

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: Proj.-Ass. 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: Proj.-Ass. 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

Ihr Text hier.

Acknowledgements

Enter your text here.

Kurzfassung

Ihr Text hier.

Abstract

Shoeprint images are one of the most commonly secured evidences on crime scenes. Even though automatic shoeprint processing is a highly researched topic, the final identification is usually done by human forensic experts.

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

The motivation of this work is to give an overview about the potential of the published technologies and to define a universal approach which is able to filter and enhance the shoeprint data despite the presence of noise and the possible low image quality. Based on the experiences acquired while investigating a new 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 given by the user, and of its multi-sized neighborhood. At the same time a model is built about the appearance of noise, to eliminate that structure from the foreground as well. Additionally a gradient based line detector is applied and the edge structures of the shoeprint are clustered to distinguish between pattern and noise edges. The experimental results show that the processed images are clearer, the pattern is sharper whereas the noise is completely eliminated in the background and suppressed in the foreground. Furthermore, according to the tests conducted with three basic image descriptor features, the enhanced shoeprints have higher matching rate to Ground Truth samples than the original images.

Contents

Kurzfassung	xi
Abstract	xiii
Contents	xv
1 Introduction	1
1.1 Problem Definition	2
1.2 Challenges	2
1.3 Contribution	3
1.4 Structure of the Work	4
2 Related Work	7
2.1 Image Enhancement	7
2.2 Noise Removal	11
2.3 Image Description	12
3 Pattern Detection	17
3.1 Methodology	17
3.2 Implementation	21
3.3 Evaluation	22
4 Fully-Automated Noise Elimination	27
4.1 Methodology	27
4.2 Implementation	29
4.3 Evaluation	31
5 Semi-Automated Shoepoint Enhancement	37
5.1 Methodology	37
5.2 Implementation	41
5.3 Evaluation	42
6 Results and Discussion	49

7 Future Work	57
8 Conclusion	61
Bibliography	63
Appendix A	73

1

CHAPTER

Introduction

Shoeprints found on crime scenes are important hints or evidence in a criminal investigation [KYZ14]. Even though on one third [Ale96] of crime scenes usable shoe patterns can be secured, there is no fully automatized algorithm available yet, which is able to identify and match those prints with the original shoe sole, thus human power is needed [WSYZ14] to recognize and analyze the found patterns. The work of forensic experts is not only time consuming and expensive, there is no guarantee about the objectivness of the final outcome[GBCN08], furthermore the stages of the human matching process are unclear and not necessarily reproducible.

There is an excessive amount of research already done [RBCP19] in order to help or replace the work of forensic experts. However, there is no algorithm published yet, which is reliably usable in varying conditions and sample quality. One reason for that are the already mentioned versatile conditions, the features and properties of the pattern on the shoe, like age, material, etc., the characteristics of the ground where the shoepoint is left and environmental conditions, like for example weather, highly influence the overall quality of the acquired sample. Those high amount of factors result in changing appearance of the prints of the same shoe causing high intra class variance while clustering. Additionally there is a lack of universal, wide ranged database [RBCP19] which correctly depicts the common scenarios occurring on real-life crime scenes.

In 2014 a new dataset, called FID-300 [KAV14], was released which aims to solve the database problem described above. It contains over 1000 reference shoepoint patterns acquired in a laboratory. Moreover the database introduces 300 new shoepoint samples collected by the police providing an insight on images forensic experts are working with on the daily basis.

1.1 Problem Definition

There are two main stages of automatic shoeprint identification, filtering, where the shoeprint pattern is separated from background and enhanced as well, and matching where the corresponding shoe is determined. Instead of automatizing the entire shoeprint recognition pipeline this work focuses on the possible ways of increasing the sample quality. Because of the mentioned absence of general, appropriate database it is difficult to compare the already available methods. Furthermore it is also challenging to estimate which one is applicable in a real-life scenario. In this thesis three prototypical enhancing techniques are developed and tested in order to create an overview of the published technologies and their performance and to find a method which handles real crime scene samples universally.

For evaluation and testing the FID-300 database is used. The dataset contains both in a laboratory acquired, synthetic, as well as on a crime secured, real, shoeprint patterns. Additionally the Ground Truth about every real sample and its synthetic pair is also available. The goal of this work is to define an image processing pipeline which correctly identifies and enhances the shoe patterns and eliminates or suppresses the noise on the pattern samples regardless of the quality of the image. A secondary objective is to gain an overview about the algorithms already published, and make an estimation which methods are applicable in real-life scenarios based on their performance on the FID-300 database.

1.2 Challenges

There are two main obstacles in the topic of shoeprint enhancement and in automatic shoeprint matching in general, the versatile image quality and appearance as well as the lack of universal and wide database. There are approaches available which build models for given structures of the shoeprint [TSKC10], [AK17], but no detailed, uniform representation for the entire shoeprint is available yet to represent the varying shoe pattern appearance. Moreover there is a high inference of noise from multiple sources. The produced shoeprint and its background highly depends on the properties of the ground where it was found, the roughness and unevenness of a given type of surface distorts the original pattern and the clutter on the ground appears as noise on the final sample. Furthermore other objects on the surface, above or behind the left shoeprint additionally cover or distort the original pattern, or prevent to secure the complete area of the original shoe sole. Besides that the pattern on the original shoe can also be distorted or modified compared to the new version. Distortions caused by usage are valuable information about the owner, on the other hand they make it more difficult to match the pattern with their unused pairs. Additional objects between the structures of the shoeprint also alter the original appearance. Lastly, there are multiple shoeprint securing methods producing different results for the same print [KS17]. The shoeprint lifting technique used depends on the properties of the ground. Those two factors, the

securing method and the floor type, also determine if the positive or the negative, the pattern or the space between them, image is captured.

The non-existing universal database causes that the published methods are difficult to compare based on their results since they are using different testing images. Moreover, the used dataset is not necessarily public [KS17], [DCC09] making it impossible to reproduce the result in those cases. Additionally the handcrafted databases can be biased, and allow such restrictions and modifications which do not correlate with real-life scenarios [RBCP19]. The used samples are either synthetically generated and computationally distorted [DCFR05], [GBCN08] or exclude low quality and noisy images [DCC09], [TSKC10]. Because of that it is difficult to compare their performance and to estimate which one of the approaches are applicable on the FID-300 database. Furthermore, it is challenging to plan a new algorithm based on the published results because their lack of a uniform baseline.

1.3 Contribution

In this thesis the possible ways of enhancing a shoeprint images are discussed. Because of the known issues on database multiple approaches are implemented, discussed and evaluated. The two ways to increase the quality of a given shoeprint sample is to enhance the pattern regardless of the noise and to suppress or eliminate the noise without losing any of pattern information. Along fully-automated methods semi-automated algorithms are also considered. Three different approaches are introduced and examined for their performance on real-life image samples.

Finally a new semi-automated framework is given which is evaluated on the FID-300 database. In the first step user input about the noise is required. The input is separated into tiles, and the subparts are compared based on the Fourier-Mellin features to the region of the user input. In that way the background is separated from the foreground and a model about the appearance of the background is calculated. Since noise appears on the pattern as well, the distorted parts are corrected based on the calculated noise model. After that gradient based line detection is performed and the results are separated into clusters where pattern and noise classes are defined and candidates of the latter are eliminated. The final image is thresholded to create a binary image, where the shoeprint is clearly visible and recognizable whereas the clutter is suppressed on the pattern and eliminated on the background. Throughout the whole processing pipeline morphological operations and small structure elimination is 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 pattern. Experimental results show that the enhanced images are clearer, the background is successfully separated and the shoeprint pattern is less noisy than on the original images. Figure 1.1 shows an example sample from the FID-300 database 1.1a and the enhanced image 1.1b with the proposed algorithm. 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

synthetic Ground-Truth using three basic image features such as Fourier-Mellin, SIFT and SURF.

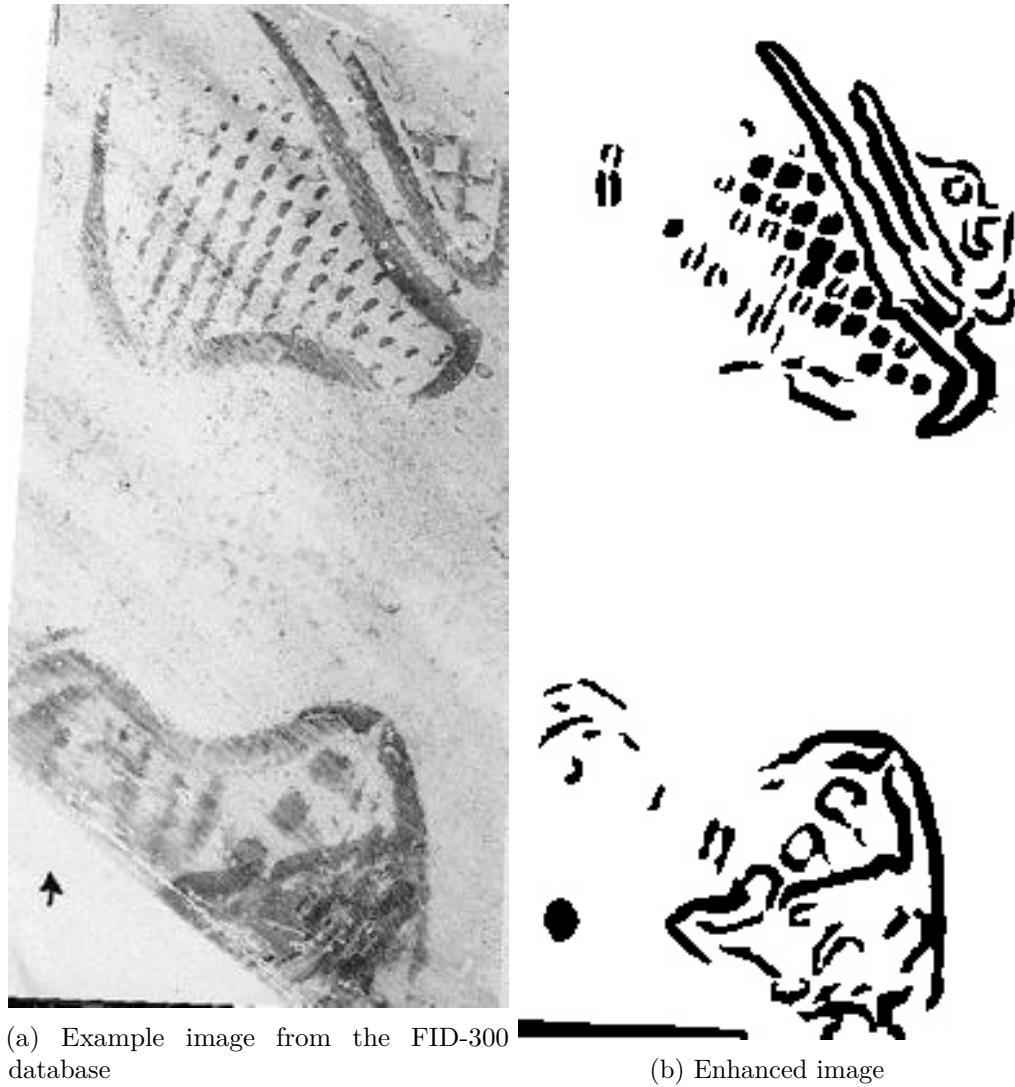


Figure 1.1: Result of the proposed algorithm

1.4 Structure of the Work

To gain an overview about the research already done the following section, Chapter 2, gives a review of the literature. Along papers published on the topic of shoepoint identification, matching and enhancement, research of similar domains is presented as well. Fields of fingerprint processing and tattoo identification 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, 4 and 5 the approaches for enhancement are given and discussed if they are applicable for real-life forensic images. Chapter 3 presents and evaluates an algorithm for detecting shoe pattern on varying conditions. Chapter 4 describes an automated noise suppression pipeline. In Chapter 5 the new algorithm for enhancing real-life crime scene shoeprint impressions is proposed. Details on implementation are revealed as well.

In Chapter 6 experimental results are shown and the proposed algorithm is evaluated. In Chapters 7 and 8 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 enhancement an overview about relevant research is made first. Along the evident literature of image enhancement and noise removal a related topic, discriminative image descriptors, is also considered, to gain the best insight possible and to define an approach which is optimized for the rest of the shoeprint identification pipeline. In this chapter the research on the domain of shoeprint identification is reviewed. Other than that publications from similar domains such as fingerprint and palmprint detection as well as tattoo identification are described. Research for fingerprint identification and tattoo recognition have been chosen for review because of their similar goal of edge structure and minimal image structure 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 four specific domains are discussed, these are shoe- and fingerprint identification, tattoo recognition and natural image enhancement.

Shoeprint Enhancement

There is an extensive research done in the field of enhancing shoeprint images. However, the problem definition and the use-case of the different publications varies strongly. Because of the absence of standard database, the discussed algorithms can be separated into two groups, techniques tested on synthetic samples and on real-life impressions.

2. RELATED WORK

Synthetic samples are scanned in images in a laboratorial environment for the purpose of building a dataset for shoeprint identification. The noise derives from scanning artifacts and computationally added distortions and modification. Furthermore many algorithms developed for real crime-scene data make restrictions about the input image and exclude noisy and poor quality images. 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. In the following discussion it is noticed repeatedly, which kind of dataset the proposed approach was tested on.

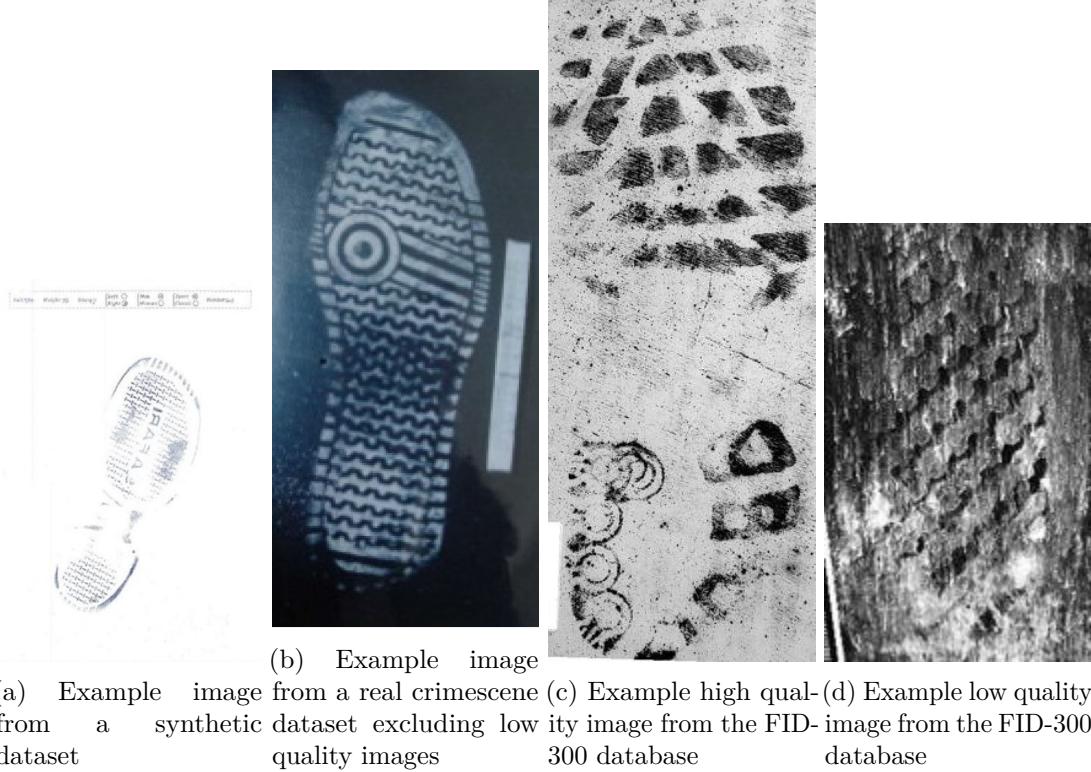


Figure 2.1: Example images of a synthetic [AK17], of a restricted [LWS14] and of the FID-300 [KAV14] dataset

Morphological Operations, Thresholding and Image Filtering are popular techniques for improving the quality of both kind, realistic and synthetic, of input data. Morphological Operations, especially Opening and Closing, is used in many cases [WSYZ14], [KYZ14], [LWS14], [TSKC10], [WWZ19], where Wang et al. [WSYZ14] uses a synthetic dataset, and other than Wu et al. [WWZ19] the forensic images are restricted to high quality data. Wang et al. [WSYZ14], Kong et al. [KYZ14] and Li et al. [LWS14] use the Morphological Operations to correct inconsistencies after thresholding. Similar to the previous approaches Wu et al. [WWZ19] applies the same pipeline on a real forensic dataset. Tang et al. [TSKC10] follow the same principle but instead of thresholding, after Canny edge detection is Opening and Closing used.

To create a binary image and eliminate noise various thresholding techniques are used. Otsu [WWZ19], [AH08], [AK17], [KYZ14] and adaptive thresholding [WSYZ14], [LWS14] are the two most popular algorithms. Algarni et al. [AH08] and Alizedah et al. [AK17] along with Wang et al. [WSYZ14] published their algorithms for synthetic datasets. Kong et al. [KYZ14] and Li et al. [LWS14] tested on restricted, whereas Wu et al. [WWZ19] developed their approach for full real forensic database. Wang et al. [WSYZ14] and Wu et al. [WWZ19] combine thresholding with a grid based approach to calculate exact thresholds for every subarea of the picture.

Another way to eliminate noise is image filtering. Alizadeh et al. [AK17] uses a simple Median filter on a synthetic dataset. Zhang et al. [ZA05] test on synthetic database as well, and they 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 an only step to enhance a real-life database. Figure 2.2 shows the output of the SMQT algorithm on an example image.

