

Enhancement of Footwear Impressions

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Visual Computing

eingereicht von

Rebeka Koszticsak, Bsc

Matrikelnummer 01325492

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Robert Sablatnig

Mitwirkung: Associate Prof. Hideki Nakayama, Ph.D.

Projektass. Dipl.-Ing. Manuel Keglevic

Wien, 4. Jänner 2020

Rebeka Koszticsak

Robert Sablatnig

Enhancement of Footwear Impressions

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Visual Computing

by

Rebeka Koszticsak, Bsc

Registration Number 01325492

to the Faculty of Informatics

at the TU Wien

Advisor: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Robert Sablatnig

Assistance: Associate Prof. Hideki Nakayama, Ph.D.

Projektass. Dipl.-Ing. Manuel Keglevic

Vienna, 4th January, 2020

Rebeka Koszticsak

Robert Sablatnig

Erklärung zur Verfassung der Arbeit

Rebeka Koszticsak, Bsc
Address

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 4. Jänner 2020

Rebeka Koszticsak

Danksagung

Ich möchte mich bei Professor Sablatnig und bei Professor Nakayama bedanken, weil Sie mir erlaubt haben auf der University of Tokyo zu studieren, und Sie mein Projekt unterstützt haben.

Ich möchte mich bei Eszter Takács bedanken für das Korrekturlesen und weil sie in der schlimmsten Zeit bei mir gestanden ist. Außerdem möchte ich mich bei ihr und bei Peter Pollak bedanken weil sie meine Stimmungsschwankungen ausgehalten haben.

Végül, a szüleimnek szeretném megköszönni, hogy megtanították, hogy mi az, ami igazán számít.

Acknowledgements

I would like to thank to Professor Sablatnig and to Professor Nakayama for allowing me to study at the University of Tokyo and for supporting my project.

I would like to thank to Eszter Takács for proofreading and for holding my hands in the worst times. Furthermore, I would like to thank her and to Peter Pollak for taking my mood swings during the last semester.

Finally, I want to thank to my parents for teaching me what really matters.

Kurzfassung

Obwohl die Erkennung von Schuhabdrücke von Tatorten ein hoch erforschtes Thema ist, erfolgt die endgültige Identifizierung durch forensische Experten. Ausserdem ist das Wissen über Preprocessing und Enhancement Techniken um die Matching Genauigkeit zu erhöhen limitiert.

Diese Arbeit untersucht Möglichkeiten zur Verbesserung von Schuhabdrucksbildern aus einem Datensatz mit forensischen Bildern aus der Praxis. Die Methodik von drei Ansätzen werden vorgestellt und bewertet. Eine Herausforderung bei automatischer Analyse ist das Muster der Schuhsole unabhängig von dem vielseitigen, möglicherweise stark strukturierten und überfüllten Rauschen, korrekt zu filtern. Neben vollautomatisierten Methoden wird auch eine halbautomatische Technik getestet, bei der Benutzereingaben für die Entfernung von Rauschen und Verzerrungen erforderlich sind.

Ziel dieser Arbeit ist es, einen State- of-the-Art Überblick zu geben und einen Algorithmus zu definieren, mit dem die Schuhabdrucksbilder trotz Rauschen und variierender Bildqualität gefiltert und verbessert werden können. Zusammen mit zwei vollautomatischen Algorithmen wird eine halbautomatische Methode zur Rauschunterdrückung und Verbesserung der Schuhabdrucksbilder vorgeschlagen. Die halbautomatische Methodik in der Arbeit identifiziert verrauschte Pixel basierend auf den Fourier-Mellin-Merkmalen eines von dem Benutzer ausgewählten Hintergrundpixels und dessen Nachbarschaft. Gleichzeitig wird ein Rauschmodell berechnet, um diese Struktur auch von dem restlichen Bild, also von dem Schuhabdruckmuster, zu eliminieren. Zusätzlich werden die Pixel basierend auf ihren Gradienten gruppiert und der Ansatz von "non-local means" wird angewendet um den Kontrast zwischen dem Vor- und dem Hintergrund zu erhöhen. Die experimentellen Ergebnisse zeigen, dass die verarbeiteten Bilder klarer sind, das Muster von der Schusohle schärfer ist und das Rauschen unterdrückt wird. Neben der qualitativen Bewertung werden unsere Ergebnisse auch anhand einer quantitativen Analyse bewertet und unter Verwendung von drei grundlegenden Bilddeskriptoren durchgeführt. Die Evaluierung zeigt, dass die verbesserten Bilder eine höhere Genauigkeit erzielt als die Originalbilder.

Abstract

Even though the recognition of shoeprint images secured at crime scenes is a highly researched topic, the final identification is usually done by human forensic experts. Furthermore, there is limited research available about preprocessing and enhancement techniques, to increase the matching accuracy.

This thesis investigates the possibilities for enhancement of shoeprint images from a real-life dataset, for which three prototypical approaches are presented and evaluated. The main challenge of this task is to correctly filter the shoe outsole pattern on the shoeprint image regardless of the versatile, possibly heavily structured and cluttered noise in the images. Among fully automated methods, a semi-automated technique is also tested, where user input is required for noise separation.

The aim of this work is to give an overview of the current state of the art and to define an approach which is able to filter and enhance the shoeprint despite the presence of noise and varying image quality. Along two fully-automated algorithms a semi-automated noise-suppression and enhancement pipeline for shoeprint images is introduced. The noisy pixels are identified based on the Fourier-Mellin features of a background pixel and its neighborhood selected by the user. At the same time a noise model is calculated to eliminate that structure on the remaining parts of the shoeprint image, thus on the shoeprint, as well. Additionally, the pixels are clustered based on their gradients and the non-local means approach is applied. The experimental results show that the processed images are clearer, the shoeprint is sharper and the noise is reduced. Furthermore besides a qualitative evaluation also a quantitative evaluation is performed using three basic image descriptor features. These results show that the enhanced images achieve higher accuracy than the original images.

Contents

Kurzfassung	xi
Abstract	xiii
Contents	xv
1 Introduction	1
1.1 Problem Definition	1
1.2 Challenges	2
1.3 Contribution	4
1.4 Structure of the Work	6
2 Related Work	9
2.1 Image Enhancement	9
2.2 Noise Removal	17
2.3 Image Descriptors	18
3 Methodology	23
3.1 Feature Learning for Shoeprint Detection	23
3.2 Fully-Automated Noise Elimination	28
3.3 Shoeprint Enhancement with Non-Local Means	32
4 Results and Discussion	39
4.1 Qualitative Evaluation	39
4.2 Quantitative Evaluation	54
5 Conclusion	59
Bibliography	63
Appendix Dataset	73
Appendix Results	83

Introduction

Shoeprints found at crime scenes are important clues or evidences during criminal investigation [KYZ14]. Even though at one third of the crime scenes usable shoeprint images are secured [Ale96], human input is required to identify and analyze the found shoeprint images [WSYZ14]. The work of forensic experts is not only time consuming and expensive, but there is no guarantee about the objectivness of the final outcome[GBCN08]. Furthermore, the stages of the human matching process are not necessarily reproducible considering the opinions of multiple experts.

There is an excessive amount of research already done in order to help or replace the work of forensic experts [RBCP19]. However this task is challenging because of the versatility of conditions: the features and properties of the shoe outsole, such as age, material, etc., the characteristics of the ground where the shoeprint is left and environmental conditions, for example weather, highly influence the overall appearance of the acquired sample. Figure 1.1 shows images of the same shoe outsole captured under different conditions from the FID-300 [KAV14] dataset. These factors result in changing appearance causing high intra-class variance while searching for common features of the same shoe outsole pattern. To make the features more robust and to lower the influence of those factors shoeprint enhancement is proposed.

In 2014 a dataset called FID-300 [KAV14] was released that consists of shoeprint samples collected by the police. Additionally, it contains over 1000 reference shoeprints acquired in controlled conditions. Moreover, the database introduces 300 new real-life shoeprint samples providing an insight to images forensic experts work with on a daily basis.

1.1 Problem Definition

This work focuses on ways to increase the sample quality in terms of the amount of noise in the image and the intensity difference between the fore- and the background. The

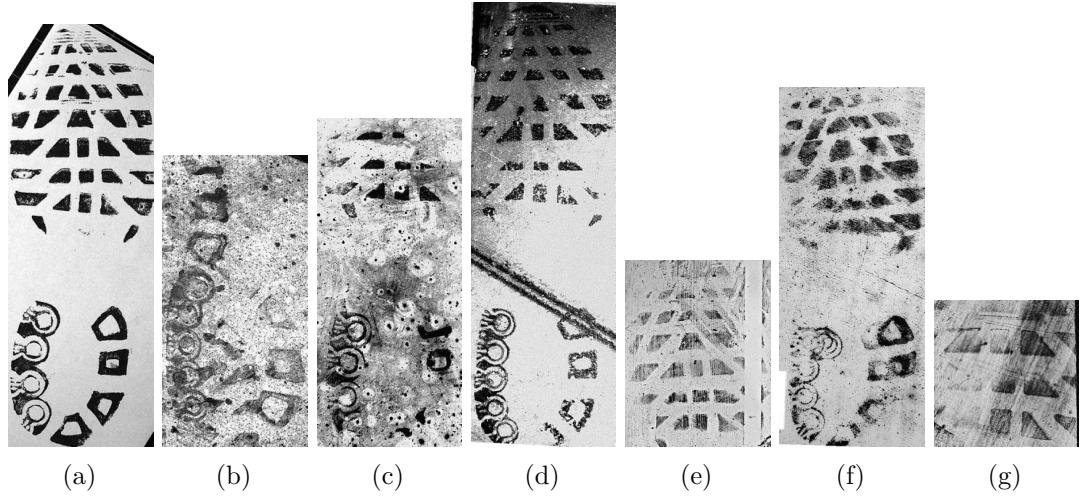


Figure 1.1: Example images from the FID-300 [KAV14] dataset, where the shoeprint is captured under different conditions.

foreground of a shoeprint image is considered as shoeprint region, those areas where the impressions of the shoe outsole pattern are visible. The goal is to enhance the shoeprint in the image which consists of parts of or the entire shoe outsole pattern of the original shoe.

Since most of the available methods are evaluated on different datasets a comparison between them is a challenging task. Furthermore, they focus on the definition of a robust descriptor and a powerful matching algorithm to overcome the problem of versatile appearance of the shoeprints. In this thesis three prototypical preprocessing techniques are developed and tested to enhance the shoeprint images and to make the extracted features more accurate.

For evaluation and testing the FID-300 database is used. The dataset contains both reference prints as well as real shoeprints secured at a crime scene. Additionally, the Ground Truth connecting each crime scene sample to a matching reference print is available. The goal of this work is to define an image processing pipeline which correctly enhances the shoeprint impression and eliminates or suppresses the noise in the shoeprint images. A secondary objective is to gain an overview of the performance of the algorithms, and make an estimation which methods are applicable in real-life scenarios based on their performance on the FID-300 database.

1.2 Challenges

An obstacle in the topic of shoeprint enhancement and in automatic shoeprint matching in general is the versatile appearance of the shoeprint. There are approaches available which build models for given structures of the shoeprint [TSKC10], [AK17], but they are

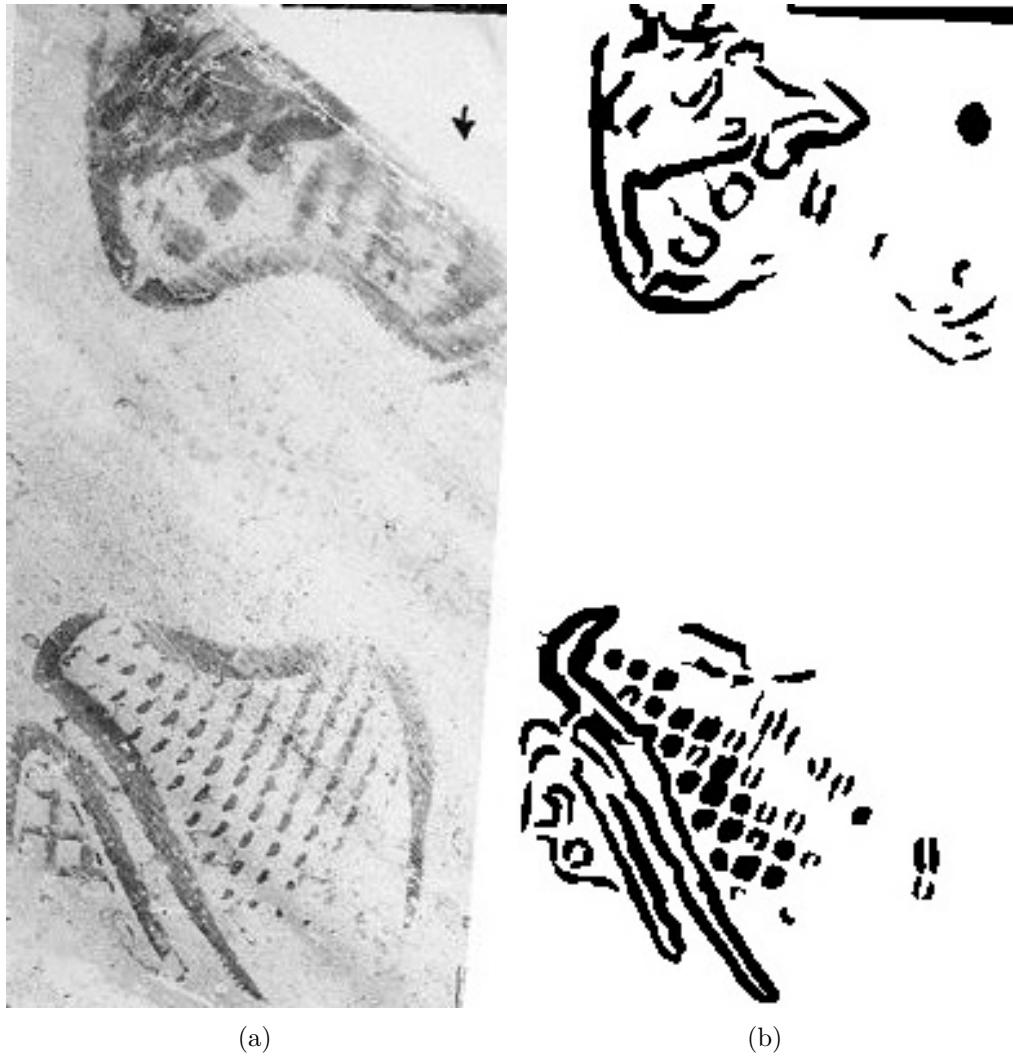


Figure 1.2: Result of the proposed algorithm. The background is correctly masked and the contrast between fore- and background is increased. On the bottom right area a weakness of the algorithms is shown, where small pattern structures were wrongly eliminated. a Example image from the FID-300 database; b Enhanced image

limited to given parts of the shoe sole pattern and are tested on a restricted database. Moreover, there is versatile noise of multiple sources. Firstly, the produced shoepoint and its background depends on the properties of the ground where the impression was made, such as the roughness and unevenness of a given type of surface distorts the original shoe sole pattern. Additionally, the possible unevenness of the ground appear as noise on the final sample. The shoepoint impressions of Figure 1.1a and 1.1d were made on an even surface whereas on Figure 1.1c and 1.1f cluttered background is visible caused by the surface properties. Furthermore, objects on the surface, above or behind the left

shoeprint potentially cover or distort the original pattern, or prevent a clear impression of the complete area of the original shoe sole. On Figures 1.1b and 1.1g only partial shoeprints were secured, on the left side of Figure 1.1g the structures of the shoe sole pattern are hardly visible because of the additional noise. Due to the noise, it is possibly ambiguous, which parts of the image are part of the shoeprint, which region is a depiction of the original outsole pattern, thus the fore- and the background of the shoeprint image is not necessarily clearly distinguishable. Since the original shoeprint image without noise is not available, it is possible that with suppressing noise the shoeprint is also damaged. Other than that, illumination changes occur as well 1.1d. Besides that, the pattern on the original shoe can also be altered over time due to wear. These alterations possibly contain valuable information about the owner, however, they make it more difficult to match the shoeprint image with its reference. Additionally, there are multiple shoeprint securing methods producing different results for the same shoeprint impression[KS17]. The shoeprint securing technique used depends on the properties of the ground. The securing method and the additional properties of the floor, for example if it was clean or dusty when making an impression, determine if the positive 1.3a or the negative 1.3b image, the shoe outsole pattern or the space between its edges, is captured. Furthermore, the majority of the publications focus on the development of robust feature sets and matching algorithms and not on the preprocessing of the shoeprint samples, therefore there is limited research available of the specific domain of shoeprint enhancement. Proposed methods for forensic shoeprint matching are discussed in the following chapter.

As mentioned previously the published methods were tested and evaluated on different datasets, thus a comparison between them is difficult. Moreover, the used datasets are not necessarily public [KS17], [DCC09] making it impossible to reproduce the result in such cases. Additionally, the handcrafted databases can be biased, and allow such restrictions and modifications that do not correlate with real-life scenarios [RBCP19], for instance [DCC09], [TSKC10]. The used samples can either be synthetically generated, they are secured in a laboratory for the purpose of evaluating forensic image processing algorithms, and computationally distorted [DCFR05], [GBCN08] or exclude images by a given criteria, for example quality and noise [DCC09], [TSKC10]. The time consuming approach of re-implementation of the available algorithms and testing them on the FID-300 is also impossible in these cases, since the original dataset of the publication is not accessible and thus any reimplementation can not be verified. Thus it is challenging to plan a new algorithm based on the published results because the lack of a uniform baseline.

1.3 Contribution

In this thesis an overview of shoeprint enhancement methods is given. Multiple approaches are implemented, discussed and evaluated. Two ways to increase the quality, defined as the contrast between the fore- and the background, of a given shoeprint image are to enhance the pattern regardless of the noise and to suppress the noise. Due to the noise, the definition of foreground, as the shoeprint in the shoeprint image, is however

possibly ambiguous. Along fully-automated methods semi-automated algorithms are also considered. Three different approaches are introduced and examined in respect to their performance on real-life image samples.

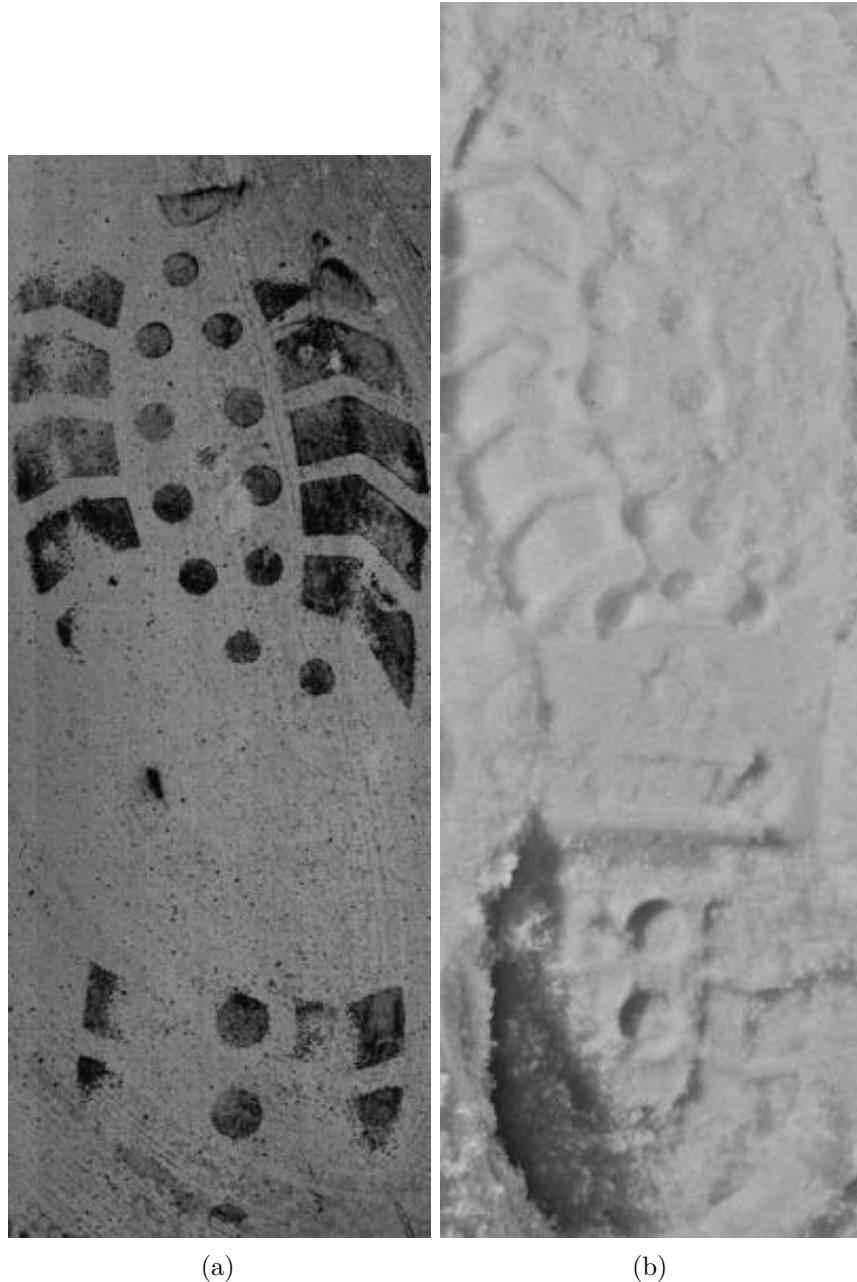


Figure 1.3: Example shoeprint impressions from FID-300 where the positive (a) and the negative (b) image of the same shoeprint is captured

Finally, a semi-automated framework is presented which is evaluated on the FID-300 database. In the first step user input specifying the noise is required. The input is separated into tiles, and the subparts are compared based on the Fourier-Mellin features of the given region and of the neighborhood selected by the user. In that way, the background is separated from the foreground and a noise model based on the background is determined. Since noise appears on the separated foreground as well, those parts are corrected according to the calculated noise model. After that, the pixels are classified based on their gradients and the non-local means algorithm is applied. According to the cardinality of each class shoeprint and noise classes are determined. Finally, candidates of the latter are eliminated. The final image is thresholded to create a binary image, where the shoeprint is better recognizable because the noise is suppressed on the foreground, where the shoeprint is assumed to be located, and eliminated on the detected background area. Throughout the whole processing pipeline morphological operations and small structure elimination are applied multiple times. First when a mask for background is built, and also in the end of the pipeline to eliminate small inconsistencies on the determined foreground and detected lines. The noise elimination is based on two assumptions. First, the area selected by the user is representative for the entire image and there is a significant contrast between background and shoeprint areas. Second, the edges of the shoeprint share similar properties, such as gradient information, frequency as well as overall area and length, and these features are different from those deriving from cluster of the image. Figure 1.2 shows an example from the FID-300 database, see 1.2a, and the enhanced image, see 1.2b, using the above algorithm.

The background of the shoeprint image is masked successfully and the contrast between the shoeprint and the background is increased. However, in the bottom right corner of the image the weakness of the enhancement method is also shown. Small line structures of the shoe sole pattern resemble the short and fine edges of the noise in the image and are wrongly eliminated. The detailed evaluation of the proposed algorithm is presented in Chapter 4. Experimental results show that the enhanced images are clearer, the background is successfully separated and the shoeprint is less noisy than on the original images. Moreover, the improved images have a better matching rate than their original version according to the experiments conducted on the enhanced images, on the original samples and on the reference images using three common image features such as Fourier-Mellin, SIFT and SURF.

1.4 Structure of the Work

To gain an overview of the published research the following section, Chapter 2, gives a review of the literature. Along papers published on the topics of shoeprint identification, matching and enhancement, research on similar domains is presented as well. Consequently, the field of fingerprint processing is also overviewed for possibilities of utilizing their solutions in the given problem space. Furthermore, natural image enhancement and denoising techniques are revised as well.

In Chapter 3 the approaches for enhancement are described. The first section presents and reviews an algorithm for detecting shoeprints on shoeprint images under varying conditions. The second one describes an automated noise suppression pipeline. In the last section an algorithm for enhancing real-life crime scene shoeprint images from FID-300 is proposed. Details on the implementation are given as well.

In Chapter 4 experimental results are shown and the proposed algorithms are evaluated whether they are applicable for real-life forensic images. In Chapters 5 prospective future work is discussed and the final conclusion is given.

CHAPTER

2

Related Work

In order to find and develop an effective algorithm for shoeprint image enhancement, an overview about relevant research is made first. Along the literature of image enhancement and noise removal other related topics are also reviewed. Discriminative image descriptors are considered as well to gain better insight and to define an approach which is optimized for the rest of the shoeprint identification pipeline. In this chapter the research in the domain of shoeprint identification is reviewed. Other than that, publications from similar domains such as fingerprint and palmprint detection are discussed as well. Research for fingerprint identification has been chosen for review because of their similar goal of edge pattern recognition. Moreover, an overview of techniques from the field of natural image enhancement and description along with general image denoising is also given. This chapter is separated into three parts, first, image enhancement techniques are described, after that algorithms developed for noise removal specifically are discussed, and lastly, proposed image descriptors are reviewed.

2.1 Image Enhancement

In this section image enhancement techniques from three specific domains are discussed, these are shoe- and fingerprint identification and natural image enhancement.

Shoeprint Enhancement

There is extensive research done in the field of enhancing shoeprint images [RBCP19]. However, the problem definition and the use-case of the different publications varies strongly. Because of the absence of a common database, the discussed algorithms are separated into two groups, techniques tested on synthetic samples, i. e. on shoeprint impressions generated for the purpose of evaluation of shoeprint identification algorithms, and on real-life shoeprint images. Furthermore, a group of algorithms developed for

2. RELATED WORK

real crime-scene shoeprint images makes restrictions about the input image and exclude images by a predefined criteria, for example noise and quality. Figure 2.1 shows example images from a synthetic 2.1a, from a restricted 2.1b and high 2.1c and low quality samples 2.1d from the FID-300 dataset. Tables 2.1 and 2.2 provide an overview of the used datasets of the discussed algorithms. The former contains further information about the accuracy of the algorithms, the latter gives keywords about the proposed methods of the publications. The algorithms presented in the "real" section of the tables were tested on the FID-300 dataset, used in this thesis as well. Table 2.1 also gives information about the matching accuracy of the discussed algorithms. The value after the at sign stands for the top portion of the bestmatches. 10% means the probability of having the corresponding reference image among the top 10% of all matched images, 10 stands for being the corresponding image among the top 10 best matches. If no value with at sign is given, the accuracy stands for the probability of providing the corresponding reference image as first match.

