Analysis of scanned documents for integrity and authenticity checking

Chaima Ben Rabah

DOCTORAT BRETAGNE LOIRE / MATHSTIC

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

SUP'COM
Ecole Supérieure des Communications de Tunis

THÈSE DE DOCTORAT DE

L'ÉCOLE NATIONALE SUPÉRIEURE
MINES-TÉLÉCOM ATLANTIQUE BRETAGNE
PAYS-DE-LA-LOIRE - IMT ATLANTIQUE
DÉLIVRÉE CONJOINTEMENT AVEC
L'ÉCOLE SUPERIEURE DES COMMUNICATIONS DE TUNIS

ÉCOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Signal, image, vision*

Par
Chaima BEN RABAH

Analysis of scanned documents for integrity and authenticity checking

Thèse présentée et soutenue à Brest, le 16/12/2021
Unité de recherche : ITI - COSIM
Thèse N° : 2021IMTA0276

Rapporteurs avant soutenance :

Imed RIADH FARAH	Professeur, ISAMM
Olivier STRAUSS	Professeur, Université Montpellier II

Composition du Jury :

Président :	Laurent NANA	Professeur, UBO
Examinateurs :	Sadok EL ASMI	Professeur, SUPCOM
	Imed RIADH FARAH	Professeur, ISAMM
	Olivier STRAUSS	Professeur, Université Montpellier II
	Johanne VINCENT	Maître de conférences, IMT Atlantique
Dir. de thèse :	Gouenou COATRIEUX	Professeur, IMT Atlantique
Co-dir. de thèse :	Riadh ABDELFATTAH	Professeur, SUPCOM

To my Mom, my Dad, and my little sister,

To my husband and my daughter,

To the memory of my beloved friend Abir Yahyaoui (1988-2018),

for their unlimited support, encouragement, and love.

Acknowledgements

First and foremost, I'd like to express my heartfelt gratitude to my supervisors Prof. Gouenou Coatrieux and Prof. Riadh Abdelfattah, who have been constantly supportive and inspiring during my Ph.D. studies. Their great personality, guidance, unlimited patience, and tolerance have educated me a lot more than scientific research. They transmitted to me the high motivation and the commitment to work. I owe them my eternal thankfulness.
I am extremely thankful to Prof. Imed Riadh Farah, and Prof. Olivier Strauss, for taking their precious time to review my thesis manuscript. I also would like to gratefully thank Prof. Laurent Nana, Prof. Sadok El Asmi, and Dr. Johanne Vincent, for being examiners on my Ph.D. defense committee. Valuable remarks provided by respectful experts in this field like them would improve the thesis's quality.
To the person who contributed directly to the completion of my thesis, foremost, I would like to express my deepest gratitude to Prof. Basel Solaiman for his valuable support. He believed in me and pointed me in the right direction. I would not have been able to complete this Ph.D. without his devoted and precious guidance.
I am greatly indebted to Dr. Wajih Ben Abdallah who has assisted me with high efficiency and availability and has accompanied me during my first year of Ph.D.
A warm thank goes to Prof. Sofiane Cherif, Prof. Fethi Tlili, and all the other SUPCOM members for supporting me in difficult moments and enjoying the good ones. I would like to deeply thank Prof. Fatma Abdelkafi and all my teachers Prof. Amel Ben Azza, Prof. Sofia Ben Jbara, Prof. Rached Hamza, Prof. Nabile Tabbane, M. Slah Bouhari, Dr. Asma Ben Ltaifa.. for their unwavering support which pushed me to achieve what I never thought possible.
I want to thank my colleagues at SUPCOM, Yosra, Emna, and Olfa for creating such a crazy and funny work atmosphere. It was the best I could hope for, especially given the stressful circumstances. I want also to thank the members of the ITI department, Corinne, Sahar, Kamilia, Yutong, Zahran, David, and Anas for the friendly and adorable environment they offered to me every time I stayed in Brest.
The work presented in this thesis is supported by the CMCU project number 17G1405, which is funded by the French Ministry of Foreign Affairs and Ministry of Higher Education and Research, as well as the Tunisian Ministry of Higher Education and Scientific Research. Thus, I would like to express my gratitude to them for the financial support, without which the SUPATLANTIQUE dataset would not exist. I am indebted also to my friend Khaled El Hmedi for designing its logo.
Finally, I would like to eternally thank my family. I owe them my endless love and gratitude. Their support, understanding, and encouragement in every single moment of my life have been and always will be the most important thing which gives me the strength to chase whatever I want.
To my husband, Mr. Jabrane Ben Othman for being by my side, and for all the inspirations and love he brings to my life.
Last but not least, to my daughter, Amel, for making my graduate life more bearable and full of fun.

Abstract

There is no universal method for detecting counterfeit images. Several techniques have been proposed but each has its limits. Among these methods, the so-called "digital image forensics" techniques offer an interesting solution. They aim either to verify image integrity or to provide proof of its authenticity by identifying the system which acquired it. To do this, they take advantage of how the acquisition systems generate their output.

In this thesis, we are particularly interested in flatbed scanners as an acquisition system and we propose to study and develop "digital image Forensics" techniques for multi-type scanned documents and suggest, based on these tools, a secure and suitable environment for fighting against any type of falsification of important digitized documents; a flexible environment offering trusted transaction services to individuals or legal entities.

The first step is to identify characteristics extracted from images, or a combination of them, which offers a good compromise in terms of ability to discriminate a scanner and also of robustness while being as generic as possible, that is to say, transposable from one scanner to another. We first proposed a technique to identify the brand of the scanner that is at the origin of a scanned JPEG document from its header. Next, we were interested in TIFF images. A first type of approach was developed on the basis of a set of signatures extracted from images and characteristic of a scanner. The principle of the first method is based on the statistical properties of the coefficients of the high frequency sub-bands of the wavelet transform. In the case of administrative papers and where the reference document is accessible, we offer another method which, taking advantage of this a priori knowledge, makes it possible to remove as much of the irrelevant content as possible and better isolate the noise specific to a scanner. This noise, when averaged according to the scanning direction, will serve as the signature of a scanned document. The second type of approach has focused on the automatic extraction of the scanners' fingerprints through 1D and 2D neural networks. These approaches have been shown to be more effective in distinguishing scanners of the same model and do not require large numbers of scanned images for training.

Next, we propose a new approach, called "Device Linking", which determines whether two images were acquired by the same scanner or not. This approach is different from other forensic approaches in that its purpose is not to identify the make or model of the source scanner, but rather to determine if the residual noises in the two images are similar. One of the advantages of this approach is that knowledge of the scanner models of the investigated images is not required.

Finally, as a continuation of these works, we provide two security mechanisms capable of detecting content manipulation of scanned data with a minimum error rate, based on certain approaches of source scanner identification proposed previously. In order to validate the solutions proposed in real situations and make comparisons between them, we have built a database of scanned documents that we have made public.

Keywords: Digital image forensics, image integrity, authenticity, flatbed scanners, falsification, digitized documents, security mechanism.

Résumé

Aucune méthode universelle de détection des contrefaçons d'images n'existe. Plusieurs techniques ont été proposées mais chacune a ses limites. Parmi celles-ci, les techniques dites "forensiques des images numériques" offrent une solution intéressante. Elles visent soit à vérifier l'intégrité de l'image, soit à apporter la preuve de son authenticité en identifiant le système qui l'a acquise. Pour ce faire, elles tirent avantage de la manière dont les systèmes d'acquisition produisent leurs données. Dans cette thèse, nous nous intéressons en particulier aux scanners à plat comme système d'acquisition et nous proposons d'étudier et de développer des techniques de "forensiques des images numériques" pour des documents multi-type scannés et suggérer, sur la base de ces outils, un environnement sécurisé et adapté pour la lutte contre tout type de falsification de documents d'importance numérisés ; un environnement flexible proposant des services de transactions de confiance à des personnes ou des entités morales.

Le premier pas consiste à identifier des caractéristiques extraites des images, ou une combinaison d'entre elles, qui offre un bon compromis en termes de capacité à discriminer un scanner et aussi de robustesse, tout en étant le plus générique possible, c'est-à-dire transposable d'un scanner à un autre. Nous avons, tout d'abord, proposé une technique permettant d'identifier la marque du scanner à l'origine d'un document scanné JPEG à partir de son entête. Ensuite, nous nous sommes intéressés à des images TIFF. Un premier type d'approches a été développé sur la base d'un ensemble de signatures extrait des images et caractéristique d'un scanner. Le principe de la première méthode est fondé sur les propriétés statistiques des coefficients des sous-bandes de hautes fréquences de la transformée en ondelettes. Dans le cas de papiers administratifs et où le document de référence est accessible, nous offrons une autre méthode qui, profitant de cette connaissance a priori, permet de supprimer le maximum du contenu non pertinent et mieux isoler le bruit propre à un scanner. Ce bruit, une fois moyenné selon la direction de numérisation, servira comme signature d'un document numérisé. Le deuxième type d'approches a focalisé sur l'extraction automatique des signatures des scanners à travers des réseaux de neurones 1D et 2D. Ces approches se sont avérées plus efficaces dans la distinction des scanners de même modèles et ne nécessitent pas des images numérisées de grand nombre pour l'apprentissage.

Ensuite, nous proposons une nouvelle approche, appelée "Device Linking", qui détermine si deux images ont été acquises par le même scanner ou non. Cette approche est différente des autres approches forensiques vu que son but n'est pas d'identifier la marque ou le modèle du scanner source, mais plutôt déterminer si les bruits résiduels dans les deux images sont similaires. L'un des avantages de cette approche est que la connaissance des modèles de scanner des images étudiées n'est pas requise.

Enfin, dans la continuité de ces travaux, nous fournissons deux mécanismes de sécurité capable de détecter les manipulations du contenu des données numérisées avec un minimum de taux d'erreur en se basant sur certaines approches d'identification du scanner source proposées précédemment. Dans le but de valider les solutions proposées dans des situations réelles et réaliser des comparaisons entre elles, nous avons construit une base de données de documents scannés que nous avons rendue publique.

Mots clés : Forensiques des images numériques, intégrité de l'image, authenticité, scanners à plat, falsification, documents numérisés, mécanisme de sécurité.

Contents

List of Figures

List of Tables

List of Abbreviations

1D-AR . . . . One-dimensional AutoRegressive
1D-CNN . . . One Dimensional Convolutional Neural Network
2D-AR . . . . Two-dimensional AutoRegressive
2D-CNN . . . Two Dimensional Convolutional Neural Network
ADC . . . . . Analog-to-Digital Converter
BDCT . . . . . Block Discrete Cosine Transform
CBMA . . . . Classical Block Based Matching Algorithm
CCD . . . . . Charge-Coupled Device
CCFL . . . . . Cold Cathode Fluorescent Lamp
CDS . . . . . Correlated Double Sampling
CE . . . . . Contrast Enhancement
CFA . . . . . Color Filter Array
CIFD . . . . . CNN-based IFD approach
CIS . . . . . Contact Image Sensor
CMDF . . . . Copy-Move Forgery Detection
DCF . . . . . Digital Content Forensics
DPI . . . . . Dots Per Inch
DPCM . . . . Differential Pulse Code Modulation
DSNU . . . . . Dark Signal Non-Uniformity
DCT . . . . . Discrete Cosine Transform
DIF . . . . . Digital Image Forensics
DL . . . . . Deep Learning
DLK . . . . . Device Linking
DWT . . . . . Discrete Wavelet Transform
EM . . . . . Expectation Maximization
EMD . . . . . Empirical Mode Decomposition
EME . . . . . Entropy Matching Estimation
EOI . . . . . End-Of-Image
EXIF . . . . . Exchangeable Image File
FPN . . . . . Fixed Pattern Noise
FT . . . . . Fourier Transform
g2NN . . . . . Generalized 2-nearest Neighbor
GLF . . . . . Global and Local Feature

GLRLM . . . . Grey-Level Run Length Matrix
GGD . . . . . Generalized Gaussian Distribution
GLCM . . . . Gray-Level Co-occurrence Matrix
GLDH . . . . . Gray-Level Difference Histogram
HIFD . . . . . Handcrafted-based IFD approach
HM . . . . . Heat Map
HOGM . . . . Histogram off Orientated Gabor Magnitude
HPF . . . . . High Pass Filter
HPS . . . . . Hard Parameter Sharing
IBMA . . . . . Improved Block-based Matching Algorithm
IFD . . . . . Image Forgery Detection
IFT . . . . . Inverse Fourier Transform
JPEG . . . . . Join Photographic Expert Group
LBP . . . . . Local Binary Pattern
LED . . . . . Light Emission Diodes
LDA . . . . . Linear Discriminant Analysis
LDD . . . . . Local Difference Descriptor
LDF . . . . . Linear Discriminant Function
LPA-ICI . . . Local Polynomial Approximation—Intersection of Confidence Intervals
LPF . . . . . Low Pass Filter
LSH . . . . . Locality Sensitive Hashing
LTP . . . . . Local Ternary Pattern
LTrP . . . . . Local Tetra Pattern
ME . . . . . Moment Estimation
MF . . . . . Median Filtering
MFF . . . . . Median Filtering Forensics
MLE . . . . . Maximum Likelihood Estimation
MLDD . . . . MultiLevel Dense Descriptor
MLP . . . . . Multi-Layer Perceptron
MROGH . . . Multi-Support Region Order-based Gradient Histogram
MSCR . . . . . Maximally Stable Color Region
MZMs . . . . . Modified Zernike Moments
NCC . . . . . Normalized Cross Correlation
NN . . . . . Neural Network
NLP . . . . . Natural Language Processing
ORB . . . . . Oriented Fast and Robust Brief
PCET . . . . . Polar Complex Exponential Transform
PNG . . . . . Portable Network Graphics
PPI . . . . . Pixels Per Inch

PRNU . . . . . Photo-Response Non-Uniformity
PST . . . . . Polar Sine Transform
QDCT . . . . . Quaternion Discrete Cosine Transforms
QT . . . . . Quantization Table
QWT . . . . . Quaternion Wavelet Transform
RLE . . . . . Run Length Encode
RNC . . . . . Réseau de Neurones Convolutif
ROI . . . . . Regions of Interest
SCI . . . . . Source Camera Identification
SIFT . . . . . Scale Invariant Feature Transform
SOI . . . . . Start-Of-Image
SPAM . . . . . Subtractive Pixel Adjacent Matrix
SPT . . . . . Steerable Pyramid Transform
SPS . . . . . Soft Parameter Sharing
SSI . . . . . Source scanner identification
SURF . . . . . Speed Up Robust Feature
SVD . . . . . Singular Value Decomposition
SVM . . . . . Support Vector Machine
SWT . . . . . Stationary Wavelet Transform
TIFF . . . . . Tagged Image File Format
TL . . . . . Transfer Learning

I've been saying for years we're gonna have to spend a lot more time on cybersecurity.

Je dis depuis des années que nous allons devoir passer beaucoup plus de temps sur la cybersécurité.

— Barack *Obama*

1 Introduction

Contents

1.1 General context and problem statement

The evolution of information and communications technologies are at the origin of major societal changes such as the dematerialization of many services and administrative documents (state - civil status, banking - account statements, notarial acts etc.). However, this transition to a fully digital world did not completely suppressed the use of paper documents: those that are/were processed in paper form before dematerialization such as the medical records of a health establishment or documents for which the regulations require passing through a paper version subsequently scanned once completed such as notarial acts. Thus, interfacing these new and old worlds involves the digitization of documents that can then be easily distributed, exchanged, collected and processed, speeding up procedures while reducing costs. Several countries have even archived their commercial, administrative and political documents in order to preserve their heritage and restore important documents. That is the case of France which has created a diplomatic archive which has nearly 180,000 documents [1]. That is also the case of Tunisia which is putting plans and strategies related to the digital transformation with the aim of modernizing the administration and guaranteeing citizens online services in line with their needs.
However, while this development makes it easier to access and process information, it also raises many security questions. For example, one might question the origin of a scanned document as well as its integrity, our interest in this thesis lie in these aspects. They are critical since documents in paper form do not include security elements (watermarks, hologram etc.) as there are for banknotes or passports.

Indeed, today, a malicious third party can quite simply falsify the content of a scanned document to his advantage using image editing software without leaving visually detectable traces [2]. This fact is causing a serious dilemma in countries like the United Kingdom (UK) which rely on importing foreign workers. According to a 2015 Guardian study on foreign labor, doctors from at least 27 nations were hired in 32 of England's 160 hospital trusts [3]. But, given the current global prevalence of document forgery (2.32% as of April 2016), it appears quite plausible that there are a considerable number of unqualified healthcare professionals working in the UK. Note that in 2020, the ministry of public health in Qatar has blacklisted 23 health practitioners: 17 doctors, two nurses, and four allied healthcare workers after it was discovered that they had provided forged certificates [4]. In fact, hundreds of fake diplomas, medical insurances and working certificates are invading daily and no industry is safe from this menace which affect not only the employment market but also the people's safety.
Different solutions and approaches have been proposed for protecting multimedia data, in terms of content. These ones can be classified into two categories: active and passive (see Fig. 1.1).

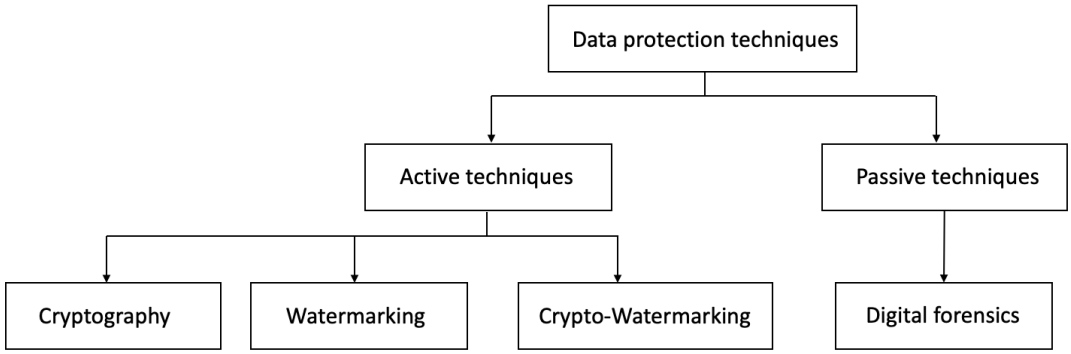


Figure 1.1: Classification of digital data protection approaches

Active techniques generally require a preprocessing step. These are cryptographic mechanisms (e.g. encryption [5], digital signatures [6]), watermarking [7] or crypto-watermarking [8]. Encryption consists of converting a clear plaintext data into a non-interpretable cipher-text data using an encryption method parameterized by encryption keys, as shown in Fig. 1.2.
While encryption is helpful for maintaining confidentiality, digital signatures will help with stored data integrity. However, those cryptographic solutions offer an a priori protection. In another word, the data is protected as long as it is encrypted or as long as its signature has not been deleted. Moreover, they constrain the data format and raise questions about sharing keys.
Herein lies the value of an a posteriori protection which is provided by watermarking. By definition, watermarking allows inserting into a document (e.g. image, video, signals, databases ...) a message, or equivalently a watermark, by means of imperceptible modifications of this latter. Taking an image as example, the watermarking process will modify the gray values of some pixels so as to encode a message. Moreover; depending on the relationship between the watermarked message and the host document, it becomes possible to know the origin of the document and/or if it has not been falsified, to embed some access right and/or granted access duration [9]. As defined, the message being inserted directly into the document, such a protection is independent of the format of storage of the data. Nevertheless, deploying this technology implies being able to watermark a

document as soon as it is produced. In our context, this requires that a watermarking module is integrated as a security component into all scanners. This is practically not feasible given the manufacturing cost and the design complexity increase of a scanner. Moreover, this operation may affect the visual quality of the digitized image while imperceptibility is an important requirement in all watermarking applications, particularly in the case of medical images.
Let us recall that a watermarking system must handle the trade-off between the capacity, robustness and imperceptibility requirements.

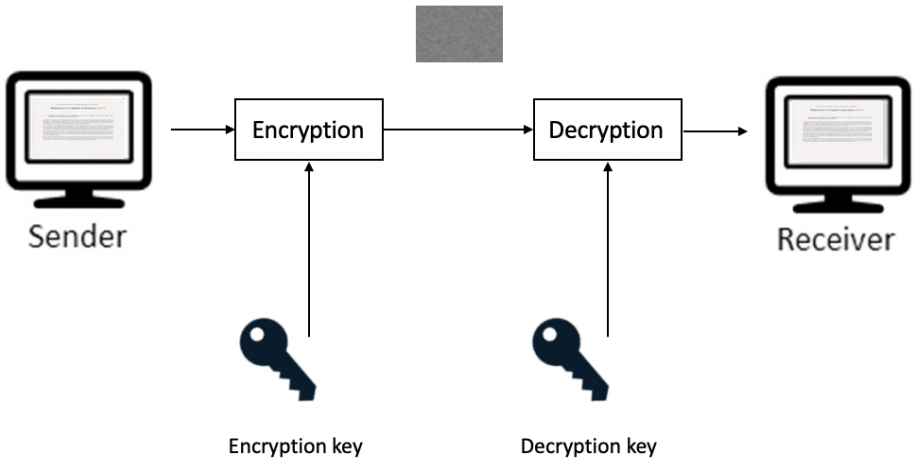


Figure 1.2: General overview of an image encryption scheme

In these perspectives, it is of paramount importance to develop solutions to ensure authenticity and integrity of digital images without relying on pre-registration or pre-embedded information. Let us recall that the authenticity brings a proof that a document is acquired by the correct source, while the purpose of integrity is to find out whether or not a document has been manipulated.
An interesting alternative is offered by the passive techniques known as "digital content forensics (DCF)" [10]. These techniques aim at verifying if an image has been modified or at providing a proof of its origin, that is to say: identifying the scanner that acquired the image. These methods work in a blind way in the sense that they do not rely on some a priori information on the data (e.g. a data signature or any other ancillary information) shared with the data to protect. When applied to images, the basic concept on which relies digital content forensics stands on the extraction of some image characteristics or features coupled or not with machine learning processes in order to constitute an image imprint that can be used by a verification process either to identify the image origins or to detect image modifications.
In the last 15 years, the number of works in the field of digital image forensics (DIF), the field of DCF devoted to digitized images, has increased considerably. However, the effectiveness of the proposed approaches remains relative and are mostly dedicated to digital cameras. On the other side, the ever-growing number of digital documents acquired every day and shared through unsecured public networks increases dramatically the number of threats. Thus, it is important to provide adequate and reliable solutions to respond to multimedia content protection issues in the digital transition and the dematerialization of data and services.
The objective of this thesis is to provide a secure and suitable environment for verifying the authenticity and detecting the falsification of important digitized documents; a flexible environment offering trusted transaction services to individuals or legal entities.

1.2 Specific objectives

Due to COVID-19 pandemic, millions of employees had to work from home for the first time which makes them an easy target for cyber criminals. Moreover, the demand for proofs of document authenticity and integrity especially in e-commerce and e-government applications has been rising faster than expected. Indeed, number of cyber-attacks has increased in the recent years leaving disastrous effects on companies and institutions as well as persons' emotional, financial, and professional life.
To counter-fight these threats, several protection and verification mechanisms have been deployed. Among them, digital image forensics has become a common practice to address data security problems. In many countries, it is used as part of the measures taken by the defense and security forces to protect their data against cyber-attacks [11]. Their strategy consists in answering a twofold question:

- Which device is at the origin of a digitized/scanned image?
- Has the content of the image been manipulated?

The first question is related to the problem of image source identification which is one of the primary goals of DIF. In fact, knowing the device that have acquired the questioned image may help verifying its integrity as well as to trace the device owner. The second question is about image integrity. The fundamental tasks of forensic analysts in this case are to detect existence of forgeries, determining the manipulation that have been performed and localizing the tampered regions within the questioned image.
Most of the research conducted in this area are not adopted to document scanners, while thousands of digitized documents are processed on a daily basis and are in danger of being stolen or forged. At the same time, existing solutions cannot claim to be successful as we will see in the following chapter. Therefore, in this thesis, we are interested in developing efficient forensics solutions to overcome information security challenges related to flatbed scanners. More specifically, the problems of source identification and forgery detection are investigated.

1.3 Contributions

To solve the first problem, we initially choose to conduct our research on documents scanned and compressed in Joint Photographic Experts Group (JPEG) owing to the fact that this format is widely used in the Internet thanks to its capability to compress images and thus saves in term of transmission and storage. We demonstrate that it is possible to use the JPEG image header for SSI. In fact, and as we will see, the scanner's fingerprint is obtained from the quantization table since we found that each scanner uses its own customized quantization matrices.
JPEG, however, is not the best choice for scanning important document such as official documents, scientific and medical images as it involves lossy and irreversible compression [12]. Moreover, once the header information got altered or lost, this solution becomes no more effective. Thus, in the rest of the thesis, we focus on the Tagged Image File Format (TIFF), a raw or lossless image format which not only preserves the quality and the clarity of the images content but also allows us to address the issues raised in this thesis by going back to the scanner acquisition chain and extracting identifiable characteristics specific to the scanner components serving as forensic evidence.
From this standpoint, we went further proposing handcrafted solutions [13–15] that can give access to the scanners traces left in the acquired images:

i) In the first solution, original handcrafted features are extracted in the wavelet domain (see appendix B) using the Discrete Wavelet Transform (DWT). In order to identify the source scanner, the Kullback–Leibler Divergence (KLD) between the features extracted from one image and those designed for each scanner is measured and the scanner that presents the lowest measure is selected. Experimental results have shown that our method offers better performance compared to the recent state-of–art method [16]. Nevertheless, giving the similarities between the manufacturing processes of scanners of the same model, it is difficult to detect near-perfect features to identify them as different sources. Indeed, one limitation of this approach is that it fails when scanners of the same make and model are used;
ii) In some cases, the original reference document (e.g. form and agreement before being filled and scanned...) is already known. For example, we may analyse a school registration form or a rental agreement. In those cases, the investigated images are scanned copies of these documents after being printed, filled and signed by someone. Considering this a priori knowledge, the second solution successfully achieved effective scanning noise extraction by employing a kind of a background subtraction to remove the maximum of image details which are at the origin of source scanner misclassification. The performance of this solution was verified along with its ability to discern similar scanners (same model). However, this can be only achieved when the reference document is available.
Due to the shortcomings of the handcrafted features extraction based solutions, we decided to work on automating the features extraction process, an original topic in scanned images forensics research. In this thesis, two following approaches [17, 18] are developed based on convolutional neural networks (CNN) that outperform the previous traditional approaches: i) Firstly, a two dimensional CNN (2D-CNN) architecture is proposed to classify images according to their sources. In order to ensure better scanner features detection and alleviate the limitations of the DWT, the Stationary Wavelet Transform (SWT) is applied to give access to its HH subband which will serve as an entry to our network and therefore reduce the effect on image content when learning features representations; ii) A more sophisticated method is proposed based on the fact that scanning is realized line by line and, thus, its noise is repeated over the rows of the digitized document. Therefore, a one dimensional CNN (1D-CNN) combined with a support vector machine (SVM) is adopted to perform an automatic learning of the scanning noise related to each scanner. In order to validate the efficiency of these methods, large-scale experiments were carried-out.
The main weaknesses of all the above mentioned methods is that they rely on the availability of the source scanner fingerprints. More clearly, if the investigated image was acquired by non-accessible scanner, it will lead to false matches. Furthermore, one may just need to compare the source similarity of a pair of images. To this end, the 2D-CNN approach proposed earlier is expanded together with a similarity score generator to check if two documents were acquired by the same scanner or not without requiring physical access to that scanner, a challenging issue known as device linking (DLK). Furthermore, it is worth noticing that DLK is a versatile strategy that can be used in a variety of practical situations such as database inconsistency verification and editing operation comparison.
In addition, we built a large-scale image dataset [19] to foster the research on scanned documents in the field of DIF and to be able to share our findings without worrying about the threat of copyright infringement. We expect that this publicly available [1] database will become a reference in this field.
In order to solve the image forgery detection (IFD) problem, the second part of this Ph.D. thesis work, we introduce two novel techniques that can not only determine if the image has been digitally forged or not, but also find the region that has been altered; the first is proposed based on a perceptual blocks' histogram comparison

[1] `https://sites.google.com/view/supatlantique-dataset/downloads`

and is inspired by the wavelet-based method introduced in Chapter 2. The second applies our 2D-CNN model to identify which blocks of the investigated image are not matching the same source of the remaining ones. Their performances are evaluated on a dozen of forged images from our public dataset.
All those goals and the techniques we developed so as to reach them are illustrated in the flowchart given in Fig. 1.3 (Please refer to the Abbreviation section for acronyms). We mentioned in purple color the chapters where each question has been answered and in blue color references to our works that have been valued via scientific publications. Our contributions are numbered in red color and we highlighted in green the systems that we developed.

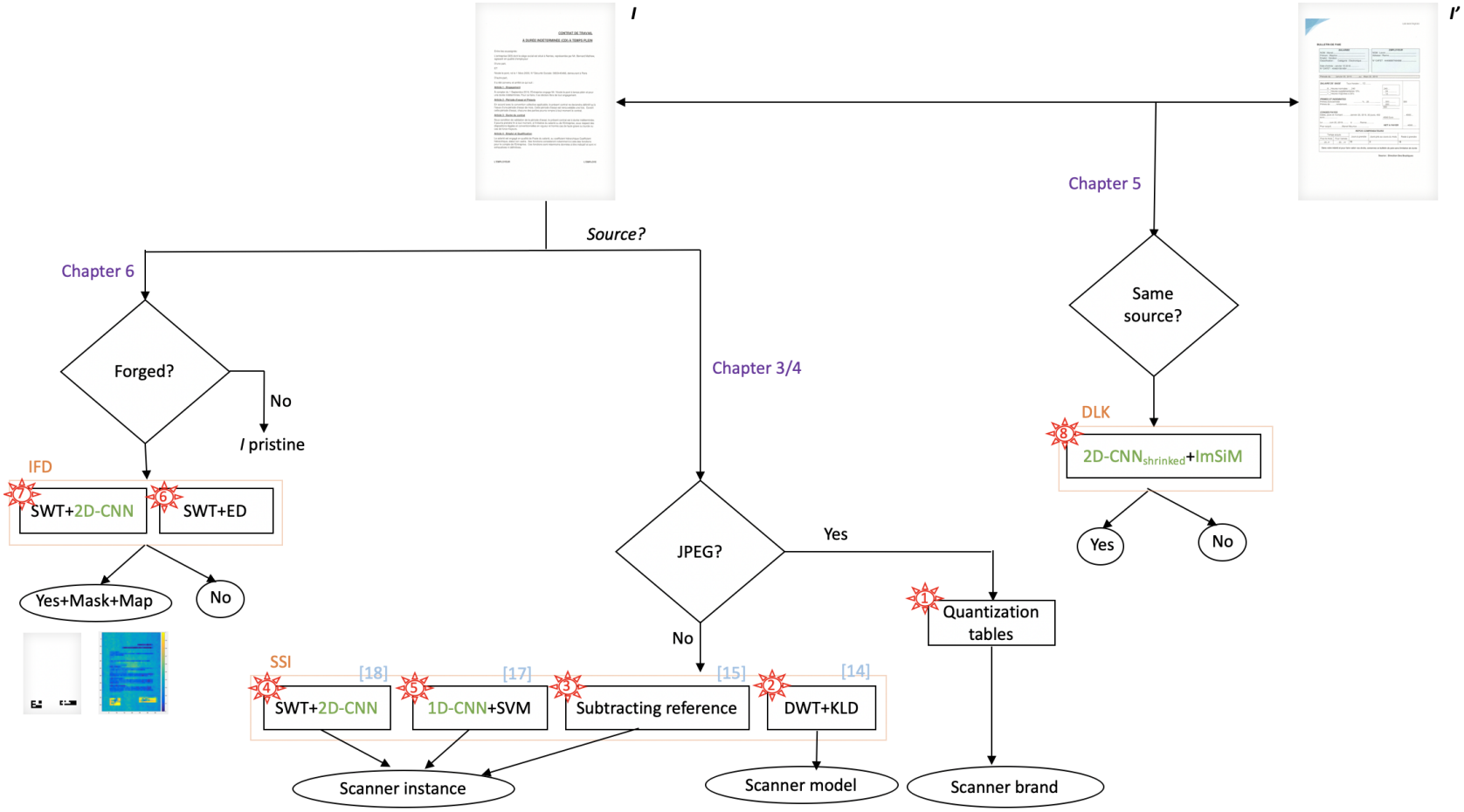


Figure 1.3: Flowchart of thesis goals and techniques used. $2D-CNN_{\text{shrinked}}$ and $ImSiM$ refers to the name given to our proposed schemes

1.4 Thesis outline

Literary forensics works which focused mainly on flatbed scanners are very few. Hence, the main purpose of this thesis is to find new and more performant forensics solutions to solve security issues related to source scanner identification and tampering detection in scanned documents.This manuscript is structured in six main chapters:
The second Chapter provides a general overview on the field of digital image forensics and reviews the state-of-art forensics methods that have been proposed to solve the device identification problem as well as forgery detection problems with a focus on those related to flatbed scanners. For that, we start by describing the scanning process to better understand how those methods were derived.
In Chapter 3, we study handcrafted features and propose different solutions to identify the scanner at the origin of a questioned image for different identification levels (brand-model-device). As we will see, one of the most challenging problems in the field is distinguishing between scanners of the same model. Performance of these solutions are experimentally validated showing identification accuracies that outperforms existing schemes.
In Chapter 4, we propose an amelioration of the handcrafted features extraction based techniques presented in the chapter 3 by exploring NN approaches. The proposed schemes originality stands on the fact that it allows identifying image source scanner using features learn automatically from input images. Those features demonstrated improving capabilities to discriminate between scanners.
In Chapter 5, a DLK scheme for matching image sources is considered. Its goal is to know whether two images were acquired by the same scanner or not which is a valuable forensic investigation tool. We demonstrated that our scheme achieved promising results.
Chapter 6 is devoted to image forgery detection. We start by presenting a novel dataset that we have created using 11 scanners as benchmark for image forensic tasks. Then, we propose two approaches to detect splicing forgery in scanned documents based on solutions proposed in Chapter 3 and 4. Performances evaluation of our approaches using the forged image contained in our dataset concludes this chapter.
Finally, Chapter 7 gives a general conclusion and discusses some perspectives for future work.

Digital forensics is an exact science -
Not the procedures but the results.

La forensique numérique est une science exacte -
Pas les procédures mais les résultats.

— Edewede *Oriwoh*

2

DIGITAL IMAGE FORENSICS: SOURCE SCANNER IDENTIFICATION AND FORGERY DETECTION

Contents

Scanners serve as a link between the physical and digital worlds, making document storage and sharing easier and faster. It lowers expenses, saves spaces and reduces the risk to lose critical documents. Companies, banks, and many other organizations use digitization to save hundreds of thousands of dollars in printing, shipping, and storing documents each year. Access to information, on the other hand, is made easier, which raises security concerns. It is obvious that our trust in images has degraded even when published in popular newspapers and official

websites as anyone can now capture, store, and process digital images due to the availability of various image editing tools that allow the manipulation of the content of digital images in an easy and unnoticeable way. Thus, the need to argue about the authenticity and integrity of those images.
Image security is a long-standing issue in the field of Digital image forensics (DIF). In recent years, this field has shown an important progress and experienced new devices such as scanners. It is derived from existing domains relating to multimedia security such as digital watermarking and steganography. As stated, image forensics relies on the intrinsic features of the "noises" embedded in the digital image by the device that has acquired it. It, then, uses those artifacts to identify the source device and detect manipulations of the digital images.
It is thus mandatory to have an adequate understanding of the basic working principles of a flatbed scanner, so as to understand how are noises introduced in the final image. This is the objective of the first part of this chapter: presents how a scanner works. In the second part, we outline the basics of the emerging field of DIF related to this device by reviewing the solutions proposed in literature to solve the problem of source scanner identification. We introduce the problem of forgery detection in the third part and how the previous techniques proposed to associate the source device can serve to solve this problem. We go also through the limitations as well as the contributions of those methods.

2.1 Image acquisition pipeline of a flatbed scanner

Understanding how a scanner works is a mandatory knowledge to get in order to develop efficient passive image forensics techniques.
The basic concept of a document scanner is to convert paper items into a digital format. We distinguish different categories of scanners depending on their intended use. The most common document scanner types are:

- Flatbed scanners (Fig. 2.1.a). They are the best-known used scanners. Such a device simply uses a glass surface over which the item to scan is placed and digitized. Most large-selling flatbed scanners are designed for A4 and letter paper. Nevertheless, there are various flatbed scanners that can handle other paper sizes such as A1, A2...

- Sheet-fed scanners (Fig. 2.1.b). They differ from flatbed scanners in terms of the format and weight of the items that they can handle. It is specifically limited to papers. It cannot be used to digitize thick objects like books and ID cards. The scan is realized by moving papers across a stationary scanning unit.

- Handheld scanners (Fig. 2.1.c). They are specifically designed to quick scanning applications like barcode reading. A data is captured by moving the handheld scanner across the surface of the item to scan. They are accurate and reliable but cannot offer good scanning quality, especially due to the operator movements.

- Drum scanners (Fig. 2.1.d). They are used to scan materials with an extremely high resolution that may exceed 10000dpi while, for the other scanners, the resolution cannot surpass 9600 dpi. During the scanning operation, a document is mounted in a glass cylinder and sent to a drum called drum scanner's photomultiplier tube (PMT) rotating at a very high speed. Regardless of their high price, drum scanners are still the ultimate solution to produce posters and other large images.

Figure 2.1: Flatbed scanner (a), Sheet-fed scanner (b), Handheld scanner (c), drum scanner (d)

In this work, our interest has been given to flatbed scanners due to their availability and to the fact that they are the most used, by professionals as well as by individuals. They also have a substantial advantage over other types of scanners as they are well known for their high quality scans.

A typical scanner imaging pipeline is illustrated in Fig. 2.2. As it can be seen, the document is in a first time placed on the glass plate under a closed lid to ensure that the external light does not enter the device and to prevent the document from moving. Next, a pass is realized by a motor slowly moving a scanning unit across the scanner's glass. The scanning unit is usually composed of a light source, mirrors, lens, and a charge Coupled Device (CCD) array. In order to ensure a constant pass without any wobble or deviation, this unit is attached to a stabilizer bar. Real photos of the scan head moving along the scanner glass are shown in Fig. 2.3.

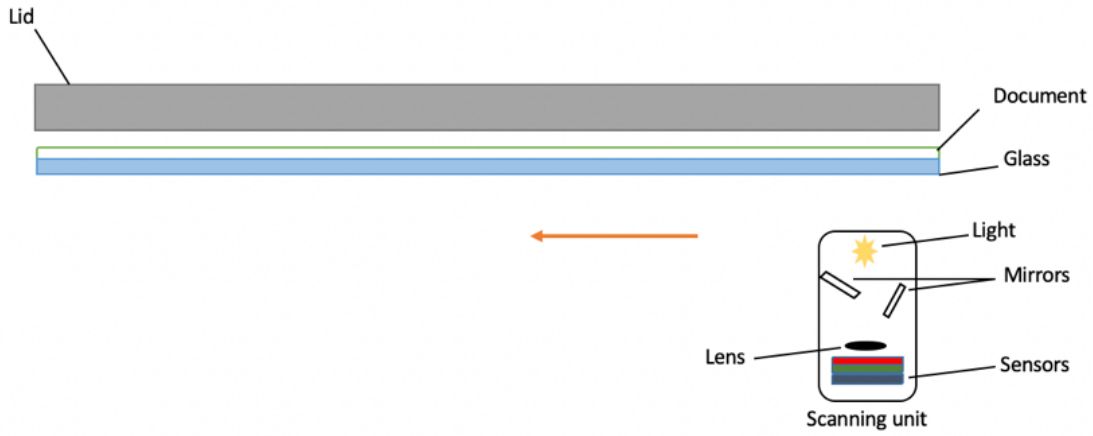


Figure 2.2: Flatbed scanner imaging pipeline: The orange arrow indicates the direction of the scan followed by the scanning unit composed of a light source, a set of mirrors, lens and the imaging sensor

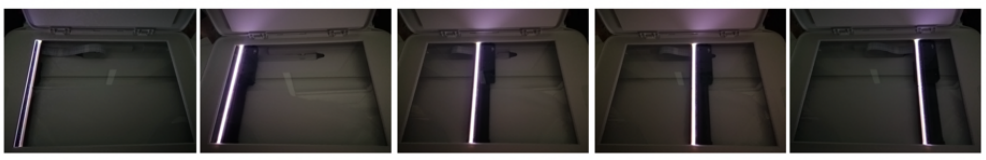


Figure 2.3: Changing positions of the scan head during the scanning process

To illuminate the document during the scanning process, scanner manufacturers used to use a cold cathode fluorescence lamp (CCFL) as a light source, but, recently, they replace it by light emission diodes (LED). Figure 2.4 depicts an illustration of each of them. Unlike traditional CCFL, LED light don't need a time to warm up enabling a quick start to scan. Moreover, scanners with LED consume less power, generate less heat and tend to last longer without color change.
Once the light reaches the document, it gets diverted by a series of slightly curved mirrors (two or three mirrors depending on the scanner). The light reflected by the last mirror goes into a lens which will focus it on the sensors array composed of a collection of tiny photosites.

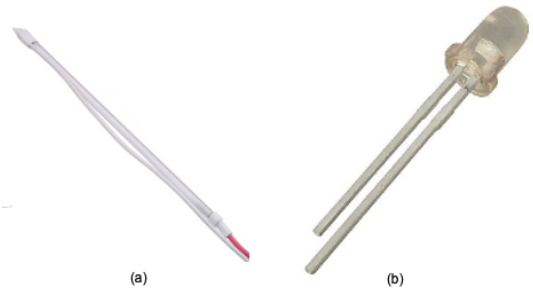


Figure 2.4: CCFL (a) vs LED (b)

Over the last 10 years, scanner manufacturers have delivered inexpensive scanners using a new technology called Contact Image Sensor (CIS) which differs a bit from the CCD.
As shown in Fig. 2.5, the main difference between CCD and CIS lies in the fact that the light reflected from the document in a CIS architecture is sent directly through an array of fiber optic lenses to the image sensors instead of using mirrors and a standard lens. Although this fact makes CIS-based scanners thinner and more energy efficient than a traditional CCD-based scanners, it prevents it from producing high quality results.

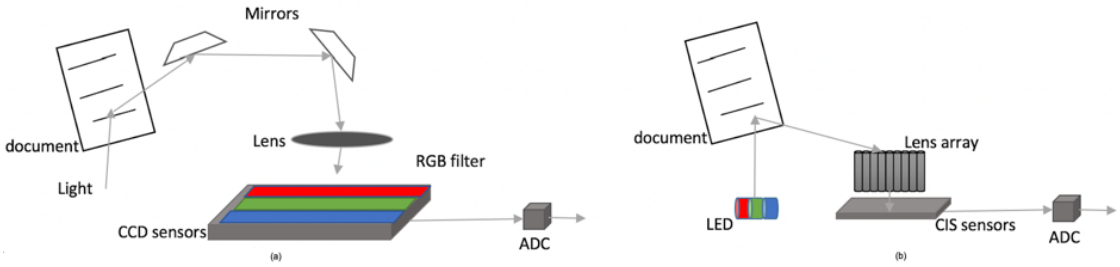


Figure 2.5: (a) CCD Scan relies on an actual lens together with a series of mirrors to reduce the entire document image onto the sensors compared to (b) CIS Scan which incorporates a lens array to transfer the document image to the array of sensors

The sensors, whatever they are CCD or CIS, convert the light into electrical signals which are later converted to a voltage proportional to the light intensity. In another word, the brighter the region of the image, the more transmitted light, which results in a higher voltage. This analogue-to-digital conversion process is a very sensitive step that may cause noise and electrical interference. Finally, the driver, a software adapted to communicate with the scanner, works on delivering a full-color image in the desired format.
One of the main characteristics of scanners is the resolution mainly expressed in

dots per inch (dpi) or pixels per inch (ppi), where 1 inch is equivalent to 2,54 cm. It is a representation of a scanner's enlargement capability. The scanners differ in their resolution and, therefore, in their cost. The higher the resolution, the more expensive the scanner is. The resolution of a scanner is measured by the number of a single row sensors in a CCD or CIS array and by the precision of the stepper motor. We distinguish the native or hardware resolution which is the maximum number of pixels per square inch a scanner can capture.

Comparison to digital cameras

Digital cameras may sometimes serve as scanners. However, despite their speed and their portability, they are not well adapted for document scanning especially when a high resolution is required. In fact, the image acquired by a camera will be affected by distortion, reflections, shadows, and blur due to the problem of stabilization of the camera during exposure. Since scanners use similar technologies as digital cameras, it is interesting to consider those similarities for forensic applications. Overall, they both shared the same fundamental components for capturing the object's color values. Both also use similar analog-to-digital converters (ADCs) and conduct similar post-processing operations such as edge enhancement and colors adjustment. The objective of the post-processing is to reduce noises and artifacts produced in the previous stages and to enhance the final image appearance.
However, it is important to notice that, compared to digital camera, flatbed scanners present distinct imaging sensors geometries as illustrated in Fig. 2.6. In fact, scanners use trilinear CCDs corresponding to the red, green and blue color channels, whereas digital cameras use 2-D periodic arrays. The scanners can directly capture all the three-color components of each raster line through the trilinear CCD array along with the line-by-line mode allowed by the motion system. Digital cameras instead use a 2-D CCD array to take one shot of the whole 2-D scene. As only the value of a single color component can be calculated by each CCD sensor, color interpolation is required for two remaining color components to be obtained. Serving as a crucial step in the imaging process of digital cameras, the coefficients of color interpolation have been used in the literature to distinguish different camera models.

Figure 2.6: Digital camera vs scanner imaging sensor (a) camera sensor (b) scanner sensor

Table 2.1 summarizes the most notable differences between scanners and digital cameras.
Since most of the artifacts investigated in source camera identification (SCI) methods are specific to the sensor design, they may not be potentially useful for the task of SSI.

	Scanner	**Digital camera**
Color sensor geometry	Trilinear	2D
Light source	Fluorescent Lamp or LED	Scene light alone or combined with flash
Acquisition time	Depends on the resolution, the input and the device.	Quick (One shot)
Input data	Hard-copy sample	Scene

Table 2.1: Comparison of the acquisition process between scanners and digital cameras

2.2 Scanned image origin identification

The common idea of passive approaches for SSI is that a scanner leaves an inherent intrinsic pattern in the image it acquires [20]. This pattern, if correctly extracted, can help discriminating between different scanner brands, models, and even between scanners of the same model. In fact, every step in the image acquisition pipeline adds some noise to the final image that can contribute to the elaboration of a unique scanner pattern. The choice of noise characteristics is the clue to an accurate and efficient SSI.
The features selected should meet the following conditions:

- Authenticate each scanner independently of document type and content.
- Enable to discriminate between different scanners whatever their make and model.
- Remain the same whatever the scanning area selected

Even though there are several approaches for SSI, the forensic investigation process, illustrated in Fig. 2.7 is almost the same. Understanding this process requires having a detailed knowledge about how the digital investigator collects traces of each device and use it to generate different patterns. He is, then, supposed to answer the question of which device has acquired each of the questioned documents by comparing these documents' patterns to those of the available scanners.

In this context, we may group digital image forensics methods related to SSI into the following five categories:

- Spatial domain features
- Transform domain features
- Color features
- Texture features
- Dust and scratches positions

We come back on the details of these main approaches in the following.

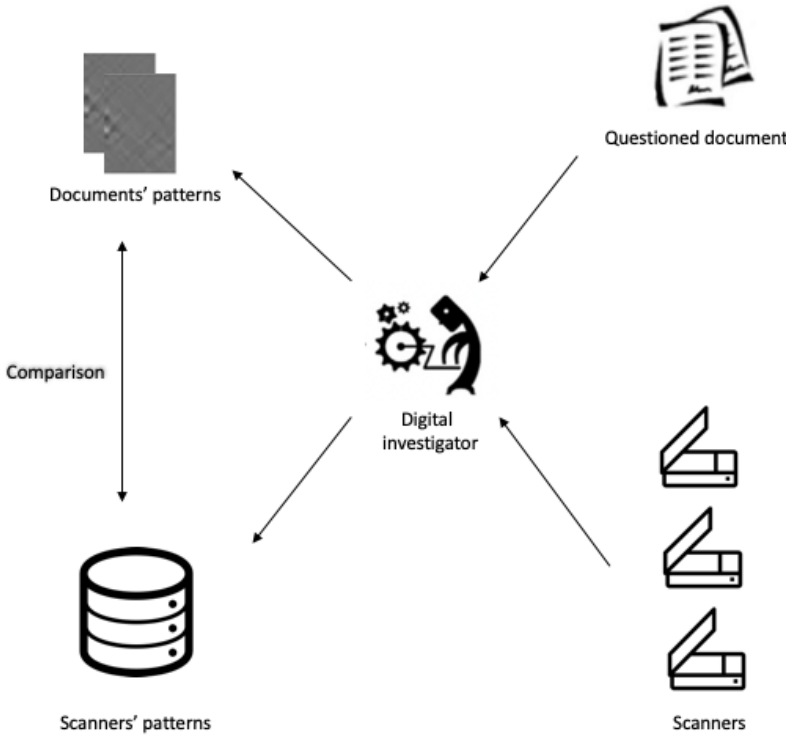


Figure 2.7: Forensics investigation process for SSI

2.2.1 Spatial domain features

For documents analysis, some digital investigators relied to an inherent mark of scanning: the noise effect which is due to its materials and components the manufacturing processes of which are imperfect.
During the scan, color sensors capture light reflected from the original paper and then translate it into digital values. Reflected light is equal to the number of photoelectrons recorded by the color sensor over each time span, which is then converted to analog signals such as voltages. Because of the random nature of photoemission, many forms of noise occur [21]:

- Shot noise: It is common when the charge carriers (for example electrons) pass through a gap resulting in random statistical changes of the electric current. If electrons flow across a barrier, their arrival times are discrete. Shot noise can be seen in those separate arrivals.

- Quantization noise: It is a quantization error model introduced by the quantization in the ADC. It is a rounding error between the ADC's analog input voltage and the digitized output value. The noise is nonlinear and dependent on the signal.

- Reset noise: It occurs when the reset of the capacitor happens before the charge accumulation cycle. In an imaging system, it might be the dominant noise source but can be fully removed by the Correlated Double Sampling (CDS) technique [22].

- $1 / f$ noise: It depends on the frequency of the detected signal. The higher the noise, the lower the frequency. It is typically very low in comparison to the other noises mentioned above. It may found in amplification systems as well as some passive elements.

- Fixed pattern noise (FPN): It refers to the spatial variation in pixel output values caused by device and interconnect parameter variations across the sensor which occur due to manufacturing imperfections.

As the shot noise and the quantization noise vary from one image to another, they cannot be used for scanners' identification. That is not the case of the FPN. It is fixed for a given sensor but varies from one to another and, consequently, is a good source of features for identifying the scanner.
If we go more into the details of FPN, it typically refers to two types of noise that are of concern for use in forensic classification; the dark signal non-uniformity (DSNU) and the photo response non-uniformity (PRNU) [21]. Together, they introduce a structured noise pattern over the scanned image that can be later explored for SSI. When the sensor is not illuminated, the DSNU creates pixel-to-pixel variations among pixels in the analog voltage. In contrast, the PRNU causes variations when the sensor is illuminated under a fixed-intensity light. It is important to know that, by device calibration and offset correction, DSNU and PRNU can be compensated to some degree.
Even though noises compensation and reduction methods are used, it has been shown that scanned images still maintain a certain amount of noise that can serve for identifying the scanner at the origin of an image. Such a method basically relies on extracting some of its statistical features.
Generally, the first step to generate the scanner's pattern is to extract the noise I_n of each image acquired using a denoising filter F :

$$I_n = F(I) = I - I_d \tag{2.1}$$

where I and I_d denote the scanned image and its denoised version, respectively. I is of size N x M (N rows and M columns).
Note that since the PRNU noise $\varkappa$ is modeled as a multiplicative noise-like signal, I may be expressed by

$$\begin{aligned} I &= I_d^0 + \varkappa.I_d^0 + \xi \\ &= (1+\varkappa).I_d^0 + \xi \end{aligned} \tag{2.2}$$

where I_d^0 is a zero-noised version of I and ξ is the sum of independent random noises which may contain other image content details, given the fact that F cannot adequately remove them. Next, the expression of I_n is rewritten as follows

$$\begin{aligned} I_n &= (1+\varkappa).I_d^0 + \xi - I_d \\ &= (1+\varkappa).I_d^0 + \xi - I_d + \varkappa.I - \varkappa.I \\ &= \varkappa.I + I_d^0 - I_d + \varkappa.(I_d^0 - I) + \xi \\ &= \varkappa.I + \psi \end{aligned} \tag{2.3}$$

where ψ is the overall combination of ξ with additional weak signals added by the denoising filter.
Working on I_n helps improving the signal to noise ratio (SNR) for the signal of interest $\varkappa.I$, thus improving the reliability of the SSI techniques.
The choice of the denoising filter is very important in order to accurately extract the fingerprint for source identification and thus to achieve high performance.
First methods in the literature adopted the same techniques already used for SCI. In [20], for example, a direct extension of the PRNU-based SCI identification algorithm [23] was used for scanner model identification. The reference pattern of each scanner was generated by averaging noise patterns from 100 training images obtained from five different scanners. The images reference patterns are obtained

by applying the Mihcak filter proposed in [24] on each image to suppress the image content. The task of SSI is then achieved by measuring ρ_{S_i} ,the normalized cross correlation (NCC), between the noise residue R of the questioned image and each of the scanners' reference patterns S_i

$$\rho_{S_i} = corr(R, S_i) = \frac{(R - \bar{R}).(S_i - \bar{S}_i)}{\|R - \bar{R}\|.\|S_i - \bar{S}_i\|} \tag{2.4}$$

where the bar " $\bar{.}$ " denotes the mean value and "$\|.\|$" stands for the L2 norm.
The source scanner will be the one that maximize this measure. All tests conducted in this work revealed a lower classification quality relative to related studies for SCI. It has also been found that the use of a 1-dimensional reference pattern provides greater classification accuracy when images are scanned at non-native resolution, whereas the 2-dimensional reference pattern gives better results when images are scanned at the scanner 's native resolution. Besides, another approach [25] for scanner model identification based on sensor pattern noise uses three sets of features that are extracted from each digitized image. It obtains the statistical features of the sensor noise estimates through the use of different denoising filters (averaging filter, Gaussian filter, median filtering, and Wiener adaptive image denoising algorithm), high-frequency wavelet coefficients, and neighborhood prediction errors. More investigations and analysis of this approach have been made in [26].
Authors in [27–29] proposed to use the statistical features of the sensor noise along with machine-learning classification such as the Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM). For that, the average row R^r and the average column R^c of the noise residue are calculated

$$R^r(1, j) = \frac{1}{N} \sum_{j=1}^{N} R(k, j) \quad , 1 \leq j \leq M \tag{2.5}$$

$$R^c(i, 1) = \frac{1}{M} \sum_{j=1}^{M} R(i, k) \quad , 1 \leq i \leq N \tag{2.6}$$

where N is the total number of rows and M is the total number of columns.
In the first approach [27], the noise residue is extracted using the anisotropic local polynomial approximation—intersection of confidence intervals (LPA-ICI) denoising scheme [30]. The statistical features such as the mean, the median, the standard deviation, the skewness and the kurtosis of the average row R^r and of each correlation between that row and each row of the noise residue R were combined and associated to its corresponding scanner. Later, in [28, 29], the average column was also considered as well as correlations between it and all the columns of R for the same features. Another measure corresponding to the relative difference in the sensor noise periodicity was added to the previously listed features. The median filter (size 3x3) and the Wiener adaptive image denoising (sizes 3x3 and 5x5) are applied along with the LPA-ICI denoising filter to extract noise residues in [29].
These methods have shown promising results for scanned photographs, however, their stability in the case of text documents is still questioned as PRNU is almost absent in dark and saturated regions. Moreover, the process of fingerprints extraction of these models is computationally extensive.