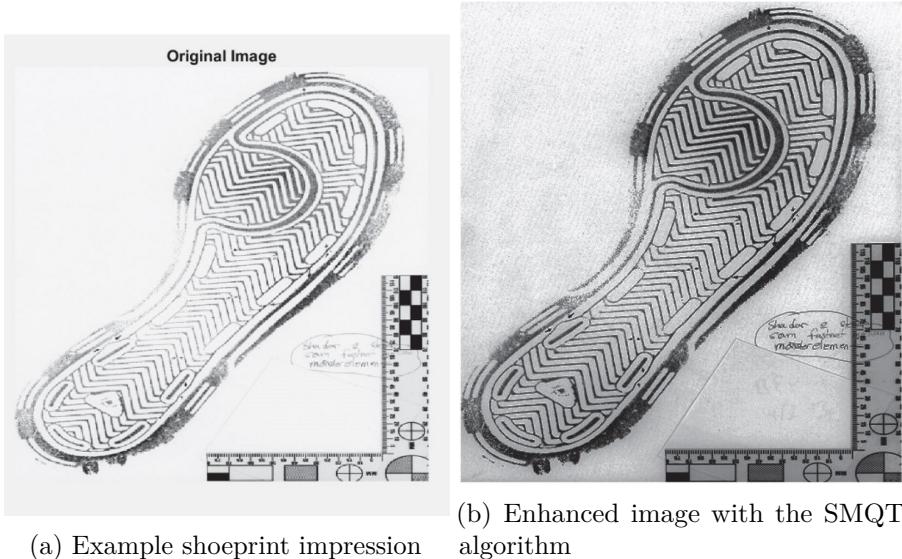


Figure 2.2: Example image about the enhancement feature of the SMQT algorithm [KS17]

Bandpass operators are also used for noise suppression. The images are converted to the frequency domain where high and low frequencies are eliminated. Gueham et al. [GBC07] and Richetelli et al. [RLL⁺17] utilize this method on a synthetic database. Li et al. [LWS14] work with a restricted real dataset, where only the lower frequencies are eliminated. Another frequency based approach was proposed 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.

Fingerprint Enhancement

Bandpass and general image filtering 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 previous thresholding of the input. Wang et al. [WLWL14] decompose the image into four subbands and process them separately, calculating the noise for every subband respectively. Li et al. [LFLY12] use Fourier transformation combined with Scale Invariant Feature Transform (SIFT) [L⁺99] to enhance the fingerprint images. With SIFT the interesting points in the Fourier domain are found and secured, while the image is filtered to suppress noise and eliminate other inconsistencies. Jahan et al. [JCI17] apply Fuzzy filtering followed by thinning. Fuzzy filter is a local method to preserve the edge information and fine line structures and suppress the noisy background of the input.

Tattoo Enhancement

For tattoo enhancement an algorithm from Han et al. [HJ13] was proposed which combines Gaussian filtering with Hysteresis thresholding. Hysteresis thresholding is a neighborhood-aware approach where a pixel is labelled when it is above a given low threshold and simultaneously connected to other pixels meeting a higher thresholding criteria. Acton et al. [AR08] propose to use Active Contour Model to find the boundaries of tattoo images and apply Opening and Closing as well to get rid of small inconsistencies.

Natural Image Enhancement

Along Signal, especially Bandpass, general Image Filtering and Thresholding, Histogram and Color Operations are also common for natural image enhancement. Maini et al. [MA10] published a review about natural image enhancing algorithms and defined two main groups of algorithms, Frequency and Spatial Domain Methods. First publications utilizing techniques from the first group are discussed, after that the usage of Spatial Domain Methods is reviewed.

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.

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.

Histogram Operations are combined apart from Bandpass filtering as Sugamya et al. [SPV16] do but also with Thresholding as proposed by Yao et al. [YZL⁺16]. Their approach separates the histogram of the input into two parts according to Otsu's method, after that the histogram of the generated subimages is equalized. Figure 2.3 shows the results of algorithm on an example image.

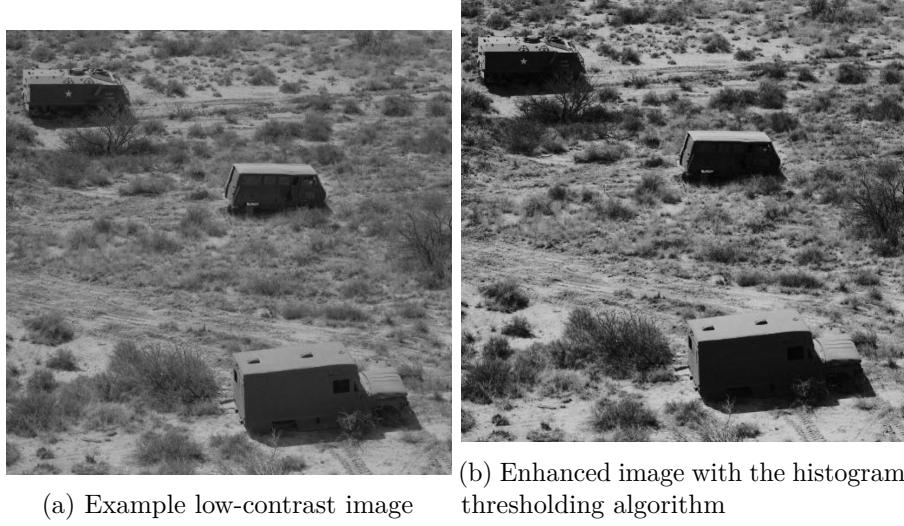


Figure 2.3: Example image [YZL⁺16] about the enhancement feature of the algorithm proposed by Yao et al. [YZL⁺16]

Color processing techniques are widespread in the topic of natural image enhancement. It is used for image dehazing for low contrast images, [SK18] and also for classical 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 processing the H channel is not modified to prevent color distortion artifacts. Although color processing is a well researched field with many promising solutions, no wider overview is given in this thesis. Even though there are shoeprint impression datasets with colored samples [KS17], FID-300 provides grayscale images thus no color processing approach can be used in this specific case.

2.2 Noise Removal

Noise Removal methods are based on estimating the original image, eliminating the deviating features of the data. One way is denoising through gradient histogram preservation [ZZSZ13]. The distribution of gradients is estimated on the original image, 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 the noisy regions are approximated. Huang et al.

[HKWL13] propose a self-learning algorithm which only considered the high frequency parts of the decomposed image. There were several approaches published, where the images are separated spatially instead of the frequency domain. [XZZ⁺15], [TM13], [CM11] and [GZZL15] are techniques based on the idea of non-local means, where the image is subsampled into tiles and the pixels are set to the mean of regions belonging to the same cluster. The main difference between the previous algorithms is how they classify the image subregions into different classes. Taleby et al. [TM13] uses an iterative shrinkage strategy. Chatterjee et al. [CM11] groups the geometrically similar regions and estimate the noise for every class separately with the Wiener filter, whereas Guo et al. [GZZL15] utilize Block Matching to determine the cluster memberships. Additionally the spatial location is also considered while calculating the mean value in a given class. The members are weighted according to the distance to the current region.

2.3 Image Description

Similar to the previous Image Enhancement section 2.1 this part is also subdivided into four domains offering solutions for the same problem in different areas. In this section the published image features are described which are proposed to be the most discriminative for their field. Similar topics are reviewed to gain insight about the most powerful descriptors and consider to port them into the domain of shoeprint enhancement. Shoeprint descriptors are described first, after that fingerprint features are discussed. Following that an overview about tattoo and natural texture descriptors is given at the end.

Shoeprint Descriptors

Shoeprint Description in varying difficulty and quality of the different datasets is a continuous issue in the research of Shoeprint Identification, therefore the properties of the database the given approach was tested on is noted repeatedly. Signal or frequency domain based image features are popular in both groups of algorithms using synthetic or real samples for testing. Gabor Transform is used in several applications tested on synthetic [PK09], on restricted [KYZ14], [LWS14] and on real forensic data [WWZ19] as well. Patil et al. [PK09] propose to use the Radon Transform to determine the dominant direction of structures on the print and process the aligned image with the Gabor Transform. Kong et al. [KYZ14] combined the Gabor features of the sample with Zernike features to describe the shapes in the pattern. 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. There are other publications along that where Fourier-Mellin Transform is proposed for feature description. Gueham et al. [GBCN08] use the classical Fourier-Mellin pipeline to compare samples from a synthetic dataset. As Wang et al. [WSYZ14] state, Fourier-Mellin transform allow multi resolution matching, so they apply it successfully on synthetic data. Richetelli et al. [RLL⁺17]

classify synthetic shoeprint impressions 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] in another paper suggest to use the basic Fourier transform before calculating the POC. Unlike the approach of Gueham et al. [GBC07], which were only tested on synthetic images, Kortylewsky et al. [KAV14] propose a Fourier Transformation based method for real forensic images. It is an unsupervised learning approach, where the periodic structures on 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 also proposed for a restricted dataset [DCC09]. It is a descriptor for random, broadband signals based on the frequency and power.

In high quality datasets shape descriptors are a popular choice for feature extraction. Algarni et al. [AH08] suggest to use Hu moments because of its robustness against noise, rotation and resolution. For restricted databases it is proposed to combine the feature descriptors, Kong et al. [KYZ14] incorporate Zernike and Gabor features whereas Tang et al. [TSKC10] define their own fundamental shapes based on common basic structures on a shoe sole through Hough Line Transform. The extracted features are then stored in an Attributed Relational Graph to represent the entire shoeprint image.

SIFT and Harris Detector are the two most popular point descriptors for matching shoeprint images. Nibouce et al. [NBC⁺09] propose to use a four-level Harris and combine with SIFT on a synthetic database. Almaadeed et al. [ABCN15] use the same combination and the Hessian Detector additionally for the same purpose on a restricted real-life dataset. Another publication [RLL⁺17] extract the SIFT features from the frequency domain after applying Fourier-Mellin Transformation and POC on the input image.

Kong et al. [KSRF17], [KSRF19] define a new descriptor for real database extracting the mid-level features from a Convolutional Neural Network. Wu et al. [WWNZ19] also use machine learning to calculate opinion scores for given matching pairs from real forensic data. 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 hierarchical composition for Active Basis Models for the same real database, FID-300, as used in this thesis, and also extended for natural image environments [KWW⁺19]. Sparse representation was also proposed [AK17] this method however was only tested on synthetic data and not on real forensic dataset.

Fingerprint Descriptors

Signal domain representations, because of their robustness against rotation, 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 get 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 directions and every subimage is decomposed into two frequencies via Wavelet Transform.

Unlike in the domain of shoeprint image representation, in case of fingerprint images several point feature descriptors were proposed. In [ZSL⁺11] SIFT features were fused with Hough Transform and Minutiae Information of the fingerprint. Chen et al. [CHZ13] also use the Hough Transform and extends with hierarchical voting score to get better 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 kind of edges on the image. Jahan et al. [JCI17] suggest to combine the Minutiae Information with Speeded Up Robust Features and compare them with a Neural Network to find the matching pairs.

For fingerprint description Local Binary Patterns (LBP) are also suitable. As mentioned previously Rida et al. [RAMM⁺18] published a combined feature vector which LBP is also a part of. Wang et al. [WBZL13] modify the usual LBP pipeline with Pixel to Patch sampling to increase the quality of descriptor without slowing down the calculation. In 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 based on gray-scale information are also considered.

Sparse representations are also popular for fingerprint description. Rida et al. [RAMM⁺18] stores the hybrid features in that way, whereas Shao et al. [SWY⁺13] represents predefined fingerprint patches, called dictionary atoms, in a sparse way.

Tattoo Descriptors

For tattoo description three main methods were introduced, signal domain features, point descriptors, especially SIFT, and shape features and Kim et al. [KPY⁺15] fuses all three of them. For local descriptor the Shape Context Features are used whereas for global descriptor multi-level SIFT and the Fourier Transform is utilized. Acton et al. [AR08] build an Active Contour Model which consists of Haar Wavelet, an HSV histogram and a Fourier Shape Descriptor.

Other than [KPY⁺15] there are multiple publications available, e.g. [DK13], which take advantage of SIFT features for tattoo image description. It is common to combine them with other shape descriptor to have a more detailed representation. Han et al. [HJ13] fuse Active Shape Models with SIFT, in [YYXK15] SURF is added and in [KLYD16] SIFT is extended with the Local Self Similarity measure. Duangphasuk et al. [DK13]

base their approach mainly on shape description and similar to [KPY⁺15] use Shape Context Features for tattoo representation.

Natural Texture Descriptors

For natural texture description signal domain features and LBPs are frequently used. 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 to take 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 to use multifractal spectrum to achieve robustness against viewpoint changes and non-rigid deformations additionally. Hayati et al. [HA18] and Ahonen et al. [AMHP09] follow the same principle by incorporating 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.

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]. It is common however, that LBP is used combined with other techniques to create a discriminative and detailed descriptor. In [HZPC14] based on the covariance matrix additional discriminative features are calculated. As mentioned previously Ahonen et al. [AMHP09] use the LBP features of the Fourier domain, similarly in [GZZ10a] LBP features are calculated twice, after the input is separated into two components, into signs and magnitudes, 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. 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 dominant orientation is stored additionally so that scale invariance is also granted. 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 to solve this problem by storing the radial and tangential information instead of 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 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, multi-level coding is used in different directions, and the differences between

2. RELATED WORK

neighbors are weighted additionally. Chahi et al. [CRT⁺18] define Local Ternary Patterns which store the directional patterns explicitly.

In [KR12] a new LBP similar descriptor is defined called Binary Statistical Image Feature (BSIF) which is also proposed to use by Crossier et al. [CG10]. In BSIF prelearned filters are used and the responses are stored in the feature, it can handle large intra-class variation when used for classification, however it varies greatly when the scale changes. For that reason Crossier et al. [CG10] suggest to calculate the features on multiple resolution to gain scale invariance as well.

Varied Local Edge Pattern Descriptors (VLEP) [YWZ16] are used to represent edge information. Every pixel of an edge is described by the angle and the magnitude of the gradient which stand for edge direction and strength respectively. Wang et al. [WZY⁺18] extend the feature to be scale invariant by combining two or more modified VLEPs, one calculated on a different scale and another calculated on different scale and resolution.

CHAPTER

3

Pattern Detection

In this and in the following two chapters three different approaches are introduced and discussed if they are usable for enhancement of real-life footwear impression. As mentioned earlier the testing data of already published algorithms varies greatly, thus it is difficult to make an estimation about their performance on the FID-300 [KAV14] dataset. For that reason three prototypical application are developed and evaluated in this thesis, and the one with most promising results is further elaborated. The goal of these three chapters is to give an overview about possible directions of development and to provide information about the effectiveness of the given techniques on a real-life forensic dataset.

There are two ways to increase the quality of a given input, find and enhance the information regardless of the properties and presence of noise or suppress any kind of noise preserving the information. In this chapter an experimental application for the first possibility is introduced in Chapter 4 an approach for noise suppression are discussed. In Chapter 5 a semi-automated enhancement algorithm is presented. The theoretical background and the structure is discussed in the first section, afterwards implementation details are reported followed by the evaluation of the results.

3.1 Methodology

To find the shoeprint pattern in any input regardless the noise a robust and discriminative descriptor is needed. Since the noise and other inference is highly variable in real-life settings the impressions from the same shoe can greatly differ depending on the circumstances causing high intra-class variance. On Figure 3.1 four samples from the same shoeprint are shown, noise occurs not only in the background but also on the shoeprint itself, the appearance of noise changes within one image and illumination changes occur as well. Furthermore on three examples only a partial shoeprint is depicted and the size and resolution of the samples also varies.

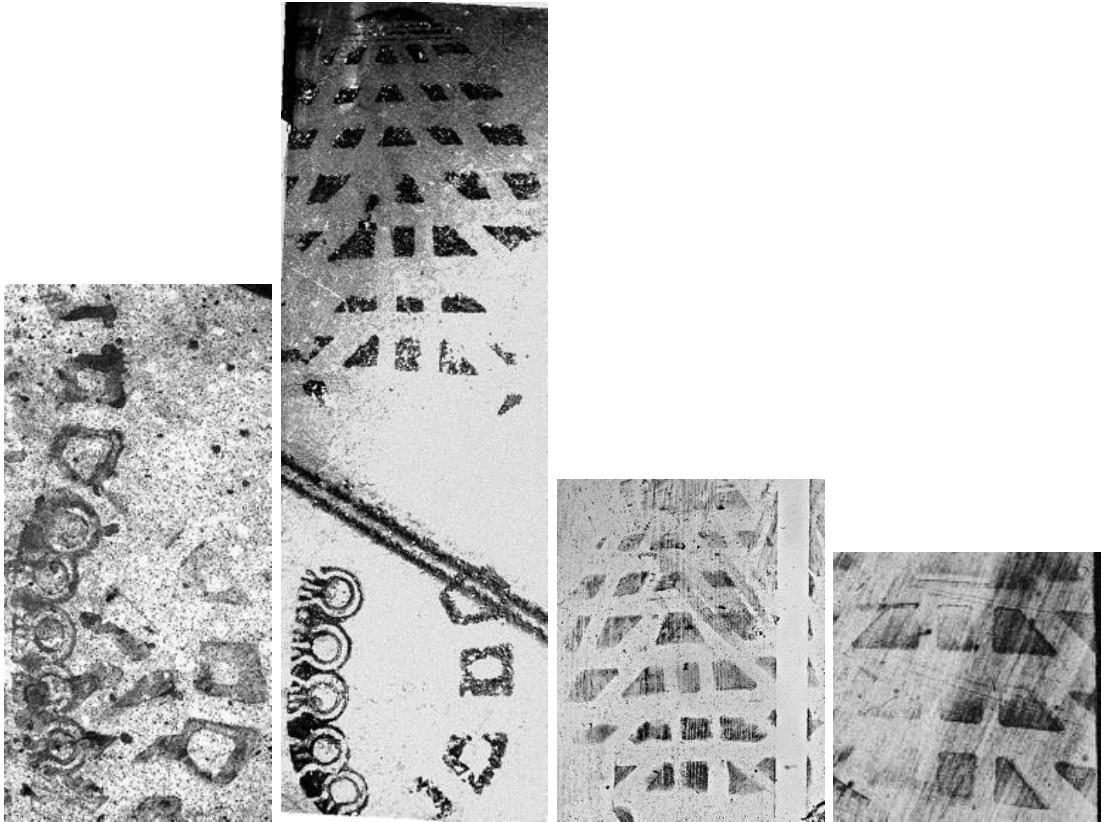


Figure 3.1: Example impressions of the same shoe from the FID-300 [KAV14] dataset

For that reason a discriminative feature learning algorithm is proposed based on the work of Guo et al. [GZP12]. In their work 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 on feature selection where only the most discriminative descriptors are chosen and propagated further. In the first layer, where robustness is granted, the features are selected which describe the majority of given texture, by finding the frequently appearing descriptors and ignoring the rare ones since they are more sensitive to noise. For that a threshold is defined, all descriptors of the given texture are selected in the order of their frequencies starting with the most frequent one. If the already selected descriptors cover a bigger region of the original texture than the given threshold, the selection process stops and the chosen descriptors are propagated for the next level. The selection of threshold is crucial in this process, 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 following two layers work according to Fisher's Separation Criteria, the second layer aims to minimize intra-class variance, whereas the third layer ensures maximum inter-