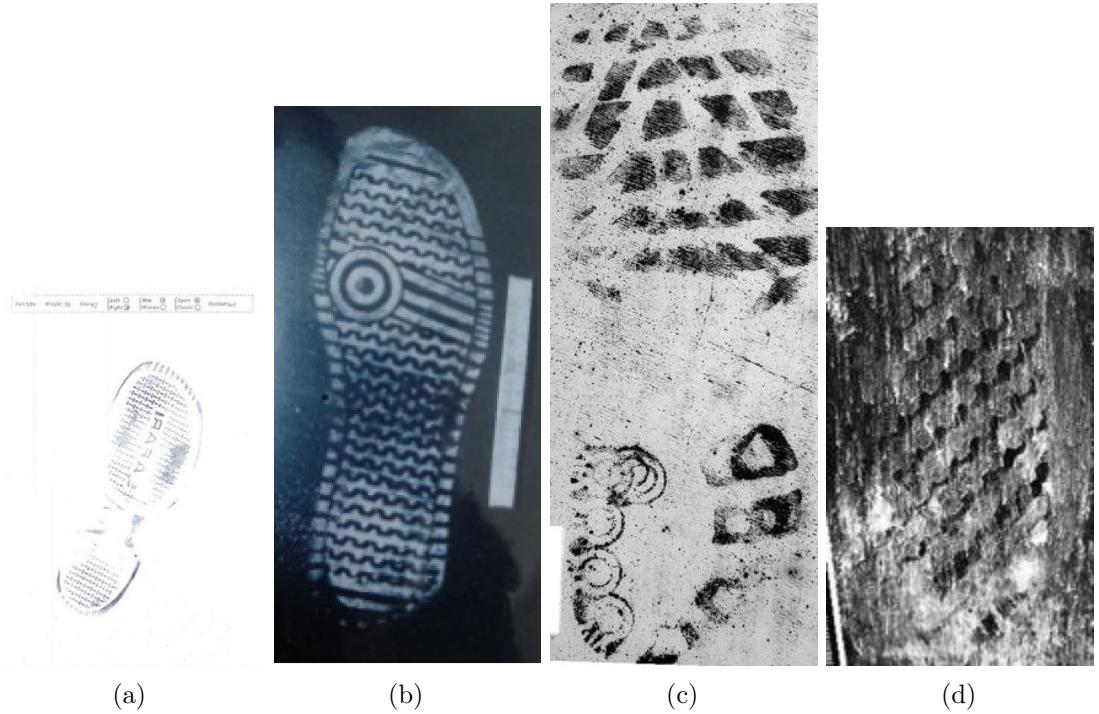


Figure 2.1: Example images of the different datasets of the forensic shoeprint matching algorithms a Example image from a synthetic dataset [AK17]; b Example image from a real crime scene dataset excluding low quality images [LWS14]; c Example high quality image from the FID-300 database [KAV14]; d Example low quality image from the FID-300 database [KAV14]

Morphological Operations, namely Opening and Closing, are popular techniques for improving the quality of both kind, realistic and synthetic, of shoeprint images, they are

used in many cases [WSYZ14], [KYZ14], [LWS14]. Wang et al. [WSYZ14], Kong et al. [KYZ14], Li et al. [LWS14] and Wu et al. [WWZ19] use the Morphological Operations to correct inconsistencies after thresholding. In case of the algorithm proposed by Tang et al. [TSKC10], the input is binarized with Canny edge detection, the inconsistencies are eliminated afterwards again with Opening and Closing. However, as seen on Table 2.2, Morphological Operations are mainly used on high quality data. They are applied on binarized data, but, except of Wu et al. [WWZ19], the other discussed algorithms tested on FID-300 use no enhancement methods and focus on the definition of robust descriptor instead. Furthermore, Morphological Operations are only used for small corrections in the presented publications, see Figure 2.1b, but on FID-300 samples noise causes bigger distortions than Opening and Closing are able to handle, see Figure 2.1d.

To create a binary image and eliminate noise various thresholding techniques are used. Otsu [WWZ19], [AH08], [AK17], [KYZ14] and adaptive thresholding [WSYZ14], [LWS14] are two popular algorithms. Wang et al. [WSYZ14] and Wu et al. [WWZ19] combine Otsu thresholding with a grid based approach to calculate exact thresholds for every subarea of the picture. Similar to the Morphological Operations, thresholding is preferred on high quality images, see Table 2.2. Wu et al. [WWZ19] adjust the algorithm for lower quality images by calculating several thresholds for subregions of the image. Nevertheless, the neglect of thresholding methods for real datasets indicates that they are not applicable for the changing conditions on the samples of the FID-300 dataset.

2. RELATED WORK

Algorithm	Kind of Dataset	Size of Dataset	Accuracy
Algarni et al. (2008) [AH08]	synthetic	500	99% *
Gueham et al. (2007) [GBC07]		100	100% *
Gueham et al. (2008) [GBCN08]		500	99% @ 10 *
Alizadeh et al. (2017) [AK17]		190	99% *
Wang et al. (2014) [WSYZ14]		210000	91% @ 2% *
Nibouche et al. (2009) [NBC ⁺ 09]		300	90% *
Zhang et al. (2005) [ZA05]		512	98% @ 4% *
Patil et al. (2009) [PK09]		1400	91% *
Almaadeed et al. (2015) [ABCN15]	restricted	300	99% *
Kong et al. (2014) [KYZ14]		104	64% @ 20
Li et al. (2014) [LWS14]		195	94%
Tang et al. (2010) [TSKC10]		2660	71% @ 1% *
Dardi et al. (2009) [DCC09]		87	73% @ 10 *
Katireddy et al. (2017) [KS17]		800	no matching done
Wu et al. (2019) [WWZ19]	FID-300	300/1175	93% @ 2%
Wu et al. (2019) [WWNZ19]		300/1175	68% @ 1%
Kong et al. (2017) [KSRF17]		300/1175	90%
Kong et al. (2019) [KSRF19]		300/1175	86% @ 1%
Kortylewski et al. (2014) [KAV14]		300/1175	86% @ 20% *
Kortylewski et al. (2016) [KV16]		300/1175	71% @ 20% *

Table 2.1: Overview about the database and the performance of the algorithms proposed for forensic image matching. The entries labeled with "*" originate from [RBCP19]

Another way to eliminate noise is image filtering. Alizadeh et al. [AK17] uses a simple Median filter. Zhang et al. [ZA05] take advantage of the Partial Differential Equations approach. In this way the edges are preserved while the background is smoothed according to a controlled curvature motion criteria. Katireddy et al. [KS17] uses Successive Mean Quantization Transform (SMQT) [Nil13] as one step to enhance the real-life database. Figure 2.2 shows the output of the SMQT algorithm on an example image. There are, thus two proposed ways for filtering, the first one is general smoothing, the second one is edge preserving filtering. Comparing the databases the proposed algorithms were tested on, the latter seems to be more suitable for forensic images of FID-300 for two reasons. Smoothing is sufficient for images such as Figure 2.1a where the sample is clear and the shoeprint pattern consists of strong edges [AK17], but it potentially damages the pattern information on lower quality samples such as in Figure 2.1d. Furthermore, Katireddy et al. [KS17] tested their algorithm on restricted real samples, which is more similar to FID-300 than synthetic images.

Bandpass operators are also used for noise suppression. The images are converted to the frequency domain where high and low frequencies are eliminated, as proposed by Gueham et al. [GBC07], Richetelli et al. [RLL⁺17]. In the algorithm of Li et al. [LWS14] only the lower frequencies are eliminated. Another frequency based approach was proposed

Algorithm	Kind of Dataset	Enhancement Method	Descriptor
Algarni et al. (2008) [AH08] Gueham et al. (2007) [GBC07]		Otsu th. Bandpass filtering	Hu moments Fourier Transf. with Phase Only
Gueham et al. (2008) [GBCN08]		-	Fourier-Mellin Transf.
Alizadeh et al. (2017) [AK17]		Otsu th., Median filter	Sparse representation
Wang et al. (2014) [WSYZ14]		Morphological op., Adaptive th.	Fourier-Mellin Transf.
Nibouche et al. (2009) [NBC ⁺ 09]		-	Harris and SIFT
Zhang et al. (2005) [ZA05]		-	Fourier Transf.
Patil et al. (2009) [PK09]		Partial Differential Equations	Gabor Transform
Almaadeed et al. (2015) [ABCN15]		-	Hessian Detector
Kong et al. (2014) [KYZ14]		Morphological op., Otsu th.	Gabor Transf., Zernike Feature
Li et al. (2014) [LWS14]		Morphological op., Adaptive th., Bandpass filtering	Gabor Transf.
Tang et al. (2010) [TSKC10]		Morphological op., Canny edge detection	Hough Line Transf.
Dardi et al. (2009) [DCC09]		-	Power Spectral Density
Katireddy et al. (2017) [KS17]		SMQIT, Daubechies wavelets	-
Wu et al. (2019) [WWZ19]		Morphological op. , Otsu th.	Fourier-Mellin Transf.
Wu et al. (2019) [WWNZ19]		-	Manifold ranking
Kong et al. (2017) [KSRF17]		-	CNN
Kong et al. (2019) [KSRF19]		-	CNN
Kortylewski et al. (2014) [KAV14]	FID-300	-	Fourier Transf
Kortylewski et al. (2016) [KV16]		-	Active Basis Model

2.1. Image Enhancement

Table 2.2: Overview about the database and the proposed methods for enhancement and feature descriptors of algorithms proposed for forensic image matching.

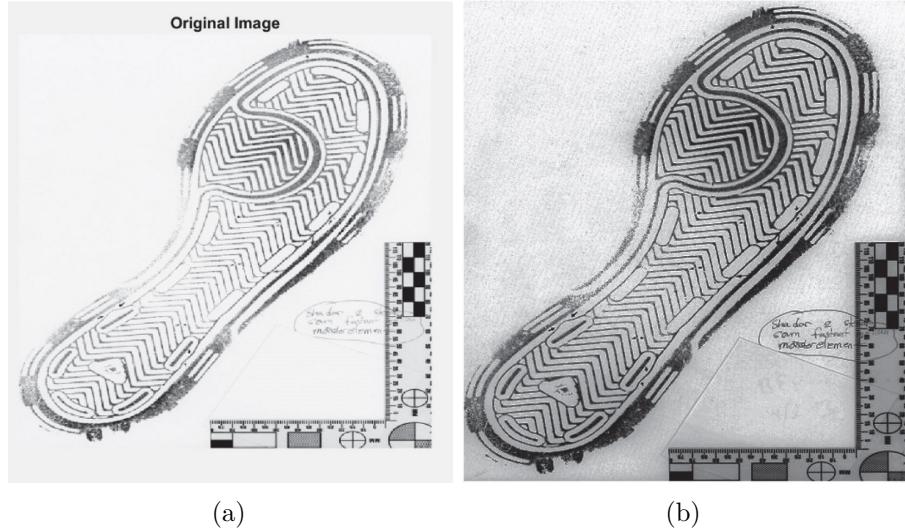


Figure 2.2: Example image presenting the enhancement feature of the SMQT algorithm [KS17]; a Example shoeprint impression; b Enhanced image with the SMQT algorithm

by Katireddy et al. [KS17] for real dataset based on Daubechies wavelets. After SMQT enhancement the Daubechies wavelets are used to separate the fore- and background and to remove the noise in the latter. In case of FID-300, it is challenging to define a heuristic for noise as well as for the shoeprint because of its versatile appearance and no previous knowledge about their properties. Therefore, Bandpass filtering potentially damages the pattern information. However, separating the fore- and the background where the background does not contain shoeprint areas, as Katireddy et al. [KS17] propose, makes it possible to process the noise and the shoeprint independently. Furthermore, it also gives a sample about the appearance of noise on a given image, thus a calculation of a noise model is possible.

It is outstanding that the majority of reviewed publications working on the FID-300 dataset do not suggest to use enhancement techniques, as shown in Table 2.2. As illustrated on Figure 1.1, the appearance of noise changes greatly, thus, it is challenging to define a universal algorithm which deals with the varying noise accurately. Additionally, the presented algorithms in Table 2.1, do not use a-priori knowledge about the samples, as no information about the appearance of noise nor of the shoeprint is available. Because of that, they focus on the definition of robust descriptor and powerful matching technique which finds the corresponding shoeprints, despite the low shoeprint shoeprint-to-noise ratio and changing conditions.

Fingerprint Enhancement

Bandpass [ZSL⁺11], [BHK15] and general image filtering [JCI17] is popular in the field of fingerprint enhancement as well. Zhou et al. [ZSL⁺11] uses a low- and a highpass

filter to eliminate striking frequencies. Baig et al. [BHK15] apply Directional Hilbert transform of Butterworth bandpass to collect the different phase shifts and eliminate the artifacts created by previously thresholding the input. Wang et al. [WLWL14] decompose the image into four subbands and process them separately, calculating the noise for every subband respectively. Jahan et al. [JCI17] apply Fuzzy filtering followed by thinning. Fuzzy filtering is a local method to preserve the edge information and fine line structures while suppressing the noisy background of the input. Similar to forensic shoeprint samples, fingerprint images also depict a 2 dimensional object where the information derives from the line pattern of the sample, see Figure 2.3. The usage of bandpass filtering and signal transform for fingerprint images shows, that the edge content of the input does not influence their performance, assuming that the noise and the content of the given image has different frequency. The proposed methods indicate that they are suitable also for fine-lined images, such as finger- or palmprint samples 2.3a, which is promising when using them on shoeprint images with detailed and fine-lined pattern as well 2.3b.



Figure 2.3: Similarity between a fingerprint sample and a shoeprint image; a Fingerprint image [VVVL16]; b fine-lined shoeprint pattern from FID-300 [KAV14]

Natural Image Enhancement

Maini et al. [MA10] published a review about natural image enhancing algorithms and defined two main groups of them, Frequency and Spatial Domain Methods. First, publications utilizing techniques from the former group are discussed, followed by a review of Spatial Domain Methods.

Xu et al. [XWYH16] combines Bandpass filtering with adaptive thresholding. Similar to Wang et al. [WLWL14] the image is separated into four subbands, and the threshold is calculated for every image separately. Sugamya et al. [SPV16] applies Subband Decomposition with two staged Histogram Equalization. The histogram of the input is equalized globally first, after that it is decomposed into subbands to equalize the values locally for every four generated subimage. In these methods the subbands are enhanced separately and reconstructed into the final output in the end. With processing the shoeprint impressions on different subbands, a method specialized for noise on the given frequency can be defined. Even though the appearance of noise varies across the samples, with decomposition it is separated into known subbands. This way a filtering method for general noise is built by breaking it down into known parts, into known subbands.

Median Filters are used not only in the domain of shoeprint enhancement [AK17], but also for natural image noise suppression. Apart from Median Filter, Li et al. [LLGF14] utilize Average and Wiener Filter as well to suppress the occurring noise and to prepare the input for neighborhood based feature extraction. Feng et al. [FJZY11] proposed a Bag-of-Words algorithm based on the Gabor wavelets of the input. For preprocessing, the Watershed Transform is used. With Wiener Filter and Bag-of-Words method a representation of the estimated noiseless image is built, and with filtering this model is approximated. Along edge preserving filter this approach is favorable, when enhancing shoeprint images with varying noise. In this way, regardless of the appearance of the clutter and non-content elements, the original image is estimated, thus the non-content, in case of this thesis the non-shoeprint, parts are suppressed. The performance of these methods, however, depends on the underlying estimation of the original image [LLGF14].

Histogram Operations are combined apart from Bandpass filtering, like Sugamya et al. [SPV16] propose, also with Thresholding as suggested by Yao et al. [YZL⁺16]. Their approach first separates the histogram of the input into two parts using the Otsu's method, then equalizes the histogram of the generated subimages. Figure 2.4 shows the results of the above algorithm on an example image. As Figure 2.4 illustrates, this method is more suitable for dehazing tasks. Even though, there are low-contrast images in FID-300 available, such as Figure 1.3b, the artifact is not derived from haze.

Color processing techniques are widely used for natural image enhancement: it is applied for image dehazing at low contrast images, [SK18] and also for regular noise removal [RLCL18], [ZSP⁺16]. Bhairannawar et al. [BPJH17] switch from RGB to HSV and use Laplace filter to detect regions with intensity changes. During the processing, the H channel is not modified to prevent color distortion artifacts. Although color processing is

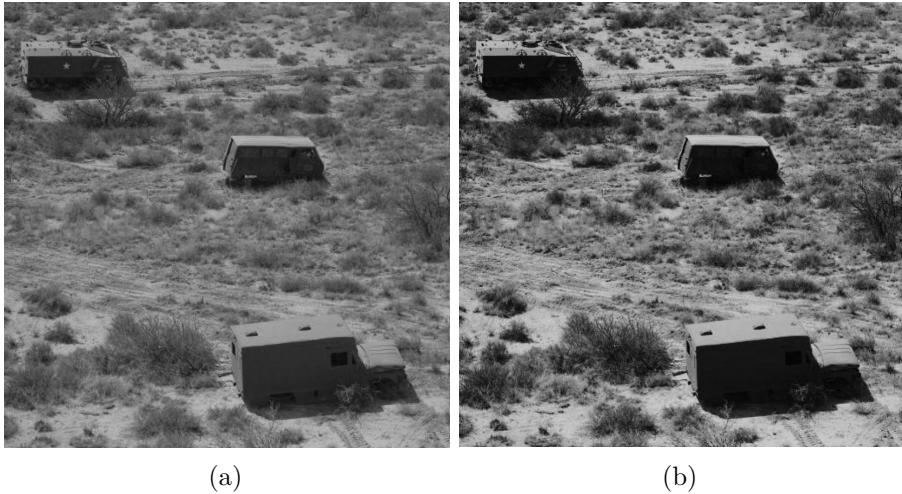


Figure 2.4: Example image of the enhancement feature of the algorithm proposed by Yao et al. [YZL⁺16]; a Example low-contrast image; b Enhanced image with the histogram thresholding algorithm

a well researched field with promising solutions, no wider overview of this topic is given in this thesis, since the here used FID-300 dataset provides only grayscale images. An example of a shoeprint impression dataset with colored samples can be found at [KS17].

2.2 Noise Removal

Noise Removal methods reviewed in this section are based on estimating the original image and eliminating the deviating features of the data. One way is denoising through gradient histogram preservation [ZZSZ13]. The distribution of gradients of the original image is estimated first, and the noisy image is adjusted to the calculated values. An alternative way is to decompose a single image and based on the clear parts, approximate the noisy regions. Huang et al. [HKWL13] propose a self-learning algorithm that only considers the high frequency parts of the decomposed image. In [XZZ⁺15], [TM13], [CM11] and [GZL15] images are separated spatially instead of in the frequency domain, the techniques are based on the idea of non-local means, where the pixels are clustered according to a given criteria and they are set to the mean of the members belonging to the same cluster. The difference between the previous algorithms is how they cluster the pixels or regions into different classes. Taleby et al. [TM13] uses an iterative shrinkage strategy. Chatterjee et al. [CM11] group the geometrically similar regions and estimate the noise for every class separately with the Wiener filter, whereas Guo et al. [GZL15] utilize Block Matching to determine cluster memberships. Additionally, the spatial location is considered as well while calculating the mean value in a given class. The members are weighted according to their distance to the current region. Offering adjustable classification criteria, the non-local means method is promising for the enhancement of

forensic shoeprint impressions. If a feature is defined which is different between noise and shoeprint elements but similar within those, with non-local means the two components are processed separately. In the next section proposed feature descriptors are reviewed, so an overview of possible classification criteria can be made.

2.3 Image Descriptors

This part is subdivided into three domains offering solutions for image description in different areas. Similar topics are reviewed to gain insight about the powerful descriptors and to consider whether they can be used in the domain of shoeprint enhancement. Shoeprint descriptors are described first, followed by the fingerprint features. Finally, at the end of the section natural texture descriptors are also reviewed.

Shoeprint Descriptors

Since the design of image descriptors depends on the data it is invented for the properties of the database the given approach was tested on is given in Tables 2.1 and 2.2. Patil et al. [PK09] propose to use the Radon Transform to determine the dominant direction of structures on the shoeprint image and to process the aligned input with Gabor Transform. A feature map is built by convolving the input image with the Gabor filter. Kong et al. [KYZ14] combine the Gabor features of the shoeprint image with Zernike features to describe the shapes in the shoeprint. Li et al. [LWS14] suggest to use the histogram extracted in the Gabor transformed domain as descriptors. In the approach published by Wu et al. [WWZ19], Gabor Filters are combined with Haar Wavelets and Fourier-Mellin Transform to get an integrated, multi-level descriptor. In other publications Fourier-Mellin Transform is proposed as well, e.g. Gueham et al. [GBCN08] where after transforming the input into the Fourier domain and mapping it into Log-Polar coordinates a correspondence scheme between images is suggested. Wang et al. [WSYZ14] apply multiresolution matching based on Fourier-Mellin Transform. The correlation coefficient is calculated by comparing the Fourier-Mellin features of multiple resolutions. Richetelli et al. [RLL⁺17] classify shoeprint images by applying Fourier-Mellin Transform following the calculation of Phase Only Correlation (POC) to determine the translative difference between two images in the frequency domain. Gueham et al. [GBC07] suggest to use the basic Fourier Transform before calculating the POC in another paper. Kortylewsky et al. [KAV14] also propose a Fourier Transformation based method. It is an unsupervised learning approach, where the periodic structures of a shoeprint are compared in the Fourier domain. Richetelli et al. [RNBS17] compares Randomly Acquired Characteristics of shoeprints, e.g. small damages, modifications and stuck objects on or in the shoesole pattern, examining their Fourier descriptor. Other than Fourier-like transformations, the use of Power Spectral Density was proposed in [DCC09]. It is a descriptor for random, broadband signals based on the frequency and power. In Table 2.1 the matching accuracy of the descriptors above is also listed. Comparing that value, the Fourier-Mellin features seem to outperform the Gabor and the Fourier Transform, having

as high accuracy as 99% and 100%. However, the evaluation was done on different datasets, the above 95% accuracy was reached on synthetic images. Two algorithms [KAV14] and [WWZ19] were tested on FID-300 and the latter achieved 93% accuracy by the combination of Gabor and Fourier-Mellin transform. This shows, that Fourier domain based features are powerful descriptors for synthetic images, and they are also applicable for lower quality data, such as the shoeprint images of FID-300.

Algarni et al. [AH08] suggest to use Hu moments because of its robustness against noise, rotation and resolution. It is also proposed to combine the feature descriptors, as Kong et al. [KYZ14] incorporate Zernike and Gabor features. Tang et al. [TSKC10] define their own fundamental shapes based on common basic structures on the shoe outsole pattern through Hough Line Transform. The extracted features are then stored in an Attributed Relational Graph to represent the entire shoeprint. Comparing the datasets the algorithms were introduced for, see Table 2.1, it is noticeable that shape descriptors were only introduced for high-quality data and not for samples from FID-300. In FID-300 images, there is a high amount of noise which distorts the original shoeprint pattern. In this case, shape descriptors are less suitable, because the original outlines of the elements are modified or covered by noise.

Point descriptors were also proposed for shoeprint impression matching. Nibouce et al. [NBC⁺09] propose to use a four-level Harris Detector and combine it with SIFT. Almaadeed et al. [ABCN15] use the same combination and uses the Hessian Detector additionally for the same purpose in case of restricted real-life datasets. Another publication [RLL⁺17] extract the SIFT features from the frequency domain after applying Fourier-Mellin Transformation and POC on the input image. Similar to the shape descriptors, point descriptors were tested only on high quality data, see Table 2.2. Even though SIFT is a robust descriptor [L⁺99], its neglect by algorithms introduced for FID-300 indicates, that it does not overcome the issue of low signal-to-noise, where the signal is the shoeprint, ratio on the shoeprint images.

Kong et al. [KSRF17], [KSRF19] define a descriptor for real crime scene shoeprint images extracting the mid-level features from a Convolutional Neural Network. Similarly, Wu et al. [WWNZ19] also use machine learning to calculate opinion scores for given matching pairs of real shoeprint images. In the learning process manifold ranking is used where the opinions of human experts as well as feature similarities are both incorporated. Kortylewsky et al. [KV16] developed a hierarchical composition of Active Basis Models, and also extended this for natural image environments [KWW⁺19]. The algorithms presented in this section were all tested on FID-300. Comparing to other descriptors previously discussed in this section, it is notable that the majority of publications utilize machine learning. That implies that for a robust descriptor training is indispensable, and to deal with the varying and high amount of noise the descriptor needs to be adjustable and adapted to the given shoeprint images.

Fingerprint Descriptors

Signal domain representations are attractive not only for shoeprint but also for fingerprint description. Multi-resolution representation through Gabor expansion was proposed by Bolle et al. [BCPR12] to achieve a localized texture descriptor. Van et al. [VVVL16] use Adaptively Adjusted Modified Finite Radon Transform after Gabor filtering to connect edges and eliminate inconsistencies. Rida et al. [RAMM⁺18] propose an ensemble descriptor consisting of Gabor filter, Local Binary Pattern, Histogram of Oriented Gradient and Line Detector to represent a palmprint impression. Other than that, Li et al. [LFLY12] suggest to use Directional Filter Banks, where the image is separated into eight subimages, which then are further decomposed into two frequencies via Wavelet Transform. In case of shoeprint images, frequency domain features, such as Fourier-Mellin features were dominating, they were proposed in several publications [GBC07], [WWZ19] and achieved high accuracy 2.1. But for fingerprint descriptors Gabor Transform and Radon Transform are preferred. This indicates that the latter is more suitable for fine-lined, detailed patterns, illustrated on Figure 2.3a, and the former is more applicable on bigger geometrical structures, Figure 1.1. This observation also foresees the performance of Fourier-Mellin features on detailed shoe outsole patterns such as shown on Figure 2.3b.

For fingerprint images several point feature descriptors were proposed. In [ZSL⁺11] SIFT [L⁺99] features were fused with Hough Transform and Minutiae Information of the fingerprint. Chen et al. [CHZ13] also use the Hough Transform and extend it with a hierarchical voting score to improve the matching information. Along [RAMM⁺18] Ghandehari et al. [GAS12] recommend to use HOG in a 3-level Gaussian pyramid to estimate the local strength of different types of edges on the image. Jahan et al. [JCI17] suggest to combine the Minutiae Information with Speeded Up Robust Features (SURF) and compare them with a Neural Network to find the matching pairs. Local descriptors are suitable for fingerprint images because of the compactness of information. That is, however, not guaranteed on forensic shoeprint samples. Depending on the original shoesole pattern, the elements of the shoeprint are possibly bigger and widespread on the input image, see Figure 1.1, compared to fingerprint sample, see Figure 2.3a. Unlike on fingerprints, where mainly minutiae information is considered for identification, the more complex geometrical structures on shoe outsole pattern are more relevant when matching shoeprint images.

Natural Texture Descriptors

For natural texture description signal and frequency domain features are reviewed. Mistry et al. [MII17] mix multiple features, mainly color and texture descriptors. For texture description, Gabor Wavelet and Binarized Statistical Images [KR12] are used simultaneously. Hatipoglu et al. [HMK00] suggest taking advantage of Dual Tree Complex Wavelet transform instead of Gabor Wavelet. Quevedo et al. [QCAC02] and Xu et al. [XJF09] suggest Fractal features, because they are invariant to intensity and scale changes. Xu et al. [XJF09] propose using a multifractal spectrum to achieve robustness against

viewpoint changes and non-rigid deformations. Hayati et al. [HA18] and Ahonen et al. [AMHP09] incorporate LBP-like features with the frequency domain representation. Ahonen et al. [AMHP09] calculate the LBP of the Fourier features of the image, whereas Hayati et al. [HA18] use Wave Inference, where the information of multiple different-sized asymmetric neighborhoods is added respectively. Fourier Transformation and Gabor wavelets were introduced for all reviewed topics so far [KAV14], [WWZ19], [BCPR12]. It implies high versatility and robustness of the descriptors, since it is possible to configure them for shoeprint, fingerprint as well as for natural texture description. These descriptors are thus powerful tools, and based on the reviewed literature it is advisable to take advantage on them.

