Computational Analysis of Writing Style in Digitised Manuscripts

Dissertation with the aim of achieving a doctoral degree at the Faculty of Mathematics, Informatics and Natural Sciences
Department of Informatics of Universität Hamburg

Submitted by Hussein Mohammed

Work done under the supervision of:
Prof. Dr. -Ing. H. Siegfried Stiehl and Dr. -Ing. Volker Märgner

2018 in Hamburg

List of Evaluators:

1. Prof. Dr. -Ing. H. Siegfried Stiehl
2. Dr. -Ing. Volker Märgner
3. Prof. Nicole Vincent

Date of Oral Defence: 22.03.2019

Declaration on Oath

Computational Analysis of Writing Style in Digitised Manuscripts

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hamburg, in:

Signature:

____

(Hussein Adnan Mohammed)

Eidesstattliche Versicherung

Computational Analysis of Writing Style in Digitised Manuscripts

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Hamburg, den:

Unterschrift:

____

(Hussein Adnan Mohammed)

Abstract

Computational Analysis of Writing Style in Digitised Manuscripts

The goal of this dissertation is to develop a novel computational method capable of analysing the handwriting styles in digitised manuscripts in order to provide supporting information for the task of handwriting style identification.

We collected and analysed the requirements from selected sub-projects within the Sonderforschungsbereich SFB 950 "Manuscript Cultures in Asia, Africa and Europe" regarding the problem of handwriting style identification. Then we analysed the state-of-the-art methods to find a starting point for the development of a novel method in order to fulfil these requirements.

In order to analyse the handwriting styles in digitised manuscripts, we developed a classifier for offline, text-independent, and segmentation-free writer identification based on the Local Naïve Bayes Nearest-Neighbour (Local NBNN) classifier. Due to scarce data, our proposed method is a learning-free approach, which takes into consideration the particularity of handwriting patterns by adding a constraint to prevent the matching of irrelevant keypoints. Furthermore, a normalisation factor is proposed to cope with the prevalent problem of unbalanced data in our case of writing style analysis of digitised manuscripts.

The performance of our proposed method has been evaluated using several public datasets, both contemporary and historical, of different writing

systems including musical scores. State-of-the-art results were obtained in all experiments with a fixed parameter set. This evaluation helps to measure the discriminative power of our proposed method w.r.t. different handwriting styles in the datasets. Furthermore, some of these standard datasets offer handwriting styles from a large number of writers and/or in many different writing systems and script types.

Furthermore, the performance of the proposed method is analysed w.r.t. typical degradation found in digitised manuscripts using samples relevant to the data used by the selected sub-projects within the SFB. Historical manuscripts from a public dataset have been used in this analysis and have been selected jointly with scholars from Humanities within the SFB 950. The selection of degradation types was based on their prevalence in digitised manuscripts and their direct influence on parameter selection of the proposed method.

Finally, an easy-to-use implementation of the proposed method has been realised as a software tool with a user-friendly GUI (graphical user interface). It presents the results in an intuitive way so that it can be easily used by scholars from manuscript research in Humanities without the aid of experts from computer science. Our software tool implementation has been used by scholars from Humanities within the SFB 950 for their research yielding very satisfying results. Several experiments and tests have been carried out in order to address their actual research problems with regards to handwritings in digitised manuscripts.

Acknowledgements

I would like to express my gratitude to my supervisors Prof. H. Siegfried Stiehl and Dr. Volker Märgner for the useful comments and remarks. In addition, a thank to Ms. Kirsten Tatam for her efforts in proof-reading this dissertation. Furthermore, I would like to thank the people in the Centre for the Study of Manuscript Cultures (CSMC) for the supportive environment they provided. Also, I would like to thank the scholars from the SFB 950 sub-projects, namely Prof. Philippe Depreux, Till Hennings, MA, and Prof. Tilman Seidensticker for their feedback and collaboration.

Last but not least, I would like to thank my wife *Duaa* for her constant support and encouragement.

This work has been funded by the German Research Foundation (DFG) of the Sonderforschungsbereich (SFB 950) within the scope of the Centre for the Study of Manuscript Cultures (CSMC) at Universität Hamburg.

To the one whom I want to spend every moment of my life with, my princess

Duaa

Contents

1 Introduction **1**

1.1 Concepts and Terminologies 2

1.2 Motivation 6

1.3 Problem Statement 8

1.4 Methodology and Workflow 11

1.5 Requirements from Sub-Projects in SFB 950 14

1.6 Challenges in Computational Analysis of Digitised Manuscripts 18

1.6.1 Standard and Public Datasets 18

1.6.2 Degradation in Digitised Manuscripts 19

1.7 Contributions of the Dissertation 20

1.8 Organisation of the Dissertation 20

2 Related Work **22**

2.1 Features Used for Writer Identification 24

2.1.1 Forensic Examiners' Features 25

2.1.2 Texture-Based Features 26

2.1.3 Allographic Features 31

2.1.4 Mixture of Texture-Based and Allographic Features . 33
2.1.5 Auto-Derived Features 34
2.2 Classifiers 35
2.2.1 Learning-Based Classifiers 36
2.2.2 Learning-Free Classifiers 37
2.3 Conclusion 42

3 Proposed Method 46

3.1 Keypoints Detection and Feature Extraction 47
3.1.1 SIFT Keypoints Detection 47
3.1.2 FAST Keypoints Detection 54
3.1.3 SIFT Descriptor 58
3.2 Matching 60
3.3 Orientation Threshold 66
3.4 Class Distance Normalisation 68
3.5 Conclusion 71

4 Performance Evaluation and Experimental Results 72

4.1 Standard and Public Datasets 73
4.2 Evaluation Criteria 76
4.3 Experimental Results 78
4.3.1 Contemporary Datasets 79
4.3.2 Historical Dataset 87
4.4 Conclusion 90

5 Performance Analysis w.r.t. Degradation Types in Digitised Manuscripts 91

5.1 Data Selection for the Analysis 92

5.2 St. Gall Sub-Set 93

5.3 Analysis Procedure 95

5.4 The Selected Degradation Types 97

5.4.1 Resolution 97

5.4.2 Contrast 100

5.4.3 Rotation 104

5.4.4 Irrelevant Information 108

5.5 Conclusion 111

6 Implementation as a Software Tool 113

6.1 Design Criteria 114

6.2 Handwriting Analysis Tool v2.0 (HAT-2) 115

6.2.1 Required Directory Structure 117

6.2.2 Parameter Settings 119

6.2.3 Results Presentation 121

6.2.4 Usage Procedure 124

6.2.5 Technical Considerations 125

6.2.6 Additional Considerations 125

6.3 Application to Research Questions of Scholars within the SFB 126

6.3.1 Use Case: Sub-Project C08 128

6.3.2 Use Case: Sub-Project B05 132
6.4 HAT-2 within the Community of Manuscript Research in Humanities 137
6.5 Conclusion 138

7 Conclusions and Future Work 139
7.1 Conclusions 139
7.2 Future Work 143

Chapter 1

Introduction

Based on the work of the DFG Research Group 963 “Manuscript Cultures in Asia and Afrika” (2008-2011) [1], the Centre for the Studies of Manuscript Cultures (CSMC) at Universität Hamburg is engaged in a fundamental research under the Sonderforschungsbereich (SFB 950) “Manuscript Cultures in Asia, Africa and Europe” [2], investigating from both a historical and comparative perspective, based on material artefacts, the empirical diversity of manuscript cultures.

As a part of the scientific services in the SFB, the Z03 Scientific Service Project “Image Processing Methods for Determining Visual Manuscript and Character Features” [3] aims to provide computer vision tools to various sub-projects of the SFB 950. Image processing methods are to be developed for determining the visual features in historical manuscripts. Furthermore, a research on the computational analysis of writing styles in digital manuscripts needs to be carried out.

The goal of this dissertation is to develop and experimentally evaluate a novel classification method and implement a software tool in order to tackle the problem of writing style analysis in digitised manuscripts and to fulfil the requirements of selected sub-projects of the SFB 950 at the CSMC. The developed method is required to cope with the problems of the lack of sufficiently large training datasets of known scribes and of the prevalence of unbalanced data in classes of scribes or writing styles. A dataset is considered as unbalanced when at least one class is represented by only a small number of samples (here, manuscripts of a particular known scribe). This work is part of the Scientific Service Project Z03 [3] of the SFB 950 [2].

1.1 Concepts and Terminologies

We use the term *writer* to refer to the person who generates the handwriting instances (samples). Scholars from manuscript research in Humanities may use different terms such as *scribe*; therefore, we use the term *scribe* in addition to the term *writer* when the context is related to a research problem within the Humanities. These two terms are used interchangeably in this dissertation.

The basic assumption behind all writer identification methods is that handwriting samples produced by a given writer have common patterns in the sense of visual features and the similarity between these handwriting samples is higher than the similarity to any handwriting sample produced by another writer; except for the case of forgery. This assumption holds under various conditions which can cause variations in the handwriting pattern of a given writer, such as ageing, physical conditions, and the context of the handwriting itself (e.g. formal script, personal letter).

In this dissertation, we refer to the intrinsic characteristics such as visual features shared by all the handwriting instances (samples) generated by the same writer as the "handwriting style" of that writer. Therefore, when we compare between handwritings of different writers to identify a specific writer, we compare their handwriting styles. We can use these characteristics to recognise and identify a given writer from her/his handwriting style. Therefore, we use the term *handwriting style identification* interchangeably with the term *writer identification* in the cases where the task is to identify a specific writer from her/his style.

A computational analysis of writing styles in digitised manuscripts consequently requires both the detection and description of visual features and the feature-based classification in order to achieve the task of handwriting style identification.

Different samples of handwritings may share similar features due to many reasons such as being produced during a certain period of time, by the same school of writing, or by the same person. These similar features set the handwriting samples apart from other handwriting samples as one group of unique handwriting style. We refer to the characteristic features of such group of handwriting samples as a *handwriting style* regardless of the reasons for this similarity.

The task of identifying a specific handwriting style (which belongs to a specific writer) is called writer identification in the scientific community of computational document analysis; see e.g. the International Conference on Document Analysis and Recognition (ICDAR) [4]. This task is typically formulated in publications and benchmarking competitions in two forms: *Writer identification* and *writer retrieval*. The task of writer identification is the process of assigning a writer with known reference handwriting samples to an unknown handwriting sample, while writer retrieval is the task of finding all relevant handwriting samples of a specific writer in a given dataset from e.g. a manuscript repository.

Identifying a handwriting style in a historical manuscript involves many aspects to be considered such as the historical background of the manuscript production, the philological evidence, and even the semantic meaning of the handwritten text. Therefore, the numerical measurements of e.g. visual features or manuscript similarity produced by computational methods should be only considered as a supporting information for scholars rather than as a decision. The numerical measurements produced by the proposed method in this dissertation are referred to as *similarity scores* or *similarity measurements*.

We refer to all the processes of extracting discriminative features from handwriting samples, comparing these features, and generating similarity measurements as handwriting style analysis. These measurements may be used by scholars as a supporting information for the task of handwriting style identification.

Computational methods for writer identification extract features relevant to the intrinsic characteristics of the handwriting style for a given writer. These features need to be as discriminative as possible such that they can be used for a classification-based identification. The same features can be used to discriminate between different schools of writing or even for the task of dating a given manuscript if the general style of handwriting changes through time for all writers within certain periods of time.

The focus of this dissertation is on analysing the handwriting samples of digitised manuscripts to generate similarity scores which can be used as a supporting information for the task of handwriting style identification.

One of the main goals for writing style analysis is to identify the writer of a given handwriting sample. In this dissertation, we focus on this goal. The identification of a given writer from his handwriting sample is based on a measure of confidence. This is true both for computational methods and for palaeographers who examine manuscripts visually. A confirmation of this concept is also provided by Hilton, the sixth president of the American Society of Questioned Document Examiners, who stated that “Any conclusion of identification derives from statistical inference, and is an expression of probability having an arithmetic value somewhere between 0 and 1.” [5].

Therefore, the computational method we develop in this dissertation does not identify writers by “yes” or “no” results, but rather it analyses handwritings by providing a measure of confidence so that the scholars from Humanities may use the results as supporting information for answering their research questions related to writer/scribe identification.

The computational method we propose in this dissertation analyses different handwriting styles and measures the similarities between them regardless of whether they belong to the same writer/scribe or to different writers/scribes. Therefore, we assume that it can also be used to discriminate between different schools of writing or even for the task of dating a given manuscript. However, no experiments are presented in this dissertation to substantiate this assumption, and the topic has to be left open, for time-out reasons, for future research.

1.2 Motivation

Analysing the style of handwriting is still a challenging task for e.g. law enforcement agencies and forensic documents analysis. Addressing this problem in digitised historical manuscripts poses additional challenges due to the nature of these documents, e.g. the different kinds of degradation. Most computational methods for the task of writer identification have been evaluated using contemporary datasets which consist of high quality images, namely high contrast, high resolution and low noise. Although some of these datasets are challenging in terms of the number of classes (writers) or even the amount of text provided per writer, they do not suffer from the typical degradation in digitised historical manuscripts. Furthermore, rather easy background-foreground separation is possible in most cases of contemporary documents. On the other hand, digitised historical manuscripts typically suffer from several kinds of degradation such as low resolution, low contrast, an arbitrary orientation of text, bleed through, textured background, varying background intensity, stains, water damage, etc. Moreover, the existence of information that is irrelevant to the handwriting of interest - such as layouts, illustration images and commentaries - can have a negative impact on computational methods.

Furthermore, the currently proposed methods for the task of writer identification are beyond the reach of the scholars from manuscript research in Humanities: Either because of the impracticality of the developed methods themselves for routine use, or because no easy-to-use implementations have been provided for them as non-experts in digital image processing and analysis.

In addition, the technological advances in terms of manuscript digitisation have given access to a large amount of digitised historical manuscripts and subsequently increased the demand for computational analysis for these manuscript collections. Evidently, manual analysis of such an increasing number of digitised manuscripts requires a significant amount of time and effort in order to provide answers to the research questions of manuscript scholars from Humanities.

Therefore, a novel computational method capable of tackling the problem of handwriting style analysis in digitised manuscripts can offer a great help and provide supporting information to scholars from Humanities. Furthermore, analysing the impact of manuscript-related degradation types is needed to better define the required quality of the images in order for the method to provide reliable results and to better understand the performance of the method. Finally, providing an easy-to-use implementation as a software tool with intuitively comprehensible GUI and results presentation can encourage, or even enable, the scholars from Humanities to integrate it into their research-driven workflow.

1.3 Problem Statement

The task of writer identification is one of the main goals for writing style analysis, and it can be defined as the process of assigning a writer with known reference handwriting samples to an unknown handwriting sample, while writer retrieval is the task of finding all relevant handwriting samples of a specific writer.

Both writer identification and retrieval methods try to explore the variations between different handwritings and use them as characteristics for writers' styles. The information on writer's general writing style can offer a valuable contribution to handwritten text recognition systems, e.g. Optical Character Recognition (OCR), by providing the ability to create writer-specific models to recognise the characters and words within a handwritten text of a certain writing style.

Generally speaking, writer identification is possible as far as the inter-variation in the handwriting of different writers exceeds the intra-variation within the handwriting of the same writer [6]. Nevertheless, similarities in the styling of writing elements (e.g. grams, bi-grams, letters, words or parts of any of these items), even when produced by different scribes, may indicate that they were written at a similar time and/or place; therefore, they can serve as a starting point for further research [7]. This is particularly important in the study of historical manuscripts in order to either date the handwriting or to identify a particular school of writing.

Writer identification methods can be classified into the following categorisations:

- **Online and offline:** Online writer identification methods use temporal, speed and acceleration data as additional features which are gathered while a text is written. This kind of data must be captured by input devices like touch screens and pen pads. Although such additional features can carry useful information like the sequence of strokes, this class of methods is not applicable when the text to be investigated is already written and/or the writer of the text is no more available. Therefore, offline writer identification methods are the class of choice when dealing with digitised historical manuscripts.
- **Text-dependent and text-independent:** A given method is considered to be text-dependent when it uses the transcription of the handwriting as an additional source of information. Such a transcription can be generated either manually or by using Optical Character Recognition (OCR) systems. In the case of digitised historical manuscripts, it

is rarely possible to have access to the transcription of a handwritten text; on the other hand, the state-of-the-art OCR systems are not reliable in regards to coping with all of the above-mentioned degradation types of the digitised historical manuscripts. Text-dependent methods are comparable to the work of forensic examiners and palaeographers with respect to comparing similar texts/words/characters from different handwriting styles. Hence text-independent methods are the first choice for our domain of application.

- **Segmentation-based and Segmentation-free:** Segmentation-based methods segment the page of a manuscript into lines and/or words, some methods even attempt to segment words into individual letters as well. Although this preprocessing step can enhance the performance of writer identification methods in many cases, even segmentation of contemporary documents can be both challenging and unreliable in many cases, whereas segmentation can be even impossible in our case of dealing with historical manuscripts for above reasons. Therefore, the proposed method in this dissertation is segmentation-free and deals with the digitised manuscripts as images with patterns formed by pixel intensities.

1.4 Methodology and Workflow

This dissertation is a part of the Scientific Service Project Z03 within the SFB 950. As a consequence, the developed method should be tailored to the requirements of scholars in Humanities, having participated in a requirements analysis, and should provide solutions for their research problems; therefore, we need to take into consideration the practicality and usability of the developed method [8].

The workflow as presented in this dissertation starts by collecting and analysing the requirements from selected sub-projects within the SFB regarding the problem of handwriting style identification; see Section 1.5. Based on these requirements, we analyse the state-of-the-art computational methods in Chapter 2 in order to find the best starting point for the development of a novel method. Then we present a novel computational method in Chapter 3 that is capable of fulfilling the requirements of the selected sub-projects within the SFB 950.

In order to compare the performance of our proposed method with the state-of-the-art methods for writer identification, we evaluate it on standard and public datasets of both contemporary and historical handwriting. The evaluation results are presented in Chapter 4. However, these standard datasets neither cover the typical degradation nor the unbalance and scarcity of handwriting samples of digitised manuscripts from the selected sub-projects within the SFB 950. Nevertheless, this evaluation helps in measuring the discriminative power of the proposed method w.r.t. different handwriting styles in the datasets. Furthermore, some of these standard datasets offer handwriting styles from a large number of writers and/or in many different writing systems and script types. Having a high performance for such varying datasets will demonstrate the generality as well as the scalability of the method.

The final step of this dissertation is to provide scholars from Humanities within the SFB with a software tool that can support them in the task of handwriting style identification by providing similarity scores as a supporting information for their research. Therefore, the performance of the proposed method is again evaluated and presented in Chapter 5 w.r.t. typical degradation types found in digitised historical manuscripts while using samples relevant to the manuscripts used by the selected sub-projects within the SFB. The selection of degradation types to be used in the analysis within this dissertation is based on their prevalence in digitised historical manuscripts and their direct influence on parameter selection of the proposed method.

Finally, we develop an easy-to-use implementation of the proposed method as a software tool which is presented in Chapter 6. This software tool is implemented with a user-friendly GUI (graphical user interface) and it presents the similarity scores in an intuitive way so that it can be used by the scholars without the aid of experts from computational document analysis.

The workflow of this dissertation is summarised as follows; see also Fig. 1.1:

- Collecting and analysing the requirements of selected SFB 950 sub-projects by Z03 team with regards to the problem of handwriting style identification.
- Developing a novel method to fulfil the collected requirements.
- Evaluating the developed method using standard datasets and comparing the performance with state-of-the-art methods.
- Analysing the developed method w.r.t. the influence of typical degradation types in digitised manuscripts.
- Implementing an easy-to-use software tool based on the developed method for scholars from manuscript research in Humanities.
- Iterative and incremental enhancements of the developed software tool based on feedback from the Humanities' participating scholars within the SFB 950.

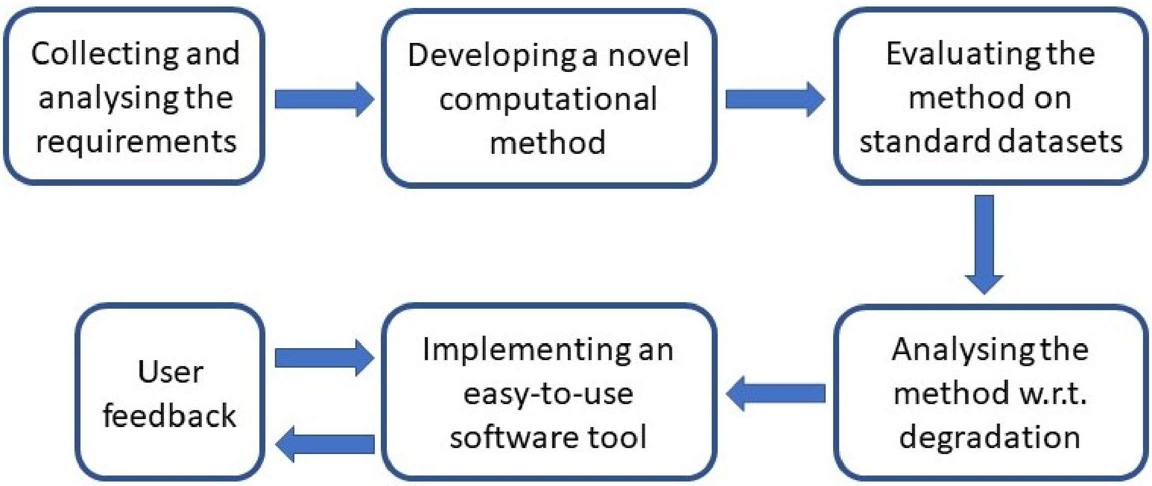


Figure 1.1: Flowchart illustrating the workflow of this dissertation.

1.5 Requirements from Sub-Projects in SFB 950

As in many interdisciplinary collaborations, having both different understanding of the scientific problems at both sides and different interpretations of the used terminology can form obstacles that researchers from both disciplines need to overcome. Therefore, one of the first steps in the Scientific Service Project Z03 was to analyse the requirements of participating scholars from Humanities and to make sure that a mutual understanding and a common language could have been established. The overall results of the requirements analysis served as a basis for deriving specific requirements for handwriting style analysis and collaborating with sub-projects to provide supporting information for their research questions.

Two effort-full requirement analysis phases took place during the second funding phase of the SFB 950 in 2016 [9]. These two phases aimed at providing a clearer understanding of the tasks and needs of the sub-projects and how the Z03 Scientific Service Project may meet the needs in an appropriate way given limited resources. These two phases were followed by personal interviews and joint discussion sessions in order to focus on detailed requirements of selected sub-projects with respect to handwriting style analysis only.

One of the outcomes from the requirements analysis process is the selection of sub-projects (see below) based on the demonstrated interest in computational methods, the availability of test data, and the ability of scholars to explain the potential role of software tools in their research. Notwithstanding the few pilot sub-projects, many of the SFB 950 sub-projects can directly benefit from a computational method to support their research with regards to handwriting style analysis and writer/scribe identification.

Some samples from the sub-projects within the SFB950 are presented in Figures 1.3, 1.4, 1.5 and 1.2. The requirements of these sub-projects can be divided into general requirements and project-specific requirements, and summarised as follows:

The *general requirements* for all selected sub-projects demand a method that can:

- Provide supporting information that can help to identify the scribe of a given handwritten text.
- Be applied to different types of scripts and character sets of different writing systems.
- Cope with a limited amount of handwritten text.
- Handle unbalanced and scarce sample data.

In the light of the mentioned problems of interdisciplinary collaboration, sub-project C08 [10] managed to clearly describe how a computational method can play a role in their tasks related to handwriting style analysis. Moreover, the scholars in this sub-project were able to provide sufficient amount of data for testing and evaluation. Therefore, this sub-project was selected as a pilot project with a realistic use case in this dissertation.

The *project-specific requirements* for sub-project C08 aim at a method that allows to:

- Sort handwriting samples by similarity to a given query.
- Provide a user-intuitive measure of similarity between the samples in question.

Thus, in general we conclude that for a method to be feasible, applicable, and reliable in real-life scenarios (especially in case of historical manuscripts), it needs to be offline, text-independent and segmentation-free. Specifically, a method meeting demands from pilot sub-project C08 has to furnish similarity measures resulting from feature-based classification of scarce and unbalanced sample data representing unknown/known scribes.

Figure 1.2: Three samples of different scribes from sub-project C08 [10] “East Frankish manuscripts with collections of formulas”.