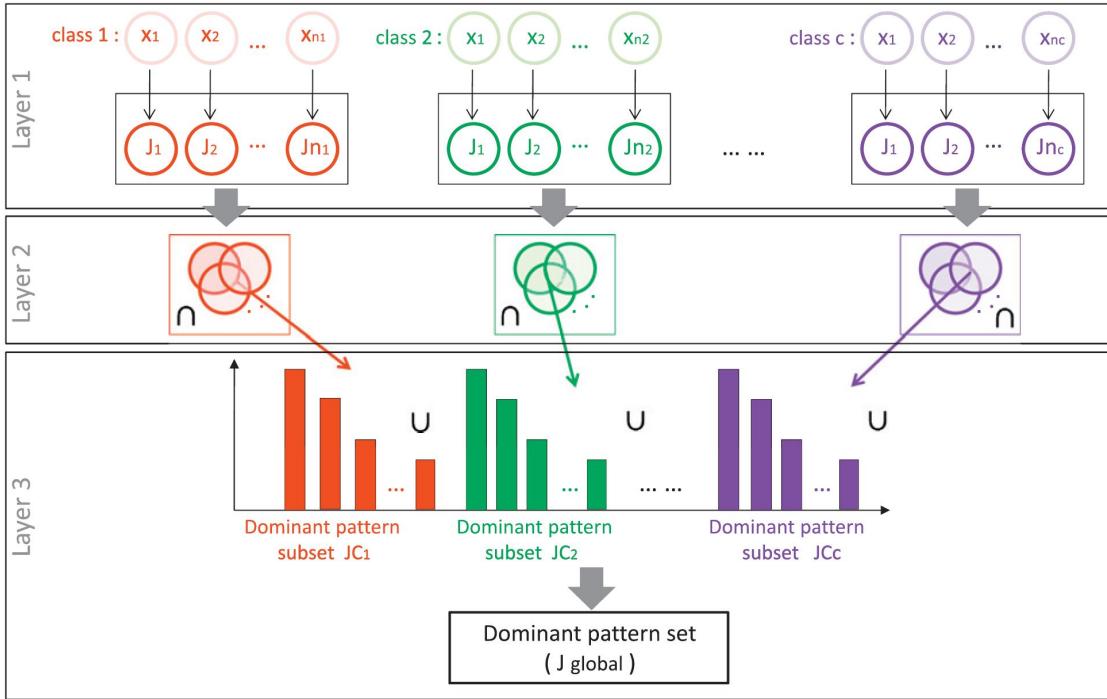


Figure 3.2: The workflow of the three-layered feature learning algorithm [GZP12].

class distance. The second layer provides discriminative power. It gathers all selected descriptors from the previous layer of the same texture on different samples and intersects those sets. The main assumption is that the same texture have similar properties under varying circumstances. In this layer features are selected which occur in every sample of the given texture.

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 is used, because the goal is to distinguish between noise and information, two classes, pattern and noise, are defined. Since the noise has no regularity along the samples, descriptors for shoe patterns are only learned. The primary assumption is that there is a fundamental similarity between every shoeprint image, e.g. regular line structures, recurring basic patterns, which is detected and learned. 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 shoeprint detection. In the research of shoeprint identification both frequency based feature descriptors, such as Fourier-Mellin Transform [WWZ19], [GBCN08], and

3. PATTERN DETECTION

SIFT [NBC⁺09], [RLL⁺17] are commonly used descriptors, thus those are also selected for feature learning.

The three-layered learning algorithm is implemented for three different feature descriptors, LBP, Fourier-Mellin and SIFT, but the learning process follows the same steps as proposed by Guo et al. [GZP12]. In the first layer Guo et al. [GZP12] propose a threshold based on the area the selected features cover on the original image. However this criteria is altered in the case of LBP and SIFT to adjust the learning algorithm to the needs of this research. Three-layered learning 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. Furthermore the goal is to distinguish between noise and shoepoint and not to find discriminative features within multiple texture images. For that reason features are firstly extracted from both fore- and background of the samples, where the foreground contains the exact shoepoint pattern and the background is the noisy ground where the shoepoint is lying. Since there is no labelled data in the FID-300 [KAV14] dataset available, the samples chosen for training were labelled manually. Figure 3.3 shows two example images the black regions of the mask show the shoepoint information whereas the white pixels indicate the background. After that, to achieve high inter-class distance despite the absence of the third layer, frequent descriptors of the noise are determined and all of them are eliminated from the pattern descriptors. The remaining pattern descriptors are propagated to the next level. To have a reasonable learning time in case of Fourier-Mellin Transform no such modification is made. After all descriptors are calculated those are selected which occurrence exceed a given threshold.

The easiest way to determine the occurrences of one descriptor is to calculate the complete matches within one sample. This method however results in high amount of descriptors with little number of occurrence. For that reason, a similarity measure is used for all three candidate features, if the similarity value reaches a predefined threshold the two descriptors are considered as the same, they are merged and the number for occurrence is incremented. For LBP histogram correlation is used for comparison, for SIFT Brute-Force matcher is applied and for the Fourier-Mellin Transform the similarity measure proposed in [GBCN08] used. The correlation coefficient between two Fourier-Mellin descriptors is given by:

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

where FM_1 and FM_2 are two Fourier-Mellin Features to compare and $\overline{FM_1}$ and $\overline{FM_2}$ are the mean values of the corresponding features.

At the end of the learning process a descriptor pool is created. The shoepoint area of a new input is determined based on those learnt features using the same similarity measures as in the learning phase.

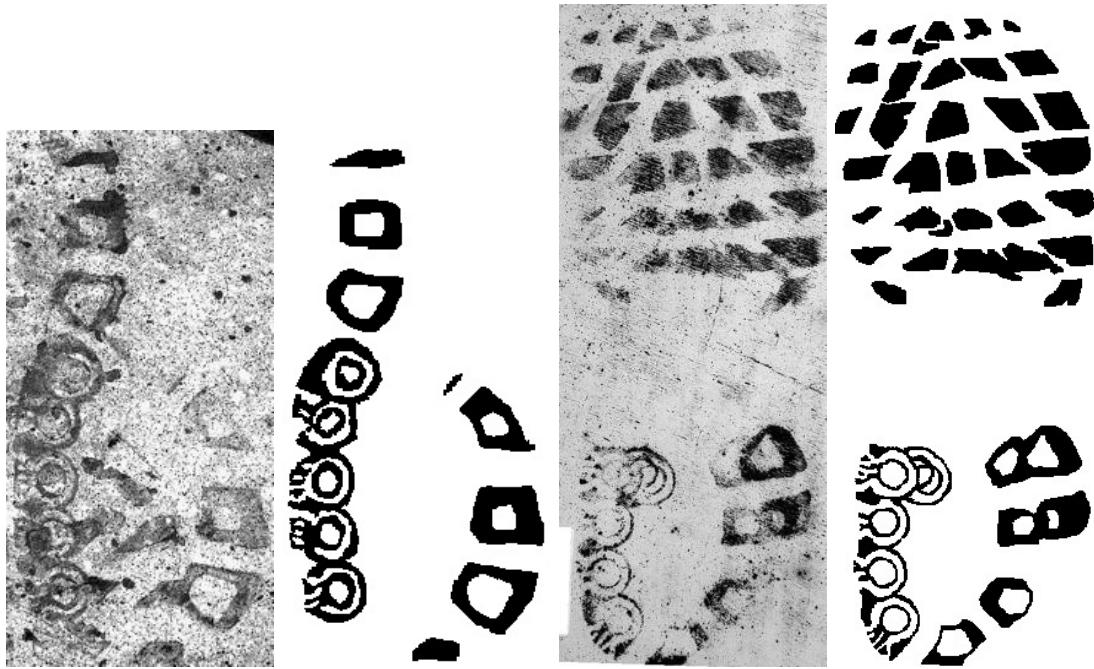


Figure 3.3: Examples from FID-300 [KAV14] dataset chosen for three-layered learning and their corresponding manually labelled masks.

3.2 Implementation

In this section implementation details of the modified three-layered learning algorithms are revealed and parameter settings are specified. This and the following algorithms are written in Python 2.7 [VRDJ95] using OpenCV 4.1.1 [Bra00]. Being a prototypical implementation, the values of the variables stated in this section are determined through experiments.

When starting the application the training data is read. It consists of the original shoeprint images and mask images where the relevant regions of the corresponding shoeprint is marked. The two sets of pictures are stored in a vector, it is assumed that the mask has the exact same size as the corresponding shoeprint image, that every image has a mask and that the images and masks are stored in the same order. For training a limited amount of seven images were selected and labelled manually. The FID-300 [KAV14] dataset consists of 300 real-life forensic samples and 1175 synthetic samples, thus not every synthetic sample has a real pair, and the majority of synthetic images with corresponding real samples have only one or two matching pairs in the database. The chosen seven images depict shoeprint images from the same shoe, one of them is a synthetic sample the rest of them is a real one.

To speed up the shoeprint recognition, the learnt features are stored in a .txt file. It is determined if a new learning process starts or the already calculated descriptors are

3. PATTERN DETECTION

read from file. When the learning starts the descriptors for all pixels are calculated. For the Fourier-Mellin calculation the images are 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. After that similar descriptors are merged and their frequency is calculated. If the histogram correlation of two LBP features is higher than 90%, if the distance between two SIFT features is lower than 300 and if the correlation between two Fourier-Mellin features is higher than 1.4 the two descriptors are combined. 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 descriptors are eliminated which have higher than 90% correlation with or have smaller than 250 distance to any dominant noise descriptor. The remaining features are then propagated to the next layer. In case of Fourier-Mellin Transform all descriptors are selected which occur at least 10 times on the given pattern.

In the second layer the dominant descriptors of multiple samples are compared. If there is a feature in every training image with a high enough similarity the feature is chosen and added to the final descriptor pool. 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 1.4, the feature is chosen and added to the final set of descriptors. At the end of this process the whole set of descriptor is written into file.

If the learned descriptors are available the input image is processed. The descriptors of the input image are determined similarly as in the training phase. For every pixel of the input is a descriptor calculated, in case of Fourier-Mellin transform the borders of the input are extended by mirroring. The output image is calculated by comparing the descriptors of the input with the ones 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. At the end of the application the resulting images are saved.

3.3 Evaluation

In this section the results of the modified three-layered learning algorithm are presented and discussed. Furthermore it is examined, if this approach is usable for real shoepoint identification algorithms.

Two kinds of experiments were conducted to test the performance of the learned descriptors. First images from training data are processed to see how many descriptors from the original image were eliminated and to identify the common features in the training dataset. After that new images are set as input to examine if the features learned on one kind of data are able to describe other kinds of shoepoint as well. Figure 3.4 and Figure

3.5 are example images from the first scenario, since Figure 3.4a and Figure 3.5a are part of the training set. The visible difference between Figure 3.4b and Figure 3.4c as well as between Figure 3.5b and Figure 3.5c shows that bigger radius is a better choice for shoeprint description. Considering the middle region of Figure 3.4, the background on Figure 3.4c is less noisy than on Figure 3.4b. This difference is more outstanding on Figure 3.5 where on Figure 3.5b no shoeprint can be recognized meanwhile on Figure 3.5c outlines of the pattern are visible. Still focusing on the middle area of Figure 3.4, Figure 3.4e shows that the SIFT descriptor is similarly robust against noise as the LBP descriptor. However, examining Figure 3.4d the Fourier-Mellin descriptor has more difficulties with the background area, where the middle regions is less homogen than on Figure 3.4c and on Figure 3.4e. On the other hand, inspecting the bottom left side of the shoeprint, it is seen that the Fourier-Mellin features managed to find the whole area of pattern, whereas only a few outlines are recognizable on Figure 3.4c and on Figure 3.4e. Interestingly that said region is more recognizable on 3.5b than on 3.5c. Based on these observations it is stated that the features less robust against noise are able to find a higher portion of pattern area than those with higher robustness. In exchange the background area stays noisy and further image processing is needed. Comparing all results of the two images, Figure 3.4 and Figure 3.5, the performance strongly depends on the input quality. On Figure 3.5 and Figure 3.5e the pattern is less recognizable than on the previous example and a higher amount of background pixels are labeled wrongly as pattern. On cluttered images Fourier-Mellin features seem to outperform both SIFT and LBP descriptors. Overall this testing scenario shows ambivalent results. On the one hand on clear samples LBP and SIFT is able to distinguish between fore- and background correctly, on the other hand, if the input data is cluttered those two descriptors are barely able to recognize the pattern and the Fourier-Mellin descriptor provides the clearest results.

On Figure 3.6 the results of the modified three-layered learning algorithm on a non-training sample are shown. The results displayed on Figure 3.6b and on Figure 3.6c strengthens the observation that bigger neighborhood LBP outperforms the smaller radius descriptor. Furthermore examining the middle, noisy part of the images in all three cases a homogenous background is seen containing some falsely labelled pixels. Observing the bottom part of the shoe a weakness of LBP and SIFT already discussed is noticeable. On the original image there are small structures which are part of the shoeprint. Similarly as on the bottom left part of Figure 3.4 the fine structures are only partially recognized, comparing Figure 3.6d to Figure 3.6c and to Figure 3.6e Fourier-Mellin outperforms LBP and SIFT again labelling the whole area correctly.

To summarize LBP and SIFT features are less responsive in the noisy area and have difficulties to find fine patterns while the Fourier-Mellin transform labels bigger parts of the shoeprint correctly whereas it more often mistakes the noise as foreground. Based on the testing images Fourier-Mellin recognize the shoeprint pixels in noisy images better than LBP and SIFT. A possible solution is to combine all three feature descriptors and propose a weighting score across them, assuming that all three feature sets were able

3. PATTERN DETECTION

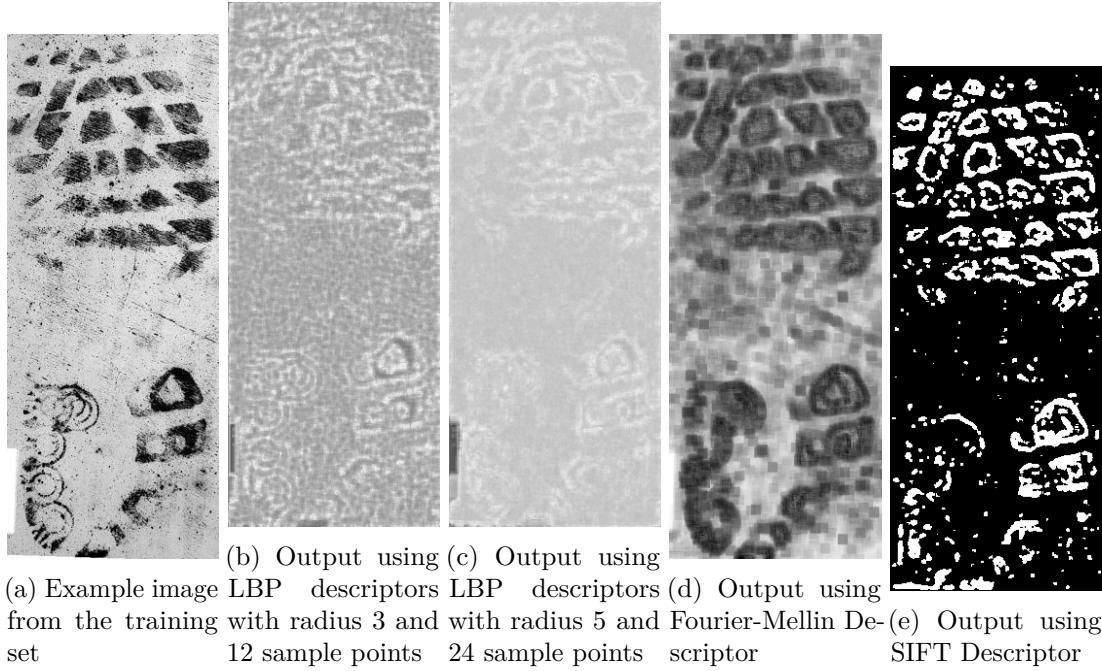


Figure 3.4: Output of the modified three-layered learning algorithm on an image from training set

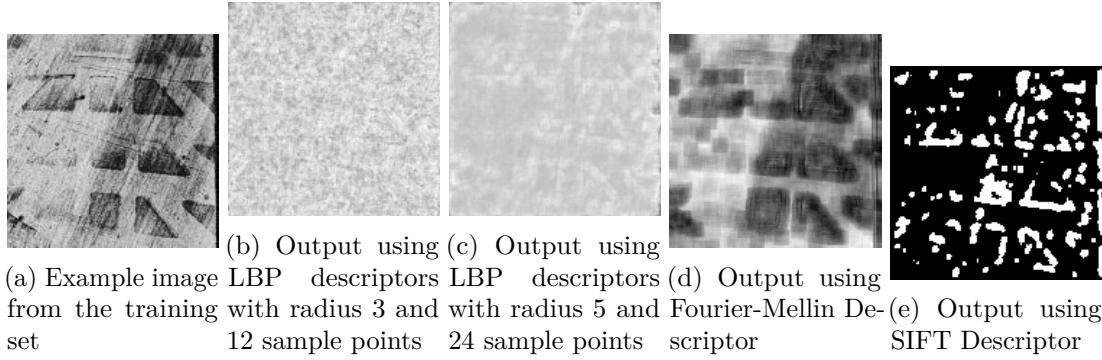


Figure 3.5: Output of the modified three-layered learning algorithm on an image from training set

to provide usable results for further processing. However, in low quality images that is not necessarily the case. Even though Figure 3.5 was part of the testing dataset, LBP and SIFT did not label the pattern correctly. That problem can originate from the learning process, where all dominant descriptors of the noise part are eliminated from the pattern descriptors. On the other hand on Figure 3.5e significant amount of background is labelled as pattern, which indicates that too many noise descriptors were selected into the descriptors pool. Furthermore there are several lower quality samples in the

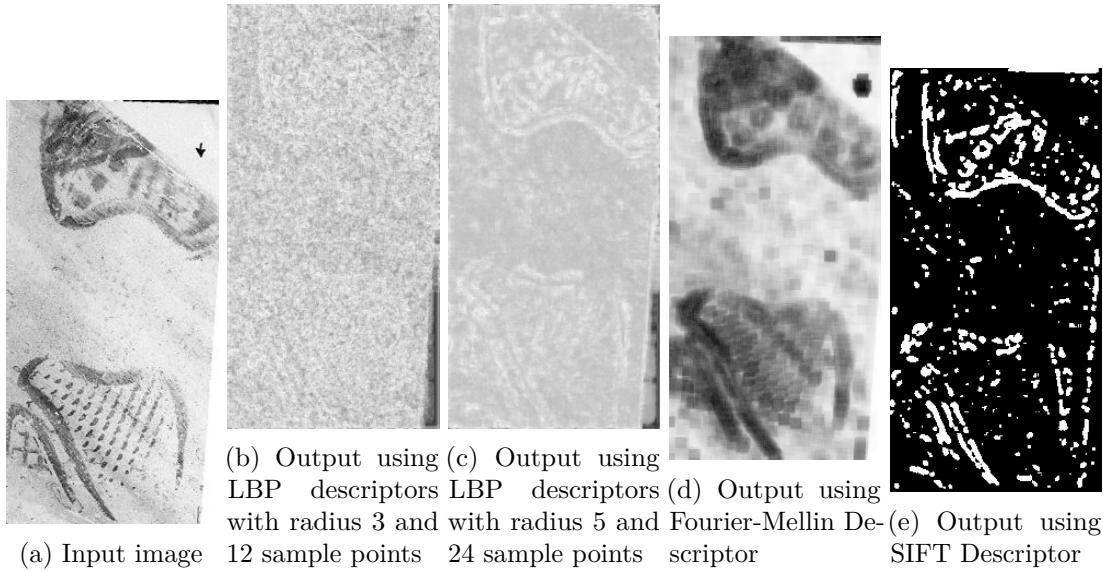


Figure 3.6: Output of the modified three-layered learning algorithm

whole dataset available then the one presented on Figure 3.5. It has to be emphasized that the descriptors were learned on a small dataset. 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, can lead to a broader texture pool. However, this also results in a higher ratio of noise descriptors in the final feature set. High noise ratio is a limitation of this algorithm in its current form. For that reason noise suppression techniques are examined in the following. In the future the modified three-layered learning algorithm can be further developed and used jointly with noise suppression techniques to solve its weakness and to take advantage of its results.

