 Test Images of Digital Camera Samsung pl120-2

No. of Classes	**Transfer Learning using GoogLeNet (%)**	**Transfer Learning using AlexNet (%)**	**Feature Extraction using AlexNet with k-NN classifier (%)**	**Feature Extraction using AlexNet with SVM classifier (%)**
C1(w1_w5)	86	89	87	85
C2(w6_w10)	87	93	88	88
C3(w11_15)	81	81	75	69
C4(w16_20)	83	83	89	75
C5(w21_25)	94	90	94	93
Overall Accuracy	86.2	**87.2**	86.6	82.0

5.7 CONCLUSION

In this chapter, the problem of acquisition time estimation is formulated using a deep learning CNN-based approach for digital images. For this reason, different experiments have been conducted using a proposed comprehensive database for temporal image forensics named the NTIF. This includes two existing CNN models used to estimate the timeline of digital pictures in a multiclass classification approach. The results have shown that the best overall accuracy of estimating the dating of pictures is reached using AlexNet CNN with pre-trained mode in which more than 82% of test images are correctly classified in their respective timeframe of classes. Overall, the estimation performance of the acquisition time of digital pictures is approximately 80% to 88% accuracy with all CNN models.

CHAPTER 6 A PROPOSED MACHINE LEARNING-BASED APPROACH FOR PICTURE ACQUISITION TIMESLOT PREDICTION USING DEFECTIVE PIXELS

6.1 INTRODUCTION

Estimating the acquisition time of digital photographs is a challenging task in temporal image forensics, but the application is highly demanded for establishing temporal order among individual pieces of evidence and deduce the causal relationship of events in a court case. The forensic investigator needs to identify the timeline of events and look for some patterns to gain a clear overview of activities associated with a crime. This chapter aims to explore the presence of defective pixels over time for estimating the acquisition date of digital pictures.

6.2 DETECTION OF DEFECTIVE PIXELS

The digital image carries two types of fingerprint about the sensor. The Photo Response Non-Uniformity (PRNU) which is intrinsic that has been used in source camera identification and image authentication (Lukáš, Fridrich and Goljan, 2006; Lawgaly, Khelifi and Bouridane, 2014; Al-Ani et al., 2015; Lawgaly and Khelifi, 2016; Gupta and Tiwari, 2018), and defective pixels which were used for image quality enhancement (El-Yamany, 2017; Dong et al., 2019). Like other microelectronic devices, digital imaging sensor develops defects over their lifetime and degrade with time gradually. Therefore, detection of defective pixels is still challenging.

As described in chapter 2, an interesting characteristic of defective pixels is that they accumulate over time and space on the sensor independently of each other. Thus, defective pixels once occurred in an imaging sensor, they continuously increased over time (Leung *et al.*, 2009). Chapman *et al.* (2018) highlighted that smart cell phones are more prone to pixel defects due to smaller size of pixels. They indicated that defective pixel accumulation is highly dependent on pixel size. In traditional digital cameras, the pixel size is about 4 to 7.5 μm but in cell phones, the pixels are 1.34 μm wide (Chapman *et al.*, 2018) which drive more defective pixels. So such defects are likely to be caused by cosmic ray damage and hence any shielding or fabrication design changes cannot prevent them from developing over time (Chapman *et al.*, 2019). Also, results shown in (Chapman *et al.*, 2019) on cell phones indicate

that defect density increases drastically when the pixel size falls below 2 microns. Such defective pixels become a permanent part of the sensor which do not heal itself (Fridrich and Goljan, 2011). Normally, it is expected for the digital camera sensor to show defects in a period of two months, approximately (Fridrich and Goljan, 2011). From this perspective, establishing temporal relationships between digital pictures taken by the same camera for dating purposes is theoretically possible when detecting the defective pixels on a temporal series of images which are spanned over a larger timespan. It's worth mentioning that the previously reviewed works (see chapter 2) on defective pixel detection were proposed for image quality enhancement unlike the current application which consists of using these potential pixels to estimate the acquisition date of images.

6.3 PICTURE DATING

As described in chapter 2 section 2.5, the research works for picture dating provide very impressive results however, the methods need the presence of particular visual elements in a scene and geographic location as a cues to reliably predict the time of images. Thus, previous studies estimate the dating of images based on the visual content of the images to identify the acquisition date of photographs. Moreover, many images were representing some historical events that directly linked the timeframes that the researchers were looking to envisage.

A very little research studies such as Fridrich and Goljan (2011) and Mao *et al.* (2009) have been devoted to the extraction of temporal localisation information from digital pictures. For temporal image forensics, the research study in Fridrich and Goljan (2011) uses natural images to identify the estimation of acquisition time of digital images. This is the only paper that uses defective pixels for picture dating in the context of temporal forensic analysis. Taking advantage of defective pixels accumulation over time, Fridrich and Goljan (2011) exploited defective pixels to establish a temporal relationship between digital pictures taken by the same camera for dating purposes. They used maximum likelihood technique to estimate the defective pixel parameters and detect the onset time to estimate the acquisition time of digital pictures. They were unable to perform accurately between two consecutive defect onsets because the likelihood is constant. That is, the technique returns a constant acquisition time, which is the average time of the two onsets, for any given picture. So, the results were not so promising. The other research framework was presented by Mao *et al.* (2009) where the time-dependent camera

parameters were assumed to be included in the PRNU. The results obtained through the correlation coefficients of PRNU estimates of different time-based image clusters showed the possibility to rank the clusters according to their acquisition date. The reported results have clearly shown that the correlation coefficient is not linearly dependent on the time span between the acquisition date of the picture under investigation and the date corresponding to the closest image cluster. Therefore, it would be theoretically incorrect to linearly map the correlation coefficient into a time span for dating purposes. Furthermore, the use of the correlation coefficient on the PRNU noise does not take the type of pixels into account and ignores their temporal evolution. That is, the correlation operation treats all the pixels equally without any prior knowledge on their temporal behaviour. Also, the dataset of images used in their studies are not in the sequential temporal order. Therefore, the researcher's work has not yet reached to accuracy results according to dating purposes.

6.4 PROPOSED APPROACH

It is worth highlighting that our proposed approach is based on the detection of defective pixels and it is not relying on the PRNU estimation. In fact, the proposed system doesn't use the PRNU in any of its stages. The principle of our technique is that the imaging sensor ages and hence pixel defects start to occur over time and this feature can be used to estimate the acquisition date of digital pictures taken in different time periods. It's believed that sensor aging is a phenomenon that has been noticed even with the latest imaging technology (Chapman *et al.*, 2019).

The problem of estimating the acquisition date of digital pictures is formulated here as a classification problem where the system aims at predicting the right timeslot in which a query image was taken. Using a training and a validation subset, this work addresses the problem by training classifiers using pixel neighbourhood and two proposed local variation (LV) features to find best candidates of defective pixels for predicting the acquisition timeslots of digital pictures. It is based on the fact that the same pixel location for a series of images will be flagged as potentially defective if there is no change in the sensitivity of the pixel over time. With the help of multi-class classification, several training images taken at different times are grouped into a number of classes where each class corresponds to a timeslot, and the query image is assigned to one of them.

The high-level overview of the proposed system as presented in Figure 6-1 is divided into three stages in order to estimate the acquisition time of digital images. The first stage, as shown in Figure 6-1a, aims at extracting a number of potential defective pixel locations from every image block and constructing reliable features using the pixel neighborhood in addition to the proposed local variation features. Such location-based features will be used to train classifiers where each classifier corresponds to a specific pixel location. Once the classifiers are trained for M actual timeslots, the subset of validation images are used to rank the performance of such individual trained classifiers and consequently identify the K best locations of potential pixel defects. This is to filter out the low-performance classifiers because they can negatively affect the overall performance if they are included at the final stage. In the second stage, as shown in Figure 6-1b (i), the system is retrained only for the best selected locations of pixel defects which show higher performance as explained earlier but for virtual sub-classes. Indeed, every actual timeslot is halved to create two virtual timeslots. This allows the system to get trained with a more challenging problem (i.e., more classes with shorter timeslots) so that when the scores get fused and reconstructed for the actual classes, performance can be boosted. As shown in Figure 6-1b (ii), a virtual time slot is predicted for every individual block using the scores in the form of virtual class labels obtained from individually trained pixel defect locations. The predicted virtual timeslots are then considered in the reconstruction step to determine the predicted actual timeslot for a query image and fused in a majority voting method. Finally, Figure 6-1c represents the last stage of the system which consists of the combination of block-based prediction scores in a second majority voting operation to boost the performance.

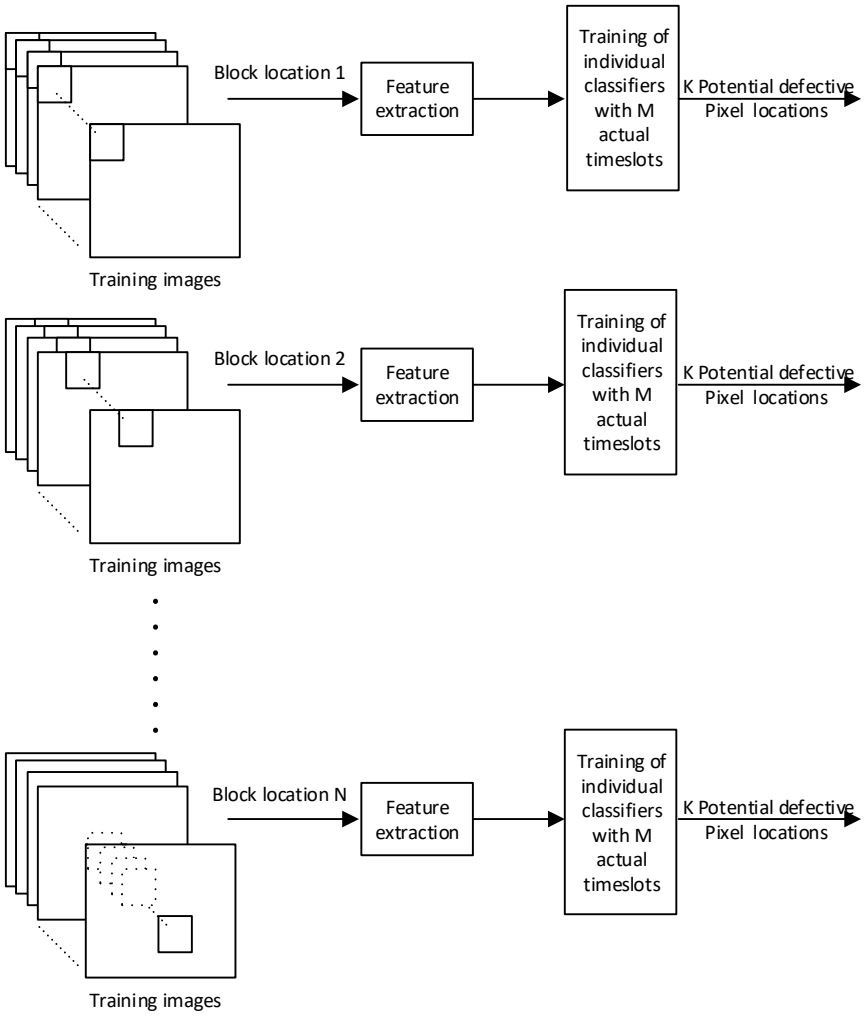


Figure 6-1 a First Stage of the Proposed System

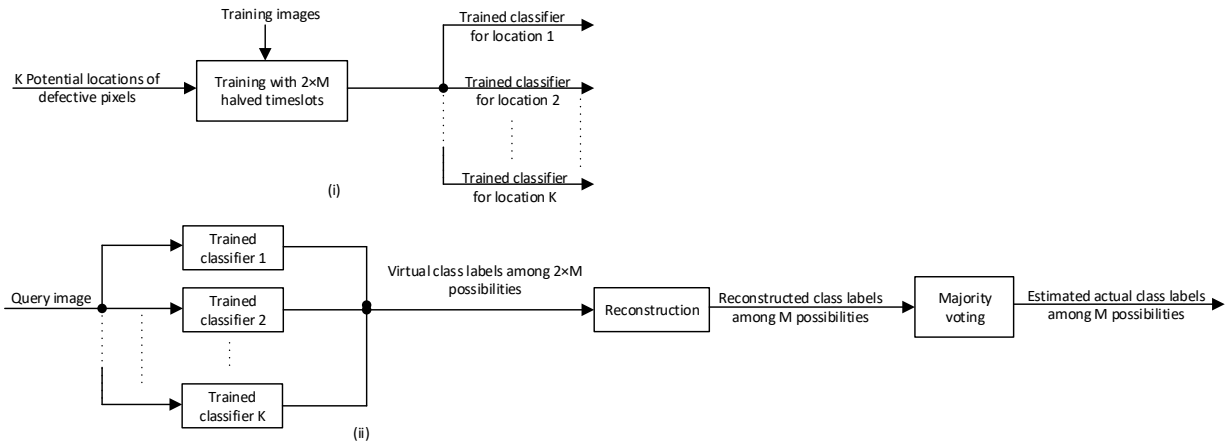


Figure 6-1 b Second Stage of the Proposed System

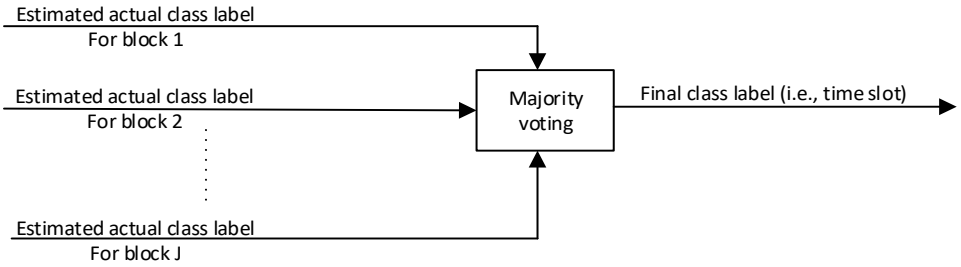


Figure 6-1 c Third Stage of the Proposed System

Figure 6- 1 High-Level of the Proposed System

The further description of the proposed system is detailed below in subsections 6.4.1 to 6.4.5.