2.2.2 Transform domain features

Choi *et al.* [16] decided to work in the Fourier-transform domain to explore different type of noises. The unsharp filter was chosen to denoise the scanned image with the mask h as follows

$$h = \frac{1}{1+\alpha}\begin{bmatrix} -\alpha & \alpha-1 & -\alpha \\ \alpha-1 & \alpha+5 & \alpha-1 \\ -\alpha & \alpha-1 & -\alpha \end{bmatrix} \tag{2.7}$$

where the parameter α is set to 0.1. The authors decided to work with a one-dimensional reference pattern by averaging the noise residue along the scanning direction following the eq. 2.5. Next, a DCT is applied to the average row. Finally, before normalizing the reference pattern, a high-pass filter is applied on the DCT coefficients. It is then possible to identify the source of the scanned image by comparing its fingerprint to the fingerprint of each of the available scanners using the Euclidian distance where a scanner's fingerprint is calculated by averaging fingerprints of images scanned by it.
While tests have yielded positive results, only distinct scanner models have been tested for 10 images per model. More tests, including multiple devices per scanner model are thus desirable.

2.2.3 Color features

In [31], Sugawara approaches the problem of SSI by suggesting a solution that identifies an image source scanner based on the variation in color features across security holograms (example is shown in Fig. 2.8) usually found in low-level currencies. This method is thus limited by the presence of holograms in the image to be questioned and has been proved efficient only under some specific conditions.

Figure 2.8: Sample of a multi-directional grating hologram [31]

2.2.4 Texture features

Most of the proposed SSI methods focused on scanned photograph versions and not scanned printed text documents. However, in the saturated areas of an image and especially in black and white documents, PRNU is almost absent. Therefore, for scanned text documents, these methods may not work well. In this context, Khanna and Delp [32] proposed a new approach to identify the scanner model through texture analysis, where textures are modeled by the gray level fluctuation

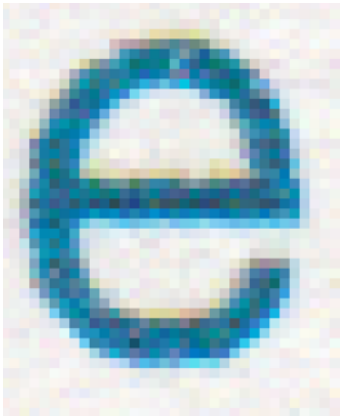


Figure 2.9: Zoomed letter 'e' from a scanned document at 300dpi resolution

in the scanned characters. These fluctuations can be observed in the zoomed letter "e" shown in Fig. 2.9.

The first step consists in determining the Gray-Level Co-occurrence Matrix (GLCM), an approximation of the second order probability density of the image pixels. Features to be considered for classification are then based on the GLCM. The isotropic gray-level difference histogram (GLDH) is also used to enhance the scanner's fingerprint.
It is easily seen that experimental results heavily depend on font shape and size and the presence of character "e" in the text. In another paper [33], authors have only focused on GLDH features to improve the robustness against variations in font sizes.
In a recent work, Joshi *et al.* [34] further proposed a more powerful approach which performs very well for varying font sizes utilizing local tetra patterns (LTrP) [35] based features. However, its performance is not totally independent from the font shapes/sizes and is affected when blurring regions are present in the digital document. Further, these methods work well only on plain text documents containing a certain amount of characters.

2.2.5 Dust and scratches positions

Due to the imperfection of the design and manufacturing process, the scanner's glass creates undesired effects in the acquired images such as dust and scratches. The position of these defects has been explored for scanner identification [36, 37]. Indeed, the plate got contaminated by dirt particles and paper debris after using the scanner for a while. Dust and scratches, when accumulated, may cause located defects in the form of white or black spots over the scanned image. Although dust particles appear in the picture as dark spots, glass scratches typically lead to light-reflections that are bright and white. In Fig 2.10, the dust and scratch positions on a scanned image are shown.

The idea is that the positions of the glass defects and scratches are strongly adhered to the scanner platen and do not change, even when the platen is manually cleaned. Thus, those positions can represent a unique scanner pattern.
However, defects may be introduced when the scanner is extensively used leading to misclassification and thus a drop in the performances. Furthermore, new scanners are equipped with filters that delete most of these defects.

Figure 2.10: Dark spots in a scanned image due to dust/scratches over the scanner platen

2.2.6 Discussion

To provide key information that can identify or narrow the possible sources of the document, several methods have been proposed. Experiments conducted on these methods, however, did not take into account all types of documents and generally failed to properly classify scanners when there is more than one instance of the same model. The increase of the number of scanners and the decrease of the number of training images may also lead to a decline in performances. Therefore, the work of Forensic Document Examiners on this task still needs more investigation. In this thesis, we worked on developing SSI methods which extract more trustful scanners fingerprints with a better forensic traceability compared with the state-of-the-art methods.

2.3 Image forgery detection

2.3.1 Introduction

Nowadays, a digital document may be presented as a proof in court of law [1] or used in a forensic investigation. If its content is manipulated, it will lead to controversial, and often tragic decisions that put authorities at risk and harm the reputation of individuals. That's why it's necessary to verify and validate the integrity of the digital document before it is used.
With the emerging image editing software like Adobe Photoshop, seashore, GIMP and Coral Draw, it has become very easy for professionals and even for amateurs to forge digital documents and hence the trust in these documents has been corroded. Digital forgery is present in many areas of everyday life (e.g. fake images of politicians, celebrities and news) for years now. For instance, in a photograph published in a Tunisian newspaper "Le Maghreb" in January 29, 2012, the number of protesters was increased by duplicating a section of the crowd (see Fig. 2.11). In fact, the significant number of forged images shared in social networks and messaging applications is the principal source of fake news propagation.

Nevertheless, image manipulations are not applied only on photographs, scanned documents are also concerned. It includes, for example, copying a person's name on

[1] In the United States, once the original documents got accurately reproduced by a scanner, the digital version becomes legally valid and can replace the physical one. Source: https://definitions.uslegal.com/u/uniform-photographic-copies-of-business-and-public-records-as-evidence-act/

Figure 2.11: First example of a digitally altered image (The original section and the duplicated one were pointed)

a document, copying and pasting a signature, or altering information on a document without its author's consent.
Recently, in March 2020, a digital document has been circulated on social media saying that travelers arriving from Egypt, Oman and Kuwait are prohibited to enter the state of Qatar and persons holding a Turkish or Iranian nationalities are exempted from this decision. In the original document, only the entrance of passengers coming from Egypt is restricted and the exception is made only for Qatari citizens. The malicious purpose of manipulating such document is probably to increase the crisis between Qatar and Egypt that has already began in 2017. Both documents are shown in Fig. 2.12. It can be seen that it is almost impossible to differentiate between them and that there are no visual clues indicating fraudulence.

Therefore, image forgery detection (IFD) schemes [38] are required to protect copyright and prevent malicious alteration of digital images. There are a plenty of approaches proposed in the literature to fight against different types of forgeries. These approaches can be mainly divided into two classes: active (non blind) approaches and passive (blind) approaches as depicted in Fig. 2.13. Approaches belonging to the first category require the insertion or the association of additional information to the digital document to be secured. Existing techniques are based on digital signature [39, 40] and watermarking [41–49], and have limited applications since the images need to be secured at their acquisition or at later stages which is no usually the case of images that are subject of forensic investigations.
Thus, passive or forensic approaches are proposed to overcome these limitations. They are called blind referred to their ability to detect alterations without any access to the original image. These techniques can be further classified into manipulation and device based methods.

2.3.2 Manipulation based methods

Manipulation based methods are sub-divided into two categories according to the type of tampering performed on the digital image. The first category, called type dependent forgery, includes splicing and copy-move forgeries. On the other hand, the second category includes forgery-type independent methods in which certain properties of the digital image are manipulated such as resampling and retouching. There has been a substantial amount of study in the area of forgery detection in the last ten years. Figure 2.14 shows the bar chart of a variety of publications according to each form of image tampering technique over the years 2010-2020.

Figure 2.12: Second example of a digitally altered image (On the left the original document and on the right the doctored one)

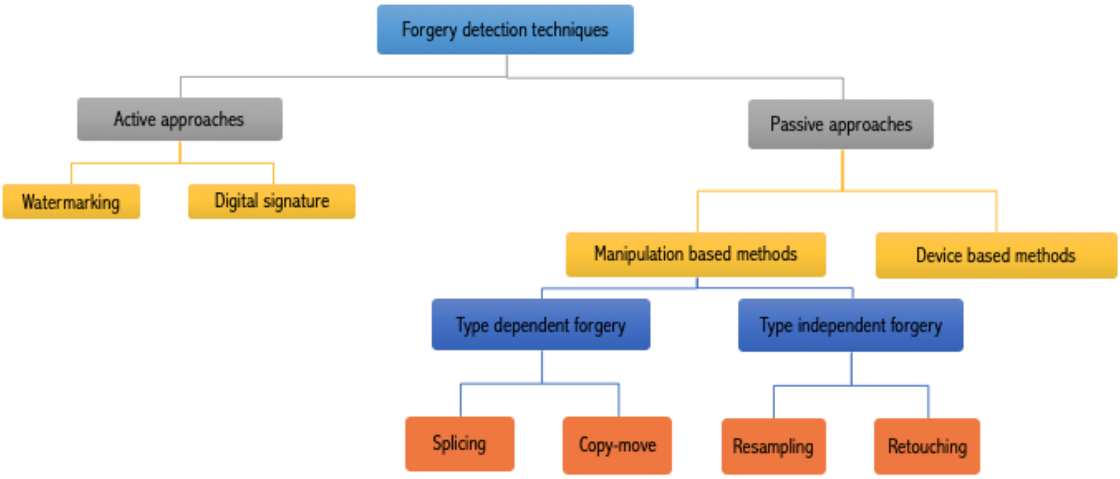


Figure 2.13: Classification of forgery detection techniques

2.3.2.1 Type dependent forgery

Image splicing

Image splicing forgery is a method which simply cuts and pastes certain portions from one or more images onto another image known as composite image. An example is given in Fig. 2.15 in which image (a) and image (b) are the original images and the image (c) is illustrating the doctored image (i.e. spliced image). In image (c), the original signature in the image (a) was replaced by the signature on the left in the image (b).

In [50], the authors extracted Markov features, which have been proved to be one of the most powerful tools for image splicing detection, in both DCT and DWT domains and captured intra-block and inter-block correlations between block

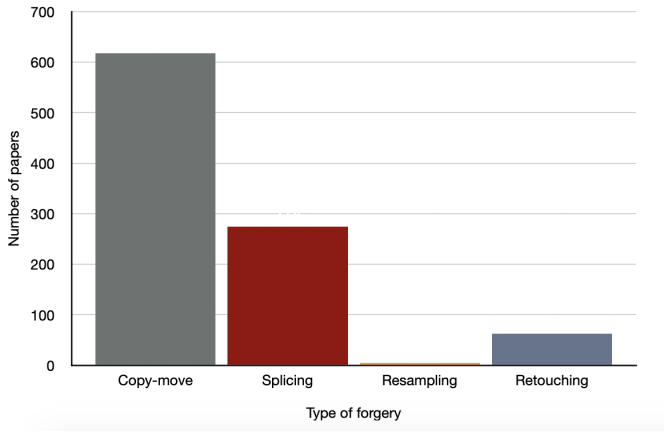


Figure 2.14: Chart with number of published papers in the field of image forgery detection over the last decade (2010-2020)

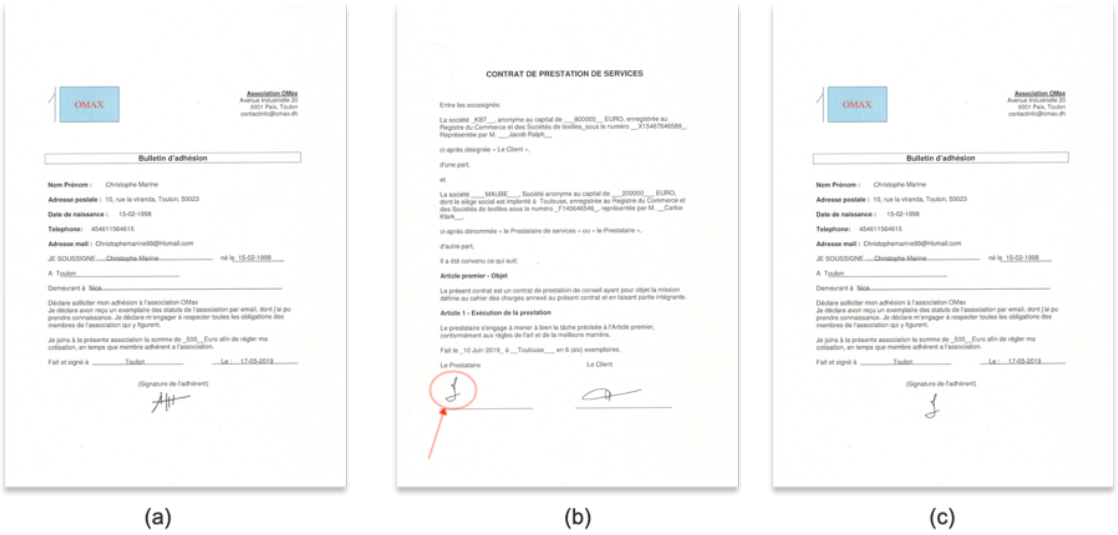


Figure 2.15: Example of splicing forgery (a) Target image (b) source image (c) Tampered image

DCT coefficients. After that, they used the most sensitive features to train an SVM classifier. In [51], Mushtaq *et al.* proposed a method based on the grey-level run length matrix (GLRLM) texture features. They used the statistical features extracted from the GLRLM to train an SVM classifier.
Moreover, there are some splicing detection methods that have used the local binary pattern (LBP) [52]. For example, Muhammad *et al.* [53] proposed an algorithm based on the LBP histogram which extracts features from the Cb and Cr color channels yielded by the Steerable Pyramid Transform (SPT). Again, the SVM is used for image classification after features reduction. LBP is combined with DWT in [54] and with DCT in [55] to detect splicing forgeries.
Traditional Markov-based approaches typically considered the image as a 1-D signal and reported the dependency between each node and its neighbor along one direction. In [56], the authors improved the model in [50] by using a 2-D non-causal Markov model instead of the traditional one and by replacing the DWT transform with the discrete Meyer transform in order to characterize the underlying dependencies of

adjacent pixels in all directions.
Zhang *et al.* [57], however, introduced the Markov model to DCT and Contourlet transform [58] domain and detected the image tampering by the Markov features extracted in both domains. Since the proposed vector is typically of high dimensionality, SVM with an ensemble classifier (EC) [59] are used for classification to solve the potential problem of overfitting.
Similarly, Li *et al.* [60] proposed a Markov based method that make use of the Quaternion Discrete Cosine Transforms (QDCT) when extracting features. Inspired by this technique, a novel image splicing detection scheme based on Markov features in QDCT and quaternion wavelet transform (QWT) is proposed in [61].
In the above-mentioned techniques, achieving high detection rates with relatively small function dimensions cannot be assured. Thus, a method [62] using textural features based on the Block Discrete Cosine Transform (BDCT) arrays and the GLCM, named TF-GLCM, has been proved to better detect textural information with a low computational complexity. Wang *et al.* [63] recently developed a way of detecting and locating the image splicing using a coarse-to-fine grained technology. Using the Laplace operator, the local noise is captured and the regions are clustered based on connected region expansion/corrosion. Kanwal *et al.* [64] investigated a method that reveals the image splicing in an image based on the local ternary count.
Contrary to handcrafted features-based methods, the latest advances have centered on deep learning based solutions [65–73] which have become very popular due to their ability to learn more general features from images.

Copy-move
Copy-move, also known as cloning forgery, is a form of image forgery in which a region is copied and pasted onto another region in the same image frame to mask or duplicate an object in the image. Figure 2.16 illustrates an example of copy-move forgery. In image (b), the signature was hidden by copying an empty block from the original image (a) and pasting it over the signature.

In the last few decades, several copy-move forgery detection (CMFD) methods have been introduced. They can be divided into three categories: keypoint-based, Block-based and hybrid-based methods.
The main idea behind the first category is to extract keypoints and their descriptors from the questioned image and then the descriptors are compared to detect forged regions where a keypoint represent a point that does not change even if a geometric transformation is applied to the image Several methods from literature applied the Scale Invariant Feature Transform (SIFT) [74, 75] and Speed Up Robust Feature (SURF) [76–79] to extract keypoints. A more computationally expensive technique is developed by Yu *et al.* [80] which combines hue histogram descriptor and the multi-support region order-based gradient histogram (MROGH). Li *et al.*[81] used the Zernike moments [82] to represent the extracted features and the Maximally Stable Color Region (MSCR) as a detector. Another way to extract keypoints based on the Oriented Fast and Robust Brief (ORB) algorithm is proposed in [83]. Similarly, the authors in [84] proposed to select keypoints by the non-maximum value suppression algorithm.
Lin *et al.* [85] proposed a method based on SIFT features and the generalized 2-nearest neighbor (g2NN) strategy [86] using the Harris-Laplace [87] and Hessian-Laplace [88]. Authors in [89] also adopted the SIFT features and the g2NN strategy for the matching process and a different clustering approach for detecting duplicated regions.
Keypoint-based approaches are more advantageous when it comes to execution time, but fail to detect keypoints in the presence of low contrast and smooth regions.
In the second categories, the suspected image is divided into overlapping blocks and then duplicated regions are located using a matching algorithm. In [90], Bi *et al.* extracted feature vectors from each block using the Multilevel Dense

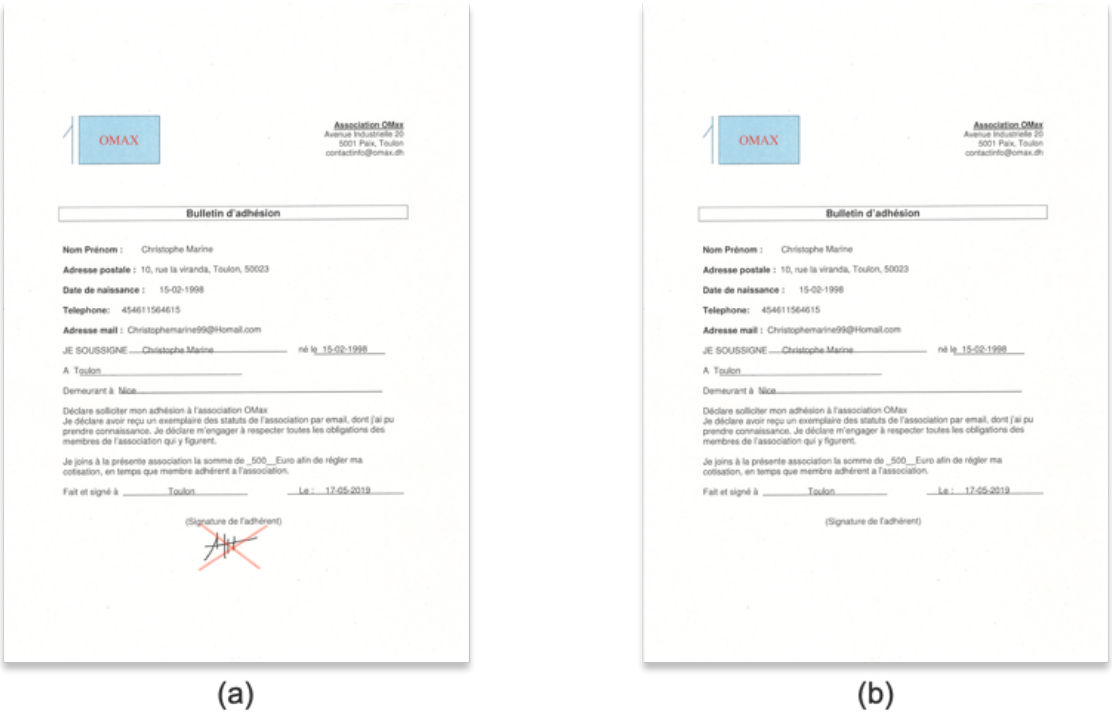


Figure 2.16: Example of copy-move forgery (a) Original image (b) Tampered image

Descriptor (MLDD) consisting of the Color Texture Descriptor and the Invariant Moment Descriptor. Experimental results have shown that this copy-move forgery detection method is robust against many attacks such as geometric transformation and downsampling... Emam *et al.* [91] applied the Polar Complex Exponential Transform (PCET) to extract features from the overlapping blocks. Then, they matched similar blocks using the Locality Sensitive Hashing (LSH). Their approach is also robust against attacks. Additionally, Cozzolino *et al.* [92] proposed to use a matching algorithm called the Patch Match algorithm [93] to efficiently compute an approximate nearest neighbor field for the suspected image.
In [94], Hsu *et al.* presented an algorithm based on the Gabor filter. They extracted features from the Histogram of Orientated Gabor Magnitude (HOGM) and calculated the maximum magnitude value of the image. Alternatively, authors in [95] developed a method based on the QDCT coefficients. Tu Huynh-Kha *et al.* [96] proposed a new approach that calculates the modified Zernike moments (MZMs) from the overlapping blocks of the LL subband resulting from the application of one level DWT on the grayscale version of the suspected image. On the other hand, Prakash *et al.* [97] proposed a new method based on DCT coefficient which works on different block size. Vivek *et al.* [98] proposed to use the local binary pattern to extract features. They, first, transform the input image into a grayscale one and, then, split it into overlapping blocks and compute the LBP histogram for each block. Finally, a Lexicographical sorting is applied for matching. An improved block-based matching algorithm (IBMA) is proposed in [99] to solve classical block based matching algorithm (CBMA). It uses the DCT, the PST (Polar Sine Transform) [100], the brightness, the Hu moments [101] and the Zernike moments for features extraction.
Because of the high-dimensional feature vectors and a number of overlapping blocks, these block-based approaches suffer from higher computational costs. Thus, to address the limitations of the block-based and keypoints-based methods, hybrid-based techniques [102–105] have been proposed.

Resampling
When copy-pasting a region of an image into another image, one may need to resize the pasted region to adapt it to the content of the host image. This operation will unavoidably leave some inconsistencies in the pasted region which helps to detect the tampering. First, it engenders periodic correlations between adjacent pixels which can be estimated using the Expectation Maximization (EM) algorithm. However, these correlations can be confused with the periodic patterns introduced by JPEG blocking artifacts if the image is compressed. In [106], the detection of resampling has been performed through the Singular Value Decomposition (SVD). This method was further improved in [107].
Moreover, resampling can be detected using a blind deconvolution [108], a linear parametric model [109], an interpolation kernel [110], or texture weight map [111]. More resampling detectors based on CNN have been proposed in [112–117].

Retouching
The image retouching forgery consists in enhancing the visual imagery contents by applying several global or local transformations on the original image using one or more filters or by manipulating a part of the image to mask or to add information to it. The image manipulation techniques involve global characteristics improvements (color or contrast enhancement, color remapping...) as well as retouching a particular region in the image. The figure 2.17 gives an example of image retouching where a filtering operation was applied on the original image (a) to make it look more recent.

Figure 2.17: An example of retouching (a) An old photo of my father taken in 1981 (b) retouched image to look that it has been taken recently

One of the most common techniques is the global contrast enhancement (CE). It generates impulsive gaps and peaks effect in the image's histogram which can be captured by analyzing the histogram characteristics [118–120], by identifying the use of histogram equalization [121], or through statistical distribution of block variance and AC DCT coefficients [122].
Complementary approaches are exploring the median filtering (MF). This filter is widely used in image processing to remove noise and preserve edges. It leaves statistical traces in the image. Therefore, MF can be detected by analyzing those traces. For example, authors in [123] use subtractive pixel adjacent matrix (SPAM) features with 686-D to detect MF manipulation, Others feature set have served in this field such as median filtering forensics (MFF) features sets [124] and the global and local feature (GLF) set [125], The probability of zero values in texture regions on the first order difference map may also serve as the MF statistical fingerprint [126]. More statistical features like the one-dimensional autoregressive (1D-AR) model of the image's median filter residual [127], the two-dimensional autoregressive

(2D-AR) model [128], the histograms features based on invariant patterns [129], the second-order local ternary pattern (LTP) [130], and local difference descriptor (LDD) features [131]. Some algorithms for the detection of MF have been proposed in the frequency domain such as the work proposed by Liu *et al.* [132]. Authors in [133] and [134] recently proposed MF detection methods based on automated feature learning and CNNs.
Face manipulation has also become a great research concern [135–140] with the growing number of face-retouched images shared via social media.

2.3.3 Device based methods

While the state-of-the-art methods discussed in the previous sub-section operate each for only one specific type of forgery, utilizing flexible and generalized algorithms that can simultaneously detect various forms of forgery seems to be a more attractive direction. The following is a brief description of the so-called passive device-based methods related to scanners that may achieve this goal. These methods detect the image manipulation by extracting features related to each scanner model and consider it as an intrinsic fingerprint of that model and then analyzing the coherence of this fingerprint in the suspected image. For example, the fingerprint can be modeled as an alteration in edges, in illumination, or in the characteristics of the scanning noise.
Most of these methods present an extension of the SSI approaches [16, 29] previously reported in Section 2.2. The basic concept behind these approaches is to split the image into non-overlapping blocks and then separately identify the source of each block. The questioned image is admitted as authentic if almost all the blocks are classified as acquired by the same scanner which will be declared as the source scanner of that image. If not, the image will be considered as forged.
One approach for detection of fallacious documents and tampering has been performed in [28] by using statistical features of image sensor pattern noise where the forged images had been synthesized using scanned images from different scanners. In [141], the method using spectral noise in the frequency domain [16] has been used to detect malicious forgery in scanned images where forging was realized using images acquired by scanners and digital cameras. The traces of dust, dirt, and scratches over different scanners platen on the scanned images are explored as unique patterns in [142]. Each block of the questioned image is compared to the acquisition scanner's template to differentiate the authentic scanned blocks and the tampered ones. If the scanner template matches all the blocks in one image, it is quite possible that this image is an authentic one coming from that source.
A more sophisticated approach has been proposed in [143]. It explores the differentiation of the illumination component histograms as a tool for slicing detection. Separating illumination from reflectance components is realized by applying homomorphic image processing on the suspicious image.
Table 2.2 summarizes a variety of distinct digital imaging forgery detection mechanisms along with their details, techniques used as well as their merits and demerits.

2.3.4 Discussion

Important documents such as contracts, educational certificates and government publications become easy to forge once digitized. Protecting those documents is not an easy task. It has even become more difficult with the tremendously advance in digital technologies.
In this section, we presented an overview of the prior work made on the task of forgery detection.
These methods were not tested on text documents and were not proved to perform well under post-processing. In addition, the accuracy of these methods actually depends on the block size that affects the outcome of the comparison between the document in question and the scanner fingerprint. Furthermore, the previously stated methods were successful in detecting forged images, but all methods have errors in determining the precise forged regions. Some of the fraudulent regions are missing or some original regions are detected as fraudulent.
Therefore, the development of methods to overcome these limitations is still an open issue.

Table 2.2: Comparison of device-based forgery detection techniques related to scanners (NN refers to Non-Numerical and NO to Non-Overlapping)

Paper	Used features	Size of the block	Type of identified forgeries	Number of scanners	Number of images	Resolution of scan	Result reported	Advantages	Limitations
Khanna *et al.* [28]	Statistical features of imaging sensor pattern noise	384x512 (NO)	Splicing	5	125	1200dpi	NN	-Able to detect splicing forgeries in image photographs	Very high computational cost -Documents are not addressed -Wrong classification in the heavily textured or saturated regions -Cannot use smaller blocks
Choi *et al.* [141]	Spectral noise	256x256 (NO)	Splicing	4	3	300dpi	NN	Able to detect splicing forgeries in image photographs where forged blocks are non-scanned photographs	Documents and scanned forged regions are not addressed
Elsharkawy *et al.* [142]	Dust and scratch positions	362x408 (NO)	Splicing	3	7	300dpi	NN	Able to detect splicing forgeries in image photographs	-Documents are not addressed -Cannot use smaller blocks
Elsharkawy *et al.* [143]	Histogram of the illumination component	Image size	Splicing	10	100	-	97.25%	-Detect tampering independently of the image source -High detection accuracy -Tested with JPEG compressed images and in the presence of noise	Documents are not addressed

2.4 Conclusion

In this chapter, a brief description of the imaging pipeline of a flatbed scanner as well as the related state-of-the-art forensic methods are introduced. Despite the fact that denoising and patterns comparison techniques are critical in the entire identification process, the selection of features utilized for source assessment plays an important role, particularly in classification methods. We also observed that various techniques have been proposed, however, a good part deal with particular scanned images and more investigations are needed to distinguish between scanners of the exactly same model. Moreover, we pointed out that a large number of images acquired by each scanner could not always be at hand to generate its fingerprint.
So far, we presented an extensive analysis of the technologies used for passive image IFD. This analysis is performed in terms of the features used, the identified forgery, the used datasets, and the limitations. As also mentioned, only few works have been devoted to deal with tampered scanned images and none of the scanner-based IFD methods have been tested on scanned documents. In addition, only splicing forgery has been investigated. Exploring these challenges may open the door to more efficient contributions.
In Chapter 3 and 4, we propose more advanced source scanner attribution techniques taking advantage of the potential of the wavelet domain and the artificial intelligence classification techniques. Later, we proposed a method for verifying whether two digital documents were obtained using the same scanner, a problem known as device linking (DLK). Our work is the only study that proves the feasibility of DLK for the case of scanned images. It will be the subject of the chapter 5. For IFD, we extended upon our works for SSI to find out if a scanned document has been manipulated. We explore small patches in order to better localize the forged regions.

Better be despised for too anxious apprehensions, than ruined by too confident security.

Mieux vaut être méprisé pour des appréhensions trop anxieuses, que ruiné par une sécurité trop confiante.

— Edmund *Burke*

3

SOURCE SCANNER IDENTIFICATION BASED ON A HANDCRAFTED FEATURES EXTRACTION

Contents

In the previous Chapter, we presented an overview of existing handcrafted scanner features extraction models. If most of these models appear to be quite useful under certain circumstances they also have some limitations that should be

considered. Natural follow-up works are to study the Joint Photographic Experts Group (JPEG) compression process as no prior model has been dedicated to this image format and also to improve, for uncompressed images, the scan artifacts extraction so as to better distinguish scanners of the same model.
To achieve those goals, in this chapter, we design new solutions to identify the scanner at the origin of a questioned digital document using fingerprints generated from features extracted manually from a large variety of scanned documents with a better forensic detectability compared with recent state-of-the-art methods as follows:
We start by exploiting the JPEG bitstream and find out that it is possible to extract a scanner fingerprint directly from the header of JPEG images. However, the extracted fingerprint gives only access to information about the brand of the device that have acquired the image. Moreover, JPEG headers are easy to manipulate. Thus, the proposed technique does not form a reliable solution for source identification. developing solutions that becomes desirable.
Considering the limitations of the JPEG based approach and convinced of the importance of working in uncompressed domain, we proposed to explore the forensic traces left by the scanner at each step of the acquisition process. The goal of the following solutions is to extract reliable and robust digital footprints in order to enhance the identification rates. The idea behind the proposed methods is to search for unique properties for each scanner model and then use those properties to identify the origin scanner of a digital image in question. To obtain these features, it is necessary to build a model that extract the scanning noise left by each device, and then, find a suitable measure to compare the features extracted from one image to the ones initially deduced from each device. Here, the model we build profits from the scanning noise contained in the high level subbands and extracts a set of features that fits the statistical distribution of those subbands. Then, a well-adapted distance measure is used to calculate the similarity between the features extracted from the investigated image and known scanners fingerprints. Note that a scanner fingerprint is generated by combining the features extracted from images acquired by it. However, neither the wavelet decomposition or any filtering operation is able to extract a perfect noise (without any image details).
When it goes to official and administrative documents, it is possible to incorporate the a priori knowledge of their content to reduce most of it and, therefore, guarantee a purer scanning noise across the whole image. This is the main contribution of the following approach.
This Chapter is organized into three parts. Firstly, we come back on the principles of JPEG compression and describe the general idea of our JPEG based technique. In its second part, we focus on Tagged Image File Format (TIFF) images and propose a wavelet-based forensics mechanism [13, 14] in which the statistical properties of the wavelet subbands are exploited so to build a scanner fingerprint one can extract from scanned documents. Then, in the third part, we introduce a novel framework to solve the problem of SSI [15] that take advantage of some prior knowledge of the type of the digitized document in order to refine the source identification process and to increase its performance.

3.1 Scanner Source Identification of JPEG Images

JPEG is one of the most commonly used standards for storing digital images. When compressing an image in JPEG, it takes up considerably less space while preserving quality. In fact, JPEG has been broadly studied for many forensics investigations. The question is: How can we use it for the sake of scanner origin identification?
As stated above, the first solution we worked on exclusively focuses on scanned images that have been stored accordingly the JPEG format. It uses ancillary information about the compressed bitstream that are stored in the JPEG file header.
Before entering into the details of our approach, let us first come back on the basic principles of the JPEG standard, that is to say the way it works so to convert a color image into a compressed bitstream. We will then see how to exploit information from the JPEG file header to find a unique fingerprint for each scanner and the performance of the proposed solution.

3.1.1 The JPEG standard

Figure 3.1 provides a summary of the essential steps required for compressing an image in the JPEG format:

- Image color transformation – The purpose of this task is to reduce information redundancy that exists in-between Red Green and Blue color channels. To do so, the digital image is converted from RGB into the YCbCr color space followed by a down sampling operation performed on the chrominance channels (Cb Cr).

- Discrete Cosine Transform (DCT) – Next, each 8x8 block of pixels is converted from raw pixels to their DCT signal representation. Note that the purpose of the DCT is to concentrate the information on a few number of low frequency coefficients and decorrelate information. To do so, pixels $P(x, y)$ from each 8x8 pixel blocks are transformed using the following formula:

$$P(u,v) = \frac{2}{N} c(u)c(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} P(x,y) \cdot \cos\left[\frac{\pi}{N} u \left(x + \frac{1}{2}\right)\right] \cdot \cos\left[\frac{\pi}{N} v \left(y + \frac{1}{z}\right)\right] \tag{3.1}$$

  where N is often equal to 8 and

$$\begin{cases} c(x) = \frac{1}{\sqrt{2}}, & \text{if } x = 0 \\ c(x) = 1, & \text{otherwise} \end{cases} \tag{3.2}$$

  Note that $P(0, 0)$ is the direct current (DC) coefficient which refers to the average brightness in the block. For u, $v > 0$, $P(u, v)$ are the alternating current (AC) coefficients.

- Quantization – Later, a quantization table is used to scale the DCT values. Due to sensitivity of human eyes to low frequencies, many of those values become zero. In fact, to get rid of the DCT coefficients that are not essential for recreating a close approximation of the original, they are independently quantized using psychovisual-dependent quantization tables (QT). More clearly, let us consider $C(u, v)$, the quantized DCT coefficient associated to the spatial frequency (u,v)

$$C(u, v) = \left\lfloor \frac{P(u, v)}{q(u, v)} \right\rfloor Q_f \tag{3.3}$$

  Where $P(u, v)$ is the original DCT coefficient, $q(u, v)$ is the quantization step and Q_f is the quality factor such that $0 \leqslant Q_f \leqslant 100$. This equation confirms that the JPEG compression is lossy and irreversible.
  This step is applied so as to turn the block-based matrix of DCT coefficients into a one dimensional (1D) sequence of coefficient. The basic idea of this transform is to ensure that two consecutive coefficients in the one sequence are closed in terms of spatial frequency.
  Notice that, as exposed in Fig. 3.1, quantized AC coefficients and quantized DC coefficients are treated separately. The Differential Pulse Code Modulation (DPCM) [144] ancomputes the difference between DC coefficients of adjacent blocks. Thus, it decreases the information entropy to encode if those coefficients are redundant. On the other hand, a Run Length Encoding (RLE) is applied to the AC coefficient after ordering them in 1D vector through a zigzag scan. Its role is to encode the coefficient vector in (run, value) pairs where run refers to the number of zeros preceding a non-zero AC coefficient, and value refers to the value of the non-zero AC coefficient.

- Entropy encoding – Finally, the encoded AC coefficients and the DC prediction errors presented in term of their difference between adjacent blocks are subjected to a Huffman entropy encoding using a set of Huffman tables.

Once DCT coefficients are entropically encoded, they are organized into a bitstream as exposed in Fig. 3.2. The bitstream consists of a sequence of segments led by markers that can be either a header or a body. The header purpose is to present some general information about the bitstream like the coder algorithm while the body represents the data itself. The body is mainly composed of six important tables which are quantization table for luminance, quantization table for chrominance, Huffman table for luminance DC, Huffman table for luminance AC, Huffman table for chrominance DC and Huffman table for chrominance AC.
Each of the segments starts with a 0xFF byte followed by a byte specifying the marker type. JPEG markers can be grouped into two general types: First, the stand-alone markers which consist of no data other than their two bytes. And then, the markers, that do not stand alone, are immediately followed by a 2-byte-long value that gives the number of bytes of data the marker contains. Most of these markers are listed in Table 3.1.
Any JPEG file in the bitstream domain must begin with a start-of-image (SOI) marker followed by an APP0 marker and end with an end-of-image (EOI) marker which must immediately follow the compressed data of the last scan in the image.

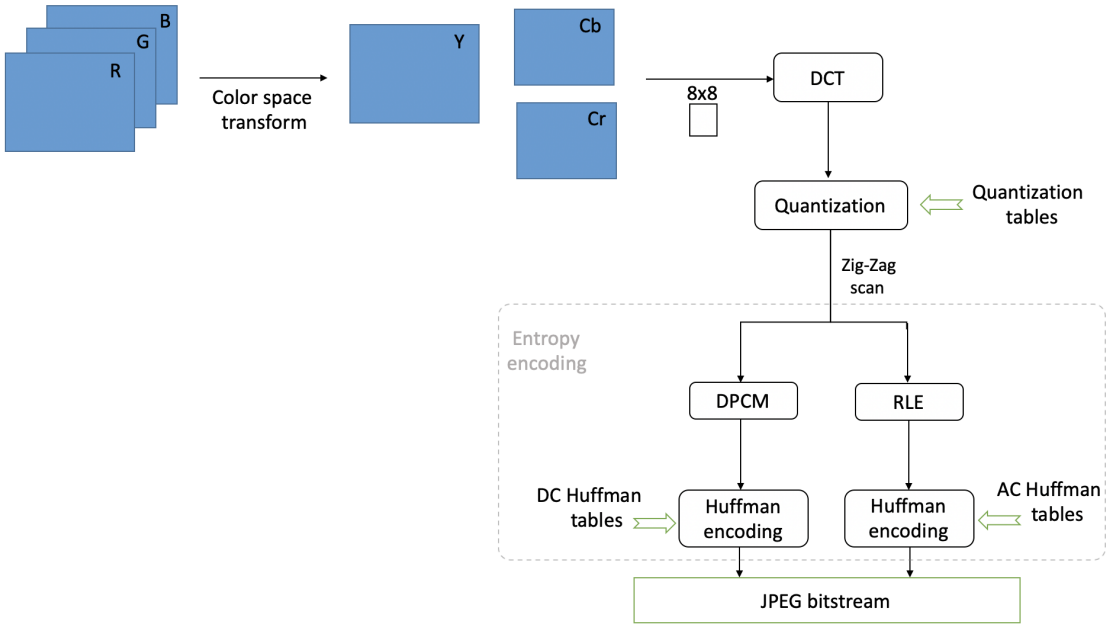


Figure 3.1: JPEG compression scheme

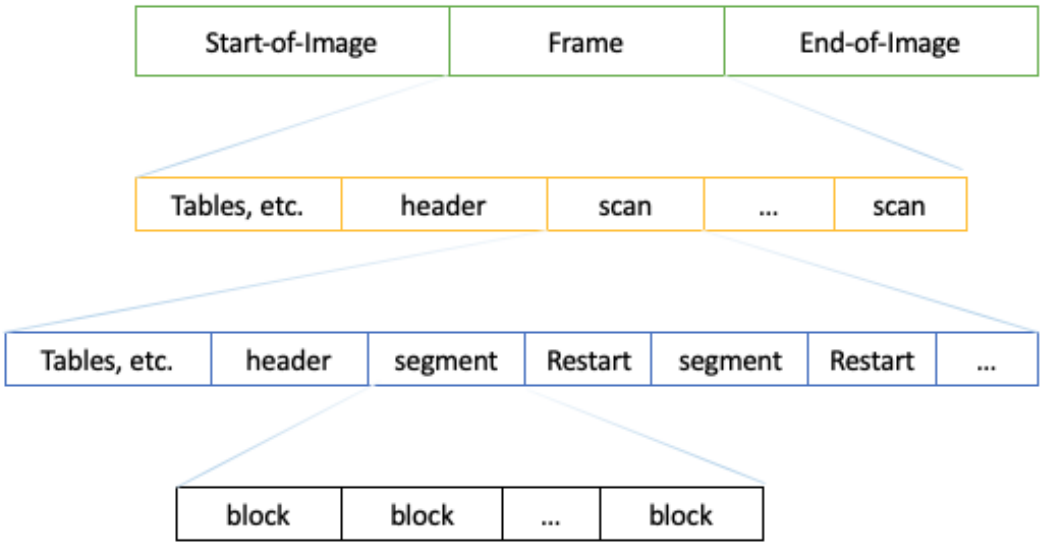


Figure 3.2: The organization of the JPEG bitstream

3.1.2 Proposed JPEG based approach

Manufacturers use different configurations when balancing the compression in their devices. Hence, we take advantage of this fact in order to find unique fingerprints in the JPEG bitstream. To do so, we have parsed bitstreams of different scanned images and find out that each manufacturer uses a different QT. The idea is, thus, to extract the QT for luminance and chrominance for which an example is illustrated in Fig. 3.3. We start by searching their marker in the JPEG bitstream and extracting a vector with 64 values for each table corresponding to the conversion of the 8x8 matrices to their zigzag ordering. The concatenation of both QT is the signature

Table 3.1: Common JPEG markers

Name	Code(HEX)	Description
SOI	FFD8	Start Of Image
SOFO	FFCO	Start Of Frame : indicates that this is a baseline DCT-based JPEG and specifies the width, height, number of components and component subsampling
SOS	FFDA	Start Of Scan : begins a top-bottom scan of image. In baseline DCT JPEG, there is generally a simple san.
DHT	FFC4	Define Huffman Table
DQT	FFDB	Define Quantization Table
APPn	FFEn	Application-specific
RSTn	FFDn	Restart : inserted every r macroblocks where r is the restart interval set by DRI marker. Not used if there was no DRI marker.
COM	FFFE	Comment
EOI	FFD9	End Of Image

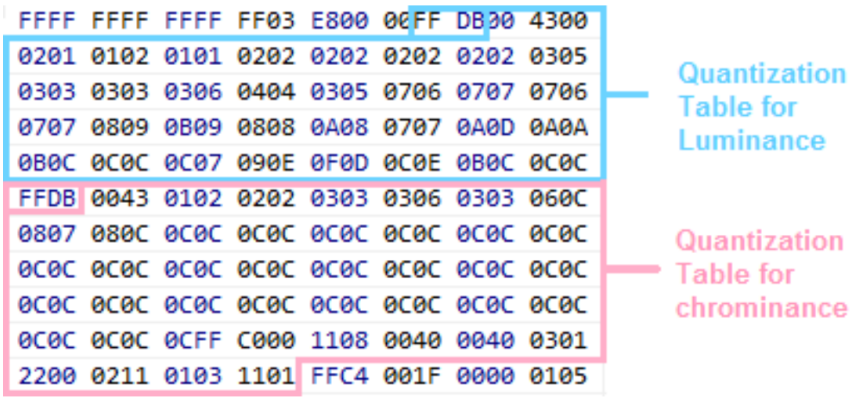


Figure 3.3: Example of QT defined by their marker in the JPEG bitstream

'tab' by which scanners will be authenticated. It is generated according to the Algorithm 1. The 64 values are ranging from 1 to 25 giving a large number of possible QT and, therefore, providing large choice of fingerprints.

The basic architecture of our verification scheme is given in Fig. 3.4 where **extractQT** is the function developed in Algorithm 1. To associate each image with the scanner that have acquired it, we first extract its luminance quantization table as well as its chrominance one. Then, a decision is made by comparing them to a database of known scanners' fingerprints (QTs). If the QTs of the questioned image and those of one of the scanners are the same, then the image

Algorithm 1: extractQT

Input: I : a JPEG image
Output: tab : the combined QT
$Bt \leftarrow Bitstream(I)$
$n, m \leftarrow size(Bt)$
`// Find starting indices of QT`
$o \leftarrow 0$
for $j \leftarrow 1 : n-1$ **do**
  if $strcmp(dec2hex(Bt(j),'FF')$ & $strcmp(dec2hex(Bt(j+1)),'DB')$ **then**
    $o \leftarrow o+1$
    $indices\_start(o) \leftarrow j+5$
`// Extract quantization table of luminance`
$j1 \leftarrow indices\_start(1)+2$
$j2 \leftarrow indices\_start(2)+2$
$TabY \leftarrow []$
for $j \leftarrow j1 : j1+63$ **do**
  $TabY(j-j1+1) \leftarrow Bt(j)$
`// Extract quantization table of chrominance`
$TabC \leftarrow []$
for $j \leftarrow j2 : j2+63$ **do**
  $TabC(j-j2+1) \leftarrow Bt(j)$
`// Generate image/scanner fingerprint`
$tab \leftarrow [TabY, TabC]$

can be linked to that scanner. Otherwise, we can confirm that none of the known scanners is at the origin of that image.

3.1.3 Experimental results

The following experiments have been conducted considering five distinct models of scanners. As depicted in Table 3.2, some of them are from the same manufacturer. We collected 45 images scanned and saved in JPEG format at the same quality factor from each scanner. By extracting and comparing images' fingerprints, we found that images acquired by the same scanner have the same fingerprint. However, and as it can be seen in the example displayed in Table 3.3, scanners of the same manufacturer share the same QT while these later differ from one brand to another. Thus, this solution can only serve for identifying the brand of the scanner and not the scanner itself.

3.1.4 Discussion

The main advantage of this method is that it is simple and fast. Indeed, a scanner's fingerprint can be extracted from only one image; no training task is needed.

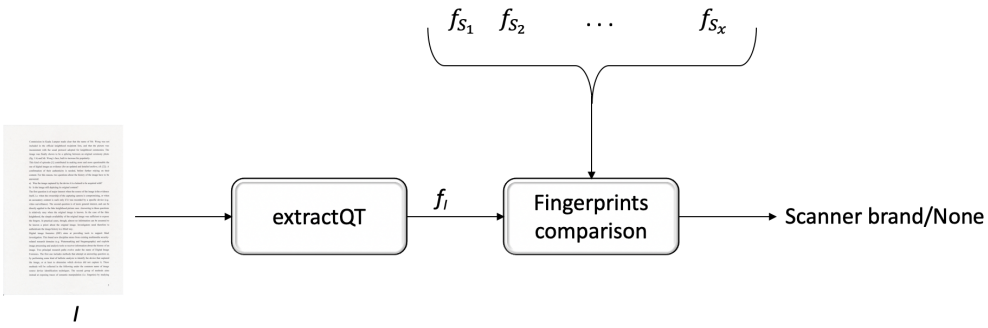


Figure 3.4: General scheme of the proposed JPEG-based verification scheme. I is the investigated JPEG image, f_I is its fingerprint which will be compared to each scanner fingerprint f_{s_i} where i:1..x and x is the number of known brands

Table 3.2: Scanners used in our experiments

Id	Brand	Model
s1	Canon	CanonScan 9000F MKII -1
s2	Epson	Perfection V39
s3	Epson	Perfection V550 Photo
s4	Canon	CanonScan 9000F MKII -2
s5	HP	Scanjet Pro 2500 F1

Moreover, the fingerprint is of fixed value and not an estimation error of which varies from an image to another. However, even the previously shown results give a first positive proof of concept of the proposed technique, a manipulation of the EXIF header or a different re-saving of the image may induce a miss-classification. This solution can mostly serve in the situation of source scanner linking where the forensic investigator need to know whether two images of unknown sources are taken by scanners of the same brand or not.
In the following, we will focus on digital images saved in the TIFF format.

3.2 Source Scanner Model Identification Based on Wavelet Features

Digitized documents can be stored in a variety of formats, including GIF, PNG, JPEG and TIFF, depending if they will be saved with or without lossy compression. In the previous section, we addressed JPEG images having some features such that entropy encoding tables and QTs contained in the JPEG header. This later often includes information about the acquisition device and its settings is thought to be the most basic approach to identify an image origin. However, it can be easily altered, resaved or deleted. Another problem is the limitation to scanner brand identification. Therefore, there is a significant deal of interest in developing forensic techniques based on intrinsic fingerprints to overcome these problems.

Table 3.3: QT of luminance and chrominance of some scanners used in our experiments

Scanner Id	QT for Luminance	QT for Chrominance
S1	3 2 2 3 5 8 10 12 2 2 3 4 5 12 12 11 3 3 3 5 8 11 14 11 3 3 4 6 10 17 16 12 4 4 7 11 14 22 21 15 5 7 11 13 16 21 23 18 10 13 16 17 21 24 24 20 14 18 19 20 22 20 21 20	3 4 5 9 20 20 20 20 4 4 5 13 20 20 20 20 5 5 11 20 20 20 20 20 9 13 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20
S2	2 3 3 3 5 5 12 11 2 3 3 8 4 12 10 16 2 4 5 4 10 11 11 22 3 4 5 12 11 11 16 23 3 10 8 13 14 17 15 24 7 6 14 17 19 12 22 21 7 18 14 20 21 20 20 20 16 13 21 21 24 18 255 219	4 5 5 5 20 20 20 20 4 9 9 20 20 20 20 20 5 20 20 20 20 20 20 20 13 13 20 20 20 20 20 20 11 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 255 192
S3	2 3 3 3 5 5 12 11 2 3 3 8 4 12 10 16 2 4 5 4 10 11 11 22 3 4 5 12 11 11 16 23 3 10 8 13 14 17 15 24 7 6 14 17 19 12 22 21 7 18 14 20 21 20 20 20 16 13 21 21 24 18 255 219	4 5 5 5 20 20 20 20 4 9 9 20 20 20 20 20 5 20 20 20 20 20 20 20 13 13 20 20 20 20 20 20 11 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 255 192

An alternative solution is the analysis of the scanner defects using the traces they left in the acquired Images. Chapter 2 has presented an overview on existing scanner model identification approaches that rely on extracting various traces from scanned images. These ones can be categorized into five classes as shown in Fig. 3.5.