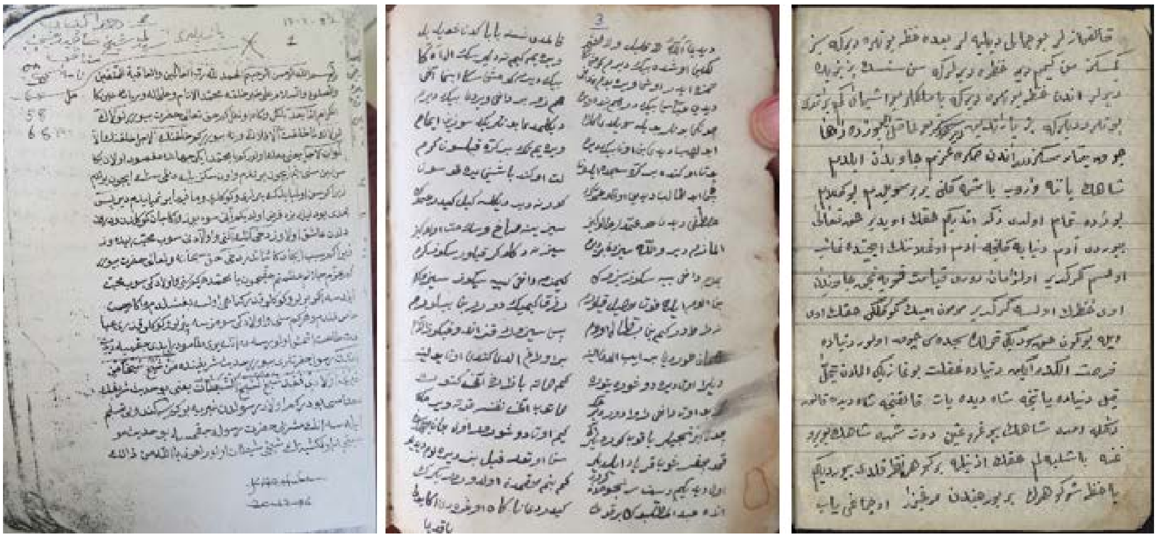


Figure 1.3: Three samples from sub-project C04 [11] "Reading, memorizing and recording: Manuscripts in Alevi village communities in Anatolia".

Figure 1.4: Three samples written by several scribes from sub-project B05 [12] "The handling of Qur'an manuscripts in Islamic-Arabic culture using the example of small and rolling Koran".

Figure 1.5: Three samples of different scribes from sub-project C06 [13] "Greek Aristotle manuscripts in teaching and interpretation practice".

1.6 Challenges in Computational Analysis of Digitised Manuscripts

1.6.1 Standard and Public Datasets

A wide range of methods has been developed for the task of writer identification; see Chapter 2 for details. Nevertheless, the vast majority of the methods has been developed and optimised to achieve high performance for synthesised contemporary datasets; furthermore, these datasets are mostly designed and created by researchers from the computational document analysis community rather than scholars from manuscript research in Humanities. Therefore, these standard datasets neither cover the typical degradation nor the unbalance and scarcity of handwriting samples of digitised historical manuscripts. In addition, there is no easy-to-use software tool currently available for handwriting style analysis in digitised manuscripts.

1.6.2 Degradation in Digitised Manuscripts

Digitised manuscripts typically contain a large amount of information that is irrelevant to the main handwriting (textual information), such as illustration images, para-texts and layout specifics. The presence of these pieces of irrelevant information can degrade the ability of any image processing and recognition task to achieve the desired results of analysing the handwriting and identifying the writing style. Even though page layout segmentation and analysis can help to extract the main text to a certain extent, other kinds of irrelevant information are much harder to deal with, such as stains and para-text in between text lines. In addition, digitised manuscripts typically suffer from different kinds of degradation such as low resolution, low contrast, high noise, and irregular orientation of text lines, etc.

Degradation in digitised manuscripts can result from e.g. poor preservation conditions, the used materials (e.g. paper or parchment), or even from the digitisation process itself. Clearly, degradation has a negative impact on the quality of the results of computational methods. This degradation can not be always attenuated or even eliminated using some semi-automatic preprocessing algorithms. Therefore, apart from selecting appropriate methods, a thorough analysis of computational methods should be carried out w.r.t. the typical degradation in digitised manuscripts in order to measure the impact of such degradation on their performance.

1.7 Contributions of the Dissertation

The main contributions in this dissertation are:

- The justified application of the Local Naïve Bayes Nearest-Neighbour (NBNN classifier) [14] with a novel descriptor matching constraint to the problem of writer analysis/identification.
- The introduction of a normalisation factor in order to cope with the problem of unbalanced data.
- The detailed analysis of the proposed method for the purpose of parameter optimisation and performance enhancement.
- The thorough analysis of the proposed method w.r.t. common degradation types in historical manuscripts.
- The implementation of the method as a software tool with an easy-to-use user interface and an intuitive presentation of results.

1.8 Organisation of the Dissertation

The rest of this dissertation is structured in chapters as follows:

- Chapter 2: The related work in the field of offline writer identification and retrieval is presented and discussed with respect to the requirements of selected sub-projects within the SFB 950.
- Chapter 3: The proposed method for handwriting style analysis is introduced in detail.

- Chapter 4: This chapter is dedicated to the experimental evaluation of the method on standard datasets, both contemporary and historical.
- Chapter 5: A detailed performance analysis of the method w.r.t. to some of the typical degradation in digitised manuscripts is provided.
- Chapter 6: An implementation of the developed method is presented along with a description of the GUI and the presentation of results. In addition, two use cases are given to demonstrate the applicability and usefulness of the implementation in actual scholars' research.
- Chapter 7: Conclusions of the presented research and possible future work are provided.

Chapter 2

Related Work

As mentioned in Chapter 1, the focus of this dissertation is on analysing handwriting samples of digitised manuscripts in order to generate similarity scores which can be used as a supporting information to scholars from manuscript research in Humanities for the task of handwriting style identification. We use the term *handwriting style identification* interchangeably with the term *writer identification* in the cases where the task is to identify a specific writer from her/his style; see Section 1.1. But in the computational document analysis community, the term *writer identification* is predominantly used to describe the task of identifying a specific handwriting style which belongs to a specific writer. Therefore, this term will be used in this chapter unless we refer to some other concepts.

Since the 1970s, the focus of computational document analysis research has been increased on the task of writer identification and retrieval. Several methods have been proposed and most are summarised until 1989 in a survey by [15]. A comprehensive review of a large number of publications in the last 20 years can be found in ([6, 16, 17]). Few recent works considered other related tasks such as handwriting style clustering [18] which involves defining groups of handwriting styles based on their similarities, manuscripts dating [19] which is the process of assigning a manuscript sample to a certain predefined period of production time and handwriting style classification [20–22] which is the process of assigning a manuscript sample to a predefined group of handwriting style. Nevertheless, the features used for writer identification (see Section 2.1) can be used for the other mentioned tasks, because of their discrimination power regarding different handwriting styles as demonstrated in [19].

Although we focus in this dissertation on digitised historical manuscripts, reviewing methods developed both for contemporary and historical handwritings can be useful. The discriminative power of a proposed method w.r.t. different contemporary handwriting styles can be useful for historical handwritings as well, because features describing the intrinsic characteristics and visual aspects shared by samples of the same handwriting style are assumed to be discriminative for both contemporary and historical handwritings.

In order to fulfil the requirements from the selected sub-projects within the SFB 950 (see Section 1.5), we need to take into account the problems of typical degradation, scarcity, and unbalance of data found in digitised manuscripts. Most of the state-of-the-art methods evaluate the performance on standard public datasets of contemporary handwritings with sufficient amount of balanced data for training; see Chapter 4. Furthermore, most of the state-of-the-art methods do not provide any analysis w.r.t. degradation, scarcity, and unbalance of data.

In this chapter, we review the used visual features in state-of-the-art methods for the task of writer identification, as well as their possible applicability to our problem at hand. Then we review the classifiers typically used in the field of writer identification. Since most of the state-of-the-art methods use learning-based classifiers in this field of research, which are not suited in our case (see Section 1.5), and in order to cope with the problem of data scarcity, we also review the state-of-the-art of learning-free classifiers from the field of image classification for natural scenes. Finally, we draw a conclusion which will be the basis for our proposed method.

2.1 Features Used for Writer Identification

As yet, the focus in writer identification research was mainly on feature selection and design rather than on classifiers. A wide variety of features has been used for the task of writer identification, such as forensic examiners' features like Quill features [23] as well as several categories of texture-based features [24, 25] such as run-length-based features [26–28], gradient- and contour-based features [29, 30]. Other researchers used allographic features [7, 31] and a mixture of texture-based and allographic features [6, 32]. On the other hand, some researchers used auto-derived features as an alternative to the manually designed features [33, 34].

2.1.1 Forensic Examiners' Features

Forensic document examiners attempt to visually extract discriminative elements of handwritings. Such elements are assumed to have the potential of distinguishing the handwriting of one writer from other writers [35]. In order to emulate this approach, the work in [36] combined features extracted from the gradients of letters' contours with character-level segmentation. For each sample of handwriting, a pseudo-alphabet is created by loosely segmenting the text into fragments of contours that consists of letters, part of letters, or parts of more than one letter. Then the distance is measured between elements of these alphabets by calculating the minimum distance required to transform one alphabet element into another alphabet element. They suggested a writer identification scheme as a possible application of their method by using the concept of inverse document frequency (IDF) to increase the significance of query letters that occur less frequently.

Another approach considered the relation between the direction of ink trace and its width as a probability distribution to construct the Quill feature [23]. Such features are particularly applicable to historical manuscripts written by using a quill.

Computational methods that extract such visually intuitive features and use a classification procedure that is comparable to the manual examination procedure are appealing to palaeographers. Nevertheless, such computational methods tend to provide poor performance when dealing with typical types of degradation in digitised manuscripts. In addition, visual features being distinctive to human vision are not necessarily distinctive for computational methods. Therefore, forensic examiners' approach is not a preferable candidate for the work in this dissertation.

2.1.2 Texture-Based Features

Instead of extracting visually intuitive features, other researchers used a quantitative description of the handwriting style in terms of pixels, gradients, and contour fragments distributions, and even the distribution of spaces within and between the letters.

A mixture of discriminative texture-based features has been extracted from text lines and paragraphs in [24] such that five categories of features are used for the writer identification task: slant and slant energy, skew, pixel distribution, curvature, and entropy.

Alternatively, a variant of Local Binary Patterns (LBP) has been used by [25]: Standard LBP is adapted and extended by applying a Sparse Radial Sampling (SRS-LBP) to cope with the particularities of handwritten texts.

Other types of texture-based features have been applied to the task of writer identification such as run-length, gradient-based and contour-based features. A brief review of these texture-based features is presented in the following sections.

Run-Length Features

The idea of calculating a run-length histogram from handwritten texts and using it in order to discriminate between two different writers has been first explored by [37]. The run-length of the background intensity value has been recorded and showed that similarity between the histograms of two samples of the same person's handwriting is greater than that between samples of two different person's handwriting. Quantitative measurements are shown for some characteristics of handwritings. Horizontal and vertical run-length are considered in that study. In a later work by the same researcher [38], additional properties have been extracted from the external structure of the handwriting, such as the outer margins of the text blocks, using horizontal and vertical run-length.

Local pixel intensity features like the second moment, variance, and entropy are extracted from separated characters of ancient Hebraic handwritings and used to identify the writers in [26]. These features are based on horizontal and vertical run-length histograms.

In order to validate the hypothesis that the writing style of an individual scribe remains constant across different scripts, handwritten texts in Greek and English are considered in [27]. Histograms of horizontal, vertical and diagonal run-length of the background and the foreground are used as features. It's worth noting here that the considered languages in that study share a large portion of their alphabets.

The General Pattern Run-Length Transform (GPRLT) is proposed in [28] as a modification to the standard run-length histogram. Their proposed algorithm can be applied to grey value images without the need for prior binarisation, but at the expense of having an additional free parameter to set the pixel intensity threshold in order to separate between the background (non-text) and the foreground (text) pixels within a given image.

Run-length features are the first category of features to be explored for the task of writer identification; nevertheless, these features proved to be not a practical choice for images with heavily textured background. Such features can only work with binarised or contemporary images where it is possible to easily and precisely separate the text from the background. The results of our preliminary experiments confirmed that run-length-based features are very sensitive to typical degradation in digitised manuscripts.

Gradient- and Contour-Based Features

Both gradient- and contour-based features try to capture the individuality of handwriting directly from the ink-trace in the image. This can be accomplished in many different ways, like describing the distribution of the intensity gradients around the ink-trace or the curvature of the segmented contour fragments.

Several gradient- and contour-based features were extracted and tested in [29], such as the contour-based features (CON) [39], the oriented basic image features (OBI) [40], the histogram of gradients (HOGs) [41] and the scale-invariant feature transform (SIFT) [42]. These features were used to identify writers of multi-page historical Arabic manuscripts [29] and resulted in high identification rates, particularly when using SIFT features. A learning-based rejection strategy is added later so that a classification decision can be rejected if no matching writer can be found [43].

A two stages method is suggested by [44]. In the coarse stage, a codebook is constructed by clustering SIFT descriptors extracted from handwriting images (while SIFT keypoints that lie on the background were eliminated). Then occurrence histograms of codebook vocabularies are calculated and used to measure the distance between images. In the fine stage, the candidate list is refined using both contour directional features and SIFT descriptors. Their method achieved state-of-the-art results in two contemporary datasets.

Writer's style has been encoded as a deviation from the mean encoding for a population of writers in [45] and an oriented Basic Image Feature Columns (oBIF) descriptor is used to encode the texture-based features. Both segmentation-based and segmentation-free implementations have been evaluated on contemporary datasets with state-of-the-art results.

In order to bridge the gap between methods based on image statistics and manual character-based methods, the writer is considered to be characterized by a stochastic pattern generator in [46], producing character fragments (Fraglets). A codebook of these fragmented *Connected Components Contours* (CO3) is constructed and used to compute a probability distribution for each writer.

Wahlberg *et al.*[30] proposed an unsupervised feature learning approach based on a dense contour descriptor sampling using Shape Context descriptors [47], combined with a learning-based approach for clustering handwriting samples from different writers, in a forensic setting. The mixed-Gaussian distribution is used to estimate the distribution of features across the different handwriting styles, and then to learn the metrics needed for the classification. The metric learning inference was based on multi-class Gaussian process classification.

Contour extraction from handwriting samples is highly sensitive to several degradation types such as noise and low resolution images. Therefore, in our case contour-based features are not a suitable candidate for digitised manuscripts. This is confirmed by our preliminary experimentations as well.

On the other hand, gradient-based features such as SIFT descriptors describe visual features in local regions of handwritings without the need for contour extraction or character segmentation. Moreover, methods with gradient-based features demonstrated state-of-the-art results on digitised manuscripts [29, 43,45]. Therefore, we use gradient-based features in this dissertation for our problem at hand.

2.1.3 Allographic Features

The extraction of allographic features is possible from images with no or little degradation because it requires character segmentation first. On the other hand, extracting parts of letters is possible in images with a low level of degradation using over-segmentation algorithms. These extracted contour fragments are used to construct a writer-specific codebook or models to be used in the classification. Such algorithms can be applied with relatively simple segmentation methods as long as they can extract repeatable (and similar) segments (contour fragments) from both training and test samples of handwriting [36].

An automatic retrieval system is developed by [7] for ancient Syriac manuscripts. A so called congealing algorithm is applied to create representative models of characters, and then an affine transformation is estimated of the actual observed characters as compared to the corresponding models. Experiments on seventy-six pages from nineteen Syriac manuscripts show that their method can identify pages written by the same hand with high precision. Nevertheless, the characters in the manuscripts are allocated and extracted manually by a human.

The work in [48] extends the idea of codebook-based writer recognition by generating two codebooks, a primary and a secondary. The text-lines are divided into small windows, four smaller adjacent windows are considered for each window. Features extracted from the main and adjacent windows are clustered separately. Their method achieved a state-of-the-art result on a contemporary dataset.

Instead of extracting graphemes from training data, the work in [49] synthesizes graphemes using the beta-elliptic model, while the work in [31] proposes the extraction of junctions instead of graphemes and the generation of a codebook from these junctions, a representation which is referred to as *Junclets*. A local descriptor is calculated at each junction by using the stroke-length distribution in every direction around a reference point inside the ink trace.

The task of writer identification in handwritten musical scores is tackled in [50] by adapting the *bag of visual words* framework using Blurred Shape Model (BSM). A Support Vector Machine (SVM) classifier is used to provide the final classification of musical scores.

Extracting characters from a handwritten text needs a character segmentation step to be done first. This process is not possible to be done automatically, or at least not reliable (see discussion in the previous sections), for the cases where handwriting is cursive (characters are connected and even overlapping) and/or suffer from degradation such as most of the digitised manuscripts from sub-projects within the SFB 950; see Figs 1.2, 1.3, 1.4 and 1.5.

2.1.4 Mixture of Texture-Based and Allographic Features

Several publications focused on combining texture-based and allographic features to obtain enhanced performance either by extracting and concatenating both types of features or by aggregating the classification results from both feature types.

Ball et al. [51] propose the combination of macro features for the task of writer identification, such as the number of interior and exterior contours, along with the gradient, structural and concavity bigrams attributes (GSC) which are first introduced in [52].

Bulacu et al. [39] proposed to combine texture features, namely contour-direction, contour-hinge and co-occurrence Probability Distribution Functions (PDFs), with grapheme emission PDF and shape codebook as allographic features for the task of writer identification. The performance of their method is evaluated on Arabic handwritings showing clearly the superiority of texture-based features over allographic features.

A thorough evaluation of both texture-based and allograph-based features for writer identification is found in [6]. Features extracted from contours, contour-hinges, and run-length histograms are used as texture features, while writer-specific grapheme emission PDF is used as an allographic feature, where the writer is characterised by a stochastic pattern generator producing graphemes. A detailed analysis of the performance of feature combinations is also included in [6]. The aforementioned comparison showed that the contribution of texture-based features to the final identification results is higher than the contribution of the allographic features.

Dondi *et al.*[32] proposed a method for short historical documents based on both allographic and texture features. Palaeographic data were used as a reference during the development of their method. Tests were conducted on Antonio Stradivari‘s relics, a collection of technical drawings and artefacts of the famous violin maker. Templates of allographic features have been created as well as augmented by resizing the templates, then matched using the normalised cross-correlation technique.

2.1.5 Auto-Derived Features

As an alternative to manually designed (or hand-crafted) features, Christlein *et al.* [33] used the activations from the penultimate deep residual network layer as features for the subsequent writing style classification task. A convolutional neural network (CNN) is trained using surrogate classes. These classes are created by clustering the training dataset, where each clustered index represents one surrogate class. Finally, the learned features are classified using Support Vector Machines (SVMs). Each SVM corresponds to a query sample, and is trained using external datasets as negative samples. Their method achieved a state-of-the-art result on a historical dataset (the samples of this dataset have been selected randomly using an automated algorithm rather than by scholars from the Humanities). Nevertheless, the validation set of the used dataset (which is provided for performance estimation) has been used and labelled as negative samples.

In a later work, Christlein *et al.* [53] compared Vector of Locally Aggregated Descriptors (VLAD) encoding with triangulation embedding. Furthermore, they investigated generalized max pooling as an alternative to sum pooling and the impact of decorrelation and SVMs.

The proposed method in [34] calculated the Probability Distribution Functions (PDFs) of some hand-crafted features from the handwritten samples and used these features as input to a convolutional neural network (CNN). The hand-crafted feature PDFs are hybridized with auto-derived CNN features. Such hybrid features are then fed into a Siamese neural network for writer verification. The experiments are carried out on an in-house Bengali offline handwritten dataset of 100 writers.

The main drawbacks of these methods are the excessive need for large amount of training data from the same domain as the test data and the need for class labels of the training data, or at least the positive/negative labelling for each sample in the data as in the case of the method in [33]. Therefore, auto-derived features are not a suitable choice in our case given the aforementioned requirements in Section 1.5.

2.2 Classifiers

Both writer identification and retrieval tasks can be viewed as an image classification problem, where the images are samples of digitised handwritings and all samples of the same writer represent a class in feature space.

Typically, there are two kinds of classifiers that can be applied to the problem of writer identification: parametric (learning-based) classifiers and non-parametric (learning-free) classifiers.

Parametric classifiers construct a model from the training data and try to estimate the parameters for that model, while non-parametric classifiers attempt to classify by comparing test data directly to the labelled data. Each method has its advantages and disadvantages. Obviously, parametric classifiers require a training phase to determine the parameters of the underlying model. Also, learning a new class typically requires re-training the entire classifier. Furthermore, parametric classifiers are usually resource-hungry and slow during training. On the other hand, the main problem with non-parametric classifiers is the inferior performance they provide. This problem of performance in non-parametric classifiers is addressed by Boiman *et al*. [54] as it will be explained in Section 2.2.2.

2.2.1 Learning-Based Classifiers

Most of the recent methods for writer identification use learning-based classifiers like Support Vector Machine (SVM) [27,33,43,50,55], CNN [56], or Gaussian Process Classification [30] and a wide range of distance measures or norms is used like Euclidean, Hamming, Chi-Square and the absolute difference.

Although this category of classifiers can provide a high identification rate, it typically requires a large amount of labelled training data. This requirement renders these classifiers useless and impractical for most of the real-life problems when dealing with historical manuscripts with small and unbalanced sets of writer class samples.

In addition, deep learning methods automatically derive and extract features that can discriminate between the labelled classes in the training datasets. Then they use these learned features to classify the unknown samples. This approach is no problem as long as class labels (or in other words ground-truth) are not subject to opinions and they are agreed upon. In the case of digitised manuscripts, it is often the case that the labels of handwriting styles or of respective writers are not necessarily agreed upon by scholars from manuscript research in Humanities, and hence the labels (e.g. writer name or style category) are subject to opinions like a school of thought or experience. Using such deep learning methods in these cases may provide results that only reinforce the opinion of the scholar or group of scholars being responsible for labelling the training data in the first place.

On the other hand, hand-crafted features are manually designed before hand; therefore, the calculated features for a given unknown sample are independent of the datasets and the labels for the known reference samples.

2.2.2 Learning-Free Classifiers

Given the task of writer identification, the number of samples per writer is usually rather small in most of the public datasets as well as in manuscript research within the SFB 950. This is especially true in the case of digitised manuscripts from sub-projects within the SFB 950 where the amount of handwritings text can be as little as only a couple of text lines per writer in some of the cases. Therefore, a learning-free classifier is better suited for our task.

The main problem with learning-free classifiers is the poor performance compared to learning-based classifiers. Since many learning-free classifiers are based on nearest-neighbour (NN) distance estimation, they inherited the bad reputation due to low classification rate. This assumption has been proved wrong and addressed by Boiman *et al*. [54].

Boiman *et al*. [54] proposed a learning-free classifier, called Naïve Bayes Nearest-Neighbour (NBNN). This classifier has demonstrated state-of-the-art results for the task of classification of natural scene images. Boiman *et al*. argued that two practices can lead to a significant degradation of performance for methods based on nearest-neighbour distance estimation; thus, these practices should be avoided. These two practices are discussed as follows:

- **Descriptor quantisation:** Reducing the number of image descriptors by keeping representative descriptors and removing all the other descriptors can cause a large loss of information for non-parametric and learning-free classifiers; such classifiers do not have a training phase to compensate for this loss.

  The quantisation error (caused by removing the non-representative descriptors) is especially outspoken for the more informative features found in less dense areas of feature space. This practice has a larger impact when the discriminative power of features of handwriting styles is inversely proportional to the frequency of occurrence of these features in a given manuscript.

- **Image-to-image distance:** In contrast to Image-to-image distance, measuring image-to-class (in our case class means handwriting style) distance will generalise the nearest-neighbour (NN) search to class-matching instead of image-matching; thus, learning-free classifiers can cope better with intra-class variations. This is particularly important for handwriting style with variations even within the same handwriting style/writer.

  Moreover, avoiding this practice enables a good generalisation beyond the provided labelled images. The NBNN classifier combines bits and pieces of information from different example images of each class. This is especially valuable in our case of digitised manuscripts because it is frequent that only a limited number of labelled samples is available.

The aforementioned attributes, namely avoiding descriptor quantisation and calculating image-to-class distance, represent the core strength of the NBNN classifier. Timofte *et al*. [57] demonstrated the importance of these attributes by replacing the NN part with several alternative representations (e.g. Local Linear Embedding (LLE) and Iterative Nearest Neighbours (INN) solving a constrained least squares (LS) problem) while keeping the good performance or even improving upon it sometimes on the expense of testing time and memory space requirements.

The NBNN classifier demonstrated state-of-the-art results on classification tasks of natural scenes [54]. However, the NBNN framework also has its requirements and limitations, which we discuss as follows:

- **The good performance of this classifier relies on dense features sampling**: The state-of-the-art performance reported in [54] is achieved by densely sampling large redundant local features for both labelled and test images, which seems necessary for good image-to-class distance estimation. This dense sampling resulted in about 15000 to 20000 features per image. Therefore, dense keypoints detectors such as the Scale-Invariant Feature Transform (SIFT) [42] and Features from Accelerated Segment Test (FAST) [58] are essential for the NBNN classifier to provide high classification rates.