6.4.1 Feature Extraction Process using Pixel Neighborhood and Local Variation (LV) features

Here the first aim is to extract the number of pixel locations used as a reliable features for acquisition time estimation. To this end, the behaviour of pixels over time is analyzed to identify the reasonable candidates of the defective pixels. Here, the detection of potential defective pixels based on local neighbourhood is adopted because colour interpolation involves neighbouring pixels and hence the effect of a pixel defect could be spread to its neighbours. The rationale behind estimating potential defective pixels using neighbourhood is that defective pixels exhibit specific and constant behaviour through images over time, unlike other ordinary pixels. Defective pixels are constant, and they did not depend on any external conditions.

The full resolution image is first divided into non-overlapping image blocks of size $W \times H$. Then, the neighbourhood of each pixel is considered in a window of size $w \times w$. Within a moving window of size $w \times w$, x_c is the centre pixel location value which is under examination. The concatenation of pixel intensities extracted for each color channel (Red, Green and Blue) within the window of size $w \times w$ represent the feature vector for a centre pixel location x_c.

Here, the centre pixel and its neighbourhood pixel values play a significant role to get the pixel location classified as a defective pixel for our proposed algorithm. In addition, we propose two attributes as the first order local variation LV1 and second order local variation LV2 for every centre pixel location and

each color channel. The local variation features for an image block are computed for the red, green and blue channel pixel values within a $w \times w$ window. This window then moves one pixel at a time over the entire image block.

The statistical features of an image such as local variation of a centre pixel within a $w \times w$ window is computed because local regions of an image exhibit more color variations than usual, and can lead to high interpolation errors that could be interpreted as being caused by a defective pixel. The centre pixel value at location (i, j) is denoted by x_c, then the local variation features in a $w \times w$ window is given as shown in Equation 6.1 and Equation 6.2:

$$LV1 = abs(x_c - \text{avg}) \tag{6.1}$$

where $avg = \frac{\sum_{i=1}^{w}\sum_{j=1}^{w} x_{i,j} - x_c}{w \times w - 1}$

$$LV2 = sqrt\left(\frac{\sum_{i=1}^{w}\sum_{j=1}^{w}(x_{i,j} - x_c)^2}{w \times w - 1}\right) \tag{6.2}$$

6.4.2 First training process of Individual Classifiers in each Image Block

Once the features are extracted at each centre pixel location in a $W \times H$ block, a classifier is trained with respect to each centre pixel location using training sample images. To detect the best candidates of defective pixel locations, a multi-class classifier is trained, and its performance is determined on a validation subset of images. It is worth noting here that each class label corresponds to a specific timeframe. Figure 6-2 is showing the extraction of R, G and B features with concatenation of local variation measures using a number of training images subset for one centre pixel location.

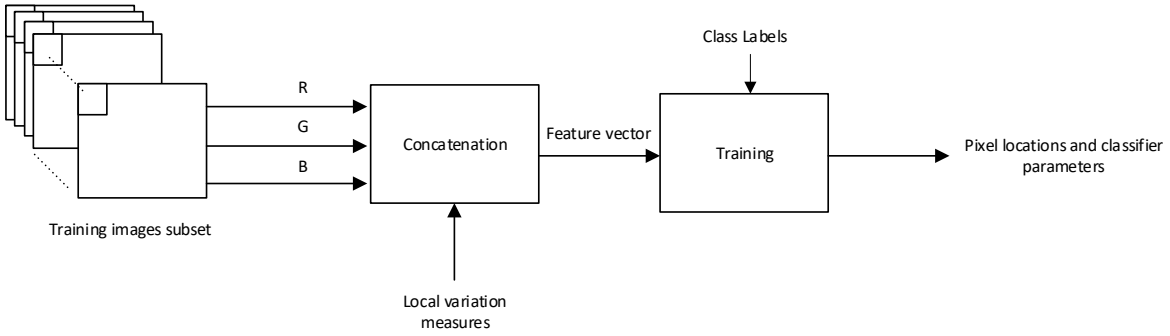


Figure 6- 2 Training of a single classifier corresponding to one fixed centre pixel location for Single Image Block

6.4.3 Detection of Potential Defective Pixel Locations over Time

In each block of size $W \times H$, all pixel locations are used to train classifiers accordingly and assess their performance on a validation subset. That is, the number of classifiers in a block is almost the same number of pixels in that block with the exception of those pixels at the borders. Once all classifiers are trained, they are validated on validation subset to find potential defective pixel locations. The best K defective pixel locations are selected corresponding to the highest performance obtained. This is to discard the weak classifiers as their contribution to the final system would be insignificant. The process is conducted by sorting the scores of the classifiers in a descending order and taking the top K locations as shown in Figure 6-3.

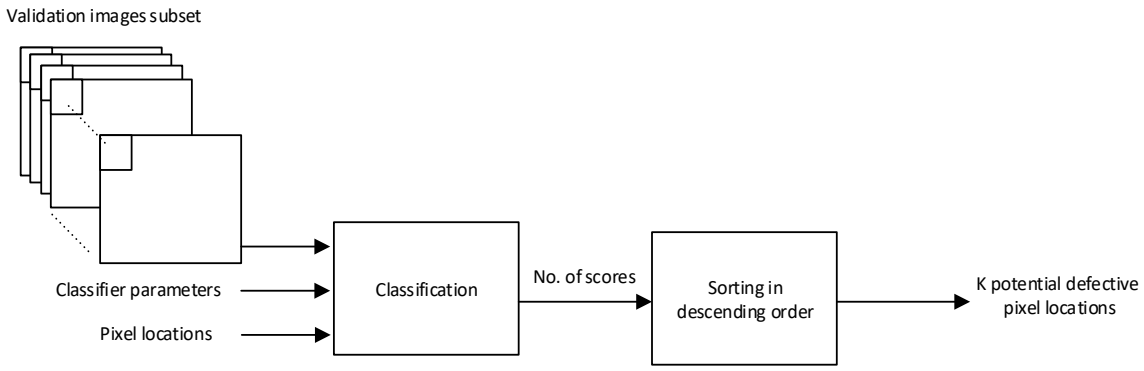


Figure 6-3 Filtering out the weak classifiers and selection of potential defective pixel locations

6.4.4 Re-Training with Virtual Timeslots and Reconstruction Process

Once the best K defective locations are detected from the previous stage, the corresponding classifiers are re-trained by combining both training and validation subset of images according to the selected K defective pixel locations. Note that this is carried out on virtual sub-classes where each sub-class represents a virtual timeslot by halving the actual timeslots (see Fig. 6-4). As will be demonstrated in experiments, this proposed idea enhances the accuracy of the system since the classifiers are trained in a more challenging scenario (shorter timeslots and less training samples for each sub-class when compared to actual classes). Each trained classifier corresponding to a specific defective pixel location can predict a virtual timeslot, accordingly.

Given a test image, non-overlapping blocks will be considered accordingly in a similar fashion as in the training stage where each block will have K location-based re-trained classifiers. The location-based decisions in the form of class labels for each image block are first obtained for virtual subclasses. The virtual timeslot is first used to reconstruct the actual class label for each location-based classifier. The representation of actual timeslots in relation to virtual timeslots is shown in Figure 6-4. The total number of timeslots used for classification is denoted by M. Each actual timeslot/class is spanned over a duration T.

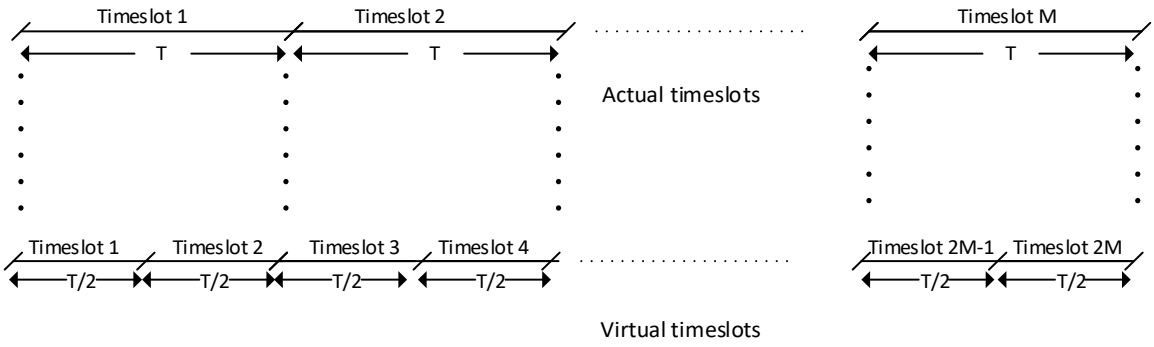


Figure 6-4 Overview of Actual and Virtual Timeslots

Once the actual class label is reconstructed for each re-trained location-based classifier, the final decision for one image block can be reached through the majority voting method, also referred to as the fusion method in Figure 6-5. This method takes the most occurred class label among all K re-trained classifiers' outputs within a given image block.

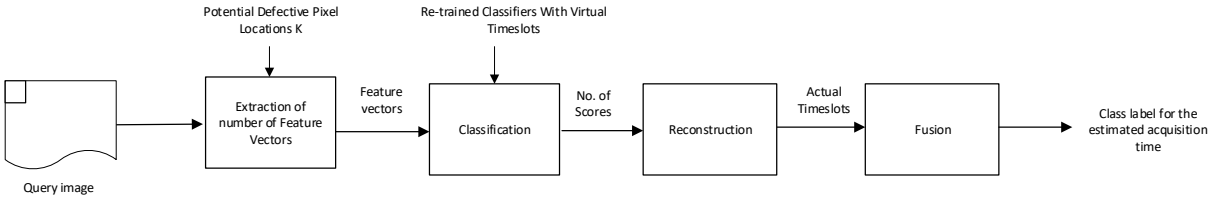


Figure 6-5 Prediction of Actual Timeslots for a Query Image from Single Image Block W×H

Figure 6-6 further depicts the process of estimating the acquisition timeslot for a single Image block $W \times H$ using K *re*-trained location-based classifiers.

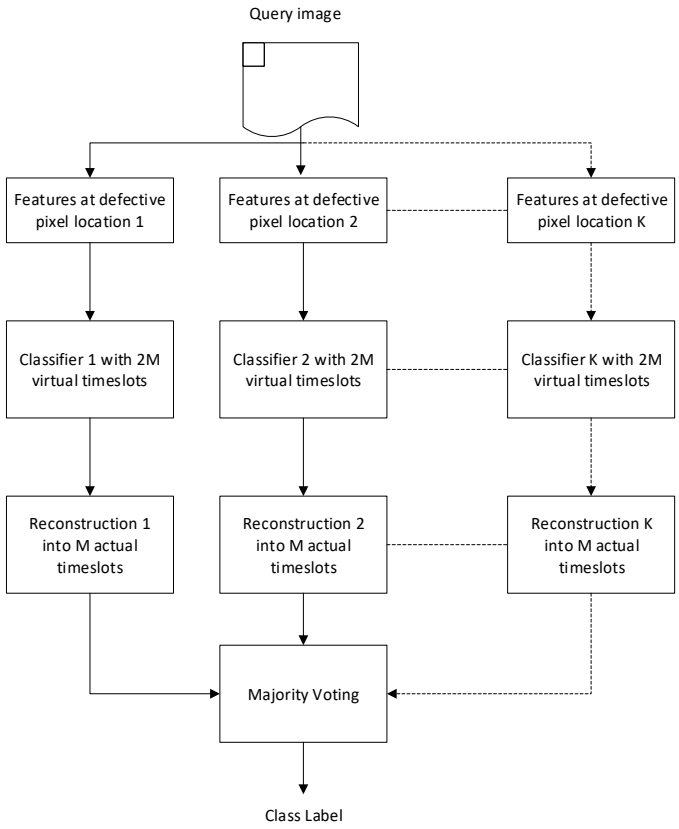


Figure 6-6 Acquisition Time Estimation of Query Image for Single Image Block W×H

6.4.5 Combination of Blocks for boosting Performance

The last stage of the system combines the predicted class labels on a number of image blocks to boost the performance of the system as will be demonstrated later. In fact, the predicted timeslot by the system is nothing but the combination of scores obtained on non-overlapping blocks through a second majority

voting approach. It is believed that potential defective pixels might occur in different spatial areas and hence including them could contribute efficiently to the picture dating process.

6.5 EXPERIMENTAL RESULTS AND ANALYSIS

The proposed system is assessed through a number of experiments with different types of classifiers in the first experiment. Mainly three multi-class classifiers were used to evaluate the system: k-nearest neighbour (KNN); error-correcting output codes (ECOC) using support vector machine (SVM) binary learner which is also known as multiclass SVM classifier; and the Naïve Bayes classifier. In KNN, the two parameters most often used in the classification model are a distance metric and the number of nearest neighbours. The optimal parameters used in experiment for KNN are Euclidean distance, which is computed to measure the similarity between two data points in a Euclidean space and the number of nearest neighbours is chosen as 1. The KNN classifier is effective if the training dataset is much larger than the number of features. ECOC is also an effective approach to deal with multi-class classification problems using a divide-and-conquer approach by decomposing the original multiclass classification problem into a series of binary classification problems. It contains many binary learners such as Naive Bayes, SVM or KNN. It solves multi-class learning with three major stages of encoding, binary classifier learning and decoding. In the encoding process, a coding matrix is used to determine multiple classes in which each row corresponds to a specific class, and each column is composed of independent binary classifiers. After dealing with each binary classification independently, a new data point is predicted by the decoding process using dichotomizers that compute the outputs of learned binary classifiers with a fusion process such as majority voting. For learning, ECOC uses different coding designs such as one-versus-all (OVA), one-versus-one (OVO), a dense random strategy or ternary designs. For SVM multiclass classifier, the best optimal parameters are chosen as a Gaussian kernel with value 1.3 and OVA coding design strategy in the experiments. The Naïve Bayes is a probabilistic classifier using a particular method of Bayesian learning. It follows the procedure based on Bayes' theorem of probability. The outcome of learning is to find the class that provides the highest probability given all input features with their values as the joint condition. The eligible parameters for Naïve Bayes are distribution names (normal or kernel), width and different types of kernels such as normal, box, epanechnikov and triangle. The default parameters of Naïve Bayes are used in the experiments as the kernel distribution with all four kernel parameters requiring a massive computation time to process the

results. In all experiments, accuracy is measured by the proportion of test images that have been correctly assigned to the right acquisition timeslot.