There are several publications available which use LBPs for texture description, e.g. [GZP12], [HZPC14], [AMHP09], or for texture classification, e.g. [Khe11], [GZZ10b], [ZLM⁺17]. Wang et al. [WBZL13] modify the usual LBP pipeline with Pixel to Patch sampling to increase the quality of the descriptor without slowing down the calculation. At Pixel to Patch sampling, the weighted average of the neighboring pixels in a given radius is calculated instead of interpolation. Additionally, the Local Neighboring Intensity Relationships subtracted from their grey-scale information are also considered. LBP is frequently used in combination with other techniques as well, examples for that are discussed in the followings. In [HZPC14] based on the covariance matrix additional discriminative features are calculated. In [GZZ10a] LBPs are calculated twice, after that the input is separated into two components, into phase and magnitude, to make the basic LBP rotation invariant as well. Zhang et al. [ZLM⁺17] propose a learning strategy for adaptively weighting the sign and magnitude LBPs to estimate their contribution in the given area. Along the sign and magnitude LBPs, a third modified LBP is also defined, where the local difference vector is determined. In this way, robustness against illumination changes is achieved. In [Khe11], LBP is calculated on multiple levels. Another solution for rotation invariance is proposed by Davarzany et al. [DMY15]. In their approach a circular neighboring radius and a dominant orientation is stored so that rotation invariance is also achieved. Li et al. [LLGF14] process the neighborhood with Rapid Transform, which is robust against cyclic permutation, to achieve the same robustness. Wang et al. [WBLL17] suggest solving rotation invariance by storing the radial and tangential information instead of the intensity values. Guo et al. [GZZ10b] defines "LBPV" instead of LBP, where V stands for variance. In their approach the LBP features with high variance are chosen as discriminative, because they indicate high frequency in the related region. In [BK16], it is proposed to calculate Local Texton Patterns, where the Value channel of the HSV input is subdivided into overlapping subblocks according to its content, and the modified LBP is then determined based on those subregions. Fadaei et al. [FAA17] published a similar approach called Local Derivative Radial Patterns. Instead of binary coding, a multi-level coding is used in multiple directions, and the differences between neighbors are weighted additionally. Chahi et al. [CRT⁺18] define Local Ternary Patterns which store the directional patterns explicitly. Even though, LBPs are extensively researched for natural textures, their applicability for forensic shoeprint images is questionable. LBPs were not suggested for shoeprint identification by

2. RELATED WORK

the publications reviewed in the previous sections. Furthermore, the presented algorithms of this section focus on invariance against rotation, e. g. [Khe11], scale, e.g. [DMY15] and illumination, e.g. [ZLM⁺17]. This is a possible issue on natural texture images, but the samples of the FID-300 dataset are rotation corrected and, as discussed in the following chapter, distortions and additional noise are the main challenges. For that reason the benefits of LBP are limited in the field of shoepoint matching.

CHAPTER 3

Methodology

In this chapter three different algorithms for shoeprint enhancement are introduced. The first one is based on feature learning, the second one uses image filtering and the third one builds a Fourier-Mellin based noise model and applies the non-local means algorithm. As mentioned earlier the testing data of already published algorithms varies, thus it is difficult to make an estimation about their performance on a FID-300 [KAV14] dataset. For that reason three prototypical applications are developed and evaluated in this thesis, and the one with most promising results is further analyzed. The goal of this chapter is to give an overview about the proposed methodology and to present the working pipeline and the implementation details of the methods.

A measure for image quality is the signal to noise ratio, a high value indicates high quality image, where the content is clearly visible, a lower value stands for low quality where the noise suppresses the content of the image. In case of forensic shoeprint images, the outlines of the shoeprint represent the content, the information of the image, and clutter or additional objects in the background are noise. The content, namely the foreground, and the noise is however not necessarily distinguishable especially in images with high noise component, see Figure 2.1d. The three proposed algorithms attempt to enhance the shoeprints by increasing the signal to noise ratio of the input, thus increasing the contrast between the fore- and the background of the image. The first algorithm focuses on finding the shoeprint pattern in varying conditions, the second one filters noise and the last one combines both of them by suppressing the noise with the calculated noise model and enhancing the shoeprint using the non-local means method.

3.1 Feature Learning for Shoeprint Detection

To find the shoeprint in a shoeprint image regardless the noise a robust and discriminative descriptor is defined. Since the noise is highly variable in real-life settings, the shoeprints of the same shoe outsole pattern differ depending on the circumstances causing high

3. METHODOLOGY

intra-class variance. Because of that, a discriminative feature learning algorithm is proposed based on the work of Guo et al. [GZP12]. Guo et al. [GZP12] proposed an adaptive feature selection method to determine discriminative features for natural image textures. While training multiple samples of the same texture captured in different circumstances are processed and the features occurring on every sample are summarized in the feature pool.

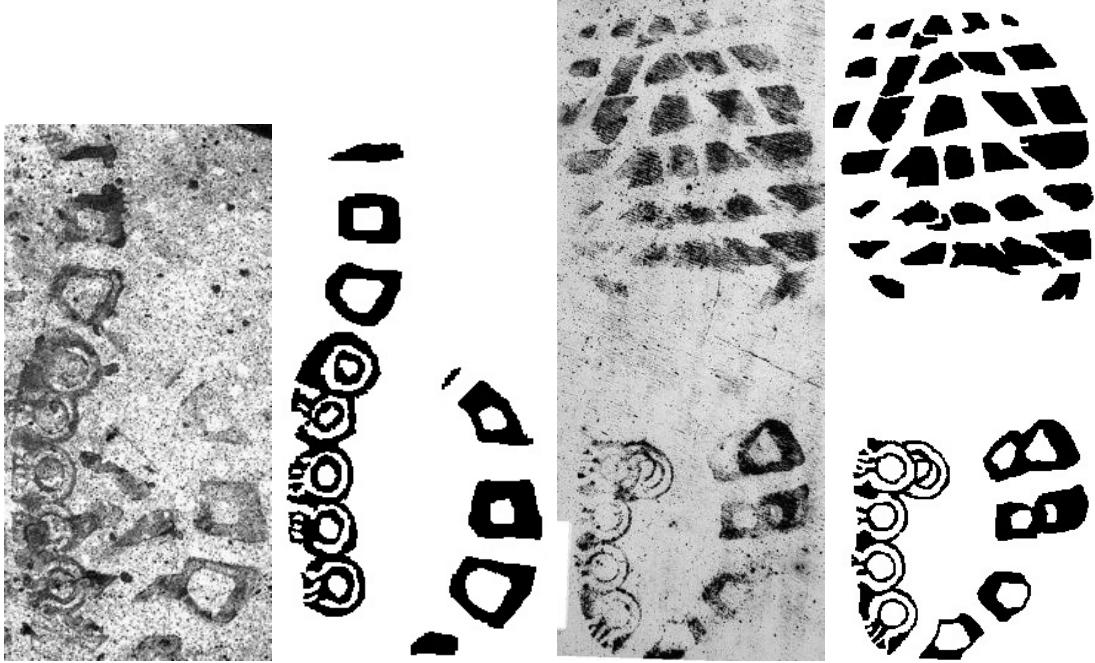


Figure 3.1: Examples from FID-300 [KAV14] dataset chosen for feature learning and their corresponding manually annotated masks.

The feature learning algorithm proposed in [GZP12] was developed originally to find a discriminative feature set for several texture images from a wide database, from [OMP⁺02], [DVGNK99], [BM01], [JNDB05] and [BNS07]. But in this project no such dataset is available since only a subset of samples from FID-300 are usable for learning. The FID-300 [KAV14] dataset consists of 300 real-life forensic shoeprint imaged and 1175 reference shoeprint images, thus not every reference image has a corresponding real image, and the majority of reference images with existing corresponding real images have only one or two examples for real shoeprint image in the database. Since the feature learning algorithm requires multiple representations for the same texture, a shoe sole pattern with the highest amount of real-life shoeprint images is chosen as training set. The training set thus contains six real-life shoeprint image and their reference image. Since there is no pixel wise labeled data in the FID-300 dataset available, the samples chosen for training were labeled manually by comparing the sample images to their reference and annotating all pixels which are deemed part of the shoeprint pattern. Figure 3.1 shows two example images. The black regions of the mask show the shoeprint, whereas the white pixels

indicate the background.

In the work of Guo et al. [GZP12] a three-layered Local Binary Pattern (LBP) feature learning algorithm is introduced for natural texture description. The schematic workflow the algorithm is shown on Figure 3.2. They utilize feature selection where only the discriminative descriptors are chosen and propagated further. In the first layer the samples of the same texture are examined separately. Robustness is granted by selecting those features that describe the majority of a texture of the given sample. This is done by finding the frequently appearing descriptors and ignoring the rare ones, since they are more sensitive to noise. This learning process assumes high quality data, where the content, in this case the depiction of the given texture, dominates the samples and noise occurs on a smaller part of the image. This is however not necessarily the case for the shoepoint images from FID-300, for that reason the learning process is altered which is described shortly. The descriptors of the given sample are selected in the order of occurrences, starting with the most frequent one. If the already selected descriptors cover a bigger region of the original texture than a given threshold, the selection process terminates and the chosen descriptors are propagated for the next level.

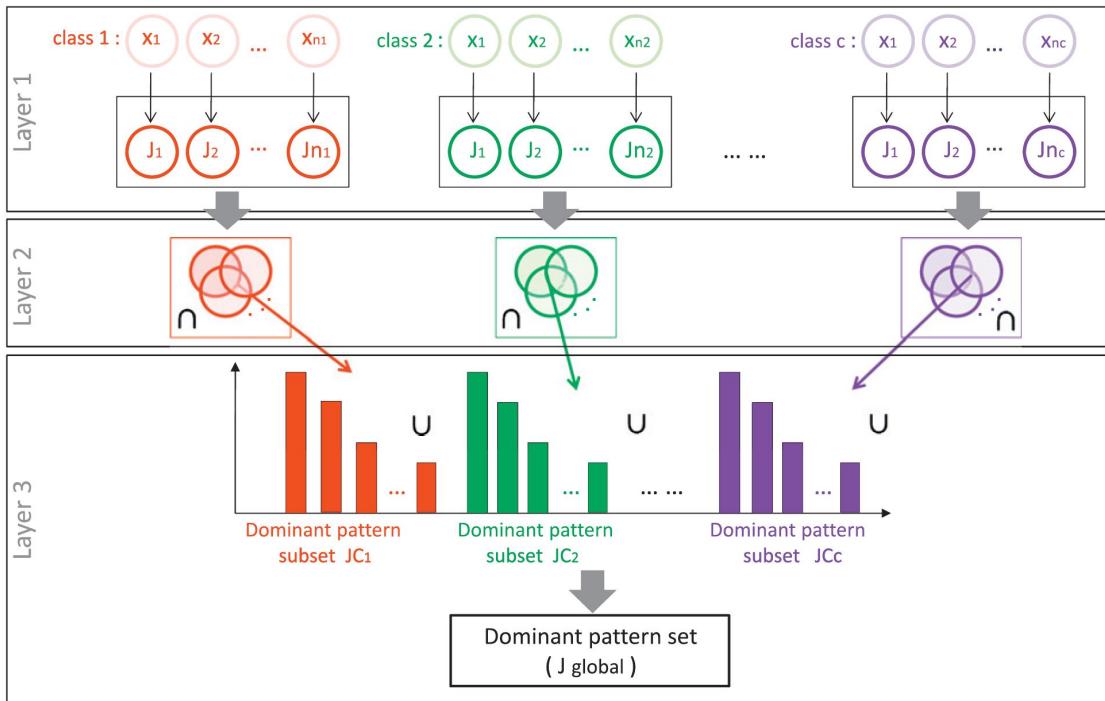


Figure 3.2: The workflow of the three-layered feature learning algorithm [GZP12].

Guo et al. [GZP12] proposed their algorithm for LBP features, and since it is a popular descriptor for natural textures [HZPC14], [AMHP09] and fingerprints [WBZL13], [RAMM⁺18], LBP is chosen to be a candidate descriptor for shoepoint detection. In the research of shoepoint identification both frequency based feature descriptors, such

3. METHODOLOGY

as Fourier-Mellin Transform [WWZ19], [GBCN08], and point descriptors, such as SIFT [NBC⁰⁹], [RLL¹⁷], were already proposed, thus these are also selected for feature learning. Therefore along LBP, the feature learning algorithm is also implemented for Fourier-Mellin and SIFT, but the selection process follows the same steps as proposed by Guo et al. [GZP12]. For the calculation of Fourier-Mellin descriptors the image is extended by mirroring the edges to have a uniform-sized, 5x5, descriptor in all cases. For LBP features two different settings are considered, one with a radius of 3 and 12 sample points and one with a radius of 5 and 24 sample points. These parameter settings were determined experimentally and their performance is discussed in the following chapter.

An easy way to determine the frequency of one descriptor is to count its occurrences within one sample. This method, however, results in high amount of descriptors with little number of occurrence. Consequently, a similarity measure is used for all three candidate features. As similarity measure histogram correlation is used for LBP, a matcher, called Brute-Force, is applied for SIFT and the Fourier-Mellin features are evaluated using the correlation coefficient proposed in [GBCN08]. Brute-Force compares a given feature to every other features of the image using the Euclidean distance and sets the closest one as match storing the distance value between them. The correlation coefficient proposed by Gueham et al. [GBCN08] is given as follows:

$$\text{corr}(FM_1, FM_2) = \frac{\sum_m \sum_n (FM_{1mn} - \overline{FM_1})(FM_{2mn} - \overline{FM_2})}{\sqrt{(\sum_m \sum_n (FM_{1mn} - \overline{FM_1})^2)(\sum_m \sum_n (FM_{2mn} - \overline{FM_2})^2)}}$$

where FM_1 and FM_2 are two Fourier-Mellin Features to compare, $\overline{FM_1}$ and $\overline{FM_2}$ are the mean values of the corresponding features and m and n represent the size of the area the Fourier-Mellin features were extracted from. The higher is the correlation coefficient, the more similar are the two compared descriptors, 1 indicates complete match 0 stands for no correlation. If the similarity value reaches a predefined threshold, the two descriptors are considered as same, and the counter for the frequency of the given descriptor is increased. The experimentally determined threshold for similarity is 90% for LBP histogram correlation, 300 for the maximal distance of two SIFT features and higher correlation than 0.95 between two Fourier-Mellin descriptors.

Guo et al. [GZP12] suggest a threshold based on the area the selected features cover on the original image. The selection of threshold is crucial; if it is too high, less robust descriptors with low frequency are also selected, when the threshold is too low, only the most frequent descriptors are considered and valuable details of the texture are ignored. The parameter settings are determined experimentally and every Fourier-Mellin feature is propagated which occurs at least 10 times on the shoeprint pattern. However, this criteria is altered in the case of LBP and SIFT to eliminate by noise distorted descriptors as follows. Features are firstly extracted both from the fore- and the background of the samples, where the foreground contains the annotated shoeprint and the background is the noisy ground where the shoeprint is lying. After this, to achieve high inter-class distance between shoeprint and noise, frequent descriptors of the noise are determined

and all of them are eliminated from the foreground descriptors. The same similarity measures defined above are utilized to determine the frequency of the noise descriptors and to eliminate them from the foreground descriptors. If the histogram correlation between two LBP noise features is higher than 90% or the matching distance between two SIFT descriptors is lower than 300 the counter for frequency of the given noise feature is increased. These thresholds and the following parameter setting are determined experimentally and the noise similarity criteria is set to be the same as the required similarity between descriptors of the shoepoint. Afterwards the most frequent noise descriptors are determined, occurring at least 100 times among LBP and at least 10 times among SIFT features. All LBP and SIFT shoepoint descriptors are eliminated which have higher than 90% histogram correlation with or have smaller than 250 distance to any dominant noise descriptor. The remaining shoepoint descriptors are propagated to the next level.

The following two layers of the original algorithm proposed by Guo et al. [GZP12] work according to Fisher's Separation Criteria. The second layer aims to minimize intra-class variance, whereas the third layer ensures maximum inter-class distance. The second layer gathers all selected descriptors from the previous layer of the same texture, that are the extracted shoepoint features, from different images and intersects those sets. The assumption is that shoepoints of the same shoe outsole pattern have similar properties under varying circumstances. During this process, features are selected that occur in every shoepoint image of the given shoe outsole pattern. If there is a feature in every shoepoint image sample of the same shoe outsole pattern with a high enough similarity the feature is chosen and added to the final descriptor pool. The same similarity measures defined in the first layer are used for feature comparison. The thresholds for similarity are determined experimentally. If the correlation between two LBP descriptor is higher than 90%, the distance between two SIFT features is smaller than 450 and the correlation between two Fourier-Mellin features is higher than 0.95, the feature is chosen and added to the final set of descriptors.

In the last, third layer representation capability is ensured. This layer merges all remaining features from the second layer into one set to create a dominant feature pool. In the second layer a feature set for a given texture is created based on the extracted descriptors from all samples of the same texture. In the third layer those feature sets of every occurring textures are collected and united. For the purpose of this prototypical implementation only the first two layers are used, because only one texture class, "shoepoint", is defined. Since the noise pattern is unique to each sample and seemingly does not follow any regularities across them, descriptors for shoe patterns are only learned. That means, that only one texture set is available, therefore no merging across multiple textures is needed.

At the end of the learning process a descriptor pool is created. The shoepoint area of the input to be tested is determined based on those learnt features using the same comparison techniques as in the learning phase. The descriptors of the input image are determined similarly as in the training phase. For every pixel of the input a descriptor is calculated. The output image is calculated by comparing the descriptors of the input with the ones

3. METHODOLOGY

in the descriptor pool. In case of LBP and Fourier-Mellin features the correlation value is written into the resulting image which is normalized at the end of calculation. The result image of the SIFT feature comparison is binary, it is set to one if the descriptor of the corresponding pixel has a smaller distance than 200 to any descriptor from the learned set. This distance is determined experimentally, its influence on the resulting image is discussed in the following chapter.

3.2 Fully-Automated Noise Elimination

Similar to the previous method the main goal is to distinguish between noise and pattern, but this time instead of finding pattern regardless of the noise, the noise is suppressed in two steps, in the background first, and on the shoepoint afterwards. The method proposed in this chapter has three main steps, first the noise is suppressed on the entire image, after that the shoepoint is enhanced and the result image is generated lastly by thresholding. A similar pipeline, namely to filter the shoepoint image first and to apply thresholding afterwards, is often used as a preprocessing step for shoepoint matching [AK17], [WSYZ14], [LWS14], [KYZ14]. Katreddy et al. [KS17] propose to enhance shoepoint impressions using Successive Mean Quantization Transform (SMQT). In the presented algorithm those two approaches are combined by adding an enhancement step between filtering and thresholding. In this way the noise is suppressed first, so that SMQT affects the relevant, shoepoint, area, and the pipeline is closed with thresholding to generate the binary image and to eliminate remaining noise parts. The workflow of the entire algorithm is shown on Figure 3.3.

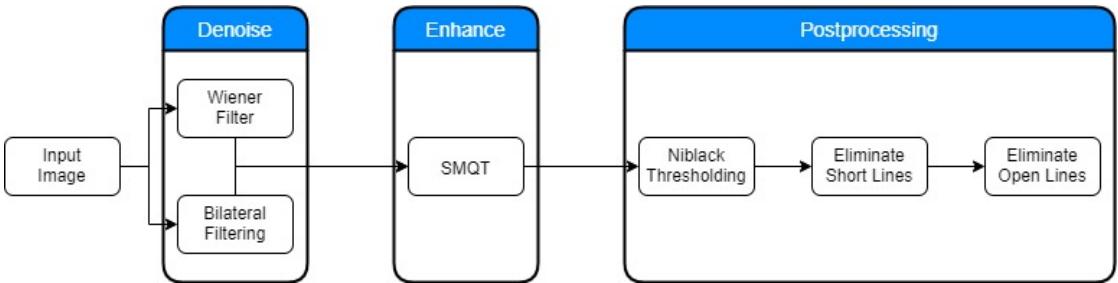


Figure 3.3: The workflow of the fully automated noise suppression algorithm.

The first part of the algorithm is responsible for noise suppression. Two filters proven to be effective in natural image noise elimination are combined, these are Wiener filter [LLGF14], [CM11] and Bilateral filter [ZSP⁺16], [HKWL13]. As Li et al. [LLGF14] state, Wiener filter is among the most popular techniques for noise reduction. Wiener filter works in the signal domain, where it estimates the original image based on the cluttered input, in this case based on the noisy real-life shoepoint image. The filter is given by following Equation [Win]:

$$G(u, v) = \frac{H^*(u, v)P_s(u, v)}{|H(u, v)|^2 P_s(u, v) + P_n(u, v)}$$

where $H(u, v)$ stands for the Fourier Transform of the point-spread function, $P_s(u, v)$ denotes the power transform of the signal, which is the Fourier Transform of the signal autocorrelation and $P_n(u, v)$ expresses the power spectrum of noise, which is the Fourier Transform of the noise autocorrelation. $P_s(u, v)$ and $P_n(u, v)$ essentially represent the two components of the image, the content and the noise, in the signal domain. The point-spread function defines the appearance of a point-like object on the images. The performance of the filter depends on the quality of P_s , on the estimated appearance of the original image. In this method the estimation is made based on the entire input image, because no further information, for example a mask about the shoeprint area, is available. However, this also means that the estimation is less accurate than using a mask on of the shoeprint area, thus the performance is lower [CM11]. To apply the Wiener Filter the input image is normalized first to the range of 0 to 1. For the Point Spread Function an image of size 5x5 is set. The balance parameter is 1100 which sets the ratio between information adequacy and prior adequacy, those parameters control frequency increment and decrement respectively. The parameter settings are determined experimentally and set a smaller kernel size, thus less aggressive filtering, intentionally. As mentioned above the estimation of the appearance of original image is limited, since the original shoeprint image without noise is not available. Therefore it is possible that apart from suppressing noise, the shoeprint is also damaged. With smaller kernel size the noise is less blurred but in exchange higher frequencies such as the shoeprint is preserved. Similar to the Wiener Filter, Bilateral Filter is also used for smoothing the image. Since it considers the weighted average of neighboring pixels, it preserves the edge information which is crucial to prevent blurring on the shoeprint [Ela02]. The kernel size of the Bilateral Filter is 5x5 which was determined experimentally and is based on the same principle as the parameter settings of the Wiener filter. A bigger kernel size results in more blurred image background, however, it is not able to preserve the fine-lined foreground area, since the bigger neighborhood outweighs the small lines occurring in the filter window. For that reason a smaller kernel is chosen, to avoid the possible blurring on the shoeprint. When both images are calculated the difference between them is propagated. In this way the blurring effect is strengthened in the background while the outlines of the shoeprint patterns are preserved. With the combination of both filtered images, the effect of the Wiener and Bilateral filters is amplified, the little blurring caused by the small filter kernels is added in the background while the foreground stays undamaged.

For enhancement Successive Mean Quantization Transform (SMQT) is used [Nil13]. The method was already proposed for shoeprint enhancement by Katireddy et al. [KS17]. SMQT is a recursive algorithm working on subsets of the entire data which splits the given set into two parts in every recursion level depending on the current value being smaller or bigger than the mean of the entire subset. In the first recursion level the data, in this case the shoeprint image, is separated into two parts, and it is noted which pixel is under (0) and which one is above (1) the current mean. The algorithm then recursively continues on both subgroups of the image, splitting and noting the relative value compared to the mean of the subgroup for every pixel. Note that the pixels in the same subgroup does not have to be neighboring, the clustering is solely based on

3. METHODOLOGY

the pixel value. The recursion stops if the predefined depth is achieved. To finish the transformation the noted relative values of the pixels are examined. Along the levels of the recursion a sequence of ones and zeros are registered for every pixel. As a last step, this sequence is considered as a binary number and its decimal value is written into the corresponding pixel. SMQT works based on the similar principle as Histogram Equalization. Because of the successive partitioning, the pixel values are spread on the range of the depth of the recursion. In this way the structure of the data is exposed.

Because of time efficiency, a speeded up version is implemented. In the first step a table of occurring values on the image is created where the zeroth column corresponds the pixel value of 0, the first column stands for pixel value of 1 and so on. In the first row of the table, the frequency of the given value is noted, that is the number of pixels having the value which the column corresponds with. In the second row the sum of frequencies up to the current column is stored. The third row represents the sum of all elements until the given column. The SMQT algorithm is then applied on this table instead on the input image. Since the entries are ordered, if the subgrouping starts, the values belonging to the same cluster will be neighboring columns. Figure 3.4 shows the first eight columns of the occurrence map of an example image. This table eases the calculation of mean value for every level of the SMQT algorithm. For the mean calculation the sum of all pixels is divided by the number of pixels, which is stored in the third and second row of the table respectively. To calculate the mean value of one subgroup two columns of the table are considered, the one before the first column of the subgroup and the last one in that given group. The values are first corrected by subtracting the elements of the former from the latter, and then dividing the third row with the second one the average value of the group is determined. This correction is needed to eliminate the offset in the table, skipping this step the mean of every pixels having lower or equal value than the biggest element of the given subgroup is calculated. When the occurrence map is ready the recursion starts, the mean of the given subgroup is determined and the occurrence map is split into two parts according to the pixel value of the given column being bigger or smaller than the mean value. Every column of the table have a binary code which is created during the recursion. In the first level the first digit of the code is written, in the second level the second digit etc., the length of this code is the same as the predefined depth of the recursion. If a pixel value is smaller or equal to the calculated mean of the given subgroup a zero is written into the belonging binary code, and a 1 is recorded if it is bigger. The manually set depth of the recursion is 8, in this way when the final binary codes are converted to a decimal number a range of 0 to 255 is covered.



	0	1	2	3	4	5	6	7
Frequency	1.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0
Sum of Frequencies	1.0	1.0	1.0	2.0	2.0	3.0	3.0	3.0
Sum of Elements	0.0	0.0	0.0	3.0	3.0	8.0	8.0	8.0

Figure 3.4: The first 8 columns of the occurrence map. The header of the table represents the possible pixel values

The last part of the application is the postprocessing step where the input is converted into a binary image and inconsistencies and remaining noise are eliminated. When binarizing footprint images Otsu's technique [AH08], [AK17], [WWZ19] or adaptive thresholding [WSYZ14] is used frequently. However in the proposed approach a local thresholding technique, called Niblack Binarization Method (NBM) [Nib85] is preferred. There are several publications available [SZG11], [Ath11], which prove that global methods, such as Otsu Thresholding [Ots79], is less feasible as their local counterparts. Although the studies mentioned were carried out on text documents with varying image quality, the two domains are considered familiar since both aim to find fine line structures on a cluttered background. Furthermore, Saxena et al. [Sax19] also state that NBM is one of the most powerful thresholding methods, outperforming the global and some local techniques as well. NBM is calculated as follows [Sax19]:

$$T_d = m(x, y) + k * s(k, y)$$

where m and s stands for mean and standard deviation in the given area respectively and k is a configuration variable which is given manually. In this thesis the window size is the same as the size of the Wiener Filter, 5x5, and the regularization parameter k is set to 0.5 which values were determined experimentally.

There is, however, one disadvantage of local binarization and that is the local window size. Since for every subwindow a new threshold is calculated NBM tends to generate salt and pepper noise on more homogeneous, e.g. background of the shoeprint, area. For that reason two other postprocessing methods are also implemented based on the connected components of the image. They are extracted considering 8-neighborhood, so diagonal connectivity is also considered. The first postprocessing step eliminates short lines on the image, based on the assumption that the outlines of the shoe pattern build longer, coherent edges. The second one deletes all remaining open structures unless the length is higher than a given threshold bigger than the one given in the previous method. Those two heuristics are based on two observations, first, when there is no shoeprint impression region in a NBM subwindow the edges of the remaining clutter and binarization artifacts are generated. Second, since NBM concentrates on the contours, if the complete structure of a pattern element is found a closed line structure is extracted. It is possible, however, that an open line structure is generated, if only parts of the shoeprint are visible or identified correctly. For that reason, only a subgroup of open lines, the ones that are shorter than a threshold, is eliminated. It is possible to set the thresholds of the two postprocessing steps to the average size of every region, however, experimental results show that such strong criterion eliminates shoe pattern lines as well. For that reason two experimentally defined, manually set thresholds are used, they are set to 50 and 60 respectively for every test image. If an open line structure is found which overall size is smaller than 60, it is deleted.