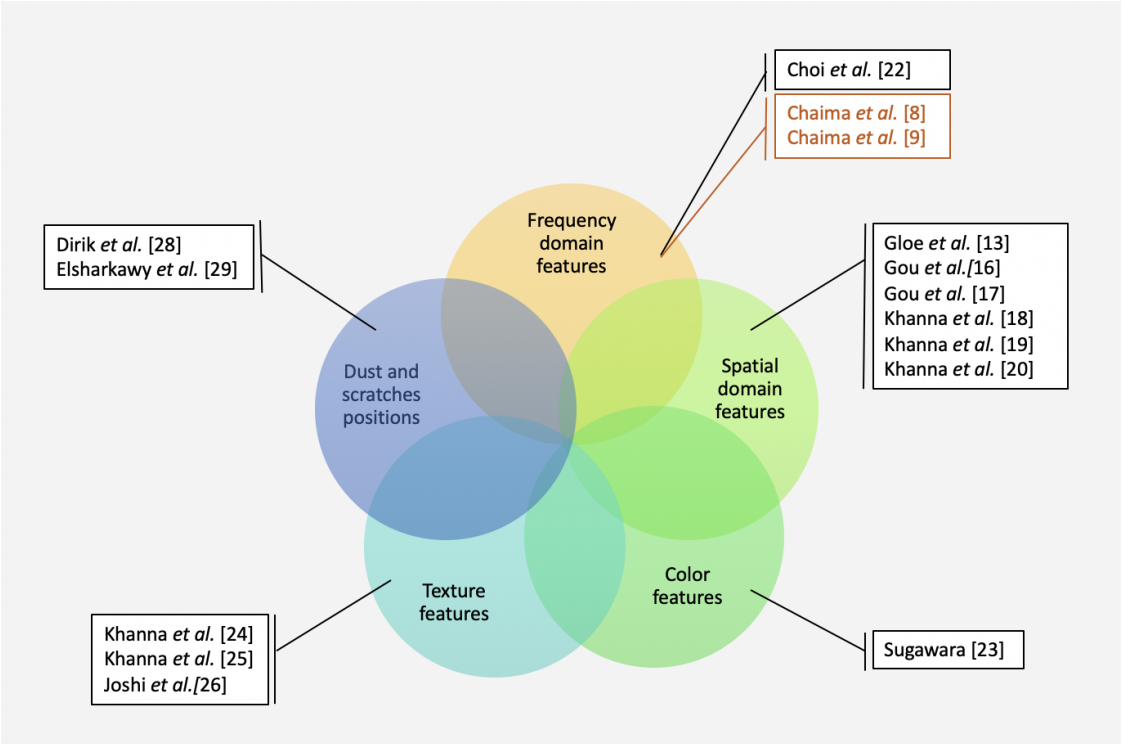


Figure 3.5: Summarizing related works on source scanner identification

As we are interested in the analysis of the details (various traces) in scanned images, our focus was oriented to the frequency domain. Thus, the study we propose in this section is performed in the frequency domain and more precisely in the wavelet domain (See Appendix B) so as to take advantage of the fact that the subbands of details of the first level of decomposition of an image give access to the scanning noise contained in its subbands. In fact, wavelets have lately emerged as a powerful analysis tool because of its ability to offer multi-scale and orientation representation via subbands. In many DIF applications, decomposition of images utilizing basis functions such as wavelets has proven to be particularly effective [145, 146]. One reason is that statistical regularities in such decompositions can be exploited. The model we propose is made up of first-order statistics that can capture the regularities found in scanned images.
Based on a supervised classification, our work consists in two stages; The first stage involves collecting and scanning documents by each of the available scanners. Those scanned documents are, then, divided into training and testing image sets and are used to evaluate the performance of the proposed system. This process, illustrated in Fig. 3.6, is adopted by almost all SSI methods.

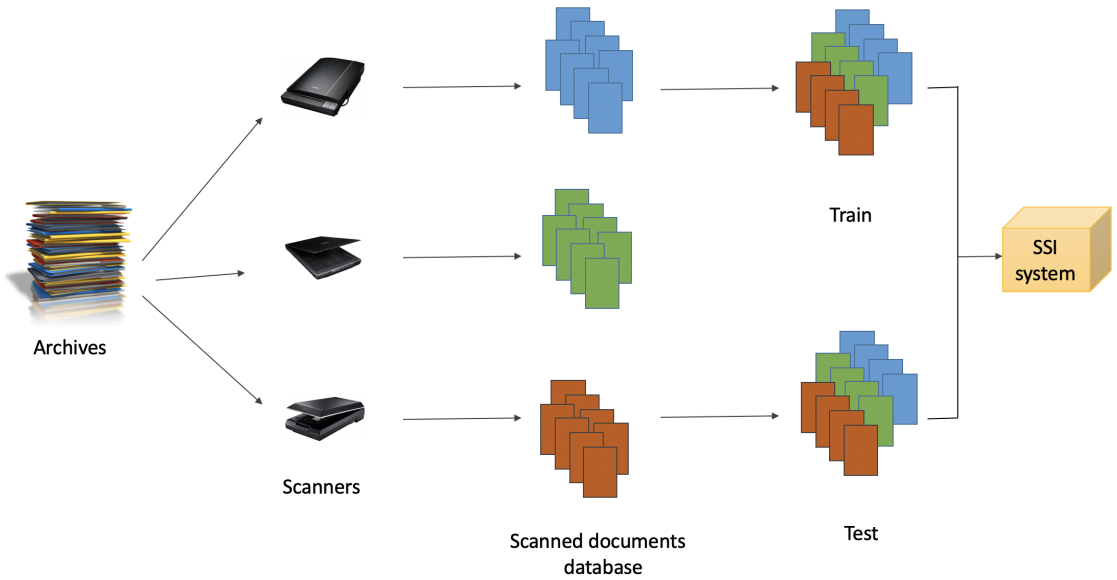


Figure 3.6: An illustration of the different stages of a typical SSI evaluation process. Once the hard-copies to scan are prepared, they are scanned by each of the scanner. The scanned images are, then, divided into two separate sets and used to train and test the SSI system.

The rest of this section is organized as follows. First, we present the proposed system architecture, as well as the process of extracting fingerprints. Then, we explain how the origin of an image can be identified. Experimental results are provided in Sub-Section 3.2.3. Finally, we draw the concluding remarks and the next challenges.

3.2.1 Image and scanner fingerprints

The overall structure of our proposal to identify a scanned image source is depicted in Fig. 3.7. It is composed of two main parts: Image scanner fingerprint extraction and scanner identification. The first aims at estimating reliable wavelet parameters

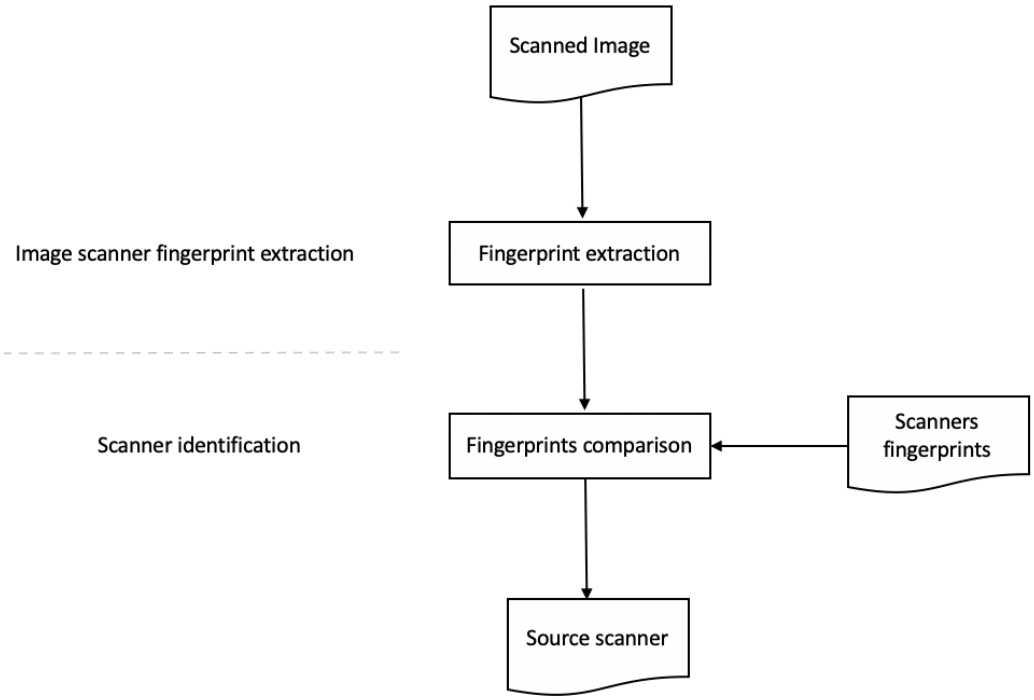


Figure 3.7: Global architecture of the proposed system

from the two dimensional high frequency wavelet subband HH, while, the role of the later is to identify, when an unknown image is presented, the scanner used to acquire it. The two steps have a common process for scanner image fingerprint extraction.

The process of extracting a scanner image fingerprint, a signature that will be used later in a scanner identification scheme, is illustrated in Fig. 3.8.

The proposed method is applied to the RGB color chanel separately since

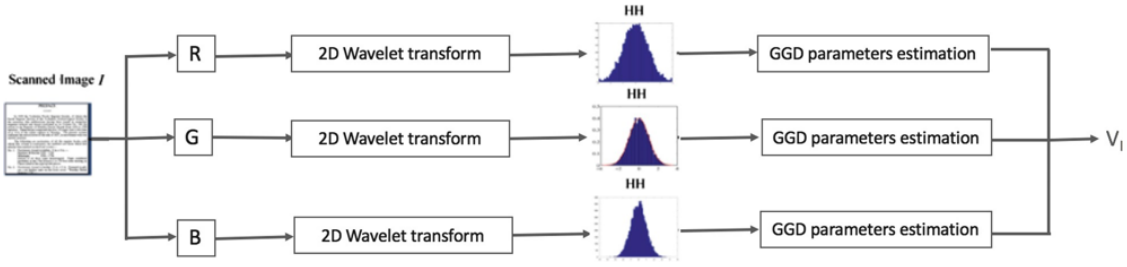


Figure 3.8: Fingerprint extraction for a scanned image

scanners use individual R, G and B sensors to scan the document. Then, the first decomposition level of the wavelet transforms of each color channel (Red R, Green G, Blue B) of a given image I is calculated to get 4 subbands: Low-Low (LL), Low-High (LH), High-Low (HL), and High-High (HH) as shown in Fig. 3.9. Among these subbands, we are interested only in the HH subbands of each color channel in which the noise is concentrated [147].

In this work, we choose to use the Symlet wavelet which showed good performance in features extraction compared to the other wavelet families particularly for printed documents. Furthermore, the Symlet4 is recommended as the most powerful in

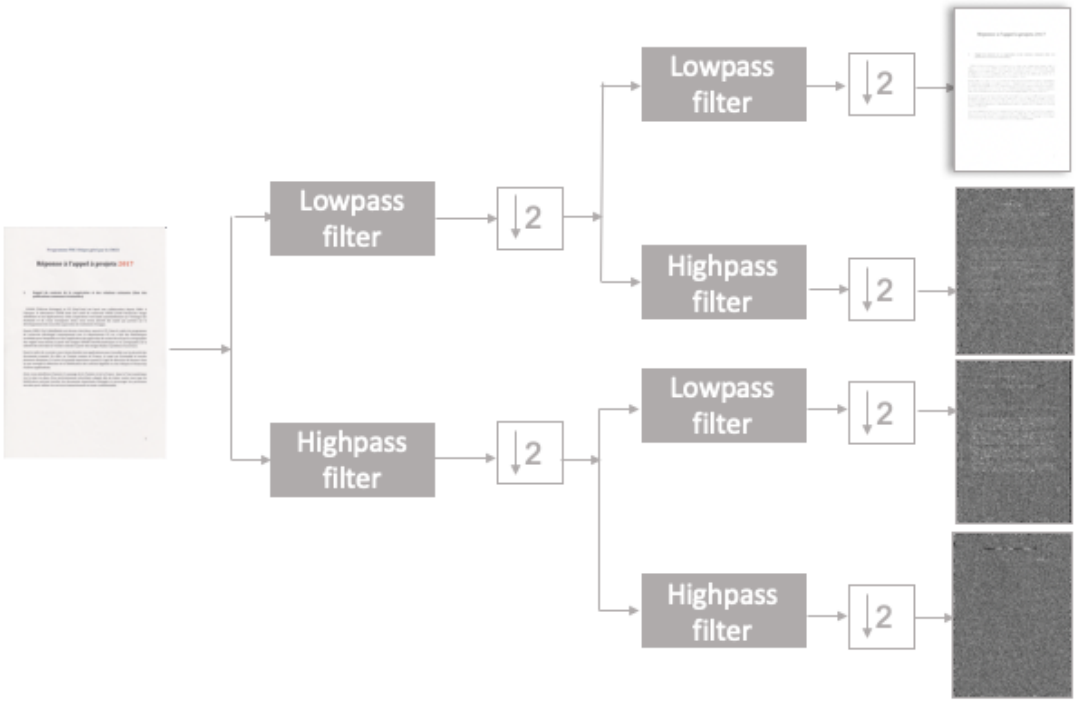


Figure 3.9: Example of a decomposition of an image into four subbands using a one lever 2D DWT

analyzing peaks that resemble Gaussian distribution curves [148]. The features of this wavelet make it a good candidate as a scanner fingerprint as shown in [149]. The constitution of our scanner signature stands on the fact that the distribution of wavelet coefficients in a subband of details is approximated by a generalized Gaussian [150]. Given a subband H and a color channel C, the density probability of that subband C_H can be approximated such as

$$f_{C_H}(x) = \frac{\beta}{2\alpha\Gamma\left(\frac{1}{\beta}\right)} \exp^{\left[-\left[\frac{|x|}{\alpha}\right)^{\beta}\right]} \tag{3.4}$$

where x, α, β and μ represent the subband samples and the parameters of scale, shape and location, respectively, while $\Gamma(.)$ is the known Gamma function defined by

$$\Gamma(z) = \int_0^{\infty} e^{-t} t^{z-1} dt, u > 0 \tag{3.5}$$

Note that α models the width of the probability density function peak and β is inversely proportional to the diminishing rate of the peak.
For our task, it is more interesting to estimate these parameters for the HH subband only since the wavelet coefficients in this subband are dominated by noise which is none other than sensor noise while the other subbands contains more image details. These parameters will constitute our signature.
Motivated by our observations, and to decrease the number of parameters to be estimated, we simply approximate the value of μ with zero. As a result, we only consider the α and β parameters to characterize the distribution of noise in a subband.
To the best of our knowledge, there are three methods for estimating the generalized

Gaussian distribution (GGD) parameters: (1) Moment estimation (ME) [151] (2) Entropy matching estimation (EME) [152] (3) maximum likelihood estimation (MLE) [153]. Almost the same accuracy can be obtained by these methods. In this work, we propose to calculate the GGD parameters using the MLE method with the Newton-Raphson's iterative algorithm [154]. On this basis, we obtain a pair of estimated parameters per color channel from each corresponding HH wavelet subband.
The image scanner fingerprint V_I have this form:

$$V_I = \begin{bmatrix} V_I^R \\ V_I^G \\ V_I^B \end{bmatrix} = \begin{bmatrix} \begin{pmatrix} \hat{\alpha}_R^{HH}, \hat{\beta}_R^{HH} \end{pmatrix} \\ \begin{pmatrix} \hat{\alpha}_G^{HH}, \hat{\beta}_G^{HH} \end{pmatrix} \\ \begin{pmatrix} \hat{\alpha}_B^{HH}, \hat{\beta}_B^{HH} \end{pmatrix} \end{bmatrix} \tag{3.6}$$

In our approach, the scanner signature is obtained via the fusion of the signatures of several images scanned with it using the median operation. The choice of the median has been demonstrated in [13]. In that case, the identifier of a scanner S from the signatures of N images is then such that

$$\begin{array}{l} V_S^R = median(V_{S,i}^R, i = 1..N) \\ V_S^G = median(V_{S,i}^G, i = 1..N) \\ V_S^B = median(V_{S,i}^B, i = 1..N) \end{array} \tag{3.7}$$

where $V_{S,i}^R$, $V_{S,i}^G$ and $V_{S,i}^B$ are the features vectors of the HH subband of the R, G and B color channels of the i^{th} image scanned with the scanner S, respectively.

3.2.2 Image origin predictor

Identifying the origin of a digitized document is based on the comparison between the signature of several scanners and the signature extracted from a questioned document.
There are numerous ways to measure the difference between two distributions. However, the Kullback-Leibler divergence (KLD) [155] is the most adapted and appropriate measure for comparing statistical features exhibiting generalized Gaussian distributions, that is why we propose to use it in our work. Herein, the source scanner of a document will be the one that minimizes this measure.
Since the coefficients of the detail subbands of the first level of the wavelet decomposition of an image follows generalized Gaussian distribution, the KLD of the color channel j is given by

$$KLD\left(\alpha_I^j, \alpha_s^j, \beta_I^j, \beta_s^j\right) = \log\left(\frac{\beta_S^f \alpha_I' r\left(\frac{1}{\beta_i}\right)}{\beta_I^j \alpha_S^j \Gamma\left(\frac{1}{\sigma_S'}\right)} + \left(\frac{\alpha_S'}{\alpha_I^j}\right)^{\beta_I^j} \frac{r\left(\frac{\beta_1^j+1}{\beta_S^J}\right)}{r\left(\frac{1}{\beta_S'}\right)} - \frac{1}{\beta_S'}\right) \tag{3.8}$$

where $j \in$ R,G,B. The global distance between the query image Q and the scanner S is given by summing along the three color channels

$$\begin{array}{l} \mathrm{KLD}_{Q,S} = \sum_{j=1}^{3} KLD(\alpha_Q^j, \alpha_S^j, \beta_Q^j, \beta_S^j) \\ \qquad = \sum_{j=1}^{3} \left[\log\left(\frac{\beta_S^j \alpha_Q^j \Gamma\left(\frac{1}{\beta_Q^j}\right)}{\beta_Q^j \alpha_S^j \Gamma\left(\frac{1}{\beta_S^j}\right)}\right) + \left(\frac{\alpha_S^j}{\alpha_Q^j}\right)^{\beta_Q^j} \frac{\Gamma\left(\frac{\beta_Q^j+1}{\beta_S^j}\right)}{\Gamma\left(\frac{1}{\beta_S^j}\right)} - \frac{1}{\beta_S^j}\right] \end{array} \tag{3.9}$$

Id	Brand	Model
s1	Canon	Lide 210
s2	Epson	Perfection V39
s3	Epson	Perfection V370 Photo
s4	Epson	Perfection V70 Photo
s5	HP	Scanjet Pro 2500 F1

Table 3.4: Scanners used in our experiments

Table 3.5: Confusion matrix for the proposed method (in %) - TIFF images

	S1	S2	S3	S4	S5
S1	87.6	12.4	0	0	0
S2	7.2	92.8	0	0	0
S3	0	0.8	99.2	0	0
S4	0	0	0	100	0
S5	0	0	0	1.2	98.8

3.2.3 Experimental results

To empirically evaluate the effectiveness of our previous method, we tested it on five different scanners of various native resolutions as shown in Table 3.4. First, we created a test database of 100 documents of varied content all scanned with each scanner. These documents contain black and white or colored text. Some of them includes figures and tables. We give some examples of this test image dataset in Fig. 3.10. The digitized documents or equivalently images were saved in TIFF format, that is to say in a format of raw data without loss of information, at a resolution of 300 dpi, leading thus to the constitution of a database of 500 images. Notice that, this resolution is the commonly used in general practice [156] whatever the activity domain; being the default parameter value of scanners. 50 out of 100 images were chosen randomly for generating each scanner fingerprint, while the remaining 50 images were used to evaluate the detection performance through a testing phase. The following results are given in average. Indeed, the training and testing phases have been repeated five times to obtain the final performance measures.

3.2.3.1 Identification performance evaluation

We give in Fig. 3.11, the distributions of the couples (α_{HH}, β_{HH}) of the HH subband of the red color channel for all test images. It can be seen that these parameters discriminate well the images according to the scanner that acquired it which proves that the noise is different from one scanner to another.

We study the performance in terms of classification accuracy by measuring the percentage of images from scanner model correctly identified. We obtain a matrix as shown in Table 3.5 and a total accuracy 95.6%. Notice also that these detection results remain the same regardless of the content of the scanned document. However, we notice a low rate of 87.6% which may be due to the close similarity between the fingerprints of S1 and S2.

To go further in this analysis, Figure 3.12 illustrates an example of an image scanned with different scanners and the HH subband issued of a one level DWT applied on each one. We can notice the differences between the HH sub-bands which tell the difference between the scanners fingerprints that somehow explains the discriminability of the parameters (α, β).

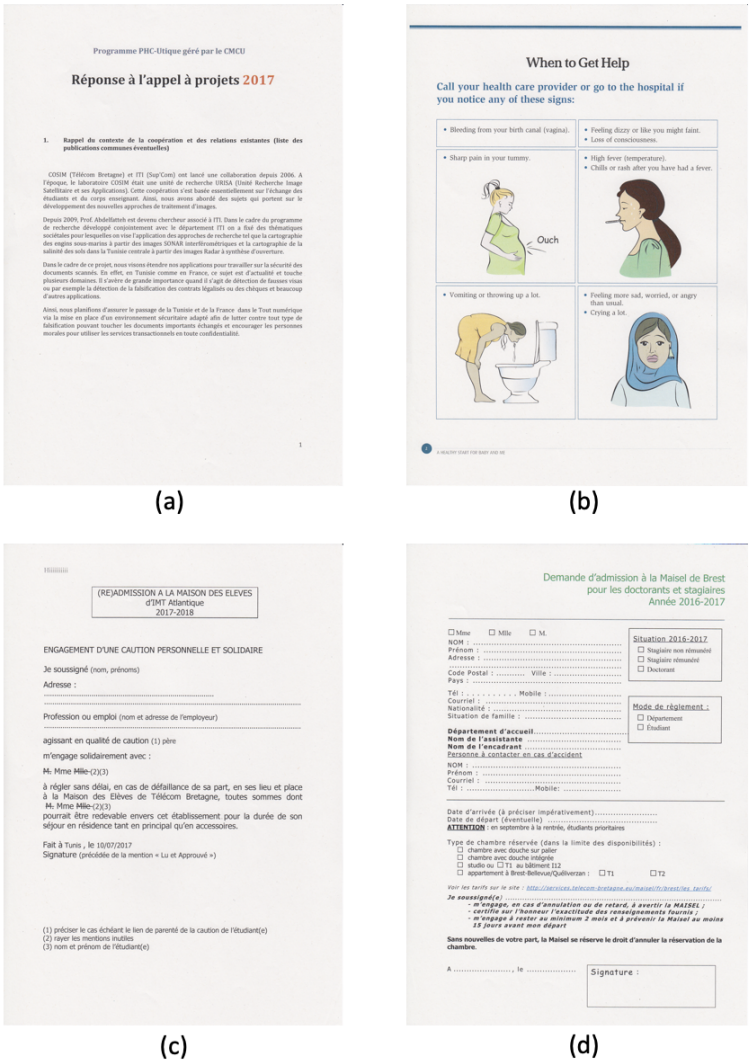


Figure 3.10: Examples of images of different types used in our experiments. (a) Text color image (b) Text color image with pictures (c) Black and white text image with shapes (d) Color text image with shapes

3.2.3.2 Evaluation of the effect of the subband choice

To highlight the relevance of the choice of the HH subband from each color channel, we choose to conduct the same experiments realized previously, but as this time, we change the subband to work on. More clearly, we extract the image scanner fingerprint from LL, HL and LH. Table 3.6 shows the identification accuracy of the proposed system for each of these subbands obtained by averaging the resulting accuracies of five different tests. It can be noted that the identification accuracy decreases with those subbands and, in particular, with the LL subband which has the lower level of noise. Indeed, this subband corresponds to an approximated version of the original document as shown in Fig. 3.13.

3.2.3.3 Evaluation of the effect of the color channels

The color channel is also an important parameter to consider. The following experiments are conducted by considering only one color channel each time which, by

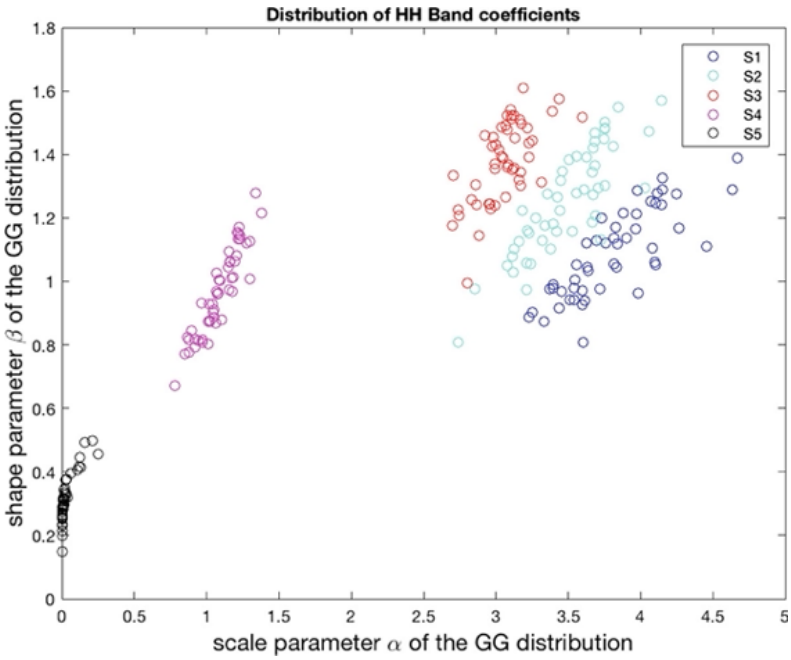


Figure 3.11: Distribution of couples (α_{HH}, β_{HH}) of the HH subband of the red color channel for different scanners computed using 50 images

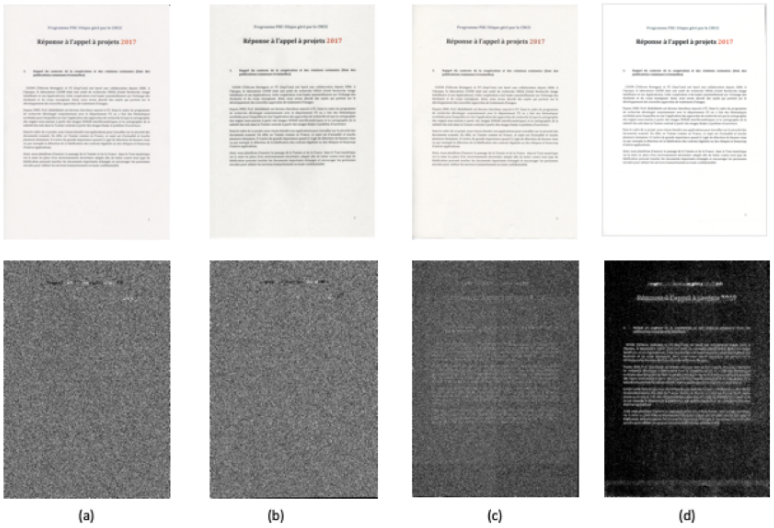


Figure 3.12: Scanned images and its corresponding HH subband of the DWT of the blue channel. (a) s1 (b) s2 (c) s4 (d) s5

the way, reduces the size of the image's fingerprint. The identification performance corresponding to those tests is shown in table 3.7. It can be noticed that the decline in the accuracy is not important especially for the green channel. This may be justified by the correlation between the color channels, and that the green channel contains more of the image information while the blue one is not so relevant. Considering the results of table 3.5, using the color channels jointly is still better to avoid miss classification. The difference of performance is around 1% compared to using the green channel separately.

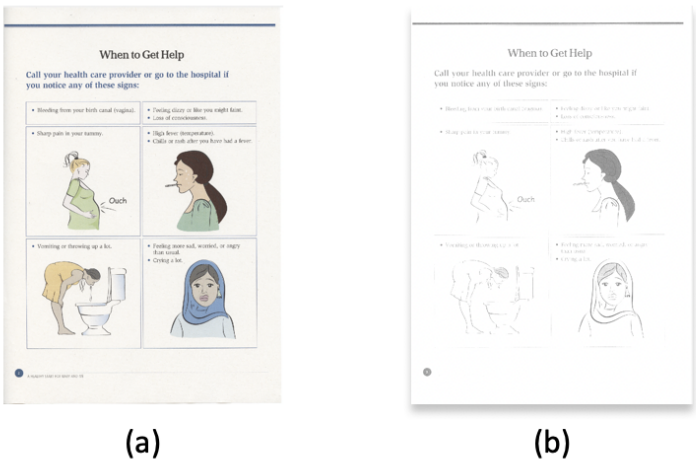


Figure 3.13: Example of an image and the LL subband of its red channel

Table 3.6: Identification accuracy for each subband

Subband	Accuracy
LL	23.52%
LH	89.44%
HL	76.96%

3.2.3.4 Comparison with a recent method

The proposed method shows greater accuracy in discriminating different scanners comparing to the last and the most powerful existing method [16] which uses spectral noise in the frequency domain.
As it can be seen in Table 3.8, Choi *et al.* [16] method is not able to correctly discriminate scanners of the same brand (Epson).

3.2.3.5 Evaluation of the robustness against lossy compression

In the following, we investigate the effect of lossy compression on the reliability of our scanner model identification technique. Indeed, it is likely that, in practice, scanned documents will be stored in the JPEG format.
To conduct this experiment, we built another dataset consisting of single JPEG images created by compressing the uncompressed previously generated TIFF image dataset with a quality factor QF equal to 75. Such a quality factor value is the default one of most image software. Table 3.9 shows average accuracy of the dedicated classifier for scanned documents saved in JPEG format.

Table 3.7: Identification accuracy for each color channel

Color channel	Accuracy
R	90.88%
G	94.88%
B	90.96%

Table 3.8: Confusion matrix for the Choi *et al.* [16] method (in %) - TIFF images

	S1	S2	S3	S4	S5
S1	100	0	0	0	0
S2	0	1.2	0	98.8	0
S3	0	0	100	0	0
S4	0	0.8	0	99.2	0
S5	0	0	0	0	100

Table 3.9: Confusion matrix for the proposed method (in %) - JPEG (QF=75) images

	S1	S2	S3	S4	S5
S1	44.4	31.6	8	17.2	0
S2	3.2	71.6	25.2	0	0
S3	2.4	20	75.6	6	0
S4	18.8	0	0.8	80.4	0
S5	0	0	0	13.6	86.4

The proposed method maintains an average classification accuracy of 67.6% despite the decline in performance. This demonstrates that our identification method is robust to JPEG compression.
The relative decline of accuracy in this experiments is due to the introduction of the JPEG quantization noise in the HH subband which deviates significantly the sensor noise characteristics from the original one.

3.2.4 Discussion

In this section, we presented a new approach to identify the scanner that was used to scan a document. We rely on a fingerprint extracted from the wavelet decomposition of the image which highlights the statistical properties of the scanning noise of a given scanner. The experimental results show a good identification accuracy independently of the content of the scanned documents.
We conclude that, for a fixed scanning resolution, the parameters (α,β) are discriminative for different scanner models. However, these parameters are almost similar for different devices of the same model. In Fig. 3.14, we show an example of the parameters distributions of four Epson scanners where two of them are of the same model "Perfection V39" (d1,d2) and the others (s1,s2) are of the same model "Perfection V370". Therefore, the proposed model can be only exploited for scanner model identification and further work is desirable for device recognition.
Moreover, another challenging part is the impact of the post-acquisition compression process. Experiments have shown that JPEG compression modifies the scanning noise model by introducing a quantization noise. The later can be suppressed by tresholding the wavelet coefficients. We consider this issue as a future work.

3.3 Semi-blind source scanner Identification

One of the key challenges in digital content forensics for images and scanners, is the capacity to extract exactly the acquisition noise from the image or equivalently to exactly remove the image content. More clearly, the more the image content remove, the greatest are the identification performance.

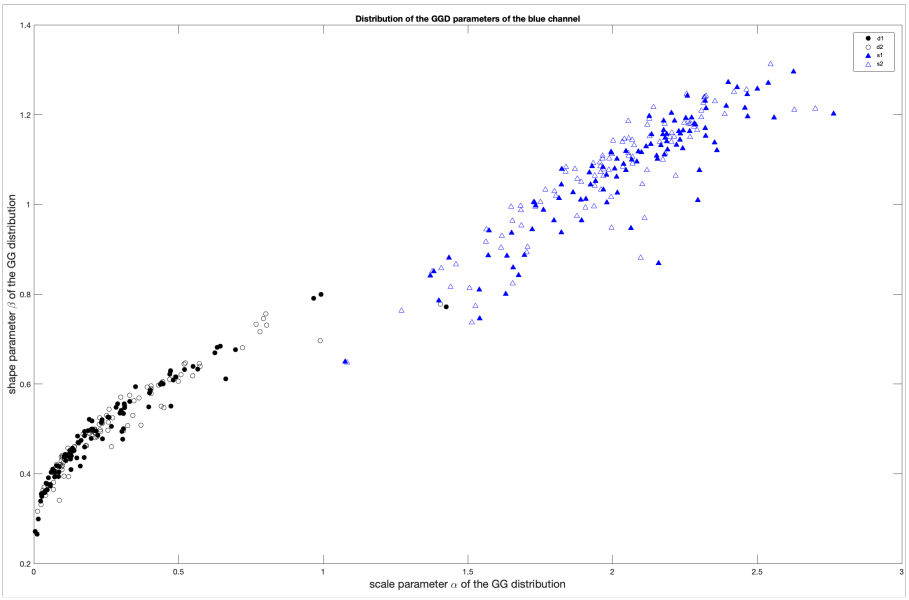


Figure 3.14: An example of the distribution of couples (α_{HH}, β_{HH}) of the HH subband of the blue channel for four scanners of the same brand (Epson). s1 and s2 as well as d2 and d2 are each of the same model

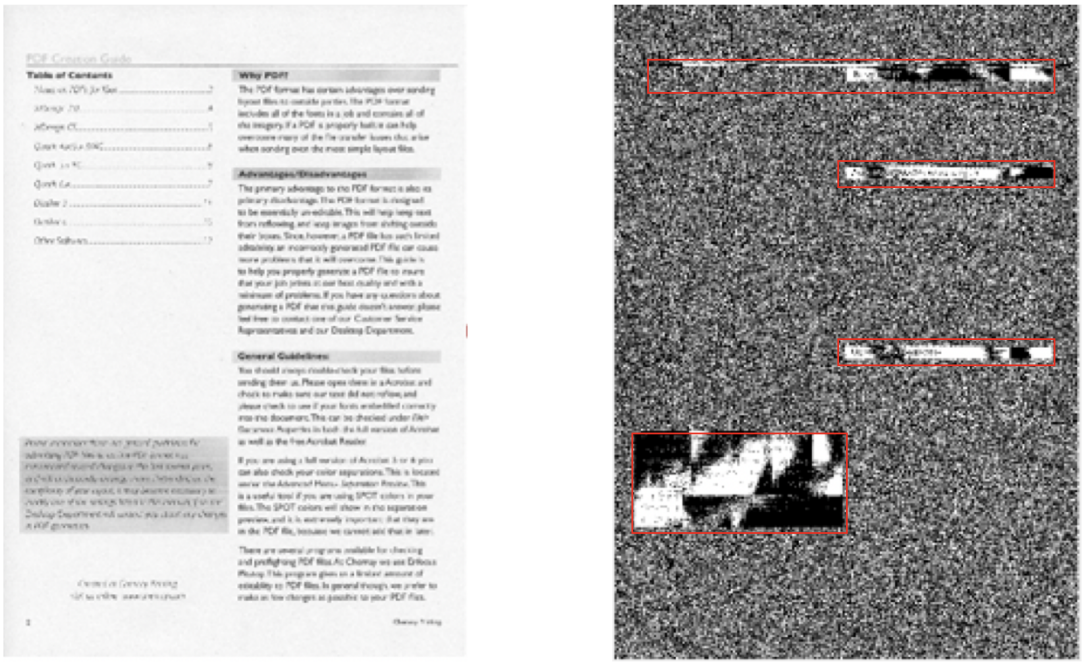
Thus, to guarantee that the parameters (α,β) defined in the previous section remain identical for different images digitized with the same scanner, we should make sure that their content is removed perfectly. In practice, working with the observed image only, those parameters are still influenced by the image content because the optimal denoising filter is difficult to obtain. For instance, Fig. 3.15 shows how the HH subband is influenced by different image details which, consequently, affect the identification process.

In many cases, especially when handling official documents, it is possible to get such prior knowledge (e.g. registration forms, contracts, agreements). The solution we present in this Section takes advantage of this proposal. More clearly, we consider a framework where some administrative documents are filled by hand by people before being scanned and transmitted. We propose thus to introduce the very first solution that associates a digital document with its source device in a semi-blind way. Again, we compare the digitized document with the original one to extract all kinds of noise resulting from the scanning process. This noise is then exploited to generate a unique fingerprint for each scanner that is by next compared to scanner fingerprint extracted from the test document. Finally, the scanner having the most similar fingerprint to the one of the test document is identified as the source scanner.

The remainder of this section is organized as follows. In Sub-section 3.3.1, we describe how we build the image and scanner fingerprints in our approach. The image origin predictor is detailed in Sub-section 3.3.2. Experimental evaluation plan and results are presented in Sub-section 3.3.3. Finally, Sub-section 3.3.4 gives our conclusions.

3.3.1 Image and scanner fingerprints

The global architecture of our scheme is shown in Fig. 3.16. It consists of three steps. First, it extracts the noise present in a scanned document. Let us recall that



(a) (b)

Figure 3.15: Example showing a scanned image (a) and its HH subband (b). We notice some image details in the HH subband that we have surrounded in red

such a document is an official form a user has to print and fulfill before scanning it. This process takes advantage of the original document; we also name reference document. Once the scanned document noise extracted, the scanner fingerprint (signature) is generated. Then, to identify the source of the questioned document, we compare this fingerprint to the ones of some scanners using an adequate metric and a decision is taken. Scanners fingerprints have been generated in the same way and stored in a database. We come back on these different steps in the sequel. The task of noise extraction is conducted as follows. Let us denote by I_r and $I_{r,i}$ the reference document and the questioned document filled and scanned by the user i, respectively. In order to reduce the amount of information to process, we first convert the color images I_r and $I_{r,i}$ into their grayscale versions $\tilde{I}_r^g$ and $\tilde{I}_{r,i}^g$ using a simple color transform [157] as follows

$$\tilde{I} = 0.21 * I^R + 0.72 * I^G + 0.07 * I^B, \tag{3.10}$$

where I^R, I^G and I^B are the red, green and blue components of the image I, respectively.
Then, due to the scanner acquisition resolution, $\tilde{I}_r^g$ is resized so as to have the same dimensions as $\tilde{I}_{r,i}^g$. By next, since the scanned and the reference documents are usually not correctly aligned, we also apply the feature-based registration algorithm F [158]. This one is based on Speeded-Up Robust Features (SURF) [159] to the resized $\tilde{I}_r^g$ as depicted in

$$\tilde{I}_r^{reg} = F(\tilde{I}_r^g), \tag{3.11}$$

where $\tilde{I}_r^{reg}$ is the registered version of $\tilde{I}_r^g$. We modify the reference document and not the scanned document so as to preserve at best the scanner signature, i.e. not

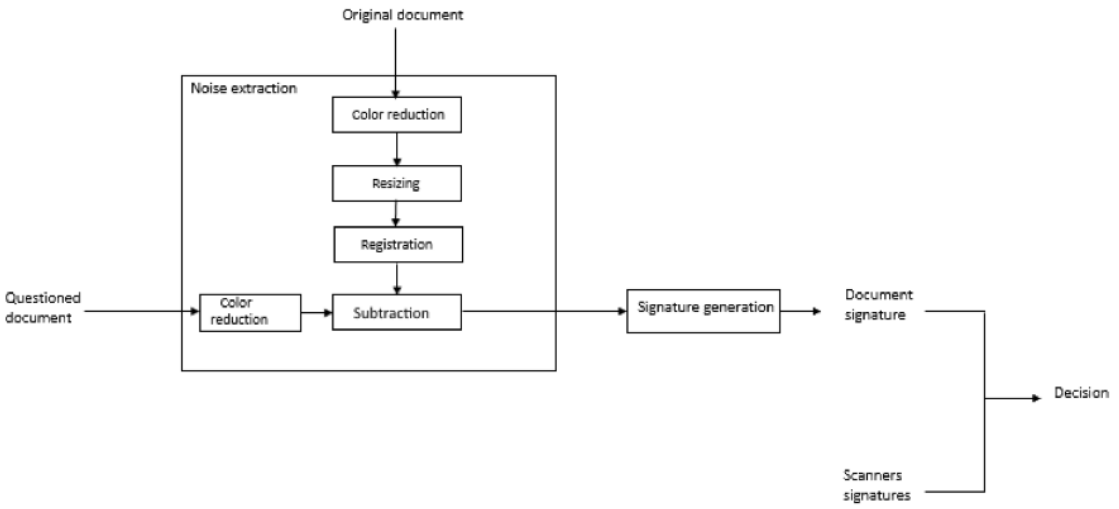


Figure 3.16: Flowchart of the proposed source scanner identification approach

altering the scanned document noise. Indeed, modifying the scanned document to make it match the original one will have as consequence to alter the scanner noise. Finally, the subtraction of the aligned image $\tilde{I}_r^{reg}$ from the scanned document image $\tilde{I}_{r,i}$ gives access to the noise image R_i :

$$R_i = \tilde{I}_{r,i} - \tilde{I}_r^{reg} \tag{3.12}$$

R_i is a combination of the different noises depicted in Chapter 2 along with some residues of the document content (e.g. modifications made by the user before he or she scanned the document).
As the scanning process is made line by line, the scanner noise is repeated over the noise image R_i along the scan direction [20]. Averaging information along this direction will reduce the power of the random noise while enhancing the one of the scanner noise. This is why we consider the fingerprint V_i of a given document as follows

$$V_i(l) = \frac{1}{K}\sum_{k=1}^{K} R_i(k,l) \quad , 1 \leq l \leq L, \tag{3.13}$$

where K and L are respectively the number of rows and the number of columns of R_i.

3.3.2 Image origin predictor

To make the source scanner identification possible, we need to build a database with the signature of several scanners. Figure 3.17 illustrates the different steps for generating a scanner fingerprint. As it can be seen, it works similarly to the one above. The main difference is that the scanner fingerprint is derived from several documents. These ones have been printed and scanned. Then, their noise images have been computed before being averaged. The scanner fingerprint corresponds to the average of the pixels' lines of this averaged noise image.
More clearly, let us consider P available scanners where $1 \leq j \leq P$.

The fingerprint S_j of the j^{th} scanner is computed as follows: From a set of M differently filled documents that have been printed and scanned, we derive an

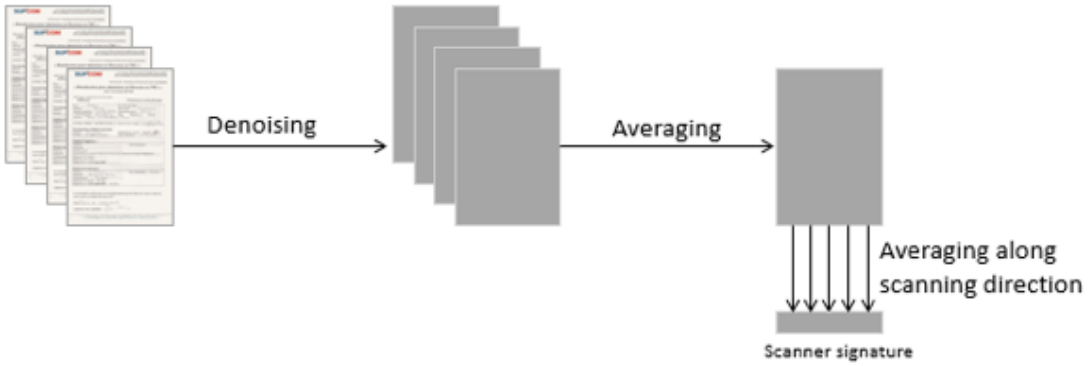


Figure 3.17: Illustration of a scanner fingerprint generation: Noise is estimated from many scanned filled forms

average noise image R_j by summing the noise images of each document and obtained accordingly the same noise extraction procedure as above. R_j is thus given by

$$R_j = \frac{1}{M}\sum_{i=1}^{M} R_i, \tag{3.14}$$

where M is the number of documents scanned by the j^{th} scanner. Then, S_j is obtained by averaging along the scanning direction

$$S_j(l) = \frac{1}{K}\sum_{k=1}^{K} R_j(k,l) \quad ,1 \leq l \leq L, \tag{3.15}$$

where K and L are respectively the number of rows and the number of columns of R_j.

As shown in Fig. 3.18, to identify the source scanner of a given document, the questioned document fingerprint is compared with the different scanners fingerprints using the Euclidean distance as a similarity measure defined by

$$d(V_i, S_j) = \sqrt{\sum_{x=1}^{L}(V_i(x) - S_j(x))^2} \quad ,1 \leq j \leq P. \tag{3.16}$$

where V_i is the questioned document fingerprint and S_j is the j^{th} scanner fingerprint. The final decision about the source scanner is performed by choosing the scanner, from the different devices in the database, having the shortest distance l with the query document

$$l = min(d(V_i, S_j)) \quad ,1 \leq j \leq P. \tag{3.17}$$

3.3.3 Experimental results

In the sections below, we discuss the performance of the proposed technique. Five scanners with various native resolutions as shown in Table 3.10 have been considered. Two of these scanners are the exact same model.

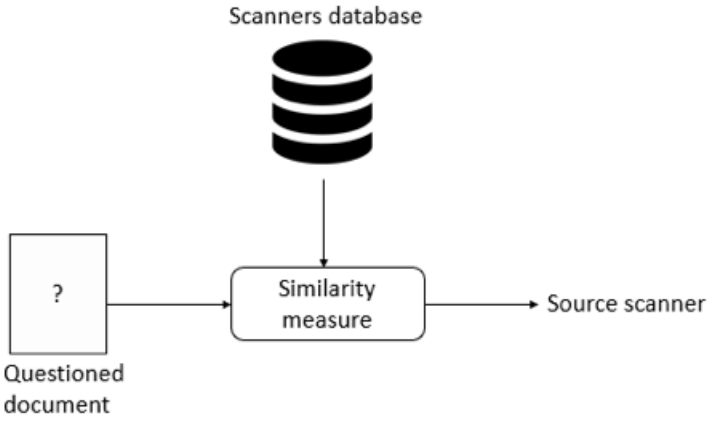


Figure 3.18: Scanner identification scheme

Table 3.10: Scanners used in experiments

Id	Model \ Brand	Resolution
S1	Canon Lide 220	4800x4800
S2	Epson Perfection V370 Photo (1)	4800x9600
S3	Epson Perfection V370 Photo (2)	4800x9600
S4	Epson Perfection V550 Photo	6400x9600
S5	HP Scanjet Pro 2500 F1	1200x1200

3.3.3.1 Creation of test images dataset

In the following experiments, the reference document corresponds to a computer generated form which contains: a logo, some text and frames as well as some other non-textual elements. It has been printed 90 times with the same printer and filled by hand by more than five persons. Some examples of the documents after scan are given in Fig. 3.19.
A dataset of 450 uncompressed images stored in TIFF format has been built by scanning all these filled forms with the five scanners at a resolution of 300dpi.

3.3.3.2 Identification performance

We analyse the performance in terms of classification accuracy by measuring the percentage of images from scanner model correctly identified. We follow a randomized multi-testing procedure in which various training and testing sets are selected to verify the consistency in reported results across experiments and, thus, limit any misleading impressions that might be given by a single experiment. We obtain an averaged accuracy of 100%. We also compare the proposed technique with the method proposed in [20] to guarantee a fair comparison since this latter works similarly to our proposed method in term of fingerprint generation and comparison while using different filtering approach. The confusion matrix of this method [20] is presented in Table 3.11. It can be seen that the proposed scheme outperforms the method in [20] for which confusion exists among scanners of the same model with an identification accuracy of 94,51%. To more investigate the use of the original document with previous methods, we implement the spectral method in [16] and the wavelet method that we have proposed in Section 3.2 by adding the noise extraction step introduced previously as a pre-processing step. Table 3.12 shows the

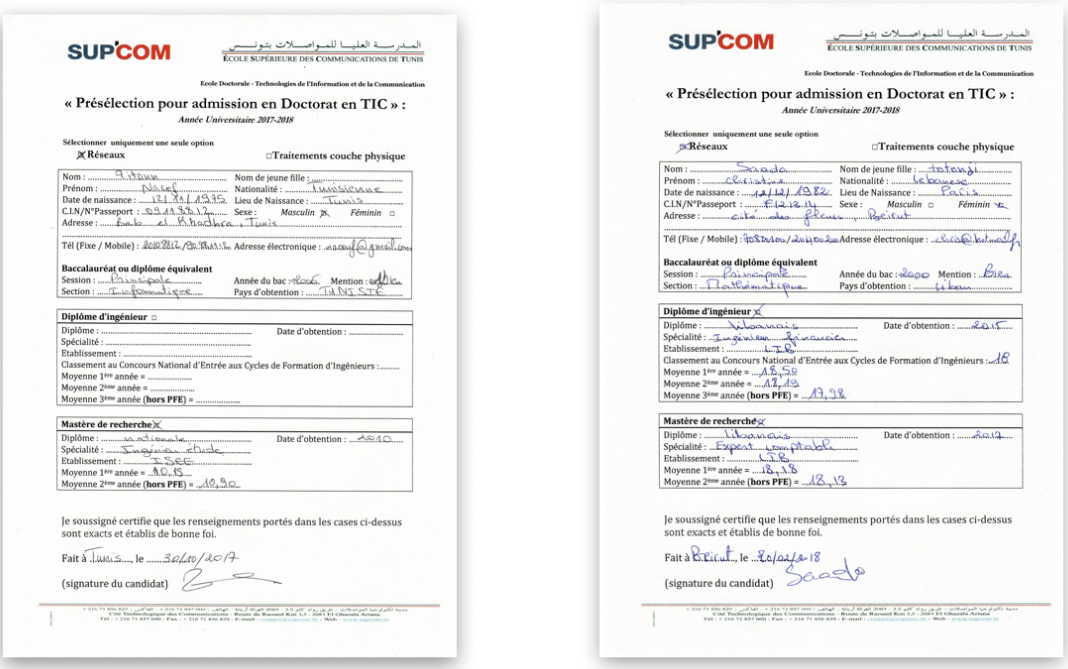


Figure 3.19: Samples of scanned filled form by different persons

Table 3.11: Confusion matrix for the method in [20] (in%) - TIFF images

	S1	S2	S3	S4	S5
S1	100	0	0	0	0
S2	0	73.33	0	14.07	12.59
S3	0	0	99.25	0.74	0
S4	0	0	0	100	0
S5	0	0	0	0	100

experimental result of performance accuracy. It can be seen that applying a wavelet decomposition to an already filtered image did not improve the identification of the source scanner. Note that we also reduced the number of features since we no longer consider the three color channel. Compared to the spectral method, the main benefit of our method is to achieve maximum accuracy with less computational steps.
To further demonstrate the robustness of our approach, we also experimented our approach considering source scanner identification after lossy JPEG compression. To do so, the original TIFF images previously generated were compressed with a quality factor equal to 75. The accuracy for classifying questioned documents is 100% for all scanners. This proves that our method is still able to capture scanning noise even when the image are post-compressed.

3.3.4 Discussion

In this section, we proposed an automatic source scanner identification method which takes advantage of the a priori knowledge of the nature of the scanned document. Such a priori knowledge is available in many domains, in particular in

Table 3.12: Overall identification accuracy for spectral methods [16] and wavelet method [14]

	Accuracy
Spectral method	100%
Wavelet method	42.44%

healthcare, banking, . . . , every activity sector where people have to fill administrative forms. By analyzing the noise resulting from the subtraction of the original document from the scanned document, a unique scanner fingerprint is built averaging signatures of documents scanned with it. Notwithstanding its simplicity, compared to methods proposed in the literature, the proposed filtering approach also brings competitive advantages. The proposed approach yields a very high accuracy even with compressed low-quality documents, which make it valuable tool in real life. It requires however to be further studied with much more scanners. In the future, we will investigate and extend the proposed method against other image post processing such as contrast stretching and sharpening.

3.4 Conclusion

Estimating scanners' noise models may be performed from single or multiple images. The very first work, presented in this chapter, allows identifying each scanner by a unique fingerprint from just one of the JPEG images. This fingerprinting is extracted from the JPEG file header where we have shown that the QTs contained in this header are unique for each scanner brand. To the best of our knowledge, the proposed approach is the only one that focuses on JPEG images to identify scanners models. It is however very simple and absolutely not robust to file format conversion and image re-compression. Also, it can not differentiate scanner form the same manufacturer.
However, for generating more robust digital evidence, we mainly focus on TIFF images and do not rely on fragile information such as data of the EXIF header. That's why, we developed two novel forensic methods. One [14] working in a fully blind way and another one [15] requires a prior knowledge about the original document template before it was printed. The first method defeats most recent and efficient approach proposed in literature when tested on our dataset containing genuine images of various content but cannot be very accurate when the image in question is acquired by a scanner having the same model as one of the known scanners in the dataset. Notwithstanding, its performance may be further improved by reducing the image content left in the HH subband. In the proposed semi-blind strategy [15], we solved the problem of scanners of the same model. Yet, the disadvantage of this method is that it is useful only if the original document is known. A hybrid framework combining these two approaches is likely to further push the performance of the forensic SSI. It is worth mentioning that proposed works in this chapter operate on the entire image. However, can we learn scanner features from small regions of the image? We will answer this question in the next chapter.

Our intelligence is what makes us human, and AI is an extension of that quality.

Notre intelligence est ce qui nous rend humains, et l'IA est une extension de cette qualité.

— Yann *LeCun*

4

SOURCE SCANNER IDENTIFICATION BASED ON AN AUTOMATIC FEATURES EXTRACTION

Contents

In the previous chapter, we addressed the problem of device identification by means of hand-crafted features. In order to perform a reliable fingerprint estimation, the traditional pipeline relies on the availability of a good number of images from each scanner which is unsuitable in scenarios where a small database is analyzed. Furthermore, the fingerprints comparison is based on distance measures that could be time consuming [160] and may become a major bootstrap when large device fingerprint databases need to be scanned. It is also important to notice that digital images may be processed with some information loss before investigation. They

can be cropped for instance. Thus, it is important to design an alternative scheme which resembles a realistic situation under which only a small part or patches of the questioned image is available.
In this chapter, still following the research line of designing a source scanner identification (SSI) model, we provide two data-driven approaches that automatically learn features from patches of digitized documents. It is organized as follows. In the first part, we give an overview of the deep learning-based methods that deal with source device identification. Then, in the second part, we present our approach based on a Two Dimensional Convolutional Neural Network (2D-CNN) and inspired from our previous work presented in 3. The third part describes another different framework using a One Dimensional Convolutional Neural Network (1D-CNN) and a Support Vector Machine (SVM). The first approach is working on non-overlapping squared patches while the second one takes linear inputs. For each one, we provide discussions and experimental comparisons with existing well-known neural networks as well as our previous method [14] proposed in 3.