6.5.1 Image dataset

To gauge the effectiveness of acquisition time estimation on digital pictures for this proposed approach, the database of natural images called Northumbria Temporal Image Forensics (NTIF) is considered (Ahmed *et al.*, 2020a). As described in chapter 3, the series of images in the NTIF database have been taken at different times with regular acquisition timeslots spanned over for 94 weeks using ten digital camera devices. A total of 41684 images were captured from 10 digital cameras with different models and brands. The number of images were captured on a weekly and bi-weekly basis for a number of timeslots for each camera, which makes the NTIF dataset unique and beneficial for this particular problem (Ahmed *et al.*, 2020a). Lawgaly and Khelifi (2016); Al-Ani and Khelifi (2016) and Lawgaly, Khelifi and Bouridane (2014) used this unique dataset of images in their studies for source camera identification (SCI). Also, the NTIF dataset has been used in a study, which is based on a comparative analysis of deep learning for SCI (Ahmed *et al.*, 2019). The details of NTIF dataset can be found in (Ahmed *et al.*, 2020a).

Ten digital cameras from NTIF have been considered in this chapter. Table 6-1 shows the image resolutions and the type of sensor for each camera. From the NTIF database, the first 40 weeks were selected for 5 actual classes as shown in Table 6.2 where 8 weeks were grouped as one class. The reason for choosing 8 weeks in one class is because the occurrence of defective pixels appear in a period of two months, approximately (Fridrich and Goljan, 2011). The 5 actual classes represent a time period that ranges from 21st October 2014 to 24th September 2015 for each camera. It is worth mentioning, however, that the term ‘timeslot’ used at the link (Ahmed et al., 2020a) for the current structure of the database refers to a small period of time (usually one week or two weeks) and this is different from what was used and indicated in this chapter.

For each actual class, the number of training images was 240, whereas 80 validation images were used and 80 test images were considered for every digital camera. As explained earlier, the validation images will be included at the re-training stage once the potential defective pixels are detected. Therefore, the re-training dataset has a total of 1600 images.

Table 6-1 Characteristics of Ten Digital Cameras

Camera Model	**Resolution**	**Type of Sensor**
Canon IXUS115HS-1	4000 × 3000	1/2.3", CMOS
Fujifilm S2950-1	4288 × 3216	1/2.3", CCD
Nikon Coolpix L330-1	5152 × 3864	1/2.3", CCD
Panasonic DMC TZ20-1	4320 × 3240	1/2.33", CMOS
Samsung p1120-1	4320 × 3240	1/2.33", CCD
Canon IXUS115HS-2	4000 × 3000	1/2.3", CMOS
Fujifilm S2950-2	4288 × 3216	1/2.3", CCD
Nikon Coolpix L330-2	5152 × 3864	1/2.3", CCD
Panasonic DMC TZ20-2	4320 × 3240	1/2.33", CMOS
Samsung p1120-2	4320 × 3240	1/2.33", CCD

Table 6-2 Summary of Datasets used in Experimental Study for 5 Actual Classes

No. of Classes	**Date of Images**	**No. of Training Images**	**No. of Validation Images**	**No. of Testing Images**
Class-1 (Week1-Week8)	21-10-2014 to 14-12-2014	240	80	80
Class-2 (Week9-Week16)	22-12-2014 to 13-02-2015	240	80	80
Class-3 (Week17-Week24)	23-02-2015 to 22-04-2015	240	80	80
Class-4 (Week25-Week 32)	07-05-2015 to 01-07-2015	240	80	80
Class-5 (Week33-Week40)	09-07-2015 to 24-09-2015	240	80	80
Total	**40 Timeslots**	**1200**	**400**	**400**

6.5.2 Single block estimation with different classifiers

In the first set of experiments, only pixel neighborhood features (i.e., the neighbourhood of the centre pixel) are considered using a single image block of size 200×200 where the window size for constructing the feature vector is 3×3. Here, the idea of virtual sub-classes is not considered since the purpose of this experiment is to assess the performance of different classifiers in order to tune the system. For every centre pixel location, 27 pixel neighborhood features (nine from each colour channel) are extracted to be used in the training stage. For each location in the block, except the ones at the borders, a classifier is trained and the trained models are validated on the 400 validation images to identify potential defective pixel locations. From the block pixel locations, the best 100 potential candidates of defective pixel locations are selected. The system is then re-trained with combined training and validation images using the selected 100 defective pixel locations. The 100 re-trained classifiers were then selected to predict the acquisition timeslot of test images and their scores are combined through majority voting. Here, the estimation of the acquisition time of 400 test images is assessed through three different multiclass classifiers namely KNN, SVM and Naïve Bayes. Table 6-3 shows the performance of picture acquisition prediction for ten digital cameras with the aforementioned classifiers. As can be seen, KNN classifier has shown a significantly better performance in predicting the actual timeslot as compared to the other tested classifiers.

Table 6-3 Accuracy in % for 5 Actual Classes using Pixel Neighborhood Features

	Pixel Neighborhood Features		
No. of Digital Cameras	**KNN**	**SVM**	**Naïve Bayes**
Canon IXUS115HS-1	43.75	43.25	31.75
Fujifilm S2950-1	39.25	39.5	28.5
Nikon Coolpix L330-1	43.25	38.75	26.5
Panasonic DMC TZ20-1	49.25	47.75	29.25
Samsung p1120-1	45.5	41	27.5
Canon IXUS115HS-2	48	46	30.5
Fujifilm S2950-2	45.5	40.25	31.25
Nikon Coolpix L330-2	45.75	40	29.25
Panasonic DMC TZ20-2	46.75	41.25	27.75
Samsung p1120-2	48	46.75	31.75

6.5.2.1 *Contribution of Local Variation (LV) features*

In this experiment, the contribution of the proposed local variation features is highlighted. Similar to the previous experiment, the training and retraining processes are repeated but this time the proposed local variation features are included for all colour channels (i.e., 6 LV features as described by Equation 6.1 and 6.2). In Table 6-4, the performance of the classifiers for picture acquisition timeslot prediction on 400 test images including the local variation features are shown for ten digital cameras.

Table 6-4 Accuracy in % for 5 Actual Classes using Proposed Local Variation Features

No. of Digital Cameras	**KNN**		**SVM**		**Naïve Bayes**	
	Pixel Neighborhood Features	**Pixel Neighborhood and LV Features**	**Pixel Neighborhood Features**	**Pixel Neighborhood and LV Features**	**Pixel Neighborhood Features**	**Pixel Neighborhood and LV Features**
Canon IXUS115HS-1	43.75	**47**	43.25	**45.75**	31.75	**32.5**
Fujifilm S2950-1	39.25	**40**	39.5	**41.75**	28.5	**30.25**
Nikon Coolpix L330-1	43.25	**44.25**	38.75	**40.75**	26.5	**29.5**
Panasonic DMC TZ20-1	49.25	**49.75**	47.75	**49**	29.25	**29.75**
Samsung p1120-1	45.5	**46.5**	41	**42.5**	27.5	**29.75**
Canon IXUS115HS-2	48	**50.25**	46	**47.75**	30.5	**32.25**
Fujifilm S2950-2	45.5	**46.5**	40.25	**41.25**	31.25	**31.5**
Nikon Coolpix L330-2	45.75	**46**	40	**43**	29.25	**31**
Panasonic DMC TZ20-2	46.75	**48.5**	41.25	**43.5**	27.75	**29.25**
Samsung p1120-2	48	**49.5**	46.75	**47.75**	31.75	**32.5**

The results showed that the combination of pixel neighborhood and LV features has enhanced the performance of the system for all the tested digital cameras and classifiers. Similar to the previous experiment, the KNN classifier is clearly superior over other classifiers. As the contribution of local variation features enhanced the performance, therefore the other experiments will be conducted using pixel neighborhood and local variation features.

Moreover, the average loss function is also computed as a classification error for all the training images used in the classifier with respect to each centre pixel's location in a single block of size 200×200. The minimal expected classification cost for loss (Min et al., 2019; Li, Li and Yao, 2005) is used in both KNN and Naïve Bayes multiclass classifiers and hinge loss is used for ECOC (SVM binary learner).

Below, the Equation 6.3 to Equation 6.5 belong to minimal expected classification cost (Min et al., 2019; Li, Li and Yao, 2005) and Equation 6.6 is used for hinge loss (Joachims, 2002; Gao and Koller, 2011; Bashir et al., 2021) computation for the observations $t = 1, \ldots, n$.

As shown below, the Equation 6.3 is used to estimate the expected misclassification cost γ_{tk} when classifying the observation X_t into the class k;

$$\gamma_{tk} = (f(X_t)'C)_k \tag{6.3}$$

where $f(X_t)$ is the column vector of class scores for multiclass classification for the observation X_t. C is the Cost property (S-by-S numeric matrix of misclassification costs) of the input model (trained classifier) and S is number of classes

The class label $\hat{y}$ is predicted for the observation t corresponding to the minimal expected misclassification cost using Equation 6.4.

$$\hat{y} = \underbrace{argmin}_{k=1,\ldots\ldots S} \gamma_{tk} \tag{6.4}$$

By using C, the cost incurred (c_t) is identified for making the prediction. So, the weighted average of the minimal expected misclassification cost loss is computed using Equation 6.5.

$$L = \sum_{t=1}^{n} e_t \, c_t \tag{6.5}$$

Where L is the weighted average classification loss, n is the sample size and e_t (n-by-1 numeric vector) is the weight for observation t.

The Equation 6.6 is showing the hinge loss L_{hinge} computation for ECOC multiclass classifier (Joachims, 2002; Gao and Koller, 2011; Bashir et al., 2021).

$$L_{hinge} = \sum_{t=1}^{n} e_t \max\{0,1 - m_t\} \tag{6.6}$$

where $m_{t=}y_t{}^* f(X_t)$ is the scalar classification score that the trained model predicts for the true, observed class.$f(X_t)$ is the length vector of class scores for multiclass classification for the observation X_t and $y_t{}^*$ is a vector of $S - 1$ zeros, with 1 in the position corresponding to the true, observed class y_t.

In Table 6-5, the average loss function for each digital camera is shown using three different classifiers. The results show that the average loss function of the KNN classifier is lower compared to those for the other classifiers for all digital cameras.

Table 6-5 The Average Loss Function for all Digital Cameras

No. of Digital Cameras	KNN	SVM	Naïve Bayes
Canon IXUS115HS-1	7.040	1.08E+04	3.20E+04
Fujifilm S2950-1	0.624	1.49E+04	3.13E+04
Nikon Coolpix L330-1	6.652	1.43E+04	3.19E+04
Panasonic DMC TZ20-1	1.605	1.04E+04	3.28E+04
Samsung p1120-1	0	1.22E+04	3.16E+04
Canon IXUS115HS-2	0.123	1.02E+04	3.08E+04
Fujifilm S2950-2	2.242	1.41E+04	3.28E+04
Nikon Coolpix L330-2	4.626	1.34E+04	3.19E+04
Panasonic DMC TZ20-2	0.588	1.02E+04	3.16E+04
Samsung p1120-2	8.175	1.28E+04	3.18E+04

6.5.3 Contribution of virtual subclasses

In this experiment, the only difference from the previous experiment is the introduction of the concept of virtual sub-classes. In this experiment, single block of size 200×200 from a test image is used to predict its acquisition timeslot. As part of the proposed system and as explained in section 6.4.4, each actual class is sub-divided into two virtual sub-classes corresponding to halved timeslots of actual classes as shown in Figure 6-7. That is, for the scenario of 5 actual classes, 10 virtual subclasses are now used to re-train the system once the potential defective pixels are determined. Therefore, the re-training dataset has a total of 1600 images for the 10 virtual subclasses with 160 images per subclass whereas the testing dataset comprises of a total of 400 images where each subclass contains 40 images. The feature extraction process is the same as discussed earlier.

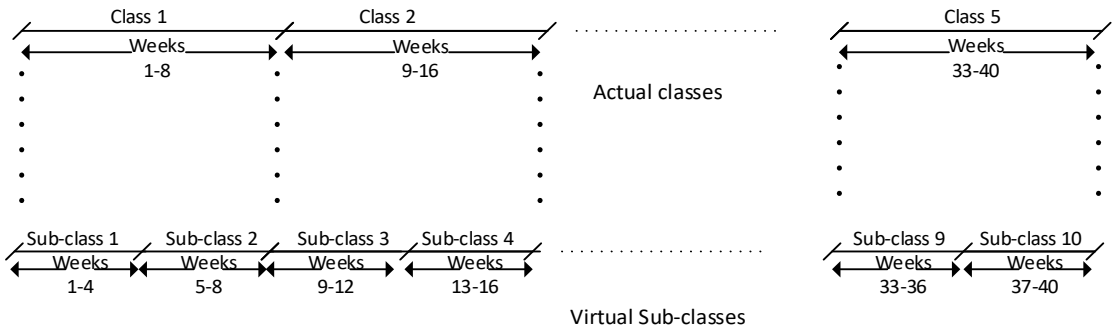


Figure 6-7 Construction of 10 virtual sub-classes from 5 actual classes

The previous types of classifiers are used again here for the aforementioned 10 digital cameras. As depicted in Table 6-6, the results show that the concept of re-training with 10 virtual subclasses (i.e., halved timeslots) prior to reconstruction into the 5 actual classes significantly enhanced the performance of the system when compared with the process of directly re-training with 5 actual classes.