CHAPTER

4

Fully-Automated Noise Elimination

Since the algorithm introduced in the previous chapter had difficulties when processing noisy or cluttered data a noise suppression technique is introduced now, a fully-automated in the following one a semi-automated approach is proposed. 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 first and the information is enhanced afterwards.

To compare the proposed methods more easily, the structure of this chapter is the same as the previous one. The methodology is discussed first, afterwards implementation details are given, finally the results are examined and evaluation of the results is discussed.

4.1 Methodology

The method proposed in this chapter has three main steps, first the noise is suppressed, after that the shoeprint pattern is enhanced and the result image is generated lastly by thresholding. The workflow of the entire algorithm is shown on Figure 4.1.

The first part of the algorithm is responsible for noise suppression. Since there is little research done on enhancement of real-life forensic images two filters proven to be effective in natural image noise elimination are combined. As Li et al. [LLGF14] state, Wiener filter is among the most popular techniques in the topic of noise reduction, however it does not reconstruct the signal data it only suppresses the noise component. Wiener filter works in the signal domain, where it estimates the original image based on the cluttered input. The filter is given by following Equation [Win]:

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

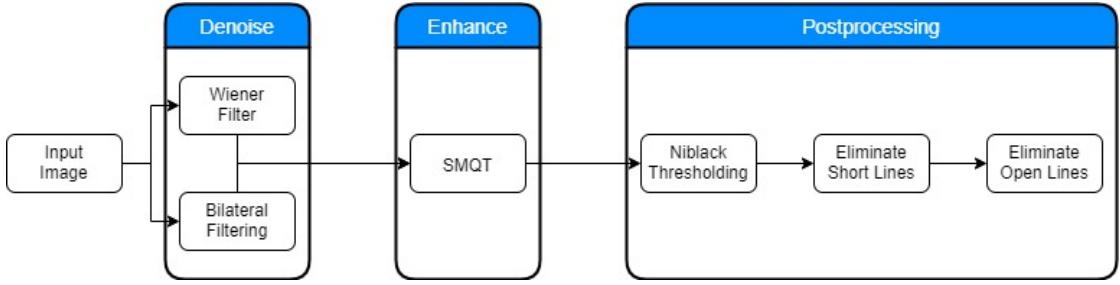


Figure 4.1: The workflow of the fully automated noise suppression algorithm.

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. 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 input image. The other technique used for noise suppression is Bilateral Filtering as proposed in [HKWL13] and in [ZSP⁺16]. 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 shoepoint pattern. 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 shoepoint patterns are preserved.

For enhancement Successive Mean Quantization Transform (SMQT) [Nil13] is used as proposed by Katireddy et al. [KS17]. SMQT is recursive algorithm which splits the data into two parts in every recursion level depending on the current data value being smaller or bigger than the mean of the given data part. In this way the structure of the data is exposed. Everytime the image is separated into two parts, it is noted first which pixel is under (0) and which one is above (1) the current mean. The algorithm then recursively continues on both subgroups of the images, splitting and noting the relative pixel value again. Note that the pixels in the same subgroup does not have to be neighboring, the clustering is solely based on the illumination value. The recursion stops if the predefined depth is achieved. To finish the transformation the noted cluster 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 after converting into decimal value it is updated as the intensity of the corresponding pixel.

The last part of the application is the postprocessing step where the input is converted into a binary image and small inconsistencies and remaining noise are eliminated. When binarizing footprint images Otsu's technique [AH08], [AK17], [WWZ19] or adaptive thresholding [WSYZ14] is used most frequently. However in the proposed enhancement approach Niblack Binarization Method (NBM) [Nib85] is preferred. NBM is a local

thresholding method and it 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. 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 as familiar since they 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. There is however one disadvantage of local binarization and that is 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, area. For that reason two other postprocessing methods are also implemented. The first one eliminates short lines on the image, based on the assumption that the actual shoe pattern outlines are recognized and they build bigger, coherent edges. The second algorithm stands on the same principle as well, from the remaining lines are all open structures eliminated unless the length is higher than a given threshold bigger than the one in the previous step. It is based on two observations, first if the complete structure of a pattern element is found a closed line structure is extracted, since NBM concentrates on the contours. Second, when there is no pattern element in a NBM subwindow only the edges of the remaining clutter and binarization artifacts are generated. Since those lines are most likely not closed, they are deleted. However it is still possible, that an open line structure is generated if only parts of the shoepoint are classified correctly. For that reason, not every open edge, only a subgroup of them, which are shorter than a threshold, is eliminated.

4.2 Implementation

Implementation details and parameter settings are discussed in the following. The programming tools are similar to the ones in the previous chapter, Python 2.7 [VRDJ95] with OpenCV 4.1.1 [Bra00] were used for development. Since all the methods presented in this thesis are prototypical applications, the parameter settings were determined experimentally.

At the start of the application the input image is read and forwarded to the denoising stage. The original image is processed separately with Wiener and Bilateral Filter. To apply the Wiener Filter correctly 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 kernel size of the Bilateral Filter is 9x9. When both filtered images are calculated their difference is determined and propagated to the enhancement step of the algorithm.

4. FULLY-AUTOMATED NOISE ELIMINATION

For SMQT calculation, 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. This table eases the calculation of mean value for every level of the SMQT algorithm. Additionally, since the entries are ordered, if the subgrouping starts, the values belonging to the same cluster will be neighboring columns. Figure 4.2 shows the first 14 columns of the occurrence map of an example image. When the occurrence map is ready the recursion starts, the mean of the given subgroup is determined and 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 cluster a zero is written in 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.

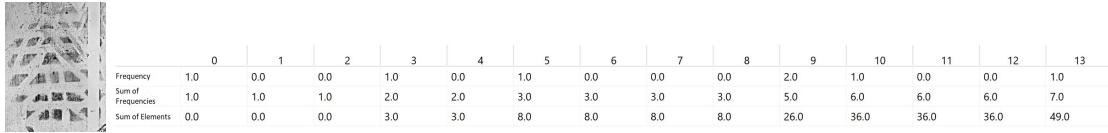


Figure 4.2: The first 14 columns of the occurrence map. The header of the table represents the possible pixel values

In the last step the image is thresholded according to NBM. The window size is the same as the size of the Wiener Filter, 5x5, the regularization parameter k of NBM is set to 0.5. After that short and not-long open edges are eliminated. The connected components of the image are extracted first considering 8-neighborhood. Afterwards every component is examined on their size, if they are smaller than a given threshold they are deleted. It is possible to set the threshold to the average size of every region, however, experimental results show that such a high criterion eliminate shoe pattern lines as well. For that reason a manually set threshold is used, which is set to 50 for every test image. The short line elimination algorithm is applied on both the fore- and on the background on the image. Lastly the remaining components are examined on their openness. If an open line structure is found which overall length is smaller than a manually set threshold it is deleted. This threshold is set to be 60 during the whole experiment.

4.3 Evaluation

In the following section the experimental results are presented. Furthermore, the performance and the usability of the proposed algorithm is discussed. Figures 4.3 and 4.4 show the output of the proposed algorithm on every stage and present the final results as well.

Figure 4.3 shows the output of the denoising pipeline on three example 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 difference between those two pictures is in the last row. Considering the background of the algorithms and the amount of noise on the images a light smoothing is done in the denoising phase. This decision was made to protect the valuable pattern information. Since no previous knowledge is available about the properties of shoe print, no aggressive filter kernel is used to make sure that the relevant information is preserved correctly. On the combined image a bright silhouette around the shoepoint image is visible, which was created by the Wiener Filter. This feature is exploited in a later stage of the algorithm, where an edge sensitive binarization method is used.

The second step of the algorithm is enhancement where SMQT is applied, and the results are shown on Figures 4.4b, 4.4f and 4.4j. The benefit is the SMQT algorithm is ambiguous. On the one hand focusing on the middle area of Figures 4.4b and 4.4f the difference between fore- and background is bigger than on the original image. The contours of the shoepoint are more outstanding than on the original image. Furthermore the halo effect is also more visible than on the output of the previous stage of the algorithm. Comparing Figure 4.4f to Figure 4.3l, the shoe pattern structures got darker and there is a brighter area around the contours than on Figure 4.3l. On the other hand not only the information but also the remaining noise elements got enhanced by SMQT. Even though the denoising stage is finished, since only moderate smoothing was applied, there is a considerable noise on the input images for SMQT. The soft filtering is able to cope with Gaussian noise like on Figure 4.3d but it does not eliminate bigger clutter particles like on the left side of Figure 4.3h or on the upper region of Figure 4.3l. The SMQT enhancement makes those clutter more obvious by increasing the color difference between the noise and the homogenous background, for example the already mentioned patches on the left side of Figure 4.3h are now more visible than on the original image 4.3e. Additionally SMQT has similar effect like Histogram Equalization, the background on Figure 4.3d 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.

The closing step of the algorithm is postprocessing, consisting of binarization (the third column of Figure 4.4) and line-clutter elimination (in the fourth column of Figure 4.4). As expected NBM is able to take advantage of the halo effect and finds the shoe pattern correctly. On all three images, Figure 4.4c, Figure 4.4g and Figure 4.4l, the contours of the patterns are clearly visible. However, the drawbacks of NBM and the enhanced noise by SMQT are evident, since among the correct outlines a high amount of noise is generated. For that reason the closing step of short line and open line structure

elimination is executed. On Figures 4.4d and 4.4h the majority of background is cleared up and the shoeprint pattern is shown correctly. On the other hand on Figure 4.4l the original pattern is not recognizable anymore, along with the background noise relevant contour edges were also eliminated, and noise lines attached to the shoeprint pattern were kept.

The presented results of the fully-automated noise elimination algorithm indicate two things. First, NBM with the current postprocessing technique works better on samples with fine lines, Figure 4.3a, and with detailed small structures, Figure 4.3e. As already discussed and shown NBM tends to generate line noise on homogenous areas, if the shoeprint consist of dense pattern, there is no big, homogenous background between the different structure elements thus no noise is generated. Additionally the heuristics which the two postprocessing steps are based on do not apply on every sample on the database. On Figure 4.4k even though there is a high amount of noise the pattern is clearly recognizable whereas on Figure 4.4l no shoeprint can be seen anymore. This implies that the settings of the line elimination algorithms are not correct and that further postprocessing is needed. Several pattern edges were eliminated because they were too short. The threshold for line 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 sample available. Alternatively looking to Figure 4.4l the outlines of the pattern were found originally, but because of the noise they were not connected. In the application 8-neighborhood is considered, a more allowing connectivity criteria, such as neighborhood in a given range, leads closed contour edges. But this causes more connectivity among noise as well. Furthermore it also magnifies the second problem occurring on Figure 4.4l, that noise lines are connected to the contours, thus are not eliminated. An additional postprocessing step is needed to examine the remaining edges and determine which connection was made unintentionally.

Second, Figures 4.4h and 4.4l show that the performance of fully-automated noise elimination algorithm strongly depends on the quality of the input image. This algorithm has a similar limitation like the modified three-layered learning approach proposed in the previous chapter. Initially the filter settings of the denoising step are not able to handle cluttered noise seen on Figures 4.3e and 4.3i. Chatterjee et al. [CM11] proposed a method which represents the estimated image for Wiener Filter based on geometrically and photometrically similar regions. Even though their approach had promising results, it was tested on data with Gaussian noise and not with irregular cluttered noise seen on the majority if FID-300 images. Later, in the enhancement step not only the information but the remaining noise elements are also enhanced. Lastly, during binarization an extensive amount of noise is generated which is not eliminated perfectly afterwards. Saxena et al. [Sax19] analyse several improvements for NBM to make it more robust and performant in poor conditions. But based on the experiments presented in this section image quality and high noise ratio is a major limitation of the method.

Furthermore, the test images showed in this chapter are high quality samples compared

to other members of the FID-300 database. The evaluation of the modified three-layered feature learning algorithm and the fully-automated noise elimination method show that these approaches are not able to overcome the limitation of the high amount of varying clutter on the sample images. The second algorithm showed promising results in ideal and very low quality in poor conditions, whereas the first one delivered average results regardless the initial properties of the input image. Furthermore, it also has to be mentioned that both algorithms were tested on high or middle quality samples, images with poor shoeprints or high noise ratio like on Figure 2.1d were excluded during evaluation. To be able to process those kind of images as well, additional knowledge about the input is needed. For that reason a semi-automated algorithm is introduced next. Since the modified three-layered feature learning algorithm is able to make an approximate estimation about the location of pattern, information about the noise is collected to successfully enhance the shoeprint impressions.