- **The assumption of features independence:** Each feature is considered independently from other features within the same image in the NBNN algorithm; therefore, information concerning the overall image composition such as spatial relations is ignored. As a result, distinguishing between different objects with similar parts is likely to be difficult for the NBNN classifier [59]. Moreover, localising objects using NBNN results in much smaller detection windows than the object instance to be detected [60]. In our case of handwriting style analysis, there are no objects to localise. Therefore, this drawback poses no problem for our task.

In addition, it has been demonstrated in [61] that the NBNN classifier is applicable even in domains where the independence assumption is violated. They showed by performing extensive evaluations on many challenging datasets that the NBNN classifier can perform well even when the assumption does not hold.

- **The high computational cost during the testing time:** The nearest neighbour (NN) search of the NBNN classifier is computationally expensive when classifying a test image. This limits the scalability to high number of classes in real-world applications, which is one of the requirements from the sub-projects within the SFB 950 (see Section 1.5), due to the required dense sampling of features and to the need for NN search in every class for every query descriptor.

  Several solutions have been proposed to speed up the classification time [59, 62]. Most of these attempts are based on adding a learning phase. On the other hand, McCann and Lowe [14] proposed an improvement of the NBNN classifier without any training phase in order to speed up the classification time by even more than 100 times. Additionally, this improvement increased the classification accuracy and even better scaled to a large number of classes (viz. run-time of improved NBNN grows with the log of the number of classes rather than linearly).

- **Bias towards classes represented by a large number of features:** The NBNN classifier assumes similar densities for all classes in feature space. In practice, this assumption is often violated, resulting in a strong bias towards one or a few classes with high densities [59]. This is a very critical limitation in our case of digitised manuscripts due to the prevalence of unbalanced data from the sub-projects within the SFB 950.

  Several methods have been proposed to adapt the NBNN classifier to work with unbalanced datasets [59,60,62]. However, all these methods proposed a learning phase to address the problem of unbalanced data.

  In this dissertation, we propose a normalisation factor to correct for the unbalanced data without any training phase.

2.3 Conclusion

The focus of computational methods for writer identification was mainly on feature selection and design rather than on classifiers. Several types of features have been used by researchers to capture the individuality of handwriting. While most of these features are selected and used for contemporary handwriting samples, some of them can be used for digitised manuscripts despite the typical degradation of such data.

Although methods using texture-based features can provide a good performance in general, run-length and contour-based features are very sensitive to typical degradation in digitised manuscripts. This is confirmed by our preliminary experimentations using data samples from sub-projects in the SFB 950.

On the other hand, methods using gradient-based features demonstrated state-of-the-art results for digitised manuscripts [29, 43, 45]. These features describe properties of local pixel intensities within handwriting samples without the need for contour extraction or character segmentation; therefore, they can cope better with typical degradation in digitised manuscripts.

Methods based on auto-derived features suffer from the need for a large amount of training data from the same domain of the test data. Moreover, such methods require prior class labelling of the training data, or at least the positive/negative labelling for each sample in the data. While this draw-back poses no problem in large datasets with a big amount of training samples, it is a major problem in our case of digitised manuscripts with scarce samples of data.

Learning-based classifiers can provide a high identification rate; nevertheless, it typically requires labelled training data. This requirement renders these classifiers useless and impractical for most of our real-life problems when dealing with historical manuscripts with small and unbalanced sets of samples per writer. In addition, learning-based classifiers based on deep learning approaches need confirmed and potentially subjective labelling of training data in order to avoid a biased classification; see discussion in Section 2.2.1.

As we mentioned earlier, image data in historical manuscript research is often sparse, unbalanced and without labels (lack of ground truth). Therefore, learning-free classifiers are a better choice to use in our case. Nevertheless, learning-free classifiers generally provide poor performance compared to learning-based classifiers. Since many learning-free classifiers are based on nearest-neighbour (NN) distance estimation, they inherited the bad reputation due to low classification rate. This assumption proved to be wrong by Boiman *et al.* [54].

As already pointed out, Boiman *et al.* [54] proposed the Naïve Bayes Nearest-Neighbour (NBNN) classifier which has demonstrated state-of-the-art results for the task of image classification for natural scenes . Nevertheless, the NBNN classifier also has its limitations which need to be addressed in order for the classifier to fulfil the requirements from the selected sub-projects within the SFB 950; see Section 1.5.

In general, the good performance of the NBNN classifier relies on dense features sampling. Both the Scale Invariant Feature Transform (SIFT) [42] and Features from Accelerated Segment Test (FAST) [58] are such dense keypoints detection algorithms. Furthermore, these two algorithms detect local changes in pixel intensity in images and do not depend on colour information in the images which thus will be ignored. Therefore, these two keypoint detection algorithms are suitable choices for our problem at hand.

Also, the NBNN classifier is computationally expensive during testing time thus limiting the usability of any practical solution that we want to provide for the mentioned scholars from Humanities. However, the Local NBNN classifier [14] addressed this problem and has been shown to provide a high classification rate with a large improvement in both run-time and scalability to a large number of classes as compared to the original NBNN classifier.

Finally, the NBNN classifier has a strong bias towards classes represented by a large number of features. Since the problem of unbalanced data is prevalent in digitised manuscripts, this limitation needs also to be addressed. All proposed methods to address this problem had to add a training phase, which requires several labelled samples from each writer as a training set. Therefore, in order to fulfil this requirement, we propose a novel learning-free solution to this problem in this dissertation.

Chapter 3

Proposed Method

As justified in the previous chapters, we propose an offline, text-independent, and segmentation-free writer identification method based on the Local NBNN classifier [14]. Since colour is not a characteristic property of the handwriting style (see discussion in Section 2.3), both query and labelled sample images of handwritten pages of manuscripts are converted to grey scale using the weighted sum of RGB channels, whereas binary images are kept as they are. Then keypoints are detected in all images and descriptors are calculated for each keypoint. In order to match the calculated descriptors from a query image with the ones calculated from the labelled images, a learning-free classifier is used due to the fact that in many practical cases (as well as in many public datasets) the number of samples per writer is very small. A non-parametric learning-free classifier is proposed by Boiman *et al.* [54] and they demonstrated state-of-the-art results for image classification tasks of natural scenes. The two main limitations of this approach are: The need to search for a nearest neighbour in each class, and the bias toward classes with more descriptors than others. While the first problem is tackled by McCann *et al.* [14], we propose a normalisation step in order to cope with the second problem. Details are presented in the following sections.

3.1 Keypoints Detection and Feature Extraction

Keypoints simply are spatial locations, or points in the image that contain interesting and discriminative features. Dense keypoints detection algorithms, namely SIFT [42] and Features from Accelerated Segment Test (FAST) [58], are used within our proposed method in order to provide a sufficient number of keypoints for reliable nearest-neighbour search. We experimented with SIFT and FAST keypoints separately and the respective results for each type of keypoints are presented in Chapter 4.

3.1.1 SIFT Keypoints Detection

Though the SIFT keypoints detection algorithm has been designed and optimised originally for images of natural scenes, we will show that it can perform well in the context of writer identification if we take the particularity of handwritings into consideration. The SIFT keypoints detection algorithm [42] consists of three main steps:

1. **Scale-space extrema detection using difference-of-Gaussian:**

   The scale-space of an image is defined as a function $L(x, y, \sigma)$ that can be generated by convolving a Gaussian kernel $G(x, y, \sigma)$ of varying scale σ with an input image $I(x, y)$; see Fig. 3.1:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{3.1}$$

where $G(x, y, \sigma) = \dfrac{1}{2\pi\sigma^2} \exp -\frac{(x^2+y^2)}{2\sigma^2}$.

In order to detect scale-space extrema, the difference-of-Gaussian is computed as an approximation of the scale-normalised Laplacian of Gaussian, which is presented in [63], as follows:

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma), \tag{3.2}$$

where k is a constant multiplicative factor that separates consecutive scales. Each octave of scale-space is divided into an integer number, s, of intervals, so $k = 2^{1/s}$; see Fig. 3.1.

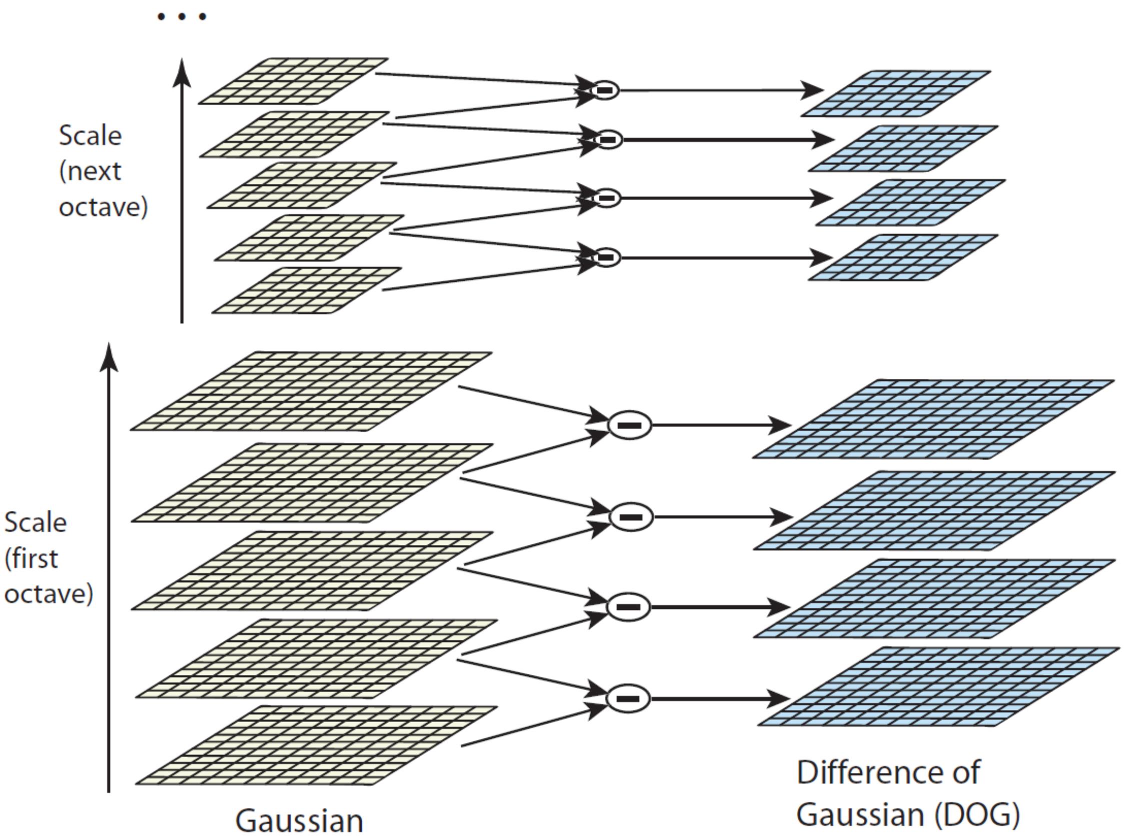


Figure 3.1: The original image is iteratively convolved with Gaussian filters $G(x, y, \sigma)$ to provide a set of scale-space images (on the left) for each octave of scale-space. Each pair of adjacent scale space images is subtracted to provide a difference-of-Gaussian result (on the right). The Gaussian images are down-sampled by a factor of 2 after each octave. Reproduced from [42].

2. **Keypoint localisation:**

   In order to detect the local maxima and minima of $D(x, y, \sigma)$, each sample point is compared to its eight neighbours in the current image and nine neighbours in the scale above and below, which results in a total of 26 neighbours to compare with; see Fig. 3.2. The sample point is selected as a candidate keypoint only if it is larger than all of these neighbours or smaller than all of them.

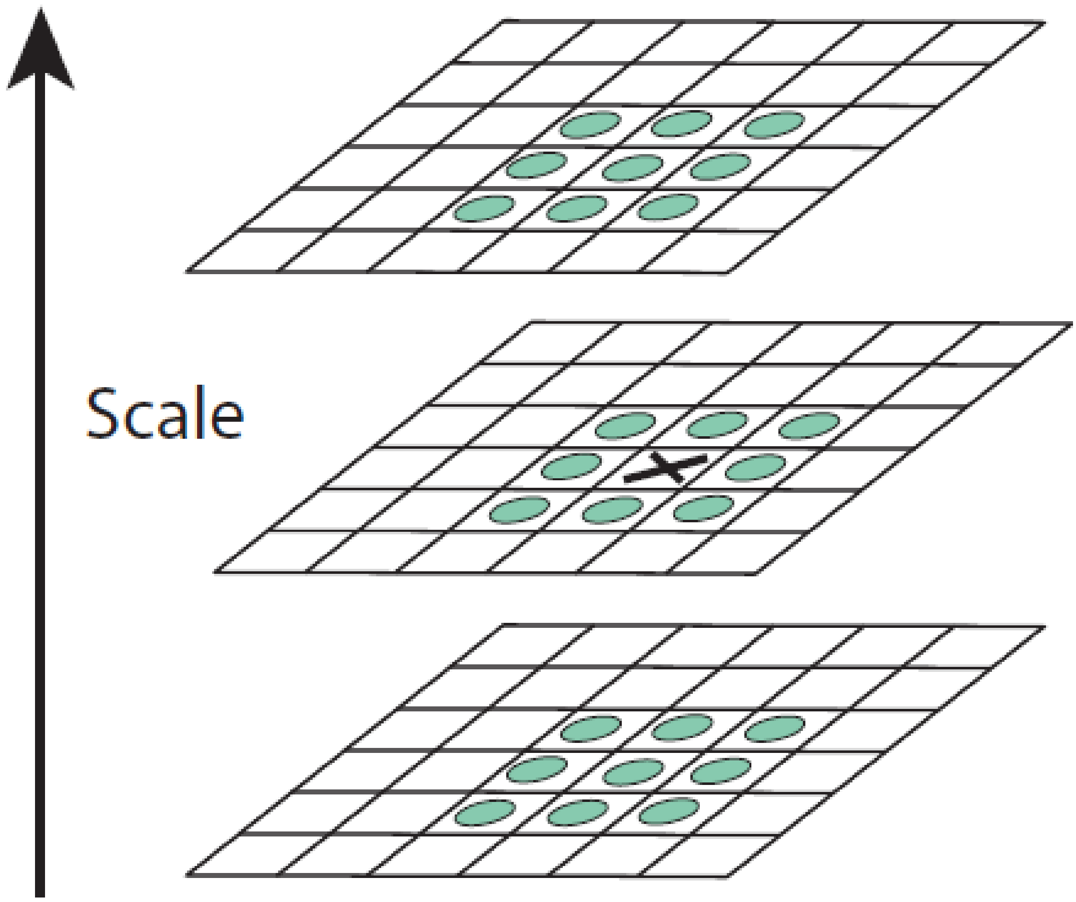


Figure 3.2: The pixel marked with X is compared to its 26 neighbours marked in circles in 3x3 local neighbourhoods at the current and adjacent scales to detect the maxima and minima in the Difference-of-Gaussian results. Reproduced from [42].

In order to enhance the accuracy of calculated localisation for the detected keypoints, a Taylor expansion of $D(x, y, \sigma)$ is applied to the local sample points, which are the candidate keypoints, to determine the interpolated location of the maximum (this approach has been proposed first in [64]), and thus improves the accuracy of localisation.

3. **Edge-response elimination:**

The difference-of-Gaussian function $D(x, y, \sigma)$ has a strong response for edges, this response can have a poorly determined location along an edge, and thus can be unstable even in case of small amounts of noise. In general, such poorly determined edge-responses have a large principal curvature across the edge but a small one in the perpendicular direction. The principal curvatures can be computed from a 2x2 Hessian matrix H at the location and scale of the keypoint:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix},$$

where the derivatives D can be estimated by taking differences of neighbouring sample points.

Since we are only concerned with the ratio of the principal curvatures of D, we can avoid computing the eigenvalues of H by using the approach proposed by Harris and Stephens in [65]. Thus, we can directly compute the ratio between the larger magnitude eigenvalue and the smaller one. In order to check that the ratio of principal curvatures is below some threshold, we need only to check for the following condition:

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r} \tag{3.3}$$

where Tr and Det are the trace and determinant operators respectively.

r is the ratio between the larger and the smaller magnitude eigenvalue.

Eliminating edge responses by the SIFT algorithm has a big yet positive impact on both the number and the quality of keypoints. In the original publication [42], the ratio of principal curvature is set to 10. The same ratio is used in this work as it yielded better results in our preliminary experiments. Generally speaking, the higher the ratio is, the better the quality of keypoints is, but the final number of keypoints will be less.

4. **Orientation assignment:**

Each keypoint is assigned an orientation based on the distribution of local gradient vectors so that it can be represented relative to this orientation thus achieving invariance to image rotation. For each sample image at a given scale $L(x, y)$, the gradient magnitude $m(x, y)$ and orientation of gradient vector $\theta(x, y)$ are approximated using numerical differences:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}, \tag{3.4}$$

$$\theta(x, y) = \tan^{-1}\left[\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}\right]. \tag{3.5}$$

An orientation histogram is then computed from the gradient orientations of sample points within a region around the keypoint. The orientation histogram has 36 bins covering the 360-degree range of orientations. Each sample added to the histogram is weighted by its gradient magnitude and by a Gaussian-weighted circular window with a σ that is 1.5 times that of the scale of the keypoint. The highest peak in the histogram is detected, and then any other local peak that is within 80% of the highest peak is used to also create a new keypoint with that orientation. Then, a parabola is fit to the 3 histogram values closest to each peak and the interpolated peak position is used as an orientation estimate of the dominant directions of local gradients.

Objects in natural scenes are subject to any amount of rotation within an image, while in handwriting images, the text usually has a limited range of differences in orientation within the same script type or alphabet. Furthermore, the handwriting orientation is a characteristic of the writing style of specific writers, and thus a discriminative property of its features.

For SIFT keypoints detection, we used the default parameters as proposed in the original publication [42]. The used number of image octaves is 3, and the used number of scales per octave is 5, the σ of the Gaussian applied to the original image is 1.6 and the edge-response threshold is 10.

3.1.2 FAST Keypoints Detection

For FAST keypoints, a circular neighbourhood of 16 pixels around every pixel p in the image has been used as proposed in [58]; see Fig. 3.3. p is classified as a keypoint if there are n contiguous pixels, where $n = 9$, in the surrounding discrete circle satisfying one of the following conditions:

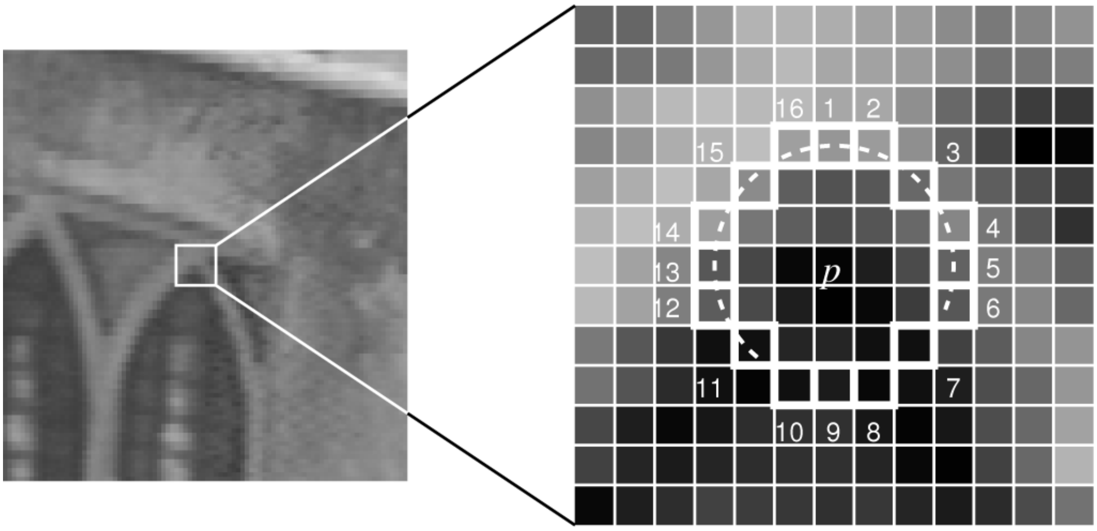


Figure 3.3: FAST keypoint detection (Reproduced from [58]).

- $\forall i \in n : I_i > I_p + t$
- $\forall i \in n : I_i < I_p - t$

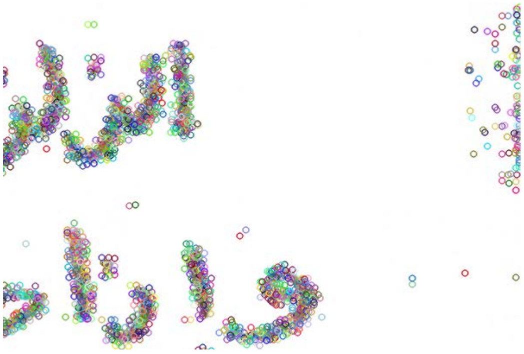
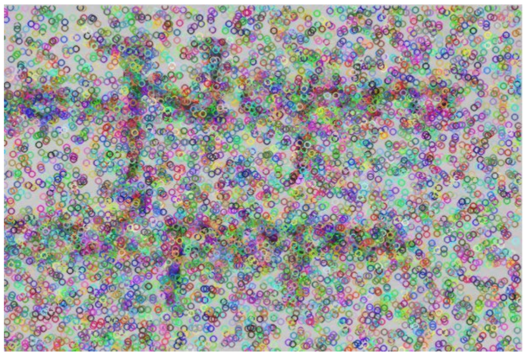
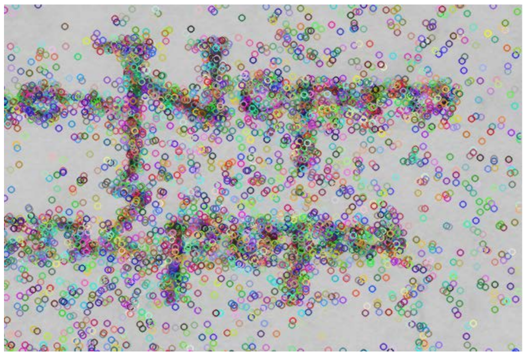
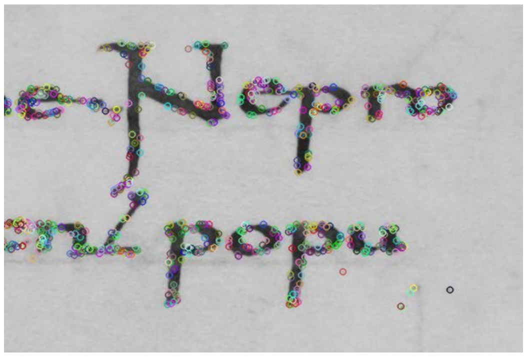
I_p is the intensity of the candidate pixel and I_i is the intensity of any pixel that belongs to n. t is a threshold to be selected manually.

The corner strength is defined in [58] as the maximum value of t for which the segment test of that corner point is passed. In our experiments, we could show that only a small percentage of the detected keypoints (yet sufficient number of keypoints) is needed to achieve high identification rates and an even smaller percentage for historical manuscripts.

We set $t = 0$ in all of our experiments so that we can detect all the keypoints in the image, then we sort the keypoints by their corner strength value. Finally, we only consider the keypoints with the highest strength values. Setting $t = 0$ does not affect the calculation of the corner strength values, because the algorithm iterates over all the possible values of t until the segment test passes, then it stores the strength value for that keypoint.

Let PCK be the percentage of considered FAST keypoints with the highest strength value. It can be noticed that the needed value of PCK to achieve the highest identification rate possible is related to the number of keypoints detected on non-textual parts of the sample images; see Table 3.1.

Table 3.1: FAST keypoints detected with different values of PCK. The first column contains part of an image from contemporary ICFHR-2016 dataset [66], while the second column contains part of an image from historical St. Gall dataset [67].

$PCK = 1$		
$PCK = 0.5$		
$PCK = 0.1$		

The impact of PCK on the identification rate, which is the ratio of correctly identified writers over all identifications in a given dataset, is presented in Fig. 3.4 and Fig. 3.5 using contemporary and historical documents, respectively. It can be noticed that much smaller values of PCK, and thus lesser number of keypoints, are needed for historical manuscripts to eliminate keypoints on non-textual information due to the complex and noisy background of the historical manuscripts. This means that a smaller number of Nearest-Neighbour (NN) searches are needed, and fewer descriptors are stored in the memory; therefore, a significant improvement can be achieved in terms of classification time and memory requirements compared with using all the detected FAST keypoints. These two aspects are important for our software tool implementation. In addition, the performance can be enhanced by selecting a suitable experimentally determined value of PCK for both contemporary and historical manuscripts as shown in Figs. 3.4 and 3.5.