Table 6-6 Accuracy in % with Reconstruction of 10 Virtual Subclasses into 5 Actual Classes using Single Image Block

No. of Digital Cameras	**KNN**		**SVM**		**Naïve Bayes**	
	5 Actual Classes	**Reconstruction of 10 virtual subclasses into 5 actual classes**	**5 Actual Classes**	**Reconstruction of 10 virtual subclasses into 5 actual classes**	**5 Actual Classes**	**Reconstruction of 10 virtual subclasses into 5 actual classes**
Canon IXUS115HS-1	47	**64.75**	45.75	**49.25**	32.5	**35.25**
Fujifilm S2950-1	40	**56.5**	41.75	**44.75**	30.25	**31.25**
Nikon Coolpix L330-1	44.25	**57**	40.75	**42.75**	29.5	**31.25**
Panasonic DMC TZ20-1	49.75	**64.75**	49	**54.75**	29.75	**30**
Samsung p1120-1	46.5	**62**	42.5	**52**	27.5	**30.25**
Canon IXUS115HS-2	50.25	**65.75**	47.75	**52.25**	32.25	**37.25**
Fujifilm S2950-2	46.5	**59**	41.25	**44**	31.5	**32.5**
Nikon Coolpix L330-2	46	**62.75**	43	**45.75**	31	**32.75**
Panasonic DMC TZ20-2	48.5	**61.5**	43.5	**46**	29.25	**30.25**
Samsung p1120-2	49.5	**61.5**	47.75	**55**	32.5	**33.75**

Furthermore, the results validate observation on the KNN classifier which again has shown a significantly better performance in predicting the actual timeslot as compared to the other tested classifiers. Consequently, the KNN classifier is adopted in the rest of the paper.

6.5.4 Contribution of multi-block score fusion

In this experiment, the contribution of multiple non-overlapping image blocks is combined through majority voting and this is assessed, accordingly using the same dataset. Up to 50 non-overlapping image blocks have been randomly chosen from each image. Results are illustrated in Figure 6-8 to Figure 6-17 for the 10 digital cameras. As can be seen, the performance of the system significantly jumps against the increasing number of combined blocks before it stabilizes at around 45 blocks.

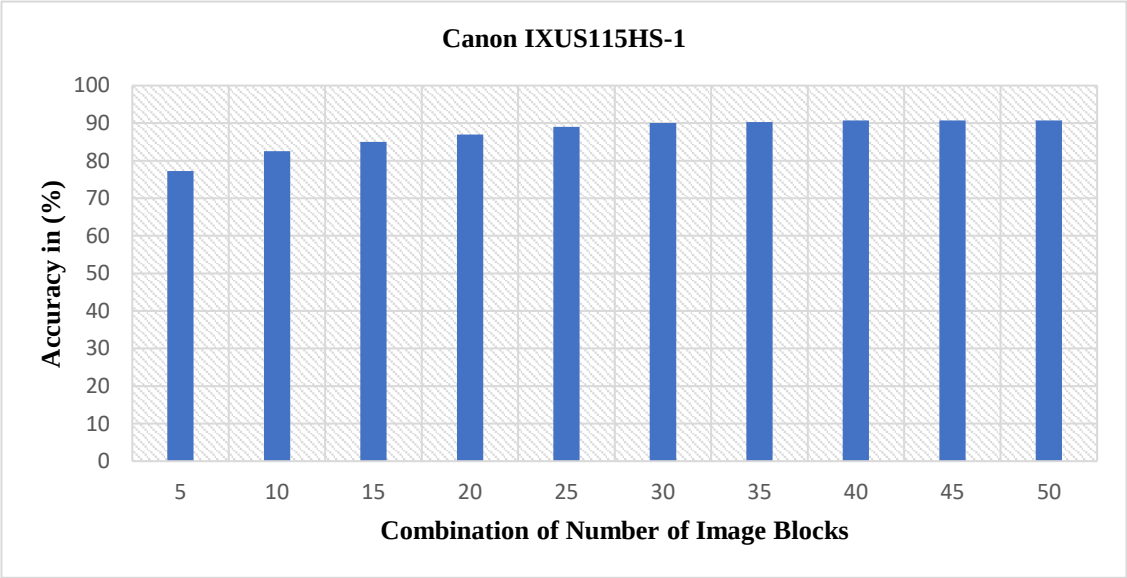


Figure 6-8 Accuracy in % on 400 Test Images using the Combination of Blocks (up to 50 blocks) of size 200×200 for Canon IXUS115HS-1

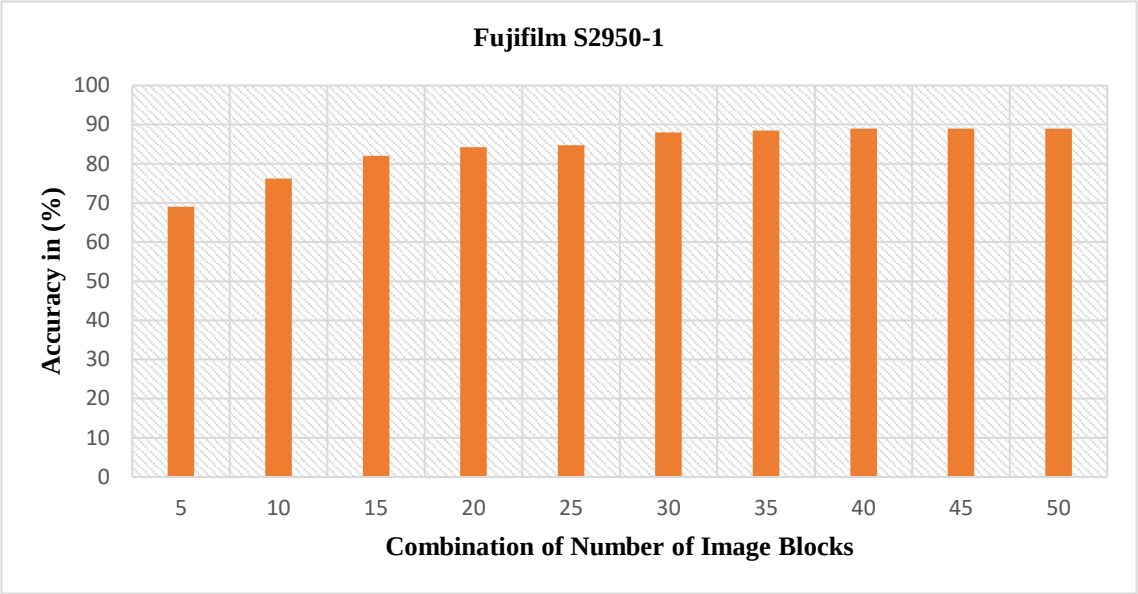


Figure 6-9 Accuracy in % on 400 Test Images using the Combination of Blocks (up to 50 blocks) of size 200×200 for Fujifilm S2950-1

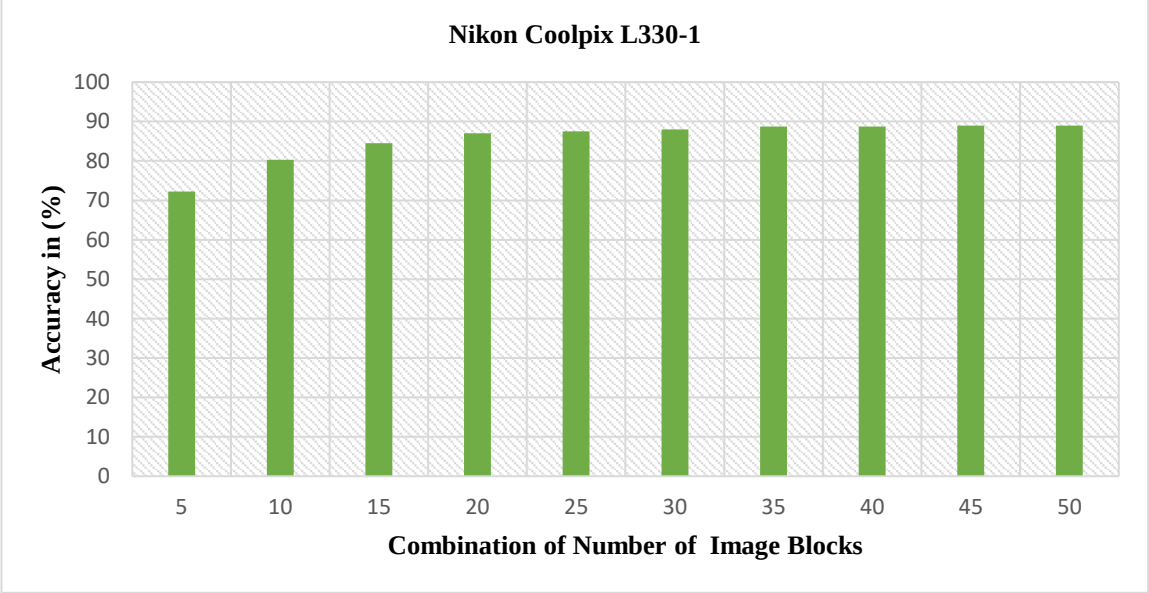


Figure 6-10 Accuracy in % on 400 Test Images using the Combination of Blocks (up to 50 blocks) of size 200×200 for Nikon Coolpix L330-1

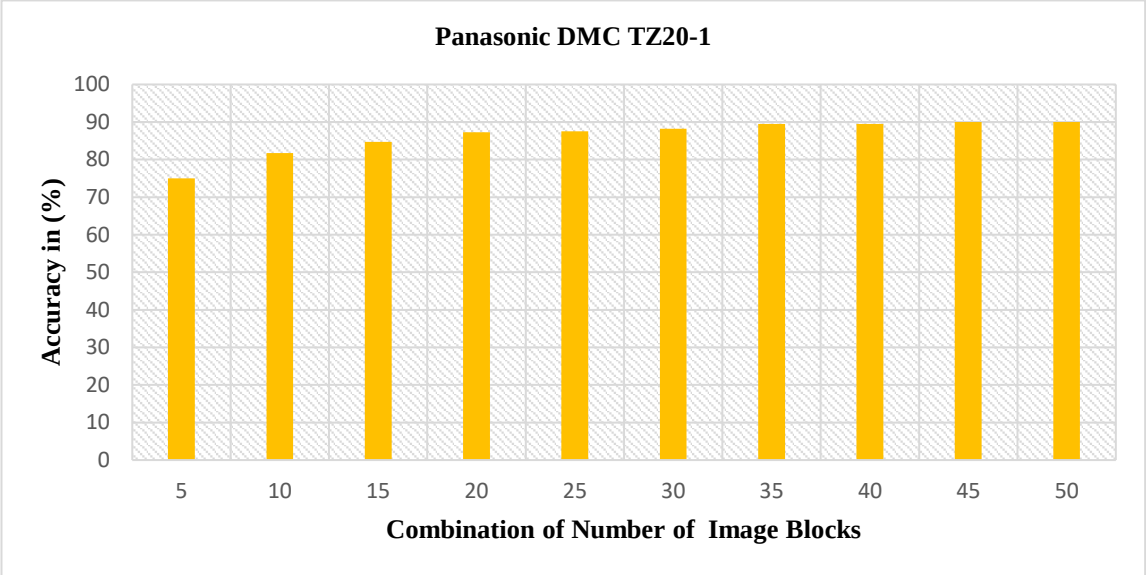


Figure 6-11 Accuracy in % on 400 Test Images using the Combination of Blocks (up to 50 blocks) of size 200×200 for Panasonic DMC TZ20-1

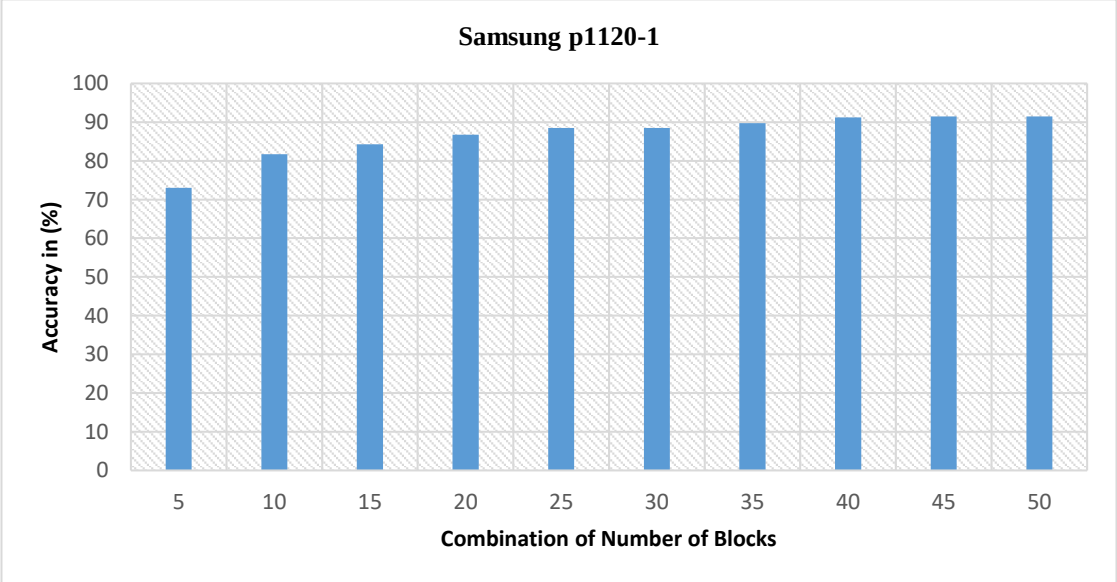


Figure 6-12 Accuracy in % on 400 Test Images using the Combination of Blocks (up to 50 blocks) of size 200×200 for Samsung p1120-1