4.1 Related work

Artificial Neural Networks (NN) are one of the trendiest research topics in 2021. They offer cutting-edge solutions to solve classification, prediction and detection problems in many fields [161]. They are a subset of Machine Learning (ML) composed of algorithms that train a system to learn pattern from data on its own and make predictions. In fact, NN simulate human decision-making to address real-world problems using ML approaches. Moreover, ML is a subset of Artificial intelligence (AI). AI refers to any computer program that performs a smart task. Notice that Deep Learning (DL) is a specific sub-field in NN that entails creating large and deep NN (generally more than three layers of neurons) to address specific problems. It may be costly and requires a large amount of data for training.
A CNN consists primarily of convolutional layers, activation functions, pooling layers and fully-connected layers stacked together to create the CNN architecture. Figure 4.1 shows a simple example of CNN. Such tools have an exceptional capability to learn accurate and convenient features representation automatically from image data (See Appendix C for further details). Recently, the use of CNNs has spread in multimedia forensics community and, in particular, for source device identification. CNN-based digital content forensic approaches can be broadly split into two categories: (i) methods that rely on stacked convolutional layers in the CNN to learn the camera patterns automatically from input images; and (ii) works relying on a pre-processing step to prevent the CNN from learning features related to the image's content and to learn only device specific features.

The first algorithm in the first category was proposed by Bondi *et al.* [162]. It utilizes a simple architecture with five layers working on 48×48 pixel patches which learns camera features directly from acquired pictures. A low accuracy of 29.8% is obtained at patch level when cameras of same models were considered and this accuracy got increased to 72.9% for model-level camera identification. Later, Huang *et al.* [163] introduced a similar architecture and improved the accuracy by using Batch Normalization and more convolutional layers. In [164], a six-layer CNN architecture was evaluated on three smartphone devices. Despite the high accuracy obtained, the results cannot be considered as reliable due to the small set of devices used. A deeper CNN architecture was proposed by Yao *et al.* [165] composed of 16 layers with 13 convolutional layers and three fully connected layers. The authors observed an average accuracy over 90%. Besides, they found that the proposed multi-classifier is not able to distinguish correctly between cameras of twinborn models.
In [166], Chen *et al.* discussed the use of a residual neural network (ResNet) composed of 26 layers. According to the reported results, ResNet performs better

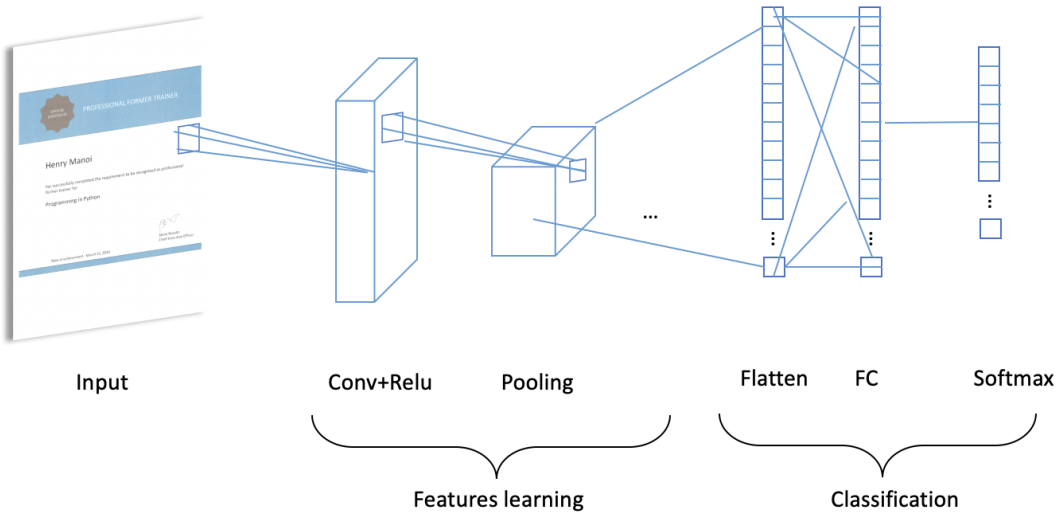


Figure 4.1: Example of a CNN. 'Conv' refers to convolutional layer, the rectified linear unit 'Relu' is an activation function, 'FC' denotes a fully-connected layer and 'Softmax' is an activation function that outputs a probability distribution

than AlexNet [167], GoogLeNet [168], and the architecture proposed by Bondi *et al.* [162]. Further works applied new architectures from computer vision such as DenseNet, XceptionNet and MobileNet [169–172].
Considering the second category, we can cite methods that use Gaussian filter residuals [173] or high-pass filter [174] prior to feeding images into CNNs. To improve the results obtained with these conventional fixed filters, Bayar and Stamm [175] have proposed a data-driven constrained convolutional layer as a preprocessing step. In [176], the authors used local binary patterns (LBP) to code the input images.
Recently, Rafi *et al.* [177] introduced a novel data-driven preprocessing block composed of several remnant blocks. In [178], strong edge components were reduced in each image patch before feeding it into the CNN.
Earlier studies [179] have reported that the camera recognition is more accurate for smooth and non-saturated images. Thus, selecting patches verifying these conditions, as a preprocessing step, can have a positive effect on the performance. Another selection criterion has been proposed in [180, 181], which seeks to find better textured pixel patches to train the CNN. Güera *et al.* [182] introduced a reliability map to estimate a value, for a given patch, indicating if it is containing valuable camera information. Interestingly, the approach in [183] has proposed to separate image patches into three subsets according to their mean and variance, and then, train each subset with its adapted CNN.
The analysis of those techniques shows the importance of the preprocessing to make the CNN invariant to image content and that using an automatic filtering is better than using manual one. However, even though the applied preprocessing operations really helped to reduce the influence of image contents, it may filter out useful information.

4.2 Scanner Source Identification using Wavelets and 2D-CNN

As previously exposed, CNNs are able to learn device features that could not be seen by hand-crafted approaches. The identification accuracy in most of the CNN-based source camera identification approaches presented in the previous section exceeds 90%. However, the training process of those networks is usually long and time-consuming. Moreover, the principal limitation of all above data-driven methods is that they mainly focus on the identification of the camera model instead of classifying devices of the same model. An additional uncontrolled factor is the complexity of the CNN architectures used in these approaches which, consequently, require a large amount of training images that may not be available in real situations. Nevertheless, due to inherent mechanical, processing and sensors differences between cameras and flatbed scanners, the vast majority of these approaches cannot be applied directly to scanners.
So far, very little attention has been paid to the role of neural networks to solve SSI problems. To the best of our knowledge, only Shao and Delp [184] proposed a SSI mechanism using CNN. They used an architecture inspired by the one proposed by Chollet in [185] and which relies on 14 layers, including 12 convolutional steps and two fully connected layers, working on 64×64 patches. As we will show it later, this mechanism does not perform efficiently once tested on an adapted dataset. This gives rise to investigating more effective CNN-based forensics scanner solutions. In addition, distinguishing scanners of same brand and model is another key issue which has not been solved yet.
In this section, we propose to take advantage of CNNs and propose a solution that identifies the scanner that has acquired a given scanned document in a blind way. It takes as input the document diagonal HH wavelet subband coefficients; coefficients that carry complementary information about the scanner noise. Our work is inspired by our methodology presented in section 3.2 which has shown advanced performance compared to the state of art. In fact, by feeding CNN with HH subbands coefficients, we expect: i) to be able to remove the scanned document content that is not relevant for SSI and, ii) that deep learning will be able to extract the noise mixture that is unique to one scanner. In this work, we opted for the stationary wavelet transform (SWT) [186] rather than for the traditional DWT due to its better performance for image denoising [187] due to the shift invariance property of SWT. In addition, SWT avoids coefficients decimation, a property of interest for small images.

4.2.1 Image and scanner fingerprints

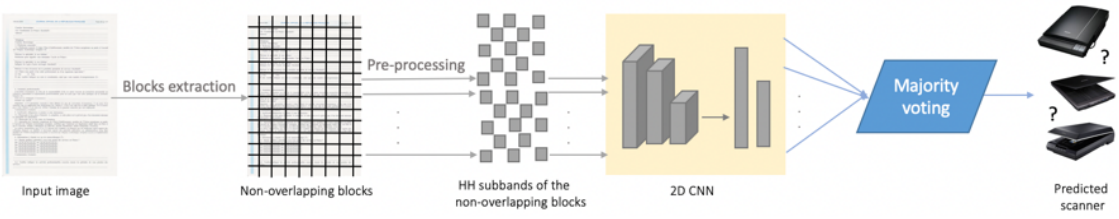


Figure 4.2: Global architecture of our system

The architecture of our system is depicted in Fig. 4.2. As it can be seen, it first cuts the image into non-overlapping blocks, that is to say into patches. This is an important step in order to obtain enough data for the training process, when the training image set is small, and to avoid memory saturation. Then, these blocks are wavelet transformed and their HH wavelet subbands are used as input of a CNN. The final decision about the source scanner relies on a majority voting considering CNN responses for all image blocks. We come back on the details and the purpose of each of these steps in the sequel.

4.2.1.1 Pre-processing: Wavelet decomposition

WT has been successfully applied in a wide variety of scientific fields. In our previous work [14], we explored the DWT transform in order to: i) suppress block content information so as to better extract the flatbed scanner mixture noise; ii) reduce the dimensions of the data to process. Whereas, in the current work, we rather use SWT, an extension of the traditional DWT, thus neglecting the downsampling step. In the following, the wavelet transforms are performed with the Symlet4 wavelet filter based on the results presented in [14].
For one decomposition level and for an image of NxM pixels, a dyadic SWT produces four subbands of coefficients of the same size NxM (LL, LH, HL, HH). As an input to our system, only the HH subband is exploited for the same reasons explained previously in Section 3.2. We will see in the experimental section that this pre-processing step is of importance in order to obtain valuable classification results.

4.2.1.2 CNN architecture

A key way to achieve high accuracy rate is to design a CNN that is adapted to the identification system desired. Figure 4.3 illustrates our CNN architecture.

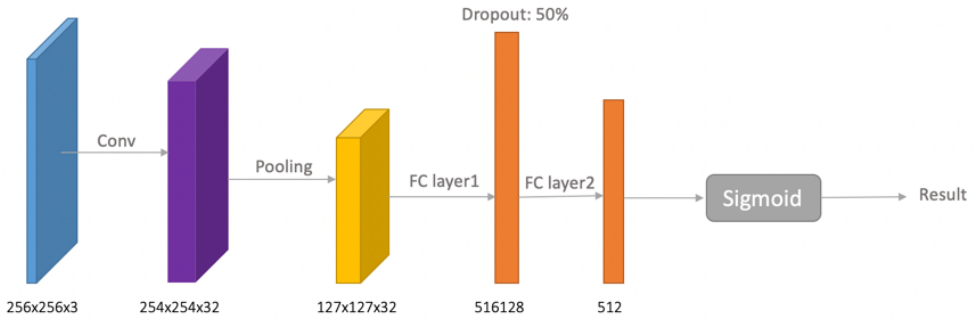


Figure 4.3: The network structure of the proposed CNN. The numbers below each colored figure is its dimension

First, to avoid increasing the number of weights/ parameters of the network, we proposed a shallow CNN with only one convolutional layer (Conv). A second main reason of not going deeper is to be able to train the network using considerably less training data. When a coefficient subband enters the network, it goes through this convolutional layer that convolves the input image with 32 kernels of size 3x3 where the kernel stride is by default set to 1. It is followed by the non-linear activation function ReLu to make our network sparse and help the training to quickly converge. The max-pooling operator of window size 2x2 is by next applied in order to reduce the spatial dimension of the input. These pieces of information are the inputs of two fully connected layers (FC layers) of 512 and R neurons, respectively, preceded by a dropout layer with a probability of 0.5 to prevent over-fitting [188]. The layer located at the very end of our network accompanied by the sigmoid activation function plays the role of a classifier which make the source prediction of the input image.
Notice that the R nodes represent the likelihood of the image to be acquired by each scanner, where R is the number of scanners. Table 4.1 sums up the hyper-parameters of our network.

Table 4.1: Structure of the proposed CNN

Id	Layer \ Name	Size
1	Convolution-ReLu	32 filters of size 3x3
2	Max-Pooling	2x2
3	Dropout 50%	-
4	FullConnected-ReLu	512
5	FullConnected-Sigmoid	N

4.2.2 Image origin predictor

In the previous section, we proposed a network that predicts the source of blocks rather than the entire image. Thus, after having classified all the blocks of an image under observation, majority voting is adopted in order to take the final decision about the specific scanner that has acquired the full image. It is calculated as follows

$$\mathrm{M}(Q) = \mathrm{k} \text{ if } \; occ(\mathrm{k}) = \max_{j=1.n}(occ(j)), \quad 1 \leq \mathrm{k} \leq \mathrm{N} \tag{4.1}$$

where: Q is the questioned image and occ is the occurrence number of a class j defined by:

$$occ(j) = \sum_{i=1}^{M} (P_i = j) \tag{4.2}$$

where P_i is the predicted class (i.e. the source scanner) of the ith block of Q and M is the total number of blocks. In other terms, the scanner the mostly predicted is identified as the source scanner of Q.

4.2.3 Experimental results

To validate the proposed SSI method, several network parameters are discussed and compared. Experiments were performed using a dataset of images of different content acquired with 8 commonly-used flatbed scanners (see Table 4.2). The image dataset corresponds to 54 documents of different types (forms, certificates, contracts, records...) of various content (text, figures, stamps...), which were printed on A4 paper with the same printer and then scanned at the same resolution (300dpi) with each scanner. Scanned documents are stored in the TIFF format.
Performance measurements are mostly based on classification accuracy, which is the ratio of correct predictions to total predictions made, and confusion matrices. In the confusion matrix, each row refers to the predicted label for each scanner instance while each column corresponds to the actual class (scanner). Each of the following schemes has been implemented in Python using Keras [189] deep learning library on a ZOTAC GeForce RTX 2070 SUPER AMP EXTREME GPU with 32 GB of RAM. The RMSprop optimizer [190] was used. The learning rate was set to 10^{-6} and the training batch size to 32 images. The final training period consists of 49 epochs which provides the smallest loss on validation blocks.
For each CNN model, we randomly selected 40% of the images for the training, 20% for the validation and 40% for the testing. During training, K-fold cross-validation technique [191, 192] is applied to guarantee generalization that is to say a low dispersion of the accuracies across the folds. In this study, K is set to 5.

Table 4.2: Scanners used to generate the database

Scanner Id	Brand	Model	Sensor type	Native resolution
S1	Canon	Lide 120	CIS	2400 x 4800
S2	Canon	Lide 220	CIS	4800 x 4800
S3	Canon	CanonScan 9000F	CCD	4800 x 4800
S4	Epson	Perfection V39	CIS	4800 x 4800
S5	Epson	Perfection V370 -1	CCD	4800 x 9600
S6	Epson	Perfection V370 -2	CCD	4800 x 9600
S7	Epson	Perfection V550	CCD	6400 x 9600
S8	HP	Scanjet Pro 2500 F1	CIS	1200 x 1200

4.2.3.1 Classification results for the proposed scheme

To train our network, all images from our dataset were split into non-overlapping blocks of size 256x256 pixels. These sub-images are annotated with their corresponding scanner candidate as label. Thus, in total, we have approximately 48000 scanned sub-images with different varieties of image details.
As stated in section 3.2, in order to make our system less sensitive to the image content and that it only learns scanner fingerprints, SWT is applied on these samples giving access to HH subbands next used as CNN input.
To test the effectiveness of the proposed neural network in distinguishing scanners, we have carried out a series of experiments. The main purpose of the first experiment is to evaluate the performance of our method when it works on single image blocks, that is to say taking a decision about the source scanner of a document from one block, only. A decent testing accuracy of 99.31% is obtained. We repeat the same experiments using the DWT instead of the SWT and a decrease in the classification accuracy by 2.31% is observed.
Then, after demonstrating the good performance of our scheme only on single block, we evaluated the entire pipeline. More clearly, we apply a majority voting on the decisions obtained from the CNN on all image blocks as presented in Fig. 4.4. The confusion matrix presented in Table 4.3 for classifying full images has an average classification accuracy of 100% using the same dataset.

To further evaluate the reliability of our scheme, we propose to use another image test set. 90 new documents were scanned using each scanner from the lsit in Table 4.2, leading thus to 720 new images next classified with our network. Again, we obtained 100% of accuracy. Thus, our system is proven to be reliable whatever the content of the processed images.
The following experiment investigated the impact of the number of convolutional layers onto the performance of our scheme. Figure 4.5 shows that the dynamics of the model with 1, 2 and 3 layers are pretty similar. But, the model converges more rapidly with only one convolutional layer, learning thus the problem more quickly.

Next, we provide experimental evaluation to prove the efficiency of feeding CNN with the HH subband of each block rather than with the block pixel, directly. As it can be seen in Fig. 4.6, better performance is achieved compared to the classification without this pre-processing step. This demonstrates the important role of this step. This result shows that, for SSI, it is more appropriate to extract features in the transformed domain than in the spatial domain. This explains the interest of considering the HH subband which isolates the high frequencies in the images, and in particular, the noise related to the scanners.

It is interesting to note that scanners embed vertical and horizontal artifacts that can be found in the HL and LH sub-bands. So, an alternative is to use these

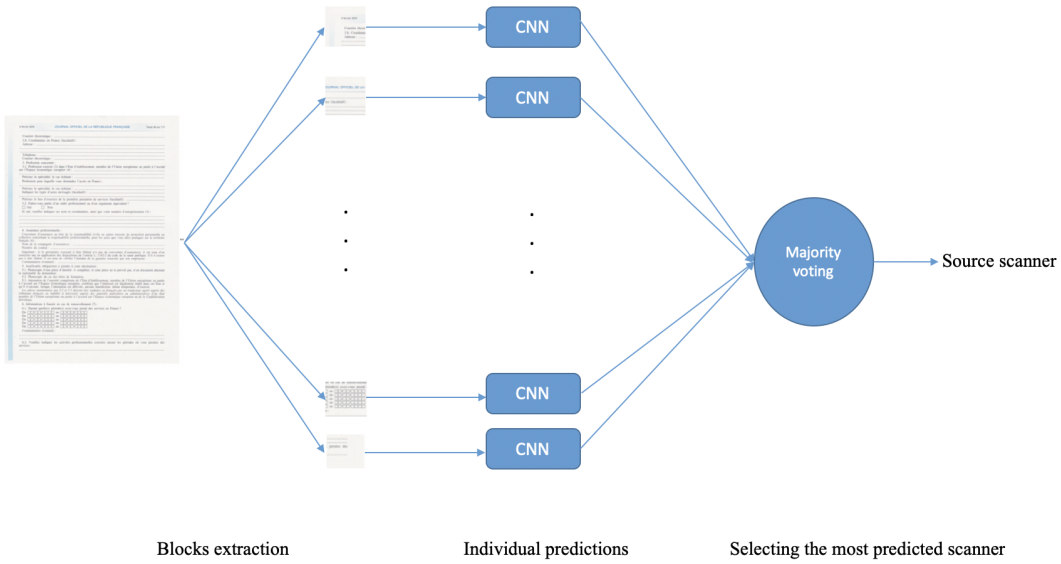
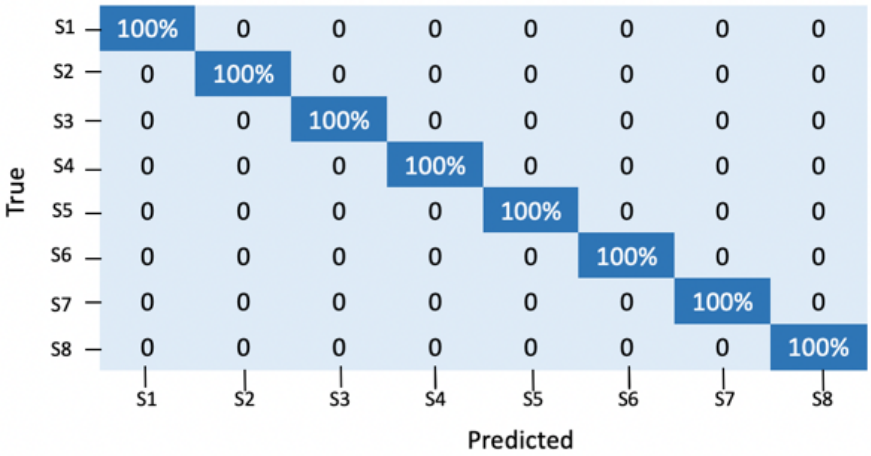


Figure 4.4: Pipeline of an image source identification based on majority voting

Table 4.3: Confusion matrix for full images using proposed method



subbands for SSI. Comparison performance are given in Fig. 4.7. The HH sub-band provides better accuracy results. The LL sub-band is also considered to confirm its non-adaptation to solve SSI problems. Notice also that the LH subband is also suitable for SSI which is related to the linearity of the scanner sensor.

Another important parameter that is crucial when analyzing the performance of the proposed CNN is the choice of the color channels. Figure 4.8 shows that a preserving all color channels dramatically improves the forecasting accuracy.

The block size is one of the most important parameter of our proposal with an impact on its classification accuracy. Thus, more experiments were carried out in order to show the effect of the block size while considering both DWT and SWT. Two different block sizes were considered. From Fig. 4.9, it appears that high classification accuracy is achieved when training the CNN with just 1000 blocks from each scanner whatever the block size. It can be also observed that using the DWT decreases the forecasting accuracy significantly for smaller

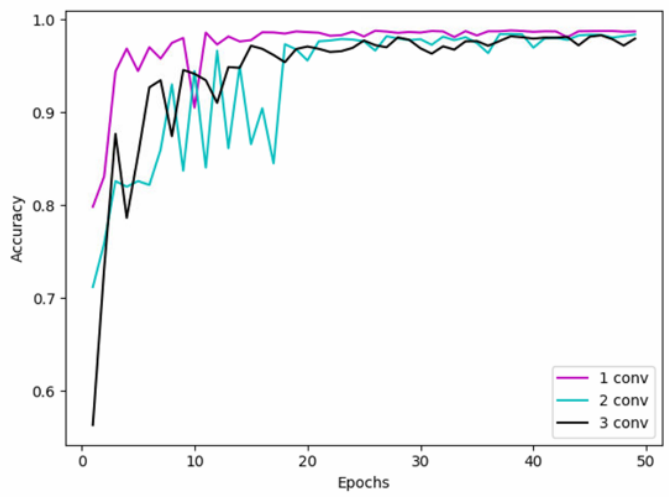


Figure 4.5: Effect of adding convolutional layers to the network on the validation accuracy

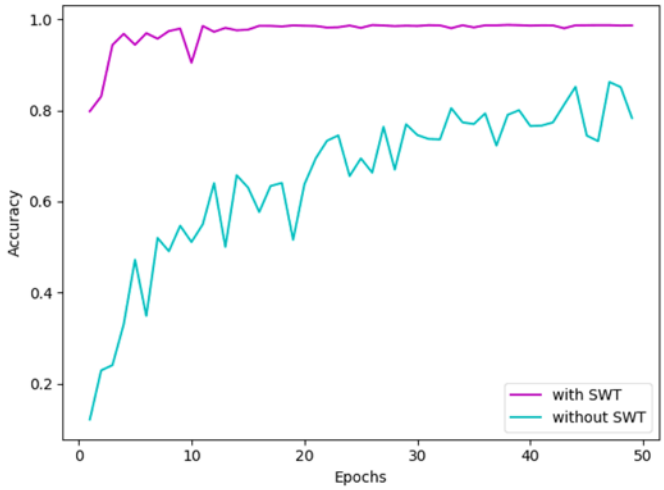


Figure 4.6: Effect of adding convolutional layers to the network

blocks. These results confirm that the SWT is more suitable than DWT for SSI based on neural networks due to its up-scaling property which likely preserves more information about the scanning noise.

4.2.3.2 Comparison assessments

A comparison of the current method with respect to all the SSI methods in literature is well beyond the scope of this work since most of them require specific testing image types and/or additional requirements. Therefore, to assess the superiority of our proposal, we propose to compare it with the KLD-based method we proposed in Chapter 3; a method that has demonstrated better behavior than other recent approaches. Let us recall that KLD-based method extracts hand-designed features in the wavelet domain. We have also implemented the CNN based method of Shao and Delp [184]. Table 4.4 shows the various methods and their respective accuracies. As it can be seen, our method outperforms the state-of-art methods by obtaining an overall accuracy of 100%. We can also notice that Shao and Delp method [184]

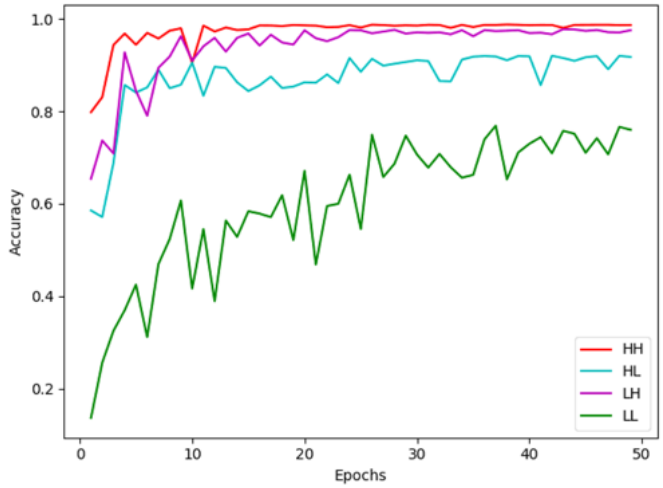


Figure 4.7: Effect of the subband choice on the classification accuracy

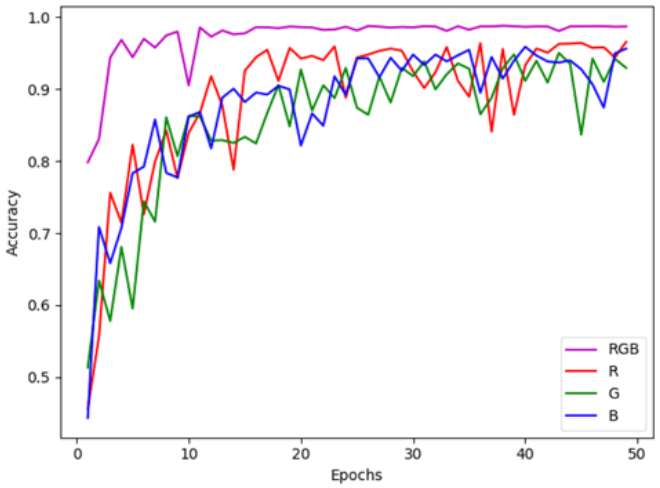


Figure 4.8: Effect of color channels on the performance accuracy

failed to correctly identify most of the scanners. This can be explained by the absence of a pre-processing denoising step which, as shown previously, is necessary to remove the image content. Note that the classification accuracy degrades also in [14] along with the increase of the number of scanners to discriminate.

To compare with common CNN, we trained AlexNet [168] and GoogLeNet [**c44432**] on our image data set, using the same pre-processing strategy. We obtained an accuracy of 94.69% and 90.64% at block level and 97.5% and 96.66% at full image level for AlexNet and GoogLeNet, respectively, as reported in Table 4.4. We further compare the time complexity of each network. Based on the results shown in Figure 4.10, AlexNet requires the shortest training time but performs the worst in term of classification accuracy. We reported 5227.77s, 4026.55s, and 36607.06s as average training time for our CNN, Alexnet and GoogLeNet respectively. Compared to the performance of our scheme, one can conclude that much better performance are achieved with a small, less time-consuming and compact CNN configuration of only one convolutional layer.

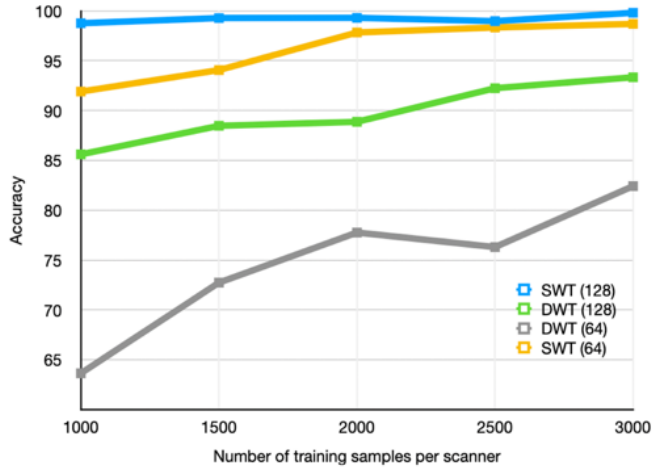


Figure 4.9: Effect of training size on testing accuracy for multiple block sizes

Table 4.4: Average classification accuracy for non-overlapping blocks and full images for different model architectures

	Accuracy	
	Block level	**Image level**
AlexNet	94.69%	97.5%
GoogLeNet	90.64%	96.66%
Proposed-DWT	97%	100%
Proposed-SWT	99.31%	100%

4.2.4 Discussion

In this section, we proposed a novel approach that exploits the scanner noise located in the wavelet HH subband while simultaneously benefiting of the capacity of CNN to automatically extract useful features from this subband.
We evaluated the advantage of using the SWT and the HH subband as input to the network. More experiments have been conducted to assess the performance gains achieved by limiting the number of convolutional layers to only one layer and, on the other side, keep using all the color channels of the images.
Classification results also demonstrate that our CNN offers superior performance than recent scanner identification techniques even under the condition of limited training samples. Another result of our scheme is that it is able to identify the source scanner even from small blocks of the image. This is a promising prospect for forensics applications particularly when only a part of the investigated image is available such as forgery detection.

4.3 Scanner Source Identification using SVM and 1D-CNN

The existing CNN-based solutions for camera source identification work by analyzing square image patches. The geometry of the patches is suitable for such task since

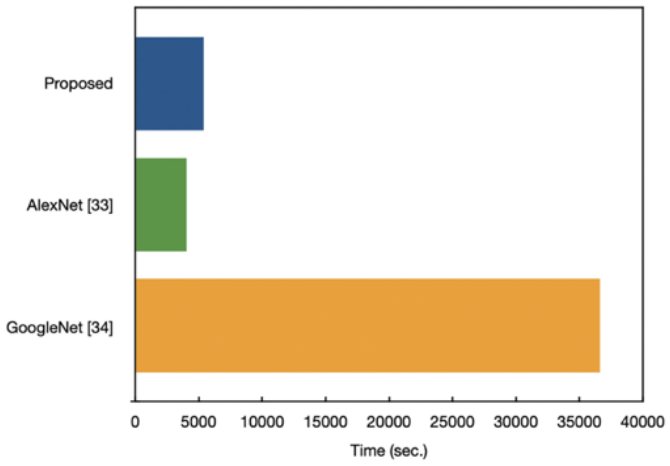


Figure 4.10: Average training time for different CNNs

it is coherent with the geometry of the camera sensors array. This interpretation, however, contrasts with the one related to scanners due to the linear geometry of its sensors. Therefore, it is better to take into consideration the linearity of the scanning noise to build more sophisticated and reliable architecture. Moreover, there is a need to meet real world conditions with a scheme capable to work whatever the digitized document content. More clearly, the effectiveness of existing methods is limited by their capability to suppress most of the image content and to only extract the scanning noise.
In recent years, sequence classification using 1D neural networks, also called sequence learning, has demonstrated groundbreaking performance in the field of machine learning tasks such as speech emotion recognition [193, 194], accent classification [195], and text classification [196].
In this section, we introduce a new sensor fingerprint-based approach for SSI. Its originality is twofold. At first, as opposed to actual approaches, it focuses on the 1D sensor characteristics of flatbed scanners with as objective to automatically extract optimal discriminative features that are next classified by an SVM (See Appendix D). By doing so, the proposed hybrid 1D-CNN-SVM approach is able to learn only from a small part of each scanned pixel lines making our solution of small complexity. In second, our scheme can discriminate scanners of same model, even they are more than one. Experiments conducted on a large set of documents and scanners demonstrate that our 1D-CNN-SVM approach is capable to provide efficient training with limited size of training data and that it surpasses the most recent state-of-the-art method. They also show the interest to take into account the 1D characteristic of scanner rather than working with a 2D-CNN model.
Our new data-driven approach works on segments extracted from each row of scanned documents. The features extracted from each of those segments are later fitted into an SVM for classification in order to achieve better performance than with a CNN model alone. The final judgment about the source scanner relies on decisions taken over all segments of an image by means of a majority voting mechanism.

4.3.1 Image and scanner fingerprints

The solution we propose focus on three critical issues in SSI: (i) discrimination or identification of scanners of same brand and model (ii) availability of only a few documents for the model training task and (iii) testing made on a portion of the document. The general structure of our framework is illustrated in Fig. 4.11. As it

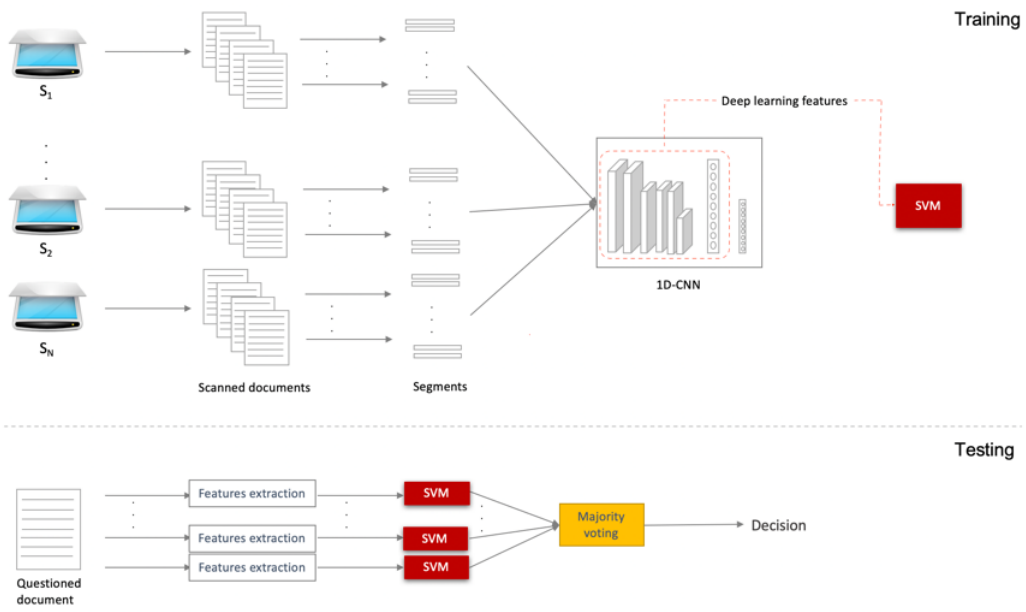


Figure 4.11: Main architecture of the proposed framework

can be seen it is constituted of two main tasks: feature extraction with the help of a 1D-CNN model followed by a SVM as a source scanner predictor.

This work addresses the scanner's sensor noise corresponding to the line-by-line scanning pattern. To meet up the one-dimensional characteristic of this noise pattern, we designed a 1D-CNN to determine the source scanner used to capture a scanned image by considering fixed-length segments taken from each of its rows as input to the CNN. Each row segment position is chosen randomly but fixedly for training and testing the proposed system. The measurement results are more or less affected by the selected position as the sensors noise is expected to be repeated over all the rows of a scanned document, as explained in Chapter 2. These multi-channel segments are then fed into a 1D-CNN.
The final model has been decided based on numerous experiments to verify the best balance between network complexity and system performance. Since the number of images used for training is limited, it is not recommended to use a more in-depth architecture [197]. Moreover, the number of filters and their corresponding kernels is chosen to provide a general presentation of the scanning noise. As shown in Fig. 4.12, the 1D-CNN is composed of 8 layers structured as follows:

- The multi-channel segment is generated from each line of the digitized document and fed as input to the first convolutional layer producing 100 feature maps as output. The purpose of this processing is to extract enough statistical properties and different types of dependencies among pixels in the noise inserted when digitizing an image.
- A second convolutional layer with another 100 filters with kernel size 10 is applied to the output of the first layer.
- The resulting feature maps are aggregated with a Max-pooling layer of kernel size 3.
- A third and fourth convolutional layer with 160 filters of size 10 are then applied.

- The number of parameters is reduced by a global average pooling layer before using a dropout layer with a probability of 0,5. The key idea of this later is to generate more robust features by randomly dropping different subsets during training.
- A last dense layer with 9 output neurons using a softmax function played the role of a classifier.

ReLu, the most used activation function in CNNs, was adopted as the activation function in all the convolutional layers to help learning complex functional mappings from data by thresholding values at 0, i.e. activating all output nodes larger than zero and suppressing output nodes smaller than zero. It is chosen to improve the non-linear problem-solving ability of our network.

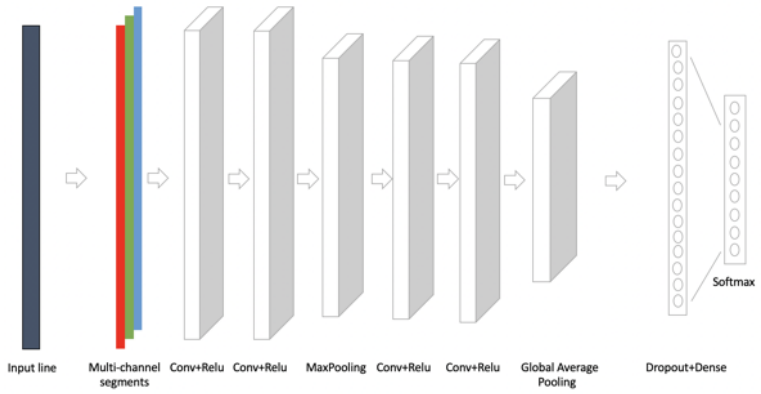


Figure 4.12: The architecture of the 1D-CNN network

The architecture of the proposed 1D-CNN was adapted according to the parameters described in Table 4.5.

Table 4.5: Configuration of the proposed 1D CNN

Layer	Type	Output shape	Kernel size
1	Convolution	119x100	10
2	Convolution	110x100	10
3	Max Pooling	36x100	3
4	Convolution	27x160	10
5	Convolution	18x160	10
6	Global average pooling	160	-
7	Dropout	160	-
8	Dense	9	-

4.3.2 Full image source scanner identification

Although Softmax performs very well in terms of classification, it has been proved recently that the SVM classifier enhances the classification accuracy [198]. In the present study, the softmax layer was replaced by an SVM classifier. Therefore, the preceding layer outputs (Layer 7) are treated as features to train the SVM. Once

trained, it performs a source identification task of segments from the testing images with those extracted features.
Since the classification is performed on segments extracted from rows of the questioned document, during testing, SVM predictions need to be aggregated to decide which scanner has acquired that document, as illustrated in the bottom of Fig. 4.11. Therefore, the final decision about the source of the questioned document is taken according to a majority voting rule.
Let $c_{i,j}$ denotes a binary metric indicating if the i^{th} segment is predicted as acquired by the j^{th} scanner:

$$c_{i,j} = \begin{cases} 1, & \text{if j is the predicted class (scanner) of the } i^{\text{th}} \text{ segment} \\ 0, & \text{otherwise} \end{cases} \tag{4.3}$$

where $i = 1,..,\ M$, $j = 1,..,\ R$, M is the number of image rows (number of segments in one image), and R is the number of scanners. For each scanner, the number of segments identified as acquired by the scanner j is denoted by L_j and defined as

$$L_j = \sum_{i=1}^{M} c_{i,j} \tag{4.4}$$

The final decision about the source scanner is chosen such that it has the highest number of segment predicted as belonging to it, thus, verifying:

$$L_j = \max_{k=1..N}(L_k) \tag{4.5}$$

4.3.3 Experimental results

Our system was implemented in Python using Keras library from the tensorflow [199] framework and executed on a computer equipped with ZOTAC GeForce RTX 2070 SUPER AMP EXTREME GPU and 32 GB of memory. A batch size of 8 samples was used for training the 1D-CNN which was trained up to 35 epochs with early stopping equal to 10. The batch size is chosen so as to reduce loss fluctuation. For optimization and weights updating in the backpropagation training stage, we used Adam [200] optimizer with a small learning rate of 0.0001 to prevent falling into local minima.
The length of the segments to be extracted from image rows is fixed to 128 and the start position within columns is 1000.
A sample on how segments are extracted from each image is shown in Fig. 4.13. A block of size Mx128 is first cropped at the given column position and then segmented horizontally. For SVM, the Python Scikit-learn library [201] was used for training and classification.

4.3.3.1 Dataset

The dataset used to train and evaluate the proposed approach has been generated by scanning 90 documents at the resolution of 300dpi with a collection of 9 scanners of different brand and model given in Table 4.6.

These documents are of different types in order to simulate real-world situations (forms, hand-written documents, certificates, medical records, reports...) leveraging various content. They were then printed at 600dpi on similar quality A4 paper using the same printer. This is crucial to get rid of the variability due to the variations in paper quality or printer. At least, we obtained a dataset composed of 810 document images saved in TIFF format. Fig. 4.14 shows several samples of the cropped blocks from which the segments were extracted.

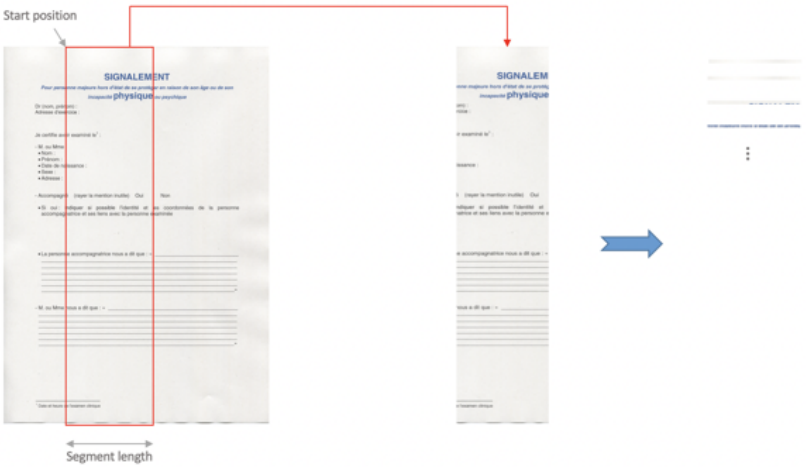


Figure 4.13: Segment extraction example

Figure 4.14: Samples of the cropped blocks from several images

4.3.3.2 Evaluation

All reported experiments used a 3-fold cross-validation procedure. In all cases, 20 documents were taken at random for training and validation and the remaining documents for testing.

Evaluation of the features extraction network
In order to evaluate the proposed architecture, we observe through various experiments the proposed 1D-CNN architecture by observing its ability to extract high discriminative features for limited training data. We proceeded by first considering different segment lengths (128, 256 and 512) to assess the impact of the input length on the classification performance.
Figure 4.15 shows that a segment of length equal to 128 pixels is the best choice since it provides the highest training accuracy even though working with other segment sizes performs well. This result is of great interest as it indicates that our system shows obvious advantages when dealing with small size images and with processor of limited capacity.

Since the digitized documents considered in these experiments are full color images, we preferred to test the effectiveness of each color channel separately. It can be seen in Fig. 4.16 that jointly handling color channels leads to higher performance.

Table 4.6: Scanners used to generate the database

Scanner Id	Brand	Model	Sensor type	Native resolution
S1	Canon	Lide 120	CIS	2400 x 4800
S2	Canon	Lide 220	CIS	4800 x 4800
S3	Canon	CanonScan 9000F -1	CCD	4800 x 4800
S4	Canon	CanonScan 9000F -2	CCD	4800 x 4800
S5	Epson	Perfection V39	CIS	4800 x 4800
S6	Epson	Perfection V370 -1	CCD	4800 x 9600
S7	Epson	Perfection V370 -2	CCD	4800 x 9600
S8	Epson	Perfection V550	CCD	6400 x 9600
S9	HP	Scanjet Pro 2500 F1	CIS	1200 x 1200

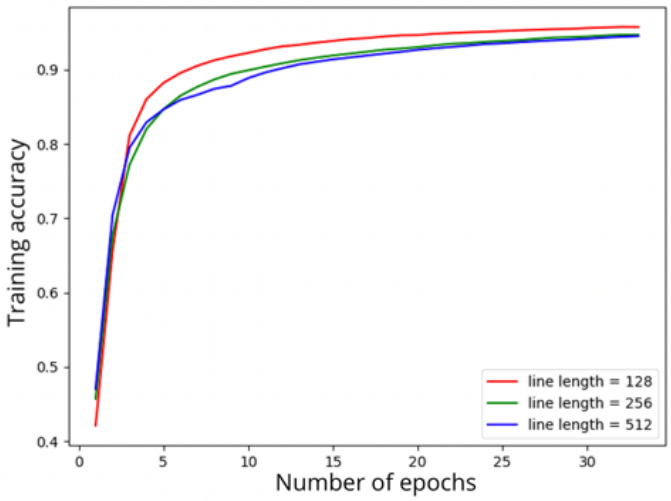


Figure 4.15: Effect of the segment length on the training accuracy

This confirms what has been reported that combining the three color channels tends to produce superior performance compared to using with each separate color channel independently due to the possible inter-channel correlation [202]. For this reason, we decided to exploit the three color channels jointly.

In order to have a better insight of the behavior of the proposed 1D- CNN during training and validation, learning curves were computed and provided in Fig. 4.17. The training curves show that our model is learning well the scanners fingerprints. On the other hand, the model generalization capacity can be confirmed by observing the validation learning curves. No over-fitting or under-fitting have been reported. To the greatest extent, fast convergence rates were observed as the training and validation loss converged in around 35 epochs. When converging, validation accuracy and loss curves slightly oscillate indicating the robustness of our network in reducing sensitivity to documents' content.

1D-CNN vs 1D-CNN-SVM

To verify the validity of the improved 1D-CNN-SVM model, we performed the following experiments: the first one aims at evaluating the source scanner prediction performance of the 1D-CNN-based classifier for separate segments as well as on

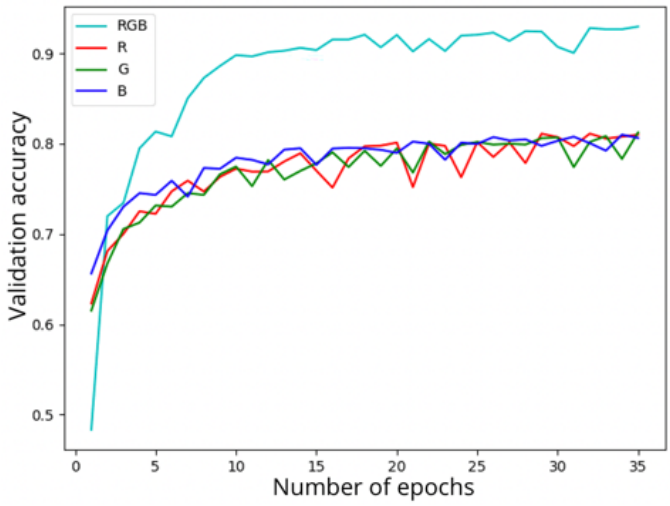


Figure 4.16: Performance comparison results for using Red (R), Green (G) and Blue (B) color channels separately and combined

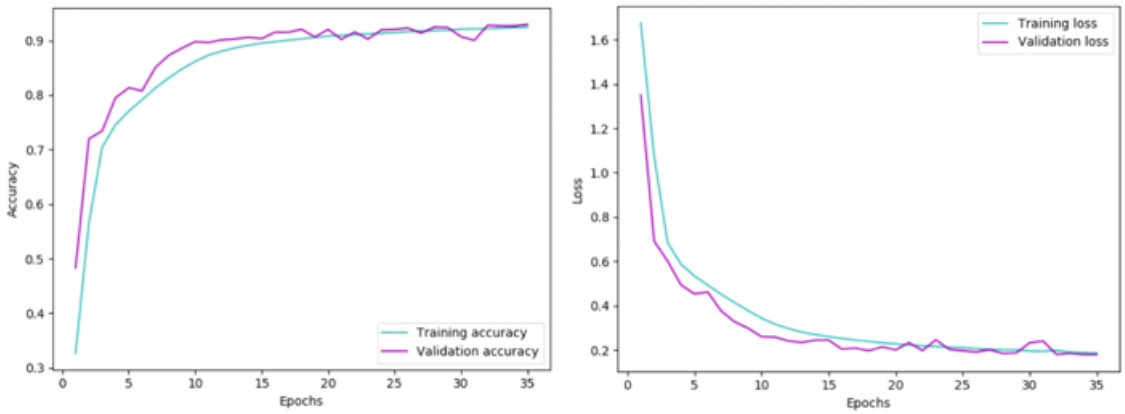


Figure 4.17: Plots showing learning curves of model loss and accuracy over each training epoch during training and validation (a) Training loss vs training accuracy (b) Validation loss vs validation accuracy

full images (i.e. with majority voting applied on segments of each testing image); the second experiment is designed to validate the performance of the SVM-based classifier. Firstly, experiments performed on segments reached almost 90% of accuracy when the softmax layer is adopted for classification. On the other hand, a total identification accuracy of 95.65% is obtained when testing on entire images. The final results are illustrated in Table 4.7. One key reason behind the high accuracy of the proposed SSI based on the 1D-CNN scheme is that this neural network is able to perform adaptive extraction of the features that characterized the sensor noise.

Next, we verified the performance of the proposed ensemble 1D-CNN-SVM model. We reported 93.13% and 98.14% at segment level and full image level respectively. Table 4.8 illustrates the confusion matrix related to this model.
It can be observed that, for both architectures, a small confusion arises between S3 and S4 which are two instances of the same model CanonScan 9000F due to the high similarity between their settings which is not the case for the Epson Perfection V370 instances. These results demonstrate the capability of the 1D-CNN-SVM model to better discriminate between scanners of the same brand and model even when they are difficult to differentiate.

Table 4.7: Confusion matrix (in%) using the 1D-CNN model over 9 scanners

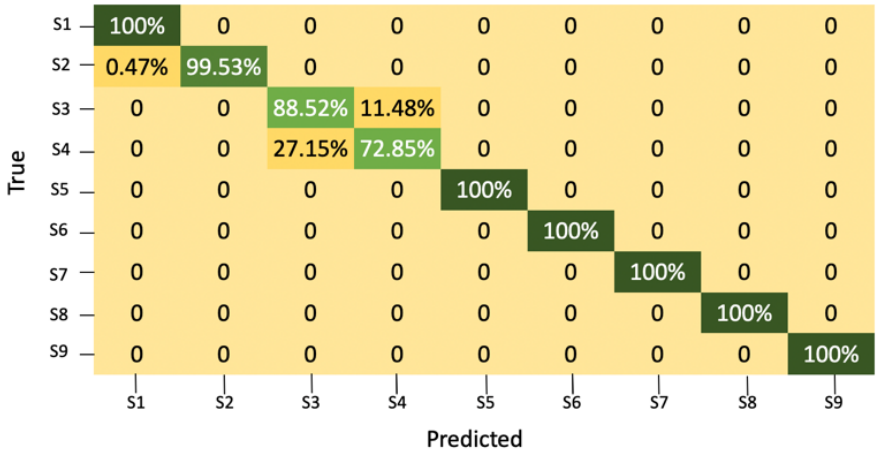
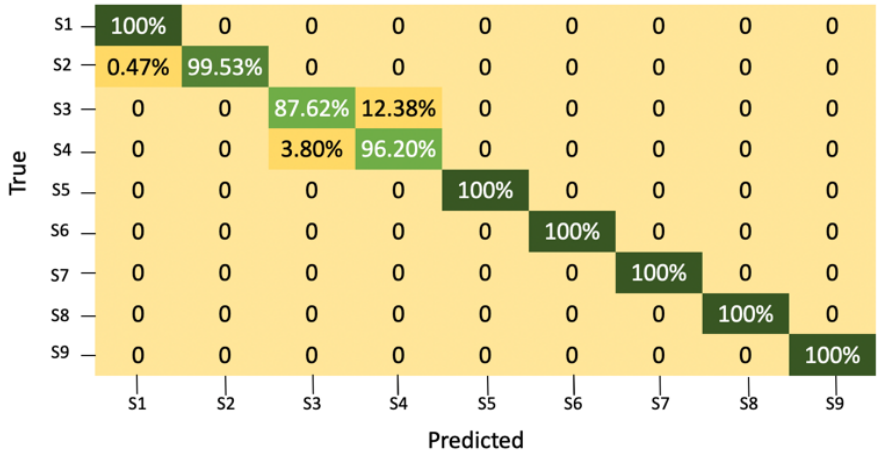


Table 4.8: Confusion matrix (in%) using the 1D-CNN-SVM model over 9 scanners



Comparison with previous work

The performance of the proposed deep learning based method is, first, compared with the most recent method (DWT-KLD) that we have presented in Section 3.2 for different identification levels. As shown in Table 4.9, the deep learning features based on 1D-CNN outperforms handcrafted features based method and achieved the best accuracy independently of the number and the models of scanners. Much of the instability in the DWT-KLD method stems for the increase in the number of scanners, the availability of scanners of same model, and the use of a small training dataset to generate the scanners' fingerprints.

In order to show the interest of working with 1D-CNN, we decided to investigate the performance of the 2D-CNN-based model proposed in the previous section. To allow fair comparison of results, we fixed the size of the input patch to 128x128. We also attempted to keep the same training parameters across experiments. The outcomes of this experimental analysis are shown in Table 4.9. A good accuracy that approaches the results obtained by the 1D-CNNs was achieved when only the scanners of different models (7 scanners) are considered. However, it appears that the 1D-CNN-SVM architecture leads to a significant increase of performance when more scanners of the same model are considered.

It is also worth noting that the 1D-CNN used in this work is more suitable to exploit the geometry of the scanner sensors array.