3.3 Shoepoint Enhancement with Non-Local Means

Since noise tends to appear on the entire shoepoint image not only in the background region, a semi-automated algorithm is introduced in this section, where user input about the noise is expected and based on that information a noise model is calculated. As mentioned earlier filtering the noise first and using a thresholding method afterwards is widely used as preprocessing step for shoepoint matching [AK17], [WSYZ14], [LWS14], [KYZ14]. Similar to that pipeline, the proposed algorithm has two main steps. First, the noise is eliminated in two stages. In the first step the background and the foreground, the assumed area of the shoepoint, is separated to mask the former region. After that, the foreground is filtered in the frequency domain to suppress the noise covering that area. Second, the shoepoint pattern is enhanced, by extracting the edges and applying the non-local means method, and binarized finally. A full separation between fore- and background is possible, because according to the user input a detailed background and noise model is built. In this way, it is possible to mask the noise in the calculated background area instead of just suppressing it with filtering. Furthermore, the masking does not modify the foreground area which is not guaranteed when using a filter on the whole input sample. As a first step of algorithm, the user has to select a pixel in the background where there is no pattern information, no foreground, in the neighboring area. The full pipeline is shown on Figure 3.5.

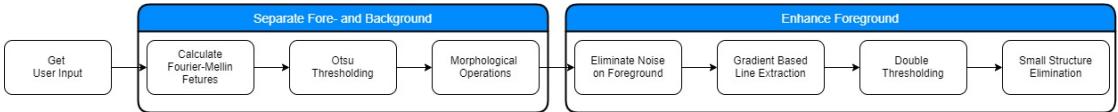


Figure 3.5: The workflow of the semi-automated noise suppression algorithm.

Based on the user input the background area is determined according to the Fourier-Mellin features [SD86]. Fourier-Mellin features were already used for shoepoint description for both synthetic [GBCN08] and real [WWZ19] shoepoint images, however, in this algorithm it is used for noise description. This decision is based on the observation that although the noise varies among samples, there are dominant features within one image which describe the majority of the clutter on a large part of the shoepoint images. The varying appearance of noise is illustrated on Figure 3.6, comparing the background of the images there is little or no similarity across the samples, however, within one shoepoint image dominant patterns are recognizable. To prepare the descriptor for the variance of the noise, three, different-sized samples are extracted around the point determined by a user input.

Fourier-Mellin features are scale, rotation and translation invariant descriptors and are calculated according to the following equation [KG11]:

$$\mathcal{M}_f(u, v) = \frac{1}{2\pi} \int_0^\infty \int_0^{2\pi} f(r, \theta) r^{-ju} e^{-jv\theta} d\theta \frac{dr}{r}$$

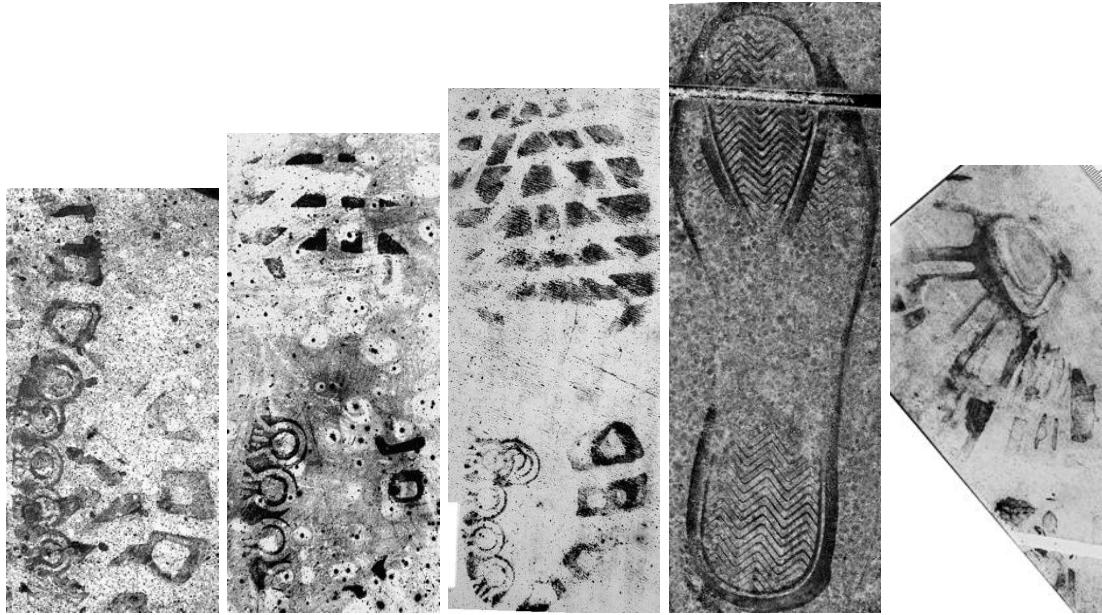


Figure 3.6: Example images from the FID-300 dataset. The noise varies across the images, but there are dominant structures within one sample.

where u and v are the Mellin and the Fourier transform parameters respectively. If the image is translated, the Fourier-Mellin transformed representation does not change. That is however not the case when the image is scaled or rotated. In both scenarios phase shift appears on the feature descriptor and the magnitude changes according to the scaling factor. Because of those reasons above, the feature descriptor is converted to Log-Polar coordinates, so that said image transformations occur as a translation in the descriptor. Log-Polar Transform is used to achieve robustness against scale and rotation [GBCN08]. The Log-Polar coordinates of a point with Cartesian coordinates are given by [SPK12]:

$$(\rho, \theta) = (\sqrt{(x - x_c)^2 + (y - y_c)^2}, \tan^{-1} \frac{y - y_c}{x - x_c})$$

where x and y are the Cartesian coordinates of the point and x_c and y_c are the center coordinates of the input image. The Log-Polar coordinates represent the radial distance, ρ , and the angle from the center, θ . This conversion is illustrated on Figure 3.7 as well [SPK12].

The Fourier-Mellin descriptors extracted from the three different sized neighborhoods around the user input represent the noise, after that every pixel of the input sample is examined for correspondence with the metric defined in the first section of this chapter 3.1. The diameter of the neighborhoods is determined experimentally and is set to 6, 12 and 18 respectively. The borders of the shoepoint image are extended by mirroring additionally to ensure that the calculated Fourier-Mellin features have uniform size for every pixel on the image. The calculated similarities are stored in a correlation map by taking the

3. METHODOLOGY

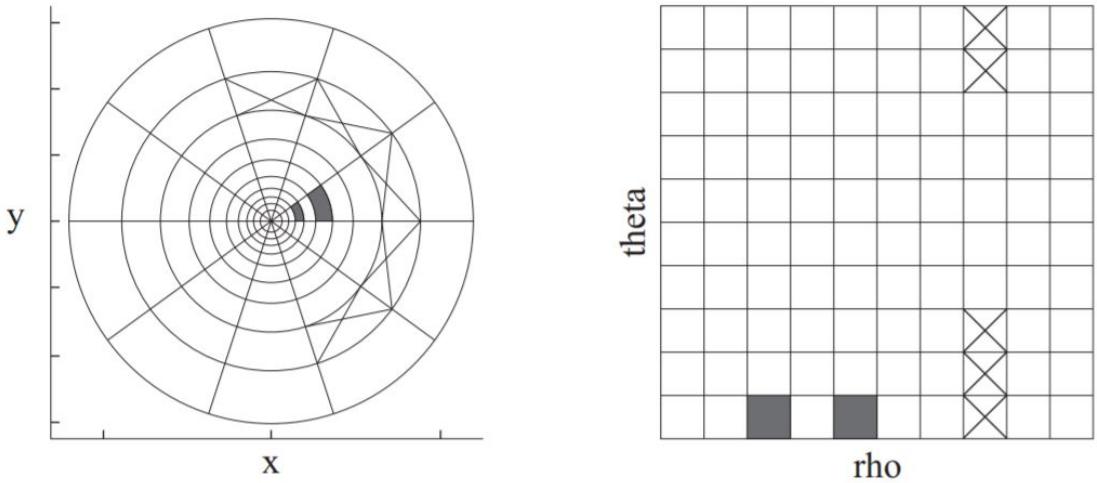


Figure 3.7: Illustration of the relation between Log-Polar and Cartesian coordinates [SPK12]

average correspondence value of the three different sized subimages, and the correlation map is binarized by Otsu's method afterwards. This thresholding method was used for two reasons. First, it is popular in the forensic image processing for all three kinds of datasets, i.e. at synthetic [AH08], [AK17], restricted [KYZ14] and also at real forensic samples [WWZ19]. Second, with a global thresholding method the pivot value between fore- and background pixels are found. To complete the background mask, morphological operations are applied, using an experimentally defined, 8x8 kernel, Opening first and Closing second, to eliminate small holes and rough contours as proposed by many approaches for shoeprint identification [WSYZ14], [KYZ14], [LWS14], [TSKC10].

Once the background mask is generated the foreground is processed to suppress the noise and enhance the shoeprint. The mask eliminates the noise in the background area, however, it does not modify the remaining region, thus this area is examined next. Noise on the shoeprint is suppressed first, afterwards the edge information is extracted and the binary image is lastly generated. The noise is possibly located on the shoeprint itself, to preserve the underlying shoeprint information frequency based noise suppression is applied. Wavelet transform is often used for natural image denoising [XWYH16], [SPV16], for fingerprint [LFLY12] and for shoeprint enhancement [KS17] as well. During wavelet transform the image is decomposed into subbands, and the subimages are processed separately. In case of four subbands, the LL image contains the edge information and the other ones the intensity values. Depending on the properties of the noise the image clutter is separated from the information and suppressed on the corresponding sub-band. After that the image is reconstructed by inverting the decomposition process.

The same principle is used in this approach, but instead of wavelet transform, Fourier transform is applied, to further exploit the previous step of the algorithm. The Fourier

representation of every pixel on the shoepoint sample is calculated using the Discrete Fourier Transform on a 6x6 window, the smallest neighborhood used in the previous step. This time the window size is fixed, since the final model is based on the neighborhood of every background pixel and not only on the area around the user input. When the background is separated two models of the appearance of the noise are built. The average noise model is calculated by taking the mean descriptor of the Fourier representations of the background pixels. The other, centorid model is determined by calculating the centroids of the k-means algorithm applied on the background area. The k-means algorithm is set to terminate after 10 iterations, whereas k equals to 8, so 8 centroids are determined which are the representations of the noise on the given shoepoint image. These parameters were set experimentally by focusing on the robustness and versatility of the submodels. Increasing k, more noise-models are calculated, thus the noise description is more versatile. But in the same time, the single models, the centroids, are based on a smaller set of pixels, therefore they are more sensitive for outliers as well. The noise is suppressed by calculating the difference between the noise model and the Fourier representations of the foreground pixels. In case of the averaged noise model, the Fourier image of noise is subtracted from the Fourier image of the foreground pixel and its neighborhood and the image is reconstructed using the Inverse Fourier Transform. When using the centroid noise models, the Fourier image of the foreground pixel is compared to all possible submodels, and the one with highest correlation is used. The correlation coefficient is determined using the Equation for Fourier-Mellin Feature comparison 3.1. When the centroid is decided the difference between both Fourier representations is calculated and the foreground is reconstructed again using the Inverse Fourier Transform. The performance and the advantages of both noise representations is discussed in the following chapter. If the background area is smaller than a given threshold, this step is skipped, because in this case no robust model about the noise can be made. In this case it is set to 20% of the image which was determined experimentally.

For line extraction and shoepoint enhancement a gradient based method published by Zeng et al. [ZL11] is used. They propose a region based non-local means algorithm to denoise natural images. Non-local means was developed for image denoising and grants better smoothing while preserving the details on the image than local mean calculation [BCM05]. This is achieved by the averaging pixels with similar properties and not the ones within the same neighborhood. In non-local means the image is separated into patches which are clustered according to a given criteria. While smoothing the average value of the patches within the same class is calculated and the members are updated with that value. Zeng et al. [ZL11] propose to use the gradient information of the pixels for classification, which is ideal for shoepoint images, because the goal is to preserve the shoepoint edges during the enhancement. Since the classification is made based on the gradients, edge properties are the essential clustering criteria. Assuming that both the shoepoint and the noise on the shoepoint image have common properties which are different to each other, pattern and noise classes are created at the end of the clustering. In this way, the outlines of the shoepoint are preserved and with the non-local menas algorithm the residual noise is smoothed.

3. METHODOLOGY

The pixels are clustered according to the eigenvalues of the tensor matrix given by [ZL11]:

$$T_\sigma = \begin{pmatrix} t_{11} & t_{12} \\ t_{12} & t_{22} \end{pmatrix} = \begin{pmatrix} G_\sigma * (g_x(i, j))^2 & G_\sigma * g_x(i, j)g_y(i, j) \\ G_\sigma * g_y(i, j)g_x(i, j) & G_\sigma * (g_y(i, j))^2 \end{pmatrix}$$

where g_x and g_y represent the gradient information in the corresponding directions and G_σ is the Gaussian kernel having σ as standard deviation. To do so, the image is blurred first with the Gaussian kernel of experimentally defined size of 11x11 to make the calculation more robust. To extract the gradients the Sobel filter is used in x and y direction separately. The eigenvalues of T_σ are then given by [ZL11]:

$$\lambda_1 = \frac{1}{2}(t_{11} + t_{22} + \sqrt{(t_{11} - t_{22})^2 + 4t_{12}^2})$$

$$\lambda_2 = \frac{1}{2}(t_{11} + t_{22} - \sqrt{(t_{11} - t_{22})^2 + 4t_{12}^2})$$

The classification is based on the difference between the two eigenvalues of the given pixel. If the difference is small, it indicates smooth region, whereas a higher value implies edge area. The proposed classification strategy is the following [ZL11]:

$$(i, j) \in \begin{cases} c_1, & \text{if } \lambda(i, j) \leq \lambda_{min} + \frac{1(\lambda_{max} - \lambda_{min})}{n} \\ c_2, & \text{if } \lambda(i, j) \leq \lambda_{min} + \frac{2(\lambda_{max} - \lambda_{min})}{n} \\ \dots \\ c_n, & \text{if } \lambda(i, j) \leq \lambda_{min} + \frac{n(\lambda_{max} - \lambda_{min})}{n} \end{cases}$$

where λ_{min} and λ_{max} stands for the minimum and maximum difference on the entire image respectively. In this thesis two parameter settings are implemented using 25 and 20 classes, which were determined experimentally. Their comparison and the discussion about their performance is given in the following chapter. For denoising, the mean illumination value of pixels in the same cluster is calculated and the original pixel value is replaced with the average of the cluster. Zeng et al. [ZL11] propose an additional weighting scheme for mean calculation as well, so that pixels in the same class with smaller geometrical distance have a higher weight than those further away on the image. However, the weighting method is skipped in this implementation, since geometrical location is not relevant in this use-case for two reasons. First, the size of the region the image was taken of is generally significantly smaller than the scene on natural images. Second, shoeprint images depict a 2 dimensional object, namely the shoesole impression of a given shoe on a given surface. Since there is no depth difference within the image, the entire area is considered as one connected region, unlike in natural images where the objects in the foreground and in the background do not belong together. But another criteria is added to the clustering algorithm. After determining the classes, if one or more cluster have more members than a given threshold, histogram equalization is applied on the image where the differences of eigenvalues are stored, and the classes are recalculated.

This threshold is set to be 60% experimentally. This optimization is needed to lower the chance that pattern and noise edges are assigned to the same cluster.

To find the cluster with the noise edges, a simple heuristic is defined. Since the clustering algorithm is applied on the entire image, not only on the foreground, the cluster with the highest amount of members is eliminated. This decision is based on the assumption that noise is found in the entire shoepoint image, whereas shoepoint edges are located only at given areas. Furthermore, it is also assumed that unlike shoepoint contours, where the edge properties change according to the original shoepoint structures and pressure, noise have similar appearance throughout the the entire sample. In case of using Histogram Equalization, however, the class memberships are more balanced, so no such heuristic is assumed.

In the next step, the image is binarized with the adaptive thresholding technique [LZ96] popular in both fields including shoepoint identification [WSYZ14], [LWS14] and natural image processing [XWYH16]. Adaptive thresholding is a local binarization technique similar to NBM. It generates multiple thresholds for every subregion of an image based on the Gaussian weighted sum in the window, which is experimentally set to 12 in this implementation.

The closing step is postprocessing, where small discontinuities and remaining noise are eliminated. Even though in the previous step noise edges belonging to the biggest class were eliminated, other noise edges assigned to other clusters are still part of the image. This problem is solved with short line elimination. To eliminate noise while preserving the shoepoint edges, the same assumption is made as the one in the previous section. The noise consists of cluttered, small components, whereas the shoepoint is built of bigger, connected structures, thus the remaining small regions, smaller than a given threshold, are eliminated. The connected components of the image using 8-connectivity are extracted and every element with an area not bigger than 15 is deleted immediately. This threshold is set manually and is significantly smaller than those in the previous section deleting every structure smaller than 50 or open structures smaller than 60. This time the majority of noise is already eliminated, since the background is masked, and adaptive thresholding does not generate such high amount of thresholding artifacts and noise clutter than NBM does. Because only the residual noise around the outlines of the shoepoint has to be deleted, no such high threshold for small structure elimination is set.

CHAPTER

4

Results and Discussion

In this chapter the experimental results of the presented algorithms are shown, and it is discussed if they are usable for enhancement of real-life shoeprint images from the FID-300 dataset. Along the qualitative evaluation of all algorithms, the quantitative evaluation of the most promising one is also given. All algorithms are implemented in Python 2.7 [VRDJ95] using OpenCV 4.1.1 [Bra00].

4.1 Qualitative Evaluation

In this section the qualitative evaluation of the three proposed algorithms is presented. Based on the results, their advantages and disadvantages are examined and possible ways for improvement are discussed.

4.1.1 Feature Learning Algorithm

Two kinds of experiments were conducted to test the performance of the descriptor. Testing on the training set and testing on an evaluation set, on shoeprint images which were not considered while training. The first scenario is discussed now, and the results of the second one afterwards. Tests are performed on the training set to see how many descriptors from the original image were eliminated, and to identify the common features in the training dataset. Figure 4.1 and Figure 4.2 show example results of this testing setup.

The visible difference between Figure 4.1b and Figure 4.1c as well as between Figure 4.2b and Figure 4.2c shows that bigger neighborhood of LBP is a better choice for shoeprint description. Considering the middle region of Figure 4.1, the background in Figure 4.1c is less noisy than in Figure 4.1b. This difference is more visible in Figure 4.2 where in Figure 4.2b no shoeprint can be recognized meanwhile in Figure 4.2c outlines of the pattern are visible. This observation is explainable with the properties of the LBP descriptor: a

4. RESULTS AND DISCUSSION

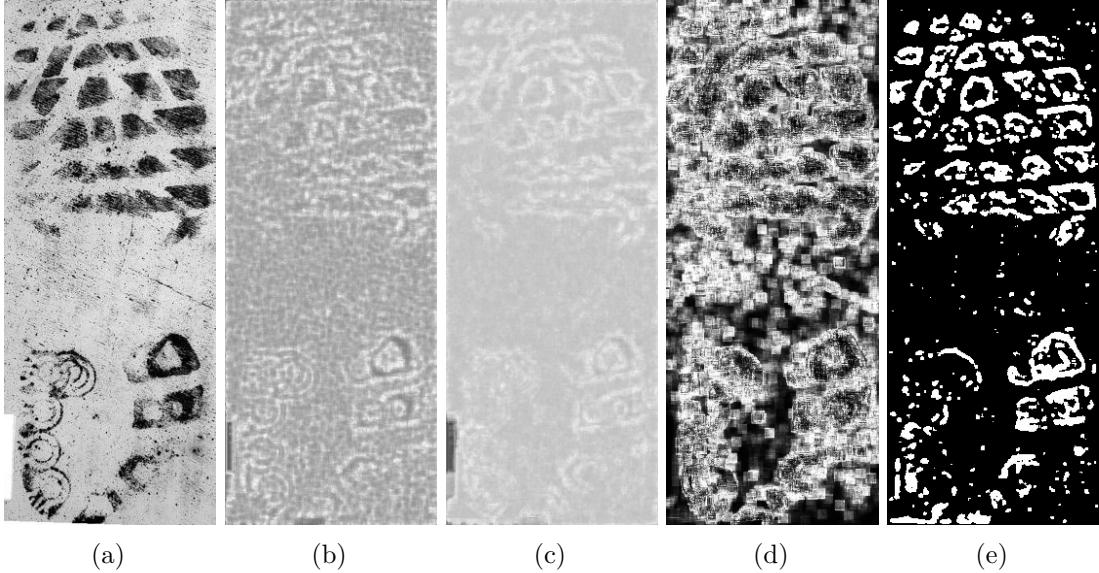


Figure 4.1: Output of the feature learning algorithm on an image from training set.a Example image from the training set b Output using LBP descriptors with radius 3 and 12 sample points c Output using LBP descriptors with radius 5 and 24 sample points d Output using Fourier-Mellin Descriptor e Output using SIFT Descriptor

smaller neighborhood is more suitable for detailed, high-frequency textures, however, it is also more sensitive for noise. A shoe outsole pattern usually consists of bigger structures, and high noise ratio in the shoepoint images is also common, thus a bigger neighborhood is more suitable. Comparing Figure 4.1b and Figure 4.2b the shoepoint pattern on the first one is better visible, since the original image, Figure 4.1, is more clear than Figure 4.2. Still focusing on the middle area of Figure 4.1, Figure 4.1e shows that the SIFT descriptor is similarly robust against noise in the background as the LBP descriptor. SIFT is a robust feature descriptor, therefore it is not surprising that it successfully distinguishes between shoepoint and background in such a clear image like Figure 4.1. However, examining Figure 4.1d the Fourier-Mellin descriptor has more difficulties with the background area, where the middle regions is less homogen than on Figure 4.1c and on Figure 4.1e. On the other hand, inspecting the bottom left side of the shoepoint, it is seen that the Fourier-Mellin features managed to find the whole area of pattern, whereas only outlines are recognizable in Figure 4.1c and in Figure 4.1e. That indicates that the Fourier-Mellin features are more detailed descriptors than LBP and SIFT. The top part of the shoepoint consist of bigger blobs, whereas on the bottom thin lines are seen. The LBP and SIFT features mistake those lines as noise, because LBP considers a bigger neighborhood and the robustness of the SIFT descriptor prevents to correctly recognize those fine edges. Fourier-Mellin works on a smaller neighborhood, thus the detailed shoepoint parts are more visible. On the other hand, it also causes significantly more noise in the middle and in the top of the image. The same phenomenon is seen

in Figures 4.1b and 4.1c, where the bottom left regions is clearer on the image with smaller neighborhood, see Figure 4.1b, but the background is more homogeneous on the latter, see 4.1c. Based on these observations, the features less robust against noise find a higher portion of shoepoint area than those with higher robustness. In exchange, the background area stays noisy and further image processing is needed.

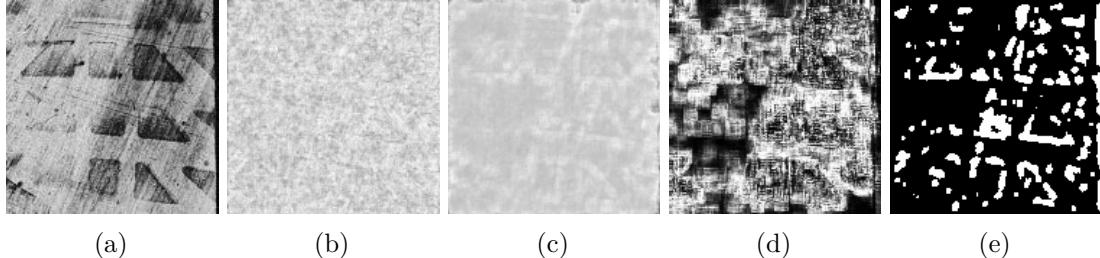


Figure 4.2: Output of the feature learning algorithm on an image from training set. a Example image from the training set; b Output using LBP descriptors with radius 3 and 12 sample points; c Output using LBP descriptors with radius 5 and 24 sample points; d Output using Fourier-Mellin Descriptor; e Output using SIFT Descriptor

In Figure 4.2 the pattern is less recognizable than on the previous example and a higher amount of background pixels are labeled wrongly as pattern. If the input image is noisy, as in Figure 4.2, the results are less clear, and LBP and SIFT seem to outperform the Fourier-Mellin descriptors. In Figure 4.2d the shoepoint is barely visible, whereas in Figures 4.2c and 4.2e outlines are recognizable. This is explainable with the different training process for Fourier-Mellin and for the other two features. The selection process for Fourier-Mellin features was the same as suggested by Guo et al. [GZP12]. However, this pipeline were modified for LBP and SIFT features where the descriptors of the annotated background area were also extracted and explicitly eliminated from the final feature pool. The results prove, that this modification provide a more robust descriptor set than the originally proposed feature training steps.

Comparing all results of the two images, the performance depends on the input quality. A noisy input, Figure 4.2, produces less exact results than a clear sample, Figure 4.1. Feature learning makes it possible to label, to indicate the shoepoint area and distinguish it from background, in lower quality images as well, see Figure 4.2, but in comparison to a higher quality input, see Figure 4.1, the accuracy is also lower. That means that the feature learning algorithm does not completely overcome the negative influence of noise, but it makes the descriptors more robust against it.

In the second testing scenario an evaluation image set is defined, which consist of non-training shoepoint images, to examine if the features learned on shoepoint images from one specific shoe outsole pattern are able to describe shoepoints of other shoes as well. In Figure 4.3 the results of the Feature Learning algorithm on a non-training sample are shown. The results displayed in Figure 4.3b and in Figure 4.3c strengthens the observation that bigger neighborhood LBP outperforms the smaller radius descriptor. In Figure 4.4

4. RESULTS AND DISCUSSION

further examples are shown, in Figures 4.4c, 4.4f the original pattern is more visible than in Figures 4.4b, 4.4e, because of the bigger neighborhood settings. Furthermore, examining the middle, noisy part of the images in all three cases a homogenous background is seen containing some falsely labeled pixels. However, even though bigger neighborhood settings provide better results compared to the smaller one, the descriptiveness of the LBP features is limited for shoeprint images. There is generally small contrast between the shoeprint and its background in the LBP images, thus the outlines of the shoeprint are potentially less visible than on the input shoeprint images.

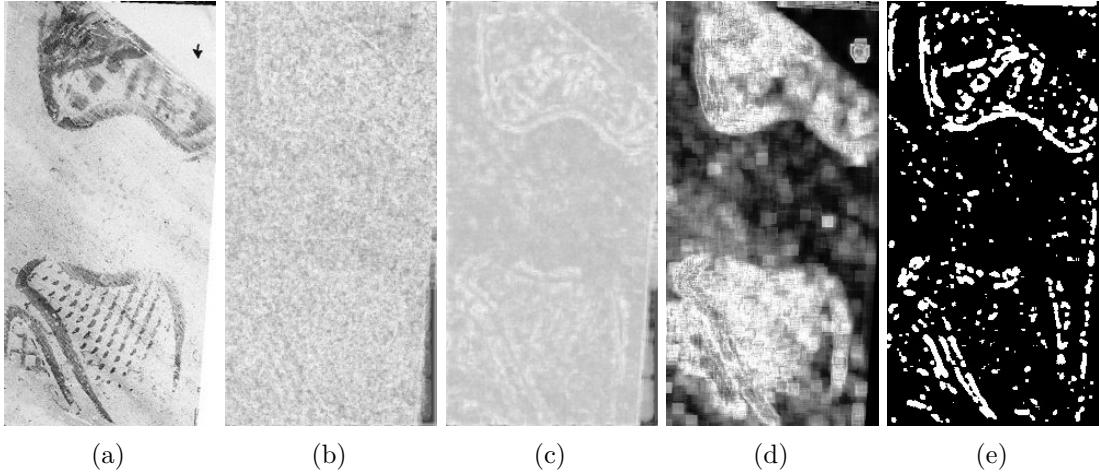


Figure 4.3: Output of the feature learning algorithm a Input image; b Output using LBP descriptors with radius 3 and 12 sample points; c Output using LBP descriptors with radius 5 and 24 sample points; d Output using Fourier-Mellin Descriptor; e Output using SIFT Descriptor

Observing the bottom part of the shoe in Figure 4.3 a weakness of LBP and SIFT already discussed is noticeable. On the original image there are small structures in the bottom area which are part of the shoeprint. Similar to the bottom left part of Figure 4.1, the fine structures are partially recognized. Comparing Figure 4.3d to Figure 4.3c and to Figure 4.3e, Fourier-Mellin outperforms LBP and SIFT again labeling the whole area and not only the outlines correctly. The output of the Fourier-Mellin features, see Figure 4.3d, is more clear than in the previous tests, as seen in Figures 4.1d and 4.2d, whereas no such change is noticeable between LBP and SIFT features. As mentioned earlier, unlike in the case of LBP and SIFT training, the noise descriptors were not eliminated from the feature pool of Fourier-Mellin descriptors. The tests on the training dataset indicated that, by producing less clear images than LBP and SIFT. However, the evaluation set contains unknown noise which were not part of the training set. Therefore, the similar shoeprint elements are recognized and the background is more distinguishable than in the previous cases.