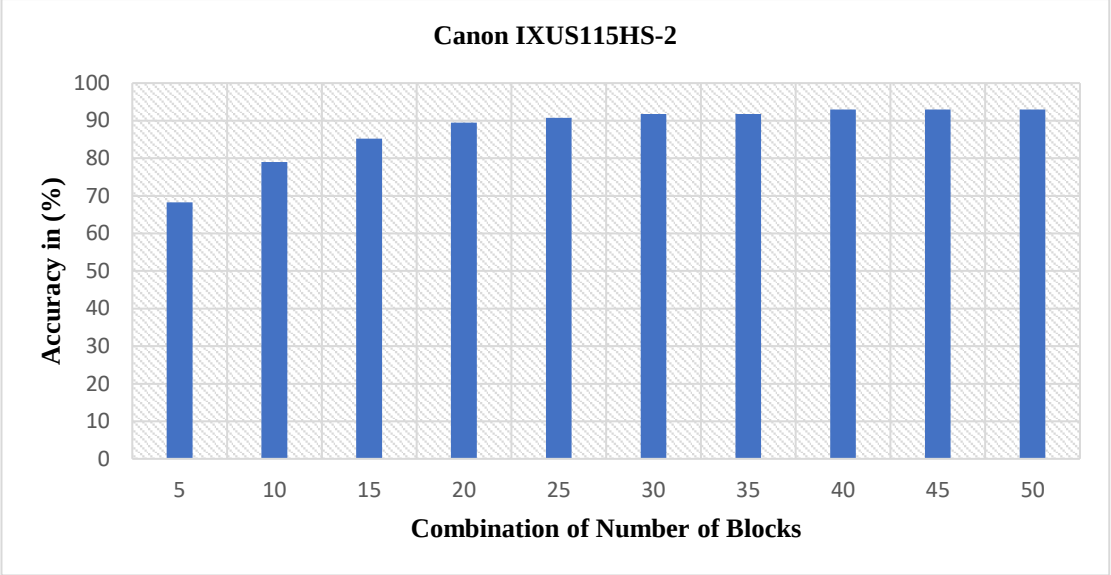


Figure 6-13 Accuracy in % on 400 Test Images using the Combination of Blocks (up to 50 blocks) of size 200×200 for Canon IXUS115HS-2

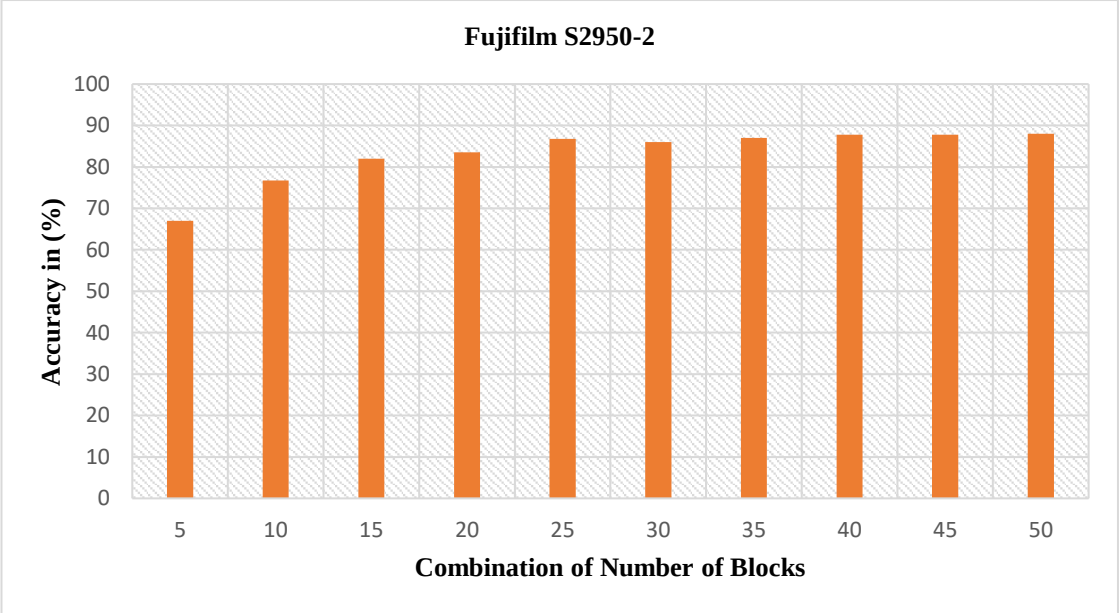


Figure 6-14 Accuracy in % on 400 Test Images using the Combination of Blocks (up to 50 blocks) of size 200×200 for Fujifilm S2950-2

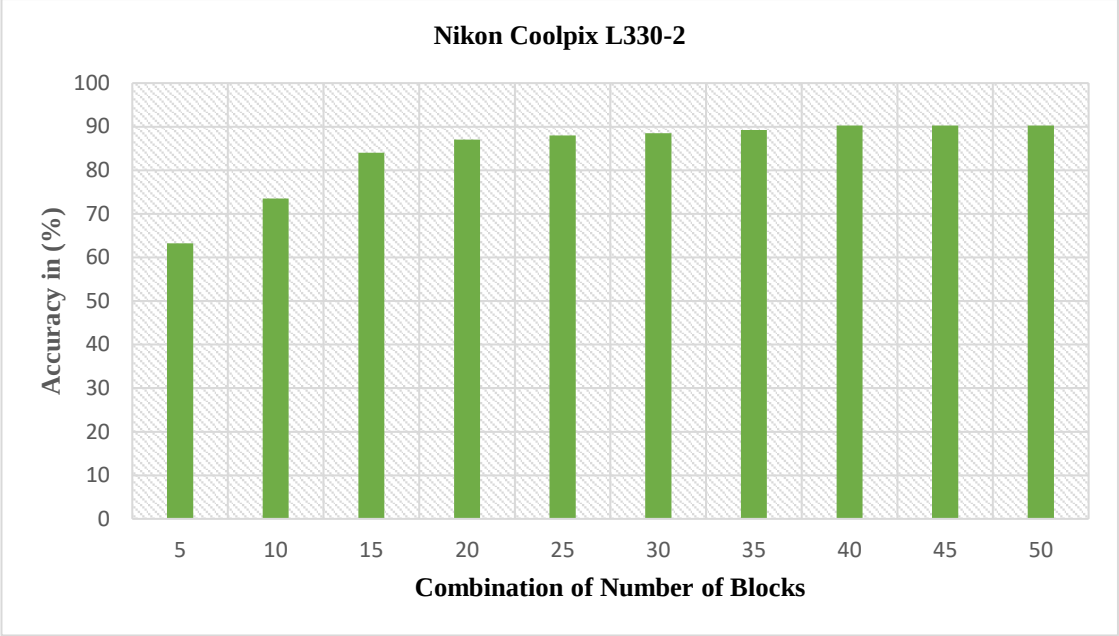


Figure 6-15 Accuracy in % on 400 Test Images using the Combination of Blocks (up to 50 blocks) of size 200×200 for Nikon Coolpix L330-2

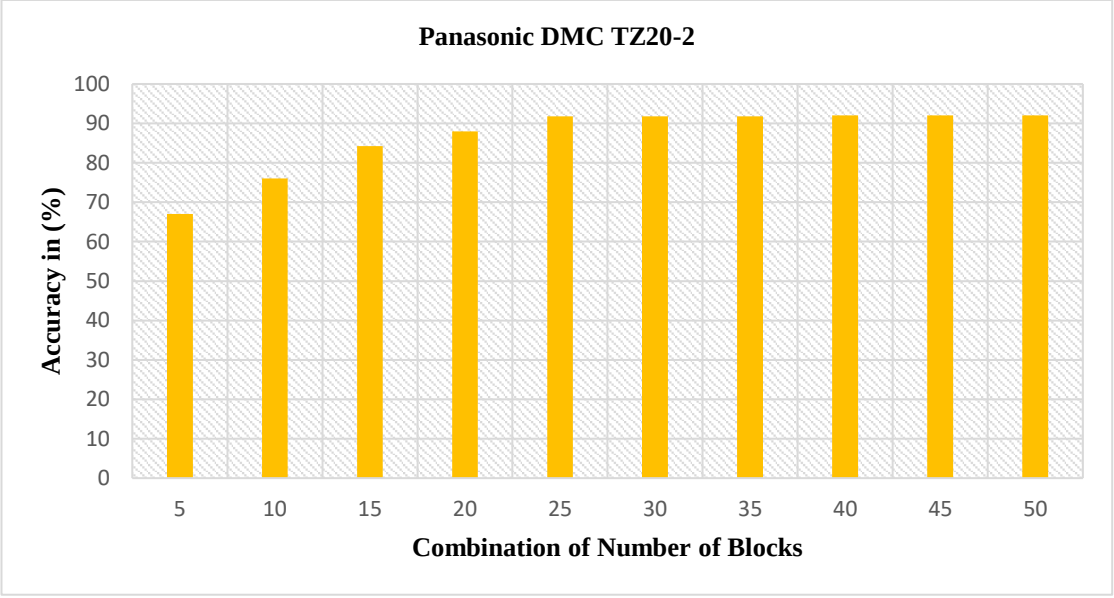


Figure 6-16 Accuracy in % on 400 Test Images using the Combination of Blocks (up to 50 blocks) of size 200×200 for Panasonic DMC TZ20-2

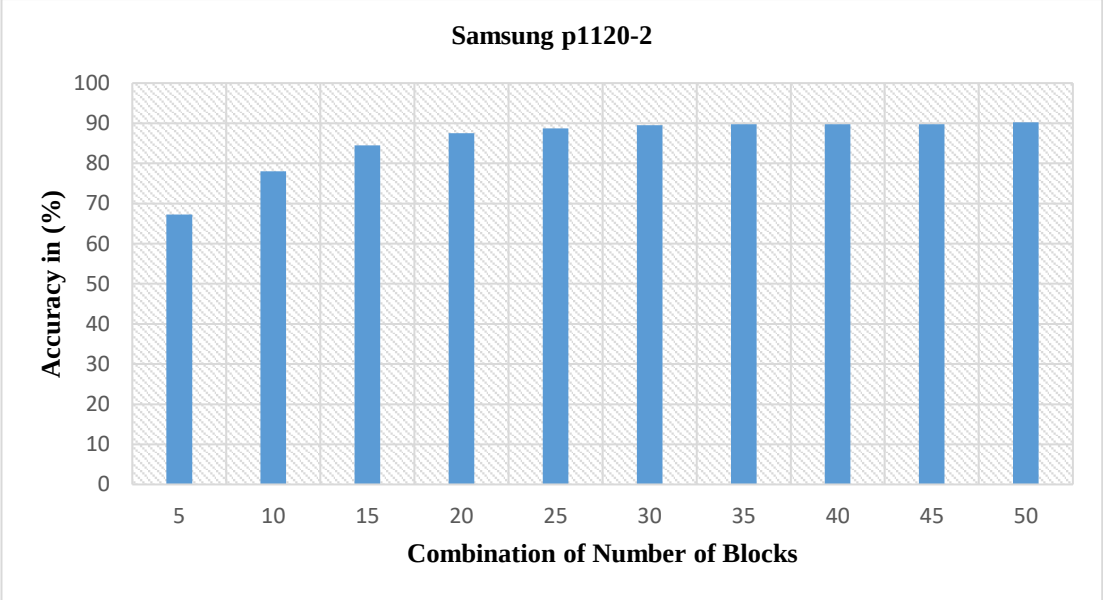


Figure 6-17 Accuracy in % on 400 Test Images using the Combination of Blocks (up to 50 blocks) of size 200×200 for Samsung p1120-2

In this experiment, it has been found that the combination of 45 image blocks through majority voting is the optimal parameter to reach the highest accuracy while giving a good tradeoff between performance and computational complexity. The majority vote method is considered fair since it gives a higher accuracy when the decisions in the form of the class/time slot from 45 image blocks for one query image are combined. Then, the most frequent time slot prediction is assigned to that query image as the 'final' time slot. Table 6-7 depicts the accuracy of the system when 45 non-overlapping image blocks were combined as compared to a single block performance for all 10 digital cameras. It can be seen that the proposed system improves considerably with an accuracy varying from 88 % to 93 %.

Table 6-7 Accuracy in % for 10 Digital Cameras using the Combination of 45 Blocks with KNN classifier

	Accuracy in %	
No. of Digital Cameras	**Single Block**	**45 Multi-Blocks**
Canon IXUS115HS-1	64.75	**90.75**
Fujifilm S2950-1	56.5	**89**
Nikon Coolpix L330-1	57	**89**
Panasonic DMC TZ20-1	64.75	**90**
Samsung p1120-1	62	**91.5**
Canon IXUS115HS-2	65.75	**93**
Fujifilm S2950-2	59	**88**
Nikon Coolpix L330-2	62.75	**90.25**
Panasonic DMC TZ20-2	61.5	**92**
Samsung p1120-2	61.5	**90.25**

Moreover, the classification evaluation using a confusion matrix for five actual classes is shown in Figure 6-18 for one digital camera model, the Canon IXUS115HS-2. It is worth noting that the 77 test images of the class 1 time slot are accurately predicted out of 80 test images as compared to other classes.