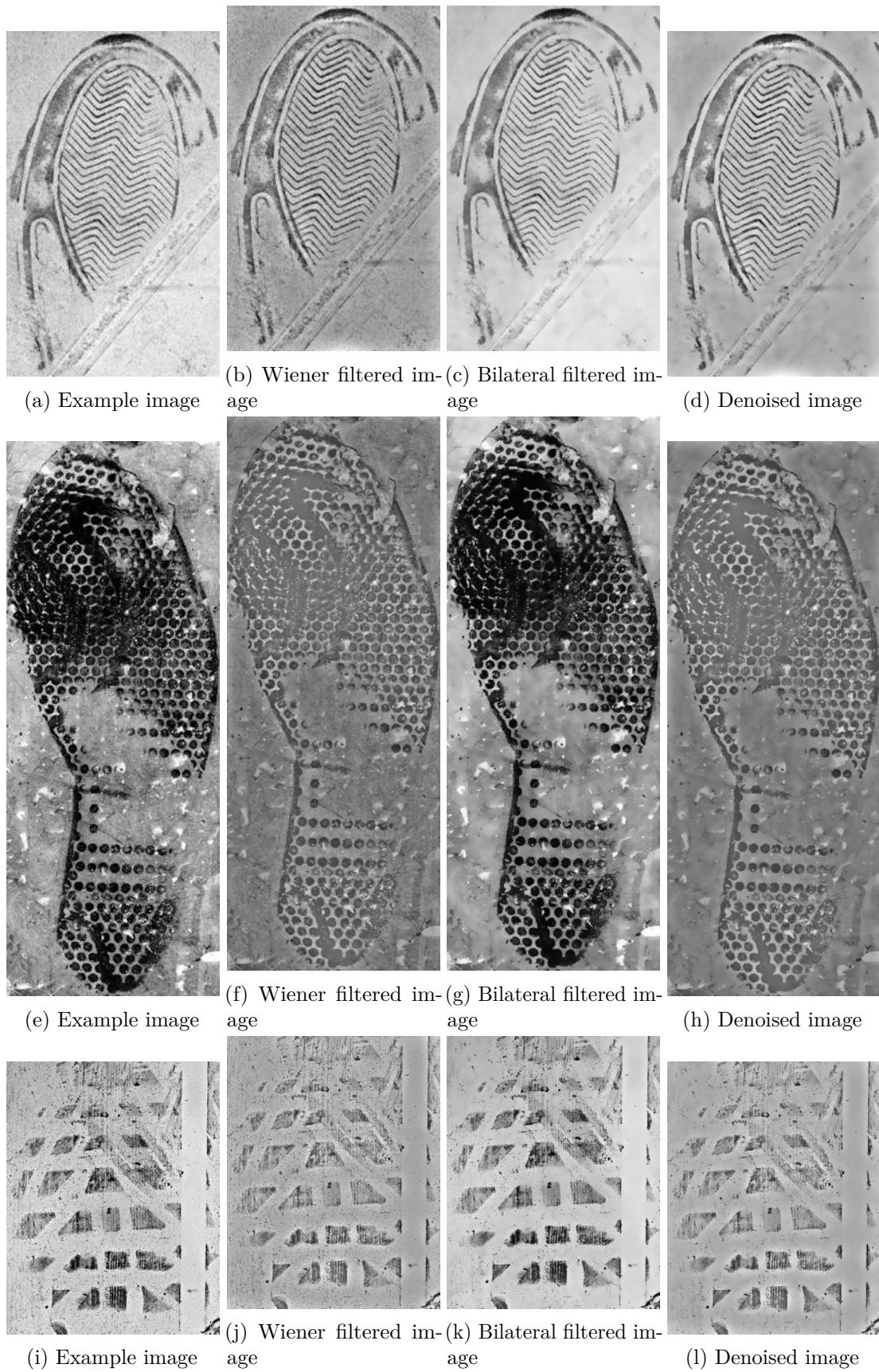


Figure 4.3: Example for the results of the denoising stage

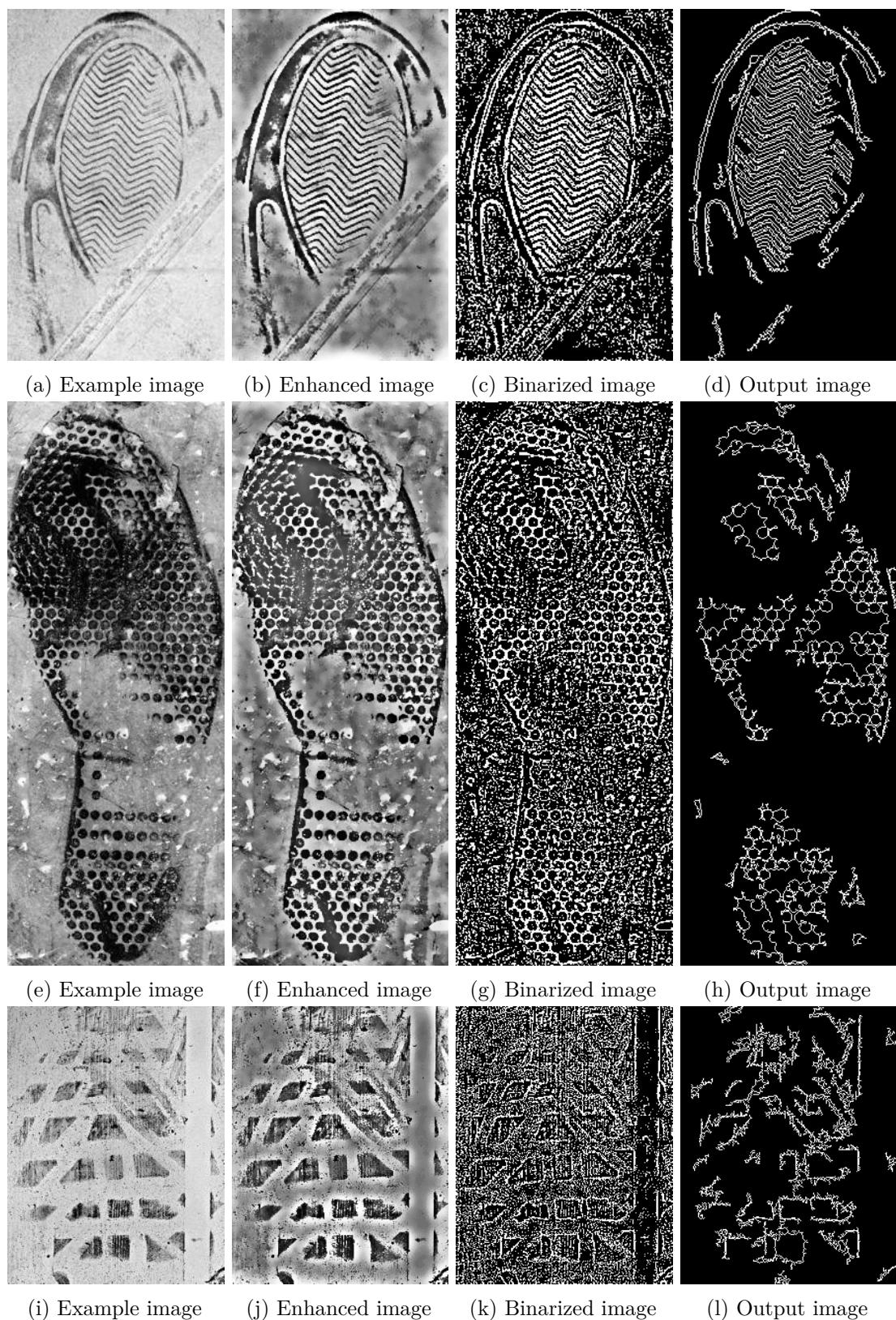


Figure 4.4: Example for the results of the enhancement and postprocessing stage

CHAPTER

5

Semi-Automated Shoeprint Enhancement

The prototypical algorithms presented and evaluated in the previous two chapters have shown that high noise ratio and the versatile appearance of clutter is a crucial difficulty in real-life shoeprint enhancement. The experiments also indicate that without previous knowledge about either the shoeprint or the noise on the sample it is difficult to distinguish between them. Since noise tends to appear both on the fore- as well as on the background region, semi-automated algorithm is introduced in this chapter, where user input about the noise is expected.

The structure of this chapter is similar to the two previous ones. First the theoretical background and overall design of the proposed method is presented, after that an overview about the implementation is given and finally the results are evaluated briefly.

5.1 Methodology

The algorithm proposed in this chapter has two main steps, first the background and foreground is separated, after that the foreground is enhanced. There is one precondition when starting the application, the user has to select a pixel in the background area where no pattern information, so foreground, is in the neighboring area. The full pipeline is shown on Figure 5.1.

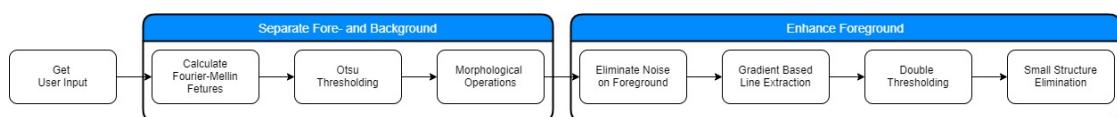


Figure 5.1: The workflow of the semi-automated noise suppression algorithm.

Based on the user input the background area is determined according to the Fourier-Mellin features [SD86]. Fourier-Mellin features are scale, rotation and translation invariant descriptors and are calculated according to the following Equation [KG11]:

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

where u and v are the Mellin and the Fourier transform parameter respectively. If the same 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 that the feature descriptor is converted to Log-Polar coordinates, so that said image transformations occur as simple translation in the descriptor. Log-Polar Transform is used to achieve robustness against scale and rotation. The Log-Polar coordinates of any 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 5.2 as well.

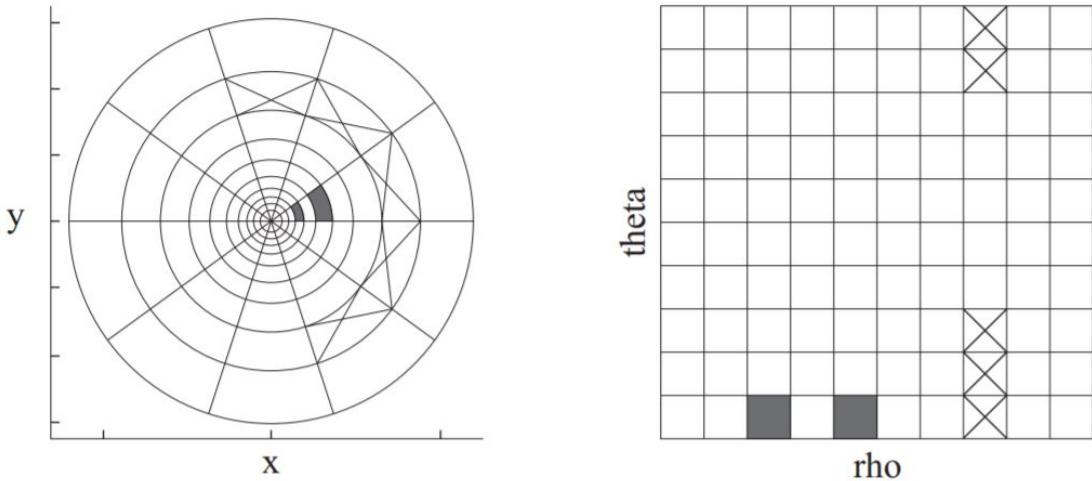


Figure 5.2: Illustration of the relation between Log-Polar and Cartesian coordinates [SPK12]

Fourier-Mellin features were already used for shoeprint description in both synthetic [GBCN08] and real [WWZ19] datasets, 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 testing images. Even though those dominant features

are determined, noise has an irregular structure by definition as well. To prepare the descriptor for that variance three-different sized samples are extracted around the point determined by user input. The Fourier-Mellin descriptors of the three subimages represent the noise, after that every pixel of the input is examined for correspondence with the same metric defined in Chapter 3 3.1. The calculated similarities are stored in a correlation map which is binarized by Otsu's method afterwards. Otsu thresholding is proven to be effective in all three kinds of data, synthetic [AH08], [AK17], restricted [KYZ14] and also on real forensic datasets [WWZ19]. To complete the background mask morphological operations are used to eliminate small holes and rough contours as proposed in many approaches for shoeprint identification [WSYZ14], [KYZ14], [LWS14], [TSKC10].

After the background mask is generated the foreground of the image is enhanced in three main steps. Noise on the shoeprint pattern is suppressed first, afterwards the edge information is extracted and the binary images is generated lastly. Wavelet transform is often used for natural image denoising [XWYH16], [SPV16], for fingerprint [LFLY12] and for shoeprint impression enhancement [KS17] as well. During wavelet transform the image is decomposed into subbands, and the subimages are processed separately. Assuming four subbands, the LL image contains the edge information and the other ones the intensity values. Depending on the properties of 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 simple Fourier transform is applied. When the background is separated a model about the appearance of the noise is built. Processing the foreground area and the created noise model with Discrete Fourier Transform, a signal domain representation of both images is created. The noise is eliminated by subtracting the noise frequencies from the foreground frequencies and inverting the resulting image with the Inverse Fourier Transform.

For line extraction 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. 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. 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} stays for the minimum and maximum difference on the entire image. For denoising the mean illumination value of pixels in the same class is calculated and the original pixel value is replaced with the average of the cluster. Zeng et al. [ZL11] propose an additional weighting scheme as well, so that pixels in the same class with smaller geometrical distance have a higher weight than those further away on the image while determining the mean. The region based non-local means algorithm was originally developed for noise removal on natural images but it is also applicable for line extraction on shoeprint impressions. Since the classification is made based on the gradients, edge properties are the essential clustering criteria. Assuming that both the shoeprint pattern and the remaining noise on the image have common properties which are different to each other, pattern and noise classes are created at the end of the clustering. However, the weighting method is skipped in this implementation, geometrical location is not relevant in this use-case for two reasons. First the size of the region the image was taken of is smaller in general than in case of natural images. Second, the shoeprint impressions are 2 Dimensional images in their original form as well. Since there is no depth difference within the image, the entire area is considered as one connected region unlike on natural images where the objects in the foreground and the background do not belong together. On the other hand another criteria is added to the clustering algorithm. After determining the classes if one or more groups have more members than a given threshold, histogram equalization is applied on the image where the differences of gradients are stored and the classes are recalculated. This optimization is needed to make the classes representative and 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 line clustering algorithm is applied on the entire image, not only on the foreground, the class with the highest amount of members is eliminated. This decision is based on the assumption that noise is found on the entire image, whereas pattern edges are located only on given areas. Furthermore, it is also assumed that unlike shoeprint contours, where the edge properties change according to the original shoeprint structures and pressure, noise have similar appearance on the entire sample. After that the image is binarized with the adaptive thresholding technique [LZ96] popular in both fields of shoeprint recognition [WSYZ14], [LWS14] and also of natural image processing [XWYH16]. Adaptive thresholding is a local binarization technique similar to NBM, generating multiple thresholds for every subregion of an image.

The closing step is postprocessing, where small discontinuities and remaining noise are eliminated. Despite in the previous step noise edges belonging to the biggest class were

eliminated, other noise lines assigned to different classes are still part of the image. The problems is solved with short line elimination. Remaining noise elimination while preserving the shoe pattern is possible because of the assumption that noise consists of cluttered small components whereas the shoepoint is built of bigger, connected structures.

5.2 Implementation

In the following implementation details of the proposed algorithm are given. Identical to the implementation of the previously discussed methods Python 2.7 [VRDJ95] with OpenCV 4.1.1 [Bra00] was used. Because of the prototypical implementation, the parameter settings were determined experimentally.

Starting up the application the user is required to select a point on the input, which neighborhood is representative for the dominant noise on the given image. Based on the information given by the user background and foreground of the sample are separated. To do so the Fourier-Mellin Features of the selected neighborhood are compared to the neighborhoods of every other pixel on the image. Three different sized neighborhoods, diameter of 6, 12 and 18, are obtained and the average correlation value is noted. The borders of the image is extended by mirroring to ensure that the calculated Fourier-Mellin Features have uniform size for every pixel on the image. The resulting correlation image is thresholded with Otsu's method. Afterwards two morphological operations are applied using an 8x8 kernel, Opening first and Closing second. The last step is normalization to the range of 0 to 1 to generate a binary image for background area.

After determining the background area the foreground is enhanced. The first step is to eliminate noise on the foreground according to the noise estimation made on the background area. If the background area is smaller than 20% of the image, this step is skipped, because in this case no robust model about the clutter can be made. To estimate the average noise appearance the original image is extended again and the Fourier image of size 6x6 of every pixel is calculated. To build a noise model two methods were used. The first one is simple average calculation. The second one represents noise by the centroids of the k-means algorithm. For that double criteria is set, the calculation terminates if the wished accuracy is achieved, however that is set to 1.0, or stop after 10 iterations. k is set to 8, so 8 centroids are determined, thus 8 submodels are used when eliminating the noise on the pattern. In case of the averaged noise model, the Fourier image of noise is subtracted from the Fourier image of the foreground pixel 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 similarity is determined using the Equation for Fourier-Mellin Feature comparison 3.1. When the centroid is decided identical to the approach of the other noise model, the difference between both Fourier images is calculated and the foreground is reconstructed using the Inverse Fourier Transform.

When the shoepoint pattern is enhanced the edges of the entire image are enhanced.

First the differences between the eigenvalues are calculated for every pixel. To do so the image is blurred first with the Gaussian kernel of size 11x11 to make the calculation more robust. To extract the gradients the Sobel filter is used in x and y direction separately. If the tensor matrix is calculated the eigenvalues are calculated as given in Equation 5.1. The differences are clustered as Zeng et al. [ZL11] suggest, two parameter settings are implemented using 25 and 20 classes respectively. Afterwards it is examined if one cluster has more than 60% of all pixels, in that case the classification is not representative and Histogram Equalization is applied on the gradient difference image and the memberships are recalculated. Finally the non-local means are calculated for every subgroup and the pixels of every cluster are updated to the corresponding mean value.

In the postprocessing step the cluster with the most members is eliminated first, when no Histogram Equalization was applied on the generated image. In case of using histogram equalization the class memberships are more balanced, so no such heuristic is assumed. After that the image is binarized with adaptive thresholding, where the Gaussian weighted sum of the neighborhood is used, the window is 12 and constant is set to 2. To prevent thresholding artifacts the background mask is applied after the binary conversion. As the last step the small line structures are eliminated in both white and black channels. The connected components of the image using 8-connectivity are extracted and every element which region not bigger than 15 is deleted immediately. To be able to use the finally extracted and enhance shoeprint image later the resulting image is saved in file.

5.3 Evaluation

In this section the experimental results are shown and the performance of the algorithm is discussed. On Figures 5.3 and 5.4 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 show that varying the algorithm parameters makes a significant difference when the input had low quality. This can be explained with two factors. First, high quality images have less noise on the pattern, Figures 5.3a and ??, or the noise is less varying, Figure 5.3a, than on a lower quality sample, Figure 5.4a. Second, using 20 and 25 classes is overstated, on high quality images there is no high variance between the properties of edges. However choosing high amount of classes was an intentional decision, first the structure of the image is better preserved, when the means are calculated on a high amount of subgroups, second using fewer classes leads to eliminating valuable pattern information when deleting the members of the most populated class. Comparing the output of the two different kinds of noise models, there is very little difference between their performance. Both models were able to preserve pattern information where the foreground is not cluttered like on Figure 5.3f. Focusing on small details like the top

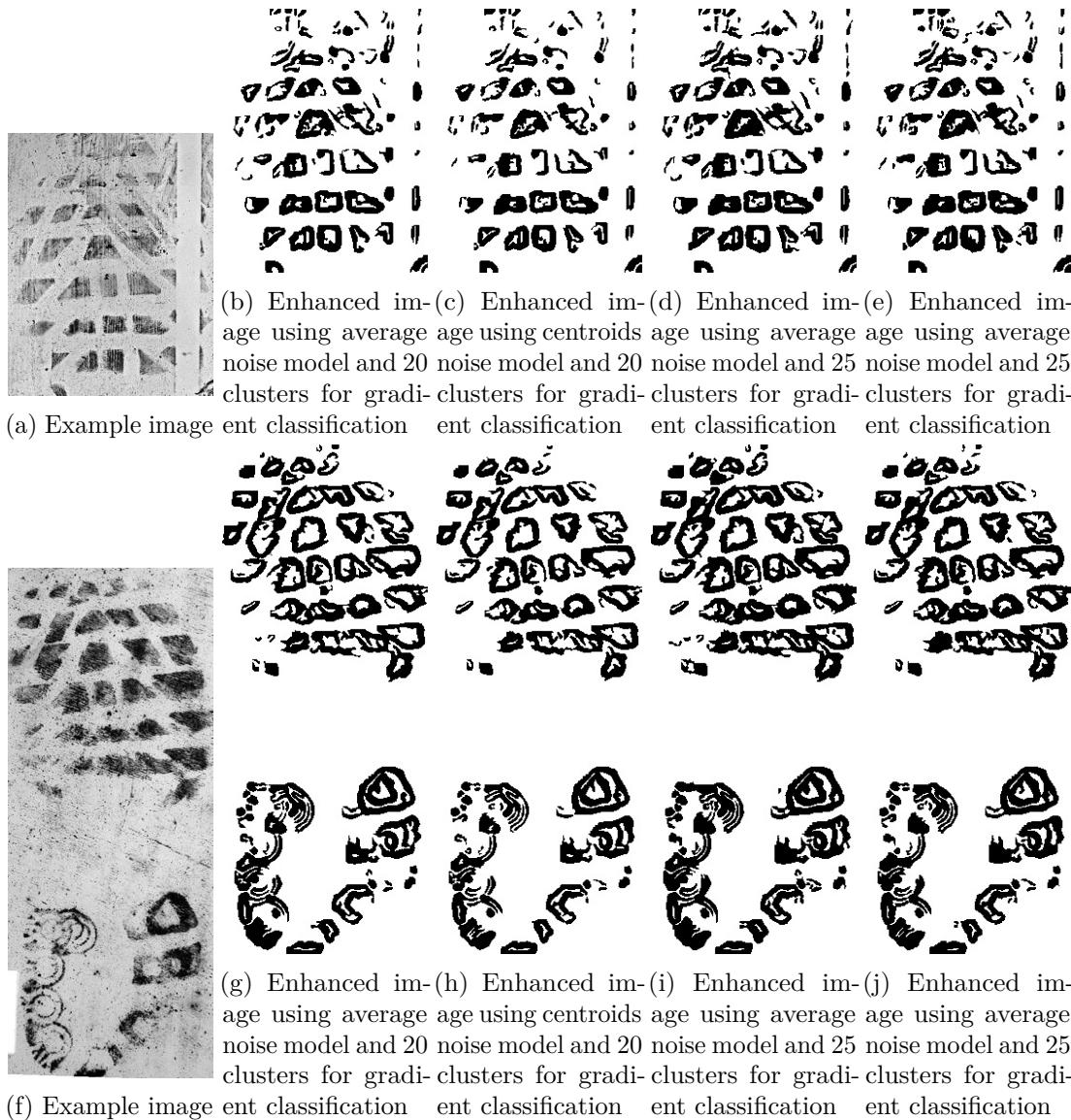


Figure 5.3: Example results of the semi-automated algorithm

of Figure 5.3a 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 every noise pixel results in a less detailed model than calculating more representatives, thus it is more destructive than the centroid representation. On the result images of Figure 5.4f significant shoeprint pattern parts are missing on the bottom area, that is the combined effect of several steps of the pipeline which is based on the heuristic that small regions are noise elements rather than relevant pattern information.