To summarize, the modified learning strategy introduced for LBP and SIFT descriptor provides more robust feature pool than the original one proposed by Guo et al. [GZP12].

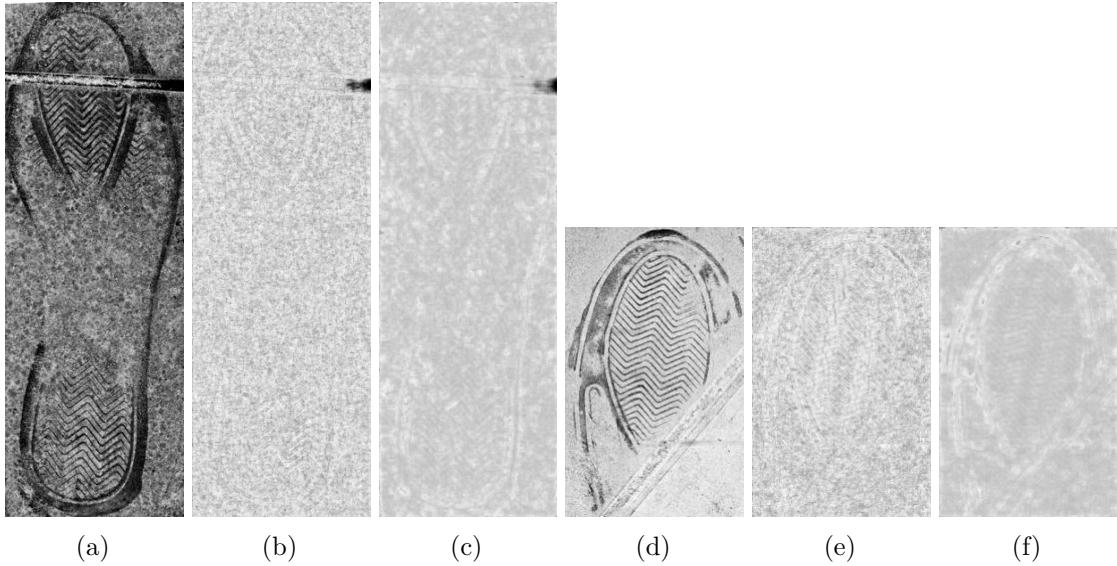


Figure 4.4: a, d Original shoeprint impressions; b, e Output using LBP descriptors with radius 3 and 12 sample points; c, f Output using LBP descriptors with radius 5 and 24 sample points

Because the Fourier-Mellin feature pool was created using the original learning strategy, it calculates clear results on the evaluation dataset and has lower performance on the noisy, training images. However, in both cases it is more responsive on detailed areas with fine line structures, such as the bottom left side of Figure 4.1 and bottom right side of Figure 4.3, than LBP and SIFT.

There are, however, lower quality samples in the dataset available than the one presented on Figure 4.2, where there is small contrast between the pattern and the background, see Figure 1.3b, or there is a high amount of noise on the fore- and the background as well, see Figure 2.1d. The feature pool is created based on a limited training set of seven images. Incorporating more samples and altering the learning criteria, e.g. the given feature has to be found in a given ratio of training images and not on every one of them, leads to a broader feature pool. Containing a higher amount of descriptors, such a feature pool is possibly more descriptive for a wider range of shoeprint images as well. However, this also increases the chance of selecting more noise descriptors into the final feature set.

4.1.2 Fully-Automated Noise Elimination Algorithm

Figures 4.8 and 4.12 show the output of the proposed algorithm on every stage and present the final results as well. The former one shows the output of the denoising step on three example shoeprint images from the FID-300 database. In the second column the Wiener filtered images are shown, in the third one the Bilateral filtered images are displayed whereas the combination of those two pictures is in the last column. Considering the

4. RESULTS AND DISCUSSION

background and the amount of noise in the images, a light smoothing is done during the denoising phase. This decision was made to protect the valuable shoeprint information. Since no previous knowledge is available about the properties of the shoeprint image, such as appearance and location of noise and the shoeprint, no aggressive filter kernel is used. The size of both kernels is small, being only 5x5, to make sure that the relevant shoeprint information is not distorted. In the combined image, the contrast between fore- and background is increased. Since both filters smooth the background area and preserve the shoeprint, their combination amplifies the blurring effect without damaging the shoeprint in the foreground. Despite the amplifying effect, bigger noise elements, such as the white clutter on Figure 4.6a, are not eliminated, because of the weak smoothing. Chatterjee et al. [CM11] proposed a method which calculates the estimated image for Wiener Filter based on geometrically and photometrically similar subregions of the input. With a better estimation strong blurring is possible without damaging the foreground area. Even though their approach had promising results, it was tested on data with Gaussian noise, similar to Figure 4.5a, and not with irregular cluttered noise seen in the majority of FID-300 images, for example in Figure 4.6a and in the top of Figure 4.7a.

The second step of the algorithm is enhancement, where SMQT is applied. The results are shown in Figures 4.9b, 4.10b and 4.11b. The benefits of the SMQT algorithm are ambiguous. On the one hand, focusing on the middle area of Figures 4.9b and 4.10b the difference between fore- and background is bigger than on the original image. The contours of the shoeprint are more outstanding than on the original image. Comparing Figure 4.10b to Figure 4.7d, the shoeprint is darker and there is a brighter area around the contours than on Figure 4.7d. On the other hand, not only the shoeprint but also the remaining noise elements became enhanced by SMQT. Even though the denoising stage is finished, since only moderate smoothing was applied, there is a considerable amount of noise on the input images for SMQT. The soft filtering is able to cope with Gaussian noise, such as on Figure 4.5d, but it does not eliminate bigger clutter particles such as in the left side of Figure 4.6d or in the upper region of Figure 4.7d. The SMQT enhancement makes these clutter more prominent by increasing the color difference between the noise and the homogeneous background. For example, the already mentioned patches on the left side of Figure 4.6d are now more visible than on the original image 4.6a. Additionally, SMQT has similar effects to Histogram Equalization: the background on Figure 4.5d is near homogeneous, but after applying SMQT, its color changes are more obvious since the original pixel values are mapped now to a wider color range. Despite its enhancing effect in the background, SMQT also increases the contrast between the outlines of the shoeprint and its surroundings which is a beneficial feature for the following, local binarization step.

The closing step of the algorithm is postprocessing, consisting of binarization (the third column of Figure 4.12) and line-clutter elimination (in the fourth column of Figure 4.12). As expected NBM, being a local edge sensitive thresholding algorithm, takes advantage of the contrast difference between the outlines of the shoeprint and its neighborhood and finds the silhouettes. On all three images, see Figure 4.9c, Figure 4.10c and Figure 4.11d,

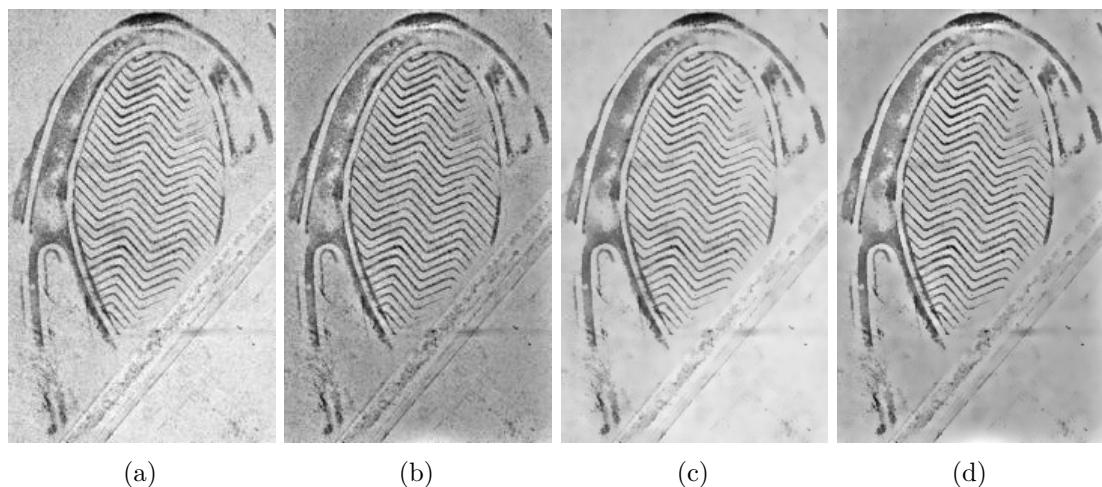


Figure 4.5

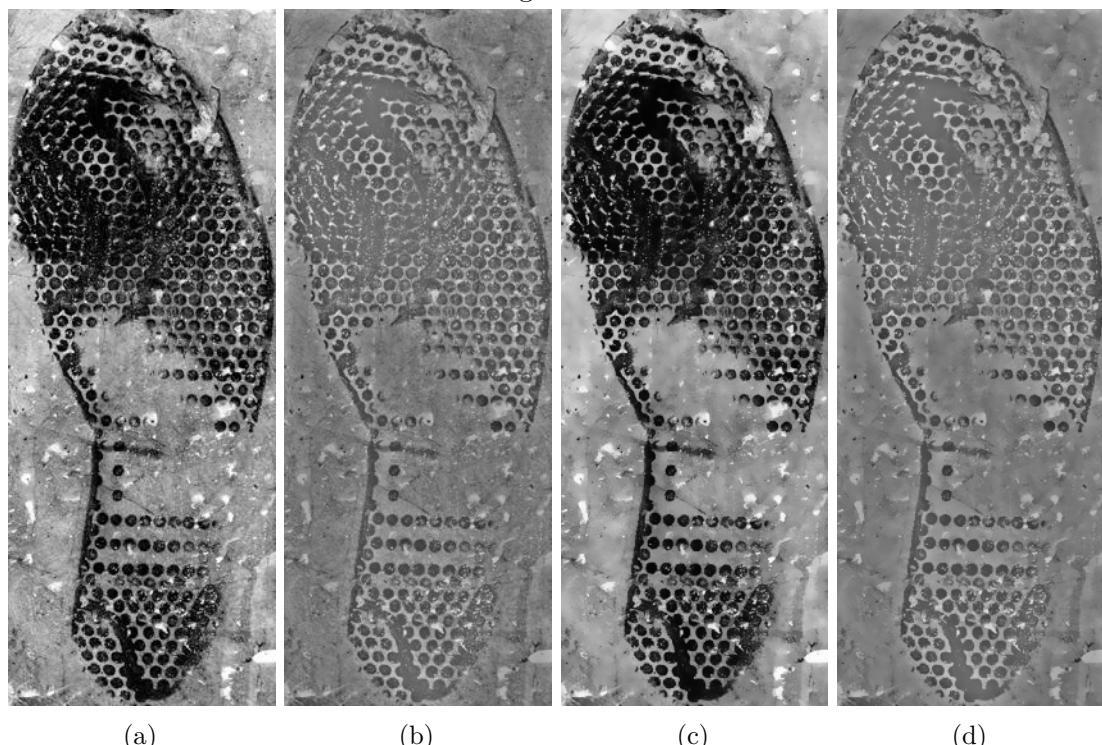


Figure 4.6

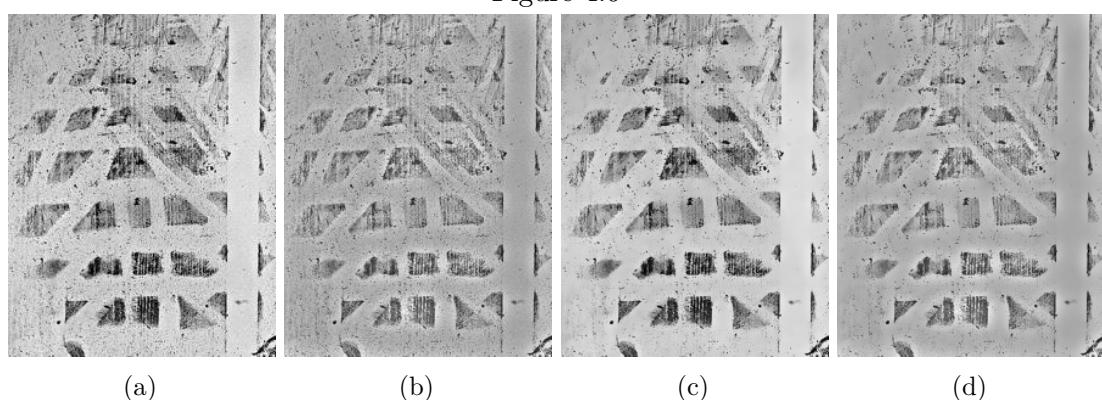


Figure 4.7

Figure 4.8: Example for the results of the denoising stage
a Example image; b Wiener filtered image;
c Bilateral filtered image; d Denoised image

4. RESULTS AND DISCUSSION

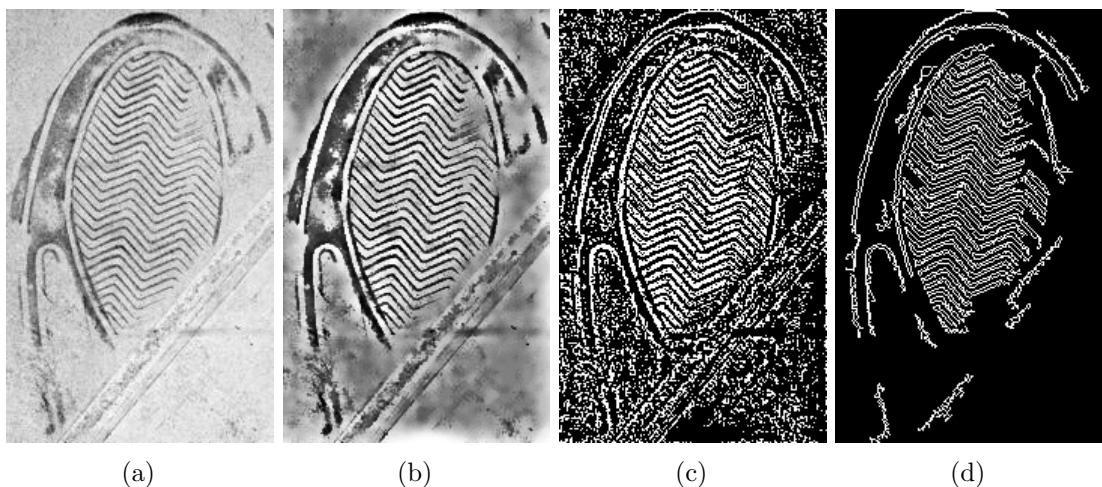


Figure 4.9

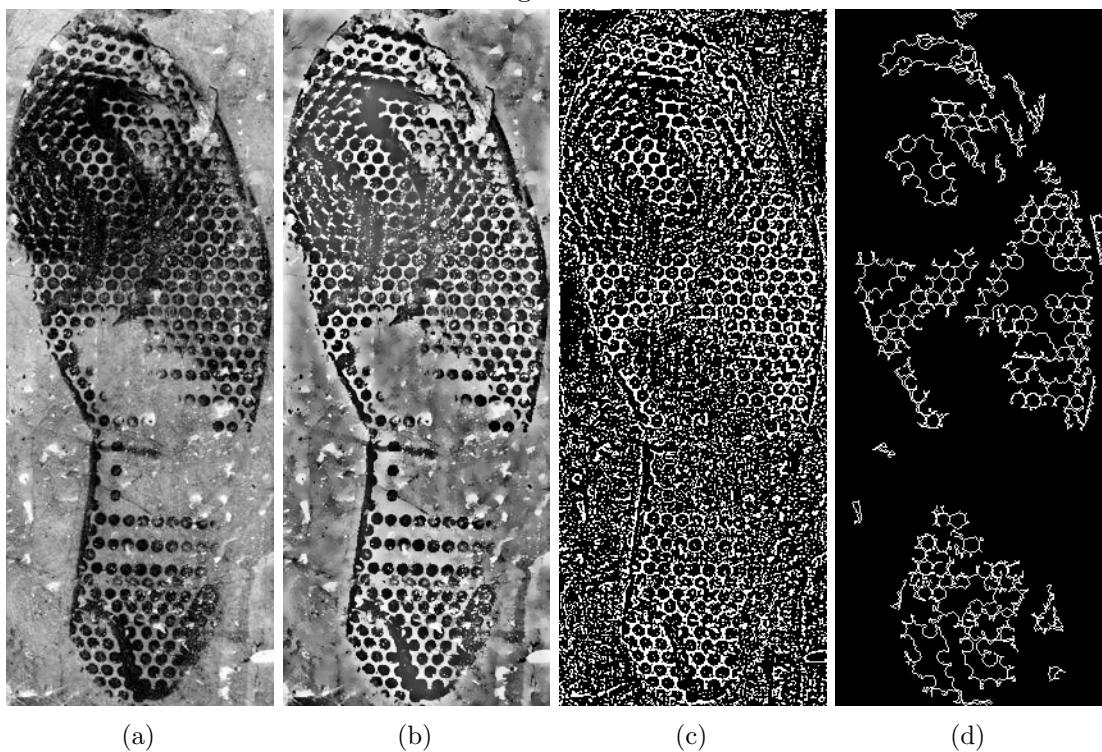


Figure 4.10

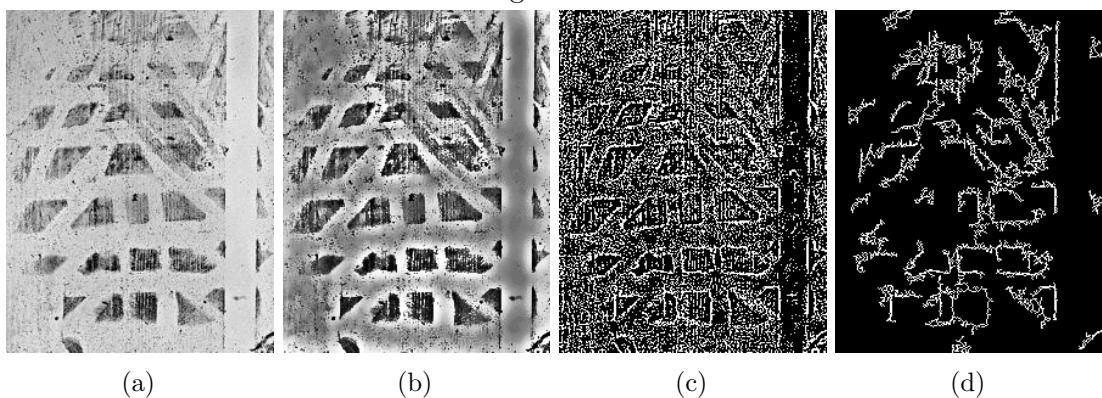


Figure 4.11

Figure 4.12: Example for the results of the enhancement and postprocessing stage a Example image; b Enhanced image; c Binarized image ; d Output image

the contours of the patterns are visible. However, the drawbacks of NBM discussed in the previous chapter and the enhanced noise by SMQT are evident, since among the correct outlines a high amount of noise is generated. In the background and between the space of the outlines of the shoeprint additional small structures are visible. For this reason, the closing step of small and open structure elimination is executed. In Figures 4.9d and 4.10d the majority of background is cleared up and the shoeprint pattern is shown clearly. On the other hand, in Figure 4.11d the original pattern is not recognizable anymore, as along with the background noise, relevant contour edges were also eliminated, and noise edges attached to the shoeprint were kept.

The presented results indicate two things. First, NBM with the current postprocessing technique works better on shoeprint images with fine lines, see Figure 4.5a, and with detailed small structures, see Figure 4.6a than on shoeprint images consisting bigger geometrical elements, such as in Figure 4.7a. As already discussed and shown, NBM tends to generate line noise on homogenous areas. If the shoeprint consists of dense pattern, there is no big, homogenous background between the different structure elements, thus no noise is generated. To solve this problem, Saxena et al. [Sax19] analyze several improvements for NBM to make it more robust in such conditions and in lower quality data. Second, the heuristics which the two postprocessing steps are based on do not apply to every shoeprint image of the database. In Figure 4.11c, even though there is a high amount of noise, the shoeprint is clearly recognizable, whereas in Figure 4.11d no shoeprint is seen anymore. This implies that the settings of the structure elimination algorithms are not detailed enough, and that more sophisticated postprocessing is needed, which covers a wide range of shoeprint images of the database. Several shoeprint edges were eliminated because they were too short. The threshold for edge length were not changed during testing, thus a possible solution for the problem is to always adjust the parameter according to the current image. This adjustment is however difficult, since there is no a-priori knowledge about the given shoeprint image available. Alternatively, looking at Figure 4.11d, the outlines of the shoeprint were found originally, but because of the noise they were not connected. In the application only 8-neighborhood is considered, a more allowing connectivity criteria, such as neighborhood in a given range, leads to closed contour edges. But this causes more connectivity among noise as well. Furthermore, it also magnifies the second problem occurring in Figure 4.11d, that noise edges are connected to the contours, thus are not eliminated. In conclusion, an additional postprocessing step is needed to examine the remaining edges and to determine which connection was made unintentionally.

Based on the experiments presented in this section two factors highly influence the output of the algorithm. These are image quality, such as the appearance of noise and contrast between fore- and background, and the properties of the shoeprint, e.g. the density and size of the unit pattern structure. In conclusion, this algorithm is specialized in a given type of shoeprint in its current form, namely the ones with dense patterns or fine-lined structures. In such cases, despite cluttered noise, Figure 4.10a, promising results were generated. To expand the usability of this algorithm for other shoeprint

4. RESULTS AND DISCUSSION

images of FID-300 further improvements, such as detailed noise model for Wiener Filter and variable settings for the postprocessing heuristics, are needed.

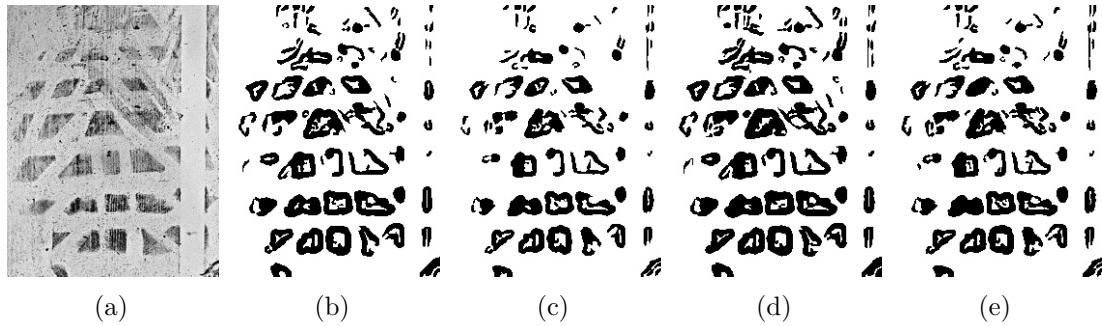


Figure 4.13

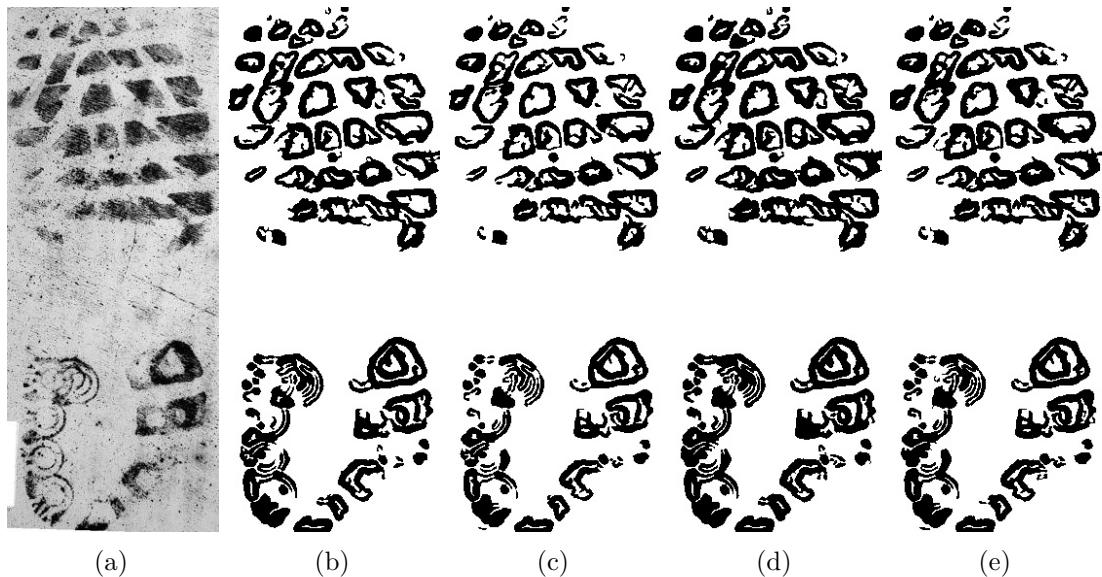


Figure 4.14

Figure 4.15: Example results of the Shoeprint Enhancement algorithm
 a Example image;
 b Enhanced image using average noise model and 20 clusters for gradient classification;
 c Enhanced image using centroids noise model and 20 clusters for gradient classification;
 d Enhanced image using average noise model and 25 clusters for gradient classification;
 e Enhanced image using centroid noise model and 25 clusters for gradient classification

4.1.3 Shoeprint Enhancement with Non-Local Means

In this section the experimental results of the Shoeprint Enhancement Algorithm with Non-Local Means are presented and the performance is discussed. In Figures 4.15 and 4.18 example results are shown, the original samples are in the first column, the results using 20 classes for gradient clustering are shown in the second and third column, the

output using 25 classes are in the fourth and fifth column. Column two and four illustrate the results when the average noise model was used for pattern enhancement, whereas column three and five represent the output in case of using the centroids of the k-means algorithm.

The example images, see 4.15 and 4.18, show that varying the algorithm parameters makes a significant difference when the input had low quality, namely in shoeprint images with low shoeprint-to-noise ratio. This is explained with two factors. First, high quality images have less noise on the shoeprint, see Figures 4.13a and 4.17a, or the noise is less varying, see Figure 4.13a, than on a lower quality sample, as in Figure 4.16a. Thus, using different noise models do not make a difference, since little noise needs to be removed. Second, using 20 and 25 classes for the non-local means algorithm is overstated. In high quality images, there is no high variance between the properties of edges, thus classes with little amount of members are generated. However, choosing high amount of classes was an intentional decision. First, the structure of the shoeprint image is better preserved, when the means are calculated on a high amount of subgroups. Second, using fewer classes leads to eliminating valuable shoeprint information when deleting the members of the most populated class. Comparing the output of the two different kinds of noise models, there is little difference between their performance. Both models were able to preserve the pattern information where there is no, see 4.14a, or little noise, see 4.13a, on the shoeprint area. Focusing on small details, such as the top of Figure 4.16a, the average noise model is more aggressive than the centroid one. That observation corresponds to the calculation of the noise representation. Taking the average of the descriptors of every noise pixel results in a less detailed model than calculating more representatives, thus it is more destructive than the centroid model.