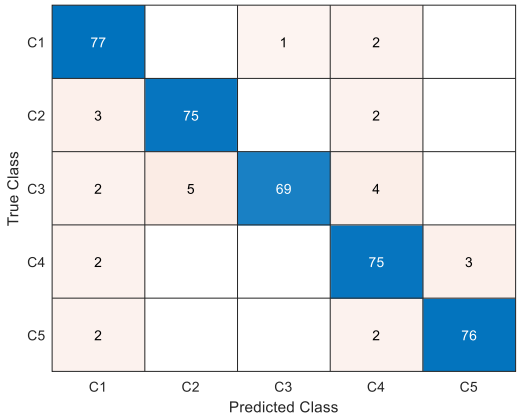


Figure 6-18 Confusion Matrix of Camera Canon IXUS115HS-2 showing Five Actual Timeslots Classification

6.5.5 Comparison with State-of-the-Art System

In this section, the proposed system is compared with an existing state-of-the-art system, namely Individual Image Placement (IIP) within an ordered cluster set (Mao *et al.*, 2009). For fair comparison, the same number of classes (timeslots) and number of training and test images have been used. Results are depicted in Table 6-8. The proposed system provides the best performance when compared with the competing technique. The proposed system appears significant more powerful in dealing with the problem of predicting image acquisition timeslot and the results obtained are very promising. In forensic investigations, it is believed that these results can provide the opportunities to the analysts to link different events of crime scenes and to determine reliably the timeframes in which evidential images were captured.

Table 6-8 Performance of state-of-the-art system and proposed system in % for 10 Digital Cameras

	Accuracy	
No. of Digital Cameras	**State-of-the-art System**	**Proposed system**
Canon IXUS115HS-1	46.25	**90.75**
Fujifilm S2950-1	41	**89**
Nikon Coolpix L330-1	44.75	**89**
Panasonic DMC TZ20-1	49	**90**
Samsung p1120-1	44	**91.5**
Canon IXUS115HS-2	44.75	**93**
Fujifilm S2950-2	40.75	**88**
Nikon Coolpix L330-2	44.25	**90.25**
Panasonic DMC TZ20-2	45	**92**
Samsung p1120-2	42.25	**90.25**

6.6 PRACTICAL IMPLICATIONS OF PROPOSED WORK

There are two possible applications for the proposed work. First, it can be used to estimate the acquisition time of a picture given a set of images taken over a period of time and whose dates are known. For instance, a practical scenario could be when a picture of interest (or a number of pictures) is found on social media (this could also be a cropped version of the original image) but there is no indication and evidence on when the picture was taken. If there are a good number of images seized at the property of the suspect and these images were taken by the same device as the picture of interest (This could be verified via PRNU estimation), the investigator can use them for training in order to

place the picture of interest in the right timeframe order. The other application is the verification of acquisition time. The date of a picture of interest is known through the file header but to confirm this information for courtroom purposes, the analyst can approve or disapprove the claimed date of acquisition by using a set of images taken over a period of time. According to our experiments, 320 images per time interval (2 months) are sufficient to achieve high accuracy estimation. This is nowadays plausible given the impressive growth of digital image usage and sharing on different social media platforms as well as the availability and easiness of the imaging technology in modern smart phones and tablets. In fact, most modern smart phones have an internal memory storage of 64 GB or more and this is mainly driven by the significant use of digital images and videos.

6.7 CONCLUSION

In this chapter, several contributions have been made for estimating the acquisition time of digital pictures. The idea uses a combination of an effective defective pixel detection and a machine learning approach to estimate the age of digital pictures. The local variation features were proposed that efficiently detect potential defective pixels for picture dating. Virtual timeslots were introduced prior to the reconstruction step into actual classes. Once the actual class label is reconstructed for each defective pixel location-based classifier, the final decision for a single image block is reached through the majority voting method. To further enhance the performance, a multi-block based machine learning model is used to combine the predicted scores from multiples blocks in a second majority voting operation. Extensive experiments have been performed to evaluate the performance of the proposed system from three different multi-class classifiers for five digital cameras. The experimental results show that KNN is the best to classify the picture acquisition timeslot prediction using defective pixels. The results showed very good performance with an estimated accuracy between 88% and 93% when compared with a related state-of-the-art system.

CHAPTER 7 CONCLUSIONS AND RECOMMENDATIONS

7.1 SUMMARY OF THE THESIS

The rapid development of digital technologies over the last decade has created significant challenges for digital forensics investigators and has shaped how forensic analysts conduct digital forensic investigations. Unfortunately, these technological advances have also become the new weapons of choice to be used in criminal acts. Different digital devices are utilized to capture and record still images and videos. So, the collection of covert videos and illegal content has taken place on a massive scale as seen in the Worldwide Web. Even unskilled users can manipulate and tamper with the content of images with only minimal effort. If such pictures or videos are related to a crime, then forensic analysts aim to investigate such digital content. Not all such contents can be protected with the help of digital watermarks or signatures. In fact, passive techniques provide great levels of protection and the authentication of digital content. Many forensic techniques have been developed to identify the history of digital images, perform forgery detection and identify the image's source device and camera model involved. PRNU has been used in forensic image analysis to determine whether a number of images were captured by the same camera. To address these issues, many high-quality forensic databases have been developed and made available in which images were collected from different devices such as digital cameras, smartphones, camcorders and social networks to help in the development of forensic algorithms. Such databases were built and presented for the forensic research community in order to benchmark forensic techniques.

Nowadays, massive and intensive forensic searches are based on the analysis of the timelines of crime events so as to contribute to investigations of murder, kidnap, or rape. The increasing importance of temporal information needed in such investigations has created a need to develop the forensic techniques for temporal forensic image analysis. The temporal forensic analysis of any given device or set of devices can help the forensic analyst to frame the links between any crime incident and other events. It mainly involves the collection of information within a particular timeframe. Through temporal forensic analysis, the investigator can determine exactly when a particular crime or other events occurred. Temporal digital evidence can provide more information on event sequences, activity levels and timing information. Based on temporal digital evidence, the forensic analyst can answer different

questions about victims or suspects (who), activities (what), places (whereabouts) and times (when). Images and other visual recordings are important means to document the condition of a subject at any instant in time. The subject of such images might include scenes related to a crime or accident, victims, suspects or any items of evidence. Such thorough and complete photographic images from recordings of a camera or any digital imaging technology of a scene, suspect or items of evidence can be helpful to find meaningful information which can lead to substantial direct or circumstantial evidence. To correlate digital artifacts, the forensic analyst needs to process the correct timeline ordering of any suspect images so as to understand the sequence of events that have occurred.

Digital pictures are mostly used to testify incidents and provide legally acceptable evidence for courtroom purposes. In such incident cases, it is essential to gather genuine information about the picture for forensic investigations. In particular, the field of digital image forensics is concerned with the originality and integrity. The analyst identifies the source camera. If many imaging devices of the same model and make happen to be under investigation, the analyst needs to identify the right device which was used to take the images under analysis. The sub-field of camera forensics has made significant advances due to efficient exploitation of the imperfections of digital imaging sensors. It has been widely accepted in the literature that PRNU is a well-established forensic method for matching an image to a source camera. In fact, each digital camera's PRNU is unique, and it can withstand lossy compression, filtering, and a variety of other image processing procedures. Therefore, chapter 2 reviews different algorithms for SCI in relation to the PRNU denoising, estimation and detection processes. With the aim of improving detection performance in source camera identification, different detectors have been investigated. Many researchers have made significant contributions to source camera identification (SCI) applications using machine learning and deep learning techniques, and many recent studies have begun to use a deep learning-based CNN approach in order to identify a source camera.

While many forensic tools and approaches for digital image analysis have been developed, the review in chapter 2 shows that scant research has been conducted in the field of the temporal forensic analysis of digital images. In forensic investigations, it is important for the analyst to know the conditions under which the picture was taken (including exposure, date, time, and other variables). The growing importance of temporal information in forensics has necessitated the development of temporal forensic image analysis. Defective pixels also constitute another sensor imperfection of forensic relevance which

can be useful for picture dating applications. Therefore, the importance of defective pixels in the sensor ageing process and for picture dating has been explored. Also, different image and video databases used for digital image forensic applications are discussed. From the review in chapter 2, it is observed that there is currently no publicly available database suitable for temporal image forensics analysis among all of the image and video datasets currently utilized for benchmarking.

In chapter 3, a temporal image database called Northumbria Temporal Image Forensics (NTIF) has been introduced specifically for use in temporal image forensic analysis. The characteristics of the NTIF database are discussed in detail in relation to each type of digital camera and image, including camera make, model, the whole acquisition time-period of images, camera resolution, image format, type of sensor, sensor size, and number of natural and blue sky images and range of exposure setting parameters. In this thesis, the relevance and importance of the NTIF database for the purpose of temporal image forensic analysis has been demonstrated through extensive experimentation. The changes in and strength of correlations detected in PRNU over time have been highlighted to represent the sensor ageing process. In addition, different patterns of correlation have been identified when the images are captured in different time slots as compared to those acquired within the same time slots for all digital cameras. The NTIF database has made freely available to benefit all researchers in the image forensics field in academia and industry.

In chapter 4, the effectiveness of the PRNU-based technique has been shown for application in SCI. A comparative analysis is conducted between state-of-the-art filter-based PRNU and deep learning convolutional neural network (CNN)-based techniques. To this end, the basic sensor pattern noise-based estimation technique using the Wiener filter in the wavelet domain is considered as well as the deep learning CNN model in order to assess their effectiveness for SCI. Different experiments have been carried out using the NTIF database of images for both approaches. From the results, it is shown that the filter-based PRNU approach for SCI is better than the deep learning CNN approach when limited numbers of images are available in digital image forensic investigations. Also, it is concluded that PRNU can distinguish images taken by different models of the same brand, while the CNN model clearly fails to do so. In deep learning, large numbers of images are required in the training process in order to achieve high performance. However, in most forensic investigations, relatively few images

acquired by the camera of interest will be available for the estimation of PRNU, and therefore a PRNU-based machine learning technique is the better choice in SCI applications.

In chapter 5, the acquisition time of digital images is predicted using a deep learning approach. The problem of picture dating has been formulated as a classification problem in which the time of the acquisition of images can be discretized into a number of classes. In fact, each class is represented as a time slot and represents a group of images captured in a period of 5 weeks' duration. In classification, the predicted modelling problems are involved to predict a class label or probability of class labels for a given input. The classification accuracy for each class (time slot) is then computed as the percentage of correctly classified predictions out of all predictions made. The results of the application of convolutional neural networks (CNNs) using the AlexNet and GoogLeNet architectures in both feature extraction and pre-trained modes have shown that these networks can successfully learn the temporal changes in the content of digital pictures acquired from the same source. It is concluded that the deep learning approach is more suitable in learning temporal changes for picture dating appliction rather than SCI application.

In chapter 6, a machine learning-based approach for picture dating using defective pixels is proposed. The temporal variations in camera sensor imperfections are investigated as a means of estimating the acquisition date of digital pictures using defective pixels and a machine learning-based approach. In this context, the technique used to predict the acquisition time slots of digital pictures is based on pixel neighbourhood correlation and local variation features to find the best candidates of defective pixels. Several contributions have been made to enhance the performance of picture dating, such as combining defective pixel detection and a machine learning technique for picture dating applications using the proposed feature called local variation that can detect the best candidate of defective pixels for picture dating, a re-training-based method using virtual time slots to more accurately estimate the time slots in which digital images were taken, and finally a multi-block-based machine learning model that combines the predicted scores from multiple blocks to boost the performance of the system. These contributions have been assessed and demonstrated through a number of experiments. The proposed system has also been shown to outperform a relevant state-of-the-art technique.

This thesis offers significant advances in the sub-field of digital camera temporal forensics due to the efficient exploitation of the imperfections of digital imaging sensors for picture dating purposes. The improved results at between 88% to 93% accuracy in picture dating can be helpful for forensic investigators in determining the full details of a timeline activity in terms of the images from a given digital device or set of devices associated with a particular crime or incident. If such images are presented in the court for investigation purposes, it is crucial that every possible measure should be taken to ensure the reliability and accuracy of picture dating.

7.2 FUTURE WORK

Although various aspects of picture dating have been considered in this thesis, there are several areas in which the current work could be extended, as follows:

7.2.1 Digital Video Forensic Dataset

With the rise in popularity of digital technologies, massive volumes of digital images and videos are captured and circulated over the internet. In addition, the range of imaging devices used has expanded from digital cameras to smart phone cameras. Also, the editing and manipulation of videos has become an easy task with the help of various software. Consequently, the capture source of images or videos provides crucial forensic evidence and important identity information for forensic investigations. Although, some video datasets are available for scientific research purposes but their contents were not captured in regular time slots suitable for temporal analysis. Therefore, a digital video forensics dataset will be built as a future project in which series of video clip sequences will be captured at different time slots. In order to facilitate future research, a large number of different digital devices will be used to capture high-resolution images and videos, such as smart phones, digital cameras, tablets, video surveillance equipment and digital camcorders.

7.2.2 Temporal Video-based Forensics

Digital videos provide a wider range of information about the sensor than images due to their rich visual and perceptual content. However, using videos in digital forensics could be more challenging due to the nature of lossy compression in which digital videos are normally presented. In future work, video forensics will be considered with the aim of video dating. This will be of great help to forensic investigators in trying to determine the sequence of events in criminal cases as videos are used as often as images nowadays in social media platforms and daily applications. By taking the advantage of accessibility to large training data in digital video forensics, deep convolution neural networks could be the ideal solution.