On Figure 5.5 the entire pipeline of the semi-automated algorithm is shown on two

example images. On Figures 5.5b 5.5h and 5.5n the binary mask is shown, where white denotes the background and black the foreground area. Inspecting the images the majority of the background is labelled correctly, however two limitations are visible. The first one is the bottom area of Figure 5.5h, where half of the fine line area of the actual shoeprint is cropped. The other one is in Figures 5.5b and 5.5n, where because of the significant change within the background the border area is labelled as foreground. The appearance of the input image after eliminating the noise in the foreground is shown in the third column. Figures 5.5h and 5.5n illustrate that the noise suppression technique does not damage the pattern area, the difference between the fore- and the background is preserved. On Figure 5.5a there is significant noise on the top area of the foreground, which is successfully suppressed on Figure 5.5c. In the next step non-local means are calculated based on the differences of the gradients. Focusing on the background of the images in the fourth column, nearly homogen background is created without applying the noise mask on the images. However, on Figure 5.5j the biggest weakness of the algorithm is also visible. The non-local mean calculation eliminates the majority of pattern information in both top and bottom area of the original shoeprint. Combined with highly structured background noise, several mixed classes are generated where noise and pattern data occur simultaneously. The artifacts of this phenomenon are visible not only on the said shoeprint pattern area, but also in the middle of the image, where noise structures were enhanced appearing as bright lines similar to the remaining pattern edges. Figures 5.5e, 5.5k and 5.5q show the binarized images after applying the background mask. Because of the nearly homogen background achieved with non-local means the thresholding method is able to preserve the extracted lines. Masking the binarized image, falsely clustered edges, such as the ones in the middle of Figure 5.5j, are deleted easily. In the last column the final results are shown after the postprocessing step of eliminated small regions. Comparing Figure 5.5e to Figure 5.5f the latter is less cluttered, many small lines around the pattern contours were eliminated successfully while no relevant edges were deleted. The same effect is also visible on Figure 5.5r where the silhouette of the pattern is clearer since the majority of remaining line noise was eliminated. However on Figure 5.5l the pattern information is further damaged by the postprocessing step. Observing the right side of the bottom area it is clearly visible that several pattern contours were eliminated. The postprocessing step assumes that the shoeprint consists of bigger structures such as filled geometries on Figure 5.5a or broad edges on Figure 5.5m, thus the detailed edge pattern of Figure 5.5g is automatically considered as noise instead of relevant information.

The experimental results show that the semi automated algorithm provides more detailed results, Figure 5.3j, than the firstly introduced modified three-layered feature learning approach, Figure 3.4d. Furthermore it copes better with noise, Figure 5.5f, than the fully-automated method, Figure 4.4l. Despite the fact that the fully-automated algorithm achieves better results on detailed, fine lined shoeprints impressions, Figures 4.4h and 4.4h then the semi-automated one, Figure 5.5l, overall the latter has higher performance than the two previous ones because it is able to deal with low-quality samples, Figure 5.4e, as well. However, along the weakness of misinterpreting fine edges 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 noise. A such case is shown on Figure 5.6. The output is computed solely based on the gradient-based non-local means method and the postprocessing step. When the calculated 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 model about the clutter on the image is calculated. In such cases several enhancement steps are not exploited, thus the quality of the output is lower.

Based on the evaluations in the last three chapters the semi-automated approach proved to be the most powerful technique from the three proposed algorithms. To get a detailed overview about its performance, additional tests are conducted in the following chapter, and the results are further discussed.

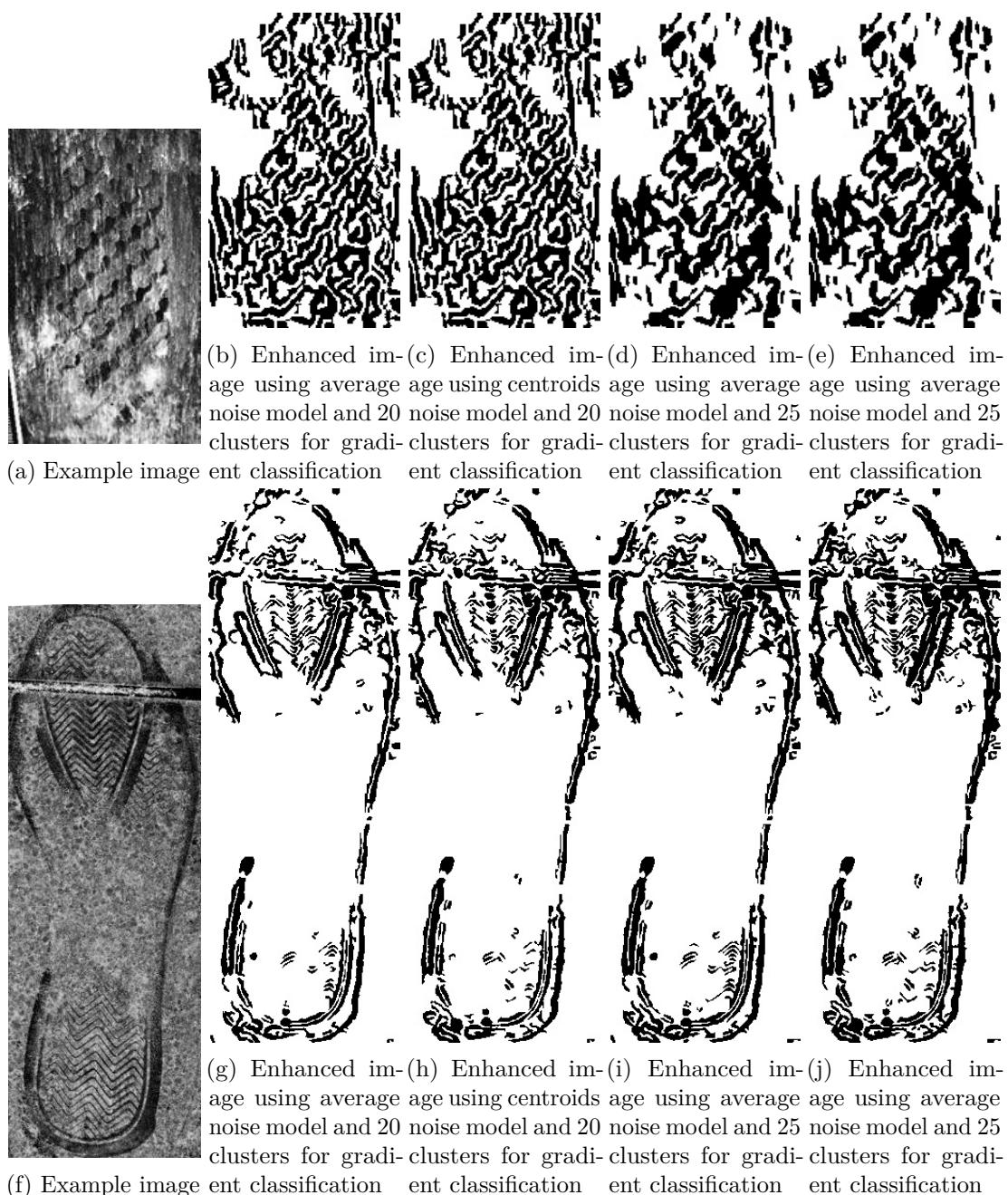


Figure 5.4: Example results of the semi-automated algorithm

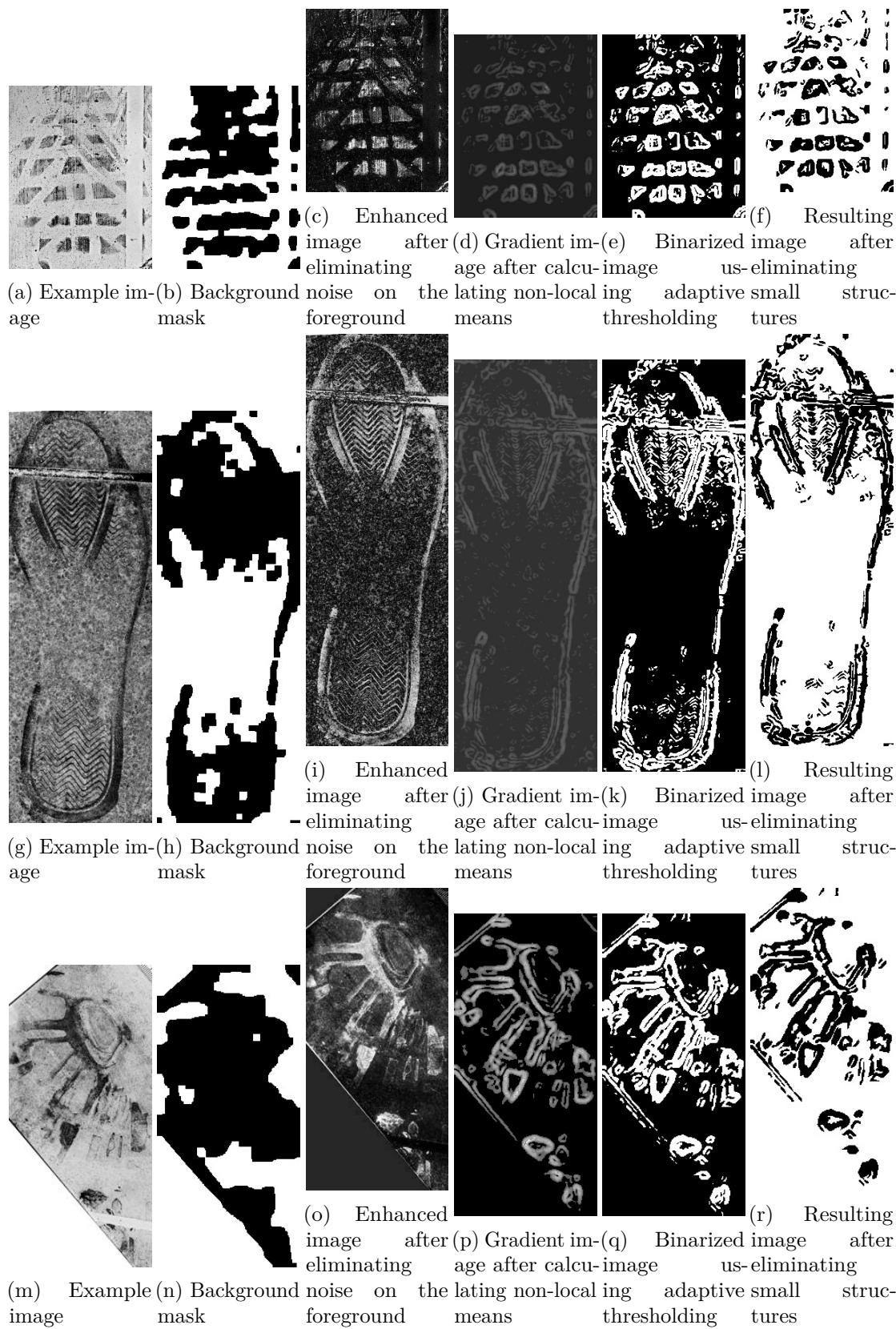
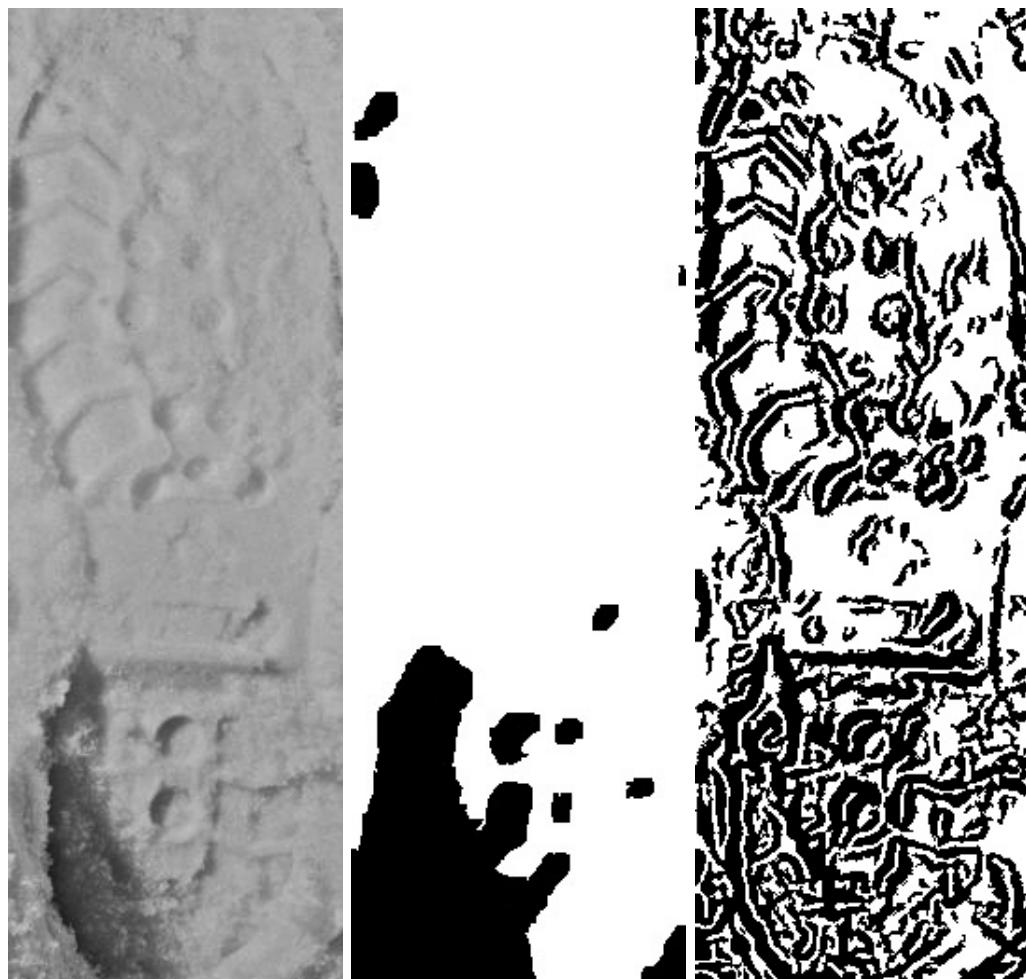


Figure 5.5: The whole pipeline of the semi-automated algorithm using 25 classes and centroids for noise elimination on pattern



(a) Example image from the (b) Calculated background (c) Output of the semi-
dataset mask automated algorithm

Figure 5.6: Output of the semi-automated algorithm, where no estimation about the background can be made

6

CHAPTER

Results and Discussion

In this chapter the experimental results of the semi-automated approach conducted on a minimal dataset from FID-300 are presented and discussed. The testing database consist of 20 images, the selected images and the output of the algorithm are shown in the Appendix 8. When selecting the testing dataset, no restrictions were made, low quality and heavily cluttered images are also part of the experimental database. Since the evaluation in the previous chapter showed that the semi-automated approach is not effective on samples with detailed, fine line structures, those images 12, 14, 19 are excluded while testing. Furthermore it was mentioned as well that in case of low intensity difference no background estimation is possible. For that reason images 13 and 15 are also disregarded while evaluation.

The visual improvement caused by the different algorithms were already discussed in the previous chapters. Now the analytical results of the semi-automated method are analyzed based on the performance of three basic feature descriptors. The goal of the algorithm is shoeprint enhancement and to increase the overall sample quality and not to develop a powerful descriptor or a matching technique. Hence for evaluation three basic descriptors were chosen, 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]. Although 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 evaluation is done by comparing the matching performance of those descriptors on the original samples and on the processed images. If the basic descriptors, without embedding in a more complex pipeline, achieve higher accuracy on the processed images than on the original ones, it implies successful enhancement. The experimental results already conducted on the field indicate [RBCP19] that using a complex approach increase

6. RESULTS AND DISCUSSION

matching accuracy, however when only one given method is used for evaluation it is not clear if the enhancement is optimized for that specific pipeline or 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.

For matching the extracted features two algorithms are used Brute-Force matching (BF) and Fast Library for Approximate Nearest Neighbors (FLANN). BF simply 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 point so it is faster than BF. The criteria for 'good match' is set to 0.8 while testing. For matching the shoeprint samples the Ground Truth images of every member of the minimal database were selected and stored, and it is examined how many 'good matches' were found between the samples and the Ground Truth images. At the end of the matching process a ranking among the Ground Truth images is made, 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 ranking of the actual Ground Truth of the given sample 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 belonging Ground Truth image of every member of the dataset had the most 'good matches', thus the highest ranking, the final score is 1. Lower performance is indicated with higher score, correct matching results in lower score.

To create a baseline the original samples were matched to the Ground Truth images 6.1. For comparison all four possible outputs of the enhancing algorithm is presented in Tables 6.2, 6.3 and 6.4, 6.5. Table 6.6 gives an overview about the overall performance of the discussed pipelines, in the upper side of the rows the calculated average score of both matching techniques is shown in Tables 6.2, 6.3 and 6.4, 6.5 and 6.1. In the lower part of the rows the average score of both matching techniques is displayed. In the last column the average score of all descriptors using both matching techniques is shown. To understand the performance of the proposed enhancing method this table is analyzed in the following.