In the images of Figure 4.17a significant shoeprint parts are missing on the bottom area. That is the combined effect of those steps where the heuristic is assumed, that small regions represent noise elements rather than relevant shoeprint information. Therefore the fine-lines of the shoeprint pattern are wrongly considered as noise, and they are eliminated.

In Figure 4.22 the entire pipeline of the Shoeprint Enhancement algorithm is shown in three example images. In Figures 4.19b, 4.20b and 4.21b the binary mask is shown, where white denotes the background and black the shoeprint area. Inspecting the images, the majority of the background is labeled correctly, however, two limitations are visible. The first one is the bottom area of Figure 4.20b, where half of the fine-lined area of the actual shoeprint is cropped. The background detection is not able to distinguish between the structured noise in the background and the detailed shoeprint in the foreground. Furthermore, the fine-lines cover a small area of the three different sized neighborhoods when calculating the correlation with the noise representation, therefore high similarity is noted. In this context the fine-lines of the shoeprint pattern are considered as "noise" when comparing them to the noise representation according to the user input. Because of the robustness of the Fourier-Mellin features, the thin lines of the shoeprint impression have little influence, thus that area is falsely labeled. The other one is in Figures 4.19b,

4. RESULTS AND DISCUSSION

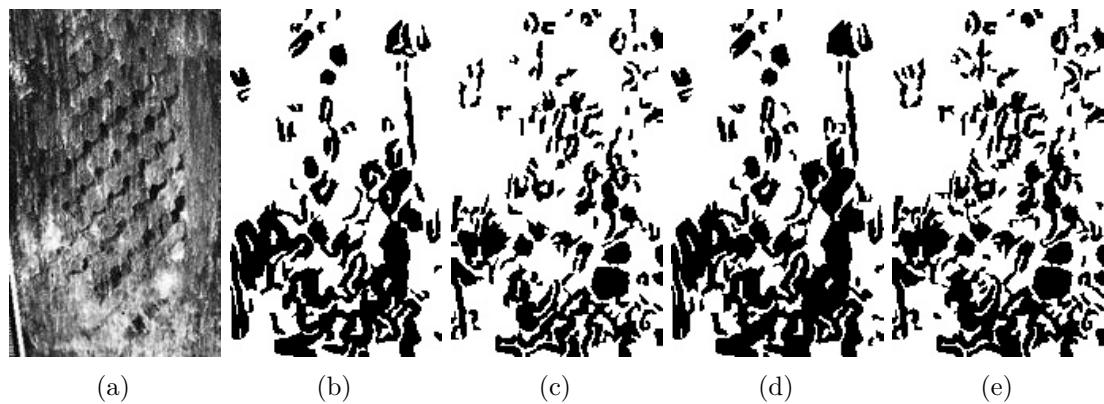


Figure 4.16

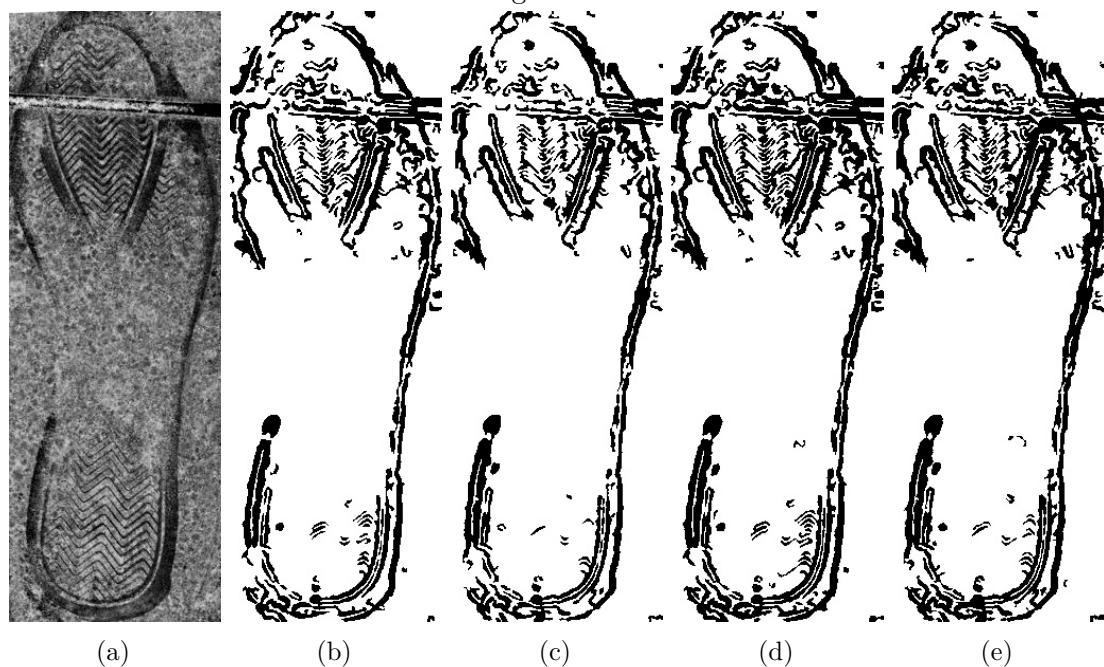


Figure 4.17

Figure 4.18: Example results of the Shoepoint Enhancement algorithm
 a Example image;
 b Enhanced image using average noise model and 20 clusters for gradient classification;
 c Enhanced image using centroids noise model and 20 clusters for gradient classification;
 d Enhanced image using average noise model and 25 clusters for gradient classification;
 e Enhanced image using centroid noise model and 25 clusters for gradient classification

where, because of the significant change within the background, parts of the top area are falsely labeled as foreground.

The appearance of the input image after eliminating the noise in the foreground is

shown in the third column. Figures 4.20b and 4.21b illustrate that the noise suppression technique does not damage the pattern area, the difference between the fore- and the background is preserved. In Figure 4.19a there is significant noise on the top region, which is successfully suppressed in Figure 4.19c. This indicates that, as expected, the noise and the shoeprint have different Fourier representations. Working in the Fourier domain, the noise frequencies are eliminated and the shoeprint is successfully reconstructed.

In the fourth column, the results after applying the non-local means algorithm are shown. Focusing to the background, nearly homogeneous background is created without applying the noise mask on the images. However, in Figure 4.20d a weakness of the algorithm is also visible. The non-local means calculation eliminates the majority of pattern information in both top and bottom area of the original shoeprint. Combined with structured background noise, several mixed classes are generated where noise and shoeprint data occur simultaneously. The artifacts of this phenomenon are visible not only on the said shoeprint area, but also in the middle of the image, where noise structures were enhanced appearing as bright edges similar to the remaining shoeprint edges.

Figures 4.19e, 4.20e and 4.21e show the binarized images after applying the background mask. Because of the nearly homogeneous background achieved by non-local means, the thresholding method is able to preserve the extracted lines. With masking, the falsely clustered edges from the previous step, such as the ones in the middle of Figure 4.20d, are deleted.

In the last column the final results are shown after the postprocessing step of eliminating small regions. Comparing Figure 4.19e to Figure 4.19f the latter is less cluttered, there is less noise in the background, small lines around the contours of the shoeprint are eliminated successfully while no relevant edges, the outlines of the shoeprint, were deleted. However, in Figure 4.20f the shoeprint information is further damaged by the postprocessing step. Observing the right side of the bottom area, it is visible that several shoeprint contours were eliminated. The postprocessing step assumes that the shoeprint consists of bigger structures such as filled geometries on Figure 4.19a or broad edges on Figure 4.21a, thus the fine-lined edges of Figure 4.20a are automatically considered as noise instead of shoeprint.

Along misinterpreting fine-line structures of the shoeprint, there is an additional limitation of the proposed method. When there is low contrast between the background and foreground, or the foreground is highly cluttered, no robust estimation can be made about the background area and about the dominant appearance of the noise. Such a case is shown in Figure 4.23. In this case the output is computed solely based on the gradient-based non-local means method and the postprocessing step. When the determined background area is too small or too big, no masking is made after binarizing the image and no noise is eliminated from the foreground, since no noise model is calculated. In such cases several enhancement steps are not exploited, thus the quality of the output is lower.

The experimental results show that the Shoeprint Enhancement algorithm provides more

4. RESULTS AND DISCUSSION

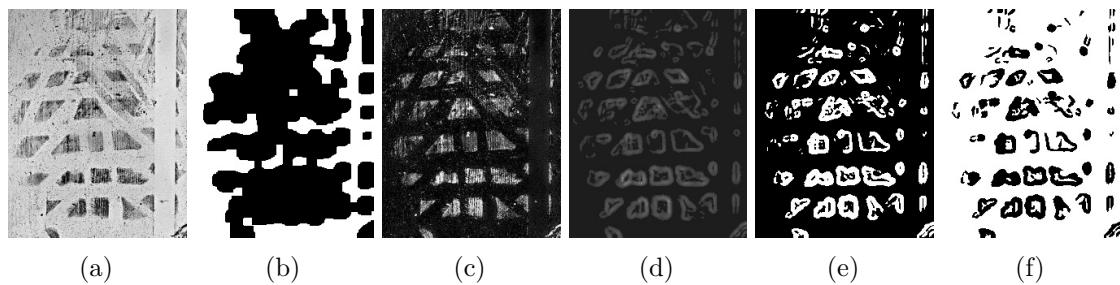


Figure 4.19

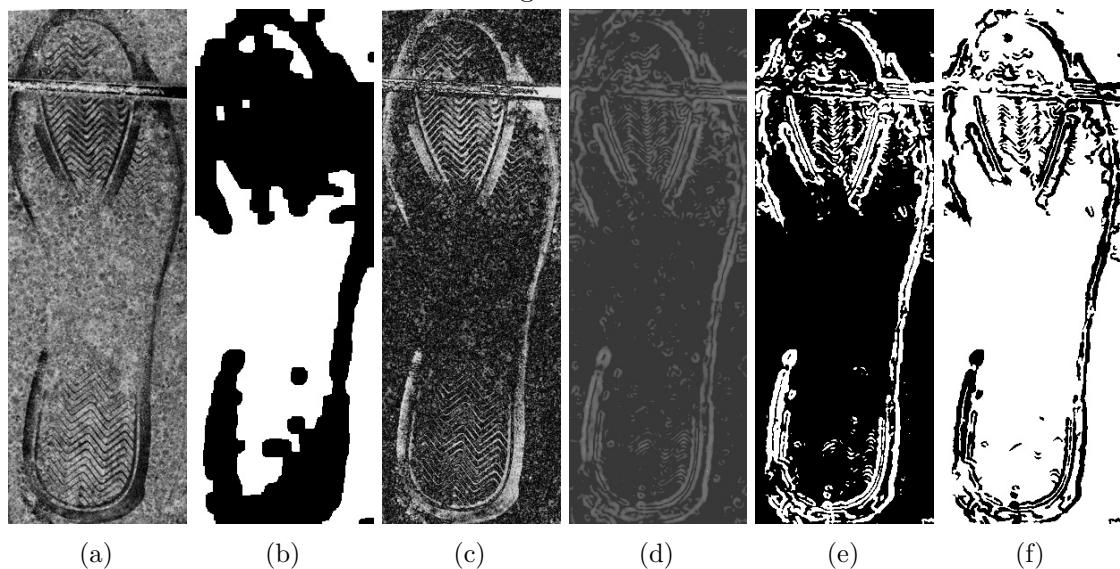


Figure 4.20

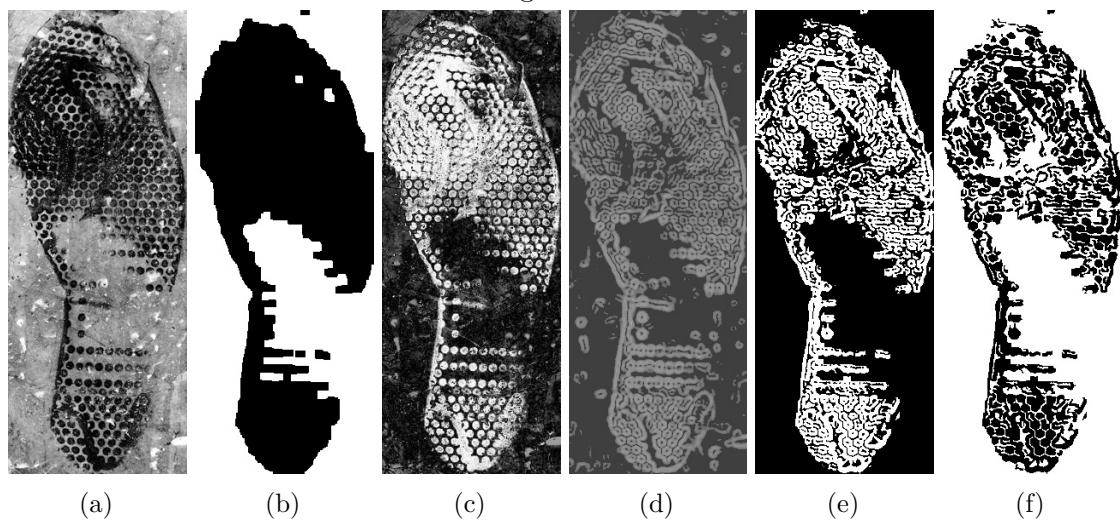


Figure 4.21

Figure 4.22: The whole pipeline of the Shoeprint Enhancement algorithm using 25 classes and centroids for noise elimination on pattern a Example image; b Background mask; c Enhanced image after eliminating noise on the foreground; d Gradient image after calculating non-local means; e Binarized image using adaptive thresholding; f Resulting image after eliminating small structures

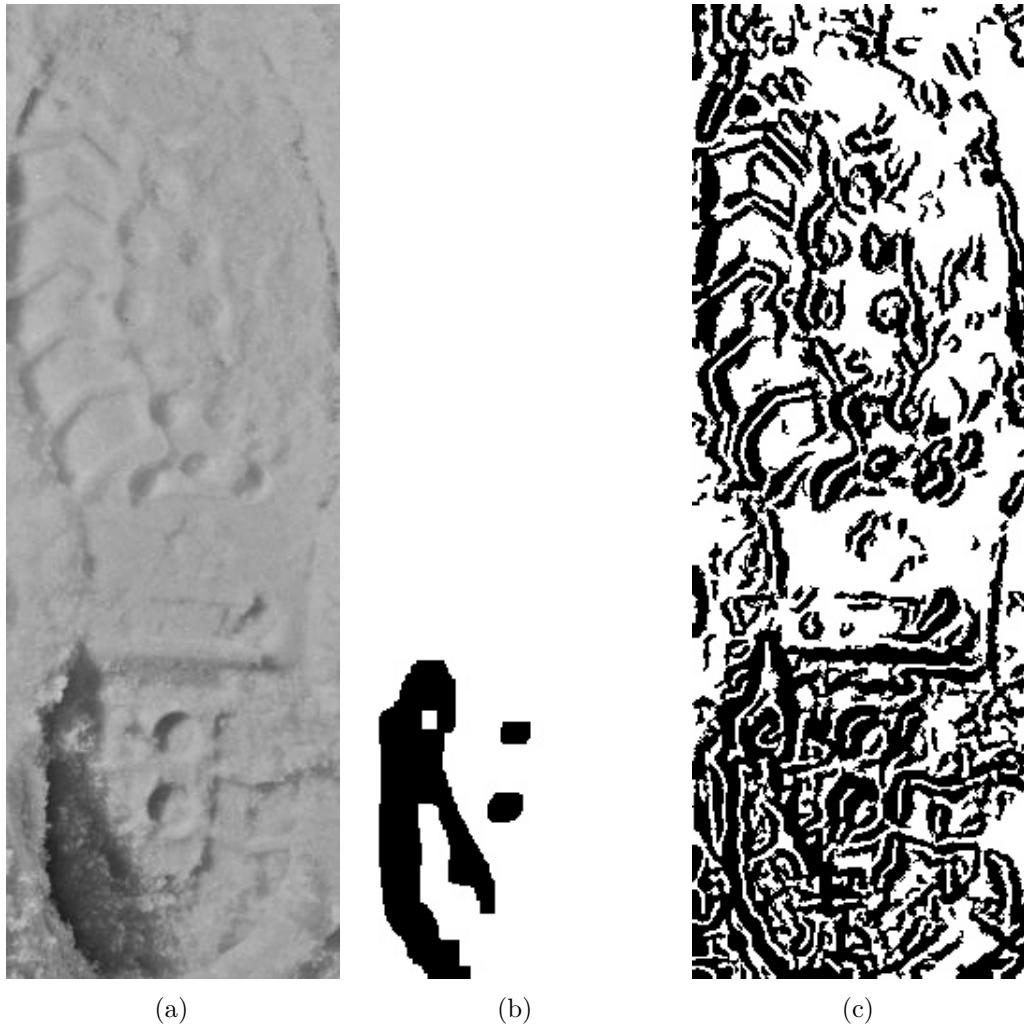


Figure 4.23: Output of the Shoepoint Enhancement algorithm, where no estimation about the background can be made; a Example image from the dataset; b Calculated background mask; c Output of the Shoepoint Enhancement algorithm

precise results, see Figure 4.14e, than the firstly introduced Feature Learning approach, see Figure 4.1d. Furthermore, as in Figure 4.19f, it copes better with noise than the fully-automated method, see Figure 4.11d. On the one hand, the Noise Elimination algorithm achieves better results on detailed, fine-lined shoepoints, as seen on Figures 4.10d and 4.9d than the Shoepoint Enhancement one, see Figure 4.20f. On the other hand, the Shoepoint Enhancement algorithm has overall higher performance than the two previous ones. First, it is able to deal with low-quality samples, see Figure 4.16e, as well. Second, its effectiveness is not limited to one kind of shoepoint outsole pattern such as the previous approach, which only provided good results on densely structured shoepoints, for example Figures 4.10d, 4.9d. Based on the evaluations in the last three sections,

the last approach proved to be the most powerful technique from the three proposed algorithms. To get a detailed overview about its performance, additional quantitative tests are conducted in the following section, and the results are further discussed.

4.2 Quantitative Evaluation

The visual improvements resulting from the three different algorithms were discussed in the previous sections. According to the qualitative evaluation, the Shoeprint Enhancement Algorithm with Non-Local Means had the most promising results, since it provided more accurate shoeprint layout than the Feature Learning algorithm and is less shoeprint specific than the Noise Elimination pipeline. Therefore further tests are conducted. In this chapter the quantitative results on a selection of images from FID-300 are presented and discussed. First, the testing dataset is described, afterwards, following the presentation of the evaluation method, the quantitative results on the dataset are discussed.

4.2.1 Testing Dataset

The testing database consist of 20 images. The selected images and the output of the algorithm are shown in the Appendix 5. Low quality and heavily cluttered images are also part of the experimental database. Since the evaluation in the previous chapter shows that the Shoeprint Enhancement approach is not effective on samples with detailed, fine line structures, those images 12, 14, 19 are excluded from the evaluation. Furthermore, it was mentioned as well that in case of low intensity difference no background estimation is possible. Therefore images 7, 13 and 15 are also disregarded during the evaluation. Additionally, the corresponding reference images of every member of the minimal database were also selected.

4.2.2 Evaluation Method

For evaluation the matching accuracy of the original and the enhanced images is compared based on three basic descriptors. If the basic descriptors, without embedding a more complex pipeline, achieve higher accuracy on the processed images than on the original ones, it implies successful enhancement. The experimental evaluations conducted in the field indicate [RBCP19] that using a complex approach increase matching accuracy. However, when one given pipeline is used for evaluation, it is not clear, if the enhancement is optimized for that specific pipeline or whether it is generally valid. Evaluating based on the feature descriptors which are the core parts of many image matching algorithms provides general results on the performance.

The descriptors used for matching are Fourier-Mellin features, SIFT and SURF, which are popular techniques for shoeprint, fingerprint and for natural image description. Fourier-Mellin transform was proposed for shoeprint description several times [GBCN08], [RLL⁺17], [WWZ19]. Along shoeprint description [NBC⁺09], [RLL⁺17] SIFT is also favored for fingerprint [ZSL⁺11] and for tattoo identification [YYXK15], [HJ13]. Al-

though SURF is not common for shoeprint recognition, it is popular in related areas, such as fingerprint description [JCI17], tattoo matching [YYXK15] and natural texture description [PK12]. The descriptors are extracted from every 4th pixel in both x and y direction.

For matching the extracted features, two algorithms are used: Brute-Force matching (BF) [SLSL93] and Fast Library for Approximate Nearest Neighbors (FLANN) [ML09]. As mentioned in the previous chapter, BF compares a feature of one image to every other features of the other one, and sets the closest one as match, whereas FLANN picks the best matches from the neighborhood, it does not examine all possible feature points, so it is faster than BF. In both cases Euclidean distance is used to compare the descriptors. To filter the provided matches Lowe's ratio test is used [Low04]. To do so, the distance ratio of the two nearest matches to a given feature is examined, and it is accepted if it is below a given threshold. The criteria for 'good match' is set to 0.8 for testing. It is examined how many 'good matches' are there between the shoeprint- and the reference images separately. During the matching process, the shoeprint image is compared to every reference image. At the end, the reference images are ordered according to the amount of 'good matches' to the same shoeprint image. The one with the most 'good matches' is the first, the one with the second most 'good matches' is the second and so on. The rank of the Ground Truth reference image of the given shoeprint image is noted, and the final score is made by averaging the noted ranking values of every image from the testing dataset. That means if the Ground Truth image of every member of the dataset had the most 'good matches', thus the best ranking, the final score is 1. Lower performance is indicated with higher score, correct matching results with lower score.

4.2.3 Evaluation

Now the quantitative results of the Shoeprint Enhancement Algorithm with Non-Local Means method are evaluated based on the performance of the three basic feature descriptors. To create a baseline the original shoeprint images were matched to the reference images. For reproducibility the scores of all four possible settings of the enhancing algorithms and of the baseline are given in the Appendix 5. Table 4.1 gives an overview of the accumulated performance of the discussed settings. The tested parameter settings are 20 classes for the non-local means clustering using average noise model (20A) or using the centroid noise model (20C) and 25 classes for the non-local means clustering using average noise model (25A) or using centorid noise model (25C). In the upper side of the cells the average score of both matching techniques, BF and FLANN, are shown separately. In the lower part of the cells the average score of both matching techniques is displayed. In the last column the average score of all descriptors, F-M, SIFT and SURF, using both matching techniques is shown. In the table, multiple cells are highlighted, the one in the first row shows the best score in the baseline, 4.43, provided by FLANN matching the SIFT descriptors. The other gray cells show the scenarios where the enhanced images did not achieve the best performance of the baseline. The cells with blue background highlight the scenarios, when the enhanced images achieved and outperformed the best

4. RESULTS AND DISCUSSION

baseline result, darker color indicates better accuracy. Bright blue is used for scores above 4, middle blue indicates results between 4 and 3.50, and better performance than that is highlighted with dark blue. The accumulated values, the average of the results, is not highlighted. To understand the performance of the proposed enhancing method, this table is analyzed in the followings.

Focusing on the first (F-M) and third (SURF) column of the table, the matching of the enhanced shoeprints always outperforms the corresponding baseline regardless the pipeline settings. Furthermore, the Fourier-Mellin descriptor achieved better results than the best baseline except for the one case when 20A is used. However, in the second (SIFT) and third (SURF) column of the table this is not the case. Even though SURF always outperformed the corresponding baseline, which is 5.64, in case of 25C and 20C the accuracy of the overall best baseline performance of 4.43 is not met. In case of SIFT, the results of the enhanced images are significantly lower than of the original images. Using BF matching, the enhanced images outperform the corresponding baseline, but achieve only two times out of four better accuracy than the overall best baseline result. In case of FLANN matching, the enhanced images always have lower accuracy than the corresponding baseline, which is the best performance noted for the original images. Since the enhanced images are binarized, the usage of SIFT and SURF is not ideal. SIFT is a gradient based descriptor and the majority of that information is lost when thresholding an image. Furthermore, local invariant descriptors, such as SIFT and SURF, usually have lower performance on binary images than on grayscale samples [WZWS17]. However, SURF is generally more robust than SIFT which also explains its better performance even using on binarized images [BTVG06].

	F-M		SIFT		SURF		Average Results
	BF	FLANN	BF	FLANN	BF	FLANN	
Baseline	5.86 5.86	5.86	5.57 5.00	4.43	6.50 6.07	5.64	5.64
25 Average (25A)	3.86 3.82	3.79	4.21	5.57	3.93	4.43	4.30
25 Centroids (25C)	3.29 3.25	3.21	4.29	5.14	4.43	5.36	4.29
20 Average (20A)	4.43 4.50	4.57	4.93	5.21	3.00	3.64	4.30
20 Centroids (20C)	3.29 3.18	3.07	4.43 5.07	5.71	3.71	4.50	4.12

Table 4.1: Overview about the performance of the analyzed pipelines

The Fourier-Mellin features (first column) always outperformed their corresponding baseline, and failed only once to meet or outperform the best accuracy achieved by the original images. The two lowest results of the Fourier-Mellin descriptors were provided using 20A settings. Furthermore, the 25A settings only outperformed 20A, but no other testing scenario. The average noise model caused, thus, lower performance. As discussed

in the previous section, the average noise model is more destructive than the centroid model. Using the centroid representation, the used noise model is adjustable to the given part of the shoeprint image. In this way the chance is lower that shoeprint information is also eliminated. Using the centroid model causes at least 0.5 points better score against the average model by comparing the best, 3.79, and the worst, 3.29, accuracy values achieved by average and centroid noise models respectively.

Considering the overall accuracy (last column) of the different settings, the enhanced matching always provides better results than the accumulated baseline. Furthermore, the overall accuracy is always higher than the best result in the baseline. Comparing the variable settings, there is small difference between the best and the worst scores. 20C provided the highest performance achieving 4.12, whereas the other scenarios achieved 4.30 and 4.29 accuracy only. That means, that using less classes for the non-local means algorithm and calculating a detailed noise model leads to better accuracy. The detailed noise model led to slightly better results, even though higher amount of classes were defined. In case of 25C the performance of 4.29 is noted, whereas the other scenarios using the average model achieved 4.30. Clustering the edges into less classes enforces the elimination of more edges. Since the cluster with the highest cardinality is assumed to be a noise class, owing the fact that using less classes the clusters have higher cardinality in average than using more classes, the chance is higher that more noise-edges are eliminated. The number of classes is, however, a crucial parameter. Lowering this value more mixed clusters are created where noise and shoeprint edges are placed in the same cluster. In that case, when the assumed noise cluster is eliminated, valuable parts of the shoeprint are also deleted. Moreover, using little number of classes further details of the input are also eliminated when applying the non-local means algorithm. The good performance using 20 classes indicates that the amount of classes is sufficient, and do not cause the elimination of potential shoeprint information.

The experimental results presented in this chapter show that the proposed image enhancement algorithm supports accuracy improvement. The performance of the Shoeprint Enhancement algorithm was tested on a versatile subset of images selected from the FID-300 dataset. However, the proposed method needs further improvement to be able to handle all possible shoeprint images of the FID-300 dataset. For known limitations of the algorithm discussed in the previous section, shoeprint images with detailed, fine line structures and low contrast between fore- and background were excluded from testing. The robustness and the high-quality of Fourier-Mellin features is promising, since many current shoeprint matching algorithms already use them for feature description [GBCN08], [RLL⁺17], [WWZ19] or are based on other frequency domain descriptors [AH08], [WSYZ14], [KS17].