Table 4.9: Comparison of classification accuracies between the proposed 1D-CNN based method with and without the SVM classifier, and existing methods [14, 18]

Method	Brand	Model	Device
DWT-KLD [14]	96.82%	81.22%	66.24%
SWT-2DCNN [18]	100%	100%	97.40%
1D-CNN	100%	99.24%	95.65%
1D-CNN-SVM	100%	99.93%	98.15%

4.3.4 Discussion

In this part, we proposed a new SSI approach applying a hybrid 1D-CNN- SVM model on digitized document lines. We have shown that features extracted by the neural network fits scanners fingerprints better than manually designed features using limited training data (only 20 documents per scanner). Assessments have demonstrated the effectiveness of the 1D-CNN to greatly suppress the content of documents and to preserve most of the sensors noise. Our network extracts more significant scanner features with inherent independence from the image content especially for scanners of different models. Moreover, replacing the softmax layer by SVM is proved to be beneficial for the generalization ability of the SSI system, overcoming the fact that CNN linear separability is not complete.

4.4 Conclusion

Long ago, the task of SSI has been dealt with many techniques based on statistical analysis and patterns comparison. However, recent advances in the capabilities of computer chips have prompted a renewed interest in ML, and in particular NN-based techniques. Thus, in this chapter, we proposed to automate the SSI process by implementing two CNN models using the one dimensional and the two dimensional architectures. We jointly search for the best features extractor-classifier pair that boosts the accuracy of the scanner identification. We prove the efficacy of the proposed models by testing on a larger number of scanners compared to the hand-designed features based methods presented in the previous chapter. Moreover, we have shown that they outperform state-of-art methods. The main benefit of these approaches is that a small number of images is required to learn a good discriminative CNN.
Nevertheless, current SSI assumes that the investigated document was acquired by one of the known scanners. Because forensic investigators frequently work in open circumstances, this is a serious constraint. The fact that new scanner models are introduced every year and that many of them are difficult to obtain makes the problem more and more complicated. Moreover, a typical investigator may be only interested to know whether or not images are acquired by the same device. That is why we propose in the next chapter to design a new forensic technique that brings a solution to this problem using only fingerprints extracted from each scanned image without any knowledge about their source scanner.

We're entering a new world in which data may be more important than software.

Nous entrons dans un nouveau monde dans lequel les données peuvent être plus importantes que les logiciels.

— Tim *O'Reilly*

5
SOURCE SCANNER LINKING

Contents

Current source scanner identification (SSI) methods can accurately identify the source device that has captured an image between a range of known scanner models. In another word, the source to identify is predefined in the database of scanners used to train these methods. This is known as a closed-set classification. It is thus natural to think about what may happen if the instigated digitized image has not been acquired by one of these known models? In this real-world situation, existing methods will associate the image to one of these scanners leading to a misclassification. As a consequence, the problem of SSI must be treated as an open-set detection problem.

On the other hand, one may just need to know if two images have been acquired by the same scanner model or not. This problem is referred as device linking (DLK). In this chapter, we aim at solving both problems by developing a novel approach with as objective to verify the similarity between two images sources without having physical access to their source acquisition devices. Consequently, this information about source similarity could be exploited to check if a suspected image was acquired by one of the few known scanners. In realistic situations, it is not possible to have full access to all source scanners and to build a unified database of scanner footprints. Therefore, it is important to know if an image is coming from a device that hasn't been used during training the SSI system with the goal to decrease false matches.

In the following, we address the problem of DLK for flatbed scanners by designing a hybrid 2D-1D-CNN architecture that, given two images of unknown sources, can decide whether they were captured by the same scanner or not. This solution is the first step towards challenging open-set SSI. One advantage of this method is that prior information about the scanner or its fingerprint is not necessary to make a DLK decision. More clearly, our system can perform matching between pairs of patches without being trained with samples from the source of those pairs. Moreover, the performance degradation for post-processed images is a critical problem in the field of digital image forensics (DIF). In order to work in real world scenarios where images may be compressed for sharing and storage purposes, it is important to devise methods that are resistant to compression operations. In this work, we have proposed a new robust deep learning approach to forensically determine if two JPEG compressed images are of the same make and model of a scanner that captured them.
This chapter is organized into three parts. First, we come back to the existing schemes that have been designed for source DLK. In the second part, we detail the main principles of our approach. Some experimental results are then presented in the third part of this chapter so as to demonstrate the effectiveness of our scheme and how it is well suited for verifying the source similarities of a pair of images.

5.1 Device linking schemes

Given two questioned images I1 and I2, DLK is a dilemma in which a forensic investigator is required to choose between the following two hypotheses:

$$\begin{cases} \text{H0 : I1 and I2 were acquired by the same device} \\ \text{H1 : I1 and I2 were acquired by different devices} \end{cases}$$

This task is based on checking the forensic similarities of the traces left by each device in those images. It has been attracting more and more attention for images acquired by general public digital cameras, while works related to scanned or digitized documents have not yet received any attention.
Two categories of DLK approaches can be distinguished according to the way the image fingerprints are compared:

Fingerprints-matching through correlation - In these schemes, the fingerprint of each image is obtained by applying a wavelet based denoising filter on it [203] or on selected regions [204]. In [203], the authors extract, for each image, the noise residuals W1 and W2 and combine them with the original images as given in Eq. 5.1. X and Y are then compared statistically using the NCC.

$$X = \frac{I_1 W_1}{\sqrt{\sigma_2^2 I_1^2 + \sigma_1^2 I_2^2}}, \quad Y = \frac{I_2 W_2}{\sqrt{\sigma_2^2 L_1^2 + \sigma_1^2 L_2^2}} \tag{5.1}$$

Next, the peak sharpness is measured using the ratio between the primary peak to the secondary peak (PSR). PSR [205] is the highest value in the NCC excluding a central region around the primary peak. If the PSR is greater than a predefined threshold, the images are declared to be from the same camera. However, this method requires that the questioned images have the same size. As a result, we may need to pad the images with zeros or crop them which reduces the percentage of correctly classified matching pairs. The second approach [204] uses a support vector machine (SVM) in conjunction with a decision boundary carving procedure which consists in moving the decision hyperplane in a way that helps deal with images issued from unknown devices. As shown in Fig. 5.1, the residual noise in each ROI (Grey blocks) for each color channel R, G and B is estimated. After that, a feature vector is created based on the correlation between the noises in each ROI of both images. Each pair of images are, later, labeled as the positive (if taken by the same camera) or as negative (if taken by a different camera) before being fed into an SVM.

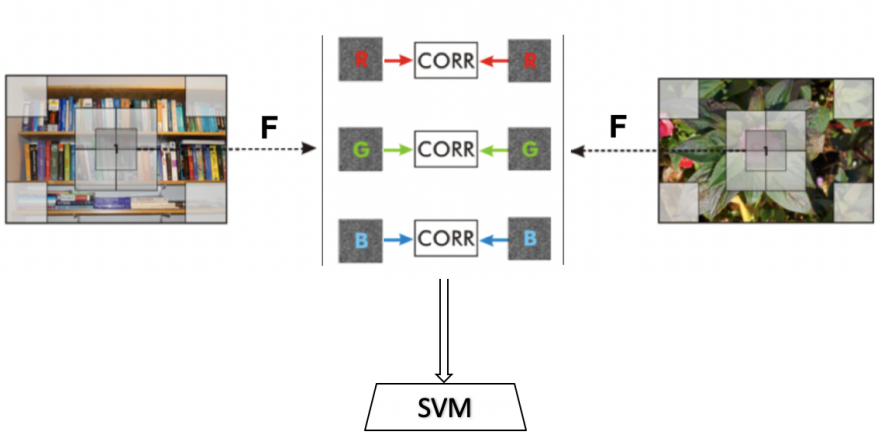


Figure 5.1: DLK scheme proposed by Costa *et al.* [204]. The correlation values between the noise residuals of each color channel R, G, and B of nine regions of interest (ROI) shown in the right and left images are used to train the SVM

Fingerprints-matching through CNN - The fingerprints-matching based schemes [206–208] are composed of a two-tiered network: One for deep feature extraction and one for similarity comparison. An overview of such schemes is given in Fig. 5.2. The first conceptual component is a network that maps images I_1 and I_2 into a low-dimensional feature vectors $f(I_1)$ and $f(I_2)$ that encodes important forensic patterns about the source device. Indeed, some recent studies proved that CNNs have shown significant improvements in features extraction from image patches [209]. Details of the network used as a features extractor in [206], [207], [208] can be found in [175]. Those vectors are, then, mapped to a similarity score by training separately the second component, a dual-input network, shown by a gray block in Fig. 5.2. This neural network is usually composed of two or three fully-connected layers followed by an output layer delivering a score that indicates whether I_1 and I_2 have different or similar forensic traces. If the score is 0 then one will consider that these images contain different forensic traces. On the contrary, a score of 1 means that both images contain similar forensic traces.

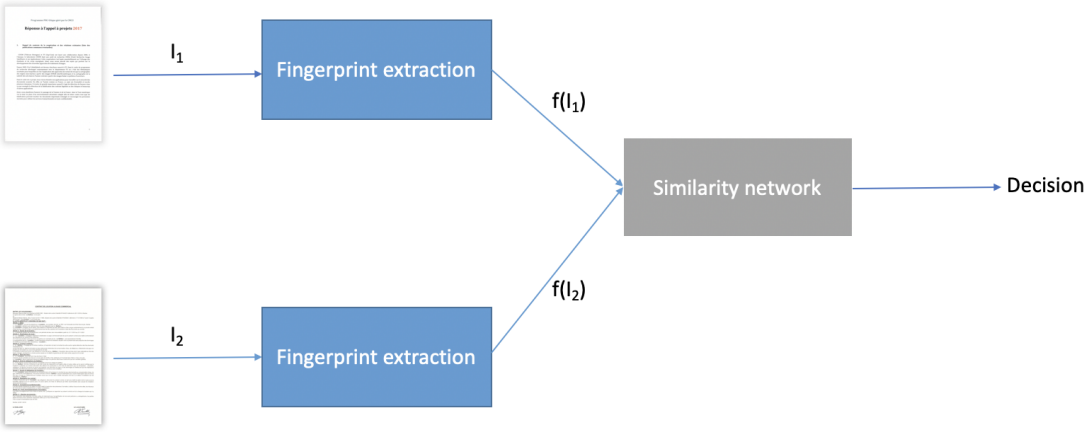


Figure 5.2: General structure of CNN-based DLK solutions

5.2 Proposed DLK Method

The approach we propose estimates the likelihood of a pair of scanned documents by using a dual-CNN architecture which will, first, extracts local forensic patterns from their associated equally sized patches via a 2D-CNN and, then, use another CNN composed of a number of fully-connected layers to compute a similarity score for each pair of patches. Later, those similarity scores are combined to produce a final unified prediction. The purpose of this study is to overcome the drawbacks of working in a closed-set context where prior knowledge of the source scanners is required which is generally not the case, given the large variety of flatbed scanners commercialized. Let us also recall that source identification systems are often not scalable, that is to say, they cannot meet the requirement of an increased number of devices. Furthermore, the fact that some forensic investigations do not require explicit identification of the particular device that has acquired a particular document. For example, in the splicing detection task, the digital investigator needs only to detect that some regions of the image have been acquired by distinct devices without specifically identifying those devices.
Figure 5.3 depicts the overall design of our DLK system. It consists of four major components. Firstly, the investigated images are decomposed into equally sized patches for model training. Secondly, a pre-processing is required so that those patches get converted to their corresponding high frequency subbands through a one-level wavelet decomposition. Thirdly, features of these subbands are compared using our system designed to estimate a similarity metric between every pair of patches. Finally, the decision is made based on a majority voting mechanism. A detailed description of the last three components are presented below.

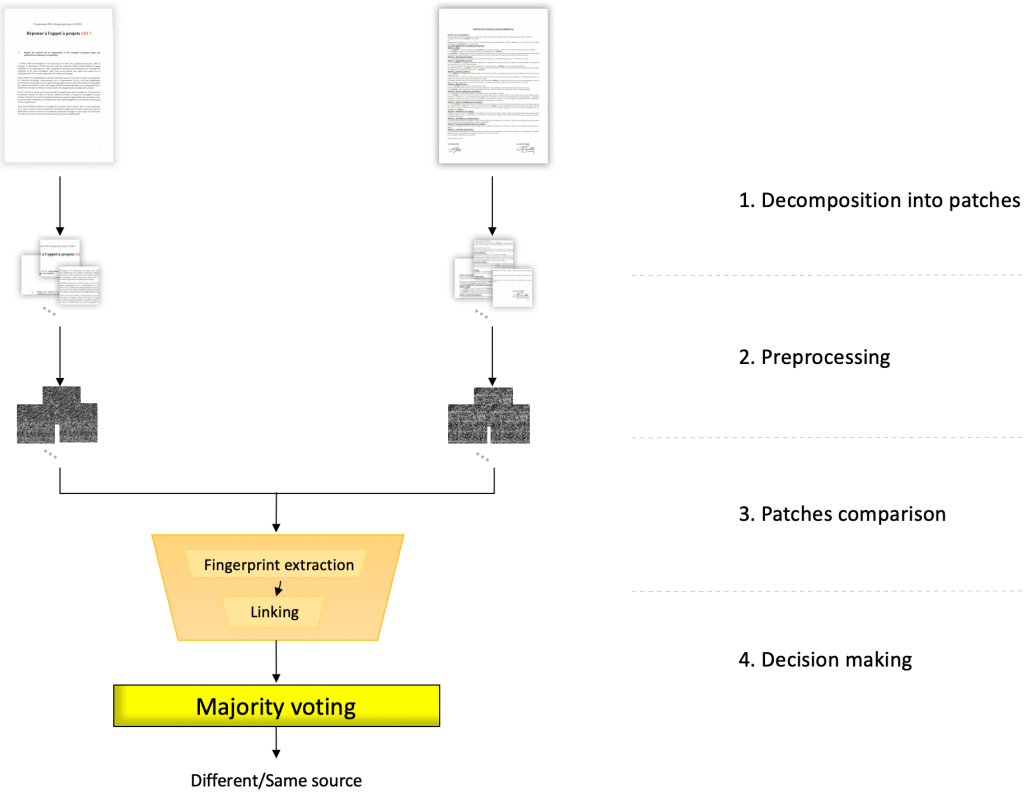


Figure 5.3: Global architecture of the proposed DLK process

5.2.1 Pre-processing

A popular method to minimize the effect of the image content is to apply a preprocessing function. In our previous works, we have shown the interest of the wavelet transform and, in particular, of SWT for achieving this goal. Moreover, we have also demonstrated that the most valuable scanning features that enable us to distinguish between scanners can be extracted from the HH subband. In this work, we propose to use this particular wavelet subband as an input to the first component of the proposed dual-CNN. This is the key behind efficient scanner fingerprint extraction which has a direct impact on DLK.

5.2.2 Patches comparison

The establishment of the proposed system involves two main subsystems, namely the fingerprints extraction module designed to map a feature vector to each patch and the classification module where each pair of patch vectors are compared. Its architecture is inspired from the existing CNN-based DLK architectures previously revealed in Fig. 5.2. As a matter of fact, it's important to note that our proposed patches comparison system uses the same position image patches each time.

5.2.2.1 Fingerprints extraction

The role of the first sub-system is to construct a representative features vector for each input image. This vector represents the forensic pattern of the device at the origin of that image and can be regarded as its fingerprint. To be effective, this fingerprint should be unrelated to the image's content.
If we refer to CNN-based studies that have been conducted in recent years, we will find out that they play an important role in underlying interpretable and powerful image features representations that can discriminate between different classes. More clearly, once a network has been trained for a specific classification task, relevant features could be extracted from one of its deep layers and then used to perform a different task. For example, features maps extracted from CNNs trained for some facial components (e.g. hair, mouth, eyes, nose, and beard) identification tasks have been demonstrated to be useful for face detection [210].
Therefore, some researchers began to develop forensics approaches to study traces left by acquisition devices using deep learning architectures. With regard to forensic features extraction, it has been shown that deep features extracted from a CNN trained for source device identification can be used for DLK [206–208] and IFD [73, 113].
In this work, we suggest conducting the fingerprints extraction using our 2D-CNN proposed in Chapter 4. Let us recall that this model consists of one convolutional layer followed by max-pooling and dropout blocks and two fully connected layers. The fingerprint is the output of our trained model truncated at layer 4 (see Table 4.1). It is composed of 512 high level features that encode forensic information about the source scanner.

5.2.2.2 Linking algorithm

After getting the HH subband from each patch, the above forensic fingerprints extractor, that we will call $2D - CNN_{shrinked}$, is used in a symmetric bilinear architecture to map the patches to their features vectors, as depicted in Fig. 5.4. Notice that both instances are employed under a hard parameter sharing paradigm. In fact, two fundamentally different types of paradigms commonly used in joint learning exist: the hard parameter sharing (HPS) and the soft parameter sharing (SPS) [211]. SPS means that each task has its own CNN and its own parameters but with the same topology, which is subsequently regularized using a similarity function. As a result, the amount of memory space used at runtime is proportional to the number of tasks. On the other side, HPS shares the hidden CNN layers between all tasks, and maintains task-specific fully connected output layers. It is one of the most common methods employed for multi-task learning [212]. HPS is able not only to minimize space complexity, but also to improve performance and reduce the risk of overfitting.
In order to accurately compare the fingerprints extracted from a pair of patches, the formulation of a novel secondary NN, that we will call "$ImSiM$", is proposed. This NN takes a pair of fingerprints as input and gives a similarity score so as to help decide if the pair of patches have similar or different forensic traces. The architecture is shown in Fig 5.4. As it can be seen, this one is composed of three Dense layers and a fusion function:

i) Two identical layers of 2048 neurons in a HPS mode with a ReLu activation function to process the outputs of the $2D - CNN_{shrinked}$ simultaneously

ii) Next, a deep features fusion module, specifically a concatenation function, is used to mix the output maps of the previous layers. This is an important step to integrate the features together and to obtain more prominent accuracy

iii) Further, a second layer of 128 neurons is followed by a parametric ReLu activation in which the fused features are embedded

iv) Finally, a single neuron layer is used to generate one score indicating if the inputs are issued from the same scanner or not. Basically, it provides a high score in the case of similar sources. We decided to choose Sigmoid as the scoring function based on experimental results we detail in the following section

Our design choices are supported by extensive experiments on a large number of image patch pairs.

5.2.2.3 Decision making

Given that investigated images' patches are classified based on the sigmoid output units, we propose to use a majority voting procedure. This later judges the source similarity between images by counting the number of times a pair of patches are classified to be acquired by the same scanner. It is the simplest strategy but it is also the most effective in terms of classification results. Based on the output of the $ImSiM$, we set a label of 1 (similarity) or 0 (dissimilarity) based on a threshold ν set based on precision-recall analysis in order to guarantee the highest accuracy. Basically, we compare the sigmoid output O to ν such that the label is

$$\begin{cases} 1 & \text{if } O \geq \nu \\ 0 & \text{if } O < \nu \end{cases}$$

A majority vote of 50% or more is required to classify the images to belong to the same device.

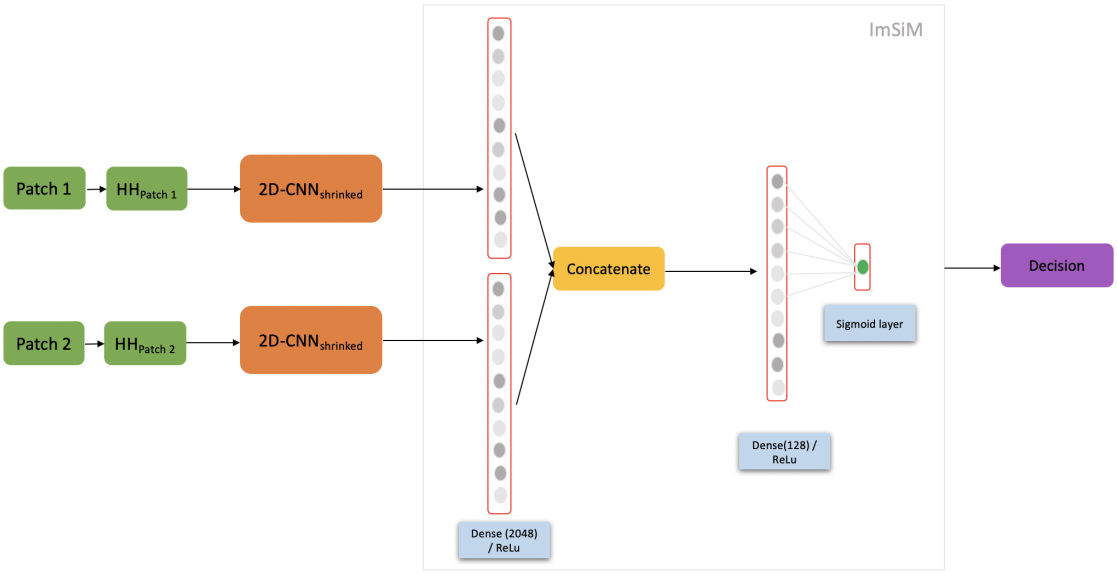


Figure 5.4: Overview of the patches comparison mechanism *ImSiM*. It is composed of a pair of layers in a hard sharing configuration, followed by a concatenation function and another layer of neurons. The decision is an output indicating the similarity score obtained from the last layer

Table 5.1: Image sources used in our experiments

Scanner Id	Brand	Model	Known\Unknown
S1	Canon	Lide 220	Known
S2	Epson	Perfection V39	Known
S3	Epson	Perfection V370 -1	Known
S4	Canon	Lide 120 -1	Known
S5	Epson	Perfection V370 -2	Known
S6	Canon	Lide 120 -2	Known
S7	Epson	Perfection V550	Unknown
S8	HP	Scanjet Pro 2500 F1	Unknown

5.3 Evaluation

Extensive experiments were conducted to examine the performance of the proposed DLK algorithm. It includes the evaluation of the influence of the fusion mechanism, the classification accuracy of our full system, the effect of JPEG post-processing and comparison performance assessment with similar methods for general public cameras.
The following series of experiments have been performed on images from 8 scanner instances where two of them were not used for training our model. More clearly these two devices are considered as unknown material as listed in Table 5.1. Note also that, we have considered scanners of the same brand and model which are S3 and S5 from the Epson brand and S4 and S6 from the Canon brand.

5.3.1 Parameters settings

In all our experiments, we have used our trained 2D-CNN model described in the previous Chapter for features vector extraction. It is important to recall that, as stated earlier, no training or fine tuning phases were additionally performed on this sub-system.
On the other side, $ImSiM$ was trained by a binary cross entropy as loss function along with an Adam optimizer. It produced high accuracy, and the validation error attended the minimum after being trained for 8 epochs, with a batch size of 4, a learning rate of 0.0001 and a decay factor of 0.1. The training was done on a ZOTAC GeForce RTX 2070 SUPER AMP EXTREME GPU.
Experimentations were conducted on the same images used in the previous chapters acquired by flatbed scanners with a variety of native resolutions. In our experiments, the documents were scanned at 300dpi resolution. We split these images into two disjoint sets: K and Uk. Images from K were used to train our system and were only acquired by scanners indicated as 'known' in Table 5.1. Both sets are then used to test the performance of our proposed model. From K, we generated 168 random pairs of patches where each pair corresponds to patches taken from the same position in two different images acquired by the same device. Next, we built, in the same way, a second set of 280 pairs from images acquired by different devices. Later, 80% of those 448 pairs were used for training and 20% for validation in a random stratified manner. For evaluating our network, we considered other pairs of patches by randomly selecting 448 image patches from K, where 35% of patch pairs were chosen from the same device, and 65% from different devices. Each pair was given a label of 0 or 1 depending on whether they were acquired by different or the same scanner.
We consider only non-overlapping 128x128 patches and we run the DLK algorithm with matching and non-matching patch pairs.
For full system testing, we experimentally determine that a threshold ν=0.7 results in more reliable accuracy scores.

5.3.2 Performance of the proposed system on image patches

Figure 5.5 depicts the evolutions of the training and validation curves of the $ImSiM$ network over 8 epochs. The training and validation accuracy are given on the right while the training and validation loss are placed on the left. One can observe that both curves jointly increase and decrease throughout the training epochs. We can also notice that, after the 5^{th} epoch, the accuracy and the loss evolution almost stop and converge to approximately the same rates. Therefore, our model is not overfitting.

5.3.2.1 Effect of the feature fusion method

To better understand the choice of the concatenation module in the $ImSiM$ network depicted in Fig. 5.4, we investigated the performance of some fusion methods on our dataset. Thus, a set of experiments were conducted in order to determine which one is the best suited for our DLK system. In fact, various feature fusion mechanisms and their application is very dependent on the desired outcome. In this section, we compared the following four different fusion methods:

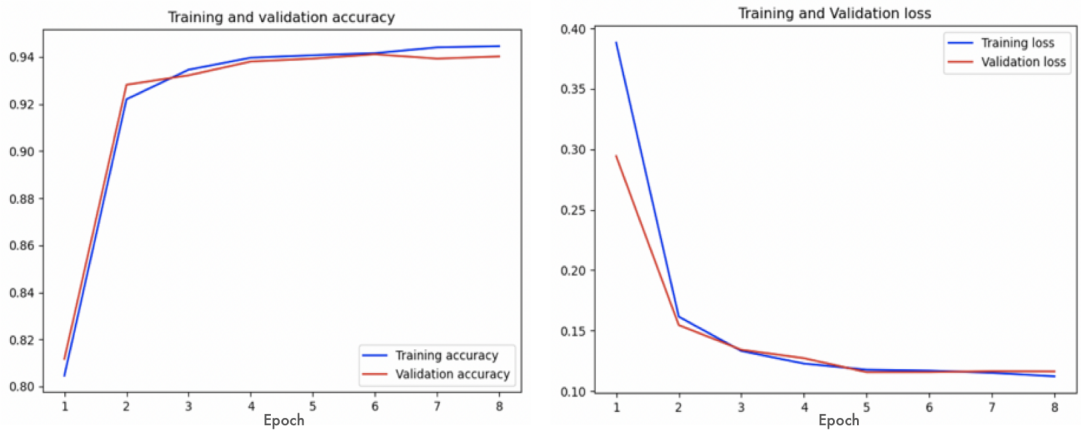


Figure 5.5: Training and validation performances of the proposed method

- **Concatenate** - Basically, this solution groups together different features. It is the most widely used because it allows the upstream network to decide how to use the data without discarding information or information loss. If we consider that the concatenation function is F_{concat} then $y_{i,j} = F^{concat}(x^1_{i,j}, x^2_{i,j})$ is such that

$$\begin{cases} y_{2i-1,j} = x^1_{i,j} \\ y_{2i,j} = x^2_{i,j} \end{cases} \tag{5.2}$$

  where $x^1_{i,j}$ and $x^2_{i,j}$ correspond to elements of the first and the second features the network has to take into account, respectively, with i=1..N and j=1..M. N and M are the number of rows and the number of columns of the concatenated inputs, respectively.

- **Subtract** - This operation is commonly used to determine how close one feature to another is. The output is obtained by subtracting one input from the other one following this formula

$$y_{i,j,k} = x^1_{i,j} - x^2_{i,j} \tag{5.3}$$

- **Multiply** - It computes the element-by-element product of two input data. If the same special position's input elements are either positive or negative, their multiplication will be positive. This information may be used to determine the similarity between the inputs. The output result is obtained in this way

$$y_{i,j} = x^1_{i,j} * x^2_{i,j} \tag{5.4}$$

- **Average** - It is a common merging function for networks that adds inputs which are in the same spatial position (i, j). The calculation of the final output $y_{i,j}$ is as follows

$$y_{i,j} = (x^1_{i,j} + x^2_{i,j})/2 \tag{5.5}$$

In terms of size, the traditional concatenation method is different from the other methods. As an example, when considering the case of two input images of same dimensions, the output of this fusion function has twice the input dimensions whereas the other methods maintain the same size, i.e. the size of their inputs.
According to the above-mentioned fusion methods, cross-validations were carried-out

Table 5.2: Accuracy comparison using different fusion methods

	Concatenate	**Subtract**	**Multiply**	**Average**
Cross-valid 1	**94.08%**	91.53%	92.73%	92.57%
Cross-valid 2	**93.78%**	91.74%	93.28%	91.53%
Cross-valid 3	**94.76%**	90.63%	91.62%	92.97%
Average accuracy	**94.02%**	91.3%	92.54%	92.35%

to get a fair accuracy rate. As shown in Table 5.2, the concatenation fusion method used in this work appears to be the best choice as it gives the highest accuracy rate of 94,02% resulting from averaging the accuracy values obtained by repeating the cross-validations three times. It can be also concluded that the multiplication method, opted by state-of-art approaches, gains a small margin over the average and subtraction fusion methods by 0.19% and 1.24%, respectively. In fact, since weights were shared when extracting fingerprints from the image patch pair, the resulting feature vectors of the two $2D-CNN_{shrinked}$ branches will have their corresponding features in the same location. Thus, information about the location is kept, which is why the concatenation function brought improvement.

5.3.2.2 Effect of the pre-processing and the scoring function

We evaluated the impact of the pre-preprocessing step on the validation accuracy, that is to say the influence of using the HH subband from the SWT decomposition of the input image patches to feed the fingerprint extractor. It can be seen from Fig. 5.6. that the images' patches directly applied as input to the CNN are less effective with an accuracy of around 75%. This is in line with our observations made in Chapter 3 and 4, where the effectiveness of this preprocessing stage has been demonstrated for scanner features extraction.
It was also observed that, when comparing the output activation functions, the proposed method achieves a relatively better performance when the Sigmoid function is used in the final layer of the proposed $ImSiM$ compared to the Softmax function. Let us recall that the Sigmoid activation function is more adapted to the problem of binary classification; this justifies the results we obtained.

5.3.2.3 Effect of lossy image compression

As already stated in Chapter 3, the identification and the extraction of device fingerprints have become more challenging when the acquired images and documents have been subject to post-processing such as compression; a hypothesis far from being unrealistic with the increasing use of image editing tools by the general public and professionals.
Here, we investigated the impact of lossy compression on DLK performance. To do so, we JPEG compressed images while varying the JPEG quality factor (QF) (see Section 3.1). We examined performance at QF $\in$ 50, 60, 70, 80, 90. What is interesting in our experiments is that only $ImSiM$ training is required, i.e. no need to train the fingerprint extractor as we will use our pre-trained 2D-CNN. Finally, the accuracy measures were computed for patches of sizes 64x64 and 128x128 and are shown in Fig.5.7.

As expected, DLK performance decreases with heavier compression (lower QF). However, our system is not too much impacted by JPEG compression. Another interesting observation is that both patch sizes show a similar performance degradation. This demonstrates strong robustness against patch size decline.

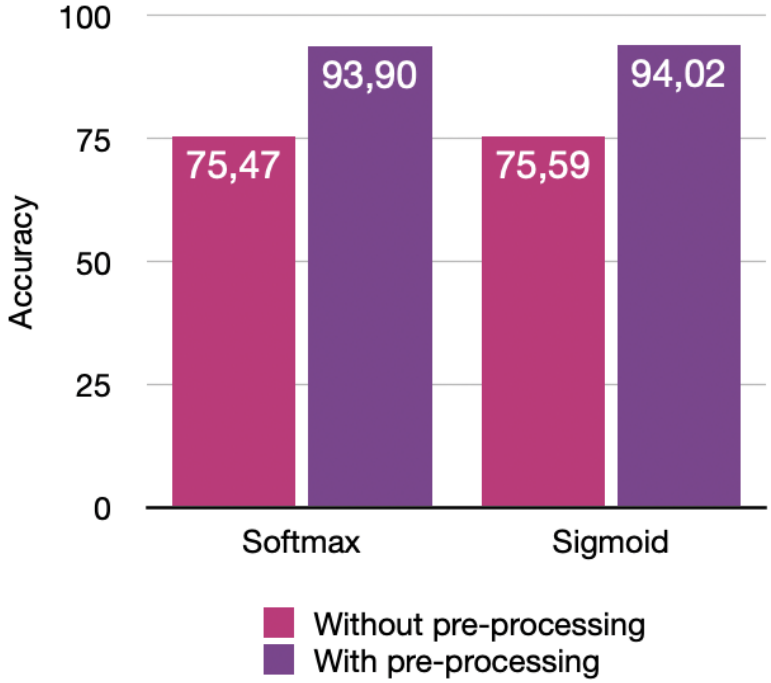


Figure 5.6: Change in the classification performance of our system when the pre-processing step was removed with a comparison of the effect of the scoring function: Softmax vs Sigmoid

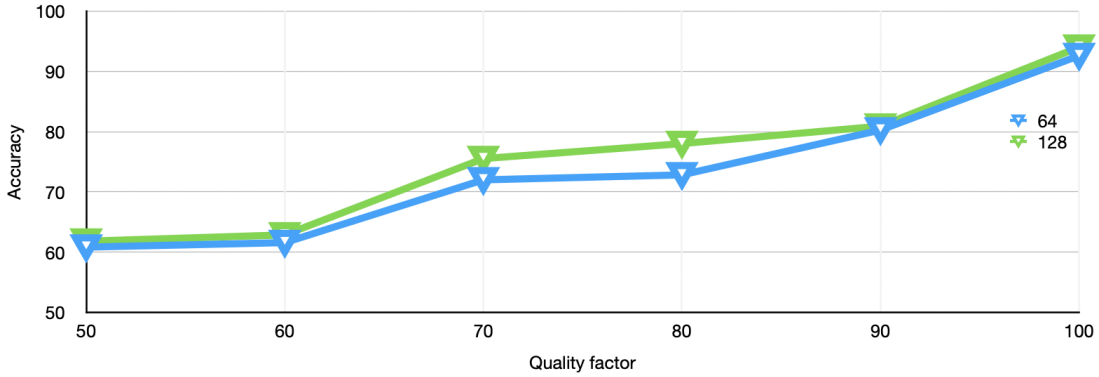


Figure 5.7: Accuracy rates with patches size 64x64 and 128x128 for JPEG compressed images

5.3.2.4 Comparison and analysis

In the next series of experiments, we applied different standard algorithms to replace the proposed $ImSiM$ for pairs matching. Experiments have been performed considering each time one of the following classifiers:

- **SVM with a linear kernel** - This one is the most basic SVM classifier type. It is preferably used to classify linearly separable data. The linear kernel function F_L is such that

$$F_L(x_1, x_2) = x_1^t.x_2 \tag{5.6}$$

where x_1 and x_2 are the data to classify.

- **SVM with a Gaussian Radial Basis Function kernel** - Such a SVM is based on a general-purpose kernel used when no prior knowledge about the data is available. Its kernel function F_G is expressed as

$$F_G(x_1, x_2) = \exp\left(-\gamma \|x_1 - x_2\|\hat{}2\right) \tag{5.7}$$

- **Linear Regression [213]** – It is a statistical technique for predictive modelling analysis used to explain the relationship between a dependent variable (output) and one or more explanatory variables using a straight line.

- **Random Forest [214]** - It is a machine learning algorithm that produces decision trees, each of which gives a vote for a certain class through a subset of training samples.

When compared with the overall accuracy of these approaches, accordingly to experimental results summarized in Table 5.3, $ImSiM$ has a significant advantage. As it can be seen, the Random Forest Classifier also provides good classification accuracy of 93,02%. In terms of complexity, our scheme is a bit slower due to the dense layer applied before the concatenation of the $ImSiM$ inputs compared to other methods that directly fuse these inputs.

Table 5.3: Performance of the DLK system with different linking algorithms (SVM, Linear Regression, Random Forest and the proposed $ImSiM$)

Topology	**Accuracy**
SVM (Linear)	78,31%
SVM (Gaussian)	88,18%
Linear Regression	73,16%
Random Forest	93,02%
$ImSiM$	**94,02%**

Our system is then compared with the most recent and efficient approach [208] proposed for DLK of images acquired by digital cameras across two different patch sizes. Figure 5.8. mentions the performance obtained from each model. A visualization of these performances clearly discerns the fact that our model outperforms Guru *et al.* [208] model with more than 30% superiority for 64x64 and 128x128 patches. In contrast to DLK methods related to cameras, our system takes advantage of the pre-processing of the patches to extract device features in a more accurate way. This may be the reason behind the huge difference between their performances. Another reason is the scalability of the fingerprint extractor. More clearly, our $2D - CNN_{shrinked}$ model may not be adapted to extract digital camera features and vice versa. Indeed, training the fingerprint extractor adopted by Guru *et al.* [208] on our dataset resulted in an accuracy of 50% which means that it did not properly learn scanners patterns as it did for digital cameras.

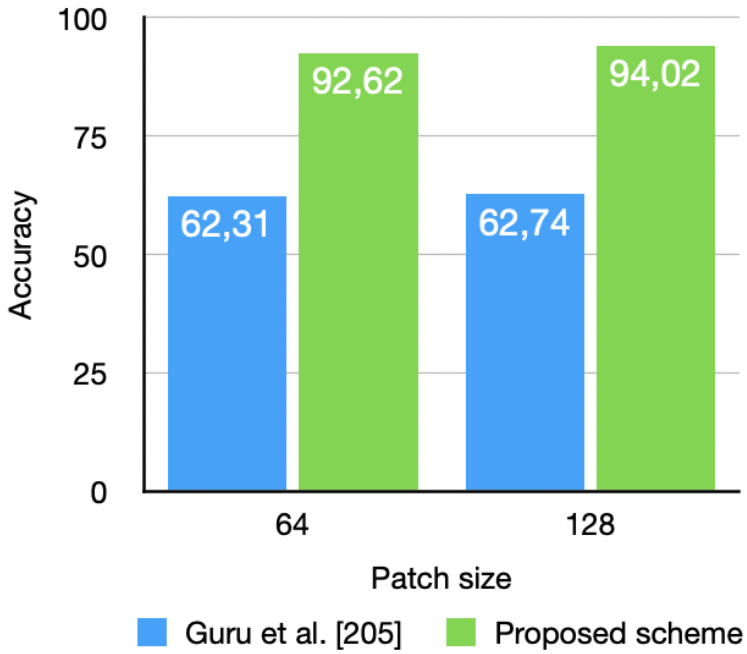


Figure 5.8: Matching rates of our proposed scheme for different patch sizes compared to an approach adapted from Guru *et al.* [208]

5.3.3 Performance of the proposed system on full images

In the previous section, we assessed the performance of our method in detecting whether a pair of patches belongs to images acquired by the same device or not. However, in real life situations, most image documents are investigated and, due to their large size, they cannot be directly tested through a neural network.
If we consider two images I_1 and I_2, the first step is to cut them out into non-overlapping patches. Then, SWT is applied to each pair of patches P_i from I_1 and L_i from I_1 located in the same position so as to get their corresponding high frequency subbands HH_i^1 and HH_i^2. These transformed patches are, next, normalized. In fact, it is an extremely common practice to scale the CNN inputs to have zero mean and unit variance [215]. To do so, we simply scaled them by 1/255. The following step is to map to their features vectors fv_1 and fv_2 using the $2D-CNN_{shrinked}$ described above. To evaluate the similarity between fv_1 and fv_2, they are fed into our $ImSiM$ network which measures a matching score between 0 and 1.
After that, this score is compared to a predefined threshold that will predict if the patches are issued by the same scanner (prediction label is 1) or not (prediction label is 0).
The final step regarding the decision for the images consists in applying a majority voting mechanism which will decide, based on the number of pairs labeled as 1, whether the inputs I_1 and I_2 are of the same or different origin. Algorithm 2

provides a pseudocode detailing this process.

Algorithm 2: Comparison of a pair of images

Input: Images I_1 and I_2 acquired by same\different scanners
Output: 1 if same source scanner and 0 if different
Sample I_1 and I_2 into patch sets S_1 and S_2
for *each patch* P_i *in* S_1 *and its corresponding patch* L_i *in* S_2 *where* $1 \leqslant i \leqslant N$ *and* N *is the total number of patches in each set* **do**
$HH_i^1 \leftarrow SWT(P_i)$ `// HH subband`
$HH_i^2 \leftarrow SWT(L_i)$
$HH_i^1 \leftarrow HH_i^1/255$ `// Normalization`
$HH_i^2 \leftarrow HH_i^2/255$
$fv_1 \leftarrow 2D-CNN_{shrinked}(HH_i^1)$
$fv_2 \leftarrow 2D-CNN_{shrinked}(HH_i^2)$
$score \leftarrow ImSiM(fv_1, fv_2)$

if $score > \nu$ **then**
  pred.append(1)
else
  pred.append(0)
final $\leftarrow Majority_Voting(pred)$
if *final == 1* **then**
  I_1 and I_2 are declared as acquired by the same scanner
else
  I_1 and I_2 are declared as acquired by different scanners

To evaluate our model on full-size images, we have used 324 pairs from both K and Uk, which were divided into 128x128 patches. Figure 5.9. displays the similarity matrix resulting from testing our model. The diagonal line shows the correct predictions rates of images acquired by the same source scanner, while off-diagonal rectangles show the correct predictions rates when images were captured by different devices. As an example, the system correctly identifies that pair of images which are both acquired by the scanner S4 with an accuracy of 85.71%. Furthermore, when the first image was captured by the scanner S4 and the second image was captured by the scanner S5, the system correctly identifies that they are issued by different devices with an accuracy of 100%. In general, our proposed system is able to achieve approximately 96% accuracy.

Generally, it can be noted that for unknown scanners the total classification accuracy is 100%. On the other hand, we can notice that there are pairs for which our system does not attain high accuracy. These situations seem to occur when the pair of patches were acquired by scanners of the same model. A notable example is when the scanner model is Epson Perfection V370, our system got no correct prediction for pairs of patches acquired by a different instance of this model. This was most likely owing to the hardware and processing pipeline similarities between the two devices, which resulted in highly comparable forensic traces. Few prediction errors also occurred when the pair of patches are acquired by the scanner S4 and S6.

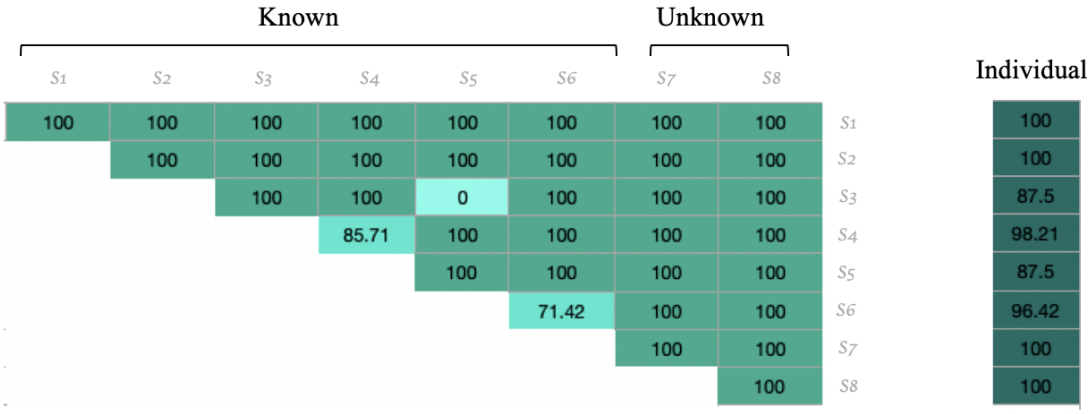


Figure 5.9: Accuracy results of our DLK system when applied to full images. Each block from the diagonal corresponds to the rate of correctly identifying a pair of images as sourced from the same scanner. The remaining blocks are the result of two images acquired by different scanners. The column on the right refers to the average accuracy for each scanner

5.4 Conclusion

In this chapter, we have focused on the problem of DLK taking advantage of our previous approach on image origin identification through a 2D-CNN. As exposed, existing SSI methods assume a closed-set of scanner models which limit their utility especially in cases when a large number of scanner instances is investigated.
The proposed approach is the first technique designed for scanned documents. It allows forensics investigators to verify if two images were acquired by the same device or not without the need to know the brand or the model of those devices. To accomplish this, we developed a hybrid 2D-1D-CNN based system that, given two questioned images, makes a decision about their source similarity by comparing their corresponding noise patches. It is composed as follows: 1) A pre-processing step to obtain the HH subband of each patch 2) The 2D-CNN for features vector extraction, 3) Patches' patterns similarity evaluation which is a 1D dual-entries CNN taking a pair of features vector at a time and finally, 4) A majority voting tool is performed by counting the pair of patches whose similarity score exceeded a certain threshold. We experimentally demonstrated that our system accurately maps pairs of patches as well as full images, although it fails to link images from some specific scanners of the same brandwith about 3% classification error.
Moreover, our simple yet effective approach may offer a new perspective on the multi-class open-set classification problem. In another word, it can be extended to find if the origin of a questioned image is not one of the known scanners, which is a common forensics scenario. This can be done by comparing that image to one image acquired by each of these scanners. That information can minimize the number of incorrectly classified pairs resulting from assigning an image of an unknown model to one of the known ones.
Given these promising results, our future work will focus on testing on different databases and exploiting the applicability of our approach to test if two images have the same processing history. In line with those goals, it is also desired to re-investigate the features extracted from scanners of the same model which are usually evaluated as similar due to the similarity between their manufacturing process.

The most important persuasion tool you have in your entire arsenal is integrity.

L'intégrité est l'outil de persuasion le plus important dont vous disposez dans tout votre arsenal.

— Zig *Ziglar*

6

NEW ANNOTATED IMAGE DATASET TAILORED FOR IMAGE FORGERY DETECTION

Contents

In this chapter, we introduce a novel image collection annotated with respect to each scanner as a useful tool for forensics investigators to test and compare scanner-based forensic techniques. We call it 'The SUPATLANTIQUE' dataset [19] referred to the names of universities I belong to. It is an image database that contains document of various content scanned with more than one resolution with 11 different scanner instances of widely known brands. Moreover, our dataset is suitable for the study of image manipulations. It comprises more than 100 forged images created using various forgery operations.
Besides, the growth of fake data has raised the attempts to detect and localize areas of forgeries in forged images. To address the second question related to forgery detection (FD) which has been announced in the introduction of this manuscript, we propose two new approaches based on the findings, previously presented in chapters 3 and 4, to detect tampering independently of the size of the forged area.
The rest of this chapter is organized as follows: In the first part, we present the protocol of acquisition and the organization of our dataset. Then, we follow up with two schemes that, unlike existing schemes in the literature, are able to accurately localize the manipulated regions in a forged scanned image. Finally, experimental results and comparison of both methods are given and discussed.

6.1 New Image dataset: Acquisition and organization

When working on source camera identification related problems, most methods can be implemented and tested quite easily thanks to a large number of available public datasets. In Fig. 6.1, we show the variation trend of the annual number of publications in the field of Digital image forensics (DIF) related to the source identification of cameras and flatbed scanners where the black line reflects scanners-related works while the green one represents works related to cameras. The availability of publicly available data collections like Flickr, the 'Dresden Image Dataset' [216] and the Raise dataset [217] show clearly why cameras have gained greater attention. It is also important to mention that the demand for public benchmark datasets is rising in wide range of applications apart from DIF such as medical [218], solar [219] and human action recognition [220].

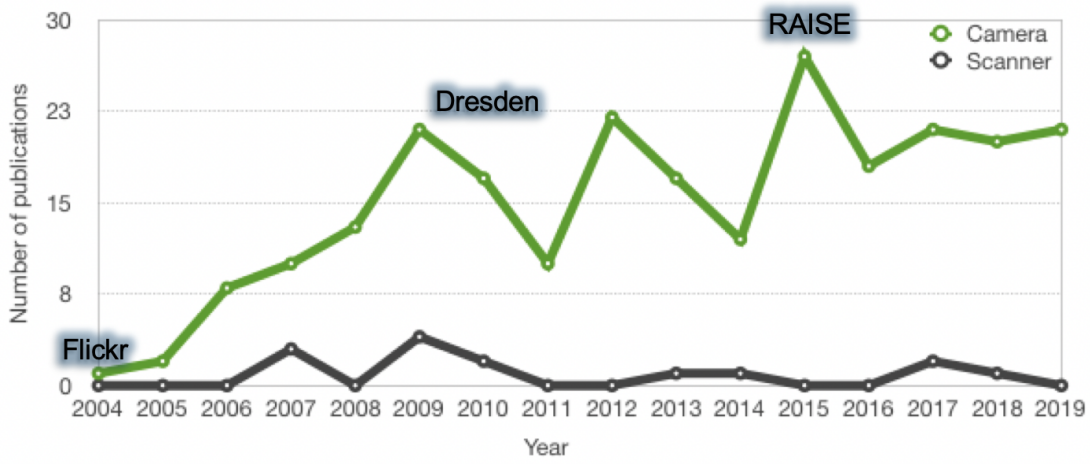


Figure 6.1: Annual number of publications related to source scanner (black) and source camera (green) identification

For scanners, there is an urgent need for a standardized image database. Researchers are wasting too much time and money in building their own datasets. Therefore, we acquired a novel and large-scale dataset using 11 flatbed scanners of different brands and models commonly used in offices. The images in our dataset are annotated according to each brand which include Canon, Epson and HP as well as other characteristics that we will explore in the following. Note that it was necessary to create an extra dataset since the images used in the previous chapters are subject to standards of privacy and confidentiality.
The full dataset is available for download upon request at `https://sites.google.com/view/supatlantique-dataset/downloads`.
An overview of the scanners used to create the dataset is given in Table 6.1. We distinguish them according to their:

- Native resolution (nominal resolution)
- Acquisition technology (CIS/CCD)
- External bit depth: Number of bits used to represent a given color pixel in the image transferred from the scanner to the host computer

Table 6.1: List of scanners used to constitute the 'SUPATLANTIQUE'

Id	Brand	Model	Technology	Resolution	External bit depth	Price
S1	Canon	LiDe 120 –d1	CIS	2400x4800	24	**
S2	Canon	LiDe 120 –d2	CIS	2400x4800	24	**
S3	Canon	LiDe 220	CIS	4800x4800	24	*
S4	Canon	CanonScan 9000F MKII –d1	CCD	9600x9600	48	**
S5	Canon	CanonScan 9000F MKII –d2	CCD	9600x9600	48	**
S6	Epson	Perfection V39 –d1	CIS	4800x4800	24	**
S7	Epson	Perfection V39 –d2	CIS	4800x4800	24	**
S8	Epson	Perfection V370 Photo –d1	CCD	4800x9600	48	***
S9	Epson	Perfection V370 Photo –d2	CCD	4800x9600	48	***
S10	Epson	Perfection V550 Photo	CCD	6400x9600	48	***
S11	HP	Scanjet Pro 2500 F1	CIS	1200x1200	24	***

- Price : Low('*'), medium('**'), high('***')

One can notice the presence of scanners of the same brand and model marked by -d1 and -d2. Those scanners will serve to solve the problem of scanner instance identification which is a more challenging and deeper lever of identification as shown in our previous chapters. All class levels are exposed in Fig. 6.3 with examples from scanners listed in Table 6.1.

The workflow of the acquisition of the dataset is shown in Fig. 6.2. We defined two types of documents: Official documents and Wikipedia documents. As official records in the real word are normally private or covered by copyrights, we have decided to create synthetic documents with artificial content to be scanned. By 'official', we mean a document which may serve as a proof in criminal and civil proceedings. We began by creating 100 realistic documents (e.g. financial statements, contracts, visa forms, civil certificates) using Microsoft Word. We ensured that these documents' contents are miscellaneous by varying setting options such as colors (i.e., full-color or grey levels), the text format as well as the font size and type (i.e., hand-written or computer-generated). Visual elements such as graphs, logos, and tables were also added. We generated logos using [1], a free online logo maker.

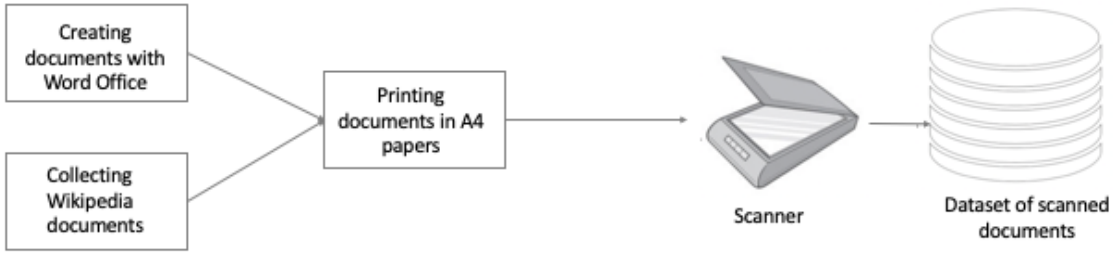


Figure 6.2: Workflow of generating our dataset

[1] https://www.ucraft.com/free-logo-maker

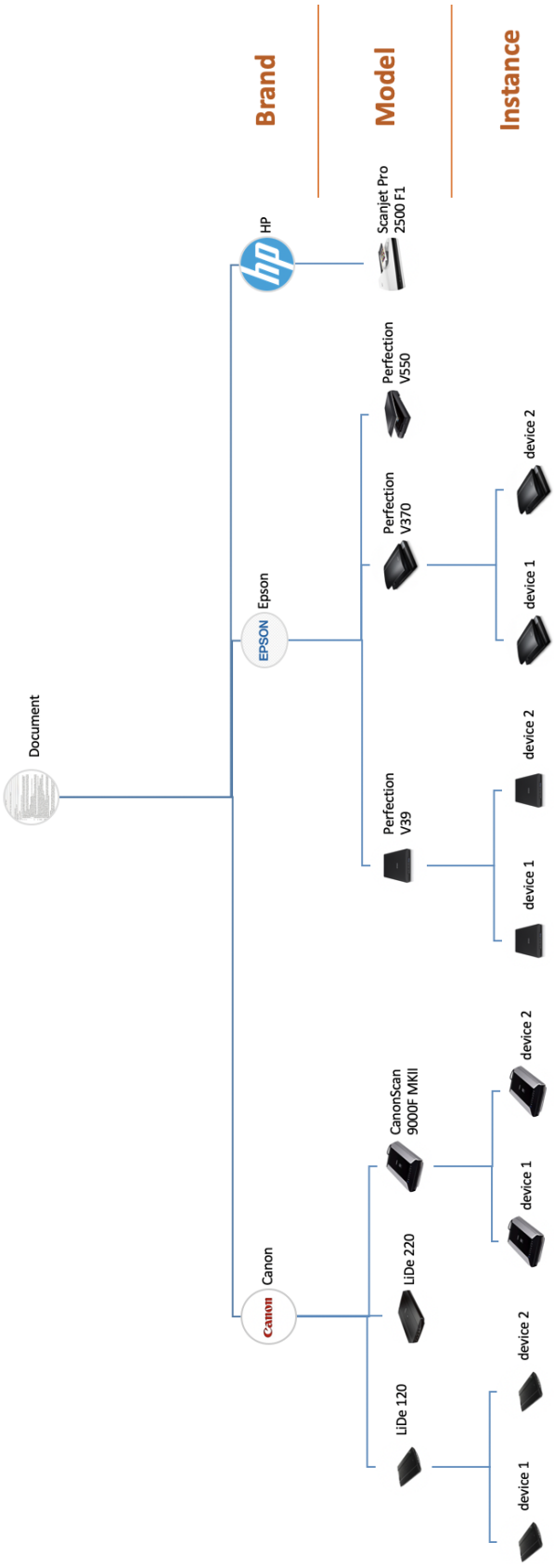


Figure 6.3: SSI levels with scanners from our dataset as examples

In addition to the manually generated documents, we downloaded some freely available file documents from the Internet.
On the other hand, we picked 108 printable versions of Wikipedia articles in different foreign languages (e.g. Arabic, French, English, Turkish) with various letter sizes, colors and fonts.
All documents are then stored in pdf format and printed in color using the same printer (RICOH MP C3004ex) in A4 paper. The use of one printer is important to avoid the interference between its defects and scanners ones. The printing resolution is fixed at 600dpi.
The last step consists in using VueScan Pro 9.7.02 to scan these documents at two different resolutions:150dpi and 300dpi. This software includes drivers of more than 6000 scanners which is time saving and recognizes almost all scanning parameters. Once digitized, the documents were saved in uncompressed TIFF format. Advanced parameters, including contrast and brightness, remain unchanged. Examples of these images are shown in Fig. 6.4.