Focusing on the first (F-M) and third (SURF) column of the table, the matching of the enhanced shoeprints always outperforms the original samples regardless the pipeline settings, except one case when FLANN matching is used on the output of 20 classes and average noise model. However, considering the second column (SIFT) of the table that is not the case, and the enhanced images are not able to achieve the performance of the original images. Wang et al. [WZWS17] state that SIFT is less effective on binarized images, since it seeks less information about the structure of the image than in the original version. On the other hand they also mention that SURF suffers from the same limitation, which is not the case in our experiments. This contradiction is explainable by examining the working pipeline of SURF, it uses a blob-detector to determine feature points which is suitable for shoeprint identification. The majority of shoeprints contain closed structures such as circles or polygons that are easily recognized by a blob-detector. Thus even though working on a binary image is not ideal for SIFT nor for SURF, the

latter is able to overcome its weakness by taking advantage of the specific features of a shoe sole pattern.

	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
7	6	6	2	8	8	9
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	5.8666666667	5.8666666667	5.3333333333	4.6666666667	6.6	5.8666666667

Table 6.1: 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	8	4	9
2	1	1	11	10	1	2
3	3	4	1	1	6	7
4	2	3	8	10	7	8
5	1	2	5	3	4	4
6	5	9	2	3	6	8
7	13	12	13	13	8	8
8	2	2	12	11	8	8
9	8	11	13	6	4	4
10	4	9	7	3	1	1
11	2	2	6	8	12	11
16	5	6	1	1	4	6
17	2	3	7	10	1	2
18	6	7	1	1	7	2
20	5	2	4	6	3	2
Score	4	4.9333333333	6.2	6.2666666667	5.0666666667	5.4666666667

Table 6.2: Results using 25 classes and the average noise model

6. RESULTS AND DISCUSSION

	F-M		SIFT		SURF	
	BF	FLANN	BF	FLANN	BF	FLANN
1	2	1	5	6	1	1
2	1	1	2	3	1	2
3	6	9	1	1	2	5
4	1	1	8	10	4	2
5	5	7	1	2	2	3
6	4	6	1	1	3	3
7	13	11	13	13	8	8
8	3	2	12	11	8	8
9	8	11	13	6	4	4
10	7	6	10	8	4	2
11	4	4	9	8	10	5
16	4	4	2	8	6	7
17	6	5	5	12	1	2
18	7	7	3	3	1	1
20	2	1	12	9	1	1
Score	4.866666667	5.066666667	6.466666667	6.733333333	3.733333333	3.6

Table 6.3: Results using 25 classes and the centroids noise model

	F-M		SIFT		SURF	
	BF	FLANN	BF	FLANN	BF	FLANN
1	1	1	5	8	3	6
2	1	1	13	13	3	5
3	1	2	1	1	5	6
4	3	5	8	9	1	3
5	3	7	3	3	1	2
6	10	13	2	3	7	7
7	13	12	13	13	8	8
8	7	9	13	12	4	4
9	8	11	13	6	4	4
10	2	5	7	5	1	1
11	3	2	8	8	10	7
16	2	2	9	6	5	5
17	3	2	9	3	1	3
18	10	9	1	3	10	10
20	4	4	4	6	2	2
Score	4.733333333	5.666666667	7.266666667	6.6	4.333333333	4.866666667

Table 6.4: Results using 20 classes and the average noise model

	F-M		SIFT		SURF	
	BF	FLANN	BF	FLANN	BF	FLANN
1	2	3	6	7	1	1
2	1	1	3	4	3	3
3	1	3	2	3	4	5
4	1	1	11	9	5	5
5	3	6	7	5	7	8
6	8	8	3	3	1	2
7	13	12	13	13	8	8
8	7	9	13	12	4	4
9	8	10	13	6	4	4
10	7	6	12	9	1	1
11	1	1	6	8	6	7
16	4	2	1	3	9	9
17	5	5	9	9	1	2
18	5	3	4	2	8	3
20	6	6	8	10	2	2
Score	4.8	5.066666667	7.4	6.866666667	4.266666667	4.266666667

Table 6.5: Results using 20 classes and the centroids noise model

Examining the first row of Table 6.6, the best score the original shoeprints achieved is 5.24 (BF matching of Fourier-Mellin features), whereas the worst accuracy is 7.80 (BF matching of SURF features). The best accuracy of the baseline is outperformed several times, the Fourier-Mellin features achieve better accuracy regardless the matching technique in every case except when working with 20 classes and average model noise model, where the scores are minimally lower, 5.44 for BF and 6.57 for FLANN. SURF features achieve the overall best scores for matching the enhanced images. Matching the results produced by using 20 classes and centroids as noise model the score of 4.27 is noted for both BF and FLANN. But in case of output generated by 25 classes and centroids as noise models a better value of 3.73 and 3.60 is achieved. Furthermore, the worst performance of the baseline algorithm is 7.80, which is outperformed in every test case. The worst score noted for the enhanced image matching is 7.40 which was provided by SIFT BF matching of the samples processed by the setting with 20 classes and using centroid models.

Considering the overall accuracy (last column) of the different settings, the enhanced matching always provide better results than the baseline. Comparing the scores of the enhanced samples to each other the pipeline using only 20 classes and the averaged noise model has the lowest performance. Using 20 classes but the centroids for noise representation results in score improvement of above 0.5 points achieving 5.44. The best scores were reached by the images where 25 classes were used earning 5.32 and 5.08 points, for average and centroid noise models respectively. These values show that the detailed algorithms lead to higher quality enhancement. Using 25 classes outperforms the case when the differences of gradients are grouped into 20 classes. Building a more

	F-M		SIFT		SURF		Average Results
	BF	FLANN	BF	FLANN	BF	FLANN	
Baseline	5.655555556	5.238888889	6.027777778	5.472222222	7.8	7.072222222	6.211111111
25 Average	4	4.933333333	6.2	6.266666667	5.066666667	5.466666667	5.322222222
25 Centroids	4.866666667	5.066666667	6.466666667	6.733333333	3.733333333	3.666666667	5.077777778
20 Average	5.43912037	5.567283951	6.634259259	5.904320988	6.630787037	6.111728395	6.047916667
20 Centroids	4.8	5.066666667	7.4	6.866666667	4.266666667	4.266666667	5.444444444

Table 6.6: Overview about the performance of the analyzed pipelines

versatile model for noise representation using the centroids of the k-means algorithm results in better scores than the images enhanced with the average noise model have. The fact that both settings using 25 classes outperformed the ones with less classes implies that detailed clustering of extracted edges is more important than the choice of noise representation in noise suppression step.

The experimental results presented in this chapter show that the proposed image enhancement algorithm supports accuracy improvement. The performance of the algorithm was tested on a versatile subset of images selected from the FID-300 dataset. However it has to be noted that the proposed method needs further improvement to be able to handle all possible samples of the FID-300 dataset. For known limitations of the algorithm samples with detailed, fine line structures and low contrast between fore- and background were excluded from testing. The matching based on the Fourier-Mellin features is the most robust one, it outperforms the baseline algorithm never having worse scores than 5.57 and the difference between the best- and worst case scenarios is only 1.57. SURF provides the best score in given cases but it also reaches above 6.00 scores multiple times, thus the difference between the best and worst accuracy is above 3.6. The robustness and the high-quality of Fourier-Mellin features is promising, since many current shoeprint matching algorithms already use it for feature description [GBCN08], [RLL⁺17], [WWZ19] or are based on other frequency domain descriptors [AH08], [WSYZ14], [KS17].

7

CHAPTER

Future Work

Forensic shoeprint identification is an extensively studied research area [RBCP19], nevertheless there are several possibilities for improvement. One problem of the field is the lack of uniform datasets, in many cases narrow datasets with low variability and high quality samples were used [RBCP19]. Additionally the testing data is not necessarily made public, so no fast comparison between existing approaches is possible.

To establish an overview and structure our knowledge in this field three prototypical approaches were introduced in this thesis. Being prototypical approaches there are several ways for improvement. The first, most obvious direction is the combination of the three techniques. Even though all three methods have different pipeline applying all of them simultaneously on the input data is not impossible. Additionally the discussed advantage of one technique can suppress the limitation of the other two. The semi-automated pipeline is currently the most powerful approach from the introduced one. However being semi-automated and not being able to cope with fine line structures is nevertheless a significant limitation. On the other hand letting the modified three-layered feature learning algorithm to determine the approximate outline of the pattern, and determining the exact background based on the semi-automated method replaces the now necessary user input at the beginning of the application. Additionally, the introduced fully automated method failed on the majority of samples, it only provided reasonable output if the input had high quality and detailed line structure. That is however exactly the case that the semi-automated pipeline is not able to handle. Enhancing any kind of input data with the semi-automated pipeline, calculating the background mask and applying the non-local means for edge detection, and finishing the calculation with the fully automated approach in case of fine-lined shoeprint pattern solves this kind of limitation. The preprocessing done with the semi-automated method increases the quality that the fully-automated noise elimination approach needs, and the latter one is able to identify the detailed pattern which the semi-automated technique eliminates immediately.

7. FUTURE WORK

Additionally there 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 judgements of the experts the neighborhood and the coefficient matrix of the given sample 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 relatively easy.

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

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 impressions. 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. Li et al. [LW15] propose a subdivision method where four regions of a shoeprint are distinguished, these are cap, sole, arch and heel and processed according to their position. 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 images.

Lastly Randomly Acquired Features (RAC) of shoeprints were already mentioned before but were not considered during feature extraction for two reasons. First, in this thesis prototypical applications are shown and the main goal is to get a basic idea about the effectiveness of the published approaches. Second, as Rida et al. [RBCP19] states in many dataset only a limited amount of samples is available for one shoeprint, and that is also the case in the FID-300 database. There are no more than six images from the same shoeprint in exceptional cases, but the majority of the dataset consist of single samples. But assuming that the demand for wider dataset increases with the active research, considering RAC features of the input provides valuable additional information about the given sample. Shor et al. [SWT⁺18] introduce a technique to distinguish between the impressions of the same shoe based on the obtained RAC features. Damary et al. [DMW⁺18] published a study about the estimated properties of a RAC based on its location of the shoeprint. In this way if a RAC is identified assumptions about the neighboring pattern are made. 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 two connected problems which have to be solved in the future. The first one is the already mentioned non-existing uniform database. 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 to other methods. Furthermore the real-life performance and usefulness of the published methods is unknown because they are often tested on synthetic and on handcrafted data. The second problem is strongly connected to the absence of uniform database and that is the lack of overview and structured knowledge in the field. 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 [?], on their testing datasets. These publications imply that the problem is already solved, and there is no further room for improvement which is not the case. Until no central, uniform, high-quality and wide dataset is available no research is done efficiently, it is not possible to take advantage of the already published approaches and it is difficult to learn from the mistakes previously done.

CHAPTER

8

Conclusion

Forensic shoeprint image retrieval and identification is an extensively researched topic in the last 20 years. A great amount of publications is available and many methods were introduced for best shoeprint pattern recognition. However, no structured knowledge is available on this field, thus results are uncertain, the solved and open problems are unclear and the publications are incoherent. Because of the absence of universal, high-quality dataset the algorithms are not comparable, the limitations and strengths of the approaches are unclear and it is ambiguous if they are applicable in real-life use-cases as well.

In 2014 a new database, called FID-300, was introduced by Kortylewski et al. [KAV14] which is a step in the right direction, containing 300 shoeprint impressions collected on real-life crime scenes. In this thesis an overview and introduction to possible ways for real-life shoeprint enhancement was given. Based on the available shoeprint recognition literature and research done in related fields such as fingerprint and tattoo identification and natural image enhancement, three prototypical algorithms were presented to increase the quality of the real-life samples. The first one is based on the idea of feature learning, the second one utilizes a traditional image enhancement pipeline and the idea between the third one is the calculation of a noise representation model and gradient clustering for relevant edge detection. The evaluation shows that many preprocessing techniques popular in the current forensic research are not robust and powerful enough to process real-life shoeprint images. All three algorithms are able to achieve reasonable results on specific samples, however many images with given properties, such as low quality images, overly cluttered shoeprints low-contrast pictures are crucial limitations of all of them. Based on the visual results of all three algorithms the last one is the most robust one. The analytical testing conducted on basic image descriptors already used in forensic image identification shows that the enhanced images achieve better matching results than the original ones in almost every cases. With further development and potential combination of the proposed method a powerful and versatile preprocessing pipeline can

8. CONCLUSION

be built which is able to enhance real-life shoeprint impressions and thus increase the matching accuracy.

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.
- [AR08] Scott T Acton and Adam Rossi. Matching and retrieval of tattoo images: Active contour cbir and glocal image features. In *2008 IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 21–24. IEEE, 2008.
- [Ath11] Maythapolnun Athimethphat. A review on global binarization algorithms for degraded document images. 2011.
- [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.
- [CG10] Michael Crosier and Lewis D Griffin. Using basic image features for texture classification. *International journal of computer vision*, 88(3):447–460, 2010.
- [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.
- [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.

- [DK13] Pruegsa Duangphasuk and Werasak Kurutach. Tattoo skin detection and segmentation using image negative method. In *2013 13th International Symposium on Communications and Information Technologies (ISCIT)*, pages 354–359. IEEE, 2013.
- [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.
- [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.
- [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.
- [GZL15] 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 decopmosition 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.
- [KLYD16] Joonsoo Kim, He Li, Jiaju Yue, and Edward J Delp. Tattoo image retrieval for region of interest. In *2016 IEEE Symposium on Technologies for Homeland Security (HST)*, pages 1–6. IEEE, 2016.
- [KPY⁺15] Joonsoo Kim, Albert Parra, Jiaju Yue, He Li, and Edward J Delp. Robust local and global shape context for tattoo image matching. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 2194–2198. IEEE, 2015.
- [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.
- [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.
- [NBC⁺09] Omar Nibouche, Ahmed Bouridane, Danny Crookes, Mourad Gueham, and Moussadek Laadjel. Rotation invariant matching of partial shoeprints. 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.
- [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.
- [SWY⁺13] Guangqi Shao, Yanping Wu, A Yong, Xiao Liu, and Tiande Guo. Finger-print compression based on sparse representation. *IEEE transactions on image processing*, 23(2):489–501, 2013.
- [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.
- [WZY⁺18] Yu Wang, Na Zhang, Huaixin Yan, Min Zuo, and Cuiling Liu. Using local edge pattern descriptors for edge detection. *International Journal of Pattern Recognition and Artificial Intelligence*, 32(03):1850006, 2018.
- [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.
- [YWZ16] Huaixin Yan, Yu Wang, and Na Zhang. Edge detection using varied local edge pattern descriptor. In *2016 International Conference on Virtual Reality and Visualization (ICVRV)*, pages 114–118. IEEE, 2016.

- [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 shoeprint 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 A