In summary, all three presented methods delivered promising results in given scenarios. Based on the evaluations, the Shoeprint Enhancement Algorithm with Non-Local Means is the most promising one, since unlike the Noise Elimination Algorithm it is not limited to one kind of shoeprint, and compared to the Feature Learning Algorithm it delivers more exact and detailed output, see Figures 4.15 and 4.1. Additionally, combining all

4. RESULTS AND DISCUSSION

three approaches the discussed advantage of one technique can suppress the limitation of the other two. Even though the Shoeprint Enhnacement Algorithm with Non-Local Means outperforms the other two methods, being semi-automated and not being able to cope with fine line structures, it has nevertheless limitations. But letting the Feature Learning algorithm to determine the approximate outline of the shoeprint, and calculating the exact background with the currently used method in the Shoeprint Enhancement Algorithm replaces the now necessary user input at the beginning of the algorithm. Additionally, the introduced Noise Elimination method only provided reasonable output, if the input had high quality and detailed line structure. That is, however, the case that the Shoeprint Enhnacement pipeline is not able to handle. In case of fine-lined shoeprints, processing the shoeprint image with the Shoeprint Enhancement method, and after applying the background mask finishing the calculation with the Noise Elimination approach, potentially solves the above issue. The preprocessing done with the Shoeprint Enhancement method increases the quality that the Noise Elimination approach needs, and the latter one is able to identify the detailed pattern which the Shoeprint Enhancement technique eliminates immediately. In this way the functionality of the proposed algorithms is extended and their current limitations are compensated.

CHAPTER 5

Conclusion

Forensic shoeprint identification is an extensively studied research area [RBCP19], nevertheless there are several possibilities for improvement. In this thesis, three different approaches for forensic shoeprint enhancement are introduced, to increase the contrast between the shoeprint and its background, and to suppress the noise in the shoeprint images. Based on the available shoeprint recognition literature and research done in related fields such as fingerprint and natural image enhancement, three prototypical methods are presented and compared to increase the quality of shoeprint images from the FID-300 dataset.

The first one is based on the idea of feature learning. Samples from the FID-300 dataset were annotated pixel-wise, to accurately distinguish between the fore- and the background area. During the training process, the features of the same shoe outsole pattern on different shoeprint images are extracted and compared. First, descriptors are extracted from the fore- and the background separately, and those foreground descriptors which are similar to at least one background feature are eliminated. Hence, by noise distorted descriptors are exuded. Afterwards, the remaining descriptors are compared and added to the final selection, if they occur in every shoeprint image of the same shoe outsole pattern. In this way the robust descriptors are selected. When processing a shoeprint image, the descriptors are extracted from every pixel and their similarity to the learnt descriptors is examined. High similarity indicates shoeprint information in that area, low similarity stands for noise and background.

The second proposed algorithm enhances the shoeprint in three steps. First, the shoeprint image is smoothed using Wiener and Bilateral filter. Using edge and information sensitive filters with small kernel settings, it is secured that the outlines of the shoeprint are preserved. Afterwards, the contrast between the shoeprint and its background is increased using the Successive Mean Quantization Transfrom. This step also prepares the shoeprint image for binarization. The image is thresholded using the local Niblack Binarization Method. To eliminate the additional edges generated by residual noise and

5. CONCLUSION

thresholding artifacts, heuristics are applied which delete small, disconnected edges in the image.

Lastly, a shoeprint enhancement method based on the non-local means algorithm is presented. At the beginning, noise descriptors are extracted based on the multiple-sized neighborhood specified by user input. Based on the descriptors, the background area of the shoeprint image is determined and a noise model is built. The noise on the foreground, where the shoeprint is assumed to be located, is suppressed by subtracting the Fourier representation of the noise model from the Fourier representation of the foreground in the frequency domain. The outlines of the shoeprint are then enhanced, contrast between them and their surroundings is increased, by applying the non-local means method. The classification criteria is the gradient information of the pixels. Gradients provide information about the edge properties of a given area, thus, assuming that the shoeprint pattern and the random noise on the image have different edge features, the elements of the shoeprint are processed separately from the noise. Finally, the binary image is calculated using adaptive thresholding.

Since there is limited research in the area of forensic shoeprint enhancement available, the presented approaches show three possibilities to develop a methodology for the enhancement of shoeprint images. The qualitative evaluation of the algorithms shows that all three methods provide quality improvement, namely, suppression of noise, and clearing or extracting the outlines of the shoeprint. Based on the qualitative evaluation, the enhancement with non-local means is the most promising, since it provides more accurate outlines, and it is not limited to given properties of the shoeprint with dense pattern and fine-line structures. The quantitative evaluation performed with descriptors already used in forensic image identification shows that the enhanced images achieve better matching results. With further development and potential combination of the proposed methods, a powerful and versatile preprocessing pipeline can be built which is able to enhance shoeprint impressions, thus increase the matching accuracy.

Future Work

There are several approaches in the research available which were not discussed in this work yet. Chen et al. [CHZ13] introduced a hierarchical voting score for finger- and palmprint identification. Wu et al. [WWNZ19] also proposed a voting technique for real-life forensic images incorporating the opinion of human experts. They defined a manifold ranking method based on the shoeprint identification of forensic specialists. Along the judgments of the experts, the neighborhood and the coefficient matrix of the given shoeprint image is also considered to develop a powerful feature descriptor. The descriptor is based on the Fourier-Mellin Transform similar to the approaches previously introduced, thus a combination is possible.

Voting scores and feature-learning are machine learning techniques, but deep learning [LBH15] was not considered either. Rida et al. [RBCP19] discuss the possibility of using Convolutional Neural Network (CNN) for end-to-end shoeprint recognition and identifi-

cation introduced in [LBB⁺98]. Kong et al. [KSRF17], [KSRF19] already published an approach for taking advantage on CNNs. However, instead of using the CNN to identify the given shoeprint, they extract mid-level features of the network to use them as feature descriptor. Even though there are publications for natural image enhancement using deep learning [GCB⁺17], [CWKC18], it has to be investigated, how their tools can be used for forensic shoeprint image.

In all approaches proposed in this thesis, the geometrical location of the extracted features is disregarded, even though it is an important information, while identifying shoeprint images. Despite that several samples depict only a partial image, as soon as there are at least two different patterns or a recognizable outline part of the shoe is visible, assumptions are possible about the geometrical location of the given pattern of the shoeprint. Li et al. [LW15] propose a subdivision method where four regions of a shoeprint are distinguished, and processed according to their position. These are cap, sole, arch and heel. Additionally, the overall silhouette of a shoeprint also contains information. Wang et al. [WSYZ14] developed a shoeprint contour model for shoeprint retrieval on high-quality samples, which helps to make their approach more robust, since it is able to correct distortions or noise on the input shoeprint images.

Lastly, Randomly Acquired Features (RAC) of shoeprints were already mentioned before, but were not considered during feature extraction. As Rida et al. [RBCP19] states, in many dataset only a limited amount of shoeprint images are available for one shoe outsole pattern which is also the case in the FID-300 database. The majority of the dataset consists of single shoeprint images, and there are no more than six shoeprint images from the same shoesole in exceptional cases. However, assuming that the demand for wider dataset increases with active research, considering RAC features of the input provides valuable additional information about the given shoeprint image. Shor et al. [SWT⁺18] introduce a technique to distinguish between the shoeprint images of the same shoesole based on the obtained RAC features. Damary et al. [DMW⁺18] published a study about the estimated properties of a RAC, according to its location on the shoeprint. This way if a RAC is identified, assumptions about the location of the neighboring pattern are possible. Similar to basic structure description for shoeprints [TSKC10], a technique for RAC description was also developed [SRF⁺16], thus incorporating RAC models into shoeprint descriptors is already possible.

Considering the general research done on the topic of shoeprint enhancement, there are general issues that have to be solved in the future. The current practice of using several datasets has several drawbacks. The proposed results are not meaningful and consequential, since the used datasets are not public in many cases, and because the results are not comparable. Furthermore, the real-life performance and usefulness of the published methods is unknown, because they are often tested on synthetic and on handcrafted data. Rida et al. [RBCP19] recently proposed a survey about the research done in the last 20 years, and made an overview about the accuracy of the published methods. Several authors claimed to have above 99% [AH08], [AK17], [ABCN15], or exactly 100% accuracy [GBC07] on their testing datasets. However, the algorithms were

5. CONCLUSION

tested on different datasets, thus an absolute comparison between their performance is not possible.

Bibliography

- [ABCN15] Somaya Almaadeed, Ahmed Bouridane, Danny Crookes, and Omar Nibouche. Partial shoeprint retrieval using multiple point-of-interest detectors and sift descriptors. *Integrated Computer-Aided Engineering*, 22(1):41–58, 2015.
- [AH08] Gharsa AlGarni and Madina Hamiane. A novel technique for automatic shoeprint image retrieval. *Forensic science international*, 181(1-3):10–14, 2008.
- [AK17] Sayyad Alizadeh and Cemal Kose. Automatic retrieval of shoeprint images using blocked sparse representation. *Forensic science international*, 277:103–114, 2017.
- [Ale96] Girod Alexandre. Computerized classification of the shoeprints of burglars' soles. *Forensic Science International*, 82(1):59–65, 1996.
- [AMHP09] Timo Ahonen, Jiří Matas, Chu He, and Matti Pietikäinen. Rotation invariant image description with local binary pattern histogram fourier features. In *Scandinavian conference on image analysis*, pages 61–70. Springer, 2009.
- [Ath11] Maythapolnun Athimethphat. A review on global binarization algorithms for degraded document images. 2011.
- [BCM05] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65. IEEE, 2005.
- [BCPR12] Rudolf Maarten Bolle, Sharat Suresh Chikkerur, Sharathchandra Umapathirao Pankanti, and Nalini Kanta Ratha. Fingerprint representation using localized texture feature, May 15 2012. US Patent 8,180,121.
- [BHK15] Abir Raza Baig, Ilyas Huqqani, and Khurram Khurshid. Enhancement of latent fingerprint images with segmentation perspective. In *2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 132–138. IEEE, 2015.

- [BK16] Anu Bala and Tajinder Kaur. Local texton xor patterns: A new feature descriptor for content-based image retrieval. *Engineering Science and Technology, an International Journal*, 19(1):101–112, 2016.
- [BM01] Michael V Boland and Robert F Murphy. A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of hela cells. *Bioinformatics*, 17(12):1213–1223, 2001.
- [BNS07] Sheryl Brahnam, Loris Nanni, and Randall Sexton. Introduction to neonatal facial pain detection using common and advanced face classification techniques. In *Advanced Computational Intelligence Paradigms in Healthcare-1*, pages 225–253. Springer, 2007.
- [BPJH17] Satish Bhairannawar, Apeksha Patil, Akshay Janmane, and Madhuri Huilgol. Color image enhancement using laplacian filter and contrast limited adaptive histogram equalization. In *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*, pages 1–5. IEEE, 2017.
- [Bra00] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.
- [BTVG06] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer, 2006.
- [CHZ13] Fanglin Chen, Xiaolin Huang, and Jie Zhou. Hierarchical minutiae matching for fingerprint and palmprint identification. *IEEE Transactions on Image Processing*, 22(12):4964–4971, 2013.
- [CM11] Priyam Chatterjee and Peyman Milanfar. Patch-based near-optimal image denoising. *IEEE Transactions on Image Processing*, 21(4):1635–1649, 2011.
- [CRT⁺18] A Chahi, Y Ruichek, R Touahni, et al. Local directional ternary pattern: A new texture descriptor for texture classification. *Computer Vision and Image Understanding*, 169:14–27, 2018.
- [CWKC18] Yu-Sheng Chen, Yu-Ching Wang, Man-Hsin Kao, and Yung-Yu Chuang. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6306–6314, 2018.
- [DCC09] Francesca Dardi, Federico Cervelli, and Sergio Carrato. A texture based shoe retrieval system for shoe marks of real crime scenes. In *International Conference on Image Analysis and Processing*, pages 384–393. Springer, 2009.

- [DCFR05] Philip De Chazal, John Flynn, and Richard B Reilly. Automated processing of shoeprint images based on the fourier transform for use in forensic science. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (3):341–350, 2005.
- [DMW⁺18] Naomi Kaplan Damary, Micha Mandel, Sarena Wiesner, Yoram Yekutieli, Yaron Shor, and Clifford Spiegelman. Dependence among randomly acquired characteristics on shoeprints and their features. *Forensic science international*, 283:173–179, 2018.
- [DMY15] Reza Davarzani, Saeed Mozaffari, and Khashayar Yaghmaie. Scale-and rotation-invariant texture description with improved local binary pattern features. *Signal Processing*, 111:274–293, 2015.
- [DVGNK99] Kristin J Dana, Bram Van Ginneken, Shree K Nayar, and Jan J Koenderink. Reflectance and texture of real-world surfaces. *ACM Transactions On Graphics (TOG)*, 18(1):1–34, 1999.
- [Ela02] Michael Elad. On the origin of the bilateral filter and ways to improve it. *IEEE Transactions on image processing*, 11(10):1141–1151, 2002.
- [FAA17] Sadegh Fadaei, Rassoul Amirkattahi, and Mohammad Reza Ahmadzadeh. Local derivative radial patterns: A new texture descriptor for content-based image retrieval. *Signal Processing*, 137:274–286, 2017.
- [FJZY11] Jie Feng, LC Jiao, Xiangrong Zhang, and Dongdong Yang. Bag-of-visual-words based on clonal selection algorithm for sar image classification. *IEEE Geoscience and Remote Sensing Letters*, 8(4):691–695, 2011.
- [GAS12] Azadeh Ghandehari, Mohammad Anvaripour, and Sima Soltanpour. Palm-print verification and identification using pyramidal hog feature and fast tree based matching. In *2012 5th IAPR International Conference on Biometrics (ICB)*, pages 421–426. IEEE, 2012.
- [GBC07] Mourad Gueham, Ahmed Bouridane, and Danny Crookes. Automatic recognition of partial shoeprints based on phase-only correlation. In *2007 IEEE International Conference on Image Processing*, volume 4, pages IV–441. IEEE, 2007.
- [GBCN08] Mourad Gueham, Ahmed Bouridane, Danny Crookes, and Omar Nibouche. Automatic recognition of shoeprints using fourier-mellin transform. In *2008 NASA/ESA Conference on Adaptive Hardware and Systems*, pages 487–491. IEEE, 2008.
- [GCB⁺17] Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédo Durand. Deep bilateral learning for real-time image enhancement. *ACM Transactions on Graphics (TOG)*, 36(4):1–12, 2017.

- [GZP12] Yimo Guo, Guoying Zhao, and Matti Pietikäinen. Discriminative features for texture description. *Pattern Recognition*, 45(10):3834–3843, 2012.
- [GZZ10a] Zhenhua Guo, Lei Zhang, and David Zhang. A completed modeling of local binary pattern operator for texture classification. *IEEE transactions on image processing*, 19(6):1657–1663, 2010.
- [GZZ10b] Zhenhua Guo, Lei Zhang, and David Zhang. Rotation invariant texture classification using lbp variance (lbpv) with global matching. *Pattern recognition*, 43(3):706–719, 2010.
- [GZZL15] Qiang Guo, Caiming Zhang, Yunfeng Zhang, and Hui Liu. An efficient svd-based method for image denoising. *IEEE transactions on Circuits and Systems for Video Technology*, 26(5):868–880, 2015.
- [HA18] Saeed Hayati and Mohammad Reza Ahmadzadeh. Wirif: Wave interference-based rotation invariant feature for texture description. *Signal Processing*, 151:160–171, 2018.
- [HJ13] Hu Han and Anil K Jain. Tattoo based identification: Sketch to image matching. In *2013 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2013.
- [HKWL13] De-An Huang, Li-Wei Kang, Yu-Chiang Frank Wang, and Chia-Wen Lin. Self-learning based image decomposition with applications to single image denoising. *IEEE Transactions on multimedia*, 16(1):83–93, 2013.
- [HMK00] Serkan Hatipoglu, Sanjit K Mitra, and Nick Kingsbury. Image texture description using complex wavelet transform. In *Proceedings 2000 International Conference on Image Processing (Cat. No. 00CH37101)*, volume 2, pages 530–533. IEEE, 2000.
- [HZPC14] Xiaopeng Hong, Guoying Zhao, Matti Pietikäinen, and Xilin Chen. Combining lbp difference and feature correlation for texture description. *IEEE Transactions on Image Processing*, 23(6):2557–2568, 2014.
- [JCI17] Sharmin Jahan, Mozammel Chowdhury, and Rafiqul Islam. Robust fingerprint verification for enhancing security in healthcare system. In *2017 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, pages 1–5. IEEE, 2017.
- [JNDB05] Jan Jantzen, Jonas Norup, Georgios Dounias, and Beth Bjerregaard. Pap-smear benchmark data for pattern classification. *Nature inspired Smart Information Systems (NiSIS 2005)*, pages 1–9, 2005.
- [KAV14] Adam Kortylewski, Thomas Albrecht, and Thomas Vetter. Unsupervised footwear impression analysis and retrieval from crime scene data. In *Asian conference on computer vision*, pages 644–658. Springer, 2014.

- [KG11] Tim Kazik and Ali Haydar Göktoğan. Visual odometry based on the fourier-mellin transform for a rover using a monocular ground-facing camera. In *2011 IEEE International Conference on Mechatronics*, pages 469–474. IEEE, 2011.
- [Khe11] Fakhry M Khellah. Texture classification using dominant neighborhood structure. *IEEE Transactions on Image Processing*, 20(11):3270–3279, 2011.
- [KR12] Juho Kannala and Esa Rahtu. Bsif: Binarized statistical image features. In *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*, pages 1363–1366. IEEE, 2012.
- [KS17] Harshitha Reddy Katireddy and Sreemanth Sidda. A novel shoeprint enhancement method for forensic evidence using sparse representation method., 2017.
- [KSRF17] Bailey Kong, J Supancic, Deva Ramanan, and Charless Fowlkes. Cross-domain forensic shoeprint matching. In *British Machine Vision Conference (BMVC)*, pages 1–5, 2017.
- [KSRF19] Bailey Kong, James Supancic, Deva Ramanan, and Charless C Fowlkes. Cross-domain image matching with deep feature maps. *International Journal of Computer Vision*, pages 1–13, 2019.
- [KV16] Adam Kortylewski and Thomas Vetter. Probabilistic compositional active basis models for robust pattern recognition. In *BMVC*, 2016.
- [KWW⁺19] Adam Kortylewski, Aleksander Wieczorek, Mario Wieser, Clemens Blumer, Sonali Parbhoo, Andreas Morel-Forster, Volker Roth, and Thomas Vetter. Greedy structure learning of hierarchical compositional models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11612–11621, 2019.
- [KYZ14] Xiangbin Kong, Chunyu Yang, and Fengde Zheng. A novel method for shoeprint recognition in crime scenes. In *Chinese Conference on Biometric Recognition*, pages 498–505. Springer, 2014.
- [L⁺99] David G Lowe et al. Object recognition from local scale-invariant features. In *iccv*, volume 99, pages 1150–1157, 1999.
- [LBB⁺98] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [LBH15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

- [LFLY12] Chaorong Li, Bo Fu, Jianping Li, and Xingchun Yang. Texture-based fingerprint recognition combining directional filter banks and wavelet. *International Journal of Pattern Recognition and Artificial Intelligence*, 26(04):1256012, 2012.
- [LLGF14] Chaorong Li, Jianping Li, Dapeng Gao, and Bo Fu. Rapid-transform based rotation invariant descriptor for texture classification under non-ideal conditions. *Pattern Recognition*, 47(1):313–325, 2014.
- [Low04] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
- [LW15] Yan Li and Hong-Xing Wang. A secondary positioning algorithm of the shoe print. In *2015 International Conference on Network and Information Systems for Computers*, pages 560–563. IEEE, 2015.
- [LWS14] Xiangyang Li, Minhua Wu, and Zhiping Shi. The retrieval of shoepoint images based on the integral histogram of the gabor transform domain. In *International Conference on Intelligent Information Processing*, pages 249–258. Springer, 2014.
- [LZ96] Andrew F Laine and Xuli Zong. Multiscale suboctave wavelet transform for denoising and enhancement. In *Wavelet Applications in Signal and Image Processing IV*, volume 2825, pages 238–249. International Society for Optics and Photonics, 1996.
- [MA10] Raman Maini and Himanshu Aggarwal. A comprehensive review of image enhancement techniques. *arXiv preprint arXiv:1003.4053*, 2010.
- [MII17] Yogita Mistry, DT Ingole, and MD Ingole. Content based image retrieval using hybrid features and various distance metric. *Journal of Electrical Systems and Information Technology*, 2017.
- [ML09] Marius Muja and David G Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. *VISAPP (1)*, 2(331-340):2, 2009.
- [NBC⁺09] Omar Nibouche, Ahmed Bouridane, Danny Crookes, Mourad Gueham, and Moussadek Laadjel. Rotation invariant matching of partial shoemarks. In *2009 13th International Machine Vision and Image Processing Conference*, pages 94–98. IEEE, 2009.
- [Nib85] Wayne Niblack. *An introduction to digital image processing*. Strandberg Publishing Company, 1985.
- [Nil13] Mikael Nilsson. Smqt-based tone mapping operators for high dynamic range images. In *VISAPP (1)*, pages 61–68, 2013.

- [OMP⁺02] Timo Ojala, Topi Maenpaa, Matti Pietikainen, Jaakko Viertola, Juha Kyllonen, and Sami Huovinen. Outex-new framework for empirical evaluation of texture analysis algorithms. In *Object recognition supported by user interaction for service robots*, volume 1, pages 701–706. IEEE, 2002.
- [Ots79] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.
- [PK09] Pradeep M Patil and Jayant V Kulkarni. Rotation and intensity invariant shoeprint matching using gabor transform with application to forensic science. *Pattern Recognition*, 42(7):1308–1317, 2009.
- [PK12] CJ Prabhakar and PU Kumar. Lbp-surf descriptor with color invariant and texture based features for underwater images. In *Proceedings of the eighth Indian conference on computer vision, graphics and image processing*, page 23. ACM, 2012.
- [QCAC02] Roberto Quevedo, López-G Carlos, José M Aguilera, and Laura Cadoche. Description of food surfaces and microstructural changes using fractal image texture analysis. *Journal of food engineering*, 53(4):361–371, 2002.
- [RAMM⁺18] Imad Rida, Somaya Al-Maadeed, Arif Mahmood, Ahmed Bouridane, and Sambit Bakshi. Palmprint identification using an ensemble of sparse representations. *IEEE Access*, 6:3241–3248, 2018.
- [RBCP19] Imad Rida, Sambit Bakshi, Xiaojun Chang, and Hugo Proenca. Forensic shoe-print identification: A brief survey. *arXiv preprint arXiv:1901.01431*, 2019.
- [RLCL18] Xutong Ren, Mading Li, Wen-Huang Cheng, and Jiaying Liu. Joint enhancement and denoising method via sequential decomposition. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2018.
- [RLL⁺17] Nicole Richetelli, Mackenzie C Lee, Carleen A Lasky, Madison E Gump, and Jacqueline A Speir. Classification of footwear outsole patterns using fourier transform and local interest points. *Forensic science international*, 275:102–109, 2017.
- [RNBS17] Nicole Richetelli, Madonna Nobel, William J Bodziak, and Jacqueline A Speir. Quantitative assessment of similarity between randomly acquired characteristics on high quality exemplars and crime scene impressions via analysis of feature size and shape. *Forensic science international*, 270:211–222, 2017.
- [Sax19] Lalit Prakash Saxena. Niblack’s binarization method and its modifications to real-time applications: a review. *Artificial Intelligence Review*, 51(4):673–705, 2019.

- [SD86] Yunlong Sheng and Jacques Duvernoy. Circular-fourier–radial-mellin transform descriptors for pattern recognition. *JOSA A*, 3(6):885–888, 1986.
- [SK18] Dilbag Singh and Vijay Kumar. Dehazing of outdoor images using notch based integral guided filter. *Multimedia Tools and Applications*, 77(20):27363–27386, 2018.
- [SLSL93] Jonathan Schaeffer, Paul Lu, Duane Szafron, and Robert Lake. A re-examination of brute-force search. In *Proceedings of the AAAI Fall Symposium on Games: Planning and Learning*, pages 51–58, 1993.
- [SPK12] Jignesh N Sarvaiya, Suprava Patnaik, and Kajal Kothari. Image registration using log polar transform and phase correlation to recover higher scale. *Journal of pattern recognition research*, 7(1):90–105, 2012.
- [SPV16] Katta Sugamya, Suresh Pabboju, and A VinayaBabu. Image enhancement using singular value decomposition. In *2016 International Conference on Research Advances in Integrated Navigation Systems (RAINS)*, pages 1–5. IEEE, 2016.
- [SRF⁺16] Jacqueline A Speir, Nicole Richetelli, Michael Fagert, Michael Hite, and William J Bodziak. Quantifying randomly acquired characteristics on outsoles in terms of shape and position. *Forensic science international*, 266:399–411, 2016.
- [SWT⁺18] Yaron Shor, Sarena Wiesner, Tsadok Tsach, Ron Gurel, and Yoram Yekutieli. Inherent variation in multiple shoe-sole test impressions. *Forensic science international*, 285:189–203, 2018.
- [SZG11] Hafizan Mat Som, Jasni Mohamad Zain, and Amzari Jihadi Ghazali. Application of threshold techniques for readability improvement of jawi historical manuscript images. *arXiv preprint arXiv:1103.5621*, 2011.
- [TM13] Hossein Talebi and Peyman Milanfar. Global image denoising. *IEEE Transactions on Image Processing*, 23(2):755–768, 2013.
- [TSKC10] Yi Tang, Sargur N Srihari, Harish Kasiviswanathan, and Jason J Corso. Footwear print retrieval system for real crime scene marks. In *International workshop on computational forensics*, pages 88–100. Springer, 2010.
- [VRDJ95] Guido Van Rossum and Fred L Drake Jr. *Python tutorial*. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands, 1995.
- [VVVL16] Hoang Thien Van, Giang Van Vu, and Thai Hoang Le. Fingerprint enhancement for direct grayscale minutiae extraction by combining mfrat and gabor filters. In *Proceedings of the Seventh Symposium on Information and Communication Technology*, pages 360–367. ACM, 2016.

- [WBLL17] Kai Wang, Charles-Edmond Bichot, Yan Li, and Bailin Li. Local binary circumferential and radial derivative pattern for texture classification. *Pattern Recognition*, 67:213–229, 2017.
- [WBZL13] Kai Wang, Charles-Edmond Bichot, Chao Zhu, and Bailin Li. Pixel to patch sampling structure and local neighboring intensity relationship patterns for texture classification. *IEEE Signal Processing Letters*, 20(9):853–856, 2013.
- [Win] The wiener filter. http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/VELDHUIZEN/node15.html. Accessed: 2020-01-10.
- [WLWL14] Jing-Wein Wang, Ngoc Tuyen Le, Chou-Chen Wang, and Jiann-Shu Lee. Enhanced ridge structure for improving fingerprint image quality based on a wavelet domain. *IEEE Signal Processing Letters*, 22(4):390–394, 2014.
- [WSYZ14] Xinnian Wang, Huihui Sun, Qing Yu, and Chi Zhang. Automatic shoeprint retrieval algorithm for real crime scenes. In *Asian Conference on Computer Vision*, pages 399–413. Springer, 2014.
- [WWNZ19] Yanjun Wu, Xinnian Wang, Namusisi Linda Nankabirwa, and Tao Zhang. Losgsr: Learned opinion score guided shoeprint retrieval. *IEEE Access*, 7:55073–55089, 2019.
- [WWZ19] Yanjun Wu, Xinnian Wang, and Tao Zhang. Crime scene shoeprint retrieval using hybrid features and neighboring images. *Information*, 10(2):45, 2019.
- [WZWS17] Xinnian Wang, Chi Zhang, Yanjun Wu, and Yingying Shu. A manifold ranking based method using hybrid features for crime scene shoeprint retrieval. *Multimedia Tools and Applications*, 76(20):21629–21649, 2017.
- [XJF09] Yong Xu, Hui Ji, and Cornelia Fermüller. Viewpoint invariant texture description using fractal analysis. *International Journal of Computer Vision*, 83(1):85–100, 2009.
- [XWYH16] Xiaojun Xu, Youren Wang, Guoshi Yang, and Yanli Hu. Image enhancement method based on fractional wavelet transform. In *2016 IEEE International Conference on Signal and Image Processing (ICSIP)*, pages 194–197. IEEE, 2016.
- [XZZ⁺15] Jun Xu, Lei Zhang, Wangmeng Zuo, David Zhang, and Xiangchu Feng. Patch group based nonlocal self-similarity prior learning for image denoising. In *Proceedings of the IEEE international conference on computer vision*, pages 244–252, 2015.
- [YYXK15] Huang Yi, Peicong Yu, Xingpeng Xu, and Adams Wai Kin Kong. The impact of tattoo segmentation on the performance of tattoo matching.