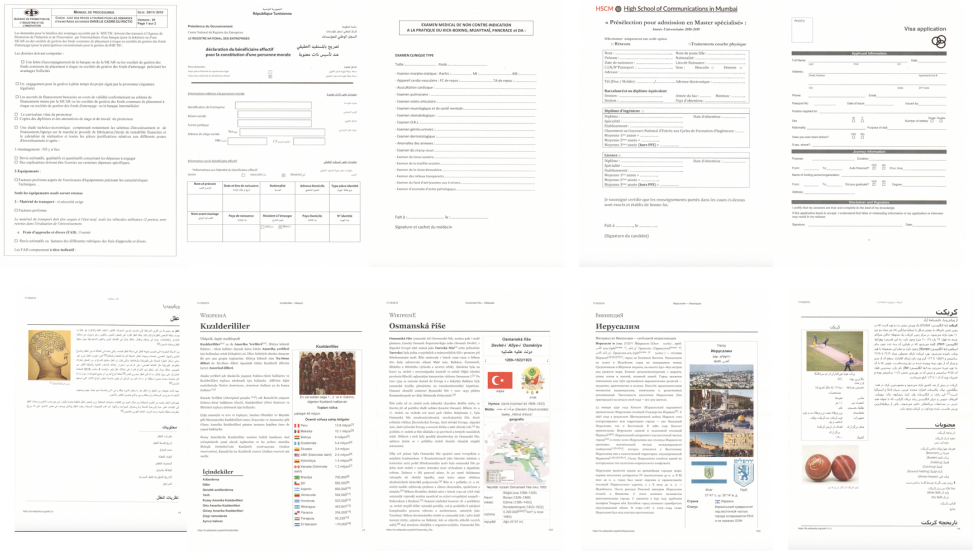


Figure 6.4: Examples of images from the SUPATLANTIQUE dataset: The first row corresponds to "official" documents and the second row corresponds to Wikipedia documents

We further included flatfield images, also called calibration images, for each scanner and each resolution.
Image forgery detection (IFD) is a very challenging task and no image dataset is yet proposed that include image manipulations of scanned documents. Thus, another directory, whose structure is shown in Figure 6.5, is built by manipulating 34 documents (scanned at 300dpi). For each one of these documents, we applied three common forgeries which are copy-move, splicing, and retouching and, for each type of forgery, we attempted to produce a large number of semantically meaningful manipulations that are not obvious to the human eye. The images were manipulated using Adobe Photoshop. Moreover, the modified pixels may correspond to realistic regions or to fixed blocks. Finally, we stored the original image, the manipulated output image, and a ground truth binary mask showing the altered pixels as well as a detailed text file in which every operation made on the original images and its related parameters are provided.

An illustrating example of a retouching operation is given in Fig. 6.6. It is important to mention that we make use of the copy-move operation when we need

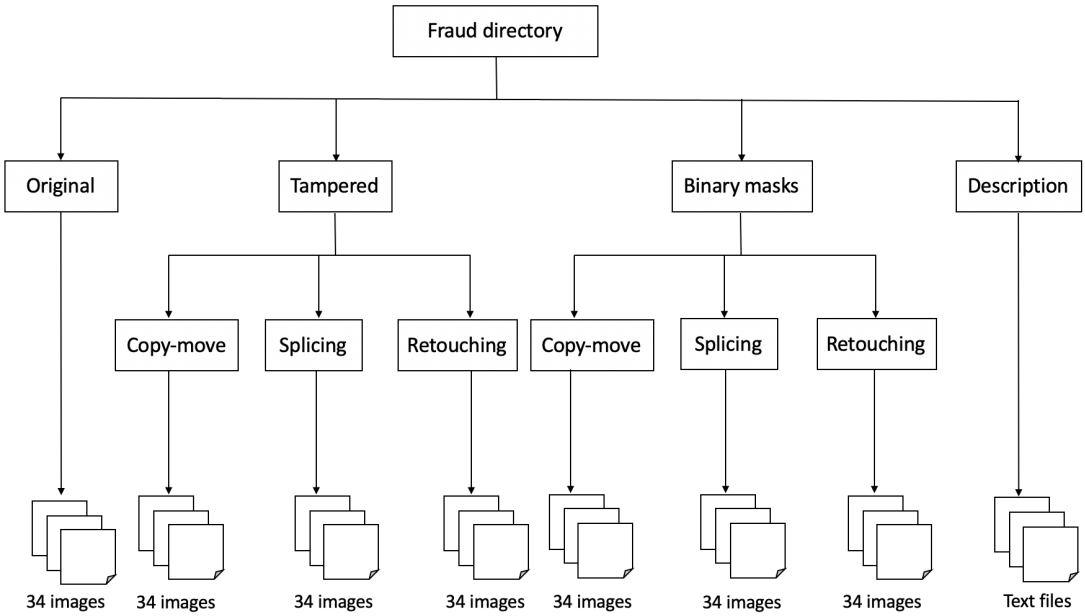


Figure 6.5: Fraud folder structure

to replace a text in the document without leaving visual traces. We created a fake image (b) by copying a white rectangular region from the same image and placing it over the parts to be manipulated (highlighted by orange boxes in Figure 6.6). Then, we add text in those regions. The added text should have the same or similar color, font and size as the original one so the manipulation could be unnoticed by the naked eye. After that, we generate the binary mask by drawing black regions for all manipulated pixels over a white background.

To summarize, the SUPATLANTIQUE database, which comprises 208 document files and a total of 4700 images, is composed of:

- Digital original official documents
- Digital original Wikipedia documents
- Flatfield images
- Official documents scanned at two different resolutions
- Wikipedia documents scanned at two different resolutions
- Fake documents with ground truth

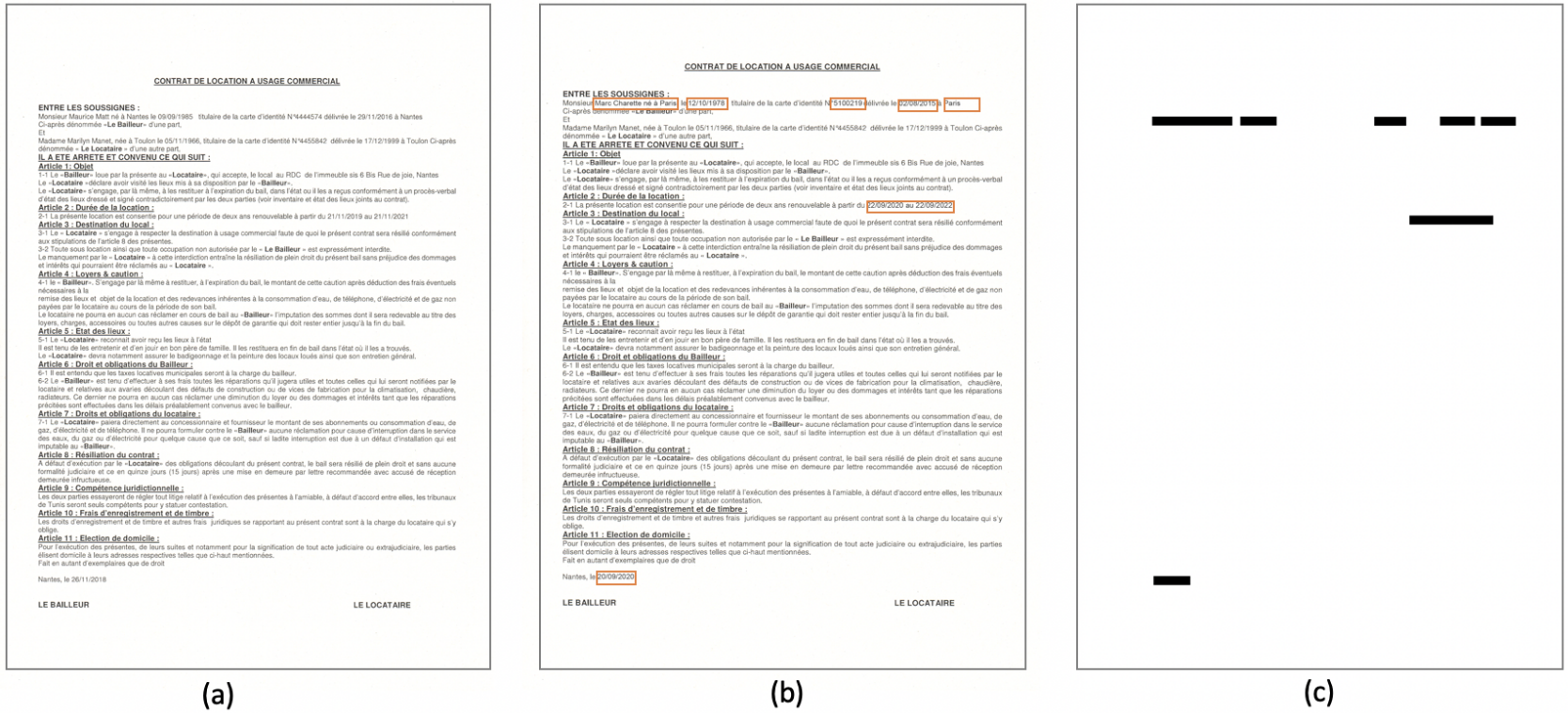


Figure 6.6: Retouching forgery example (a) original image (b) forged image with the manipulated regions marked by orange rectangles (c) binary mask

6.2 Optimal forgery detection

Nowadays, even one with amateur editing skills can create realistic documents by combining a variety of pictures, deleting objects, adding text, etc. These manipulated images are often used to spread fake news, forge signatures and lead to law violation. In many countries, altering a document and using it as authentic is a crime and the forger can be sentenced in prison. Unfortunately, new techniques for creating forged documents make the task of IFD very hard and, until recently, proposed IFD methods are far from being reliable. Furthermore, forensics methods related to scanned images received much less attention from the forensics researchers' community. As shown earlier, those methods can either identify if an image is forged or not [221] or assume that it is forged and try to detect the anomalies in it. More clearly, as they are focusing on splicing detection where an image is composite of two or more images, the localization is made based on finding inconsistencies between the investigated image regions. Their experimental results are very preliminary and only use handcrafted features. Moreover, they are limited by the size of the block (region) as reported in Table 2.2. It is also important to mention that none of these methods has been tested on text documents.
This section provides details on the proposed studies for detection and localization of splicing forgeries. For further details on this type of forgery and its related state-of-art methods, please refer to Chapter 2.
The key idea behind our approaches is that both forgeries detection and localization are based on fingerprints discrimination among different scanners. A summary of the proposed methods is as follows. The generated frameworks usually start by a pre-processing step. It will, then, subdivide the processed image into patches. The third step is the patches matching. This step plays a critical role and thus, we will mainly focus on it. Finally, three significant detection maps are generated to localize the forged regions. Figure 6.7 illustrates the general pipeline for the proposed methods. Notice that the order of the image division and the pre-processing may be inverted.

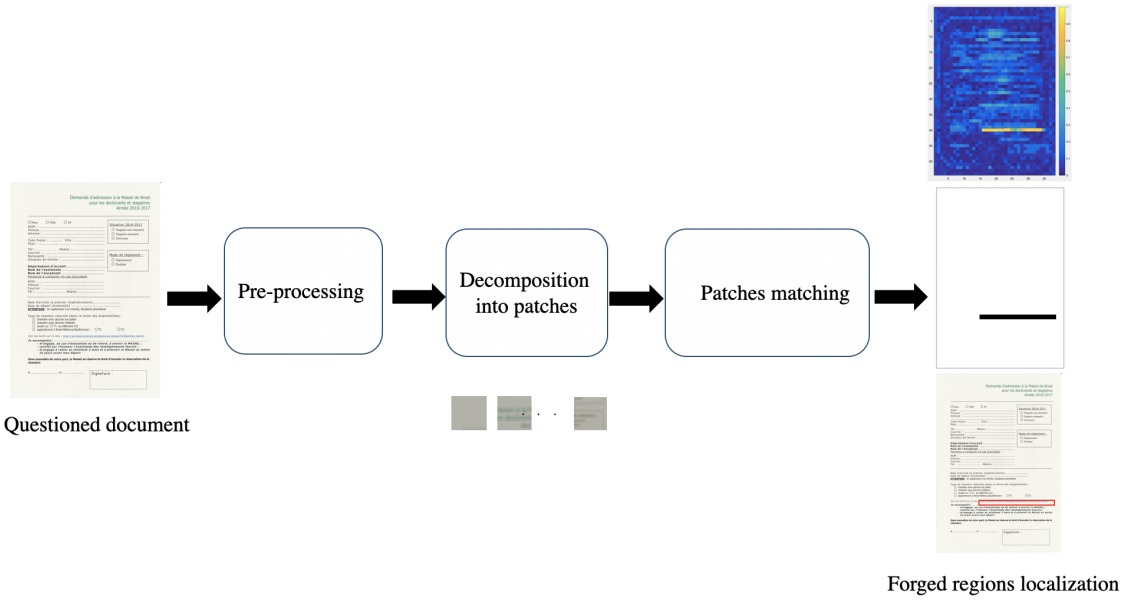


Figure 6.7: The general pipeline of the proposed schemes

The remainder of this section is organized as follows. In Sub-section 6.2.1, we describe our first approach based on handcrafted features matching. The second approach relying on CNN for fingerprints extraction is detailed in Sub-section 6.2.2. Experimental evaluation plan and results are presented in Section 6.3 followed by some discussions.

6.2.1 First method: a Handcrafted-based IFD approach (HIFD)

The proposed detection method is illustrated in Fig. 6.8. and can be summarized as follows:

- Apply SWT transform on the blue channel of each suspected image I to get its HH subband as pre-processing step
- If not available, generate the scanner's signature that we will call the reference signature
- Estimate the histograms of each block of the reference signature and its corresponding block of size 64x64 from the HH subband of the suspected image
- Compare the histograms block by block using the Euclidian distance (ED) and generate a heat map by comparing the ED measures to a threshold

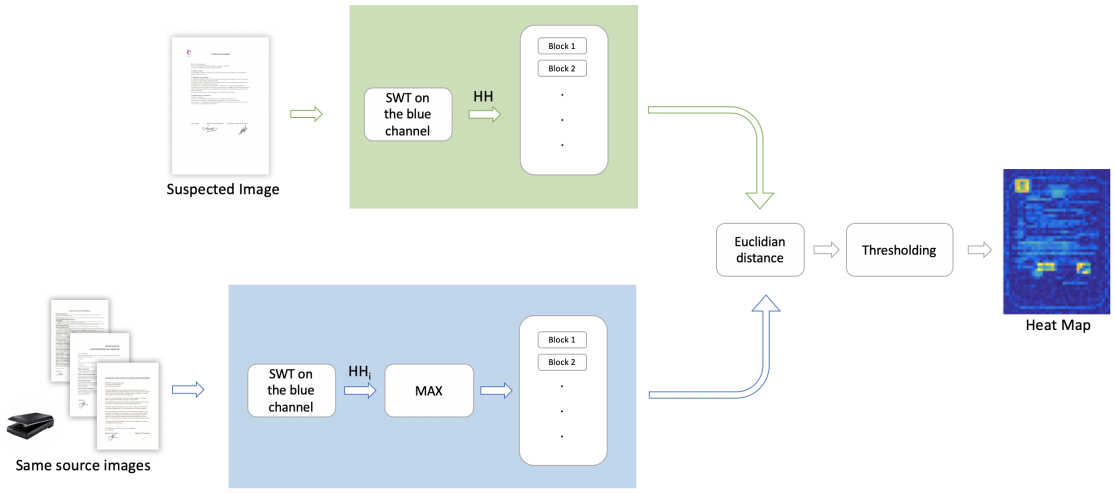


Figure 6.8: Forgery detection process using the proposed approach (i=1..K, K is the number of images acquired by the scanner)

As in our previous chapters, we proposed to use the wavelet transformation as a pre-processing step. The wavelet domain has been shown to be effective to extract relevant scanner noise estimation. In this method, the HH subband obtained from the one level stationary wavelet transform (SWT) of the blue channel of an image is used for IFD. Since, the SWT does not perform down-sampling, the HH subband keeps the same size as the image and is supposed to encode the scanner's noise that is distorted every time the image got manipulated. In fact, tampering traces will be hidden by adding noise around the edges. Therefore, HH subband patches should be affected. Furthermore, the choice of the blue channel for this work is taken after a visual and quantitative comparison of the tampering detection results carried out on each color channel as it will be seen in the following evaluation section.
The core idea behind our proposal is that patches of similar noise are supposed to have identical histograms. Thus, any manipulation of the noise will result in a change of the histogram. This assumption is validated by the pair of histograms

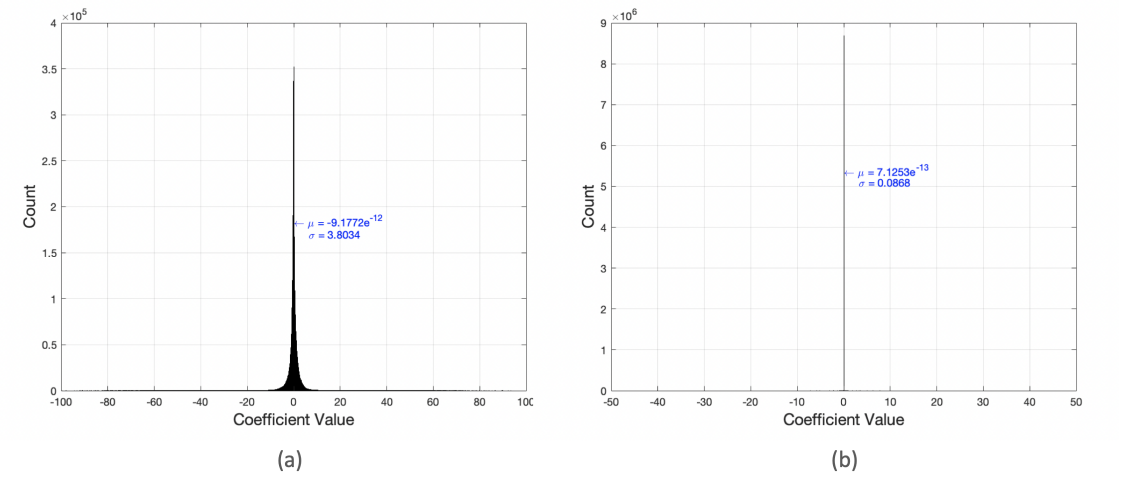


Figure 6.9: An illustration of a pair of histograms (a) before tampering (b) after tampering

given in Fig. 6.9. It shows the histograms produced before and after splicing tampering. We can notice the significant influence of the tampering on the mean and standard deviation of the histograms.

To generate a scanner reference signature, first, the blue channel of each image acquired by this scanner is decomposed into four subbands using a one level 2D-SWT and then the obtained HH subbands are combined by applying a suitable fusion rule. The most well-known rules are:

i) The Simple Average Rule (SAR): It computes the average of the coefficients of the two subbands as follows:

$$\text{SAR}\left(HH_I, HH_S\right) = \text{SAR}(i,j) = \left(HH_I(i,j), HH_S(i,j)\right)/2 \quad , \quad \text{i} = 1..\text{N and j} = 1..\text{M} \tag{6.1}$$

where HH_I and HH_S are the HH subbands of the suspected image and the scanner's signature, respectively. The indices i and j correspond to the i^{th} row and j^{th} column of each element. N is the number of rows and M is the number of columns of HH_I.

ii) The MAX Rule (MR): It compares the coefficients of the two subbands and selects the ones with the largest magnitude between them as follows:

$$MR\left(HH_1, HH_S\right) = MR(i,j) = \max\left(HH_I(i,j), HH_S(i,j)\right) \tag{6.2}$$

In this work, the max rule was selected for this task. The details of the reference

signature generation are given in Algorithm 3.

Algorithm 3: Reference signature generation

Input: Images acquired by the scanner S
Output: Scanner signature HH_S
$HH_S \leftarrow []$
for i *in* S **do**
  $I_b \leftarrow I(:,:,0)$ `// Blue channel`
  $HH_I \leftarrow SWT(I_b)$
  $[k1, k2] \leftarrow size(HH_I)$
  for k *from 1 to* $k1$ **do**
    for j *from 1 to* $k2$ **do**
      $HH_S \leftarrow MAX(HH_S(k,j), HH_I(k,j))$

The ED measurements are used to create a heat map (HM), which can help the forensics investigator to have a visual inspection. HM is a data matrix that uses a sequential colormap to visualize the values in its cells. Large measurements are shown as yellow or orange in the HM announcing that a forgery exists while indicating the positions of the forged areas at the same time. On the other hand, non-tampered regions appear in blue and correspond to small ED. Figure 6.10 shows an example case of a HM with its corresponding colormap. A suspected image is labeled as non-authentic if at least one of the ED measurements is higher than an experimentally pre-defined threshold ν.
To have a better visibility of the IFD accuracy and to remove falsely detected blocks, a binary decision map M is generated such that blocks with normalized ED smaller than are considered as authentic and marked by white blocks and the rest is marked by black blocks (the ones detected as forged). Algorithm 4 gives the step followed to generate M.

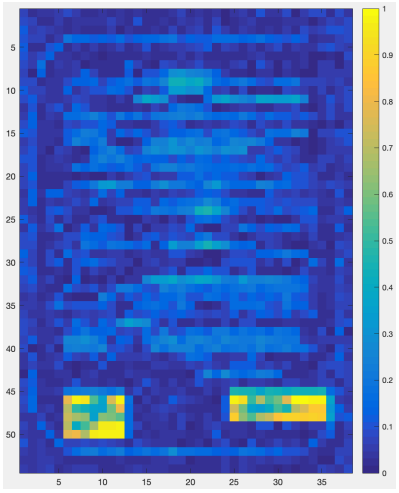


Figure 6.10: A heat map sample

Algorithm 4: Binary mask generation

Input: Normalized euclidian distance N_{ED}
Block size bs
Threshold ν
Output: Binary mask M
$[k1, k2] \leftarrow sizeofthesuspectedimage$
for k *from 1 to* $k1/bs$ **do**
  for j *from 1 to* $k2/bs$ **do**
    if $N_{ED} > \nu$ **then**
      $M((bs*(i-1)+1 : bs*(i-1)+bs), (bs*(j-1)+1 : bs*(j-1)+bs)) \leftarrow 0$
    else
      $M((bs*(i-1)+1 : bs*(i-1)+bs), (bs*(j-1)+1 : bs*(j-1)+bs)) \leftarrow 1$

6.2.2 Second method: a CNN-based IFD approach (CIFD)

The previous section proposes a handcrafted-based IFD method that rely on the dissimilarities between the HH subbands of an image and of its source scanner reference pattern to find the tampered regions. However, the source scanner is not expected to be known in realistic situations. Another limitation of this method is that it may fail if the spliced area is taken from an image acquired by the same scanner model as the suspected image due to the similarity between the distributions of the HH subband coefficients in this case.
In this section, we address the IFD problem the same way it has been tackled in most of the state-of-art methods. More precisely, we propose a new approach that applies a SSI system on image blocks rather than on the full image in order to determine which patches of the suspected image did not originate from the same scanner as the other patches or did not contain the scanning noise of the device that has been used to capture it. In contrast to earlier findings, our approach adopted a data-driven strategy which has been successfully applied in our previous works. As mentioned earlier, we consider the 2D-CNN system proposed in chapter 4.
This is the first CNN-based architecture specifically dedicated to IFD. This decision was made with the goal to achieve high source scanner attribution accuracy with a relatively small network architecture. Our choice was motivated by the fact that our 2D-CNN is efficient in working at patch level facilitating, for instance, tampering localization.
The idea is to divide the suspected image it into non-overlapping blocks and then to find the ones that are predicted to be acquired or generated with a different device. To do so, each block is fed into the 2D-CNN network, after being pre-processed, in order to associate it to one of the known scanners. Next, a majority vote is applied to find out the source scanner of the suspected image. We may consider that the image is forged if the number of blocks identified as acquired by a different scanner is important. The proposed forgery detection scheme is illustrated in Fig. 6.11. Incorrectly identified blocks, marked by bounding blue boxes, indicates predicted forged area.

Once forged blocks are predicted, we create a tampering mask that will be compared to the ground truth for evaluating the reliability of the proposed approach. The mask is a binary matrix of the same size as the suspected image where authentic blocks are presented by white blocks (block values are zeros) and the forged one are in black (block values are ones).

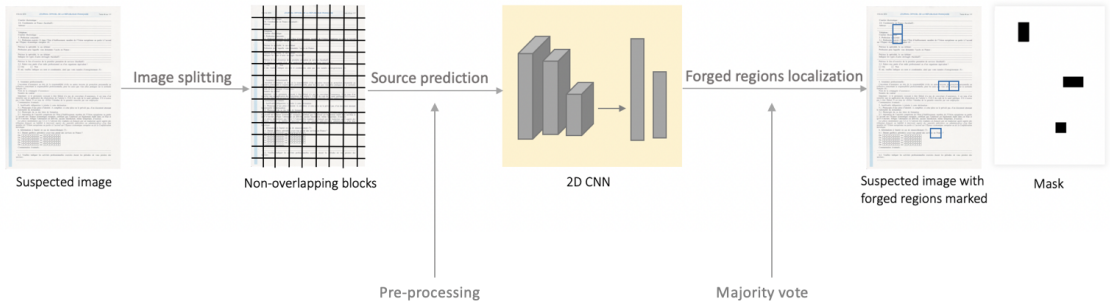


Figure 6.11: Pipeline of the data-driven forgery detection scheme

6.3 Evaluation

To test the performance of the proposed IFD methods, the steps outlined above are performed on images from the SUPATLANTIQUE dataset which provides realistic forgeries.
The presented results have been evaluated visually through masks and heat maps, as well as, numerically using the following metrics:

$$\text{Sensitivity} = \text{true positive}/(\text{true positive} + \text{false negative})$$

$$\text{Specificity} = \text{true negative}/(\text{true negative} + \text{false positive})$$

$$\text{Predictability} = \text{true positive}/(\text{true positive} + \text{false positive})$$

where i) true positive is the number of forged blocks correctly detected as such; ii) true negative is the number of pristine blocks correctly detected as such; iii) false positive is the number of non-forged blocks detected as forged; and, iv) false negative is the number of forged blocks that have not been detected as such.

6.3.1 Evaluation of HIFD

To evaluate the performance of the first approach, it was evaluated on different test images that were forged using splicing operations. It was implemented in Matlab. We selected block size of 64x64 pixels, which gives a total of 2052 blocks per image. The optimal threshold was experimentally determined as ν =0,7 so as to have the lowest false positive rate.
Considering a supervised setting, we generate a signature (See Section 6.2.1) for each scanner using 50 images.
First, to justify the choice of the blue component in our approach, we compared the results we obtained taking as input of pipeline the different color channels independently. Figure 6.12 shows the HH subbands of each color channel of one of our forged test image and their corresponding estimated forgery masks. In the tampered image (e), two signatures has been added at the bottom of the document. We notice the superiority of working on the blue channel in IFD which offers less false predictions.

Testing results demonstrate that the proposed approach can be used for IFD in digital scanned documents. Again, we have treated the HH subband successfully and proved that the noise and textures concentrated in this high frequency subband are intended to be a good discriminator between flatbed scanners. Examples of

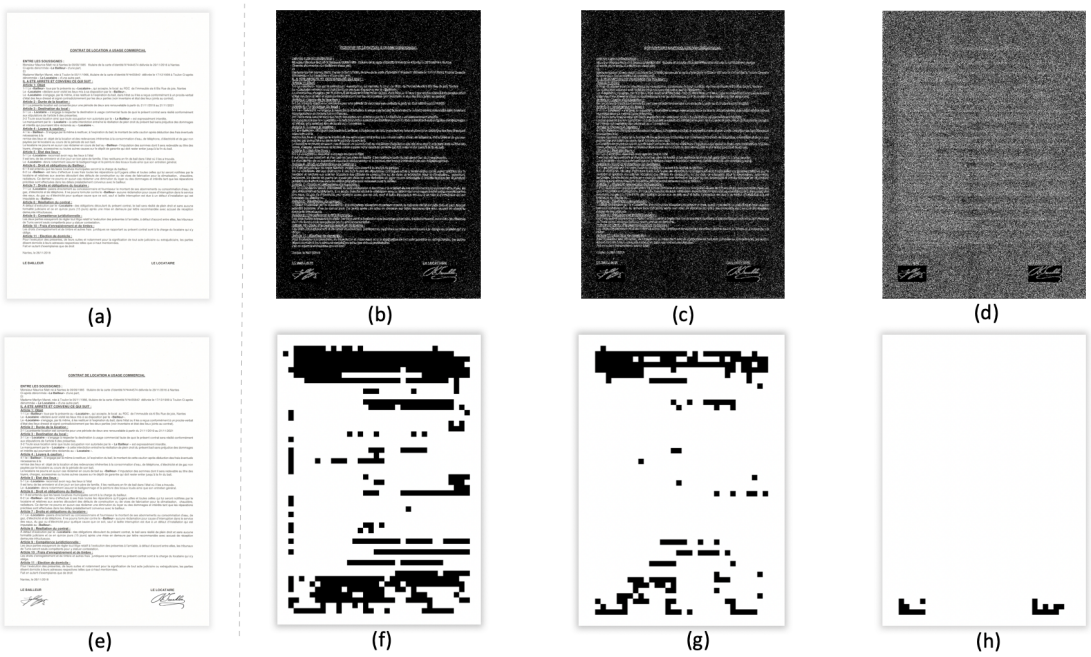


Figure 6.12: Sample HH subbands ((b), (c), (d)) of each color channel of an RGB image and their estimated forgery masks ((f), (g), (h)). From left to right are presented R, G and B, respectively. (a) is the original image and (e) is the tampered one

detection results are presented in Fig. 6.13. We observed that the detection is more reliable for large forged areas and for copied dense regions (such as logos).
In a second experiment, we considered testing the fusion rules (SAR and MR) defined in section 6.2.1 for computing the scanners 'fingerprints. It is clear from the binary masks shown in Fig. 6.13 that both rules are providing a good performance. However, it is worth noting that there is a slight decrease in the number of predicted forged blocks in the case of SAR. Thus, MR is more appropriate for combining the scanning noise from different HH subbands without any loss while catching the most relevant features.
In the last experiment, we intended to test the feasibility of the proposed approach in the case of an unknown source scanner. For that, we perform several tests by using each time a different scanner.
Figure 6.14 gives a comparison between the binary masks resulting from applying our IFD method on the same image but with three different reference signatures. It can be found that the proposed method is efficient independently of the scanner used to generate the reference signature. Therefore, we can conclude that our system is more interested in finding similarities and dissimilarities between blocks of the compared HH subbands than looking for the origin of each block and, thus, can work without an a priori knowledge of the source scanner, that is to say, the source of the suspected image is unknown.

6.3.2 Evaluation of CIFD

In this work, our 2D-CNN model has been re-trained on the three scanner models S1, S9 and S11. A total of 9000x3=27000 image patches of size 128x128 are involved in the training and evaluation phases.
To show the capability of the proposed approach to distinguish between forged and non-forged patches, we performed a set of tests on some manipulated images of size

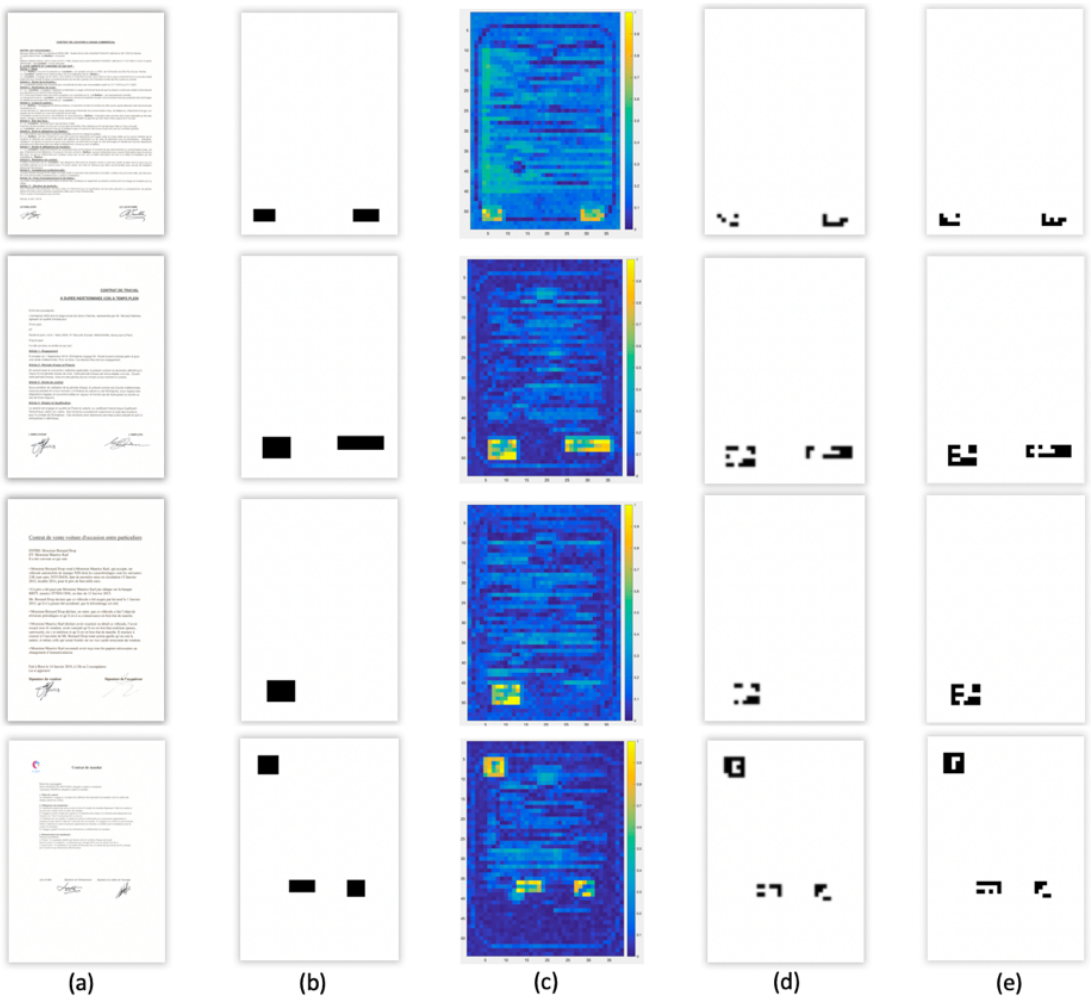


Figure 6.13: Examples of forged images and the outputs of the proposed handcrafted-based FD method (a) Forged images (b) Ground truths (c) Heat maps (d) Masks using the SAR (e) Masks using the MR

2480×3508 in which we fixed the block size to 128x128. Each image is, thus, split into 1026 non-overlapping blocks. Figure 6.15 illustrates the visual results obtained by applying the current method on four forgeries examples. From a subjective perspective, it can be seen that our method is effective in the detection of forged areas.
We observe some falsely detected blocks in the second image. Such a false positive detection may be due to different reasons: low scanning noise, saturated regions...

As far as tampering localization accuracy is concerned, we decided to examine the effect of reducing the testing block size on the proposed method. Therefore, the CNN was again trained over 64x64 blocks. A comparison between the previous result and the current ones is presented in Fig. 6.16. Obviously, we can observe that the tampered regions are located with more precision at the cost of the appearance of a number of falsely predicted blocks.

To further validate the performance of our approach, we applied three different forgeries on an image to create three manipulated images using a different editing tool than the one used to produce the tampered images of the SUPATLANTIQUE dataset. The example is illustrated in Fig. 6.17. The left image represents the original image. In the middle, each case of forgery is given such that:

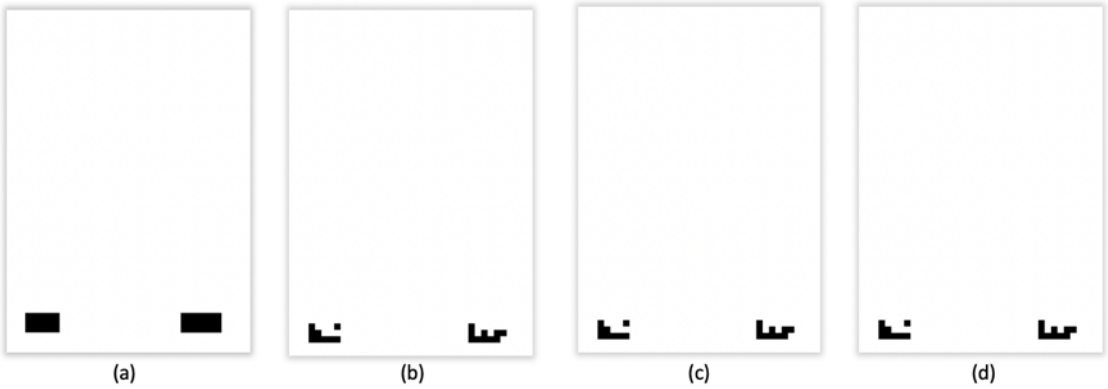


Figure 6.14: Comparison between the binary masks obtained by using different reference signatures in our method (a) Ground truth (b) Mask obtained using the first scanner (source scanner) to generate the reference signature (b) Mask obtained using the second scanner to generate the reference signature (b) Mask obtained using the third scanner to generate the reference signature

Table 6.2: Sensitivity, Specificity, and Predictability of the proposed models

	Sensitivity	**Specificity**	**Predictability**
HIFD	63.81%	**100%**	**100%**
CIFD (64)	**83.39%**	98.55%	76.87%
CIFD (128)	39.43%	99.64%	92.39%

- In the first image (a), a blank area is copied and moved within the same image. It has been manipulated to cover up the desired area
- In the second image (b), a block from another image scanned with a different scanner has been paste
- In the third image (c), a block from another image captured with a camera has been pasted

One can immediately observe that altered areas, marked with blue rectangles, are detected with a relatively high precision.

To numerically evaluate the ability of our proposed model to distinguish between pristine and forged blocks, we used the true positive, false positive, true negative and false negative rates to derive the sensitivity, specificity and predictability of each model in forgery detection. Table 6.2 shows results in terms of those metrics for both block sizes stated previously that we compare to the first method. As for predictability and specificity, we surprisingly found that the HIFD indicated 100% for all the investigated images. However, the sensitivity of the CIFD with 64x64 blocks surpasses the HIFD and the CIFD with 128x128 blocks with 63,81% and 39,43%, respectively. Notice also that the CIFD method with block size 128x128 has the lowest sensitivity due to high number of non-predicted forged regions compared to the predicted ones.

Figure 6.15: Examples of forged images and the outputs of the proposed CNN-based FD method (a) Forged images (b) Ground truths (c) Forged blocks marked by blue boxes (d) Estimated binary masks

6.3.3 Discussion

We proposed two blind splicing scanned document forgery detection methods with SWT as a pre-processing procedure. We utilized the HH subbands to find dissimilarities between the image regions. The first method was based on handcrafted features while the second one used CNN for scanner fingerprints extraction. We evaluated them on forged images from the SUPATLANTIQUE dataset. The efficacy of the proposed methods was demonstrated visually as well as quantitatively.
Both techniques succeed to detect areas in a suspected image resulting from replacing an object with another one downloaded from the Internet or copied from an image

acquired by a different scanner (splicing). Copy-moved objects can also be detected if they have received some geometrical transformations (rotation, resizing...). However, the advantage of using the CNN-based approach is its adaptability to automatically extract characteristics and its ability to detect manipulated patches when the copied patch was acquired by a scanner of the same model as the source of the suspected image. Moreover, given the challenges in training CNNs with a small labeled dataset, our network, trained on few samples from three scanners only, has been successful in addressing the IFD problem. On its side, the handcrafted-based approach provides good result on small size patches. More interesting, these approach are able to detect forgeries even when tested on images acquired by unknown scanners.
At last, one should also notice that these approaches face difficulties to detect relatively small forgeries and tampering in regions for which the scanning noise is almost absent such as saturated or heavily textured areas. These areas are at the origin of false alarms as previously shown in Fig. 3.15.

6.4 Conclusion

In this chapter, we first presented the SUPATLANTIQUE database, the only document image collection available in public domain which contains scanned documents suitable for evaluating the performance of SSI and FD techniques. It combines official and Wikipedia documents, flatfield images as well as forged digitized images.
Next, in order to achieve the second specific objective of this thesis, we propose to take benefit of the SSI models presented in the previous chapters to develop two new transform-based approaches for IFD. The parameters characterizing the scanning noise are extracted from the HH subband and are investigated to detect traces of tampering. The proposed methods are more relevant to localize forgeries in images acquired using flatbed scanners than other existing methods in the literature. We evaluated the performance of the proposed schemes using the SUPATLANTIQUE database. Experimental results clearly prove the significantly high accuracy of these schemes to discern pristine areas from forged ones, though they may occasionally result in false positive.
Furthermore, in the hand-crafted approach, ED is used to evaluate the dissimilarity between the histograms of the HH subbands blocks. However, despite its effectiveness, it is worth testing more distance measures like the Bhattacharyya distance [222], the Matusita distance [223] and the K-L metric [224].
One limitation of our methods is that they cannot detect all kinds of forgeries. Thus, a more solid study for other types of forgery is highly desirable as future research. Besides, it would be interesting to use our DLK solution proposed in the previous chapter for the purpose of IFD. This has not been completed due to time constraints.

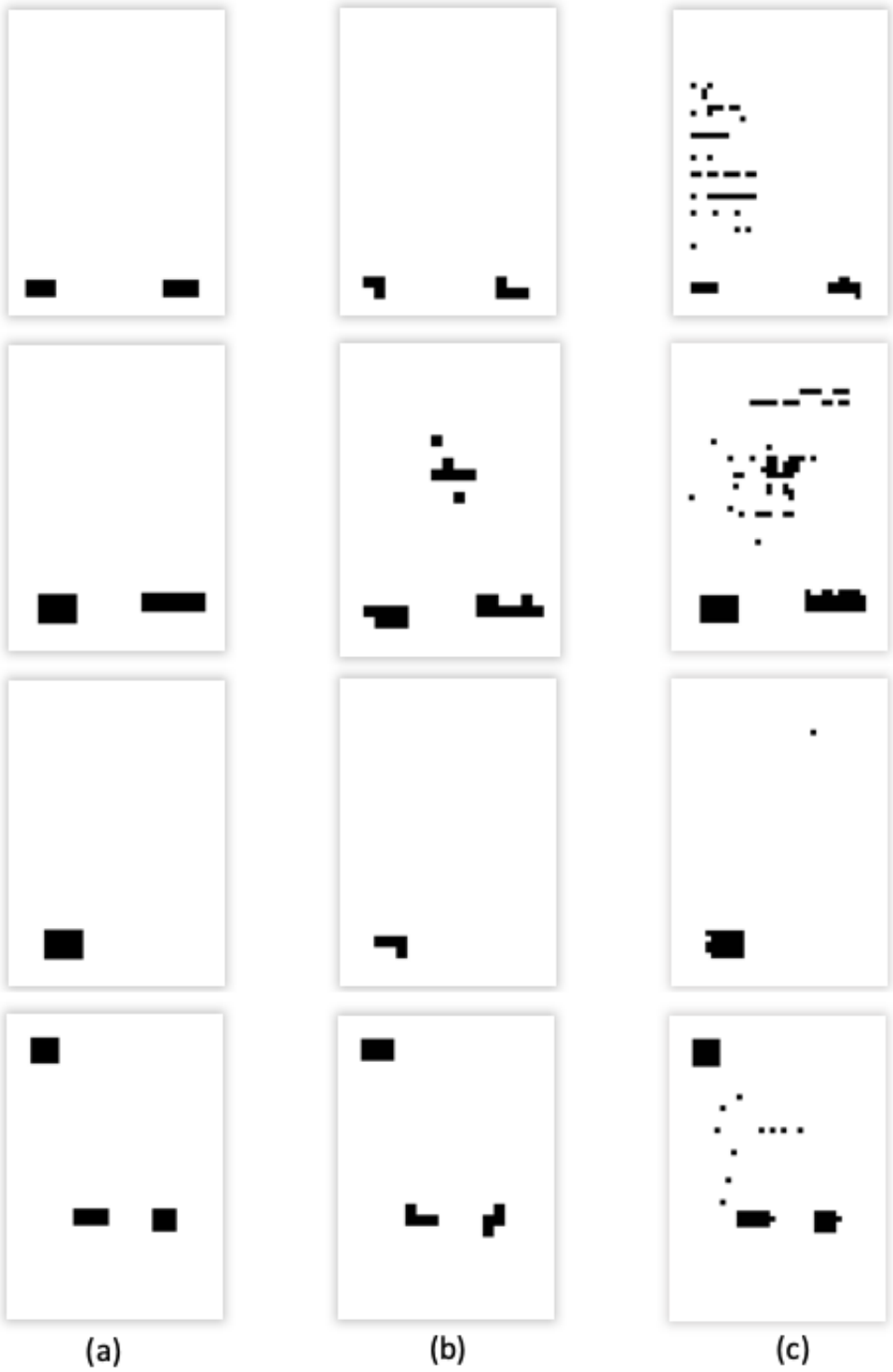


Figure 6.16: Comparison between forgeries detection performances when using different size patches (a) Ground truth (b) Mask in the case of 128x128 patches (c) Mask in the case of 64x64 patches

Figure 6.17: Test case. (a) Original. ((b), (c), and (d)) Tampered. ((e), (f), and (g)) Tampered regions detected

A conclusion is simply the place where you got tired of thinking

Une conclusion est simplement l'endroit où vous en avez assez de penser.

— Dan *Chaon*

7 CONCLUSION AND PERSPECTIVES

Contents

7.1 Conclusion

This manuscript presents a number of digital image forensics methods which take advantage of the fact that every acquisition system leaves residual noises called fingerprint on every image it acquires. As demonstrated, each scanner's fingerprint is unique and the likelihood two scanners have similar fingerprints is relatively low. Thus, this fingerprint has become an asset for several forensic tasks, including device identification, device linking, and forgery detection.
From that perspective, the thesis' contributions can be summarized as follows.

Proposing new source scanner identification approaches: The different works conducted in this thesis were first dedicated to solve the SSI problem. The main goal was to design an adaptive model for estimating the scanner that has acquired a suspect digital document where all scanners are supposed to be known. The idea was to find a unique pattern that can be used as a scanner fingerprint. An analysis of the state-of-art shows that existing models mainly focus on a particular image type or structure and, thus, present unsatisfactory results in realistic cases. Besides, there was a limited exploitation of the scanning noise compared with the methods related to other acquisition devices. Moreover, most of the benchmark datasets employed by those works do not take into account the fact that a scanner model may have more than one instance. It is also important to mention that scanner manufacturers are continuously improving the resolution of their devices, which require looking for more robust and flexible solutions. Therefore, we considered all these challenges in the aim to establish new SSI frameworks based on more accurate scanning noise models. For that, we investigated the problem in many directions, depending on the image format, the image content and the techniques

to be used for features computation. Approaches based on handcrafted features are firstly exploited in Chapter 3. In a first time, we were interested in the scanned images stored in the JPEG format. We established a solution that, based on the quantization tables contained in the JPEG header, is able to identify the brand of scanner that has acquired the investigated image. However, this technique is limited to brand identification and is not credible since the header can be easily replaced or deleted. Furthermore, the literature review conducted has shown that all the state-of-art methods are commonly using the TIFF format to evaluate the performances of their methods. In fact, this format is the most ubiquitous image format used by scanners since it is better suited for producing high quality images especially in the medical field [225]. Thus, we proposed an adaptive approach [14] for scanner model identification with fingerprints extraction in the wavelet domain based on the GGD modeling of subband coefficients. This approach achieves good performance compared to existing methods. An improved forensic performance is achieved by taking advantage of the a priori knowledge of the reference image, the original noise-free image, in the case when the scanned document is fillable [15]. This is implemented by subtracting the reference image from the filled scanned one as a way of denoising. Indeed, the previously wavelet-based proposed method [14] is able to extract good scanners fingerprints from HH subbands. However, these subbands are often preserving image details leading to identification errors. Two main limitations of these approaches are the decline of performance in case scanners of the same model are considered and the dependence to the type and the number of the training images. To overcome these limitations, two novel approaches based on neural networks (NN) are proposed in Chapter 4. Their originality is two-fold. First, they perform an automatic features extraction from small patches of the scanned documents. Second, they exploit the scanning noise in two different dimensions (1D [17] and 2D [18]). These methods have been demonstrated to be effective for large number of scanners even when two instances of the same model are considered for evaluation. Their main highlight is that they can be trained from scratch using a few number of images. It is also important to mention that CNN based systems for SSI are proposed for the first time in the DIF field.

Image collection and its application to DLK: To date, all SSI algorithms have taken into account the form of the closed set scenario, where the scanners that may have issued the investigated images are all available at the training stage. A more realistic scenario for forensics applications is the open set that takes into consideration the possibility of an unknown source of image, i.e. a scanner non-identified a priori. The purpose of the following study was to propose a new research line in which we consider two images of unknown sources aiming at finding out if they have been acquired by the same scanner or not. Such a study is known as the device linking (DLK) and can be generalized to solve the open-set recognition problem. The proposed architecture is composed essentially of two NN; The first network is a fingerprints extractor which maps an image patch to a vector of features so as to feed the second network that takes as input a pair of the produced feature vectors and returns a similarity score indicating whether the patches were issued by the same device or not. Experiments conducted on image pairs issued from 8 different scanners, where two of them were not used to train the linking network or equivalently considered as unknown sources, confirm the efficiency of our system. To the best of our knowledge, we are the first to conduct scanned images forensics DLK. We expect this new research direction would be useful in other DIF problems such as database consistency verification and images manipulation detection.

Designing IFD approaches: The development of digital imaging has raised a number of challenges in terms of information security. Because of the wide availability of low-cost image editing tools, trustworthiness of digital images can be more and more questioned. Examples of image tampering for various purposes are

numerous. The field of DIF has been established to rebuild trust in media content. Though IFD through forensic techniques is the second main research topic of this thesis. Herein, we proposed two optimal detectors. Both integrate patch-based algorithms and are able to precisely detect falsifications in the scanned image. The basic idea of the first one is to detect image splicing using SWT and histograms of HH subbands patches based on the fact that a simple tampering operation disrupts the scanning noise contained in these subbands. A more promising research direction revisits IFD with a data-driven paradigm. Thus, as a second step, we designed a method leveraging on our prior 2D-CNN model proposed in Chapter 4 for the identification of the source of each patch of the investigated image. The rationale is to distinguish between homogeneous and heterogeneous image patches in terms of scanner noise. More clearly, heterogeneous patches notify the presence of forgeries in it. To evaluate the performance of our models, we introduced a new public image collection for digital forensics purposes related to flatbed scanners. We demonstrated the effectiveness of our approaches for splicing detection. Given examples validate the accuracy of the forgeries localization. Although the problem of IFD has received considerable attention, the efficacy of published device based prior-art approaches (given in Table 2.2) strongly depends upon the size of image patches. In contrast, the models we propose rely on a smaller patch size which is a very important parameter for forgery detection precision.

To sum up, the above works aim at contributing to the field of DIF for flatbed scanners. Their main novelty lies on several aspects: i) they deal with realistic situations, ii) they allow the development of advanced methods that provide an accurate estimation of the forensic trace of scanners, iii) they model the scanning noise in different dimensions (1D, 2D) and features spaces (spatial, wavelet), iv) they are at the origin of NN based data-driven solutions, the firsts that deal with scanned images, v) they are the first addressing the problem of DLK related to scanners, vi) and, most importantly, they are adapted to any type of images including scanned hard-copy documents which present a major challenge for state-of-art methods related to both, scanners and digital cameras, because of their pseudo binary nature and the presence of large white (saturated) regions.