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

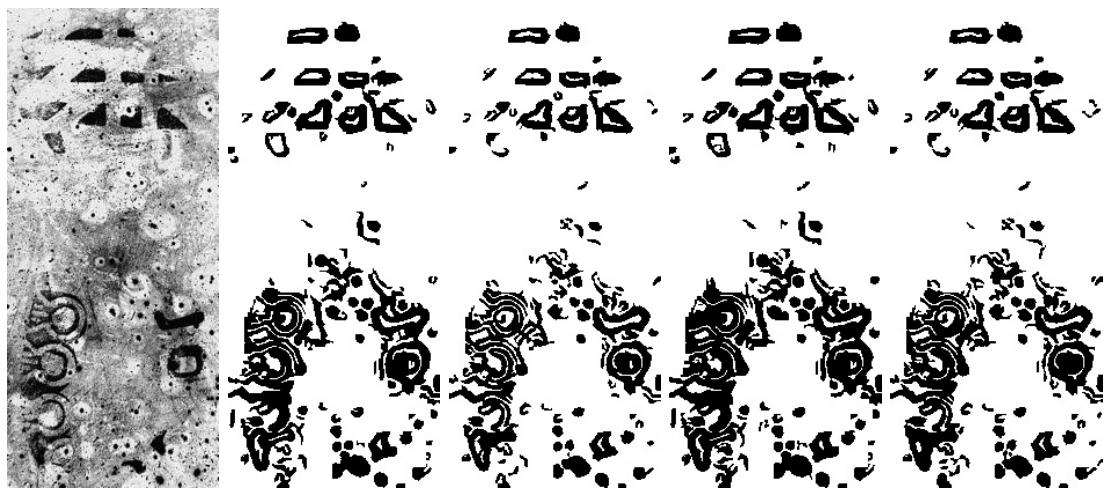


Figure 2

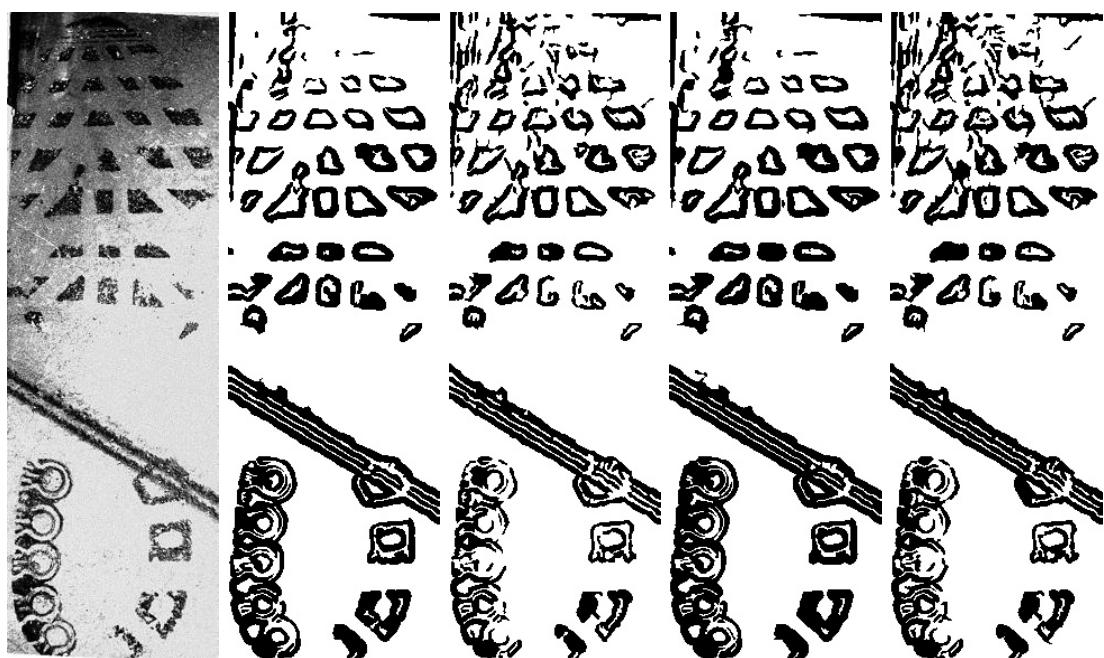


Figure 3



Figure 4



Figure 5



Figure 6

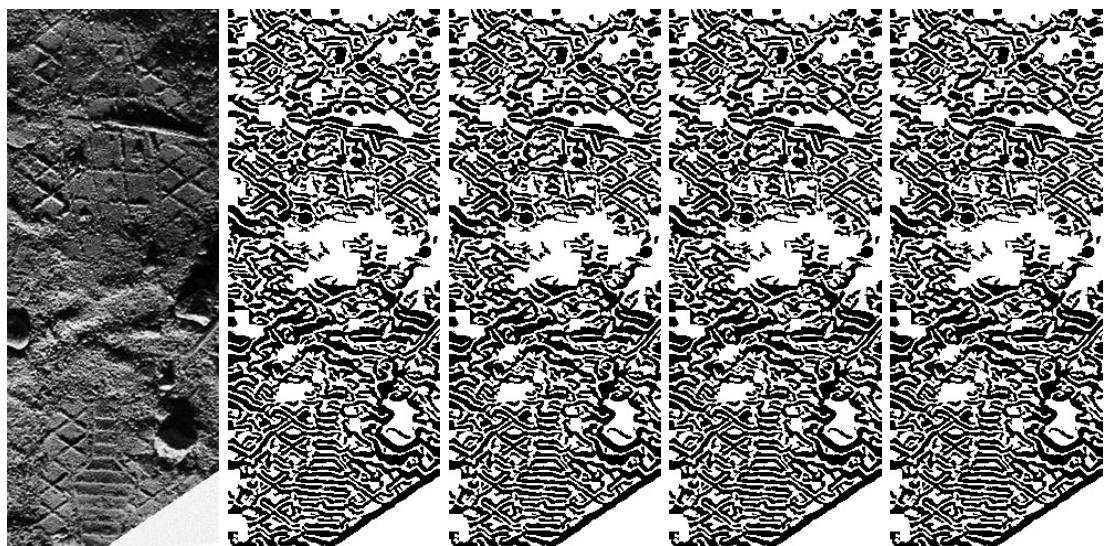


Figure 7



Figure 8

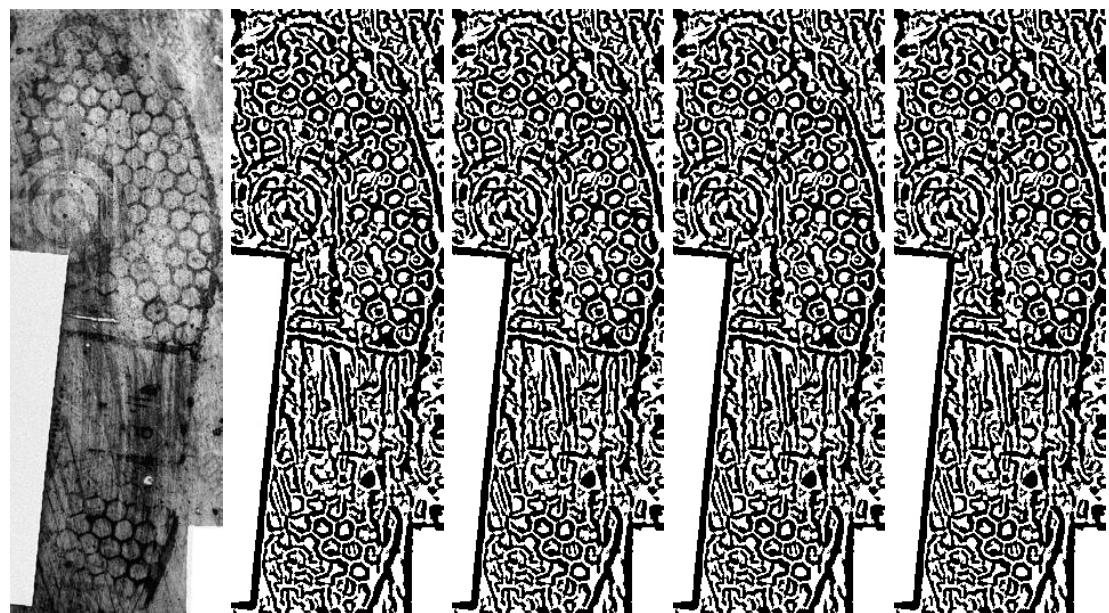


Figure 9

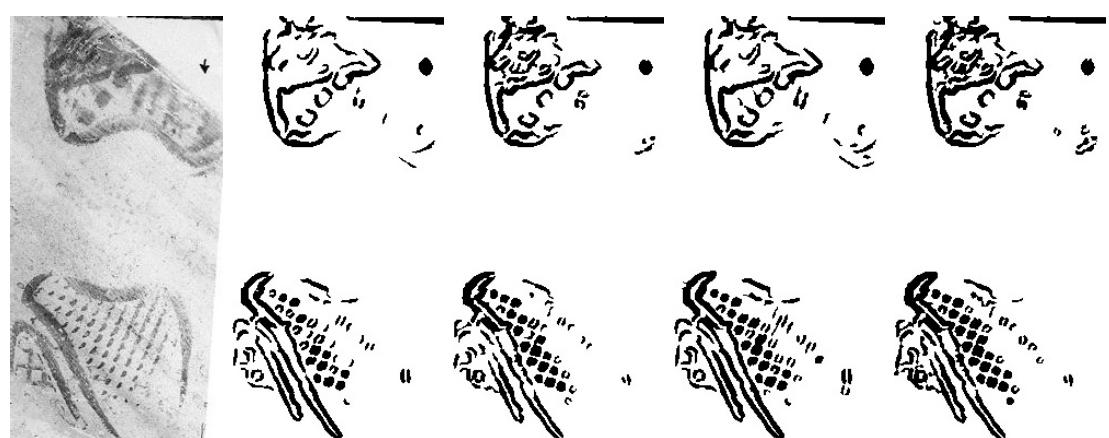


Figure 10



Figure 11

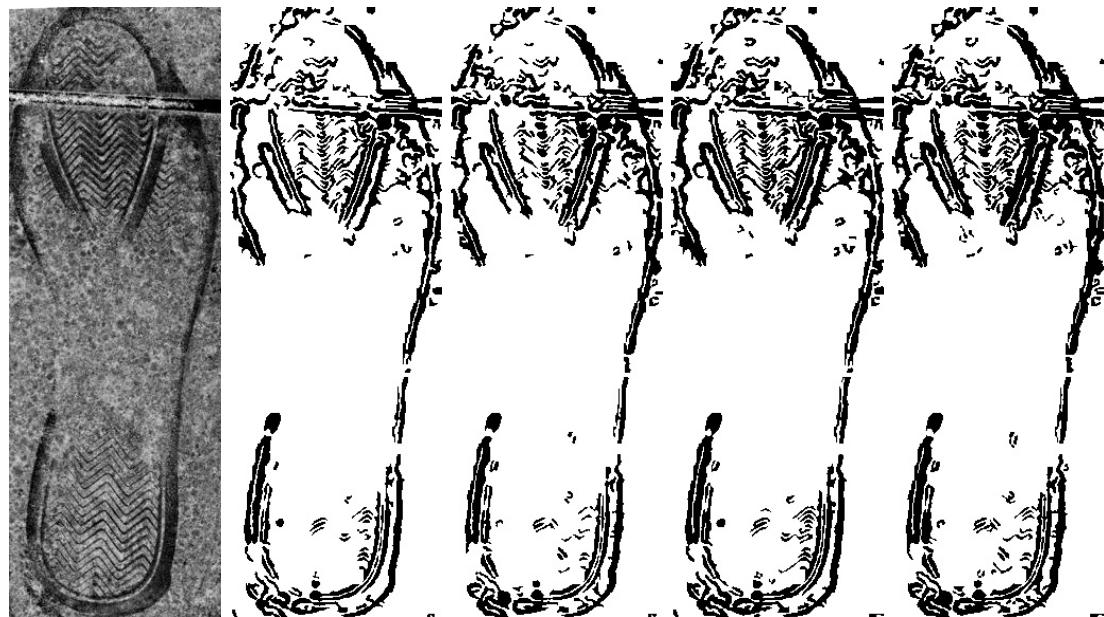


Figure 12

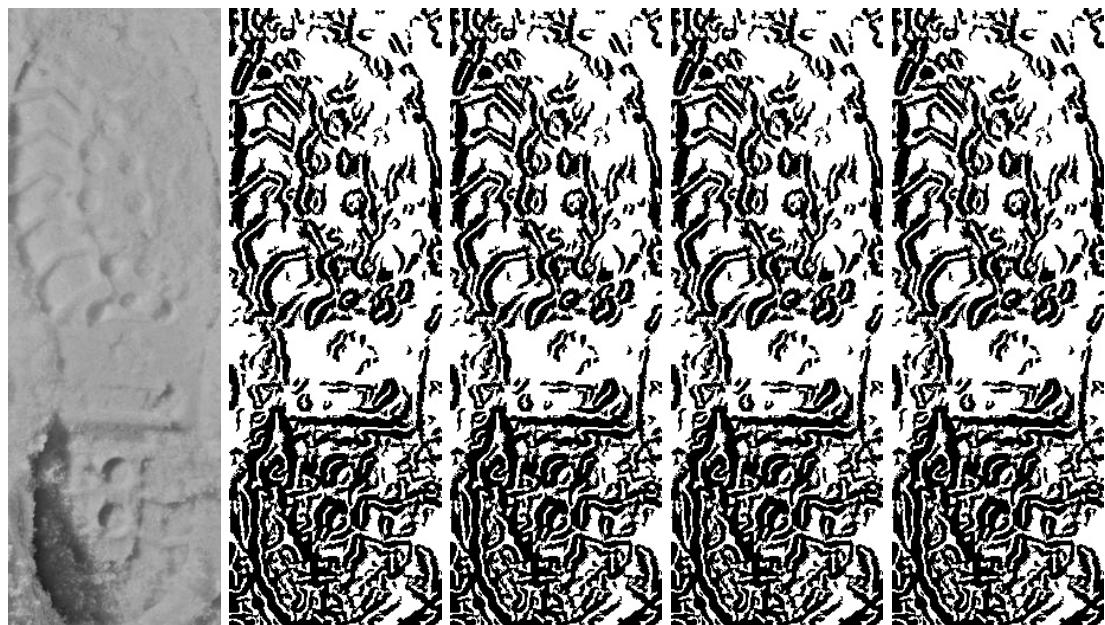


Figure 13



Figure 14



Figure 15

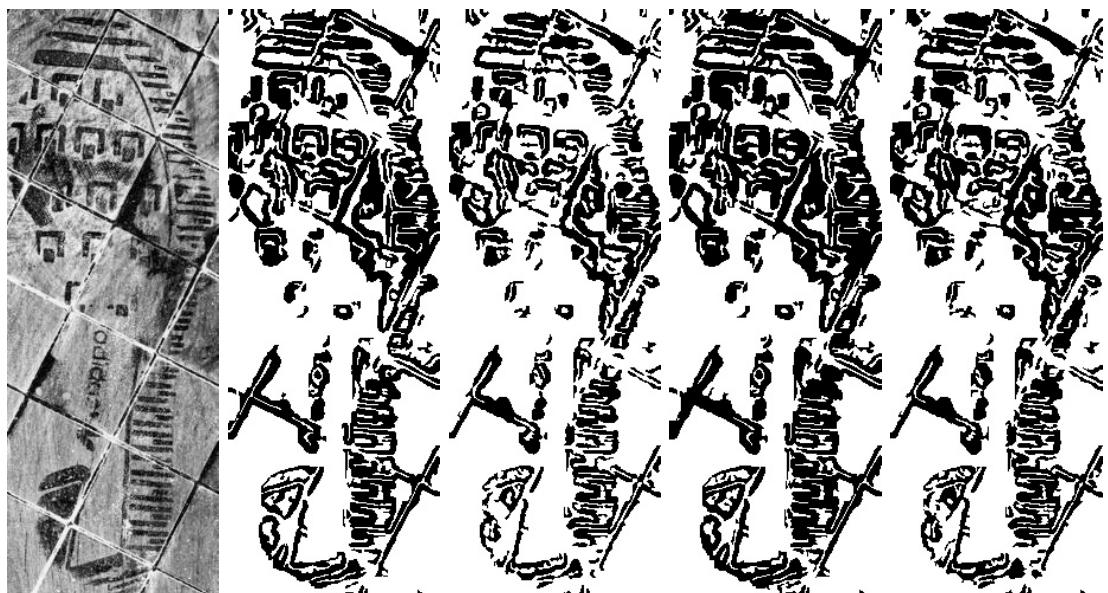


Figure 16

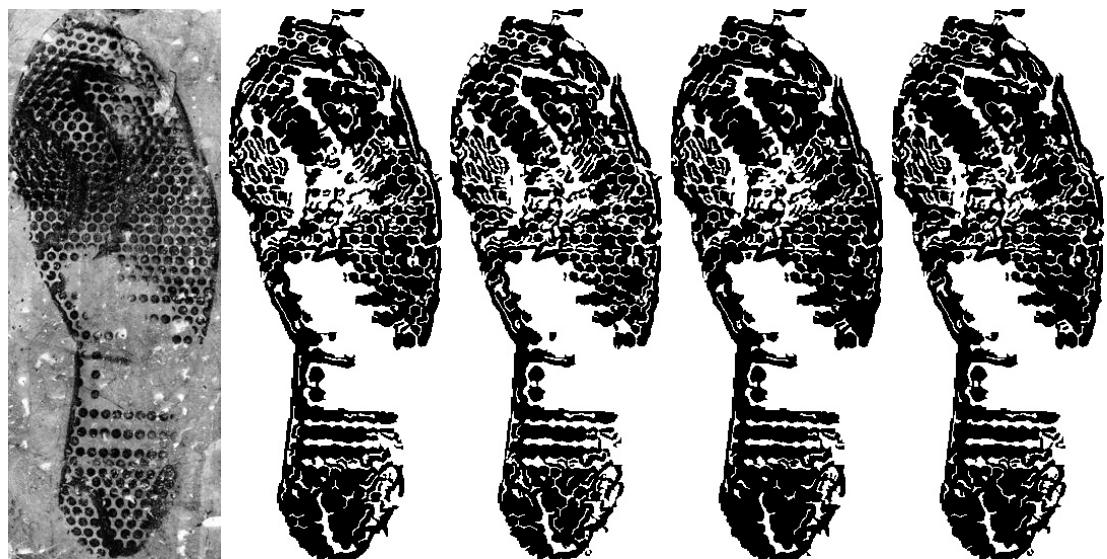


Figure 17



Figure 18

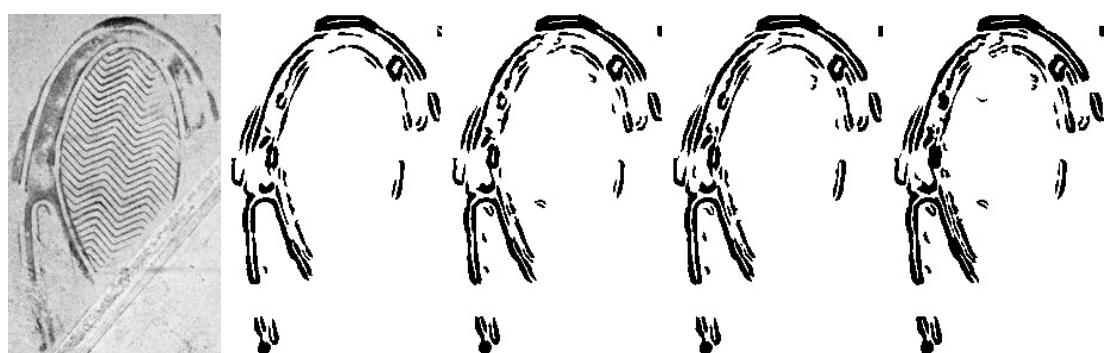


Figure 19

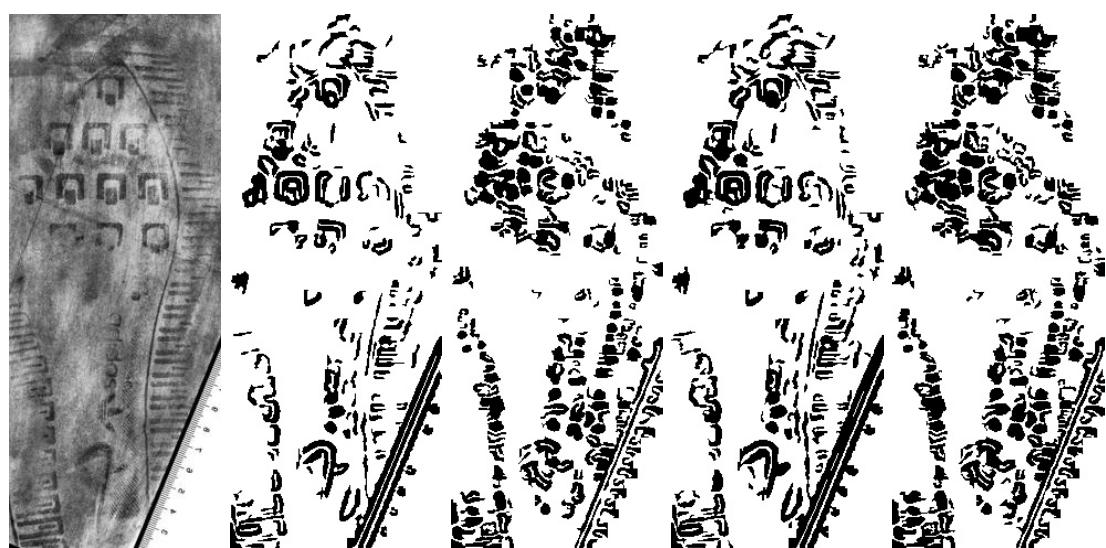


Figure 20