7.2.3 Additional Sensor Parameters as Features for Picture Dating

In relation to defective pixels, it has been reported in the literature (Leung *et al.*, 2010; Leung *et al.*, 2009; Leung *et al.*, 2008b; Leung *et al.*, 2007; Chapman *et al.*, 2011; Chapman *et al.*, 2012; Chapman *et al.*, 2013; Dudas *et al.*, 2007; Chapman *et al.*, 2010; Kauba and Uhl, 2016) that not only the cosmic-ray radiation is the main cause of sensor defects but also parameters such as sensor area and type (CCD or CMOS), pixel size and exposure settings. All of these factors can have an impact on the growth rate of the defective pixels which increases linearly over time. Previous research in digital image forensics has shown that there is a huge impact of exposure setting parameters. Also, the effect on image output is predicated upon temperature, ISO (gain) and exposure time. Most digital cameras are set in auto-mode to select all exposure settings, but these sensor parameters are also responsible for defects in pixels if not exposed correctly. For instance, higher ISO produces more digital noise and can also amplify the number of defects. Lower shutter speed produces more blurred images. Increasing the exposure time or shutter speed can reduce blurring in the picture but consequently enhance the development of defective pixels (Leung *et al.*, 2009). Also, when the shutter speed is slower, the amount of light hitting the sensor increases. Thus, the image looks brighter and may have the possibility of high saturated defective pixels. In future work, additional features related to camera sensor parameters need to be considered for defective pixel detection and picture dating.

7.2.4 Prediction of a Location

Finally, it would be interesting to consider in the future forensic analysis and prediction of the location in which a digital photo was taken by processing their content. This would involve deep learning and large datasets. Such an application would be of great help to forensic investigators in order to establish links between events and gather pieces of evidence.

7.2.5 Audio Forensic Analysis

Apart from image and video analysis, the audio forensic evidence is also the critical evidence in a criminal or civil case. The primary purpose of audio forensics is to establish the authenticity of audio evidence, to enhance the speech intelligibility and to reconstruct the crime scenes or timelines. Therefore, the audio signals/data can be analyzed using deep learning models as a future task. Various processing and compression techniques can be looked which is the cause of introducing unwanted noise in the audio data. To authenticate the audio, different audio waveforms and frequency information using spectrogram can be inspected in the future work.

REFERENCES

Adams, J. E. and Pillman, B. (2013) 'Digital Camera Image Formation: Introduction and Hardware', *Digital Image Forensics*: Springer, pp. 3-44.

Ahmed, F., Khelifi, F., Lawgalv, A. and Bouridane, A. 'Comparative Analysis of a Deep Convolutional Neural Network for Source Camera Identification'. *2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3)*: IEEE, 1-6.

Ahmed, F., Khelifi, F., Lawgaly, A. and Bouridane, A. 'The 'Northumbria Temporal Image Forensics' Database: Description and Analysis'. *2020 7th International Conference on Control, Decision and Information Technologies (CoDIT)*: IEEE, 982-987.

Ahmed, F., Khelifi, F., Lawgaly, A. and Bouridane, A. 'Temporal Image Forensic Analysis for Picture Dating with Deep Learning'. *2020 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*: IEEE, 109-114.

Al-Ani, M. and Khelifi, F. (2016) 'On the SPN estimation in image forensics: a systematic empirical evaluation', *IEEE Transactions on Information Forensics and Security,* 12(5), pp. 1067-1081.

Al-Ani, M., Khelifi, F., Lawgaly, A. and Bouridane, A. 'A novel image filtering approach for sensor fingerprint estimation in source camera identification'. *2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*: IEEE, 1-5.

Al Banna, M. H., Haider, M. A., Al Nahian, M. J., Islam, M. M., Taher, K. A. and Kaiser, M. S. 'Camera model identification using deep CNN and transfer learning approach'. *2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*: IEEE, 626-630.

An, J., Lee, W. and Kim, J. 'Adaptive detection and concealment algorithm of defective pixel'. *2007 IEEE Workshop on Signal Processing Systems*: IEEE, 651-656.

Athanasiadou, E., Geradts, Z. and Van Eijk, E. (2018) 'Camera recognition with deep learning', *Forensic Sci Res,* 3(3), pp. 210-218.

Baroffio, L., Bondi, L., Bestagini, P. and Tubaro, S. (2016) 'Camera identification with deep convolutional networks', *arXiv preprint arXiv:1603.01068*.

Bashir, A. K., Khan, S., Prabadevi, B., Deepa, N., Alnumay, W. S., Gadekallu, T. R. and Maddikunta, P. K. R. (2021) 'Comparative analysis of machine learning algorithms for prediction of smart grid stability', *International Transactions on Electrical Energy Systems*, pp. e12706.

Basly, H., Ouarda, W., Sayadi, F. E., Ouni, B. and Alimi, A. M. 'CNN-SVM learning approach based human activity recognition'. *International Conference on Image and Signal Processing*: Springer, 271-281.

Bayram, S., Sencar, H., Memon, N. and Avcibas, I. 'Source camera identification based on CFA interpolation'. *IEEE International Conference on Image Processing 2005*: IEEE, III-69.

Bergmüller, T., Debiasi, L., Uhl, A. and Sun, Z. 'Impact of sensor ageing on iris recognition'. *IEEE International Joint Conference on Biometrics*: IEEE, 1-8.

Bondi, L., Baroffio, L., Güera, D., Bestagini, P., Delp, E. J. and Tubaro, S. (2016) 'First steps toward camera model identification with convolutional neural networks', *IEEE Signal Processing Letters*, 24(3), pp. 259-263.

Borole, M. and Kolhe, S. 'Statistical Feature Based Digital Camera Identification'. *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*: IEEE, 1019-1023.

Casey, E. (2009) *Handbook of digital forensics and investigation.* Academic Press.

Casey, E. (2020) 'Standardization of forming and expressing preliminary evaluative opinions on digital evidence', *Forensic Science International: Digital Investigation*, 32, pp. 200888.

Çeliktutan, O., Sankur, B. and Avcibas, I. (2008) 'Blind identification of source cell-phone model', *IEEE Trans. Information Forensics and Security*, 3(3), pp. 553-566.

Chan, C.-H. 2009. Dead pixel real-time detection method for image. Google Patents.

Chapman, G. H., Leung, J., Koren, I. and Koren, Z. 'Tradeoffs in imager design with respect to pixel defect rates'. *2010 IEEE 25th International Symposium on Defect and Fault Tolerance in VLSI Systems*: IEEE, 231-239.

Chapman, G. H., Leung, J., Namburete, A., Koren, I. and Koren, Z. 'Predicting pixel defect rates based on image sensor parameters'. *2011 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems*: IEEE, 408-416.

Chapman, G. H., Leung, J., Thomas, R., Namburete, A., Koren, Z. and Koren, I. 'Projecting the rate of in-field pixel defects based on pixel size, sensor area, and ISO'. *Sensors, Cameras, and Systems for Industrial and Scientific Applications XIII*: International Society for Optics and Photonics, 82980E.

Chapman, G. H., Thomas, R., Koren, I. and Koren, Z. (2017) 'Hot Pixel Behavior as Pixel Size Reduces to 1 micron', *Electronic Imaging*, 2017(11), pp. 39-45.

Chapman, G. H., Thomas, R., Koren, Z. and Koren, I. 'Empirical formula for rates of hot pixel defects based on pixel size, sensor area, and ISO'. *Sensors, Cameras, and Systems for Industrial and Scientific Applications XIV*: International Society for Optics and Photonics, 86590C.

Chapman, G. H., Thomas, R., Meneses, K. J., Koren, I. and Koren, Z. (2019) 'Image degradation from hot pixel defects with pixel size shrinkage', *Electronic Imaging*, 2019(9), pp. 359-1-359-7.

Chapman, G. H., Thomas, R., Meneses, K. J., Purbakht, P., Koren, I. and Koren, Z. (2018) 'Exploring Hot Pixel Characteristics for 7 to 1.3 micron Pixels', *Electronic Imaging*, 2018(11), pp. 401-1-401-6.

Chen, C. and Stamm, M. C. 'Camera model identification framework using an ensemble of demosaicing features'. *2015 IEEE international workshop on information forensics and security (WIFS)*: IEEE, 1-6.

Chen, M., Fridrich, J., Goljan, M. and Lukás, J. (2008) 'Determining image origin and integrity using sensor noise', *IEEE Transactions on information forensics and security*, 3(1), pp. 74-90.

Chen, S.-H. and Hsu, C.-T. 'Source camera identification based on camera gain histogram'. *2007 IEEE International Conference on Image Processing*: IEEE, IV-429-IV-432.

Cho, C.-Y., Chen, T.-M., Wang, W.-S. and Liu, C.-N. 'Real-time photo sensor dead pixel detection for embedded devices'. *2011 International Conference on Digital Image Computing: Techniques and Applications*: IEEE, 164-169.

Dabov, K., Foi, A., Katkovnik, V. and Egiazarian, K. (2007) 'Image denoising by sparse 3-D transform-domain collaborative filtering', *IEEE Transactions on image processing*, 16(8), pp. 2080-2095.

Dang-Nguyen, D.-T., Pasquini, C., Conotter, V. and Boato, G. 'Raise: A raw images dataset for digital image forensics'. *Proceedings of the 6th ACM multimedia systems conference*, 219-224.

Danielyan, A., Katkovnik, V. and Egiazarian, K. (2011) 'BM3D frames and variational image deblurring', *IEEE Transactions on Image Processing*, 21(4), pp. 1715-1728.

Deng, Z., Gijsenij, A. and Zhang, J. 'Source camera identification using auto-white balance approximation'. *2011 International Conference on Computer Vision*: IEEE, 57-64.

Dirik, A. E., Sencar, H. T. and Memon, N. (2008) 'Digital single lens reflex camera identification from traces of sensor dust', *IEEE Transactions on Information Forensics and Security*, 3(3), pp. 539-552.

Dong, S., Sun, X., Xie, S. and Wang, M. (2019) 'Automatic defect identification technology of digital image of pipeline weld', *Natural Gas Industry B*, 6(4), pp. 399-403.

Dudas, J., Wu, L. M., Jung, C., Chapman, G. H., Koren, Z. and Koren, I. 'Identification of in-field defect development in digital image sensors'. *Digital Photography III*: International Society for Optics and Photonics, 65020Y.

El-Yamany, N. (2017) 'Robust Defect Pixel Detection and Correction for Bayer Imaging Systems', *Electronic Imaging*, 2017(15), pp. 46-51.

Farid, H. (2006) 'Digital image ballistics from JPEG quantization'.

Farid, H. (2019) 'Image forensics', *Annual review of vision science*.

Fernando, B., Muselet, D., Khan, R. and Tuytelaars, T. 'Color features for dating historical color images'. *2014 IEEE International Conference on Image Processing (ICIP)*: IEEE, 2589-2593.

Filler, T., Fridrich, J. and Goljan, M. 'Using sensor pattern noise for camera model identification'. *2008 15th IEEE International Conference on Image Processing*: IEEE, 1296-1299.

Flickr (2018) 'Flickr 'Find your inspiration''.

Focus, F. (2020) 'Timelines In Digital Forensic Investigation: From Investigation To Court'.

Forcina, A. and Carbone, P. (2020) 'Modelling dark current and hot pixels in imaging sensors', *Statistical Modelling*, 20(1), pp. 30-41.

Freire-Obregón, D., Narducci, F., Barra, S. and Castrillón-Santana, M. (2019) 'Deep learning for source camera identification on mobile devices', *Pattern Recognition Letters*, 126, pp. 86-91.

Fridrich, J. and Goljan, M. 'Determining approximate age of digital images using sensor defects'. *Media Watermarking, Security, and Forensics III*: International Society for Optics and Photonics, 788006.

Galdi, C., Hartung, F. and Dugelay, J.-L. 'SOCRatES: A Database of Realistic Data for SOurce Camera REcognition on Smartphones'. *ICPRAM*, 648-655.

Gao, T. and Koller, D. 'Multiclass boosting with hinge loss based on output coding'. *ICML*.

Geradts, Z. and Bijhold, J. (2000) 'Overview of pattern recognition and image processing in forensic science', *Anil Aggrawal's Internet Journal of Forensic Medicine and Toxicology*, 1(2).

Geradts, Z. J., Bijhold, J., Kieft, M., Kurosawa, K., Kuroki, K. and Saitoh, N. 'Methods for identification of images acquired with digital cameras'. *Enabling technologies for law enforcement and security*: International Society for Optics and Photonics, 505-512.

Ginosar, S., Rakelly, K., Sachs, S., Yin, B. and Efros, A. A. 'A century of portraits: A visual historical record of american high school yearbooks'. *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 1-7.

Gloe, T. 'Forensic analysis of ordered data structures on the example of JPEG files'. *2012 IEEE International Workshop on Information Forensics and Security (WIFS)*: IEEE, 139-144.

Gloe, T. and Böhme, R. 'The'Dresden Image Database'for benchmarking digital image forensics'. *Proceedings of the 2010 ACM Symposium on Applied Computing*: Acm, 1584-1590.

Goljan, M. 'Digital camera identification from images–estimating false acceptance probability'. *International workshop on digital watermarking*: Springer, 454-468.

Guan, H., Kozak, M., Robertson, E., Lee, Y., Yates, A. N., Delgado, A., Zhou, D., Kheyrkhah, T., Smith, J. and Fiscus, J. 'MFC datasets: Large-scale benchmark datasets for media forensic challenge evaluation'. *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*: IEEE, 63-72.

Gupta, B. and Tiwari, M. (2018) 'Improving source camera identification performance using DCT based image frequency components dependent sensor pattern noise extraction method', *Digital Investigation*, 24, pp. 121-127.

He, S., Samara, P., Burgers, J. and Schomaker, L. (2016) 'Image-based historical manuscript dating using contour and stroke fragments', *Pattern Recognition*, 58, pp. 159-171.