7.2 Perspectives

In addition, all along this thesis, we spotlight a number of aspects of important interest for future research in the field of DIF. Our perspectives follow three major axis: scanned documents authentication, digitized documents integrity, and finally anti-forensics and counter anti-forensics.
Image authentication is one of the most investigated fields of DIF where the origin of the suspected image is verified in a blind way. Even if this is a field has reached a certain degree of maturity, we demonstrated that there are still some intriguing issues to be addressed. In this thesis, we have focused on non-compressed image authentication with a less successful attempt to identify the source of JPEG compressed image from its header. Thus, there is a great interest to conduct more studies on noise-dependent fingerprints to be extracted from the JPEG images. Furthermore, previous works analyzing scanners artifacts produced during the documents acquisition have focused on the sensor defects. However, other scanner components may leave their unique trace such as optical defects that need to be considered. In addition, some new scanners can eliminate defective pixels using on-board post-processing, leading to harder fingerprint extraction. Thus, there is a motivation to provide solutions to assess these situations. Moreover, a specific limitation of SSI approaches is the knowledge of only a few scanners for training, whereas in practical situations, the document may be digitized by an unknown scanner. It would be interesting to design one system that can classify an image to

one of the known devices and tell if its source is just unknown. For this, we can make use of the DLK model we proposed in Chapter 5 with a possible improvement of this later. Another crucial issue with noise-based approaches is that the noise pattern is affected by the desynchronization produced by post-processing operations like compression and geometrical transformations (e.g. rotation, scaling...). This motivates us to consider using cropping and rescaling techniques in our future work to solve this problem.

To address the difficulties faced by Forensics investigators in tampering detection, we proposed different block-based solutions that surpasses the existing ones. However, despite the reported promising results, the performance of those techniques is far from being trustworthy and there is still a lot of room for improvement. In fact, most of the existing works are not effective for forgeries other than splicing and when dealing with real-world documents. In addition, it is important that further research studies take into consideration the case when many different tampering operations were applied to the image. More clearly, there is a need to develop a universal manipulation detection method which is capable of detecting multiple manipulations in the same scanned document. We think that this objective can be achieved by combining passive with\or actives forensics state-of-art approaches, or by relying on new techniques based on data mining and NN. For example, a combination of image forensics and cryptography might link a scanned document to its digital fingerprint with a higher level of assurance. It is also worth noting that the methods proposed in literature present a certain complexity and require human intervention. If this issue has been solved for SSI, it will be more practical to investigate creative solutions for IFD following the same research line. This is particularly important to reduce the computational time and to improve the accuracy of manipulation detection methods. To go further, it is also important to say that there is an interest to find a solution which can not only detect forgeries but also tell the type of manipulation. For example, one can decide if the manipulation is a copy-move operation by checking if the suspected object or block is repeated in a different area of the image. Addressing these challenges will open the door to more effective contributions.

The two sides of the coin in image authentication are forensics and anti-forensics. This is the same as we talk about cryptography and cryptanalysis or steganography and steganalysis. Digital image anti-forensics [226] is a methodology for exposing the limitations of existing forensic methods to build more reliable forensics. The fundamental goal of image anti-forensics is to execute certain operations on digital images to mask traces left by image manipulations so that forensic algorithms cannot detect them. Notice that anti-forensics techniques aim at helping forensics researchers to know the vulnerabilities of the approaches they propose and, thus, prompt them to reinforce them against any possible attacks by introducing the so-called "counter anti-forensics" methods. During this period, we had not time to address such an issue. Anyway, it must know that image anti-forensics and counter anti-forensics are still in their infancy. Based on a literature review, it is clear that the number of publications related to those research fields are far less in number than those related to image forensics. Furthermore, existing anti-forensic and counter anti-forensics approaches bore some drawbacks. Therefore, there are still numerous research lines in this field that should be pursued. To this end, some advanced methods could be introduced by combining anti-forensic strategies with the methods proposed in this thesis.

Appendices

A
RÉSUMÉ DE LA THÉSE

A.1 Introduction

L'évolution des technologies de l'information et des communications sont à l'origine de changements sociétaux majeurs avec notamment la dématérialisation de nombreux services et de documents administratifs. Cette transition vers un monde entièrement numérique implique la numérisation de documents : ceux qui étaient traités sous forme papier avant dématérialisation, comme les dossiers médicaux d'un établissement de santé ou encore les documents pour lesquels la réglementation impose le passage par une version papier et une numérisation une fois ceux-ci complétés tel que les actes notariés, d'état civils ou bancaires, etc. Une fois scannés, ces documents peuvent alors être aisément distribués, échangés, collectés et traités, accélérant les procédures tout en réduisant les coûts. Plusieurs pays ont même procédé à l'archivage de leur documents commerciaux, administratifs et politiques afin de préserver leur patrimoine et restaurer les documents d'importances tel que le cas de la France qui a créé une archive diplomatique qui compte près de 180 000 documents [1].
Cependant, si cette évolution facilite l'accès et le traitement de l'information, elle soulève beaucoup questions en matière de sécurité. Par exemple, on pourra s'interroger sur l'origine d'un document numérisé comme également de son intégrité. Ces aspects sont critiques dès lors que les documents sous forme papier n'inclus pas d'éléments de sécurité (filigranes, hologramme etc.) comme il en existe pour les billets de banques ou les passeports. En effet, un tiers malveillant peut aujourd'hui assez facilement falsifier le contenu d'un document scanné à son avantage à l'aide d'un logiciel d'édition d'images sans laisser des traces visuellement détectables.
Différentes solutions permettent de protéger les données multimédia. Ces solutions peuvent être classées en deux catégories: actives et passives (voir Fig. A.1).

Les techniques actives nécessitent généralement une étape de prétraitements. Il s'agit de mécanismes cryptographiques (ex. chiffrement [6], signatures numériques [6]), de tatouage [7] ou de crypto-tatouage [8]. Le chiffrement de données et les signatures numériques peuvent empêcher la falsification de données, mais elles contraignent, au même temps, le format de données et soulèvent des questions de partage des clés.
Le chiffrement et les signatures peuvent être vus comme une protection a priori, l'information étant protégée tant qu'elle est chiffrée ou que sa signature numérique n'a pas été supprimée. A contrario, le tatouage offre une protection a posteriori. Il

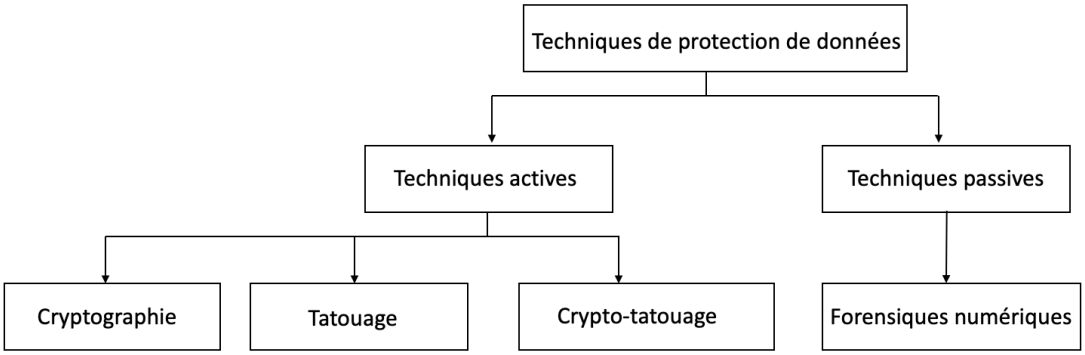


Figure A.1: Classification of digital data protection approaches

permet d'insérer dans le document scanné, par modification imperceptible de ce dernier, une marque ou un message permettant de savoir son origine et s'il n'a pas été falsifié. Cependant, cette protection a posteriori implique de pouvoir tatouer un document dès sa numérisation, et donc impose qu'un module de marquage soit intégré dans tous les scanners ce qui n'est pratiquement pas faisable vu le coût de fabrication et la complexité de conception. De plus, cette opération affectera la qualité visuelle de l'image numérisée.
Une alternative intéressante est offerte par les techniques passives dites « forensiques des images numériques » [10]. Ces techniques visent à vérifier si une image a été modifiée ou à apporter la preuve de son origine (i.e. identification du scanner qui a acquis l'image) d'une manière aveugle, sans information a priori sur l'image (i.e. sans signature ou information ancillaire partagée avec l'image). Elles s'appuient sur l'extraction de caractéristiques de l'image couplées ou non avec des processus d'apprentissage pour constituer une empreinte de l'image qui peut servir soit à identifier son origine soit à détecter des modifications.
Au cours des 15 dernières années, le nombre de travaux dans le domaine du « forensiques des images numériques » liés aux images numérisées a considérablement augmenté. Cependant, l'efficacité de ces approches proposées reste relative. Ainsi, il est important de fournir des solutions adéquates et fiables pour répondre à des problématiques de protection des contenus multimédias dans la transition numérique et la dématérialisation des données et des services.
L'objectif de cette thèse est donc de proposer un environnement sécurisé et adapté pour la vérification de l'authenticité et la lutte contre la falsification de documents numérisés d'importance ; un environnement flexible proposant des services de transactions de confiance à des personnes ou des entités morales.
Le reste de ce résumé est organisé de la façon suivante:
La section A.2 présente quelques notions sur le forensique des images numériques et les scanners, et continue avec une introduction des solutions forensiques proposées dans la littérature qui répondent aux questions liées à l'authenticité et l'intégrité des images numériques.
La section A.3 propose une approche qui identifie la marque du scanner ayant numérisé le document en question quand celui-ci est sauvegardé au format JPEG. Vu que la majorité des documents scannés sont au format TIFF, deux nouvelles approches sont ainsi proposées prenant en compte ce format. Cependant, ces méthodes ne sont pas efficaces pour des scanners de même modèle et présentent certaines autres limitations. De ce fait, elles seront encore améliorées dans la section A.4.
La section A.4 propose deux approches d'identification du scanner source, en se basant sur les résultats trouvés précédemment. Ces approches prennent avantage des réseaux de neurones pour extraire les caractéristiques des scanners. Outre

l'excellent taux de précision atteint, les réseaux proposés ne nécessitent pas un nombre important d'images pour leur apprentissage.
La section A.5 résout le problème de la vérification de la correspondance des sources d'une paire d'images. En d'autres termes, l'objectif de la solution proposée est d'apporter preuve que deux images sont acquises par le même scanner ou par des scanners différents. Cette information est très utile pour un expert en forensique comme elle peut aussi servir à adresser la problématique liée aux documents acquis par des scanners 'inconnus'.
La section A.6 présente une nouvelle collection de documents numérisés réalisée dans le cadre de la thèse et son exploitation pour vérifier les performances de deux nouvelles approches de détection de falsification. Nous avons proposé ces approches en s'appuyant sur les approches que nous avons précédemment proposées pour l'identification du scanner source.
La section A.7 conclut ce manuscrit en présentant brièvement les contributions et les perspectives de recherche.

A.2 Forensique des images numériques

Les scanners servent de lien entre les mondes physique et numérique, facilitant et accélérant ainsi le stockage et le partage de documents. En effet, la numérisation des documents réduit les dépenses, économise de l'espace et réduit le risque de perdre des documents critiques. Les entreprises, les banques et de nombreuses autres organisations utilisent la numérisation pour économiser des centaines de milliers de dollars en impression, expédition et stockage de documents chaque année. L'accès à l'information, en revanche, est facilité, ce qui pose des problèmes de sécurité. Il est évident que notre confiance dans les images s'est dégradée même lorsqu'elles sont publiées dans des journaux populaires et des sites Web officiels, car tout le monde peut désormais acquérir, stocker et modifier des images numériques en raison de la disponibilité de divers outils d'édition d'images qui permettent la manipulation du contenu d'une manière simple et imperceptible. Ces fausses images peuvent parfois causer de sérieuses pertes financières, ou même, être à l'origine d'une crise au sein de la société. D'où la nécessité d'établir de nouvelles techniques qui permettent de retracer l'historique de ces images afin de vérifier leur authenticité et leur intégrité. Dans la littérature, il existe plusieurs approches en forensiques qui permettent d'identifier la source d'une image numérique. L'idée derrière ces méthodes est basée sur le fait que le processus d'acquisition d'un document laisse certaines traces dans l'image crée qui peuvent, si correctement exploitées, servir comme signature pour le scanner. Dans son principe un scanner convertit un document, initialement présent sous forme papier, en sa version numérique. C'est vrai que la configuration d'un scanner à plat dépend du fabricant et du modèle mais ses composants sont généralement les mêmes. La Fig. A.2 illustre les composants de base constituants un scanner. Le document, une fois placé sur la vitre, est balayé ligne par ligne par une unité de scan. Il est à noter qu'il existe deux types de technologies utilisés par les scanners: CCD ou CIS.

Il est possible de grouper les méthodes forensiques d'identification du scanner source en cinq catégories:

- Approches qui extraient les caractéristiques du scanner dans le domaine spatial [16, 24, 27–30]
- Approches qui extraient les caractéristiques du scanner dans le domaine fréquentiel [32]

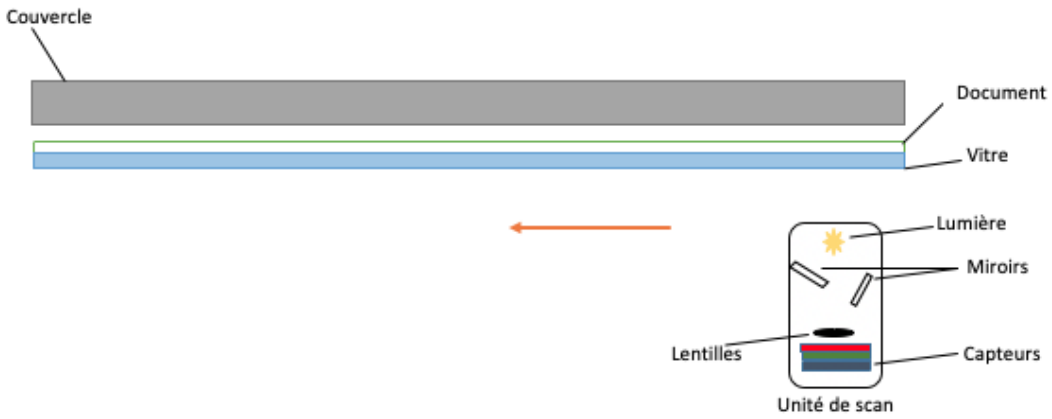


Figure A.2: Processus d'acquisition d'un document numérique par un scanner à plat

- Approches qui extraient les caractéristiques du scanner à partir des holo-grammes [33]
- Approches qui extraient les caractéristiques du scanner à partir des textures [34–36]
- Approches qui extraient les caractéristiques du scanner depuis les traces fixes des fissures et de la poussière situées sur le vitre du scanner [38, 39]

Même si ces méthodes offrent de bonnes performances pour certains types d'images, elles restent insuffisantes. Tout d'abord, elles ne réussissent pas à distinguer les scanners de même modèle. Deuxièmement, elles ne répondent pas aux conditions réelles de numérisation et ne sont pas capable de fonctionner quel que soit le contenu du document numérisé.
D'autre part, le nombre de documents falsifiés est entrain de croître. Aujourd'hui encore plus qu'hier, ils sont utilisés sur les réseaux sociaux dans le but de partager des fausses informations qui peuvent, dans certains cas, alimenter des débats et des opinions politiques ayant un effet néfaste sur la stabilité de tout un pays. Il s'agit donc de faire face à ces fraudes et de proposer des méthodes qui, dans un premier temps, vérifie si un document a été falsifié ou non, et par la suite, localise les régions falsifiées une fois une falsification a été confirmée.
Au cours de la dernière décennie, plusieurs approches de détection de falsification ont été proposées mais peu de celles-ci qui prennent en considération les spécificités des scanners et sont applicables à tout type de document. Par ailleurs, on distingue principalement quatre types de falsification:

- Le «splicing»: C'est une méthode qui consiste simplement à couper et coller certaines parties d'une ou plusieurs images sur une autre image appelée image composite
- Le copier-coller: La falsification par cette technique consiste à copier une partie de l'image et la coller dans un autre emplacement dans la même image
- Le ré-échantillonnage: Généralement, la falsification est réalisée en copiant-collant une partie d'une image sur une autre image. Pour cela, il est souvent nécessaire de redimensionner la partie collée de l'image pour l'adapter à la grille d'échantillonnage de l'image hôte. Cette opération est ainsi appelée ré-échantillonnage

- Les retouches: La falsification par cette technique consiste à améliorer le contenu de l'image en appliquant plusieurs transformations globales ou locales sur cette image à l'aide d'un ou plusieurs filtres ou en manipulant une partie de cette image pour la masquer ou lui ajouter des informations

Dans cette thèse, nous nous intéressons aux approches basées sur les propriétés des scanners. Les méthodes de l'état de l'art relatives à ce type d'approches sont listées dans la table A.1. Dans cette table, les quatre approches traitent le même type de falsification et sont limitées par la taille du bloc de traitement.

A.3 Identification du scanner source basée sur une extraction manuelle des caractéristiques

L'objectif de cette partie est de proposer de nouvelles solutions permettant l'identification du scanner à l'origine d'un document en question. En premier lieu, nous nous sommes intéressés aux images sauvegardées au format JPEG. Ce format a des spécificités liées à son entête que nous avons exploitées afin de cerner le fabricant du scanner recherché.
Pour les images TIFF, nous avons proposé une méthode qui exploite les propriétés de la gaussienne généralisée caractérisant la distribution des coefficients d'ondelettes de sous-bande HH issue de la décomposition en ondelettes de l'image numérique. Cette solution s'est montrée plus fiable par rapport aux approches de l'état de l'art mais ses performances décroissent en présence d'un nombre important de scanners. Cette chute de performance peut s'expliquer par la présence dans notre jeu de scanners, de scanners de même marque et de même modèle. Ces scanners, ayant un processus de fonctionnement identique, présentent très peu de différences ce qui fait qu'ils sont difficiles à discriminer. Pour faire face à ce problème, nous nous sommes focalisés sur les documents officiels. En effet, ces derniers ont généralement la même structure et sont juste remplis différemment avant d'être numérisés. La méthode que nous proposons soustrait le document original (vierge) du document scanné afin de supprimer le maximum du contenu du document empêtrant le bruit du scanner.

A.4 Identification du scanner source basée sur une extraction automatique des caractéristiques

Les travaux de cette partie visent essentiellement à trouver les caractéristiques qui permettront de mieux distinguer les scanners, en particulier ceux de même modèle. L'ambition est aussi d'exploiter des pistes permettant d'automatiser le processus d'identification du scanner à l'origine d'un document qui est quasi-exclusivement réalisée manuellement. Pour cela, nous avons proposé deux approches basées sur le CNN, réseau de neurones convolutif, qui différent principalement dans leurs dimensions (2D et 1D). Les résultats issus de ces approches sont extrêmement favorables.
Dans la première approche, une architecture CNN bidimensionnelle (2D-CNN) est proposée pour classer les images selon leurs sources. Afin d'assurer une meilleure détection des caractéristiques du scanner, la transformée en ondelettes stationnaires est appliquée pour donner accès à sa sous-bande HH qui servira d'entrée à notre réseau et donc de réduire l'effet du contenu de l'image lors de l'apprentissage du réseau. La deuxième approche est plus sophistiquée. Elle est basée sur le fait que le balayage du document à scanner est réalisé ligne par ligne et, ainsi, son bruit est répété sur toutes les lignes du document numérisé. Par conséquent, un CNN unidimensionnel (1D-CNN) combiné à une machine à vecteurs de support (SVM) est adopté pour effectuer un apprentissage automatique du bruit de balayage lié à chaque scanner.

Table A.1: Comparaison des techniques de falsification basées sur les propriétés des scanners (NN correspond à Non-Numérique et SC à Sans-Chevauchement)

Publication	Caractéristiques utilisées	Taille du bloc	Type des falsifications	Nombre des scanners	Nombre des images	Résolution du scan	Résultats rapportés	Avantages	Limitations
Khanna *et al.* [28]	Caractéristiques statistiques du bruit des capteurs	384x512 (SC)	Splicing	5	125	1200dpi	NN	Capable de détecter la falsification de type splicing dans les photos	-Coûteuse en termes de calcul -Les documents ne sont pas étudiés -Fausse classification dans les régions fortement texturées ou saturées -Impossible d'utiliser des blocs de pixels plus petits
Choi *et al.* [141]	Bruit spectral	256x256 (SC)	Splicing	4	3	300dpi	NN	Capable de détecter les falsifications de type splicing dans des photos où les blocs falsifiés sont prises de photos non scannées	Les documents et les régions falsifiées numérisées ne sont pas étudiés
Elsharkawy *et al.* [142]	Les positions des poussiére et des fissures	362x408 (SC)	Splicing	3	7	300dpi	NN	Capable de détecter la falsification de type splicing dans les photos	-Les documents ne sont pas étudiés -Impossible d'utiliser des blocs plus petits
Elsharkawy *et al.* [143]	Histogramme de la composante de luminance	Taille de l'image	Splicing	10	100	-	97.25%	-Détecte la falsification indépendamment de la source de limage -Haute précision de détection -Testée avec des images compressées JPEG et en présence de bruit	Les documents ne sont pas étudiés

A.5 Correspondance des scanners sources

Souvent, le document en question n'est pas acquis par l'un des scanners « connus », c.-à-d., non disponible lors de l'apprentissage du système d'identification du scanner source. Dans une telle situation réelle, l'expert forensique va assumer que le document a été acquis par l'un des scanners connus. Cependant, cette hypothèse est fausse et engendre de mauvaises classifications. Une solution possible est de vérifier si le document n'a pas été acquis par l'un des scanners connus en comparant son origine avec l'origine d'images acquises par ces scanners. Pour cela, nous avons proposé une nouvelle technique de correspondance de sources. Elle permet de faire face à la situation dans laquelle l'expert en forensique a besoin de savoir si une paire d'images a été acquise par le même modèle de scanner ou non. La technique est basée, tout d'abord, sur l'extraction des signatures des scanners sources de chaque image en exploitant le système 2D-CNN proposé précédemment. Ensuite, elle compare ces signatures à travers un réseau de neurones à double entrée qui génère des scores indiquant le taux de similarités entre les deux signatures qui n'est autre que le taux de similarités entre les deux images en question.

A.6 Nouvelle base d'images annotées adaptée à la détection de la falsification des images

Du fait qu'il n'existe aucune base d'images publique permettant de tester les approches forensiques relatives aux scanners, nous avons préparé une collection de documents scannés pouvant servir dans plusieurs études de recherche liées aux documents numérisés. Ensuite, inspiré par les modèles d'identification du scanner source présentés dans les sections précédentes, nous avons proposé deux nouvelles approches de détection de falsification en s'appuyant sur l'importance de l'exploitation la sous-bande HH. Les méthodes proposées sont plus pertinentes pour localiser les contrefaçons dans les images acquises à l'aide de scanners à plat que les autres méthodes existantes dans la littérature. Nous avons validé l'efficacité de ces approches en effectuant des tests sur les images falsifiées présentes dans la nouvelle base crée.
Dans la première approche, nous comparons les histogrammes des sous-bandes HH des différents blocs de l'image en question avec ceux du scanner ayant la même localisation. Nous avons opté pour la distance euclidienne comme mesure de ressemblance. La deuxième approche opère directement sur l'image en suspect sans avoir recours à une signature de scanner comme référence. L'idée de diviser l'image scannée en des blocs de même taille et identifier le scanner source de chacun de ces blocs. Une diversité importance des sources identifiées fait que l'image a été falsifiée et il est ainsi possible de discerner les régions manipulées.

A.7 Conclusion et perspectives

Pour conclure, différents travaux ont été menés durant cette thèse pour la résolution du problème d'identification du scanner source. L'objectif principal était de concevoir un modèle adaptatif pour estimer le scanner qui a acquis un document numérique suspect où tous les scanners sont censés être connus. Nous avons pris en compte tous ces défis dans le but d'établir de nouvelles approches fondées sur des modèles capables d'extraire un bruit de scanner plus précis. Des approches basées sur une extraction manuelle des caractéristiques du scanner ont d'abord été exploitées. Les limitations principales de ces approches sont la baisse des performances dans le cas où des scanners de même modèle sont considérés et la dépendance au nombre

d'images d'apprentissage. Pour surmonter ces limitations, des empreintes digitales extraites automatiquement à l'aide de réseaux de neurones sont proposées.
Par la suite, nous avons proposé une nouvelle piste de recherche dans laquelle nous considérons deux images de sources inconnues et visons à savoir si elles ont été acquises par le même scanner ou non. Finalement, nous avons introduit une nouvelle collection d'images à des fins d'investigations forensiques liées aux scanners. Nous avons exploitée cette base d'images pour vérifier l'efficacité de nouvelles architectures que nous avons introduites dans le but de détecter les régions falsifiées dans un document scanné.
Comme travaux futures, nous avons intérêt à exploiter d'avantages les propriétés des images scannées JPEG. Par ailleurs, il serait intéressant de concevoir un système unique qui peut non seulement identifier le scanner source d'une image s'il est connu mais aussi alerter si le scanner est inconnu. Il est également possible de penser à créer de nouvelles signatures basées sur des propriétés du scanner autres que celles liées aux défauts de ses capteurs. Un autre problème crucial avec les approches proposées est qu'elles sont basées sur le bruit est qui peut être affecté par la désynchronisation produite par les opérations de post-traitement telles que la compression et les transformations géométriques (par exemple, rotation, mise à l'échelle ...). Cela motive à envisager d'utiliser des techniques de recadrage et de redimensionnement dans de futurs travaux pour résoudre ce problème.
En ce qui concerne la détection de falsification, il est important de penser à des solutions qui restent fiables dans le cas où de nombreuses opérations de fraudes ont été appliquées à l'image. De plus, les approches proposées dans la littérature présentent une certaine complexité et n'ont de sens que dans le cas où l'image en question est alignée avec la signature du scanner. Les performances de ces approches dépendent également de la taille du bloc de test et ne peuvent pas informer du type de manipulation réalisée sur l'image. Il y a également un intérêt à trouver une solution qui puisse, non seulement détecter les contrefaçons, mais aussi indiquer le type de manipulation. Par exemple, on peut décider si la manipulation est une opération de type « copier-coller » en vérifiant si l'objet ou le bloc suspect est répété dans une zone différente de l'image. Relever ces défis ouvrira la porte à des contributions plus efficaces.
Récemment, la notion d'anti-forensiques [226] a été introduite dans la littérature. C'est une méthodologie permettant d'examiner les faiblesses des méthodes forensiques existantes afin d'améliorer leur pertinence et leur fiabilité. L'objectif fondamental est donc d'exécuter certaines opérations sur des images numériques pour masquer les traces laissées par quelconque manipulation afin que les algorithmes forensiques ne puissent pas les détecter. Notons que ces techniques visent à aider les chercheurs en criminalistique à connaître les vulnérabilités des approches qu'elles proposent et, ainsi, les inciter à les renforcer contre d'éventuelles attaques en introduisant les méthodes de contre-mesures (Anti-anti-forensiques). A partir d'une analyse de publications abordant ces thèmes, il est clair que le nombre de publications liées à ces domaines de recherche est encore faible. Par conséquent, nous pouvons exploiter nos travaux comme un bon point de départ pour de futures recherches sur des problèmes anti-forensiques et leurs contre-mesures.

B
WAVELET TRANSFORM

The Fourier Transform (FT) projects a signal from the time domain to the frequency domain by dividing it into sinusoidal basis functions with varying frequencies. This transformation does not lose any information, that is to say, we can fully recover the original signal from its Inverse Fourier Transform (IFT) representation. However, it provides only precise information on the signal's spectral components, but not about the temporal localization of these components. Therefore, the wavelet transform (WT) has been proposed to overcome the limits of the FT by dividing the unstationary signals into discrete intervals that are short enough to obtain stationary signals in each interval. In fact, the WT is a mathematical transformation technique that aims to represent any data as a wavelet superposition. The wavelet subbands are further decomposed with a certain resolution at each level. Moreover, it is able to extract temporal and spectral information simultaneously. Unlike the FT, the DWT refers to a collection of transforms, each with its own set of wavelet basis functions such as the Haar, the Daubechies and the Symlet. In fact, given the wide variety of wavelets and the possibility to go through many levels, the WT gives the possibility to work with different subband depending on the application that it will serve.
A one level 2D wavelet transform is depicted in Fig. B.1. Four new images are obtained, which are: the approximation subband (Low-Low LL), vertical details subband (Low-High LH), horizontal details subband (High-Low HL) and diagonal details subband (High-High HH). It is important to note that this later isolates the localized high frequency features. Further levels of decomposition can be obtained by repeating this decomposition each time on the LL subband as shown in Fig B.1.

B.1 The discrete wavelet transform (DWT)

As we are particularly interested in images, in this section, we will present the algorithm for computing the two-dimensional DWT, which is very useful in image processing and computer vision applications. An image can be analyzed using the DWT algorithm by passing it through an analysis filter bank followed by a decimation operation. In each decomposition stage, the analytical filter bank includes a Low Pass Filter (LPF) and a High Pass Filter (HPF). When a signal crosses these filters, it got divided into two bands. After each decomposition, the image's most energy is contained in the low frequency subband while the image's details and edges are represented by the other subbands.
Note that for a 2D image of M rows and N columns, the produced subbands are of size $M\backslash 2$ rows and $N\backslash 2$ columns due to the decimation operation.

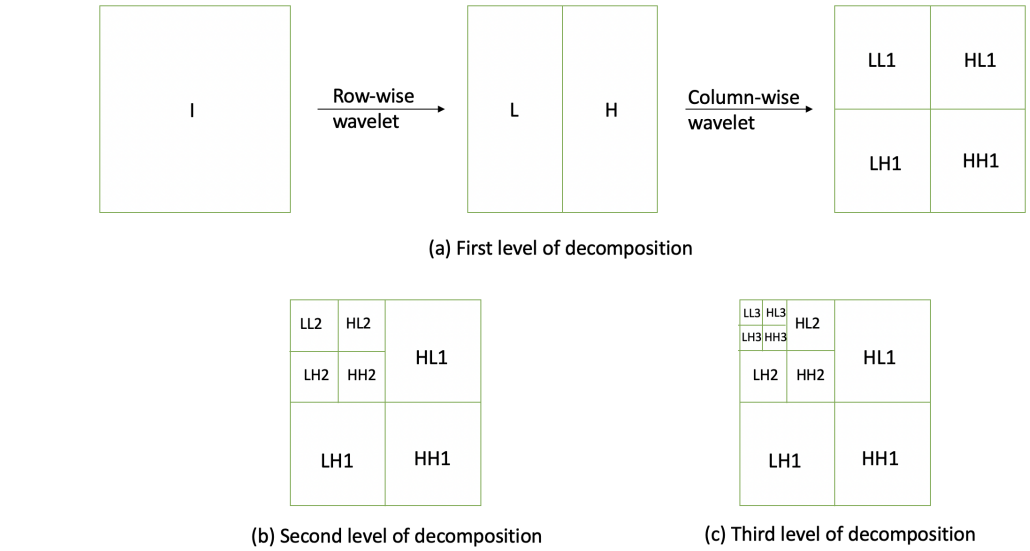


Figure B.1: Illustration of different wavelet decomposition levels

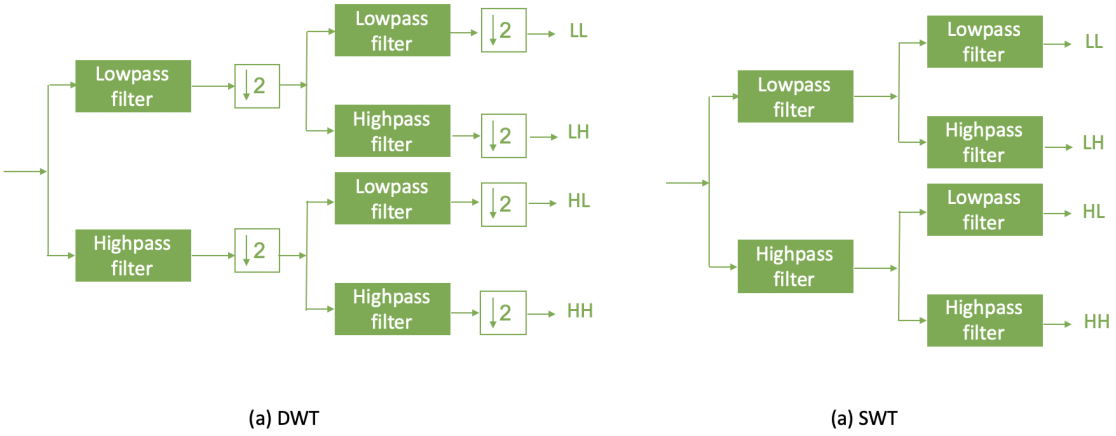


Figure B.2: Difference between DWT and SWT decomposition trees

B.2 The stationary wavelet transform (SWT)

DWT is not a transform that is shift-invariant. The SWT, also known as undecimated wavelet transform, can overcome this limitation. The downsampling stage is eliminated at every transformation level, that is to say, the subband coefficients are not decimated. Thus, the number of columns and number of rows of different subbands are the same as the original image with this decomposition. Figure B.2 illustrates the difference DWT and SWT decomposition process.

C
NEURAL NETWORKS

In the past few years, deep learning (DL) and, in particular Convolutional Neural Network (CNN), have been widely applied in the image processing field. CNN not only surpasses prior handcrafted based methods, allowing to achieve high precision in image classification, but it also serves to automatically extract features from images. It is, thus, necessary to understand its architecture and to know some classic CNN models that are commonly used in various fields. Since we use CNNs extensively in this thesis, we introduce important aspects in this appendix.

C.1 Two-Dimensional Convolutional Neural Networks (2D-CNNs)

C.1.1 Overall structure

A 2D-CNN is a deep learning architecture capable of taking an input image, learning various features (e.g. aspects, objects, edges...) in that image, and then using these learned features for image classification. Figure C.1 shows an overall basic structure of 2D-CNN composed of several layers.

It includes convolutional layers (Conv), activation (Actv), pooling, and fully connected (FC) layers. The basic purpose of the convolutional layer is to extract the features of the input image. The pooling layer acts as a down sampling operation. It is mostly used to lower the resolution of features maps. On the other side, activation layers are used to introduce nonlinear elements and boost the neural network's expression ability. Then, a set of FC layers are integrated to minimize the size of feature maps and to classify them. The last layer outputs maps the input image to one of the classes of the dataset.
In the following, we describe each of these layers:

Convolution Layer
The convolutional layer is the core building block of a CNN. It mainly applies a convolution filter on the input image using a sliding window algorithm. For each window, an element-wise product between each input and kernel element is calculated and summed to obtain the output value in the corresponding position of the output feature map. This later has a relative relationship with the number of

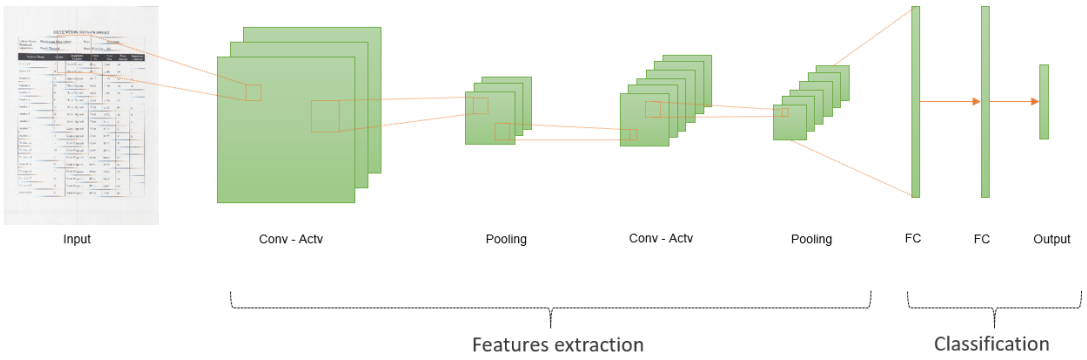


Figure C.1: Example of a 2D-CNN

kernels, the kernel size, the stride size and the padding size. The number of kernels and the kernel size represent the depth of the output features map and the size of the filter applied, respectively. The stride, by default 1, indicates how much the convolution filter is moved at each step. Big strides are usually used to make small feature maps. On the other side, padding is the process of adding pixels to the image's edge in order to ensure that the feature map has the same shape as the input of the convolutional layer.
For example, if we perform a convolutional operation over the image shown in Fig. C.2 with the 6 filters of size 3x3 for each color channel (R, G and B) depicted in Fig. C.3, then, we will obtain the feature maps shown in Fig. C.4.

Figure C.2: An image sample

Here, only the first 16 feature maps are displayed. We can notice a lot of versions of the input image were produced but with different features highlighted.

Activation
The activation function aids in introducing non-linearity into the network by permitting certain inputs to be transmitted forward. This unit is, traditionally, a rectified linear unit (ReLU), which employs the non-saturating activation function indicated in the equation below

$$f(x) = \max(0, x) \tag{C.1}$$

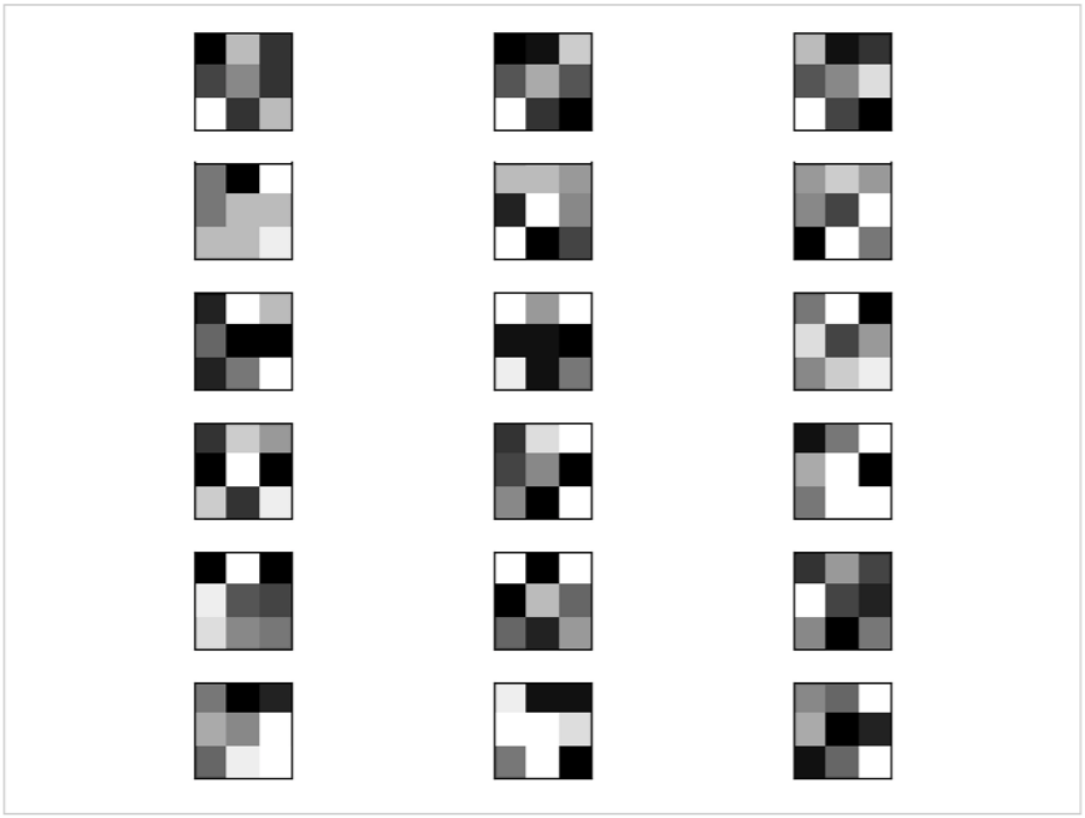


Figure C.3: Normalized filters used for each color channel R, G and B

Recently, Sigmoid and Tanh have become more and more used as activation functions.
The equation of Sigmoid is

$$\text{sigmoid } (x) = \frac{1}{1 + e^{-x}} \tag{C.2}$$

The equation of Tanh is

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{C.3}$$

However, ReLU is still the most widely utilized activation function in the field of image processing. This function, compared to Sigmoid or Tanh, can converge faster. Notice that a variety of enhanced activation functions such as Leaky ReLU, PReLU, Randomized ReLU (RReLU) as well as the scaled exponential linear units (SELU) have been derived from the ReLu. Each should be selected according to the actual issues to be addressed.

Pooling Layer
The pooling layer is a kind of down-sampling usually used after a convolutional layer. It is mainly performed to reduce the dimensionality without losing too much important information. In fact, there are two common types of pooling methods among which the maximum pooling is the most used. This later slides a window over the input and outputs the maximum value for each window. Figure C.5 gives an example of a max pooling operation where each color refers to a different non-overlapping window. The second type of pooling is the average pooling where the output is the average of all values in each window.

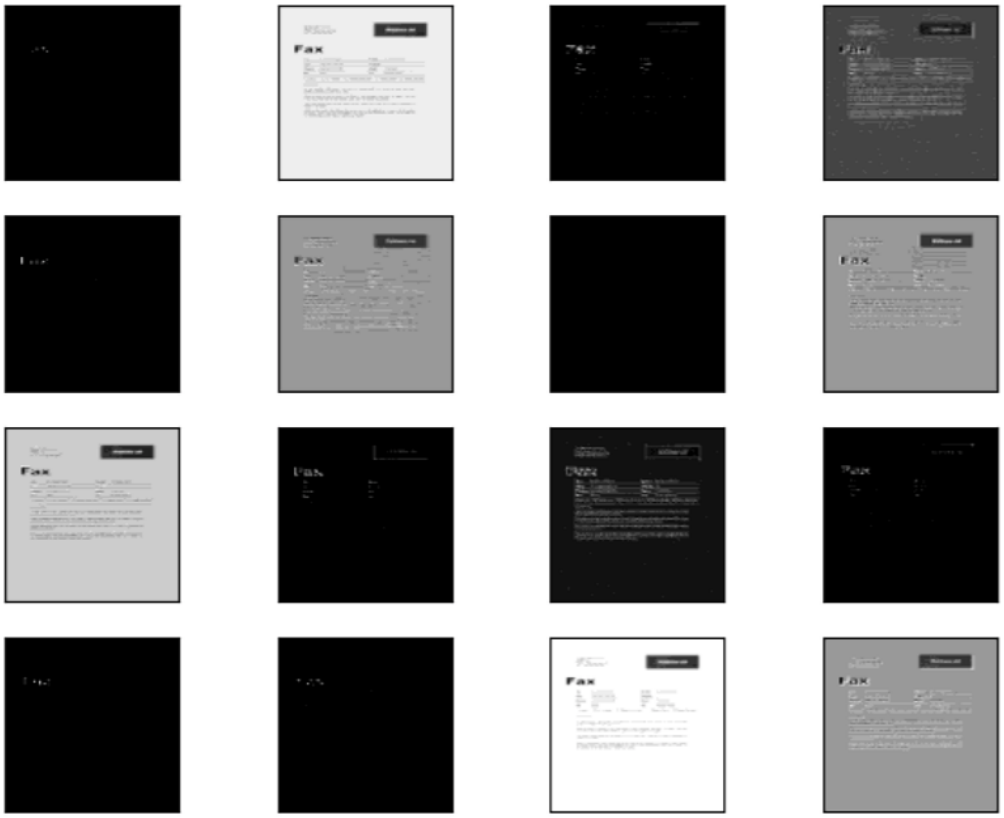


Figure C.4: Visualization of the feature maps obtained by a convolutional layer of shape (3,3,3,32)

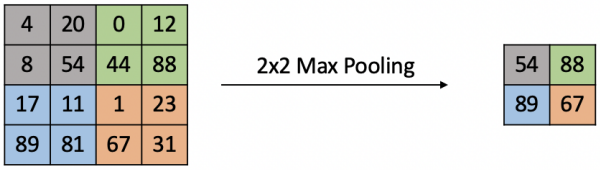


Figure C.5: An example of a 2x2 Max pooling operation: A max function is applied on each 2x2 windows with a stride 2 returning the highest value in each window

Fully-connected Layer

In a CNN, the fully connected layer serves as a classifier. While convolutional layers, pooling layers, and activation function layers help to extract the representative features from the original data, the fully connected layers map the learned features to a number of classification probabilities. Notice that the output of a fully connected layer is a 1D vector obtained by flattening the output of the previous layer.
Usually more than one fully connected layers is necessary to construct the classification part of the network. The choice of the parameters of these layers is very important to avoid overfitting.

C.1.2 Classical CNN models

In this section, we introduce some classic CNN models that have been used in this thesis. On the one hand, these models were sensational at that time and played a significant role in promoting the development of deep neural network. On the other hand, many of them are pre-trained on large datasets and are widely used in transfer learning as will be explained below.

AlexNet
Large datasets like ImageNet [227], which include hundreds of thousands to millions of annotated pictures, have increased the requirement for a highly competent deep learning model such as AlexNet. This later is a tremendously strong model capable of attaining high accuracies on extremely challenging databases. Removing any of the convolutional layers, on the other hand, will significantly reduce its performance. This leading architecture has potential applications in computer vision and artificial intelligence related issues. As it is shown by Fig. C.6, AlexNet adopts three convolutional layers with different kernel sizes (11×11, 5×5 and 3×3). At each of these convolutional layers, features got extracted and the size of the resulting feature maps is reduced with max pooling layers of size 3x3. Finally, three fully connected layers are used to generate the output.

GoogleNet
Another notable DL architecture is the so-called GoogLeNet. This deep network is composed of 22 layers of which an inception structure, shown in Fig. C.8, is repeated 9 times. The inception acts as a single layer in the network by stacking three convolutional cores of sizes 5x5, 3x3 and 1x1 simultaneously in parallel with an occasional max pooling layer. The full architecture of this CNN is presented in Fig. C.7.
It is interesting to know that GoogLeNet is has twelve times less parameters than the AlexNet and, thus, it is faster to train.

C.2 One-Dimensional Convolutional Neural Networks (1D-CNNs)

As stated above, 2D-CNNs are able to learn complicated objects and patterns if trained on a vast size visual database. Thus, when they are properly trained, they may serve as the principal tool for many applications related to 2D signals such as digital images and video frames. However, this may not be the best solution in applications involving 1D signals, particularly when training data is sparse or application specific.
During the last decade, 1D-CNN has become applicable to a wide range of fields [228] due to its theoretical appeal and impressive performance. Such networks appear suitable to address issues related to 1D data. Let us recall that the general architecture of a 1D-CNN consists of several layers of different kinds as shown in Fig. C.9. The most used layers are as follows:

- 1D convolutional layer - It conducts convolution operations in order to extract feature maps by sliding a set of kernels over the input data. Its output is, then, passed through a non-linear activation function such as sigmoid, hyberbolic tangent (tanh), Rectified Linear Unit (ReLU) [167],..

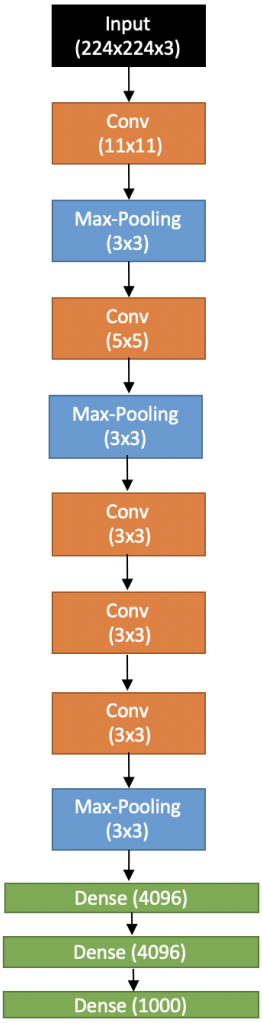


Figure C.6: AlexNet architecture

- Max pooling layer - It is usually used after a convolution layer for data complexity and dimensionality reduction. Its role is to down-sample the feature maps resulting from the previous layer in order to reduce the amount of parameters to learn and the network computational cost. The output of this layer is generated by scanning several regions of the feature map and computing the maximum of each of those regions.

- Global average pooling layer - This layer plays an important role in avoiding overfitting. As the max pooling layer, its main goal is the data dimensionality reduction but, this time, by computing average of every incoming feature map.

- Dropout layer [188] - It is also used to reduce overfitting. With this layer, the generalization of the CNN got improved by becoming less sensitive to small data variations.

- Dense layer - This layer is a Multi-Layer Perceptron (MLP) that classifies the input data by computing scores of each class where the number of classes is equal to the number of neurons. Scores should sum up to one.

C.3 Transfer learning

When dealing with computer vision problems, one circumstance happens regularly; the need to establish a large database is required which is both complex and expensive. In other situations, we want to learn something new from a previously solved problem and, thus, move on to the next assignment as fast as possible. This is the idea behind Transfer Learning (TL). This later is simply a way to transfer knowledge from one field to another. For instance, knowledge gained while learning to recognize devices could apply when trying to recognize scanners.
A simple illustration of TL is given in Fig. C.10 .The CNN model developed for a specific task is reused with different data to solve a different problem.

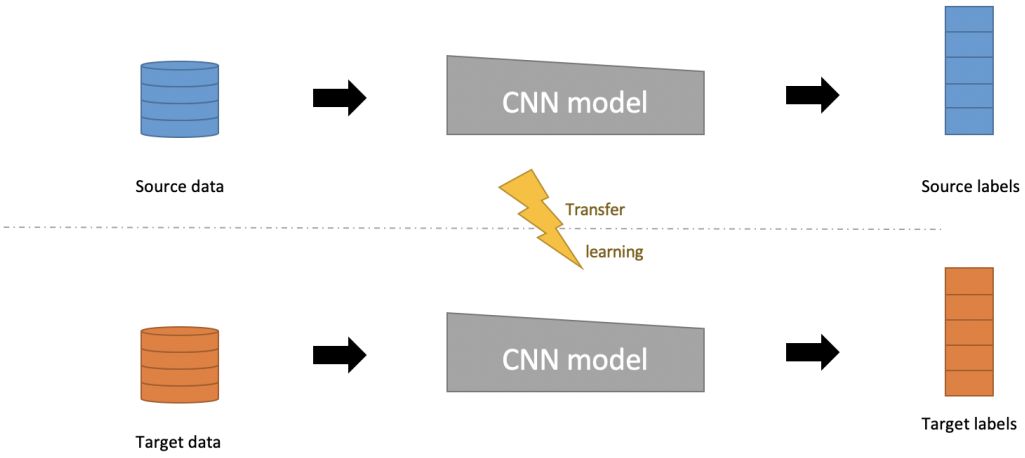


Figure C.10: Illustration of transfer learning

In recent years, one of the most common strategies of applying TL is to utilize a classic convolutional neural network as a pre-trained model, freeze some layers and then retrain a few layers by the target data. Another popular strategy is to use some layers from the pre-trained model as a feature extractor and, then, add a new classifier such as Support Vector Machine (SVM).

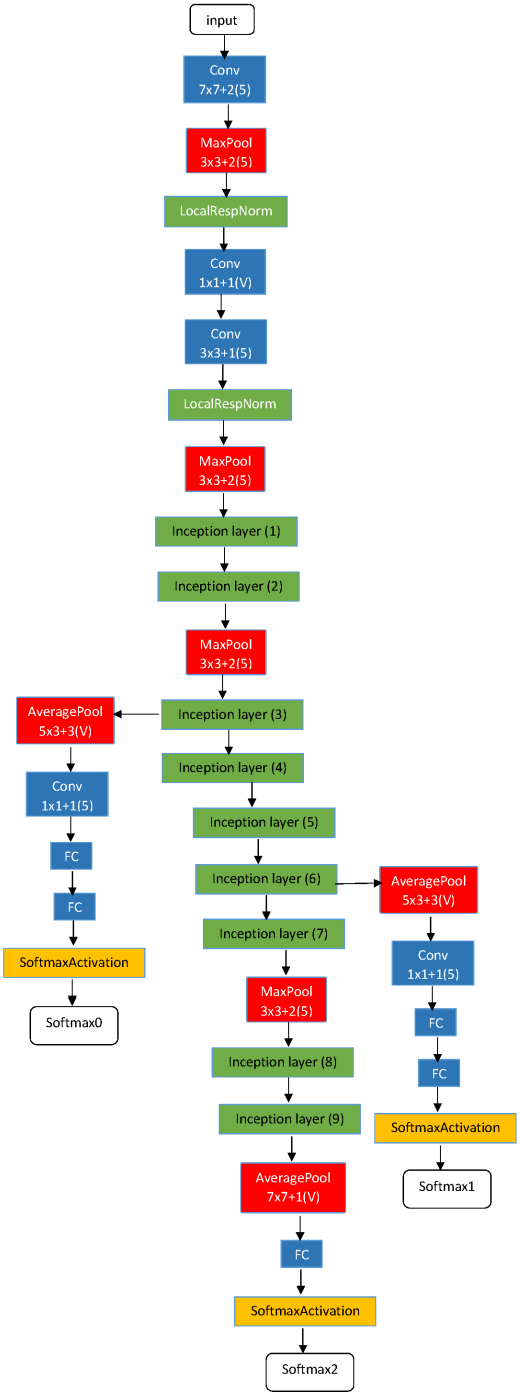


Figure C.7: GoogleNet architecture

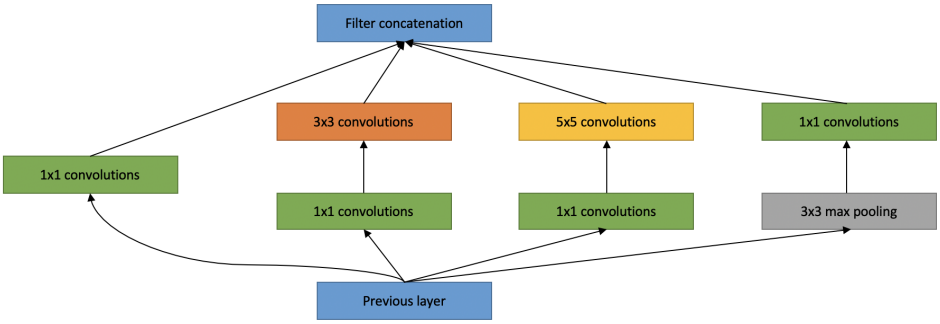


Figure C.8: GoogleNet Inception module

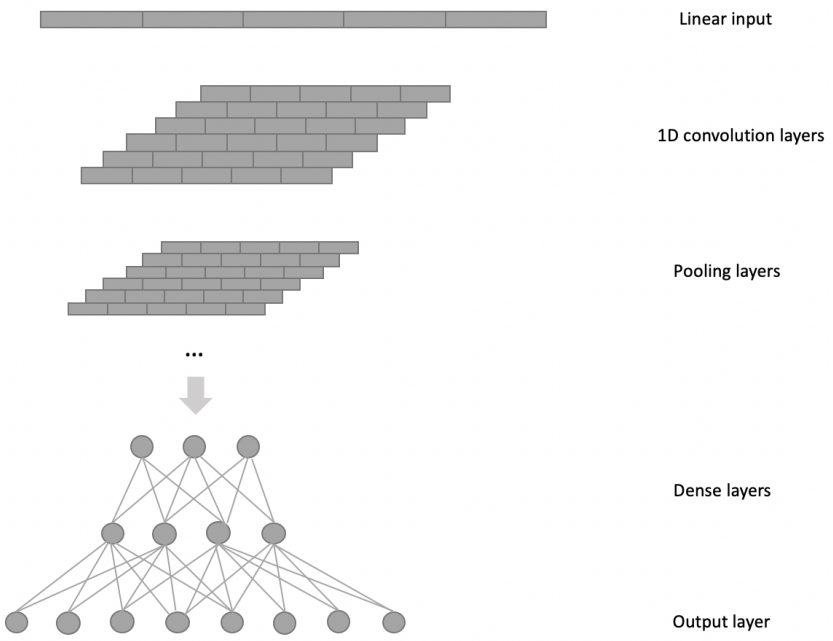


Figure C.9: General structure of 1D CNN

D

SUPPORT VECTOR MACHINE CLASSIFIER

The Support Vector Machine (SVM) is a popular supervised machine learning method used for the classification of data belonging to two classes. It can also be extended to solve regression problems. The objective of the SVM classifier is to maximize the margin which is the distance from the hyperplane, also called decision boundary, to the closest elements from each class on either side. The hyperplane can be written as the set of points x satisfying the Linear Discriminant Function (LDF)

$$w^\top \cdot x - b = 0 \tag{D.1}$$

where w is the weight vector normal to the hyperplane.