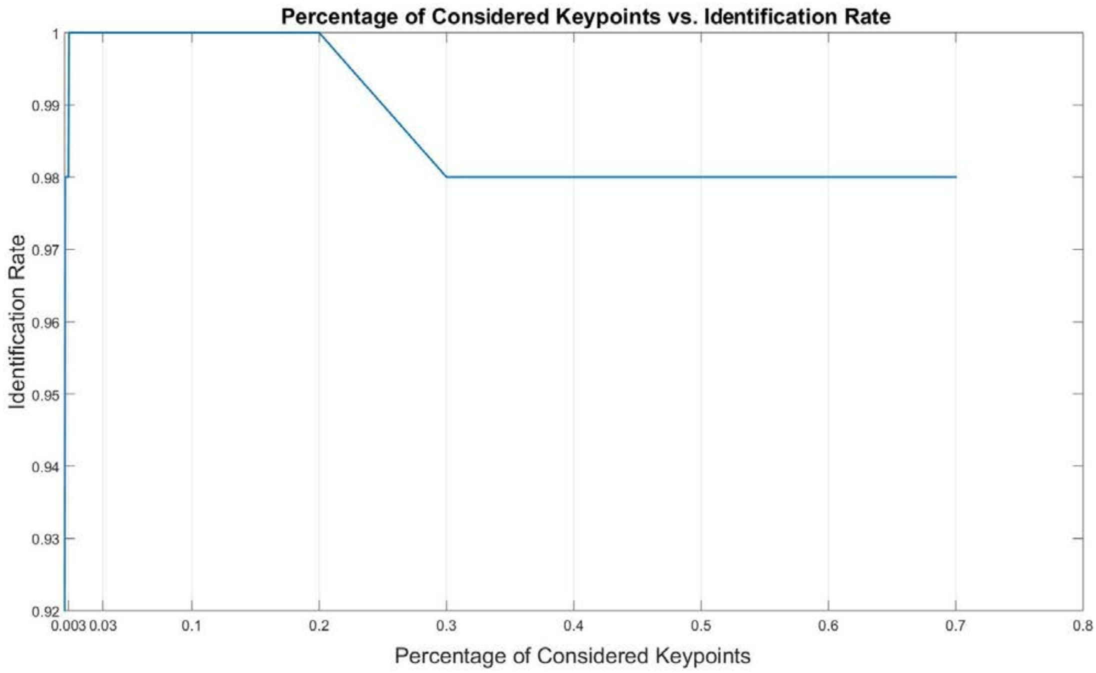


Figure 3.4: PCK vs. the identification rate using a subset of the historical manuscripts of St. Gall dataset [67].

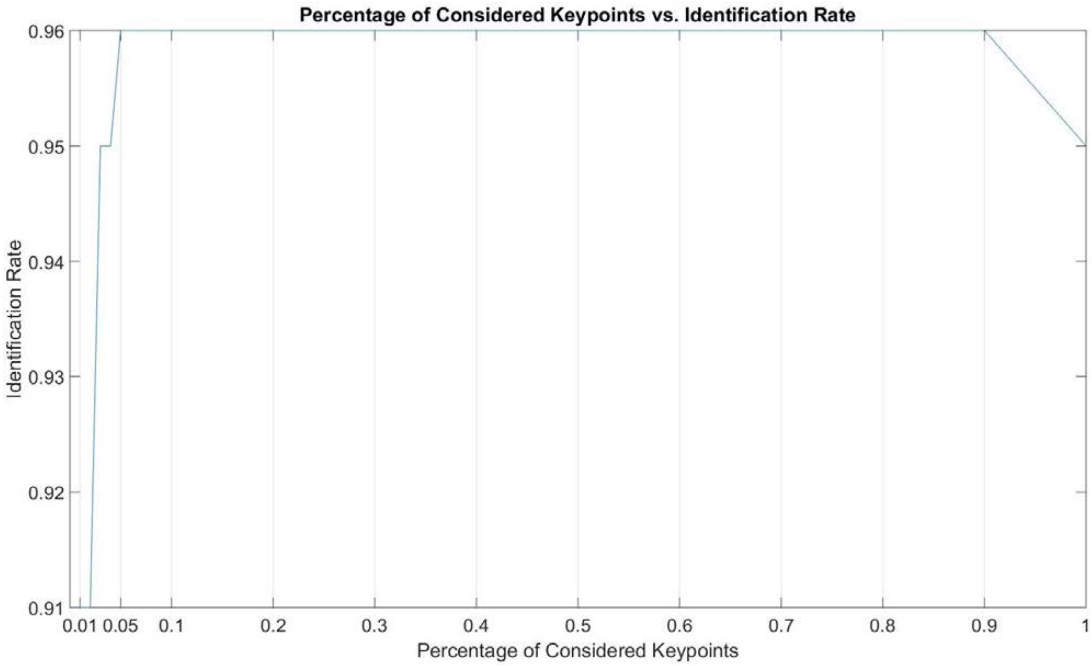


Figure 3.5: PCK vs. the identification rate using the contemporary datasets of the validation set from ICFHR-2016 competition of writer identification, task 1A [66].

3.1.3 SIFT Descriptor

After detecting keypoints by SIFT and FAST algorithms, we compute a descriptor for the local image region around each of the detected keypoints. The image descriptors present local visual features of an image and their elementary visual characteristics as a unique set of numeric values which can be used by computational methods, e.g. to calculate similarities between two images.

Following the justification in Chapter 2, we use SIFT descriptor [42] to describe the local image gradients in digitised manuscripts. The SIFT descriptor is calculated as follows:

The gradient magnitudes and orientations are sampled around the local neighbourhood of each keypoint, which is the centre of the image array where the descriptor is calculated. Then the magnitudes are weighted by a Gaussian window to give less emphasis to gradients that are far from the centre of the descriptor. These samples are then accumulated into orientation histograms; see Fig 3.6. The descriptor used in this dissertation consists of a 4x4 array of histograms with 8 orientation bins in each, which results in a 4x4x8 $= 128$ numerical value as a feature vector for each keypoint. This SIFT descriptor is used for both SIFT and FAST keypoints representation.

In the case of SIFT keypoints, the SIFT descriptor is calculated relative to the SIFT keypoint orientation (see Section 3.1.1), thus achieving invariance to image rotation. In the case of FAST keypoints, no orientation is calculated (see Section 3.1.2); therefore, the SIFT descriptor is not relative to the keypoint orientation.

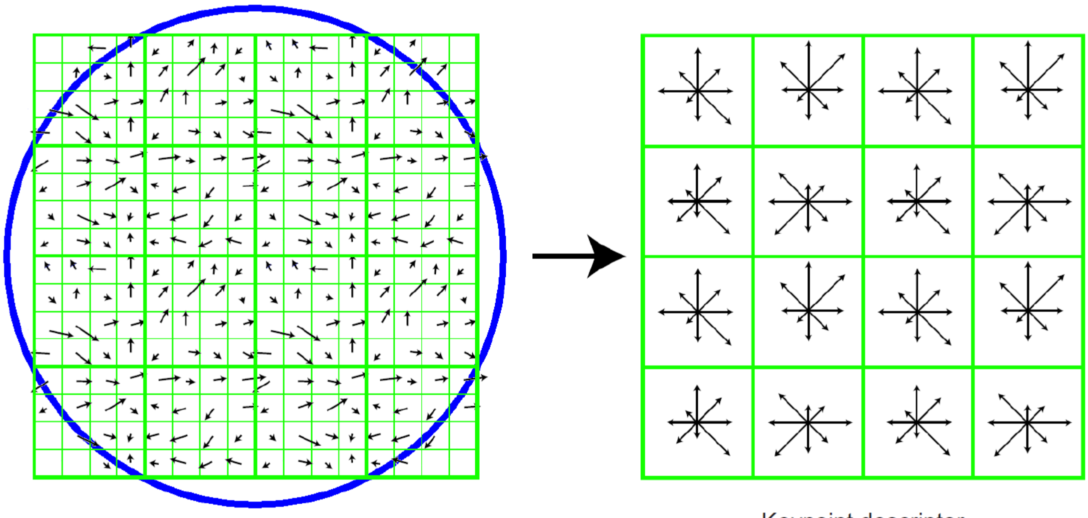


Figure 3.6: The gradient magnitude and orientation is computed at each sample point in the local neighbourhood around the keypoint, as shown on the left. Then they are weighted by a Gaussian window, indicated by the overlaid circle. These samples are then accumulated into orientation histograms summarizing the contents over 4x4 subregions, as shown on the right, with the length of each arrow corresponding to the sum of the gradient magnitudes near that direction within the region. This figure shows a 4x4 descriptor array computed from a 16x16 set of samples. Reproduced from [42].

3.2 Matching

A state-of-the-art Naïve Bayes Nearest-Neighbour (NBNN) classifier has been proposed by Boiman *et al.* [54] for the task of classification of natural scenes. They showed that conditional class probabilities can be well approximated by the squared Euclidean distance of the query descriptor to the nearest feature vector belonging to the correct class. Their mathematical derivation presented in [54] is shown below:

Given a query image q represented by a set of local features d and a set of classes C, q can be classified as belonging to class $c \in C$ according to the conditional probability

$$c = \underset{C}{\text{argmax}}\ p(C|q). \tag{3.6}$$

By applying Bayes' theorem and assuming a uniform (equal) prior probability over classes we obtain

$$c = \underset{C}{\text{argmax}}\ p(q|C). \tag{3.7}$$

If the n descriptors d_i, extracted from image q, are assumed to be independent, the equation can be re-written as

$$c = \underset{C}{\text{argmax}} \left[\log(\prod_{i=1}^{n} p(d_i|C)) \right] \tag{3.8}$$

$$= \underset{C}{\text{argmax}} \left[\sum_{i=1}^{n} \log p(d_i|C) \right]. \tag{3.9}$$

The probability $p(d_i|C)$ in Eqn. 3.9 can be approximated using a Parzen window estimator, with kernel K, i.e.,

$$\hat{p}(d_i|C) = \frac{1}{L} \sum_{j=1}^{L} K(d_i - d_j^c), \tag{3.10}$$

where L is the number of descriptors that belong to class c in the labelled set, and d_j^c is the j-th nearest descriptor to d_i in class c. A further approximation can be done by using only the r nearest-neighbours,

$$\hat{p}_r(d_i|C) = \frac{1}{L} \sum_{j=1}^{r} K(d_i - d_j^c). \tag{3.11}$$

It can be approximated further by considering only the single nearest-neighbour ($\mathrm{NN}_c(d_i)$) by setting r to 1:

$$\hat{p}_1(d_i|C) = \frac{1}{L} K(d_i - \mathrm{NN}_c(d_i)). \tag{3.12}$$

Substituting Eqn. 3.12 into Eqn. 3.9 and using a Gaussian kernel for K gives

$$c = \underset{C}{\operatorname{argmax}} \left[\sum_{i=1}^{n} \log \frac{1}{L} e^{-\frac{1}{2\sigma^2}\|d_i - \mathrm{NN}_c(d_i)\|^2} \right] \tag{3.13}$$

$$= \underset{C}{\operatorname{argmin}} \left[\sum_{i=1}^{n} \| \ d_i - \mathrm{NN}_c(d_i) \ \|^2 \right], \tag{3.14}$$

where ($\log$) is the natural logarithm.

The last Equation 3.14 shows that conditional class probabilities can be approximated by the squared Euclidean distance of query descriptor to the nearest feature belonging to the correct class. In other words, it suffices to find the class with the minimum Euclidean distance of its features to those of the query image.

An illustration of NBNN algorithm using images of digitised manuscripts is presented as a simplified example in Fig. 3.7. The NBNN algorithm searches for the nearest neighbour of each descriptor in the query Q (only 3 descriptors are shown in this illustration) in all the classes (writers), namely $W1, W2, W3$. Then the algorithm accumulates the distances to each class separately. The query sample finally assigned the label of the class with the smallest total distance, which is $W1$ in this case.

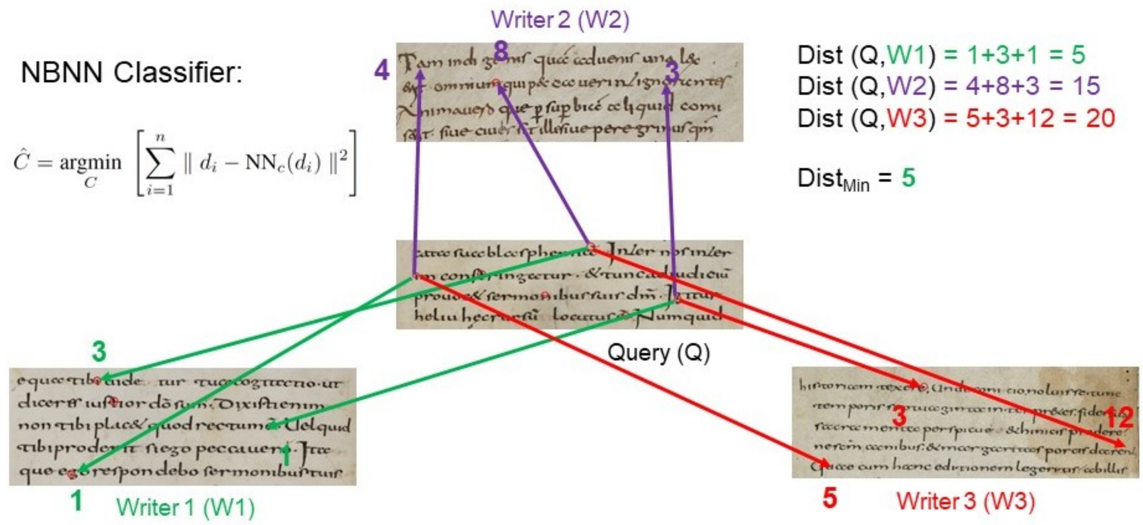


Figure 3.7: Illustration of NBNN classifier using images of digitised manuscripts.

An NBNN algorithm can be summarised as follows:

1. Compute n descriptors $d_1, ..., d_n$ of the query image q.

2. $\forall d_i\ \forall c \in C$ compute the NN of d_i in c: $NN_c(d_i)$, where NN_c is the Nearest-Neighbour in class c.

3. $\hat{C} = \underset{C}{\text{argmin}} \sum_{i=1}^{n} \parallel d_i - NN_c(d_i) \parallel^2$, where $\hat{C}$ is the class with minimum total distance.

Later, McCann *et al.* [14] presented the Local Naïve Bayes Nearest-Neighbour (Local NBNN) algorithm as an improvement to the NBNN algorithm. This improvement involved increasing both the classification accuracy and the classification speed for images of natural scenes; therefore it can also better scale to a large number of classes.

The basic idea of Local NBNN is eliminating the need to search for a nearest-neighbour match in all classes; instead, only the classes within a certain neighbourhood of the query descriptor in feature space are considered. Fig. 3.8 illustrates the main difference between NBNN and Local NBNN.

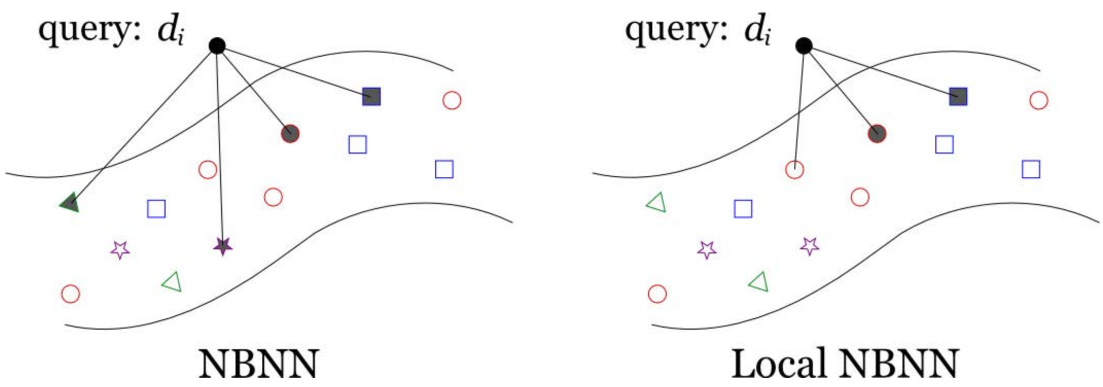


Figure 3.8: The difference between NBNN and Local NBNN. NBNN forces a query descriptor d_i to search for its closest neighbour in every class (given as filled icons). Local NBNN requires the query descriptor to search for its closest neighbour only from the closest classes, where the neighbourhood is defined by the number of nearest neighbours to the query descriptor (also given as filled icons). Reproduced from [14].

McCann *et al.* [14] went one step further and showed that the effect of each descriptor in a query image Q can be expressed as a log-odds update. This formulation allows us to be selective about which log-odds updates to apply. The proposed classification rule by [14] is:

$$\hat{C} = \underset{C}{\text{argmax}} \left[\sum_{i=1}^{n} \log \frac{p(C|d_i)P(\bar{C})}{p(\bar{C}|d_i)P(C)} + \log \frac{P(C)}{P(\bar{C})} \right], \tag{3.15}$$

where C is any given class and $\bar{C}$ is the set of all other classes. The prior term can be dropped if one assumes equal class priors. Significant log-odds updates can be used to adjust the class posteriors for which the descriptor gives a positive contribution to the sum [14].

For the selected increments where the posterior odds are greater than the prior odds, the NBNN classification rule is applied [14] as in equation 3.14.

The proposed method in this work is based on the Local NBNN Algorithm (2) in [14], which, by ignoring priors as explained above, we reformulate in equations as follows:

$$Dist^{c}_{local} = \sum_{i=1}^{n} \Big[(\parallel d_i - \phi(\mathrm{NN}_c(d_i)) \parallel^2 - \parallel d_i - \mathrm{N}_{k+1}(d_i) \parallel^2) \Big], \quad (3.16)$$

$$\hat{C} = \underset{C}{\operatorname{argmin}} \left(Dist^{c}_{local} \right), \quad (3.17)$$

where

$$\phi(\mathrm{NN}_c(d_i)) = \begin{cases} \mathrm{NN}_c(d_i) & \text{if } \mathrm{NN}_c(d_i) \leq \mathrm{N}_{k+1}(d_i) \\ \mathrm{N}_{k+1}(d_i) & \text{if } \mathrm{NN}_c(d_i) > \mathrm{N}_{k+1}(d_i), \end{cases}$$

and $\mathrm{N}_{k+1}(d_i)$ is the neighbour $(k+1)$ of d_i.

One search index is created for all the classes using the kd-trees implementation provided by the FLANN (Fast Library for Approximate Nearest

Neighbours) library [68] to have efficient nearest-neighbour search. Then the closest 10 neighbours (the parameter value is determined experimentally by [14] and confirmed by all of our experiments with handwriting images) are retrieved for each descriptor in the query handwriting image. As in [14], we used the distance to the $k + 1$ nearest neighbours ($k = 10$) as a ”background distance” to estimate the distances of classes which were not found in the k nearest neighbours.

In order to avoid the matching of descriptors with different keypoint orientations, we neglected any match between descriptors with a keypoint orientation difference larger than a pre-defined threshold by adding a matching condition; see Subsection 3.3. Then we normalise the total class distance by using the number of keypoints for each class in order to cope with the problem of unbalanced data; see Subsection 3.4.

3.3 Orientation Threshold

Typically, handwriting patterns yield many keypoints with similar features but different orientations. As the keypoint orientation of features is a characteristic of the writing style of specific writers, the orientation is a discriminative property of these features. In order to match only features with similar orientation, we propose the following matching condition:

$$|Ort_{kpt1} - Ort_{kpt2}| \leq T_r, \tag{3.18}$$

where Ort_{kpt1} and Ort_{kpt2} are the orientations of two keypoints (in degrees) which features to be matched, and T_r is the orientation-difference threshold.

In other words, features with orientation differences larger than a pre-defined threshold are not considered as valid for a match. The orientation-difference threshold can be estimated from the amount of rotation in handwriting due to line-skew or image rotation, which can be calculated automatically using run-length (or any other) skew-estimation method. From both considerations and the result of the tests with a challenging dataset shown in Fig 3.9, where the best identification rate can be obtained from a 10 to 13 degrees difference, we were able to fix the value of this parameter to 10 degrees in all of our experiments.

Note that this matching condition is not used for FAST keypoints, as the original work in [58], that we use, does not calculate any orientations for the detected keypoints. Therefore, the SIFT descriptor is calculated for the detected FAST keypoints without any rotation for the described local region.

The plot in Fig. 3.9 shows the impact of our matching condition in Eq. 3.18 on the identification rate. The identification rate is defined as the ratio of correctly identified samples over the total number of samples.

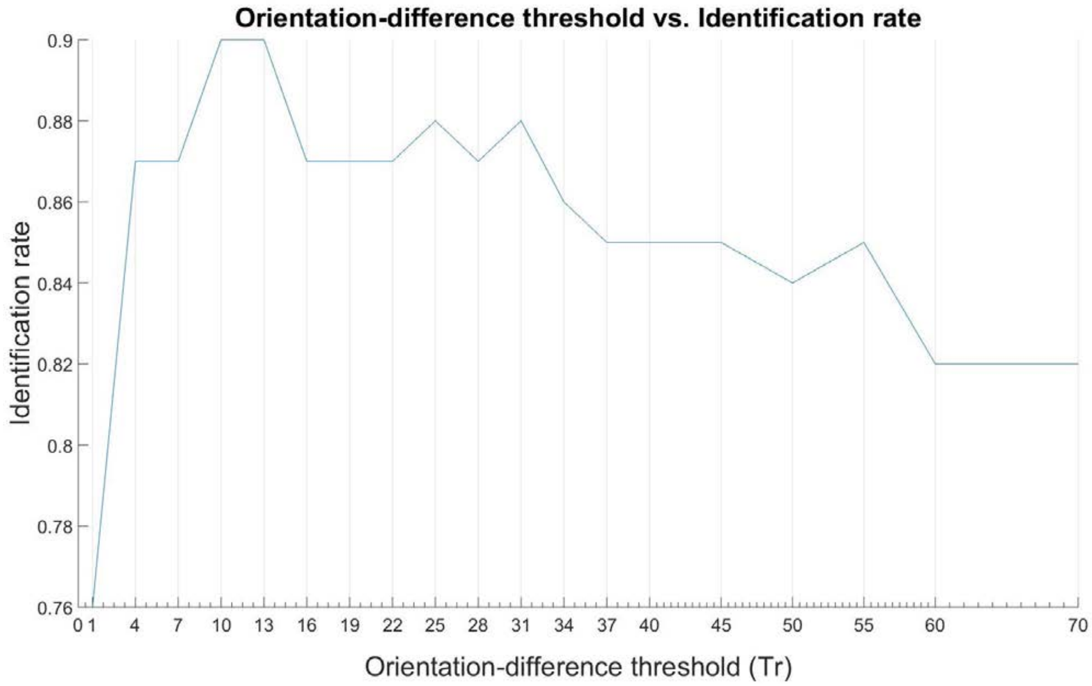


Figure 3.9: The identification rate versus orientation-difference threshold using SIFT keypoints. The validation set from ICFHR-2016 competition of writer identification, task 1A [66] is used for this test. Details and Sample images of this dataset are provided in Chapter 4, Fig. 4.6.

3.4 Class Distance Normalisation

Data sets like samples of different handwriting styles are considered as unbalanced when at least one class is represented by only a small number of samples. Typically in the case of writer identification, the labelled samples are not equally distributed among the writers (classes) in many practical scenarios. One of the main limitations of NBNN-based methods is the bias towards classes with a large number of keypoints; this limitation can reduce the identification rate significantly in the case of unbalanced data. Therefore, we normalise the final distance of each class $Dist^c_{local}$ in equation 3.17

by the number of keypoints in the respective class:

$$\hat{C} = \underset{C}{\operatorname{argmin}} \left(\frac{Dist^{c}_{local}}{K_c} \right), \tag{3.19}$$

where K_c is the number of keypoints for each class c.

In order to demonstrate the impact of the proposed normalisation, we measured the identification rate while we reduce the number of samples per writer for half of the dataset. We used the ICDAR-2011 dataset for musical scores [69] due to the fact that this dataset has a large number of samples (10 samples) per writer for testing; see Chapter 4, Section 4.3.1 for details. Sample images of this dataset are shown in Fig. 4.2. The graphs of these experiments in Fig. 3.10 markedly show the positive effect of the normalisation: The identification rate drops much slower with normalised class distance as the difference between the number of samples per writer increases.