- In *2015 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*, pages 43–46. IEEE, 2015.
- [YZL⁺16] Zhijun Yao, Quan Zhou, Zhongyuan Lai, Zhiming Ren, and Liming Liu. Image enhancement based on bi-histogram equalization with non-parametric modified technology. In *2016 IEEE 22nd International Conference on Parallel and Distributed Systems (ICPADS)*, pages 1211–1215. IEEE, 2016.
 - [ZA05] Lin Zhang and Nigel Allinson. Automatic shoepoint retrieval system for use in forensic investigations. In *UK workshop on computational intelligence*, volume 99, pages 137–142, 2005.
 - [ZL11] WL Zeng and XB Lu. Region-based non-local means algorithm for noise removal. *Electronics letters*, 47(20):1125–1127, 2011.
 - [ZLM⁺17] Zhong Zhang, Shuang Liu, Xing Mei, Baihua Xiao, and Liang Zheng. Learning completed discriminative local features for texture classification. *Pattern Recognition*, 67:263–275, 2017.
 - [ZSL⁺11] Ru Zhou, SangWoo Sin, Dongju Li, Tsuyoshi Isshiki, and Hiroaki Kunieda. Adaptive sift-based algorithm for specific fingerprint verification. In *2011 International Conference on Hand-Based Biometrics*, pages 1–6. IEEE, 2011.
 - [ZSP⁺16] Liang Zhang, Peiyi Shen, Xilu Peng, Guangming Zhu, Juan Song, Wei Wei, and Houbing Song. Simultaneous enhancement and noise reduction of a single low-light image. *IET Image Processing*, 10(11):840–847, 2016.
 - [ZZSZ13] Wangmeng Zuo, Lei Zhang, Chunwei Song, and David Zhang. Texture enhanced image denoising via gradient histogram preservation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1203–1210, 2013.

Appendix Dataset

This section contains the images selected for testing. The samples were chosen from FID-300, and are representative for the database. No restriction were made while selecting the images, low quality images are also part of the experimental setting. In the first column, the original images are shown, in the second and third columns the results using 20 classes for classification of the gradient difference image is visible, whereas in the fourth and fifth column the output with 25 classes is displayed. The results using the mean noise model are in columns two and four, the output using the centroids are in columns three and five.

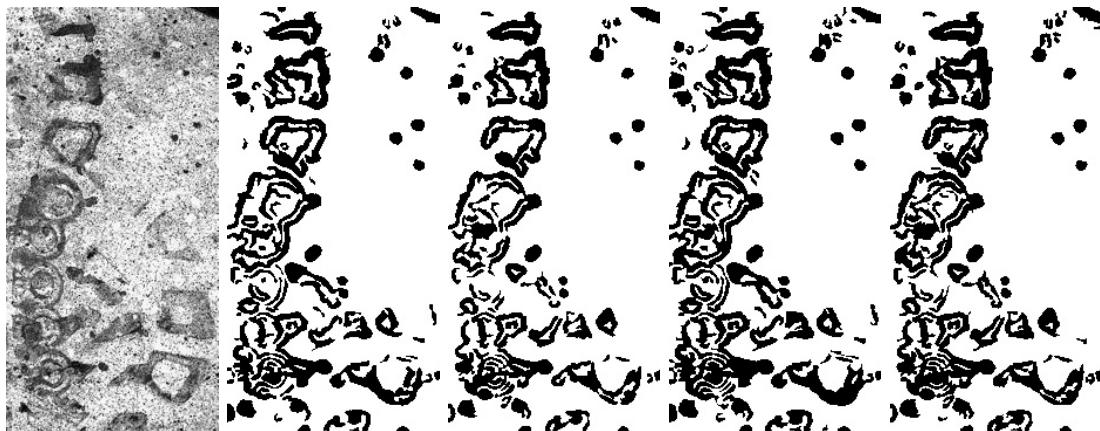


Figure 1

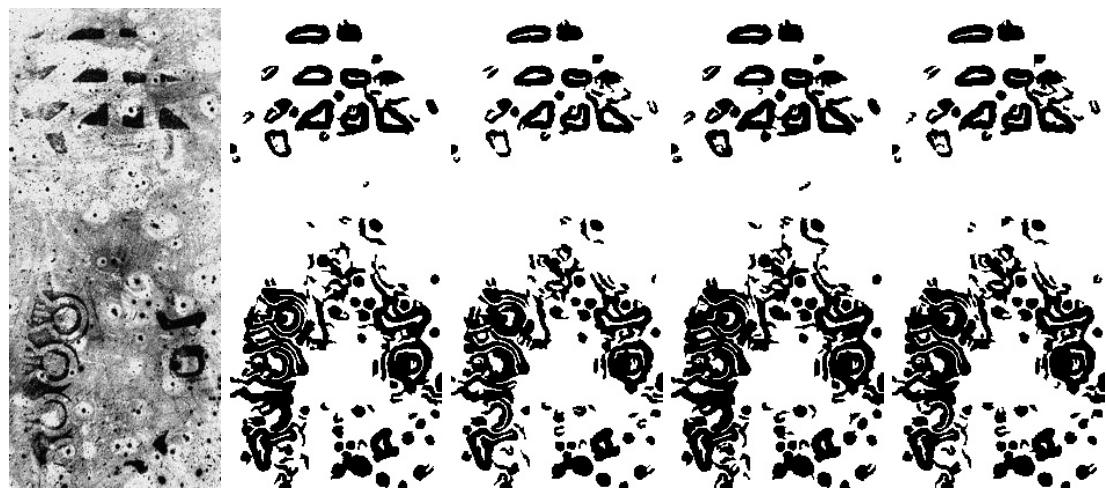


Figure 2

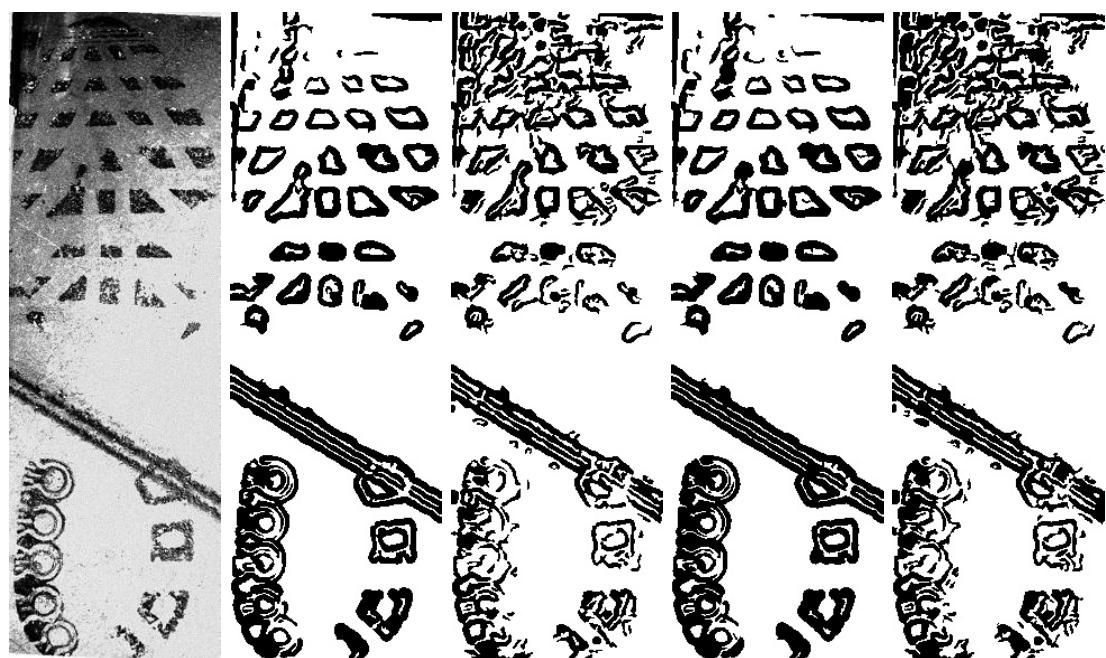


Figure 3



Figure 4



Figure 5



Figure 6

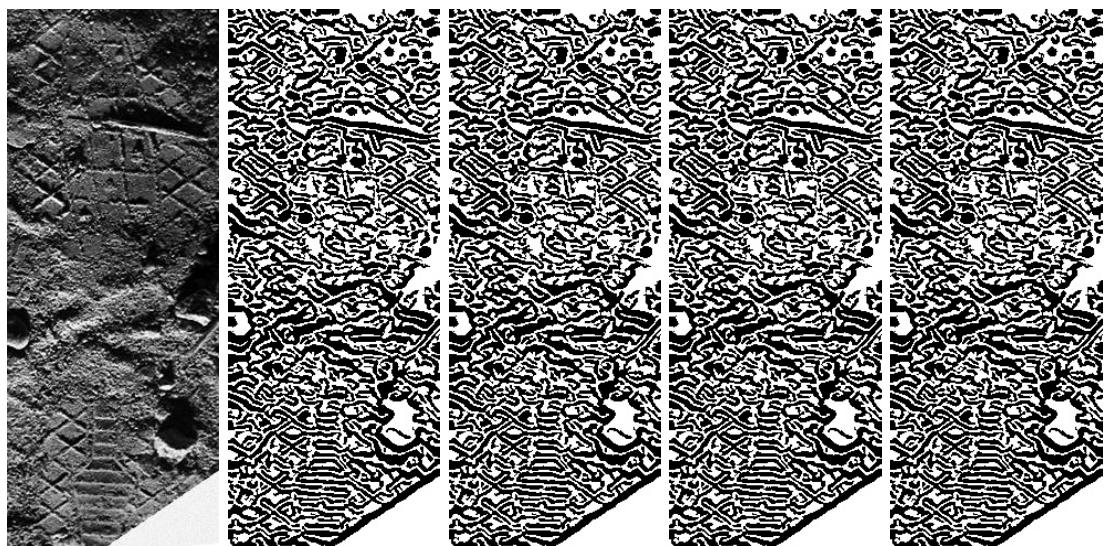


Figure 7



Figure 8

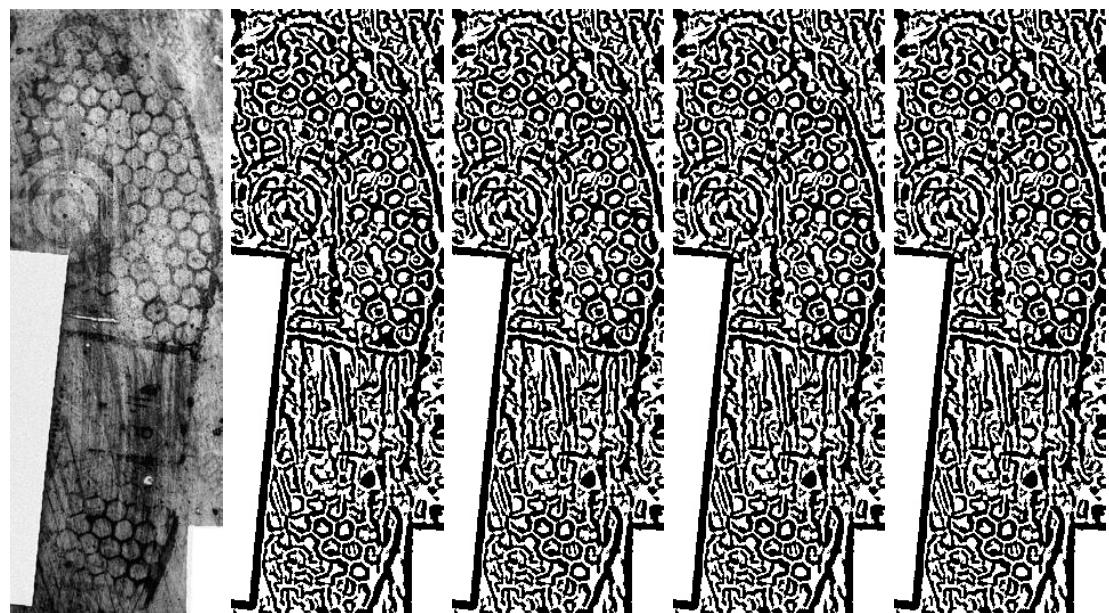


Figure 9

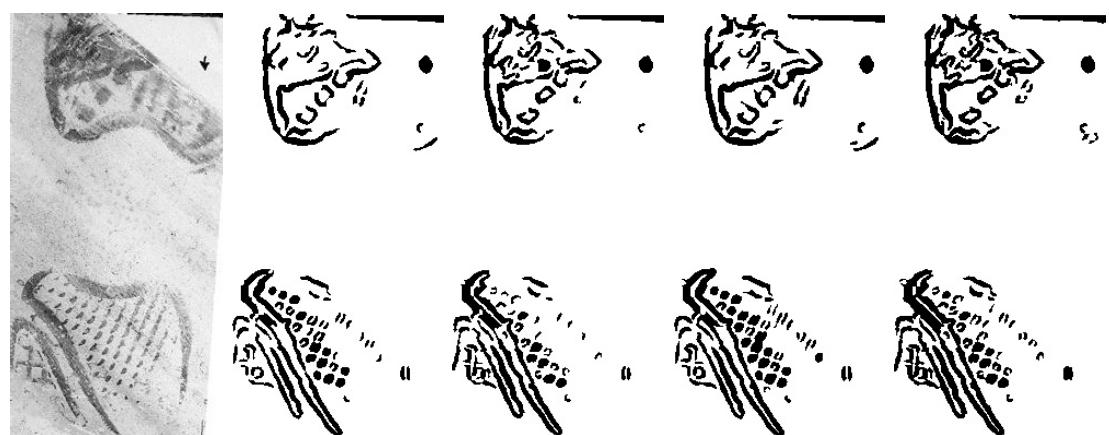


Figure 10

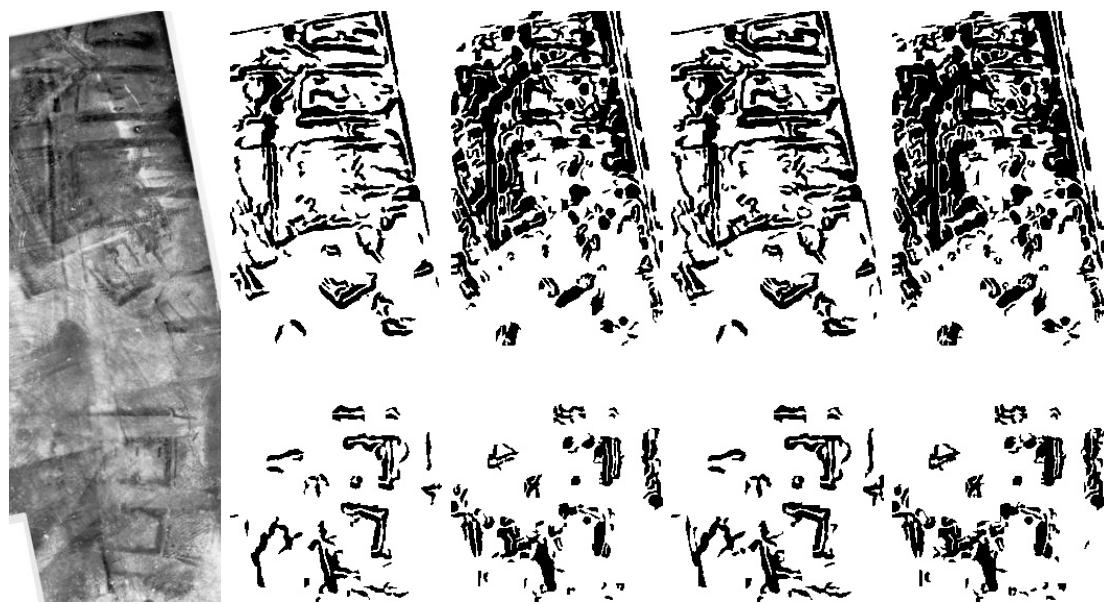


Figure 11

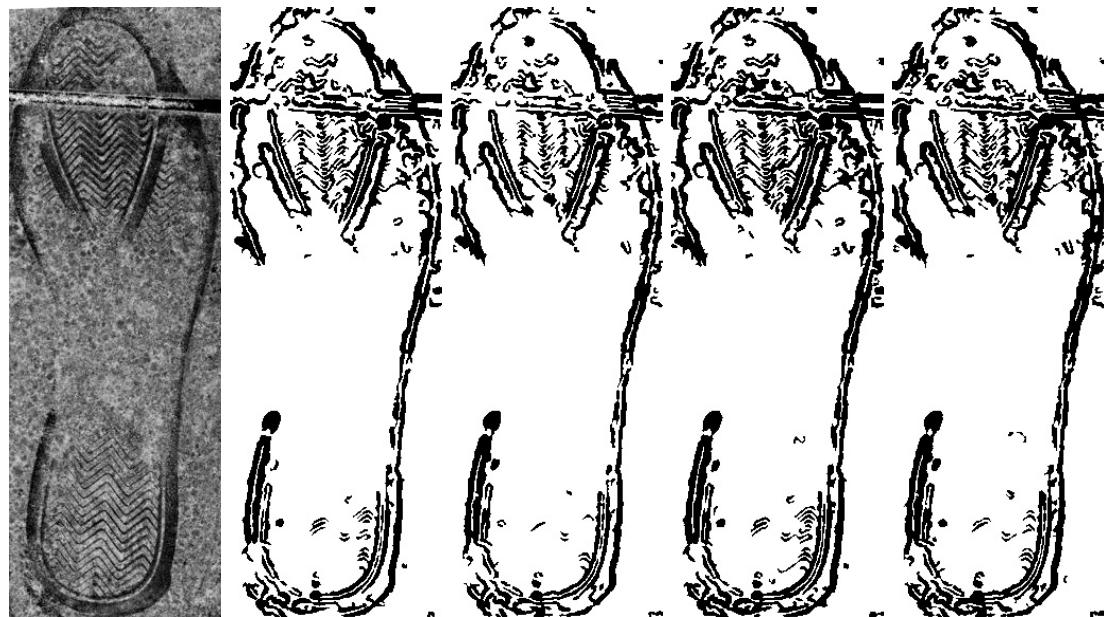


Figure 12

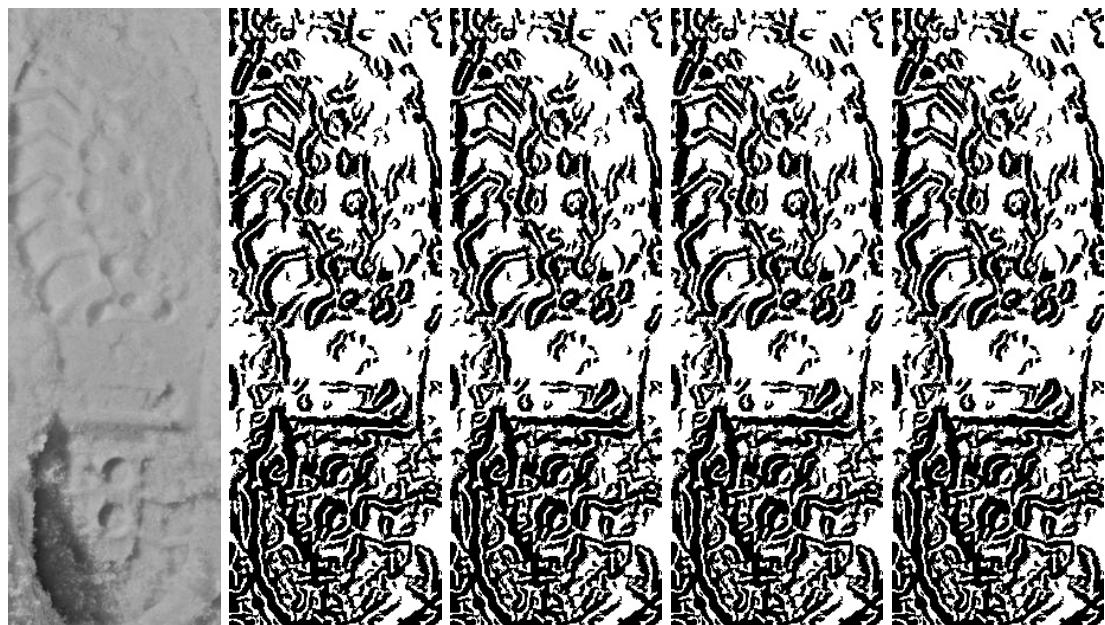


Figure 13



Figure 14



Figure 15

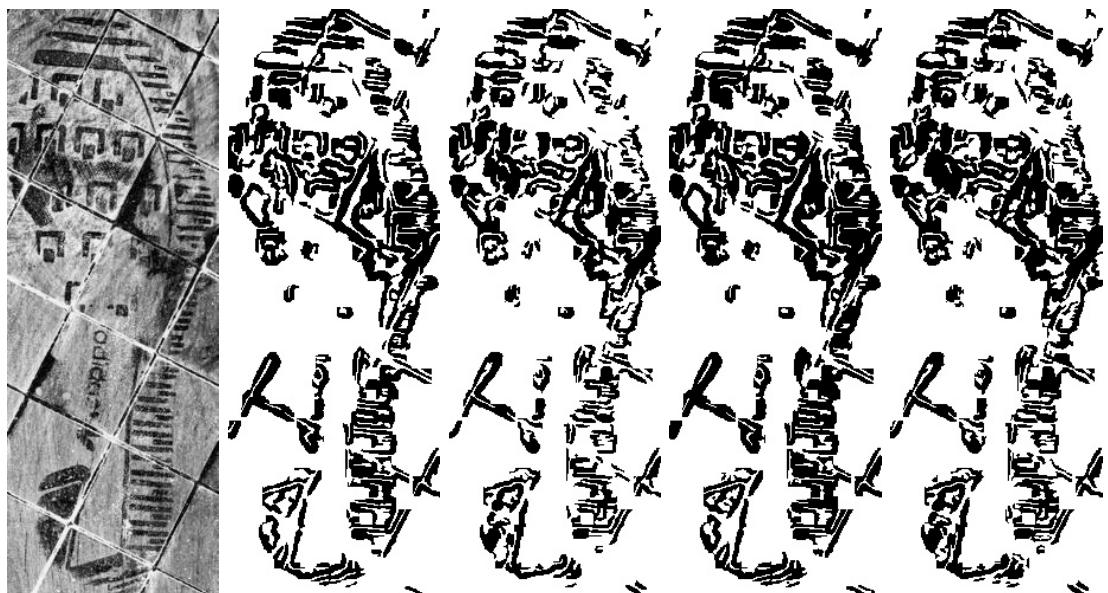


Figure 16



Figure 17



Figure 18



Figure 19



Figure 20

Appendix Results

This section contains the quantitative results of the Shoepoint Enhancement Algorithm with Non-Local Means. The testing dataset is presented in 5. The accumulated results are presented and discussed in Chapter 4 4. The four tables, see 2, 3 and 4, 5, present the matching scores of the different settings separately and are given to enlighten the reproducibility of the algorithm. Furthermore, to create a baseline, the original samples were matched to the reference images 1.

	F-M		SIFT		SURF	
	BF	FLANN	BF	FLANN	BF	FLANN
1	12	13	9	4	1	1
2	4	5	1	3	2	2
3	5	7	2	1	4	4
4	5	6	1	1	5	2
5	6	1	1	1	5	4
6	6	6	4	6	6	6
8	3	4	4	3	8	8
9	2	1	7	4	7	7
10	9	10	12	9	13	12
11	2	1	12	13	11	11
16	6	5	1	1	9	6
17	1	1	6	5	1	1
18	13	11	12	7	12	12
20	8	11	6	4	7	3
Score	5.86	5.86	5.57	4.43	6.5	6.65

Table 1: Baseline results matching the original images with the unprocessed Ground Truth images

	F-M		SIFT		SURF	
	BF	FLANN	BF	FLANN	BF	FLANN
1	1	1	2	6	4	5
2	1	1	3	6	4	4
3	1	2	1	1	3	4
4	1	1	4	8	1	2
5	4	5	1	1	4	6
6	7	5	5	5	8	8
7	13	13	3	1	1	1
8	3	3	13	12	6	4
9	7	5	3	4	1	3
10	5	3	2	4	3	5
11	2	3	3	5	10	12
16	2	2	6	5	5	4
17	4	4	8	11	1	1
18	11	12	4	5	2	1
20	5	6	4	5	3	3
Score	3.86	3.79	4.21	5.57	3.93	4.43

Table 2: Results using 25 classes and the average noise model

	F-M		SIFT		SURF	
	BF	FLANN	BF	FLANN	BF	FLANN
1	2	3	1	2	3	4
2	1	3	3	4	6	8
3	1	2	8	6	2	5
4	1	1	2	5	3	5
5	1	1	1	1	4	5
6	5	4	6	9	7	7
7	13	13	3	1	1	1
8	3	4	12	13	8	9
9	7	5	3	4	1	3
10	6	5	1	1	4	4
11	3	2	9	6	11	11
16	4	4	2	8	6	7
17	1	1	4	8	1	2
18	9	9	5	1	4	1
20	2	1	3	4	2	4
Score	3.29	3.21	4.29	5.14	4.43	5.36

Table 3: Results using 25 classes and the centroids noise model

	F-M		SIFT		SURF	
	BF	FLANN	BF	FLANN	BF	FLANN
1	1	1	2	4	3	4
2	3	3	6	3	2	3
3	1	2	1	1	4	5
4	2	5	6	7	1	1
5	2	3	1	1	2	4
6	4	8	10	12	6	8
8	5	5	11	12	3	4
9	7	5	3	4	1	3
10	7	4	1	2	1	4
11	2	1	6	6	3	2
16	10	9	3	7	4	5
17	2	2	11	6	1	1
18	12	12	4	3	9	5
20	4	4	4	5	2	2
Score	4.43	4.57	4.93	5.21	3.00	3.64

Table 4: Results using 20 classes and the average noise model

	F-M		SIFT		SURF	
	BF	FLANN	BF	FLANN	BF	FLANN
1	1	1	1	3	1	2
2	1	1	2	2	5	8
3	1	2	4	8	5	5
4	1	1	3	8	2	2
5	3	3	1	1	4	4
6	2	2	8	12	5	5
8	6	5	9	12	7	6
9	7	5	3	4	1	3
10	4	2	8	5	1	1
11	6	2	6	8	11	10
16	1	1	5	6	1	2
17	2	3	3	4	1	6
18	9	11	7	4	3	1
20	2	4	2	3	5	8
Score	3.29	3.07	4.43	5.71	3.71	4.50

Table 5: Results using 20 classes and the centroids noise model