Figure D.1 represents an illustration of SVM. Given a set of data points, each of which belongs to one of two classes, the goal is to determine which of the two classes a new data point belongs to. For that, we want to separate such points with a hyperplane that reflects the greatest separation, or margin, between the two classes because the larger the margin, the smaller the classifier's generalization error. The support vectors and the margins are used to find the optimal hyperplane where the support vectors are points on each of the lines $w^\top.x$ - b = - 1 and $w^\top.x$ - $b = 1$. Notice that deleting some of the support vectors may change the position of the hyperplane. The distance between these two hyperplanes, which is 2 / $|w|$, should be minimized.

To maximize the margin of separation, these hyperplanes are represented by the following equations

$$\begin{aligned} w^\top.x_i - b >= 1, \text{for } x_i \text{ points of the first class} \\ w^\top.x_i - b <= -1, \text{for } x_i \text{ points of the second class} \end{aligned} \tag{D.2}$$

After training the SVM classifier, a data point x_p is classified to the first class if $w.x_p - b > 0$, otherwise it is classified as belonging to the second class. When there are more than two classes the simplest solution is to create Q two-class problems where Q is the number of classes, then, separate each class from all other classes combined, thus learning Q SVMs.

We distinguish two types of SVM: Linear and non-linear, depending on how the

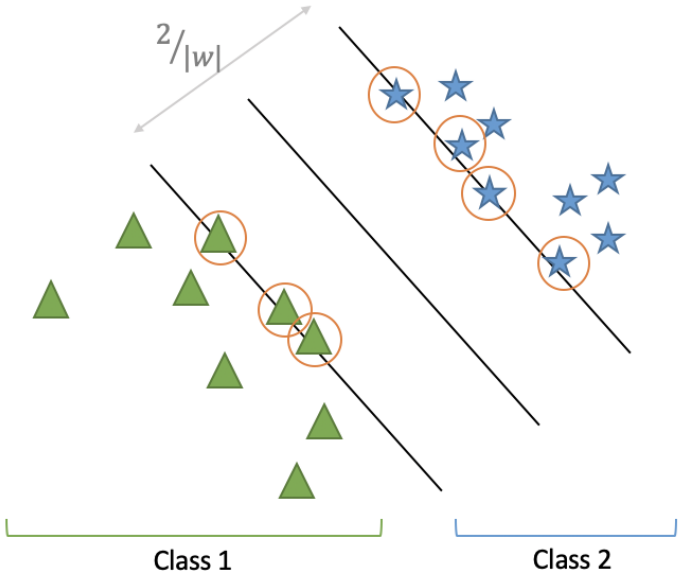


Figure D.1: A linear SVM example

data are separated. More clearly, if the data is linearly separable, then the SVM is linear. However, in the case where the dataset cannot be separated using a straight line, the SVM is said non-linear.
Furthermore, the SVM is using a set of mathematical functions called kernels. The role of the kernel is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions such as linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid. We introduce, in the following, these four basic kernels:

Linear
This kernel is used when the data can be divided linearly, that is to say, with a single line. It is one of the most often utilized kernels for text classification.

$$K\left(\mathbf{x}_i, \mathbf{x}_j\right) = \mathbf{x}_i^T \mathbf{x}_j \tag{D.3}$$

where x_i and x_j are vectors of features computed from training or test samples.

Polynomial
This kernel is quite practical for natural language processing (NLP). It is used to learn non-linear models by representing the similarity of vectors in a feature space over polynomials of the original samples.

$$K\left(\mathbf{x}_i, \mathbf{x}_j\right) = \left(\gamma {\mathbf{x}_i}^T \mathbf{x}_j + r\right)^d, \gamma > 0 \tag{D.4}$$

where $r \geqslant 0$ is a free parameter in the polynomial that balances the effect of higher-order vs lower-order terms.

RBF
In general, the RBF kernel is the most used type. It is preferred when there is no prior knowledge about the data. It maps data into a higher dimensional space, allowing it to handle cases where the relationship between class labels and attributes is nonlinear. Its equation is defined as

$$K\left(\mathbf{x}_i, \mathbf{x}_j\right) = \exp\left(-\gamma \left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2\right), \gamma > 0 \tag{D.5}$$

When training an SVM with an RBF kernel, two parameters must be considered: C and γ. The parameter C, which is shared by all SVM kernels, trades off misclassification of training examples versus decision surface simplicity. A low C smoothes the decision surface, whereas a high C attempts to correctly classify all training samples. On the other hand, Gamma refers to the degree of influence a single training sample has. The selection of C and γ is critical to the performance of the SVM. To select convenient values, GridSearchCV with C and gamma spaced exponentially is recommended.

Sigmoid
It is originated from NN. Its equation is such that

$$\tanh\left(\gamma {\mathbf{x}_i}^T \mathbf{x}_j + r\right) \tag{D.6}$$

List of Publications

International Conferences

Chaima Ben Rabah, Gouenou Coatrieux, and Riadh Abdelfattah. "Scanner model identification of official documents using noise parameters estimation in the wavelet domain". In: *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer. 2018, pp. 598–608.

Chaima Ben Rabah, Gouenou Coatrieux, and Riadh Abdelfattah. "The SUPATLANTIQUE scanned document database for digital image forensics purposes". In: *2020 IEEE International Conference on Image Processing (ICIP)* (2020).

International Journals

Chaima Ben Rabah, Gouenou Coatrieux, and Riadh Abdelfattah. "Automatic source scanner identification using 1D convolutional neural network". In: *Multimedia Tools and Applications* (2021), pp. 1–18.

Chaima Ben Rabah, Gouenou Coatrieux, and Riadh Abdelfattah. "Boosting up Source Scanner Identification Using Wavelets and Convolutional Neural Networks". In: *Traitement du Signal* 37.6 (2020).

National Conferences

Chaima Ben Rabah et al. "Identification de l'Origine d'un Document Numérisé sur la Base d'une Empreinte de Scanner Dans le Domaine des Ondelettes". In: *Computer & Electronics Security Applications Rendez-vous (C&ESAR)*. 2017.

Chaima Ben Rabah, Gouenou Coatrieux, and Riadh Abdelfattah. "Semi-Blind Source Scanner Identification". In: *2019 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC)*. IEEE. 2019, pp. 220–225.

References

[1] *Ministère de l'Europe et des Affaires étrangères. (n.d.). Victoires des Acteurs publics 2020.* June 2020. URL: https://www.diplomatie.gouv.fr/fr/archives-diplomatiques/action-scientifique-et-culturelle/cabinet-des-decouvertes/article/victoires-des-acteurs-publics-2020.

[2] Hany Farid. "Digital doctoring: how to tell the real from the fake". In: *Significance* 3.4 (2006), pp. 162–166.

[3] Mayce Y. Khoury. *Document Forgery in Healthcare: The Integral Role of Primary Source Verification as a Solution.* https://corp.dataflowgroup.com/wp-content/uploads/2017/01/Document_Forgery_in_Healthcare_-_The_Integral_Role_of_PSV_as_a_Solution.pdf. 2016.

[4] *Certificate Forgery: QATAR Blacklists 17 Doctors, 2 Nurses and 4 Allied Health Workers.* URL: www.qatar-tribune.com/latestnews-article/mid/506/articleid/278/certificate-forgery-qatar-blacklists-17-doctors-2-nurses-and-4-allied-health-workers (visited on 08/26/2021).

[5] Lanxiang Chen et al. "Blockchain based searchable encryption for electronic health record sharing". In: *Future Generation Computer Systems* 95 (2019), pp. 420–429.

[6] Budi Triand et al. "Digital Document Security on Legalize Higher Education Diplomas with Digital Signature and SHA-1 Algorithm". In: *2019 7th International Conference on Cyber and IT Service Management (CITSM).* Vol. 7. IEEE. 2019, pp. 1–5.

[7] Ashwani Kumar. "A Review on Implementation of Digital Image Watermarking Techniques Using LSB and DWT". In: *Information and Communication Technology for Sustainable Development.* Springer, 2020, pp. 595–602.

[8] Surekha Borra and Rohit Thanki. "Crypto-watermarking scheme for tamper detection of medical images". In: *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* 8.4 (2020), pp. 345–355.

[9] Amit Kumar Singh et al. "Joint encryption and compression-based watermarking technique for security of digital documents". In: *ACM Transactions on Internet Technology (TOIT)* 21.1 (2021), pp. 1–20.

[10] William D Ferreira et al. "A review of digital image forensics". In: *Computers & Electrical Engineering* 85 (2020), p. 106685.

[11] Jangirala Srinivas, Ashok Kumar Das, and Neeraj Kumar. "Government regulations in cyber security: Framework, standards and recommendations". In: *Future Generation Computer Systems* 92 (2019), pp. 178–188.

[12] Kishore Sakariya. “DIGITAL PRESERVATION: PROCESS AND STANDARDS”. In: *INTERNET OF THINGS AND CURRENT TRENDS IN LIBRARIES* (2018), p. 75.

[13] Chaima Ben Rabah et al. “Identification de l’Origine d’un Document Numérisé sur la Base d’une Empreinte de Scanner Dans le Domaine des Ondelettes”. In: *Computer & Electronics Security Applications Rendez-vous (C&ESAR)*. 2017.

[14] Chaima Ben Rabah, Gouenou Coatrieux, and Riadh Abdelfattah. “Scanner model identification of official documents using noise parameters estimation in the wavelet domain”. In: *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer. 2018, pp. 598–608.

[15] Chaima Ben Rabah, Gouenou Coatrieux, and Riadh Abdelfattah. “Semi-Blind Source Scanner Identification”. In: *2019 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC)*. IEEE. 2019, pp. 220–225.

[16] Chang-Hee Choi, Min-Jeong Lee, and Heung-Kyu Lee. “Scanner identification using spectral noise in the frequency domain”. In: *Image Processing (ICIP), 2010 17th IEEE Int. Conference on*. IEEE. 2010, pp. 2121–2124.

[17] Chaima Ben Rabah, Gouenou Coatrieux, and Riadh Abdelfattah. “Automatic source scanner identification using 1D convolutional neural network”. In: *Multimedia Tools and Applications* (2021), pp. 1–18.

[18] Chaima Ben Rabah, Gouenou Coatrieux, and Riadh Abdelfattah. “Boosting up Source Scanner Identification Using Wavelets and Convolutional Neural Networks”. In: *Traitement du Signal* 37.6 (2020).

[19] Chaima Ben Rabah, Gouenou Coatrieux, and Riadh Abdelfattah. “The SUPATLANTIQUE scanned document database for digital image forensics purposes”. In: *2020 IEEE International Conference on Image Processing (ICIP)* (2020).

[20] Thomas Gloe, Elke Franz, and Antje Winkler. “Forensics for flatbed scanners”. In: *Security, Steganography, and Watermarking of Multimedia Contents IX*. Vol. 6505. International Society for Optics and Photonics. 2007, p. 65051I.

[21] Gerald C Holst. “CCD arrays, cameras, and displays”. In: (1998).

[22] Marvin H White et al. “Characterization of surface channel CCD image arrays at low light levels”. In: *IEEE Journal of Solid-State Circuits* 9.1 (1974), pp. 1–12.

[23] Jan Lukas, Jessica Fridrich, and Miroslav Goljan. “Determining digital image origin using sensor imperfections”. In: *Image and Video Communications and Processing 2005*. Vol. 5685. International Society for Optics and Photonics. 2005, pp. 249–260.

[24] M Kivanc Mihcak, Igor Kozintsev, and Kannan Ramchandran. “Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising”. In: *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258)*. Vol. 6. IEEE. 1999, pp. 3253–3256.

[25] Hongmei Gou, Ashwin Swaminathan, and Min Wu. "Robust scanner identification based on noise features". In: *Security, steganography, and watermarking of multimedia contents IX*. Vol. 6505. International Society for Optics and Photonics. 2007, 65050S.

[26] Hongmei Gou, Ashwin Swaminathan, and Min Wu. "Intrinsic sensor noise features for forensic analysis on scanners and scanned images". In: *IEEE Transactions on Information Forensics and Security* 4.3 (2009), pp. 476–491.

[27] Nitin Khanna et al. "Scanner identification using sensor pattern noise". In: *Security, steganography, and watermarking of multimedia contents IX*. Vol. 6505. International Society for Optics and Photonics. 2007, 65051K.

[28] Nitin Khanna et al. "Scanner identification with extension to forgery detection". In: *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*. Vol. 6819. International Society for Optics and Photonics. 2008, 68190G.

[29] Nitin Khanna, Aravind K Mikkilineni, and Edward J Delp. "Scanner identification using feature-based processing and analysis". In: *IEEE Transactions on Information Forensics and Security* 4.1 (2009), pp. 123–139.

[30] Alessandro Foi et al. "A novel anisotropic local polynomial estimator based on directional multiscale optimizations". In: *Proc. 6th IMA int. conf. math. in signal processing*. 2004, pp. 79–82.

[31] Shigeru Sugawara. "Identification of scanner models by comparison of scanned hologram images". In: *Forensic science international* 241 (2014), pp. 69–83.

[32] Nitin Khanna and Edward J Delp. "Source scanner identification for scanned documents". In: *2009 First IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE. 2009, pp. 166–170.

[33] Nitin Khanna and Edward J Delp. "Intrinsic signatures for scanned documents forensics: effect of font shape and size". In: *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*. IEEE. 2010, pp. 3060–3063.

[34] Sharad Joshi, Gaurav Gupta, and Nitin Khanna. "Source classification using document images from smartphones and flatbed scanners". In: *Computer Vision, Pattern Recognition, Image Processing, and Graphics: 6th National Conference, NCVPRIPG 2017, Mandi, India, December 16-19, 2017, Revised Selected Papers 6*. Springer. 2018, pp. 281–292.

[35] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns". In: *IEEE Transactions on pattern analysis and machine intelligence* 24.7 (2002), pp. 971–987.

[36] Ahmet Emir Dirik, Husrev Taha Sencar, and Nasir Memon. "Flatbed scanner identification based on dust and scratches over scanner platen". In: *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE. 2009, pp. 1385–1388.

[37] Zeinab F Elsharkawy et al. "Identifying Unique Flatbed Scanner Characteristics for Matching a Scanned Image to its Source". In: *Digital Image Processing* 5.9 (2013), pp. 397–403.

[38] Zankhana J Barad and Mukesh M Goswami. “Image Forgery Detection using Deep Learning: A Survey”. In: *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*. IEEE. 2020, pp. 571–576.

[39] Manpreet Singh, Harpreet Kaur, and Ajay Kakkar. “Digital signature verification scheme for image authentication”. In: (2015), pp. 1–5.

[40] Sahib Khan et al. “Forgery Detection and Localization of Modifications at the Pixel Level”. In: *Symmetry* 12.1 (2020), p. 137.

[41] Oussama Benrhouma et al. “Chaotic watermark for blind forgery detection in images”. In: *Multimedia Tools and Applications* 75.14 (2016), pp. 8695–8718.

[42] Wu-Chih Hu et al. “Effective image forgery detection of tampered foreground or background image based on image watermarking and alpha mattes”. In: *Multimedia Tools and Applications* 75.6 (2016), pp. 3495–3516.

[43] Radu Ovidiu Preda, DRAGOŞ NICOLAE Vizireanu, and Simona Halunga. “Active image forgery detection scheme based on semi-fragile watermarking”. In: *Revue Roumaine des Sciences Techniques-Serie Electrotechnique et Energetique, Romania* 61.1 (2016), pp. 58–62.

[44] Nikodim Lazarov and Zlatoliliya Ilcheva. “A fragile watermarking algorithm for image tamper detection based on chaotic maps”. In: *2016 IEEE 8th International Conference on Intelligent Systems (IS)*. IEEE. 2016, pp. 723–728.

[45] Heng Zhang, Cheng-You Wang, and Xiao Zhou. “Fragile Watermarking Based on LBP for Blind Tamper Detection in Images.” In: *JIPS* 13.2 (2017), pp. 385–399.

[46] Yinyin Peng et al. “Image authentication scheme based on reversible fragile watermarking with two images”. In: *Journal of information security and applications* 40 (2018), pp. 236–246.

[47] Ertugrul Gul and Serkan Ozturk. “A novel hash function based fragile watermarking method for image integrity”. In: *Multimedia Tools and Applications* 78.13 (2019), pp. 17701–17718.

[48] Shiv Prasad and Arup Kumar Pal. “Hamming code and logistic-map based pixel-level active forgery detection scheme using fragile watermarking”. In: *MULTIMEDIA TOOLS AND APPLICATIONS* (2020).

[49] Xinhui Gong et al. “A secure image authentication scheme based on dual fragile watermark”. In: *Multimedia Tools and Applications* (2020), pp. 1–18.

[50] Zhongwei He et al. “Digital image splicing detection based on Markov features in DCT and DWT domain”. In: *Pattern recognition* 45.12 (2012), pp. 4292–4299.

[51] Saba Mushtaq and Ajaz Hussain Mir. “Novel method for image splicing detection”. In: *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE. 2014, pp. 2398–2403.

[52] Timo Ojala, Matti Pietikäinen, and David Harwood. “A comparative study of texture measures with classification based on featured distributions”. In: *Pattern recognition* 29.1 (1996), pp. 51–59.

[53] Ghulam Muhammad et al. “Image forgery detection using steerable pyramid transform and local binary pattern”. In: *Machine Vision and Applications* 25.4 (2014), pp. 985–995.

[54] Fahime Hakimi, Mahdi Hariri, and Farhad GharehBaghi. "Image splicing forgery detection using local binary pattern and discrete wavelet transform". In: *2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI)*. IEEE. 2015, pp. 1074–1077.

[55] Amani Alahmadi et al. "Passive detection of image forgery using DCT and local binary pattern". In: *Signal, Image and Video Processing* 11.1 (2017), pp. 81–88.

[56] Xudong Zhao et al. "Passive image-splicing detection by a 2-D noncausal Markov model". In: *IEEE Transactions on Circuits and Systems for Video Technology* 25.2 (2014), pp. 185–199.

[57] Qingbo Zhang, Wei Lu, and Jian Weng. "Joint image splicing detection in dct and contourlet transform domain". In: *Journal of Visual Communication and Image Representation* 40 (2016), pp. 449–458.

[58] Minh N Do and Martin Vetterli. "The contourlet transform: an efficient directional multiresolution image representation". In: *IEEE Transactions on image processing* 14.12 (2005), pp. 2091–2106.

[59] TG Dietterich et al. "Ensemble learning. The handbook of brain theory and neural networks". In: *Arbib MA* (2002).

[60] Ce Li et al. "Image splicing detection based on Markov features in QDCT domain". In: *Neurocomputing* 228 (2017), pp. 29–36.

[61] Ruxin Wang et al. "Digital image splicing detection based on Markov features in QDCT and QWT domain". In: *Digital Forensics and Forensic Investigations: Breakthroughs in Research and Practice*. IGI Global, 2020, pp. 61–79.

[62] Xuanjing Shen, Zenan Shi, and Haipeng Chen. "Splicing image forgery detection using textural features based on the grey level co-occurrence matrices". In: *IET Image Processing* 11.1 (2017), pp. 44–53.

[63] Xiaofeng Wang et al. "Coarse-to-fine Grained Image Splicing Localization Method Based on Noise Level Inconsistency". In: *2020 International Conference on Computing, Networking and Communications (ICNC)*. IEEE. 2020, pp. 79–83.

[64] Navdeep Kanwal et al. "Digital image splicing detection technique using optimal threshold based local ternary pattern". In: *Multimedia Tools and Applications* 79.19 (2020), pp. 12829–12846.

[65] Yuan Rao and Jiangqun Ni. "A deep learning approach to detection of splicing and copy-move forgeries in images". In: *2016 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE. 2016, pp. 1–6.

[66] Jason Bunk et al. "Detection and localization of image forgeries using resampling features and deep learning". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE. 2017, pp. 1881–1889.

[67] Bo Liu and Chi-Man Pun. "Locating splicing forgery by fully convolutional networks and conditional random field". In: *Signal Processing: Image Communication* 66 (2018), pp. 103–112.

[68] Ronald Salloum, Yuzhuo Ren, and C-C Jay Kuo. "Image splicing localization using a multi-task fully convolutional network (MFCN)". In: *Journal of Visual Communication and Image Representation* 51 (2018), pp. 201–209.

[69] Beijing Chen et al. “An improved splicing localization method by fully convolutional networks”. In: *IEEE Access* 6 (2018), pp. 69472–69480.

[70] Bin Xiao et al. “Image splicing forgery detection combining coarse to refined convolutional neural network and adaptive clustering”. In: *Information Sciences* 511 (2020), pp. 172–191.

[71] Yuan Rao, Jiangqun Ni, and Huimin Zhao. “Deep learning local descriptor for image splicing detection and localization”. In: *IEEE Access* 8 (2020), pp. 25611–25625.

[72] Belal Ahmed, T Aaron Gulliver, et al. “Image splicing detection using mask-RCNN”. In: *Signal, Image and Video Processing* (2020), pp. 1–8.

[73] Kunj Bihari Meena and Vipin Tyagi. “A Deep Learning based Method for Image Splicing Detection”. In: *Journal of Physics: Conference Series*. Vol. 1714. 1. IOP Publishing. 2021, p. 012038.

[74] Chien-Chang Chen, Wei-Yu Lu, and Chung-Hsuan Chou. “Rotational copy-move forgery detection using SIFT and region growing strategies”. In: *Multimedia Tools and Applications* 78.13 (2019), pp. 18293–18308.

[75] Guonian Jin and Xiaoxia Wan. “An improved method for SIFT-based copy–move forgery detection using non-maximum value suppression and optimized J-Linkage”. In: *Signal Processing: Image Communication* 57 (2017), pp. 113–125.

[76] S Dhivya, J Sangeetha, and B Sudhakar. “Copy-move forgery detection using SURF feature extraction and SVM supervised learning technique”. In: *Soft Computing* (2020), pp. 1–12.

[77] VT Manu and Babu M Mehtre. “Detection of copy-move forgery in images using segmentation and SURF”. In: *Advances in signal processing and intelligent recognition systems*. Springer, 2016, pp. 645–654.

[78] Chengyou Wang, Zhi Zhang, and Xiao Zhou. “An Image Copy-Move Forgery Detection Scheme Based on A-KAZE and SURF Features”. In: *Symmetry* 10.12 (2018), p. 706.

[79] Chengyou Wang et al. “An Image Copy-Move Forgery Detection Method Based on SURF and PCET”. In: *IEEE Access* 7 (2019), pp. 170032–170047.

[80] Liyang Yu, Qi Han, and Xiamu Niu. “Feature point-based copy-move forgery detection: covering the non-textured areas”. In: *Multimedia Tools and Applications* 75.2 (2014), pp. 1159–1176.

[81] Jingwei Li et al. “Keypoint-based copy-move detection scheme by adopting MSCRs and improved feature matching”. In: *Multimedia Tools and Applications* 76.20 (2016), pp. 20483–20497.

[82] Seung-Jin Ryu, Min-Jeong Lee, and Heung-Kyu Lee. “Detection of Copy-Rotate-Move Forgery Using Zernike Moments”. In: *Information Hiding Lecture Notes in Computer Science* (2010), pp. 51–65.

[83] Ye Zhu, Xuanjing Shen, and Haipeng Chen. “Copy-move forgery detection based on scaled ORB”. In: *Multimedia Tools and Applications* 75.6 (2015), pp. 3221–3233.

[84] Guonian Jin and Xiaoxia Wan. "An improved method for SIFT-based copy–move forgery detection using non-maximum value suppression and optimized J-Linkage". In: *Signal Processing: Image Communication* 57 (2017), pp. 113–125.

[85] Cong Lin et al. "Region duplication detection based on hybrid feature and evaluative clustering". In: *Multimedia Tools and Applications* 78.15 (2019), pp. 20739–20763.

[86] Irene Amerini et al. "A sift-based forensic method for copy–move attack detection and transformation recovery". In: *IEEE transactions on information forensics and security* 6.3 (2011), pp. 1099–1110.

[87] Krystian Mikolajczyk and Cordelia Schmid. "Indexing based on scale invariant interest points". In: *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001.* Vol. 1. IEEE. 2001, pp. 525–531.

[88] Krystian Mikolajczyk and Cordelia Schmid. "Scale & affine invariant interest point detectors". In: *International journal of computer vision* 60.1 (2004), pp. 63–86.

[89] Rahul Dixit and Ruchira Naskar. "Region duplication detection in digital images based on Centroid Linkage Clustering of key–points and graph similarity matching". In: *Multimedia Tools and Applications* 78.10 (2019), pp. 13819–13840.

[90] Xiuli Bi, Chi-Man Pun, and Xiao-Chen Yuan. "Multi-level dense descriptor and hierarchical feature matching for copy–move forgery detection". In: *Information Sciences* 345 (2016), pp. 226–242.

[91] Mahmoud Emam, Qi Han, and Xiamu Niu. "PCET based copy-move forgery detection in images under geometric transforms". In: *Multimedia Tools and Applications* 75.18 (2016), pp. 11513–11527.

[92] Davide Cozzolino, Giovanni Poggi, and Luisa Verdoliva. "Efficient dense-field copy-move forgery detection". In: *IEEE Transactions on Information Forensics and Security* 10.11 (2015), pp. 2284–2297.

[93] Connelly Barnes et al. "PatchMatch: A randomized correspondence algorithm for structural image editing". In: *ACM Trans. Graph.* 28.3 (2009), p. 24.

[94] Chen-Ming Hsu, Jen-Chun Lee, and Wei-Kuei Chen. "An efficient detection algorithm for copy-move forgery". In: *2015 10th Asia Joint Conference on Information Security*. IEEE. 2015, pp. 33–36.

[95] Ce Li et al. "An image copy move forgery detection method using QDCT". In: *Proceedings of the International Conference on Internet Multimedia Computing and Service*. 2016, pp. 5–8.

[96] Tu Huynh-Kha et al. "Improving the computational cost for copied region detection in forensic images". In: *J Sci Technol: Issue Inf Commun Technol* 2.1 (2016), pp. 55–63.

[97] Choudhary Shyam Prakash, Kumar Vijay Anand, and Sushila Maheshkar. "Detection of copy-move image forgery using DCT". In: *Advances in Computational Intelligence*. Springer, 2017, pp. 257–265.

[98] Vivek H Mahale et al. "Image inconsistency detection using local binary pattern (LBP)". In: *Procedia computer science* 115 (2017), pp. 501–508.

[99] Yuecong Lai et al. “An improved block-based matching algorithm of copy-move forgery detection”. In: *Multimedia Tools and Applications* 77.12 (2018), pp. 15093–15110.

[100] Pew-Thian Yap, Xudong Jiang, and Alex Chichung Kot. “Two-dimensional polar harmonic transforms for invariant image representation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.7 (2009), pp. 1259–1270.

[101] Ming-Kuei Hu. “Visual pattern recognition by moment invariants”. In: *IRE transactions on information theory* 8.2 (1962), pp. 179–187.

[102] Jiangbin Zheng et al. “Fusion of block and keypoints based approaches for effective copy-move image forgery detection”. In: *Multidimensional Systems and Signal Processing* 27.4 (2016), pp. 989–1005.

[103] Kunj Bihari Meena and Vipin Tyagi. “A hybrid copy-move image forgery detection technique based on Fourier-Mellin and scale invariant feature transforms”. In: *Multimedia Tools and Applications* 79.11 (2020), pp. 8197–8212.

[104] Sreenivasu Tinnathi and G Sudhavani. “An efficient copy move forgery detection using adaptive watershed segmentation with AGSO and hybrid feature extraction”. In: *Journal of Visual Communication and Image Representation* 74 (2021), p. 102966.

[105] Patrick Niyishaka and Chakravarthy Bhagvati. “Copy-move forgery detection using image blobs and BRISK feature”. In: *Multimedia Tools and Applications* 79.35 (2020), pp. 26045–26059.

[106] David Vázquez-Padın, Pedro Comesana, and Fernando Pérez-González. “An SVD approach to forensic image resampling detection”. In: *2015 23rd European signal processing conference (EUSIPCO)*. IEEE. 2015, pp. 2067–2071.

[107] David Vázquez-Padın, Fernando Pérez-González, and Pedro Comesana-Alfaro. “A random matrix approach to the forensic analysis of upscaled images”. In: *IEEE Transactions on Information Forensics and Security* 12.9 (2017), pp. 2115–2130.

[108] Yuting Su et al. “Hierarchical image resampling detection based on blind deconvolution”. In: *Journal of Visual Communication and Image Representation* 48 (2017), pp. 480–490.

[109] Tong Qiao, Aichun Zhu, and Florent Retraint. “Exposing image resampling forgery by using linear parametric model”. In: *Multimedia Tools and Applications* 77.2 (2018), pp. 1501–1523.

[110] Alaa Hilal. “Image re-sampling detection through a novel interpolation kernel”. In: *Forensic science international* 287 (2018), pp. 25–35.

[111] Tong Qiao et al. “Statistical model-based detector via texture weight map: Application in re-sampling authentication”. In: *IEEE Transactions on Multimedia* 21.5 (2018), pp. 1077–1092.

[112] Belhassen Bayar and Matthew C Stamm. “On the robustness of constrained convolutional neural networks to jpeg post-compression for image resampling detection”. In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2017, pp. 2152–2156.

[113] Belhassen Bayar and Matthew C Stamm. "Constrained convolutional neural networks: A new approach towards general purpose image manipulation detection". In: *IEEE Transactions on Information Forensics and Security* 13.11 (2018), pp. 2691–2706.

[114] Jason Bunk et al. "Detection and localization of image forgeries using resampling features and deep learning". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE. 2017, pp. 1881–1889.

[115] Arjuna Flenner et al. "Resampling forgery detection using deep learning and a-contrario analysis". In: *Electronic Imaging* 2018.7 (2018), pp. 212–1.

[116] Gang Cao et al. "Resampling detection of recompressed images via dual-stream convolutional neural network". In: *arXiv preprint arXiv:1901.04637* (2019).

[117] Anatol Maier, Benedikt Lorch, and Christian Riess. "Toward Reliable Models For Authenticating Multimedia Content: Detecting Resampling Artifacts With Bayesian Neural Networks". In: *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2020, pp. 1251–1255.

[118] Gang Cao, Yao Zhao, and Rongrong Ni. "Forensic estimation of gamma correction in digital images". In: *2010 IEEE International Conference on Image Processing*. IEEE. 2010, pp. 2097–2100.

[119] Longyin Wen, Honggang Qi, and Siwei Lyu. "Contrast enhancement estimation for digital image forensics". In: *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 14.2 (2018), pp. 1–21.

[120] Gang Cao et al. "Contrast enhancement-based forensics in digital images". In: *IEEE transactions on information forensics and security* 9.3 (2014), pp. 515–525.

[121] Matthew C Stamm and KJ Ray Liu. "Forensic detection of image manipulation using statistical intrinsic fingerprints". In: *IEEE Transactions on Information Forensics and Security* 5.3 (2010), pp. 492–506.

[122] Neetu Singh, Abhinav Gupta, and Roop Chand Jain. "Global contrast enhancement based image forensics using statistical features". In: *Advances in Electrical and Electronic Engineering* 15.3 (2017), pp. 509–516.

[123] Matthias Kirchner and Jessica Fridrich. "On detection of median filtering in digital images". In: *Media forensics and security II*. Vol. 7541. International Society for Optics and Photonics. 2010, p. 754110.

[124] Hai-Dong Yuan. "Blind forensics of median filtering in digital images". In: *IEEE Transactions on Information Forensics and Security* 6.4 (2011), pp. 1335–1345.

[125] Chenglong Chen, Jiangqun Ni, and Jiwu Huang. "Blind detection of median filtering in digital images: A difference domain based approach". In: *IEEE Transactions on Image Processing* 22.12 (2013), pp. 4699–4710.

[126] Gang Cao et al. "Forensic detection of median filtering in digital images". In: *2010 IEEE International Conference on Multimedia and Expo*. IEEE. 2010, pp. 89–94.

[127] Xiangui Kang et al. "Robust median filtering forensics using an autoregressive model". In: *IEEE Transactions on Information Forensics and Security* 8.9 (2013), pp. 1456–1468.

[128] Jianquan Yang et al. “Detecting median filtering via two-dimensional AR models of multiple filtered residuals”. In: *Multimedia Tools and Applications* 77.7 (2018), pp. 7931–7953.

[129] Xinlu Gui et al. “Blind median filtering detection based on histogram features”. In: *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*. IEEE. 2014, pp. 1–4.

[130] Yujin Zhang et al. “Revealing the traces of median filtering using high-order local ternary patterns”. In: *IEEE Signal Processing Letters* 21.3 (2014), pp. 275–279.

[131] Yakun Niu, Yao Zhao, and Rongrong Ni. “Robust median filtering detection based on local difference descriptor”. In: *Signal Processing: Image Communication* 53 (2017), pp. 65–72.

[132] Anan Liu et al. “Median filtering forensics in digital images based on frequency-domain features”. In: *Multimedia tools and applications* 76.21 (2017), pp. 22119–22132.

[133] Zunli Hu and Shilin Wang. “Median filtering forensics based on discriminative multi-scale sparse coding”. In: *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE. 2017, pp. 141–145.

[134] Jiansheng Chen et al. “Median filtering forensics based on convolutional neural networks”. In: *IEEE Signal Processing Letters* 22.11 (2015), pp. 1849–1853.

[135] Aparna Bharati et al. “Detecting facial retouching using supervised deep learning”. In: *IEEE Transactions on Information Forensics and Security* 11.9 (2016), pp. 1903–1913.

[136] Aparna Bharati et al. “Demography-based facial retouching detection using subclass supervised sparse autoencoder”. In: *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE. 2017, pp. 474–482.

[137] Anubhav Jain, Richa Singh, and Mayank Vatsa. “On detecting gans and retouching based synthetic alterations”. In: *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE. 2018, pp. 1–7.

[138] Anubhav Jain et al. “Detecting GANs and retouching based digital alterations via DAD-HCNN”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2020, pp. 672–673.

[139] Sheng-Yu Wang et al. “Detecting photoshopped faces by scripting photoshop”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 10072–10081.

[140] Christian Rathgeb et al. “Differential detection of facial retouching: A multi-biometric approach”. In: *IEEE Access* 8 (2020), pp. 106373–106385.

[141] Chang-Hee Choi et al. “Forged region detection for scanned images”. In: *Computer Science and Convergence*. Springer, 2012, pp. 687–694.

[142] Zeinab F Elsharkawy et al. “Accurate and robust identifying forged region method in scanned images”. In: *International Journal of Computer Applications* 83.1 (2013).

[143] Zeinab F Elsharkawy et al. "New and efficient blind detection algorithm for digital image forgery using homomorphic image processing". In: *Multimedia Tools and Applications* 78.15 (2019), pp. 21585–21611.

[144] Sofian Rizal. "Compression Method in JPEG Standard". In: *IOP Conference Series: Materials Science and Engineering*. Vol. 553. 1. IOP Publishing. 2019, p. 012013.

[145] Matthew C Stamm and KJ Ray Liu. "Anti-forensics of digital image compression". In: *IEEE Transactions on Information Forensics and Security* 6.3 (2011), pp. 1050–1065.

[146] Vladimir I Ivanov and John S Baras. "Authentication of fingerprint scanners". In: *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2011, pp. 1912–1915.

[147] Mei Yu et al. "Improved WαsH feature matching based on 2D-DWT for stereo remote sensing images". In: *Sensors* 18.10 (2018), p. 3494.

[148] Simon Prikler and Jürgen W Einax. "Improving detection power in trace analysis using wavelet transform". In: *Analytical and bioanalytical chemistry* 403.9 (2012), pp. 2563–2567.

[149] Ingrid Daubechies. *Ten lectures on wavelets*. SIAM, 1992.

[150] S Grace Chang, Bin Yu, and Martin Vetterli. "Adaptive wavelet thresholding for image denoising and compression". In: *IEEE transactions on image processing* 9.9 (2000), pp. 1532–1546.

[151] Karnran Sharifi and Alberto Leon-Garcia. "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video". In: *IEEE Transactions on Circuits and Systems for Video Technology* 5.1 (1995), pp. 52–56.

[152] Kostas Kokkinakis and Asoke K Nandi. "Exponent parameter estimation for generalized Gaussian probability density functions with application to speech modeling". In: *Signal Processing* 85.9 (2005), pp. 1852–1858.

[153] Minh N Do and Martin Vetterli. "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance". In: *IEEE transactions on image processing* 11.2 (2002), pp. 146–158.

[154] Johan Verbeke and Ronald Cools. "The Newton-Raphson method". In: *International Journal of Mathematical Education in Science and Technology* 26.2 (1995), pp. 177–193.

[155] Solomon Kullback and Richard A Leibler. "On information and sufficiency". In: *The annals of mathematical statistics* 22.1 (1951), pp. 79–86.

[156] Mitchell Cochran. "A proposed standard procedure to define minimum scanning attribute levels for hard copy documents". In: *2014 47th Hawaii International Conference on System Sciences*. IEEE. 2014, pp. 2036–2043.

[157] Keith Jack. "Video Demystified-forth edition". In: *Newnes: Elsevier* (2007).

[158] Arthur Ardeshir Goshtasby. *2-D and 3-D image registration: for medical, remote sensing, and industrial applications*. John Wiley & Sons, 2005.

[159] Herbert Bay et al. "Speeded-up robust features (SURF)". In: *Computer vision and image understanding* 110.3 (2008), pp. 346–359.

[160] Marko Helén and Tuomas Virtanen. "Audio query by example using similarity measures between probability density functions of features". In: *EURASIP Journal on Audio, Speech, and Music Processing* 2010 (2009), pp. 1–12.

[161] DT Mane and Uday V Kulkarni. "A survey on supervised convolutional neural network and its major applications". In: *Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications*. IGI Global, 2020, pp. 1058–1071.

[162] Luca Baroffio et al. "Camera identification with deep convolutional networks". In: *arXiv preprint arXiv:1603.01068* (2016).

[163] Na Huang et al. "Identification of the source camera of images based on convolutional neural network". In: *Digital Investigation* 26 (2018), pp. 72–80.

[164] David Freire-Obregón et al. "Deep learning for source camera identification on mobile devices". In: *Pattern Recognition Letters* 126 (2019), pp. 86–91.

[165] Hongwei Yao et al. "Robust multi-classifier for camera model identification based on convolution neural network". In: *IEEE Access* 6 (2018), pp. 24973–24982.

[166] Yunshu Chen, Yue Huang, and Xinghao Ding. "Camera model identification with residual neural network". In: *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2017, pp. 4337–4341.

[167] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems* 25 (2012), pp. 1097–1105.

[168] Christian Szegedy et al. "Going deeper with convolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9.

[169] Abdul Muntakim Rafi et al. "Application of DenseNet in Camera Model Identification and Post-processing Detection." In: *CVPR workshops*. 2019, pp. 19–28.

[170] Artur Kuzin et al. "Camera model identification using convolutional neural networks". In: *2018 IEEE International Conference on Big Data (Big Data)*. IEEE. 2018, pp. 3107–3110.

[171] Anselmo Ferreira et al. "An inception-based data-driven ensemble approach to camera model identification". In: *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE. 2018, pp. 1–7.

[172] Md Hasan Al Banna et al. "Camera model identification using deep CNN and transfer learning approach". In: *2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*. IEEE. 2019, pp. 626–630.

[173] Xinghao Ding et al. "Camera identification based on domain knowledge-driven deep multi-task learning". In: *IEEE Access* 7 (2019), pp. 25878–25890.

[174] Amel Tuama, Frédéric Comby, and Marc Chaumont. "Camera model identification with the use of deep convolutional neural networks". In: *2016 IEEE International workshop on information forensics and security (WIFS)*. IEEE. 2016, pp. 1–6.

[175] Belhassen Bayar and Matthew C Stamm. "Augmented convolutional feature maps for robust cnn-based camera model identification". In: *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2017, pp. 4098–4102.

[176] Bo Wang et al. “Source camera model identification based on convolutional neural networks with local binary patterns coding”. In: *Signal Processing: Image Communication* 68 (2018), pp. 162–168.

[177] Abdul Muntakim Rafi et al. “RemNet: remnant convolutional neural network for camera model identification”. In: *Neural Computing and Applications* 33.8 (2021), pp. 3655–3670.

[178] Changhee Kang and Sang-ug Kang. “Camera model identification using a deep network and a reduced edge dataset”. In: *Neural Computing and Applications* 32.17 (2020), pp. 13139–13146.

[179] Mo Chen, Jessica Fridrich, and Miroslav Goljan. “Digital imaging sensor identification (further study)”. In: *Security, steganography, and watermarking of multimedia contents IX*. Vol. 6505. International Society for Optics and Photonics. 2007, 65050P.

[180] Luca Bondi et al. “First steps toward camera model identification with convolutional neural networks”. In: *IEEE Signal Processing Letters* 24.3 (2016), pp. 259–263.

[181] Luca Bondi et al. “A preliminary study on convolutional neural networks for camera model identification”. In: *Electronic Imaging* 2017.7 (2017), pp. 67–76.

[182] David Güera et al. “Reliability map estimation for CNN-based camera model attribution”. In: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE. 2018, pp. 964–973.

[183] Pengpeng Yang et al. “Source camera identification based on content-adaptive fusion residual networks”. In: *Pattern Recognition Letters* 119 (2019), pp. 195–204.

[184] Ruiting Shao and Edward J Delp. “Forensic scanner identification using machine learning”. In: *2020 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*. IEEE. 2020, pp. 1–4.

[185] François Chollet. “Xception: Deep learning with depthwise separable convolutions”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1251–1258.

[186] James E Fowler. “The redundant discrete wavelet transform and additive noise”. In: *IEEE Signal Processing Letters* 12.9 (2005), pp. 629–632.

[187] Ramesh Kumar and Prabhat Patel. “Signal denoising with interval dependent thresholding using DWT and SWT”. In: *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* 1.6 (2012), pp. 47–52.

[188] Nitish Srivastava et al. “Dropout: a simple way to prevent neural networks from overfitting”. In: *The journal of machine learning research* 15.1 (2014), pp. 1929–1958.

[189] François Chollet et al. *keras*. 2015.

[190] Tijmen Tieleman, Geoffrey Hinton, et al. “Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude”. In: *COURSERA: Neural networks for machine learning* 4.2 (2012), pp. 26–31.

[191] Juan D Rodriguez, Aritz Perez, and Jose A Lozano. “Sensitivity analysis of k-fold cross validation in prediction error estimation”. In: *IEEE transactions on pattern analysis and machine intelligence* 32.3 (2009), pp. 569–575.

[192] Gaël Varoquaux. “Cross-validation failure: small sample sizes lead to large error bars”. In: *Neuroimage* 180 (2018), pp. 68–77.

[193] Yulan Li et al. “Speech emotion recognition using 1d cnn with no attention”. In: *2019 23rd international computer science and engineering conference (ICSEC)*. IEEE. 2019, pp. 351–356.

[194] Jianfeng Zhao, Xia Mao, and Lijiang Chen. “Speech emotion recognition using deep 1D & 2D CNN LSTM networks”. In: *Biomedical Signal Processing and Control* 47 (2019), pp. 312–323.

[195] Ayodeji Olalekan Salau, Tilewa David Olowoyo, and Solomon Oluwole Akinola. “Accent classification of the three major nigerian indigenous languages using 1d cnn lstm network model”. In: *Advances in Computational Intelligence Techniques*. Springer, Singapore, 2020, pp. 1–16.

[196] Muhammad Gumilang and Ayu Purwarianti. “Experiments on character and word level features for text classification using deep neural network”. In: *2018 Third International Conference on Informatics and Computing (ICIC)*. IEEE. 2018, pp. 1–6.

[197] Xiaohang Xu et al. “SDD-CNN: Small data-driven convolution neural networks for subtle roller defect inspection”. In: *Applied Sciences* 9.7 (2019), p. 1364.

[198] Yichuan Tang. “Deep learning using linear support vector machines”. In: *arXiv preprint arXiv:1306.0239* (2013).

[199] Martın Abadi et al. “Tensorflow: A system for large-scale machine learning”. In: *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*. 2016, pp. 265–283.

[200] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).

[201] Fabian Pedregosa et al. “Scikit-learn: Machine learning in Python”. In: *the Journal of machine Learning research* 12 (2011), pp. 2825–2830.

[202] Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. “Pointwise shape-adaptive DCT denoising with structure preservation in luminance-chrominance space”. In: *Proc. of the 2nd international workshop on video processing and quality metrics for consumer electronics, VPQM*. 2006.

[203] Miroslav Goljan, Mo Chen, and Jessica Fridrich. “Identifying common source digital camera from image pairs”. In: *2007 IEEE International Conference on Image Processing*. Vol. 6. IEEE. 2007, pp. VI–125.

[204] Filipe de O Costa et al. “Open set source camera attribution and device linking”. In: *Pattern Recognition Letters* 39 (2014), pp. 92–101.

[205] BVK Vijaya Kumar and Laurence Hassebrook. “Performance measures for correlation filters”. In: *Applied optics* 29.20 (1990), pp. 2997–3006.

[206] Owen Mayer and Matthew C Stamm. “Learned forensic source similarity for unknown camera models”. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2018, pp. 2012–2016.

[207] Owen Mayer and Matthew C Stamm. “Forensic similarity for digital images”. In: *IEEE Transactions on Information Forensics and Security* 15 (2019), pp. 1331–1346.

[208] Guru Swaroop Bennabhaktula et al. "Device-based image matching with similarity learning by convolutional neural networks that exploit the underlying camera sensor pattern noise". In: *arXiv preprint arXiv:2004.11443* (2020).

[209] Belhassen Bayar and Matthew C Stamm. "Design principles of convolutional neural networks for multimedia forensics". In: *Electronic Imaging* 2017.7 (2017), pp. 77–86.

[210] Shuo Yang et al. "From facial parts responses to face detection: A deep learning approach". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 3676–3684.

[211] Pedro Savarese and Michael Maire. "Learning implicitly recurrent CNNs through parameter sharing". In: *arXiv preprint arXiv:1902.09701* (2019).

[212] R Caruana. "Multi-task learning Machine Learning 28 (1), 41-75 (1997) 10.1023". In: *A: 1007379606734* ().

[213] Jürgen Groß. *Linear regression*. Vol. 175. Springer Science & Business Media, 2012.

[214] Gérard Biau and Erwan Scornet. "A random forest guided tour". In: *Test* 25.2 (2016), pp. 197–227.

[215] Yann A LeCun et al. "Efficient backprop". In: *Neural networks: Tricks of the trade*. Springer, 2012, pp. 9–48.

[216] Thomas Gloe and Rainer Böhme. "The'Dresden Image Database'for benchmarking digital image forensics". In: *Proceedings of the 2010 ACM Symposium on Applied Computing*. 2010, pp. 1584–1590.

[217] Duc-Tien Dang-Nguyen et al. "Raise: A raw images dataset for digital image forensics". In: *Proceedings of the 6th ACM multimedia systems conference*. 2015, pp. 219–224.

[218] Jayanthi Sivaswamy et al. "Drishti-gs: Retinal image dataset for optic nerve head (onh) segmentation". In: *2014 IEEE 11th international symposium on biomedical imaging (ISBI)*. IEEE. 2014, pp. 53–56.

[219] Michael A Schuh et al. "A large-scale solar image dataset with labeled event regions". In: *2013 IEEE International Conference on Image Processing*. IEEE. 2013, pp. 4349–4353.

[220] Chen Chen, Roozbeh Jafari, and Nasser Kehtarnavaz. "UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor". In: *2015 IEEE International conference on image processing (ICIP)*. IEEE. 2015, pp. 168–172.

[221] Zeinab F Elsharkawy et al. "New and efficient blind detection algorithm for digital image forgery using homomorphic image processing". In: *Multimedia Tools and Applications* 78.15 (2019), pp. 21585–21611.

[222] Thomas Kailath. "The divergence and Bhattacharyya distance measures in signal selection". In: *IEEE transactions on communication technology* 15.1 (1967), pp. 52–60.

[223] Sung-Hyuk Cha. "Comprehensive survey on distance/similarity measures between probability density functions". In: *City* 1.2 (2007), p. 1.

[224] Yossi Rubner et al. “Empirical evaluation of dissimilarity measures for color and texture”. In: *Computer vision and image understanding* 84.1 (2001), pp. 25–43.

[225] Mario Vaneechoutte and Marc Heyndrickx. “Application and analysis of ARDRA patterns in bacterial identification, taxonomy and phylogeny.” In: *New approaches for analysis of microbial typing data. Editors: L. Dijkshoorn, K. Towner, and M. Struelens. ISBN Hardbound 0444 507 40X., 2001. Elsevier. 357 p. Chapter 9.* 2001, pp. 211–247.

[226] Murat Gül and Emin Kugu. “A survey on anti-forensics techniques”. In: *2017 International Artificial Intelligence and Data Processing Symposium (IDAP).* IEEE. 2017, pp. 1–6.

[227] Jia Deng et al. “Imagenet: A large-scale hierarchical image database”. In: *2009 IEEE conference on computer vision and pattern recognition.* Ieee. 2009, pp. 248–255.

[228] Serkan Kiranyaz et al. “1D convolutional neural networks and applications: A survey”. In: *Mechanical systems and signal processing* 151 (2021), p. 107398.

DOCTORAT BRETAGNE LOIRE MATHSTIC

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

SUP'COM
Ecole Supérieure des Communications de Tunis

Titre : Analyse des documenents numérisés à des fins de contrôle d'intégrité et d'authenticité

Mot clés : Forensiques des images numériques, intégrité de l'image, authenticité, scanners à plat, falsification, documents numérisés, mécanisme de sécurité

Résumé : Aucune méthode universelle de détection des contrefaçons d'images n'existe. Plusieurs techniques ont été proposées mais chacune a ses limites. Parmi ces méthodes, les techniques dites "forensiques des images numériques" offrent une solution intéressante. Elles visent soit à vérifier l'intégrité de l'image, soit à apporter la preuve de son authenticité en identifiant le système qui l'a acquise. Pour ce faire, elles tirent avantage de la manière dont les systèmes d'acquisition génèrent leur sortie. Dans cette thèse, nous nous intéressons en particulier aux scanners à plat comme système d'acquisition et nous proposons d'étudier et de développer des techniques de forensiques pour des documents multi-type scannés. Nous avons, tout d'abord, proposé des techniques permettant d'identifier le scanner à l'origine d'un document scanné sur la base d'un ensemble de signatures extraites "manuellement" des images. Ensuite, pour faire face aux limites de ces approches, nous nous somme focalisé sur l'extraction automatique des signatures des scanners à travers des réseaux de neurones 1D et 2D. Par la suite, nous avons développé une nouvelle approche, appelée "Device Linking", qui détermine si deux images ont été acquises par le même scanner ou non. Enfin, nous fournissons deux mécanismes de sécurité capable de détecter les manipulations du contenu des données numérisées en se basant sur certaines approches d'identification du scanner source proposées précédemment. Dans le but de valider les solutions proposées dans des situations réelles et réaliser des comparaisons entre elles, nous avons construit une base de données de documents scannés que nous avons rendue publique.

Title: Analysis of scanned documents for integrity and authenticity checking

Keywords: Digital image forensics, image integrity, authenticity, flatbed scanners, falsification, digitized documents, security mechanism

Abstract: There is no universal method for detecting counterfeit images. Several techniques have been proposed but each has its limits. Among these methods, the so-called "forensic digital image" techniques offer an interesting solution. They aim either to verify the integrity of the image, or to provide proof of its authenticity by identifying the system which acquired it. To do this, they take advantage of how the acquisition systems generate their output. In this thesis, we are particularly interested in flatbed scanners as an acquisition system and we propose to study and develop "digital image forensics" techniques for scanned multi-type documents. We first proposed techniques to identify the scanner behind a scanned document based on a set of signatures "manually" extracted from images. Then, to face the limitations of these approaches, we focused on the automatic extraction of signatures from scanners through 1D and 2D neural networks. Subsequently, we developed a new approach, called "Device Linking", which determines whether two images were acquired by the same scanner or not. Finally, we provide two security mechanisms capable of detecting content manipulation of scanned data based on certain approaches of source scanner identification proposed previously. In order to validate the solutions proposed in real situations and make comparisons between them, we have built a database of scanned documents that we have made public.