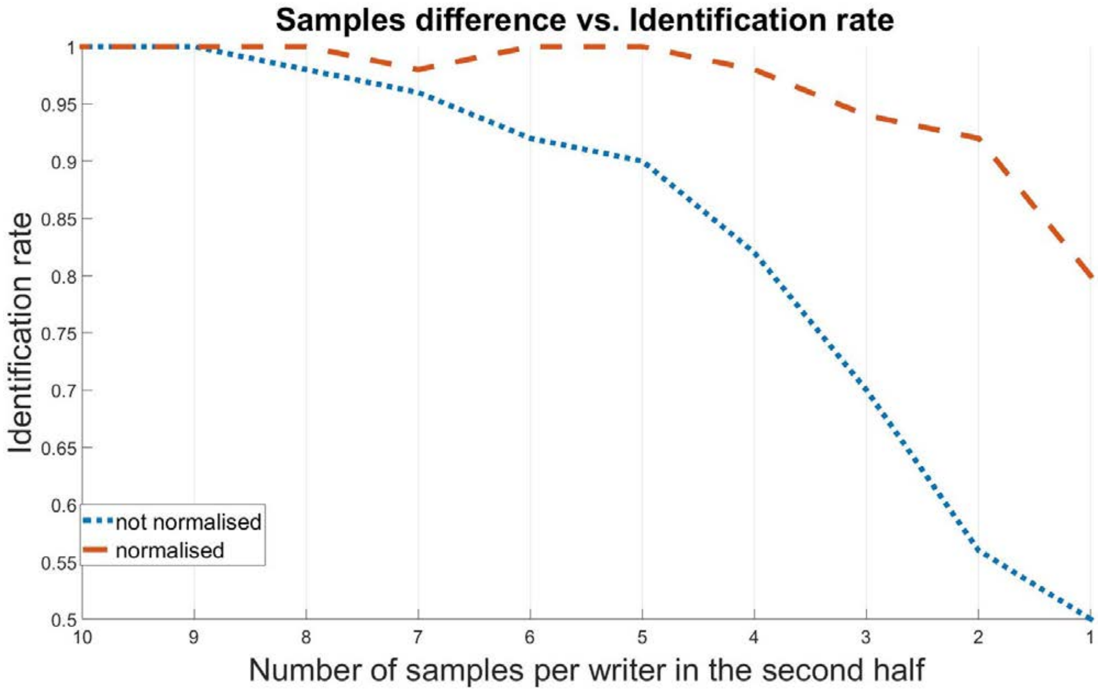


Figure 3.10: Comparison between the identification rate with and without normalisation using SIFT keypoints. 10 samples for each of 50 writers are used for the test from the ICDAR-2011 dataset for musical scores [69]. The number of samples for the randomised half of the writers is fixed, while we decrement the number of samples for the other half from 10 to 1. The x-axis represents the number of samples per writer for the second half of writers.

The superiority in performance of the Local NBNN classifier over the NBNN classifier is confirmed in this work using handwriting images as well for both SIFT and FAST keypoints. The orientation-difference threshold and the normalisation are applied to both classifiers; see Table 3.2.

Table 3.2: Comparison between the identification rate of Normalised NBNN and Normalised Local NBNN with the orientation-difference threshold using SIFT and FAST keypoints detection algorithms. The dataset from ICFHR-2016 competition of writer identification, task 1A [66] is used for this test; see Section 4.3.1 for details of this dataset. Sample images of this dataset are shown in Chapter 4, Fig. 4.6.

Classifier	SIFT	FAST
Normalised NBNN with orientation-difference threshold	85%	97%
Normalised Local NBNN with orientation-difference threshold	97%	100%

3.5 Conclusion

We present an improved Local NBNN classification method for the task of writer identification given small sets of unbalanced sample data. The orientations of SIFT keypoints are used to restrict the matching between descriptors to only those with similar orientations. Distances to classes are normalised by the number of keypoints for each class. The method has been tested with several public datasets of different writing systems including musical scores as will be presented in Chapter 4, and state-of-the-art results were obtained in all experiments with a fixed parameter set [70]. The key parameter PCK of FAST keypoint detection algorithm has been analysed and optimised to enhance the performance for historical manuscripts in Section 3.1.2.

Chapter 4

Performance Evaluation and Experimental Results

In this chapter, we evaluate the performance of the proposed method on standard and public datasets of both contemporary and historical handwriting in order to compare it with the state-of-the-art methods for the task of writer identification. These standard datasets neither represent the typical degradation nor the unbalance and scarcity of handwriting samples in digitised manuscripts of the selected sub-projects within the SFB 950. Nevertheless, this evaluation helps in assessing the discriminative power of the proposed method w.r.t. different handwriting styles in the datasets. Moreover, some of these datasets offer handwriting styles from a large number of writers and/or in many different writing systems and script types. Having a high performance for such datasets demonstrates the generality and the scalability (to a large number of classes) of our proposed method.

4.1 Standard and Public Datasets

Several public datasets have been proposed by the community of computational documents analysis for the task of writer identification in recent years, each with its own evaluation procedure and performance measures; see Section 4.2 for details. These datasets contain different character sets, languages, and even musical scores. In Table 4.1 we provide a summary of recent public datasets for the task of writer identification in the last seven years. Some of them are provided through international competitions for the task of writer identification.

Table 4.1: Summary of the used datasets in the evaluation of our proposed method for the task of writer identification. Sample images for these datasets can be found in the corresponding references and in the following sections as well.

Datasets	No. of writers	Total no. of pages	Pages per writer	Offered presentation	Pages per language per writer
ICDAR-2011 [71]	26	208	8	Binary	2 English 2 French 2 German 2 Greek
ICDAR 2011 [69] Musical Scores	50	1000	20	Binary	Musical scores
ICFHR-2012 [72]	100	400	4	Binary	2 English 2 Greek
ICDAR-2013 [73]	250	1000	4	Binary	2 English 2 Greek
CVL [74]	310	1550	5	RGB Colour	4 English 1 German
ICFHR-2016 [66] Task 1A	400	800	2	RGB Colour	2 Arabic
ICFHR-2016 [66] Task 1B	400	800	2	RGB Colour	2 English
ICDAR-2017 Historical-WI [75]	720	3600	5	2 RGB Colour and Binary	5 (Mostly) English

The first international writer identification contest ICDAR-2011 [71] consists of 208 samples written by 26 writers in 4 different languages, the writers are asked to copy the same fixed text. Another dataset has been created from the same samples by keeping only two lines of text from each sample.; see sample images in Fig. 4.1.

The ICDAR 2011 competition for musical scores [69] used sample images from CVC-MUSCIMA database [76], all the 50 writers in this competition dataset are selected to be adult musicians in order to ensure that they have their own characteristic handwriting music style. Furthermore, the provided sample images in this dataset for the task of writer identification are without the staff lines (the straight horizontal lines) in order to ensure that the published results are not dependent on the performance of a particular staff removal technique.

The contest of ICFHR-2012 [72] was created for the task of writer identification with the help of 100 writers that were asked to copy four parts of the text in two languages (English and Greek). These parts of the text were the same for all users. Only the Greek documents were written in the native language of the writer. Following the same criteria of ICFHR-2012 dataset, the competition dataset of ICDAR-2013 has been created with the help of 250 writers.

In contrast to the aforementioned datasets, CVL dataset [74] consists of RGB colour images of 300 dpi instead of binary images. Moreover, 310 writers were asked to copy German and English texts which have been chosen from literary works. This dataset offers ground-truth for word spotting as well.

The dataset of ICFHR-2016 competition [66] is based on the QUWI database [77] for Arabic and English offline handwritings. This handwriting database consists of RGB colour images of 300 dpi. This competition comprises twelve different tasks, only two of them are relevant to our research in this dissertation, namely Task 1B and Task 1A: Task 1A is writer identification on Arabic handwritings, and Task 1B is writer identification on English handwriting.

Only recently, the ICDAR-2017 Historical-WI [75] dataset has been published using handwritten historical manuscripts. The image samples of this dataset have been taken from the digital archive of the Universitätsbibliothek Basel (`http://www.e-manuscripta.ch/`) which contains manuscript samples originating from the 13^{th} to the 20^{th} century. This dataset consists of colour as well as binary images of 300 dpi.

4.2 Evaluation Criteria

Evaluation procedures w.r.t. the task of writer identification used in competitions from 2011 to 2017 for public datasets are:

- **Leave-one-out:** Each image in the dataset searches for the best match within the other images in the dataset.
- **Training set and test set:** The dataset is divided into two sets, a training set and a test set. For each image in the test set, the best match is searched within the images of the training set.

The used performance measures for writer identification methods are:

- **Identification rate:** the ratio of correctly identified writers over all identifications in a given dataset. This value is typically calculated in the following ways:

- **Soft TopN:** An identification is considered as correct when at least one document image of the same writer is included in the N most similar document images.
- **Hard TopN:** An identification is considered as correct when all N most similar document images are written by the same writer.
- **Training set and test set - Top1:** An identification is considered as correct when the most similar document image in the test set is from the same writer of the query document image in the training set.

- **mean Average Precision (mAP):** It is one of the standard evaluation metrics for information retrieval which takes into account both the precision measure and the ranking of retrieved samples as follows: Let $P(k)$ be the precision of the method in retrieving samples relevant to a query sample from k samples in a dataset, and Rel_k is the number of samples relevant to the query from k retrieved samples, then

$$P(k) = \frac{Rel_k}{k}. \tag{4.1}$$

If n is the total number of retrieved samples and Rel_{total} is the total number of samples relevant to a query in a dataset, then the average precision Ave_P can be calculated as follows:

$$Ave_P = \frac{\sum_{k=1}^{n} [P(k) Rel_k]}{Rel_{total}}, \tag{4.2}$$

and the mean average precision mAP is calculated as follows:

$$mAP = \frac{\sum_{q=1}^{Q} [Ave_P(q)]}{Q}, \tag{4.3}$$

where q is the current query sample and Q is the total number of queries.

The proposed method in this dissertation is a learning-free method that does not require training data; therefore, we can carry out tests under the aforementioned evaluation procedures.

4.3 Experimental Results

We evaluated our method on several public datasets with different character sets, languages, and even musical scores to demonstrate the generality of the proposed approach. The samples of ICDAR-2011 [71], ICDAR-2011 for musical scores [69], ICFHR-2012 [72] and ICDAR-2013 [73] datasets are binarised, while the samples in CVL [74], ICFHR-2016 [66] and ICDAR-2017 Historical-WI [75] datasets are given in RGB format. Important properties of these datasets are the variation of the number of writers (from 26 to 720), the variation of the number of pages per writer (from 2 to 20), and the variation of the amount of handwritten text per page (from only two lines to a full page).

4.3.1 Contemporary Datasets

A comparison with the state-of-the-art results is presented for each dataset separately. It is important to note that the method's parameters were kept constant for all experiments as follows: For SIFT keypoints, the orientation-difference threshold is 10 and all other parameters are as described in Section 3.1.1; for FAST keypoints, PCK is set to 5%; see Section 3.1.2 for details.

We followed the exact evaluation criteria for each dataset to provide a fair comparison (each evaluation criteria is mentioned in the table of the corresponding dataset). Results with different evaluation criteria are not considered; for example, the parse radial sampling of Local Binary Patterns (SRS-LBP) method [25] partitioned the datasets and used the average performance of the cross-validation for each partition, whereas the contour-Zernike method [78] partitions ICDAR-2013 and CVL datasets into training and test sets.

Although we propose a segmentation-free method (see the justification in Chapter 2), we considered segmentation-based methods (e.g. Fisher Vector method [50]) for the comparison as well (the methods are denoted in the tables); see Tables 4.2, 4.1, 4.3, 4.4, 4.5, 4.6 and 4.7. All the results we present in these tables are for the Normalised Local NBNN with orientation threshold, unless stated otherwise.

In Table 4.7, we present the official result of our participation in the

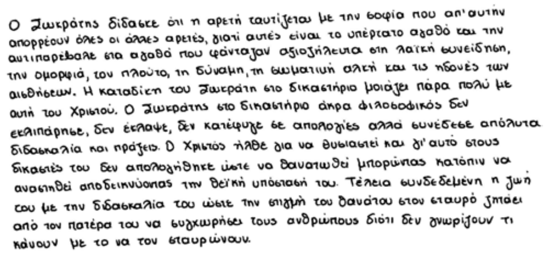
ICFHR-2016 competition [66] with SIFT keypoints but without normalisation, because this part of our method was then not developed; see Section 3.4. The results of our method using SIFT keypoints and FAST keypoints with normalisation are presented in the table as well.

Since a large number of keypoints per writer (class) is needed for reliable nearest neighbour search in our method (see details in Chapter 2), it is expected that identifying writers (classes) represented by a small number of keypoints will be less accurate.

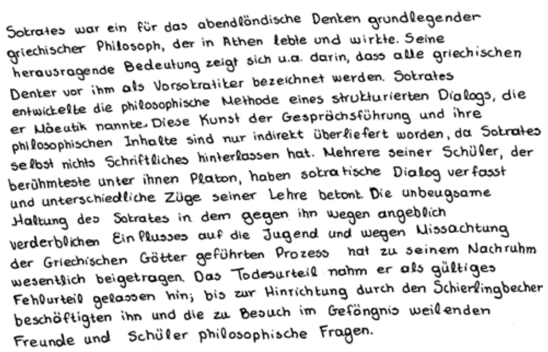
Although the number of samples is the same for all writers in ICFHR-2016 competition, the amount of handwritten text varies significantly between the samples; see Fig. 4.6. This results in a varying number of detected keypoints between different samples, and leads to an unbalanced representation of classes in feature space. Therefore, the normalisation step has a larger positive impact in such cases. Furthermore, a very high identification rate is obtained for the CVL dataset despite the large number of classes (writers) which clearly shows the scalability of our method to a high number of classes.

Method	Identification Rate Full / Two lines	Dataset details
Proposed Method using SIFT keypoints	**100**/96.6	26 writer 208 pages 8 pages per writer (2 English, 2 French, 2 German, 2 Greek) Leave-one-out Top-1
Proposed Method using FAST keypoints	**100/98.6**	
TSINGHUA [71] 1st in competition	99.5/90.9	
CS-UMD [71] 2nd in competition	99.5/66.8	
MCS-NUST [71] 3nd in competition	99.0/82.2	
Lehigh [24]	97.1/—	

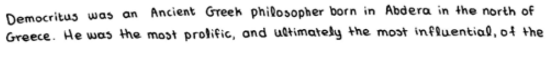
Table 4.2: ICDAR-2011 [71], using full text / using only two lines per image. See Fig. 4.1 for sample images.



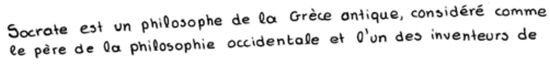
(a) First sample, full text.



(b) Second sample, full text.



(c) Third sample, two lines.

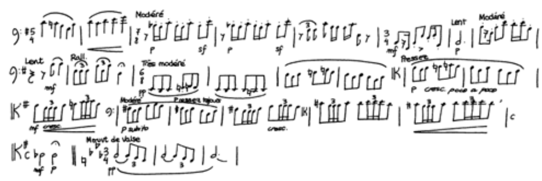


(d) Fourth sample, two lines.

Figure 4.1: Samples from ICDAR-2011 dataset.

Method	Identification Rate	Dataset details
Proposed Method using SIFT keypoints	98.2	50 writer 1000 pages 20 pages per writer Musical scores Training and Test sets Top1
Proposed Method using FAST keypoints	99.4	
PRIP02-combination [69] 1st in competition	77	
TUA03-SVMOAA [69] 2nd in competition	76.6	
Fisher Vector [50]	**99.5** Segmentation-based	

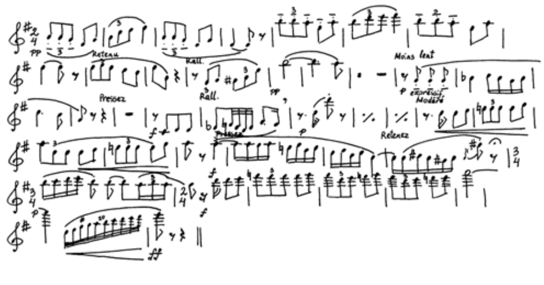
Table 4.3: ICDAR-2011 for musical scores [69], see Fig. 4.2 for sample image.



(a) First sample.

(b) Second sample.

(c) Third sample.

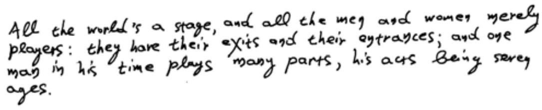


(d) Fourth sample.

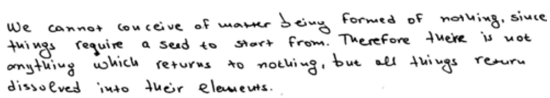
Figure 4.2: Sample from ICDAR-2011, musical scores.

Method	Identification Rate	Dataset details
Proposed Method using SIFT keypoints	96	100 writer 400 pages 4 pages per writer (2 English, 2 Greek) Leave-one-out Top-1
Proposed Method using FAST keypoints	**98.8**	
TEBESSA-c 1st in competition [72]	94.5	
TSINGHUA 2nd in competition [72]	92.8	
SIFT+Contour-directional [44]	96.8	

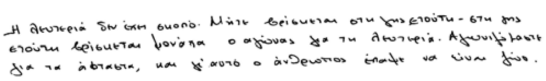
Table 4.4: ICFHR-2012 [72], see Fig. 4.3 for sample image.



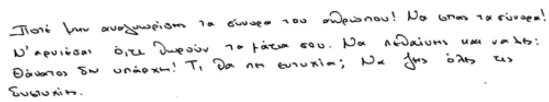
(a) First sample.



(b) Second sample.



(c) Third sample.

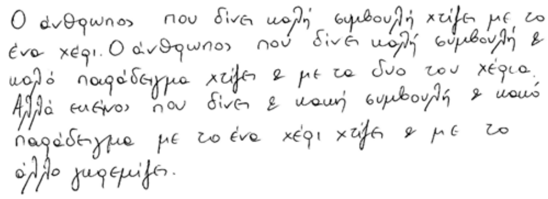


(d) Fourth sample.

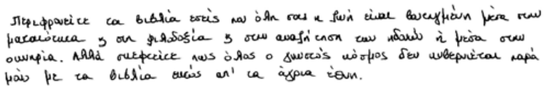
Figure 4.3: Sample from ICFHR-2012.

Method	Identification Rate	Dataset details
Proposed Method using SIFT keypoints	92.4	250 writer 1000 pages 4 pages per writer (2 English, 2 Greek) Leave-one-out Top-1
Proposed Method using FAST keypoints	**97.9**	
CS-UMD-a 1st in competition [73]	95.1	
CS-UMD-b 2nd in competition [73]	95	
SIFT+Contour-directional [44]	96.2	
SRS-LBP metric [25]	96.9	

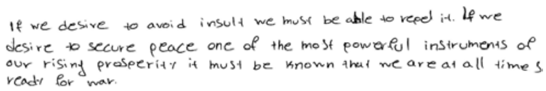
Table 4.5: ICDAR-2013 [73], see Fig. 4.4 for sample image.



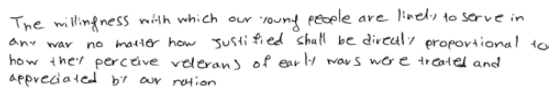
(a) First sample.



(b) Second sample.



(c) Third sample.

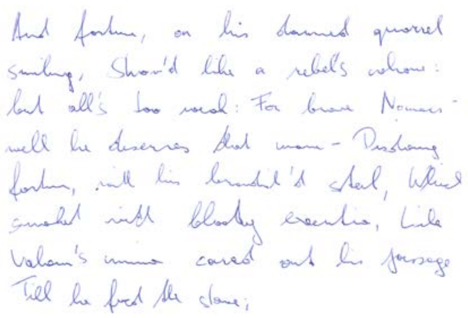


(d) Fourth sample.

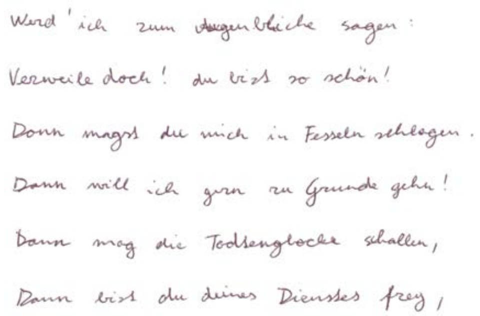
Figure 4.4: Sample from ICDAR-2013.

Method	Identification Rate	Dataset details
Proposed Method using SIFT keypoints	99.3	311 writer 1609 pages 5 pages per writer English Leave-one-out Top-1
Proposed Method using FAST keypoints	**99.8**	
CS-UMD 1st in competition [74]	97.9	
TSINGHUA 2nd in competition [74]	97.7	
SRS-LBP metric [25]	98.6	

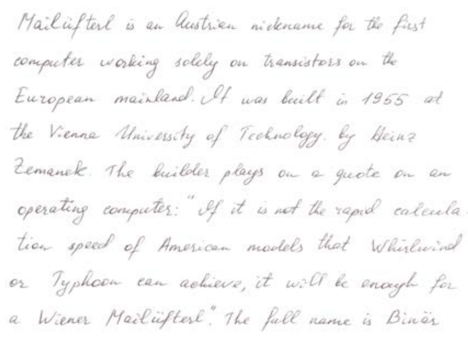
Table 4.6: CVL [74], see Fig. 4.5 for sample image.



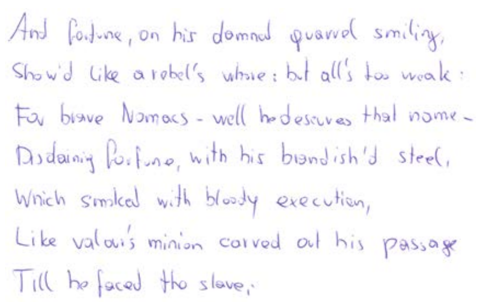
(a) First sample.



(b) Second sample.



(c) Third sample.



(d) Fourth sample.

Figure 4.5: Sample from CVL.

Method	Identification Rate 1A/1B	Dataset details
Proposed Method using SIFT keypoints [66] but without normalisation	90.33/87.67	400 writer 800 pages 2 pages per writer (2 Arabic / 2 English) Training and Test sets Top1
Proposed Method using SIFT keypoints with normalisation	91.67/87.67	
Proposed Method using FAST keypoints	**99.7/97.7**	
Nuremberg [66]	89.33/84.67	
CVC [66]	80.67/80.33	

Table 4.7: ICFHR-2016 competition, tasks 1A and 1B [66], see Fig. 4.6 for sample images.

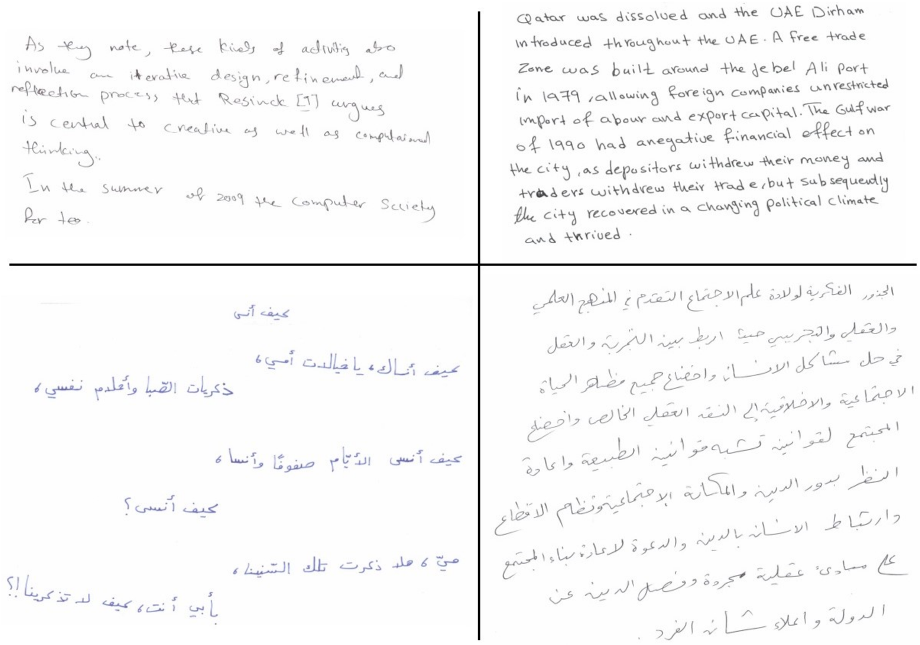


Figure 4.6: These four samples show that the amount of handwritten text varies significantly between the samples in ICFHR-2016 dataset.

4.3.2 Historical Dataset

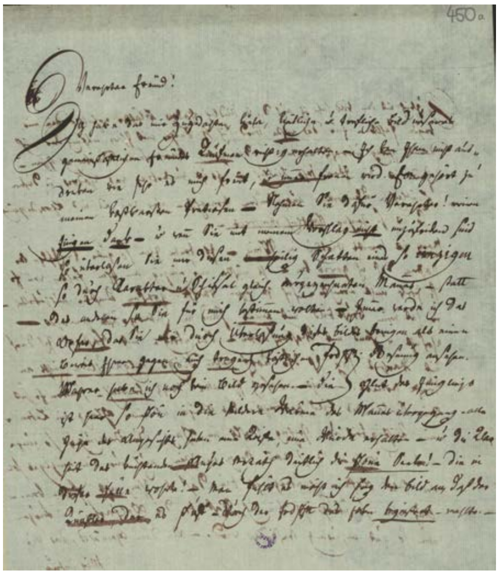
As shown in the previous section, several public datasets for contemporary handwritings have been made available. Hence we were forced to use them for our first set of experiments. However, there was no public dataset for historical manuscripts available until recently, when the dataset for the writer identification Historical-WI competition [75] was published in ICDAR-2017 conference.