Hosler, B. C., Zhao, X., Mayer, O., Chen, C., Shackleford, J. A. and Stamm, M. C. (2019) 'The video authentication and camera identification database: A new database for video forensics', *IEEE Access*, 7, pp. 76937-76948.

Huang, N., He, J., Zhu, N., Xuan, X., Liu, G. and Chang, C. (2018) 'Identification of the source camera of images based on convolutional neural network', *Digital Investigation*, 26, pp. 72-80.

Jae Lee, Y., Efros, A. A. and Hebert, M. 'Style-aware mid-level representation for discovering visual connections in space and time'. *Proceedings of the IEEE international conference on computer vision*, 1857-1864.

Joachims, T. (2002) *Learning to classify text using support vector machines.* Springer Science & Business Media.

Julliand, T., Nozick, V. and Talbot, H. 'Image noise and digital image forensics'. *International Workshop on Digital Watermarking*: Springer, 3-17.

Kakar, P. and Sudha, N. (2012) 'Verifying temporal data in geotagged images via sun azimuth estimation', *IEEE Transactions on Information Forensics and Security*, 7(3), pp. 1029-1039.

Kang, X., Chen, J., Lin, K. and Anjie, P. (2014) 'A context-adaptive SPN predictor for trustworthy source camera identification', *EURASIP Journal on Image and video Processing*, 2014(1), pp. 19.

Kang, X., Li, Y., Qu, Z. and Huang, J. 'Enhancing ROC performance of trustworthy camera source identification'. *Media Watermarking, Security, and Forensics III*: International Society for Optics and Photonics, 788009.

Kang, X., Li, Y., Qu, Z. and Huang, J. (2011b) 'Enhancing source camera identification performance with a camera reference phase sensor pattern noise', *IEEE Transactions on Information Forensics and Security*, 7(2), pp. 393-402.

Katkovnik, V., Danielyan, A. and Egiazarian, K. 'Decoupled inverse and denoising for image deblurring: variational BM3D-frame technique'. *2011 18th IEEE International Conference on Image Processing*: IEEE, 3453-3456.

Kauba, C. and Uhl, A. 'Prnu-based image alignment for defective pixel detection'. *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*: IEEE, 1-7.

Kee, E. and Farid, H. 'Digital image authentication from thumbnails'. *Media Forensics and Security II*: International Society for Optics and Photonics, 75410E.

Kharrazi, M., Sencar, H. T. and Memon, N. 'Blind source camera identification'. *2004 International Conference on Image Processing, 2004. ICIP'04.*: IEEE, 709-712.

Kirchner, M. and Böhme, R. 'Synthesis of color filter array pattern in digital images'. *Media Forensics and Security*: International Society for Optics and Photonics, 72540K.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. 'Imagenet classification with deep convolutional neural networks'. *Advances in neural information processing systems*, 1097-1105.

Kurosawa, K., Kuroki, K. and Saitoh, N. 'CCD fingerprint method-identification of a video camera from videotaped images'. *Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348)*: IEEE, 537-540.

Lawgaly, A. and Khelifi, F. (2016) 'Sensor pattern noise estimation based on improved locally adaptive DCT filtering and weighted averaging for source camera identification and verification', *IEEE Transactions on Information Forensics and Security*, 12(2), pp. 392-404.

Lawgaly, A., Khelifi, F. and Bouridane, A. 'Image sharpening for efficient source camera identification based on sensor pattern noise estimation'. *2013 Fourth International Conference on Emerging Security Technologies*: IEEE, 113-116.

Lawgaly, A., Khelifi, F. and Bouridane, A. 'Weighted averaging-based sensor pattern noise estimation for source camera identification'. *2014 IEEE International Conference on Image Processing (ICIP)*: IEEE, 5357-5361.

Lee, S., Maisonneuve, N., Crandall, D., Efros, A. A. and Sivic, J. 'Linking past to present: Discovering style in two centuries of architecture'. *IEEE International Conference on Computational Photography*.

Leung, J., Chapman, G. H., Choi, Y. H., Thomas, R., Koren, I. and Koren, Z. 'Analyzing the impact of ISO on digital imager defects with an automatic defect trace algorithm'. *Sensors, Cameras, and Systems for Industrial/Scientific Applications XI*: International Society for Optics and Photonics, 75360F.

Leung, J., Chapman, G. H., Koren, I. and Koren, Z. 'Automatic detection of in-field defect growth in image sensors'. *2008 IEEE International Symposium on Defect and Fault Tolerance of VLSI Systems*: IEEE, 305-313.

Leung, J., Chapman, G. H., Koren, Z. and Koren, I. 'Statistical identification and analysis of defect development in digital imagers'. *Digital Photography V*: International Society for Optics and Photonics, 72500W.

Leung, J., Dudas, J., Chapman, G. H., Koren, I. and Koren, Z. 'Quantitative analysis of in-field defects in image sensor arrays'. *22nd IEEE International Symposium on Defect and Fault-Tolerance in VLSI Systems (DFT 2007)*: IEEE, 526-534.

Leung, J., Dudas, J., Chapman, G. H., Koren, Z. and Koren, I. 'Characterization of pixel defect development during digital imager lifetime'. *Sensors, Cameras, and Systems for Industrial/Scientific Applications IX*: International Society for Optics and Photonics, 68160A.

Li, C.-T. (2010) 'Source camera identification using enhanced sensor pattern noise', *IEEE Transactions on Information Forensics and Security*, 5(2), pp. 280-287.

Li, C.-T. and Li, Y. (2011) 'Color-decoupled photo response non-uniformity for digital image forensics', *IEEE Transactions on Circuits and Systems for Video Technology*, 22(2), pp. 260-271.

Li, J., Li, X. and Yao, X. 'Cost-sensitive classification with genetic programming'. *2005 IEEE congress on evolutionary computation*: IEEE, 2114-2121.

Li, R., Guan, Y. and Li, C.-T. 'PCA-based denoising of sensor pattern noise for source camera identification'. *2014 IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP)*: IEEE, 436-440.

Li, R., Li, C.-T. and Guan, Y. (2018) 'Inference of a compact representation of sensor fingerprint for source camera identification', *Pattern Recognition*, 74, pp. 556-567.

Li, X., Xu, W., Wang, S. and Qu, X. 'Are you lying: Validating the time-location of outdoor images'. *International Conference on Applied Cryptography and Network Security*: Springer, 103-123.

Li, Y., Genzel, D., Fujii, Y. and Popat, A. C. 'Publication date estimation for printed historical documents using convolutional neural networks'. *Proceedings of the 3rd international workshop on historical document imaging and processing*, 99-106.

Lin, X. and Li, C.-T. (2016) 'Enhancing sensor pattern noise via filtering distortion removal', *IEEE Signal Processing Letters*, 23(3), pp. 381-385.

Lin, X., Li, J.-H., Wang, S.-L., Cheng, F. and Huang, X.-S. (2018) 'Recent advances in passive digital image security forensics: A brief review', *Engineering*, 4(1), pp. 29-39.

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

Mao, J., Bulan, O., Sharma, G. and Datta, S. 'Device temporal forensics: An information theoretic approach'. *2009 16th IEEE International Conference on Image Processing (ICIP)*: IEEE, 1501-1504.

Martin, P., Doucet, A. and Jurie, F. 'Dating color images with ordinal classification'. *Proceedings of International Conference on Multimedia Retrieval*: ACM, 447.

Mehrish, A., Subramanyam, A. V. and Emmanuel, S. (2018) 'Robust PRNU estimation from probabilistic raw measurements', *Signal Processing: Image Communication*, 66, pp. 30-41.

Mihcak, M. K., Kozintsev, I., Ramchandran, K. and Moulin, P. (1999) 'Low-complexity image denoising based on statistical modeling of wavelet coefficients', *IEEE Signal Processing Letters*, 6(12), pp. 300-303.

Milani, S., Tagliasacchi, M. and Tubaro, S. (2014) 'Discriminating multiple JPEG compressions using first digit features', *APSIPA Transactions on Signal and Information Processing*, 3.

Min, F., Liu, F.-L., Wen, L.-Y. and Zhang, Z.-H. (2019) 'Tri-partition cost-sensitive active learning through kNN', *Soft Computing*, 23(5), pp. 1557-1572.

Müller, E., Springstein, M. and Ewerth, R. '"When was this picture taken?"–Image date estimation in the wild'. *European Conference on Information Retrieval*: Springer, 619-625.

Nair, V. and Hinton, G. E. 'Rectified linear units improve restricted boltzmann machines'. *Proceedings of the 27th international conference on machine learning (ICML-10)*, 807-814.

Padilha, R., Andaló, F. A., Lavi, B., Pereira, L. A. and Rocha, A. (2021a) 'Temporally sorting images from real-world events', *Pattern Recognition Letters,* 147, pp. 212-219.

Padilha, R., Salem, T., Workman, S., Andaló, F. A., Rocha, A. and Jacobs, N. (2021b) 'Content-Based Detection of Temporal Metadata Manipulation', *arXiv preprint arXiv:2103.04736*.

Palermo, F., Hays, J. and Efros, A. A. 'Dating historical color images'. *European Conference on Computer Vision*: Springer, 499-512.

Qadir, A. M. and Varol, A. 'The role of machine learning in digital forensics'. *2020 8th International Symposium on Digital Forensics and Security (ISDFS)*: IEEE, 1-5.

Qadir, G., Yahaya, S. and Ho, A. T. (2012) 'Surrey university library for forensic analysis (SULFA) of video content'.

Quan, Y., Li, C.-T., Zhou, Y. and Li, L. 'Warwick image forensics dataset for device fingerprinting in multimedia forensics'. *2020 IEEE International Conference on Multimedia and Expo (ICME)*: IEEE, 1-6.

Redi, J. A., Taktak, W. and Dugelay, J.-L. (2011) 'Digital image forensics: a booklet for beginners', *Multimedia Tools and Applications,* 51(1), pp. 133-162.

Ruder, S. (2016) 'An overview of gradient descent optimization algorithms', *arXiv preprint arXiv:1609.04747*.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986) 'Learning representations by back-propagating errors', *nature,* 323(6088), pp. 533-536.

Ryser, E., Spichiger, H. and Casey, E. (2020) 'Structured decision making in investigations involving digital and multimedia evidence', *Forensic Science International: Digital Investigation,* 34, pp. 301015.

Salem, T., Workman, S., Zhai, M. and Jacobs, N. 'Analyzing human appearance as a cue for dating images'. *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*: IEEE, 1-8.

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

Schaefer, G. and Stich, M. 'UCID: An uncompressed color image database'. *Storage and Retrieval Methods and Applications for Multimedia 2004*: International Society for Optics and Photonics, 472-480.

Schindler, G., Dellaert, F. and Kang, S. B. 'Inferring temporal order of images from 3D structure'. *2007 IEEE Conference on Computer Vision and Pattern Recognition*: IEEE, 1-7.

Shi, Z., Hao, H., Zhao, M., Feng, Y., He, L., Wang, Y. and Suzuki, K. (2019) 'A deep CNN based transfer learning method for false positive reduction', *Multimedia Tools and Applications*, 78(1), pp. 1017-1033.

Shullani, D., Fontani, M., Iuliani, M., Al Shaya, O. and Piva, A. (2017) 'VISION: a video and image dataset for source identification', *EURASIP Journal on Information Security*, 2017(1), pp. 1-16.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014) 'Dropout: a simple way to prevent neural networks from overfitting', *The journal of machine learning research*, 15(1), pp. 1929-1958.

Swaminathan, A., Wu, M. and Liu, K. R. (2008) 'Digital image forensics via intrinsic fingerprints', *IEEE transactions on information forensics and security*, 3(1), pp. 101-117.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. 'Going deeper with convolutions'. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1-9.

Tchendjou, G. T. and Simeu, E. (2020) 'Detection, Location and Concealment of Defective Pixels in Image Sensors', *IEEE Transactions on Emerging Topics in Computing*.

Theuwissen, A. J. (2007) 'Influence of terrestrial cosmic rays on the reliability of CCD image sensors—Part 1: Experiments at room temperature', *IEEE Transactions on Electron Devices*, 54(12), pp. 3260-3266.

Tian, H., Xiao, Y., Cao, G., Zhang, Y., Xu, Z. and Zhao, Y. (2019) 'Daxing smartphone identification dataset', *IEEE Access*, 7, pp. 101046-101053.

Tiwari, M. and Gupta, B. (2018) 'Image features dependant correlation-weighting function for efficient PRNU based source camera identification', *Forensic science international*, 285, pp. 111-120.

Tsai, M.-J. and Wu, G.-H. 'Using image features to identify camera sources'. *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*: IEEE, II-II.

Tuama, A., Comby, F. and Chaumont, M. 'Camera model identification with the use of deep convolutional neural networks'. *2016 IEEE International workshop on information forensics and security (WIFS)*: IEEE, 1-6.

Van, L. T., Emmanuel, S. and Kankanhalli, M. S. 'Identifying source cell phone using chromatic aberration'. *2007 IEEE International Conference on Multimedia and Expo*: IEEE, 883-886.

Vittayakorn, S., Berg, A. C. and Berg, T. L. 'When was that made?'. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*: IEEE, 715-724.

Wang, W., Zhu, T., Fu, Y.-t. and Dong, F. 'A new blind-pixel detection method for 384× 288 long-wave infrared focal plane arrays images'. *9th International Symposium on Advanced Optical Manufacturing and Testing Technologies: Optoelectronic Materials and Devices for Sensing and Imaging*: International Society for Optics and Photonics, 108431T.

Wu, G., Kang, X. and Liu, K. R. 'A context adaptive predictor of sensor pattern noise for camera source identification'. *2012 19th IEEE International Conference on Image Processing*: IEEE, 237-240.

Xu, B., Wang, X., Zhou, X., Xi, J. and Wang, S. (2016) 'Source camera identification from image texture features', *Neurocomputing*, 207, pp. 131-140.