This dataset mostly contains samples of English language, but also some of other languages (e.g. Greek and Latin) because it has been selected randomly (was created from the digital archive of the Universitätsbibliothek Basel from 13th to 20th century `https://www.e-manuscripta.ch/`) using an automated algorithm [75] rather than by scholars from the Humanities. Hence no research problems and no work-flow of scholars are reflected in this dataset. Nevertheless, we evaluated our method (the Normalised Local NBNN classifier) on this dataset using FAST keypoints with $PCK = 0.05$ and submitted our results' file to the competition server. It can be seen from the results in Table 4.8 that the Normalised Local NBNN method significantly outperforms the winner method of the competition. Only the methods that follow the criteria of the competition [75] are mentioned in the comparison table.

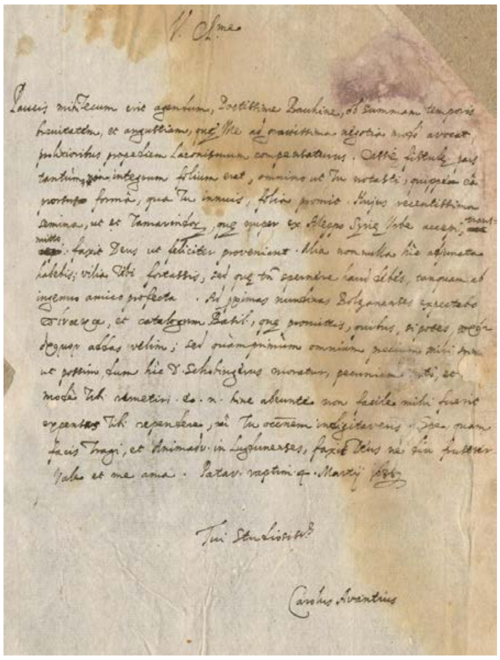
Though only the recently published method in [33] achieved better results (Top-1: 88.9, mAP: 76.2), these results have been obtained by using the validation set of the competition (which is provided for performance estimation only) and labelled as negative samples to train SVM classifier. Therefore, it is not considered in our state-of-the-art results comparison as it used these data samples for training purposes. Furthermore, their method incorporates a complicated procedure which requires preprocessing steps and training of a deep residual net. Therefore, it is not a practical candidate for an efficient (in terms of processing power and memory space) and easy-to-use software tool, especially if there were only a few lines of handwritings per writer available with no additional training samples from the same data domain like the case in many research questions of scholars of manuscript research from Humanities within the SFB 950.

Method	**Identification Rate**	**mAP**	**Dataset details**
Proposed Method using FAST keypoints	**85.6**	**68.3**	720 writers 3600 pages, 5 pages per writer (Mostly English) Leave-one-out, mAP and Top-1
Tebessa II [75] 1st in competition	76.4	55.6	
Groningen [75] 2nd in competition	76.1	54.2	
Tebessa I [75] 3rd in competition	74.4	52.5	

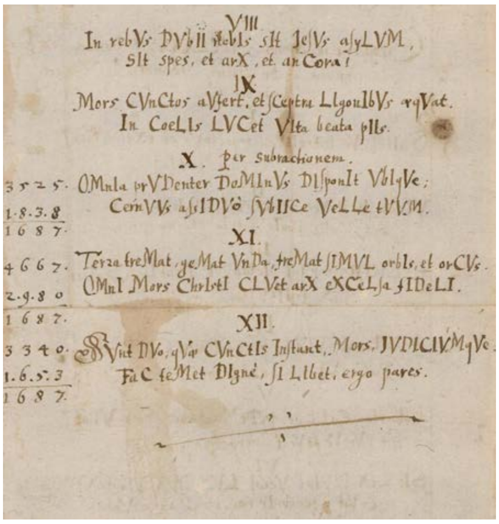
Table 4.8: ICDAR-2017 WI-Historical competition [75], see Fig. 4.8 for sample images.



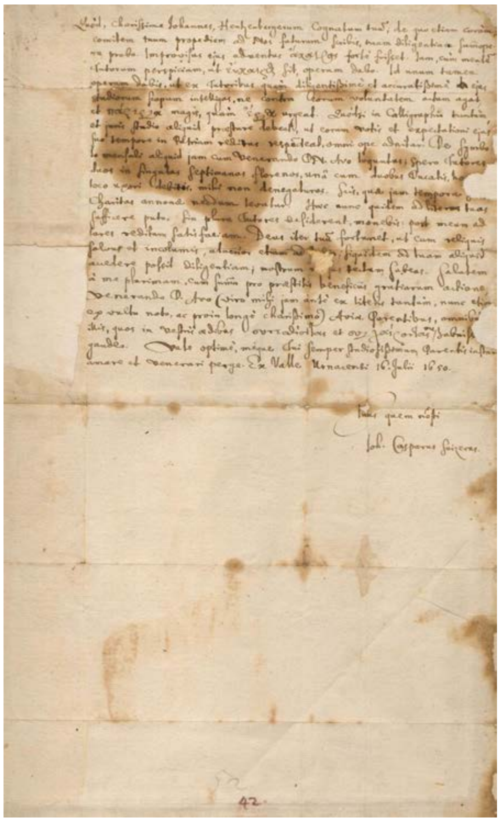
(a) First sample.



(b) Second sample.



(c) Third sample.



(d) Fourth sample.

Figure 4.7: Samples from ICDAR-2017, historical dataset.

4.4 Conclusion

Our proposed method has been tested with several public datasets of different writing systems including musical scores. State-of-the-art results were obtained in all experiments with a fixed parameter set [70]. This evaluation demonstrated the discriminative power of the proposed method w.r.t. different handwriting styles (or respectively writers) in the standard and public datasets, both contemporary and historical.

In addition, some of these standard datasets offer handwriting styles from a large number of writers and/or in many different writing systems and script types. Therefore, the high performance of our proposed method in all these datasets demonstrated the generality over several writing systems and the scalability to a high number of classes (writers). In the following chapter, we analyse the proposed method w.r.t. typical degradation in digitised manuscripts.

Chapter 5

Performance Analysis w.r.t. Degradation Types in Digitised Manuscripts

Degradation in digitised manuscripts results from e.g. the poor preservation conditions, the used materials, or even from the digitisation process itself. Such degradation has a negative impact on the performance of computational methods. In addition, the amount of available handwritten text in historical manuscripts vary greatly between different scribes from a couple of lines via few pages up to several manuscripts.

In order to measure the impact of some degradation types on the identification rate, we analyse the proposed method using systematically generated degradation on digitised manuscripts. The selection of degradation types used in this analysis is based on their prevalence in digitised manuscripts from sub-projects within the SFB 950 and their direct influence on parameter selection of the proposed method. The results of this analysis can be used to better define the required quality of the images in order for the method to provide reliable results.

5.1 Data Selection for the Analysis

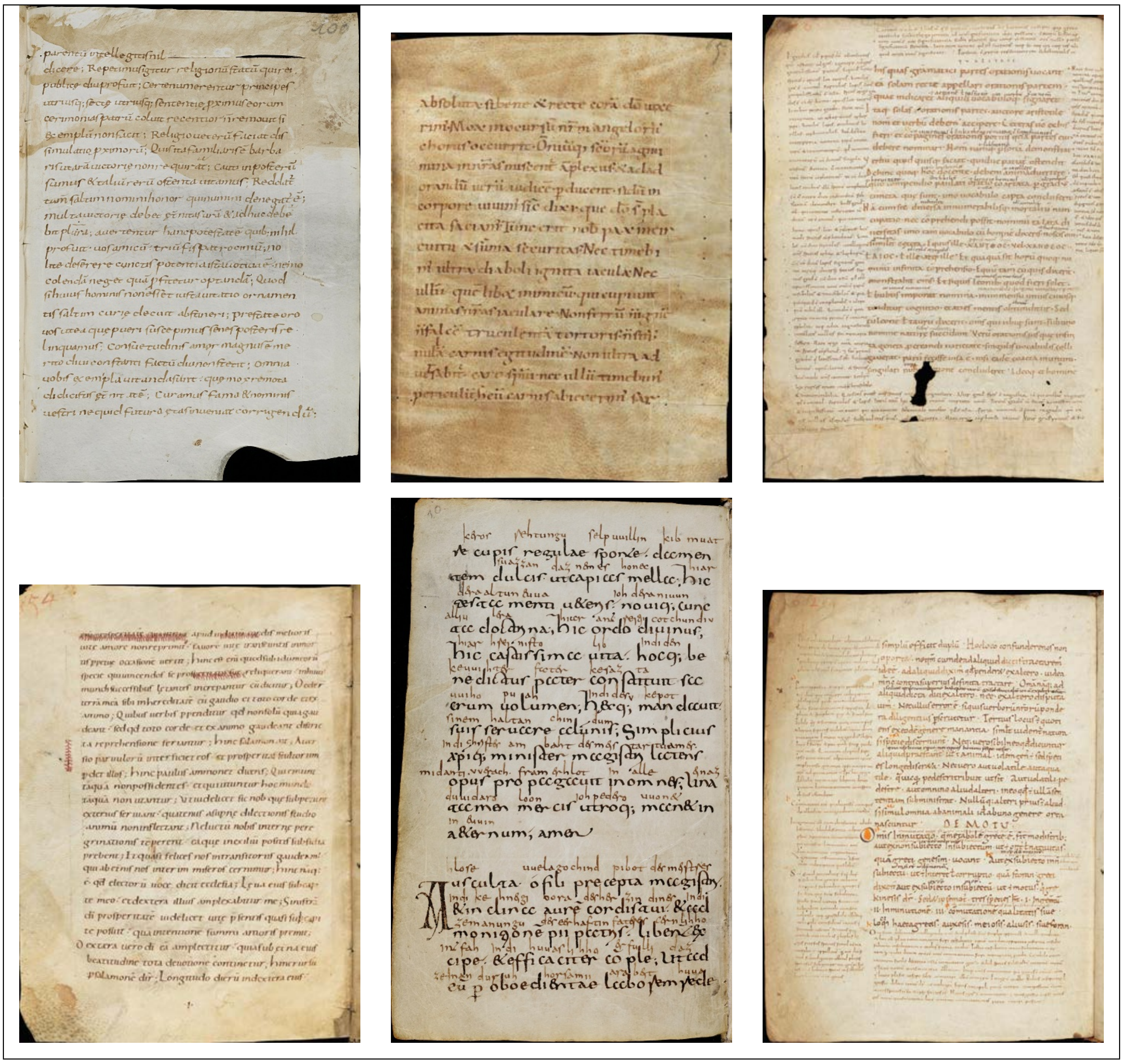
Acquiring confirmed ground truth for historical manuscripts is a critical issue, because in many cases the name of the scribes (which are given in the so-called "colophons") are missing, incomplete, or doubtful. Therefore, preparing an evaluation dataset should be done jointly with scholars from manuscript research in Humanities. After having prepared a validation dataset with confirmed ground truth, we can better tune the proposed method to work best with a certain domain-specific dataset (e.g. a certain writing school, writing style, writing material, etc.), and we will have a clear indication for the performance of the proposed method within that domain. In this chapter, we focus on the Carolingian Minuscule script because of its relevance to the digitised manuscripts of sub-project C08 [10] in the SFB 950.

In addition, selecting representative data samples from a certain domain is an important decision to be made when creating an evaluation dataset; otherwise, it will be hard to provide a realistic and quantitative estimation of the method's performance using data samples from that given domain. Therefore, sample pages are selected from different parts of the same manuscripts and scribe (e.g. begin, middle and end section). In addition, samples with different visual features are selected; these differences include e.g. ink, writing material, digitisation process, and degradation level.

5.2 St. Gall Sub-Set

Taking into account the aforementioned considerations, 100 pages from the "Stiftsbibliothek" library of St. Gall collection [67] have been selected for the analysis in this chapter: 10 scribes, 10 pages per scribe; see Table 5.1. The selection has been done jointly with the scholars in the SFB sub-project C08 [10] "East Frankish manuscripts with collections of formulas", namely Philippe Depreux and Till Hennings. This selection is based on the following reasons:

Table 5.1: Samples from St. Gall dataset [67].



- Both the manuscripts under research in the sub-project and the St. Gall sub-set share the same script type, writing material, and production period. They are both Latin script, more specifically the Carolingian Minuscule script type from the 9^{th} century using ink on a parchment.
- All handwritings in the selected sub-set share the same script type of Carolingian Minuscule. This makes it more interesting and challenging

than distinguishing between different script types like Anglo-Saxon and Gothic script.

- The scribe identity of each hand has been already confirmed previously by palaeographers and cataloguers; furthermore, the samples of different scribes are set apart in time (ca. 750 - 950 B.C.E.). Hence ground truth is available.

- The St. Gall library collection contains high-quality images (RGB colour images, 300 DPI, the spatial resolution ranges from 1875 x 3290 to 3833 x 5055 pixels (with an average of 2896 x 4192)). This high resolution is required to allow us to analyse the impact of reducing the resolution systematically.

- The images in this library collection are under a free license of access and usage for research and education purposes.

5.3 Analysis Procedure

The matching algorithm of the proposed method (Normalised Local NBNN) performs a one direction NN search, meaning that each descriptor from the query sample image searches for nearest neighbours in the labelled sample images. In other words, the descriptors from the labelled sample images do not search in the opposite direction for nearest neighbours in the query sample image. This means also that query descriptors from image parts which are irrelevant to the handwritings in question are forced to have matches (nearest neighbours) in the labelled images; this can have a negative impact on the identification rate. Furthermore, more than one query descriptor may be matched to the same labelled descriptor as a nearest neighbour (NN).

On the other hand, the descriptors of labelled images are matched only to query descriptors if they are the nearest neighbours of a given query descriptor; otherwise, they will never be considered.

In addition, the computational description of keypoints in local image regions is different in many cases under different resolution and contrast. This implies that the overall difference in contrast or resolution between the query images and the labelled images can have a large negative impact on the performance of our method.

In order to consider the aforementioned issues, the dataset (100 pages of 10 scribes) is split into an unlabelled set and a labelled set; 50 pages are assigned to each set. Degradation types have been applied in three different scenarios for both resolution and contrast analysis: to all 100 samples (pages), to 50 test samples only, and to 50 labelled samples only. Finally, identification rates are compared between the mentioned scenarios. Since the goal of these tests is to compare the results for different degradation levels, randomised image selection is avoided to ensure valid comparison. The images have been selected, jointly with the aforementioned scholars, so that both sets contain as similar quality and degradation level as possible to avoid getting biased results.

5.4 The Selected Degradation Types

The considered degradation types in this analysis were selected based on their prevalence in digitised manuscripts and their direct influence on parameter selection of our proposed method. In our experiments, these degradation types, namely varying image resolution, sample rotation and image contrast, and impact of irrelevant information in samples, are systematically produced and controlled, and they are selected to be relevant to the possible degradation types found in historical manuscripts from the selected sub-projects within the SFB 950; see Figs. 1.2, 1.3, 1.4, and 1.5.

5.4.1 Resolution

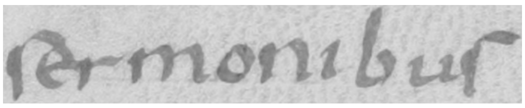
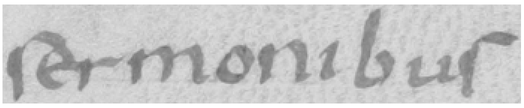
The resolution in historical manuscripts differs between samples due to the different settings of the digitisation process and acquisition equipments. Although it does not necessarily impose any difficulties for palaeographers in most cases, the resulting image resolution can have a significant impact on computational methods.

In order to investigate the impact of resolution on the identification rate, we systematically reduce the resolution of the images and recalculate the identification rate as follows:

Let $I_{N,M}$ be the image with the original resolution (N x M) and $\bar{I}_{K,L}$ the image with reduced resolution (K x L), where $K = N$ x r and $L = M$ x r. The value of the decimation factor r starts with 1.0 and is decremented iteratively by 0.1 until no more keypoints can be detected. The case of $r = 0.05$ is selected manually to investigate the performance limits of our method. The pixel intensities of the resulting image are determined by simply averaging the pixel intensities of the neighbouring pixels; see the figures in Table. 5.2.

For this test, we set PCK (Percentage of Considered Keypoints) to 1 for the keypoints detected in the original resolution of the images. In order to prevent the influence of resolution reduction on the number of detected keypoints, the number of considered keypoints in the original resolution for every image is stored and used for all values of r.

Table 5.2: A sample from St. Gall dataset with different values of the decimation factor r.

r = 1.0	**r = 0.5**
Resolution = 691 x 130	**Resolution = 347 x 66**
	
r = 0.1	**r = 0.05 (manually selected)**
Resolution = 70 x 15	**Resolution = 37 x 9**
	

The results in Fig. 5.1 and Fig. 5.2 show that whether we use SIFT or FAST keypoints, the identification rate drops much slower with respect to r if the resolution of both test and labelled images is decimated by the same amount (same value of r). As yet, given our experimental data, the reason is assumed to be due to the fact that the calculated descriptor of the same local image region varies under different resolutions. In addition, the identification rate drops slower in all scenarios when using SIFT keypoints, a possible reason for this is the scale-invariance property of SIFT keypoints.

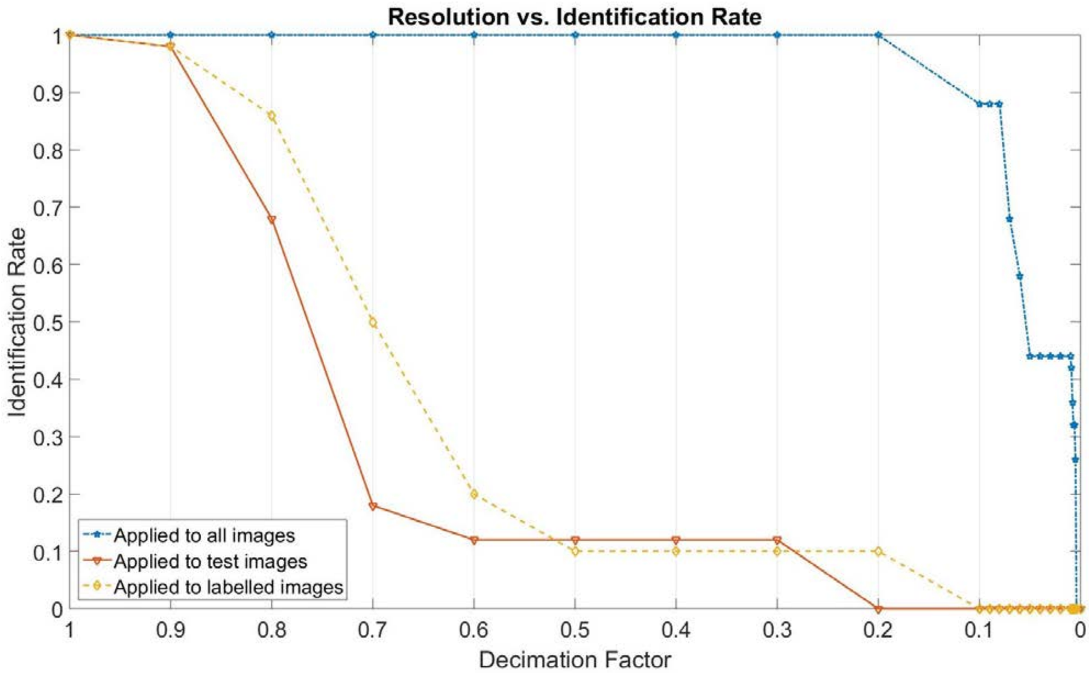


Figure 5.1: Resolution vs. Identification Rate using FAST keypoints.

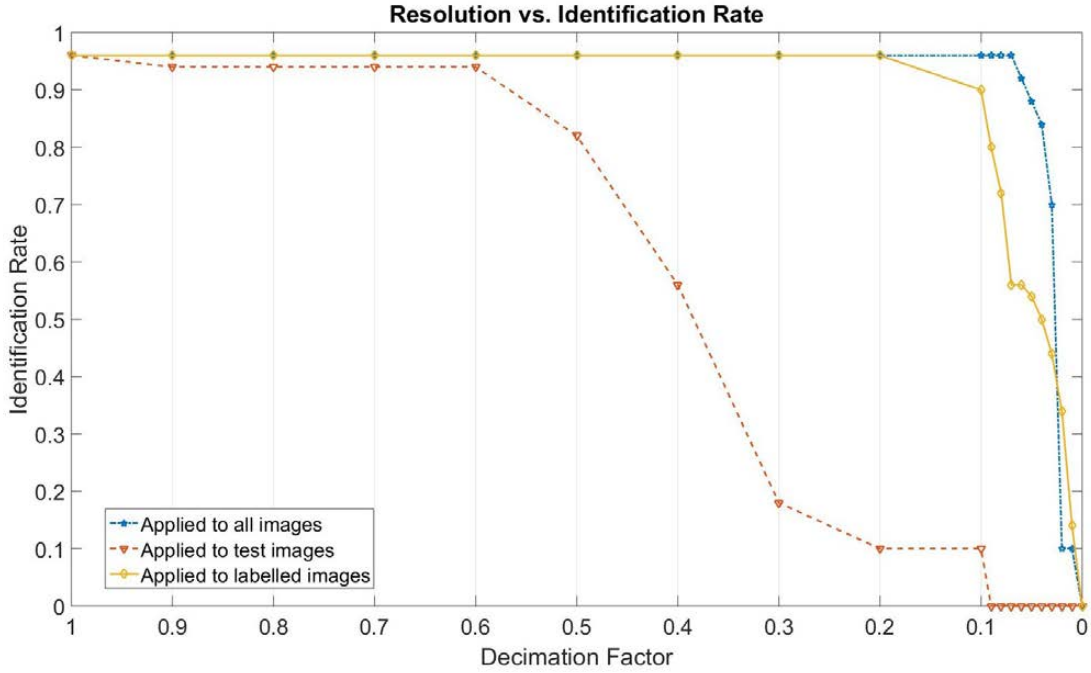


Figure 5.2: Resolution vs. Identification Rate using SIFT keypoints.

5.4.2 Contrast

The contrast in a digital image can be perceived in a small local area from the difference between the parts with high- and low-intensity values, where less contrast gives a ”flatter” image, and more contrast gives a ”deeper” image. There are many other definitions of contrast in literature, such as the difference in visual properties that makes an object distinguishable or simply the difference in intensities from pixel to pixel [79], e.g. at edges. In this dissertation, contrast refers to the global difference between the maximum and minimum pixel intensity in an image. Therefore, we use the term *contrast* in this chapter to refer to the global contrast defined above.

Having a low contrast is a very common problem in historical manuscripts due to degradation factors like the preservation conditions and the type of writing materials. The following test has been conducted to analyse the impact of reducing the contrast on the identification rate:

1. The selected image samples of St. Gall library are converted to grey values as described in Chapter 3.

2. Given an image I, let the lowest grey value of that image be I_{low}.

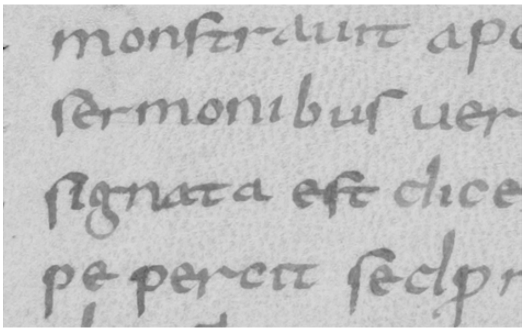
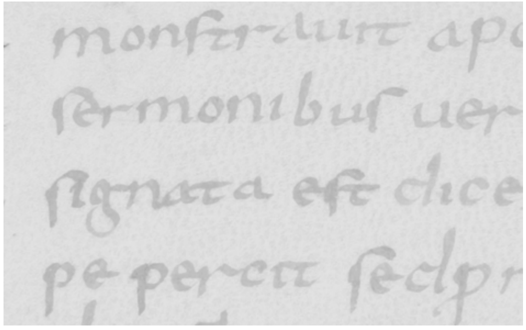
3. All image grey values are shifted so that the minimum value equals zero (for better visualisation, for easier visual inspection, and for simpler mathematical formulation). This can be accomplished by setting the value of I_v to $I_v - I_{low}$, where I_v is any given grey value in the image I.

4. The highest grey value of all the images is determined to be used as the starting Maximum Contrast Threshold (MCT).

5. In each iteration, MCT is systematically reduced as follows: If the highest grey value I_M in any image is smaller than the MCT, then nothing is done; otherwise, grey values in that image are normalised to the range between zero and the current MCT as follows: $\bar{I}_v = I_v$ x $\dfrac{MCT}{I_M}$; see Table 5.3.

6. Finally, the identification rate is calculated.

Table 5.3: A sample from St. Gall dataset with different values of the contrast threshold (MCT). For better visualisation, we used the intensity value of zero for white and 255 for black.

MCT = 234	**MCT = 100**
	
MCT = 50	**MCT = 15**

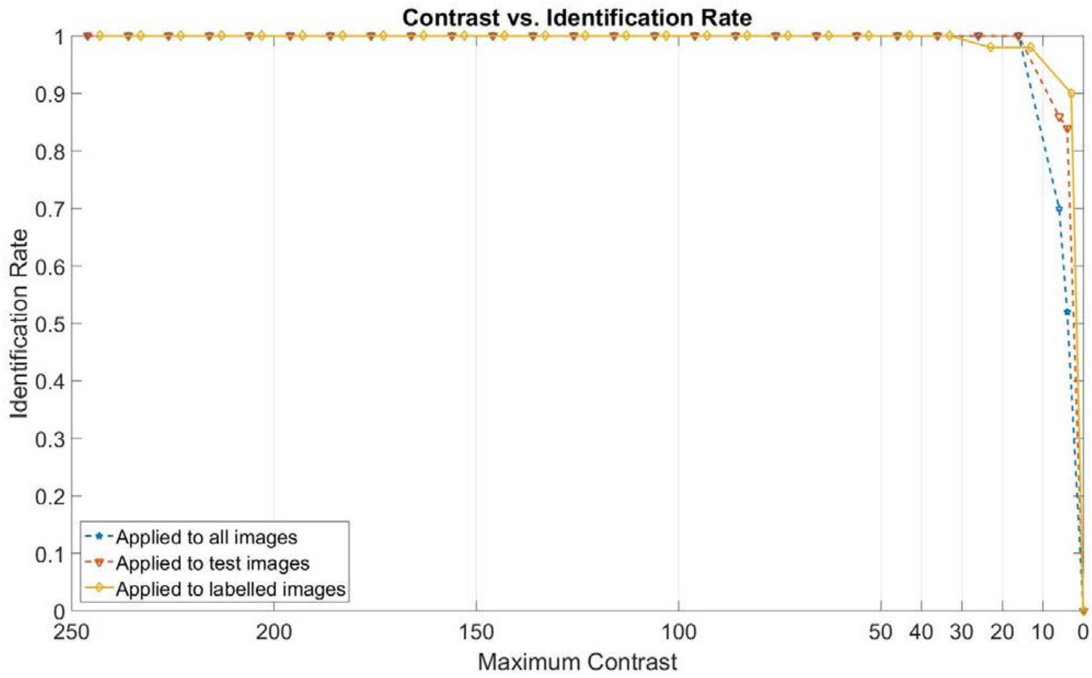


Figure 5.3: Contrast vs. Identification Rate using FAST keypoints.

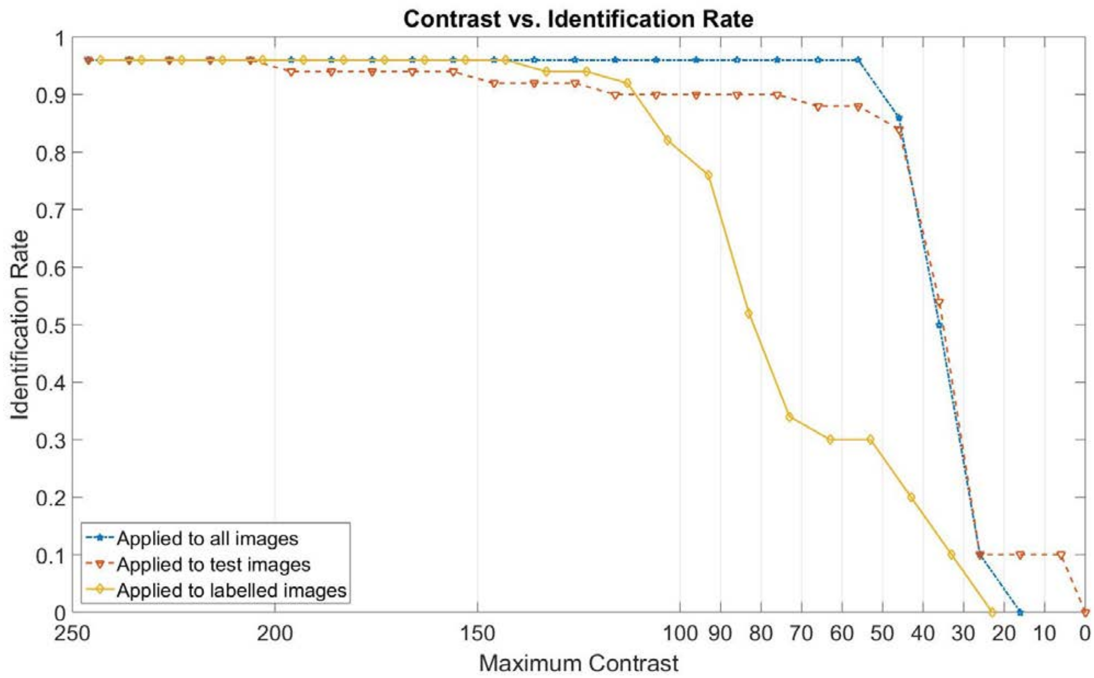


Figure 5.4: Contrast vs. Identification Rate using SIFT keypoints.

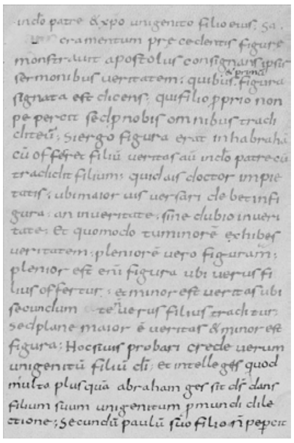
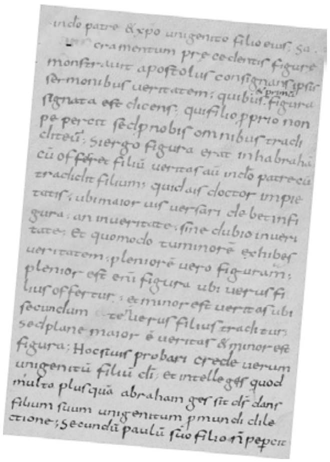
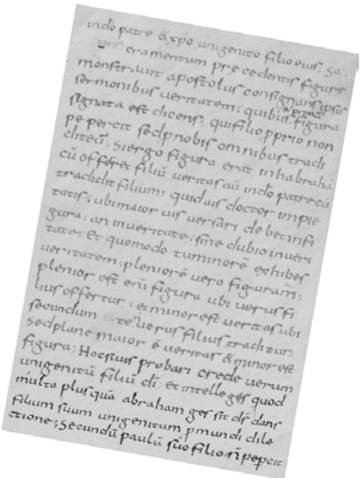
FAST keypoints are evidently very robust against contrast reduction; see Fig. 5.3. High identification rates can be obtained in all scenarios with an as low value of MCT as 15. On the other hand, using SIFT keypoints results in a much faster drop of the identification rate; see Fig. 5.4. From these results, it follows that using FAST keypoints is recommended in case of having low contrast images.

5.4.3 Rotation

This kind of degradation usually occurs during the digitisation process of the manuscript pages. Although handwriting orientation itself depends on the script type, it is compared to handwritings of the same script type for the task of writer identification. Therefore, what matters when applying the proposed method is the relative rotation between different samples.

In order to understand the effect of having a relative rotation between samples with different orientations, we rotate the test samples and increase the rotation angle in each iteration while fixing the orientation of the labelled samples; see the figures in Table 5.4.

Table 5.4: A sample from St. Gall dataset with rotation around their centres.

0 Degrees Rotation	**5 Degrees Rotation**
	
10 Degrees Rotation	**45 Degrees Rotation**
	

Each test image is rotated around its centre by following the steps below:

1. A rotation matrix is calculated for each test image using the centre coordinate of that image and the current rotation angle as follows:

Let M be a 2x3 rotation matrix of six elements as follows:

$$M = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \end{bmatrix}.$$

For an image with centre coordinate p_{centre} at (x_c, y_c) and a rotation angle θ_r, the rotation matrix is:

$$M_{p_{centre}} = \begin{bmatrix} \alpha & \beta & (1-\alpha)\cdot x_c - \beta \cdot y_c \\ -\beta & \alpha & \beta \cdot x_c + (1-\alpha)\cdot y_c \end{bmatrix},$$

where

$$\alpha = \cos\theta_r \quad and \quad \beta = \sin\theta_r$$

2. A linear transformation is applied to the coordinate (x, y) of each pixel p in the test image to obtain the rotated position T_p by simply multiplying the rotation matrix M by the (x, y) coordinate of pixel p as follows:

   Let the coordinates of the current pixel be a 2x1 matrix:

$$P = \begin{bmatrix} x \\ y \end{bmatrix}.$$

   Then

$$T_P = M \cdot [x, y, 1]^T \implies T_P = \begin{bmatrix} a_{00}x + a_{01}y + a_{02} \\ a_{10}x + a_{11}y + a_{12} \end{bmatrix},$$

where T_P is the linear transformation.

3. The process is repeated for each test image in each iteration, and the rotation angle is incremented by 5 degrees in each iteration within a range of degrees from 0 to 45.

The experimental results of rotating the test images show a graceful decline of the identification rate as the rotation angle is increased. The behaviour of the method is very similar in both cases of using SIFT or FAST keypoints as can be seen in Fig. 5.5. High identification rate is achievable within a maximum rotation of 5 to 10 degrees.

Typically, the main text in historical manuscripts is digitised with no or very little rotation that imposes no problem for the proposed method. However, in some cases, para-texts and comments are the handwritings in question. Para-text does not necessarily have the normal orientation of the script. In fact, some para-texts can have any orientation such as commentaries in Arabic manuscripts. In those cases, correcting the orientation of para-text handwritings to horizontal is a necessary step before applying our method.

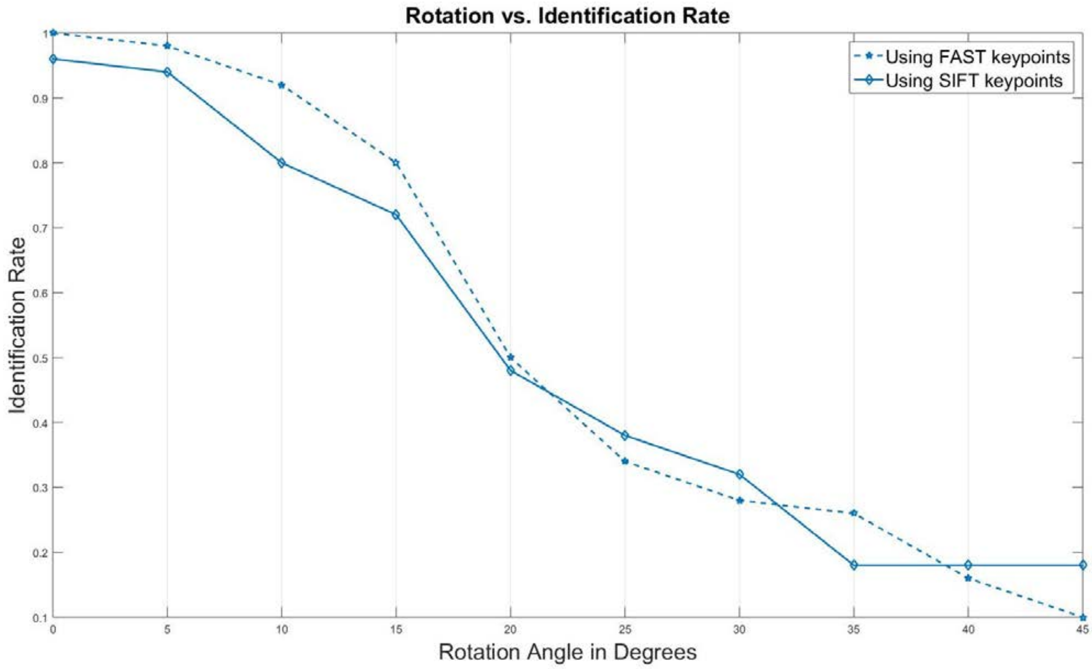


Figure 5.5: Rotation vs. Identification Rate.

5.4.4 Irrelevant Information

Our proposed method is a segmentation-free method that classifies the image as a whole with all what it might contain. This means that any additional text, illustration images, or layouts existing in the image are used in the classification and have an influence on the performance of our method. Therefore, it is important to keep only the desired text written by the scribe in question, in other words, only information relevant to the handwriting style of a specific scribe should be kept, all other information is considered irrelevant and should be removed (cropped out) as much as possible.

In order to quantify the effect of having irrelevant information on the identification rate, a test has been conducted to compare the results with and without the existence of irrelevant information. Two datasets have been created for this test: The first one contains the originally selected 100 sample images with all the layouts, para-texts and other irrelevant information (see Section 5.1), while the second one contains the same sample images but with the main text region only (relevant information); three examples of cropped images are presented in Tables 5.5, 5.6, and 5.7. This has been accomplished simply by manually cropping the region of the main text in the image. Everything within this region is kept as it is, including para-texts and commentaries in between lines. The test results on the two datasets can be found in Table 5.8.

Table 5.5: First sample from St. Gall dataset. Sample image before and after cropping.

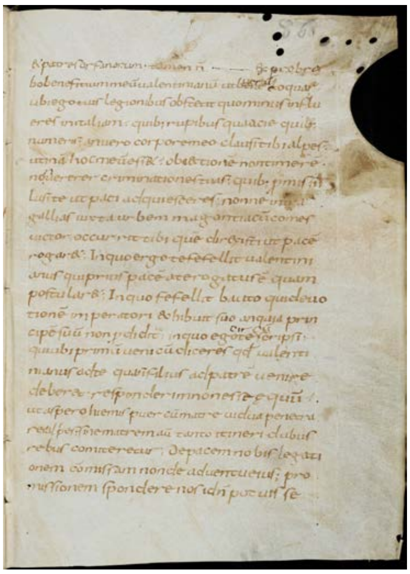
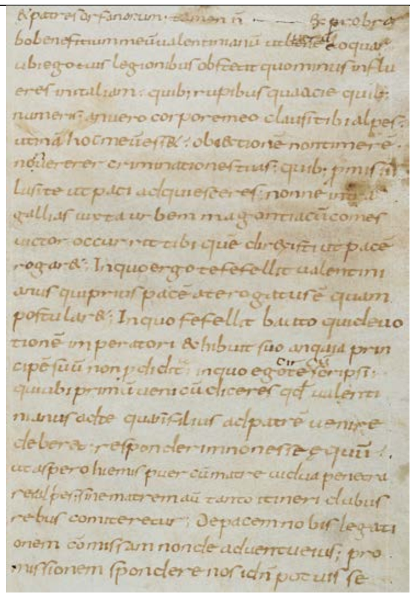
Original Image	**Cropped Image**
	

Table 5.6: Second sample from St. Gall dataset. Sample image before and after cropping.

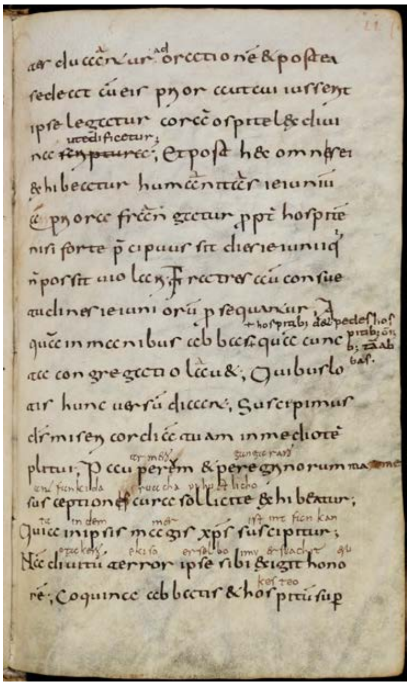
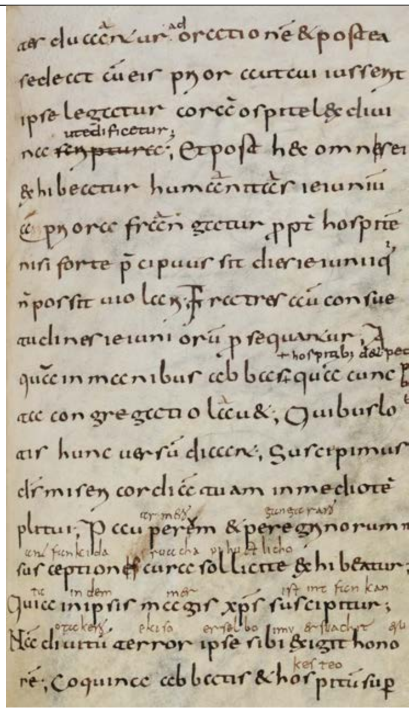
Original Image	Cropped Image
	

Table 5.7: Third sample from St. Gall dataset. Sample image before and after cropping.

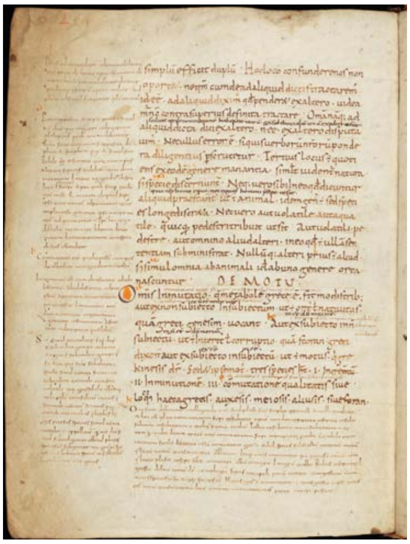
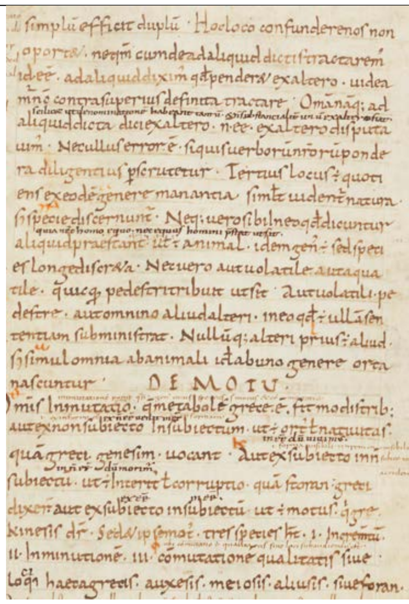
Original Image	Cropped Image
	

Table 5.8: Comparison between the identification rates of our proposed method before and after image cropping.

Before and after cropping	SIFT	FAST
Original images	86%	96%
cropped images	96%	100%

In many cases, para-texts, commentaries, and corrections can be found in between the lines of the main text. Although such text is usually written by another scribe, it has not been removed in the analysis to keep the procedure as simple as possible.

Theoretically, the number of descriptors that describe the handwriting of a given scribe needs to be larger than the number of descriptors of irrelevant information, in order for the method to classify the style correctly; otherwise, the layout and para-text will be classified instead, leading to unintended classification results.

5.5 Conclusion

The proposed method (Normalised Local NBNN) has been analysed w.r.t. some of the common degradation types in digitised manuscripts in order to define the required quality of images and to evaluate the performance of our method w.r.t. the selected degradation types.

Images from a public historical dataset have been selected jointly with the scholars in the SFB sub-project C08 [10] as an evaluation dataset with confirmed ground truth. This dataset is relevant to the data used in the sub-project C08 and representative for the typical degradation types they have.

Our analysis shows that having a similar degradation level of image resolution, contrast, and rotation in both the query and the labelled images provides higher identification rates. Nevertheless, our experiments have shown that SIFT keypoints can cope better with samples of different resolutions. On the other hand, FAST keypoints can cope better with samples of a very low contrast or a very low resolution.

In addition, the impact of the relative rotations between query and labelled samples is analysed and the experimental outcomes indicate that the typical range (from 0 to 10 degrees) of rotation found in digitised manuscripts does not have any significant impact on the identification rate.

Finally, removing any information that is not relevant to the handwriting in question can improve the identification rate regardless of which keypoint detection algorithm is used.

Chapter 6

Implementation as a Software Tool

A software tool has been developed and implemented based on the proposed method in this work with the option to change the main parameters according to the quality of the images as we will explain in the following sections. An intuitive graphical user interface (see Fig. 6.1) has been implemented in order for the scholars from Humanities within the SFB 950 and beyond to be able to integrate the results from our computational method in their research work flow without the aid of experts from the community of computational document analysis. The recommendations and guidelines in the user manual are based on the conclusions drawn from our analysis in this dissertation; see Chapters 3 and 5. The installation and the usage procedure is kept as simple as possible so that the tool can be used by users with limited technical experience.

The implementation of the proposed method has been developed iteratively and incrementally based on feedback from Humanities' scholars in the CSMC within the SFB 950, mainly from sub-project [10]. The first implementation of the proposed method was a command line interface without the feature of parameter change. The second implementation was a simple GUI that can handle a single sample per query and it was without the feature of parameter change as well. Eventually, the third implementation is HAT-2 [80], which has been made public on the website of the Centre of the Study of Manuscript Cultures (CSMC).

6.1 Design Criteria

In order to develop a practical and easy-to-use software tool for the scholars from manuscript research within the SFB 950, several points need to be taken into account. The main criteria which have been considered in the development and design of our software tool are:

- The installation of the software tool should be straight-forward and easy. No additional libraries should be required.
- The sample images of manuscripts should be processed locally to avoid copyrights and ownership issues.
- The overall downloadable package size should be small enough for a typical connection speed and local storage systems.
- The developed software tool should be compatible with the operating systems used by the scholars within the SFB 950 in the CSMC. The operating systems installed in the PCs of the centre are Windows x64 and Windows x32.

- The users should interact with the software tool through an easy-to-use Graphical User Interface (GUI).

In order to develop an easy-to-use GUI, the following design criteria have been considered:

- All GUI elements should be accessible from the main window of the software tool for simplicity.
- The GUI elements should be distributed according to the usage procedure which the user follows.
- The control elements of the GUI, such as buttons, should be disabled whenever not needed to avoid any unintended action by the user.
- In the case of wrong usage or invalidity of data, messages should be provided for the user with easy-to-understand descriptions.

6.2 Handwriting Analysis Tool v2.0 (HAT-2)

Our implementation of the proposed method in this dissertation can be installed as a standalone software without the need to install additional software packages and libraries; furthermore, it can be used by the scholars from Humanities without the aid of experts from the community of computational document analysis as a decision support tool. Therefore, we refer to this implementation as a software tool in this dissertation.

The Handwriting Analysis Tool v2.0 [80] (HAT-2) is developed as a Windows Forms project using C# within the .NET framework from Microsoft. It is an open source project that is licensed under the Creative Commons Attribution-NonCommercial 4.0 International Public License. This software tool has been tested using manuscripts from sub-projects within the SFB 950, namely C08 [10], B05 [12] and C06 [13], and it has been used for the tasks related to writer/scribe identification by scholars from Humanities. In Section 6.3, we will present two use cases to demonstrate the applicability of this tool to research questions of scholars in Humanities within the SFB 950.

The Handwriting Analysis Tool v2.0 (HAT-2) can be used to analyse handwritings of known scribes and sort them according to their similarity to unknown handwritings. A similarity score is produced for each style (scribe) so that the user can have a relative comparison between the styles with respect to a given unknown handwriting. A description of this similarity score will be presented in Section 6.2.3.

The main goal of this tool is to provide supporting information for the scholars from Humanities within the SFB 950 regarding their research questions related to writer/scribe identification and handwriting style analysis.

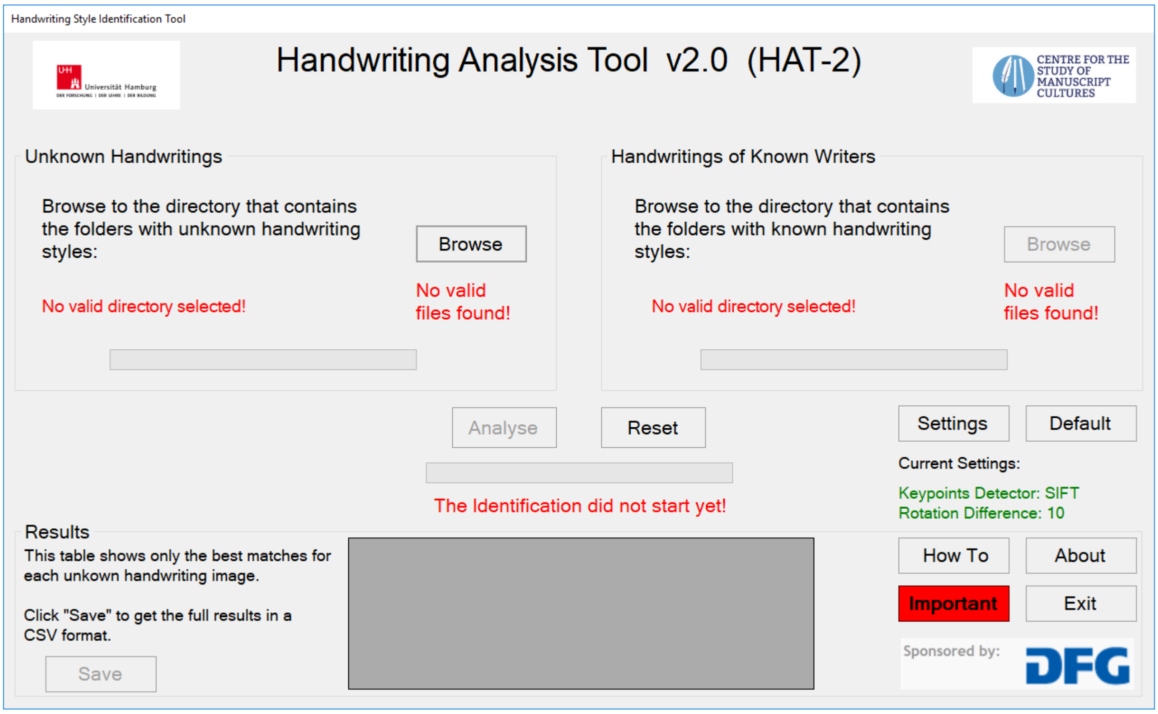


Figure 6.1: HAT-2 Graphical User Interface (GUI).

6.2.1 Required Directory Structure

The handwriting images need to be structured in a certain way so that the tool can process them correctly; an example of the needed directory structure is presented in Fig 6.2. The following guidelines need to be considered when the directory structure is created:

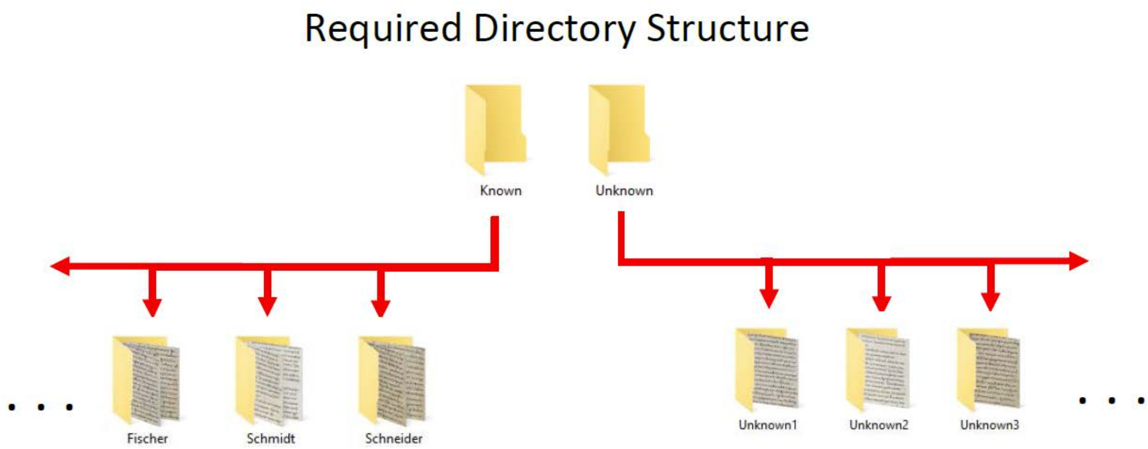


Figure 6.2: Example of the required directory structure in order for the software tool to process the images correctly.

- The (Known) folder must contain at least 2 sub-folders for handwritings from different styles (scribes).
- The name of the folders can be any valid string under the Windows operating system, as far as it is distinguishable by the user.
- There is no upper limit neither for the number of sub-directories nor for the number of images inside them.
- Several unknown handwritings (queries) can be tested simultaneously.
- In the directory of unknown handwritings, all images within the same sub-folder are treated as one query (one image). This is particularly useful when dealing with a heavily degraded or fragmented piece of handwriting; parts with clear handwritings can be cropped and saved as individual images in the same sub-folder as one unknown handwriting (query). The same procedure is applicable to the known handwritings.

6.2.2 Parameter Settings

The default settings of the HAT-2 software tool applies the proposed method using SIFT keypoints detection algorithm with an orientation-difference threshold of 10 degrees. The user can apply these default settings whenever she/he clicks the button *Default*. By clicking the button *Settings*, the user can choose between two different keypoints detection algorithms and can modify the corresponding parameters of the chosen algorithm; see Fig. 6.3.

The user can choose the suitable settings for her/his sample images based on the recommendations we offer in the manual of HAT-2 [80], as well as in the following sections. The recommendations mentioned in the following sections are based on the method analysis presented in Chapters 5 and 3, and in [81] as well.

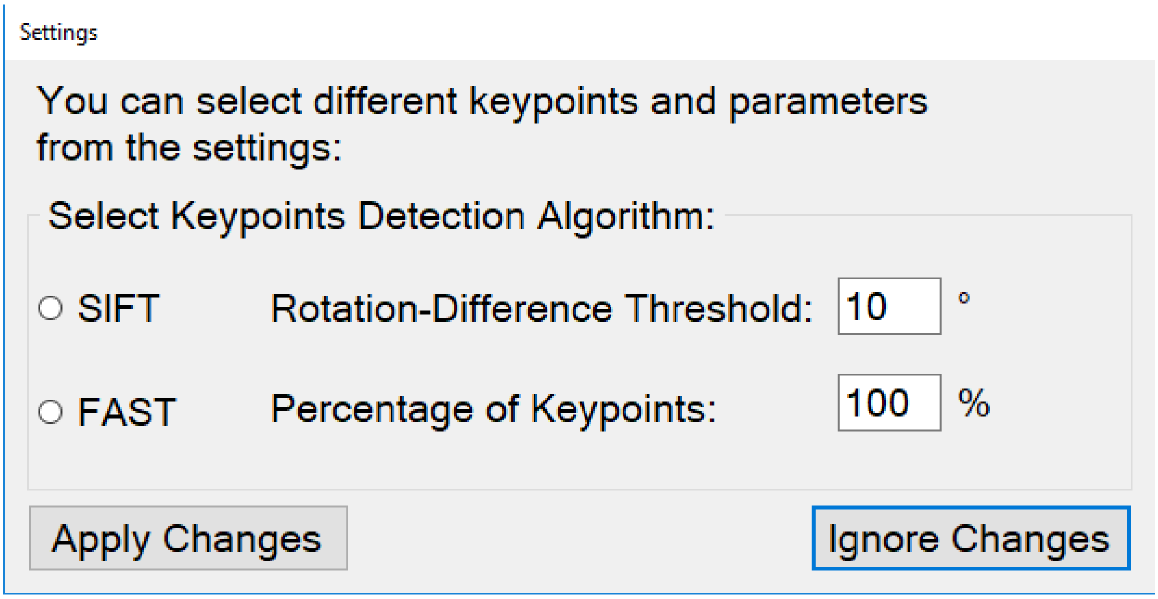


Figure 6.3: The *Windows Dialog Box* from which the settings of HAT-2 can be modified.

Scale Invariant Feature Transform (SIFT)

This keypoint detection algorithm can cope better with images of large difference in resolution; see Chapter 5. Furthermore, one can specify the amount of rotation that can be tolerated between images; see Chapter 3. A rotation-difference threshold of 10 degrees is typically enough. The allowed values are integers between 0 and 90.

Features from Accelerated Segment Test (FAST)

This keypoint detection algorithm can cope better with very low-contrast or very low-resolution images; see Chapter 5. Furthermore, one can specify the percentage of keypoints to be considered in the analysis; see Chapter 3. Only the specified top percentage of keypoints with the highest response is considered. This parameter selection can greatly speed up the processing time, which could be of high importance when dealing with a large collection of manuscripts. The recommended value for this parameter depends on the ratio of the relevant information to the irrelevant information with respect to the targeted handwriting. For handwritings on heavily degraded non-contemporary material such as parchments, the parameter value of which the best results can be obtained can be as low as 5%. In general, a percentage of 10% or less was suitable for all of the manuscript images we tested within the SFB 950 sub-projects. The allowed values are decimals between 0.01 and 100.

6.2.3 Results Presentation

The calculated similarity scores by HAT-2 software tool are measures of relative similarity, the tool calculates how similar an unknown style is to a given known style relative to the other known styles. These similarity scores should be used by scholars in Humanities as indicators of handwriting style similarity and as a supporting information for their research questions. The similarity scores are calculated as follows:

Let D_s be the absolute value of the distance to the handwriting style s calculated by the proposed method in Chapter 3, equ. 3.19: $D_s = |\frac{Dist^c_{local}}{K_c}|$, and $Sum_D = \sum_{s=1}^{n} D_s$, where n is the number of known styles. The relative score S for a given style s is $S_s = \frac{D_s}{Sum_D}100$.

A brief version of the results is displayed automatically as a *summary table*. This *summary table* shows only the best handwriting style match for every unknown handwriting style, while in the results file, all the styles are ranked according to their similarity to the unknown handwriting as *full results*; see Fig. 6.4 for details.

Example of summary results:

File	Best Match	Score
Unknown1	Fischer	71.3
Unknown2	Schmidt	80.7
Unknown3	Schneider	48.1

Example of a full results file:

	A	B	C
1	Results for Unknown1		
2	Rank	Directory	Score
3	1	Fischer	71.3
4	2	Schneider	15.3
5	3	Schmidt	13.2
6			
7	Results for Unknown2		
8	Rank	Directory	Score
9	1	Schmidt	80.7
10	2	Fischer	10.4
11	3	Schneider	8.7
12			
13	Results for Unknown3		
14	Rank	Directory	Score
15	1	Schneider	48.1
16	2	Fischer	31.8
17	3	Schmidt	19.9

Figure 6.4: Illustrative example of the two versions of results produced by the software tool HAT-2.

An illustration of a possible scenario is presented in Fig. 6.4. The similarity scores of three known (labelled) handwriting styles (writers), namely *Fischer*, *Schmidt* and *Schneider*, are calculated by the HAT-2 software tool w.r.t. three unknown (query) handwriting styles (*Unknown1*, *Unknown2* and *Unknown3*). The summary results only provide the name and the similarity score of the most similar known handwriting style to each unknown handwriting style. On the other hand, the full results file provides the similarity scores of all the known handwriting styles to each of the unknown handwriting styles in a separate table.

For example, we can see from the summary results that the known handwriting style *Fischer* is the most similar handwriting style to *Unknown1* with a similarity score of *71.3*. In order to have more detailed results and better understanding for the meaning of this numerical value, we refer to the full results file. In the full results file, we can see that the similarity scores of the known handwriting styles w.r.t. *Unknown1* are given as follows: *Fischer 71.3*, *Schneider 15.3* and *Schmidt 13.2*. One can have an indication from these numerical values for the similarity of *Fischer* handwriting style to *Unknown1* relative to *Schneider* and *Schmidt* handwriting styles, for which they both have much smaller similarity values than *Fischer* to *Unknown1*.

The HAT-2 software tool produces two versions of the generated similarity scores:

- Summary: a brief version of the similarity scores is displayed automatically as a table in the tool window. This summary shows only the best match for every unknown handwriting along with their relative similarity score S.
- Full: a complete version of the results can be obtained by saving the results to a file. One can save it in a (.csv) or (.txt) format. In the saved file, one can find all the styles ranked according to their relative similarity to the unknown handwriting.

In order to avoid any possible confusion by the presented result values, it is worth mentioning that the results can vary slightly for a repeated test.

The typical variation range is less than 1% due to the application of the Fast Library for Approximate Nearest Neighbours (FLANN)[68].

6.2.4 Usage Procedure

Creating a user-friendly graphical user interface that is easy to use was one of the main goals for creating this tool. Therefore, the procedure that the scholars need to follow has been kept simple in the user manual [80]. It can be summarised in the form of a user guide as follows:

1. If the user wants to change the default settings, she/he can select the desired keypoints detection algorithm and enter the corresponding parameter from the Settings; see Section 6.2.2 for details. If this step is skipped, the default settings will be applied; see Section 6.2.2.

2. The user can browse to the folder that contains unknown handwriting styles (each style must be stored in a separated sub-folder).

3. The user can browse to the folder that contains known handwriting styles (each style must be stored in a separated sub-folder).

4. The user can click the button **Analyse** to analyse the known and unknown handwriting styles and produce the similarity scores.

5. The user can check the summary version of the results in the **Results** table to see the best match for every unknown handwriting. She/he can save the full version of the results to a file to see the full results

with all the similarity scores produced by the HAT-2 software tool; see Section 6.2.3 for details.

6.2.5 Technical Considerations

The main technical requirements to install and use the tool are:

- The required target system (platform) is Windows (x64 and x32).
- The supported file extensions for input images are: .jpg/.jpeg, .tif/.tiff, .png and .bmp.
- The possible formats for results file are: (.csv) file, which can be opened by any spreadsheet application like Microsoft Excel, and (.txt) file, which is a plain text format.

6.2.6 Additional Considerations

In general, the following remarks are important to be considered when using this software tool:

- In the directory of Known handwriting styles, the name of the sub-folder will be used as the name of the style for the images in that sub-folder.
- The **Reset** button deletes any stored information and prepares the software tool to start a new test. The previous test results will be deleted; therefore, the full version of results should be saved to a file before resetting.

- Any information that is irrelevant to the targeted handwriting should be removed as much as possible. As yet, this can be done by simply cropping the image region(s) with the targeted handwritings parts only; see the discussion in Section 5.4.4.

6.3 Application to Research Questions of Scholars within the SFB

The design of HAT-2 software tool allows for several usage scenarios depending on the scholar's approach and the problem at hand. Moreover, there is no distinction internally between the labelled samples in the "Known" sub-folders and the unlabelled samples in the "Unknown" sub-folders. Therefore, all the writers/scribes of the sample images can be unknown in a given test, it is only needed to add labels (distinctive names) to the samples in the "Known" sub-folders to discriminate between the different writing styles. The similarity scores then can be interpreted in a meaningful and useful way.

For the cases when only two handwriting styles (S1 and S2) need to be compared, samples from one of the styles can be split into two parts (S1.1 and S1.2), then the problem can be formulated as follows:

The "Unknown" sub-folder (query) contains S1.1 samples, while the "Known" sub-folders contain S1.2 samples and S2 samples, each with a different label. The higher the difference in the score between the two styles, the more they are different.

The results produced by this software tool should be considered as a supporting information to the palaeographers rather than a final scribe identification result. Furthermore, the HAT-2 design allows the user to interact by modifying the set of handwriting style to compare with, the settings, and the parts of handwritings that they are interested in. Therefore, this software tool is a research tool that provides supporting information related to handwriting style analysis for the scholars from Humanities. This supporting information can be used in the process of identifying (or hypothesising) the scribe of a given manuscript.

In summary, this software is designed as a useful tool to be used by scholars with sufficient knowledge about the handwritings in question. The research questions (see the following sections) need to be formulated by the scholars in order for the tests to make any scientific sense. Furthermore, the results need to be used by experts from manuscript research in Humanities to ensure the careful considerations of the textual, philological, and historical context of the handwriting materials.

Some of the usage possibilities (options) for this software tool are:

- Validating proposed hypotheses by the scholars related to handwriting style analysis and scribe identification.
- Indicating similarities between handwritings in a collection of questioned manuscripts to be further investigated by the scholars.
- Providing similarity values between questioned handwritings in manuscripts which can lead to further investigations.

- Providing a ranked list of similarities which can be used to help tracking gradual changes in a handwriting style (e.g. due to the ageing of a given scribe or the increasing physical fatigue from the begin to the end of a manuscript).

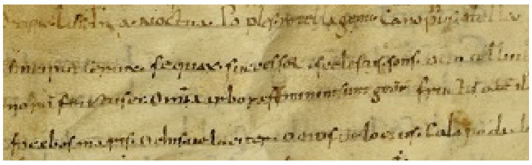
In the following sections, use cases from two sub-projects in the SFB 950 are presented. In order to demonstrate the generality and applicability of our proposed method, we selected these two use cases based on their orthogonality in terms of script types and research questions.

6.3.1 Use Case: Sub-Project C08

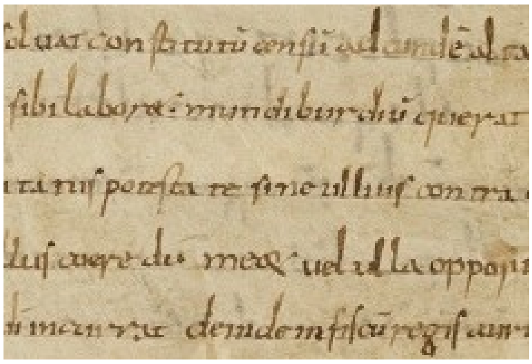
The sub-project C08 [10] "East Frankish manuscripts with collections of formulas", led by Philippe Depreux, investigates manuscripts written in the Eastern regions of the Frankish Empire during the 9th and 10th century which contain collections of formulae, i.e. sample letters and charters. A collaboration has been established with this sub-project [3] in order to carry out handwriting style analyses of these manuscripts in order to provide supporting information for the task of scribe identification and handwriting style comparison. The intention is not only to assign manuscripts (or parts of manuscripts) to a specific scribe or scriptorium but to document the compilation and growth of the manuscripts.

A scholar in this sub-project, namely Till Hennings, isolated each handwriting style hypothesized to be different in a separate sub-folder and gave a label for each sub-folder. After that, the formulation of the questions (tests) refer only to the labels they assigned. The scholars formulated the tests to measure the similarity between these handwriting styles. One of their research questions was the following:

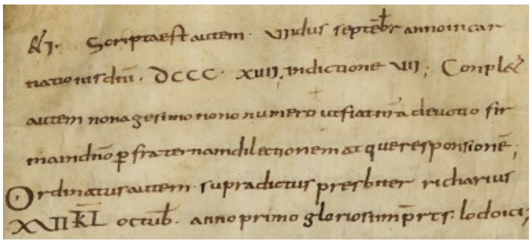
Do the handwriting samples in B, E and P sub-folders belong to the same style/scribe? Handwriting samples from the mentioned sub-folders are given in Fig. 6.5 along with their assigned labels.



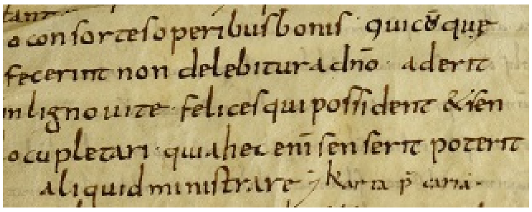
(a) Sample from sub-folder B.



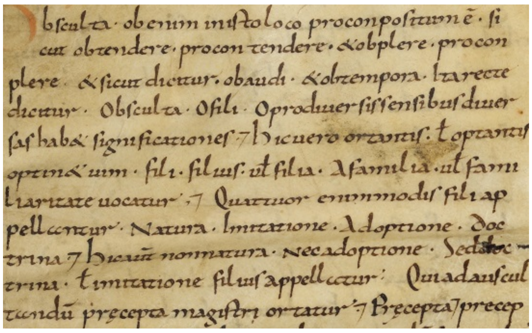
(b) Sample from sub-folder E.



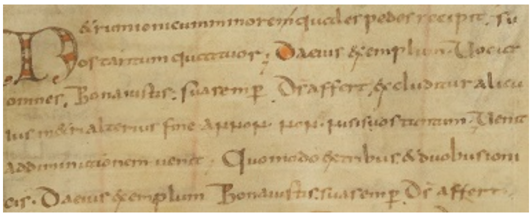
(c) Sample from sub-folder P.



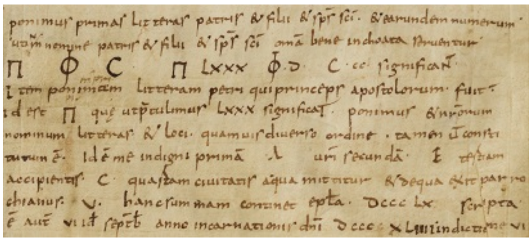
(d) Sample from sub-folder C.



(e) Sample from sub-folder F.



(f) Sample from sub-folder O.



(g) Sample from sub-folder Q.

Figure 6.5: Handwriting samples from sub-folders B,E,P,C,F,O and Q. Paris BNF. Latin 763 (http://www.bnf.fr/fr/acc/x.accueil.html)

In order to answer this question, the scholars in sub-project C08 performed three tests: The first one measures the similarity between B and all the other styles including E and P, the second one measures the similarity between E and all the other styles including B and P, the third test measures the similarity between P and all the other styles including B and E. HAT-2 is used to perform these tests using FAST keypoints with $PCK = 10\%$. The FAST keypoints detection algorithm has been selected because all the samples in this test have similar image resolution, and the global image contrast is very low in some images. The full results are shown in Figs. 6.6, 6.7 and 6.8.

Results for B		
Rank	Directory	Score
1	P	42.3
2	E	32.4
3	O	13.7
4	Q	4.0
5	C	3.7
6	F	3.7

Figure 6.6: The test result for query B obtained by scholars from sub-project C08 using HAT-2.

Results for E		
Rank	Directory	Score
1	B	33.5
2	P	20.5
3	C	17.7
4	O	10.8
5	F	8.8
6	Q	8.3

Figure 6.7: The test results for query E obtained by scholars from sub-project C08 using HAT-2.

Results for P		
Rank	Directory	Score
1	B	35.1
2	E	23.6
3	O	17.7
4	C	12.0
5	F	6.3
6	Q	5.0

Figure 6.8: The test results for query P obtained by scholars from sub-project C08 using HAT-2.

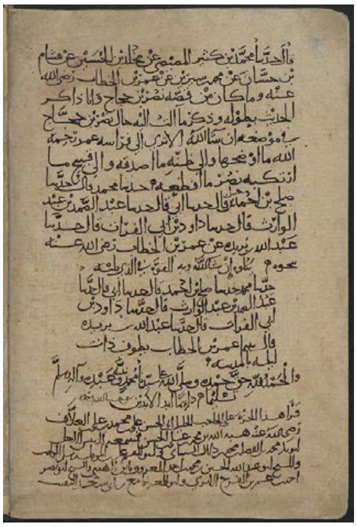
The test results for query B in Fig. 6.6 show that the similarity scores of handwriting styles P and E (42.3% and 32.4% respectively) are clearly higher than the following similarity score of handwriting style O (13.7%) in the ordered list of results. These similarity scores are relative to the similarity scores of the other handwriting styles, namely Q, C, and F. The test results for queries E and P also demonstrate clearly the similarities between the handwriting styles of B, E, and P.

In conclusion, test results of the queries B, E and P show that the most similar pair of handwritings for any given query is the other two queries. These results confirm the preliminary hypothesis stated by the scholars in the sub-project based on palaeographical pieces of evidence.

6.3.2 Use Case: Sub-Project B05

Tilman Seidensticker from sub-project B05 [12] presented an investigation [82] of a case study of Arabic audience certificates contained in the manuscript Ms. orient. A-627 from the Forschungsbibliothek Gotha, Germany. He provided in his research a hypothesis about the different hands of audience certificates based on manual examination of text by comparison of similar text contents across different handwriting styles in addition to the textual context and content.

In order to validate the hypothesis presented in [82], we jointly carried out a test to measure the similarities of handwriting styles between the different audience certificates. Each handwritten paragraph is cropped and considered to be written by a unique writer and is given a numerical label from *writer 1* to *writer 13*; the samples used in the tests are shown in Fig. 6.9. Then we measured the similarity between all the writing styles using HAT-2 to see if some styles are indeed more similar to each other which indicate that they may belong to the same scribe. The FAST keypoints detector was used with $PCK = 10\%$. The FAST keypoints detection algorithm has been selected because all the samples in this test have the same image resolution, and the contrast is very low in some parts of the sample images.



(a) First sample.

(b) Second sample.

(c) Third sample.

(d) Fourth sample.

(e) Fifth sample.

(f) Sixth sample.

Figure 6.9: Samples used in the test from Arabic audience certificates from manuscript Ms. orient. A-627.

The results of the test confirmed the hypothesis presented in [82]. Furthermore, the similarity scores provided additional information to the research and paved the way for further investigation. Three examples of the confirmed hypotheses using HAT-2 are presented in figures 6.10, 6.11 and 6.12. For example, the test results in Fig. 6.10 clearly show that the similarity between the handwriting pairs (*writer 2* and *writer 8*) as queries, which are hypothesised to be written by the same hand, is indeed much higher compared to the other handwritten paragraphs in the same test by a large margin. The same is also true for the other two pairs of handwritten paragraphs (*writer 4*, *writer 10*) as queries and (*writer 5*, *writer 12*) as queries.

Writer 2

Rank, Writer, Score	
1,writer 8	26,04
2,Writer 3	9,54
3,writer 4	8,44
4,writer 1	7,99
5,writer 10	7,97
6,writer 13	7,35
7,writer 5	7,26
8,writer 9	6,66
9,writer 12	5,19
10,writer 7	4,97
11,writer 6	4,63
12,writer 11	3,97

Writer 8

Rank, Writer, Score	
1,writer 2	23,79
2,writer 13	11,05
3,Writer 3	9,63
4,writer 10	9,28
5,writer 4	8,36
6,writer 1	7,08
7,writer 5	6,94
8,writer 12	5,64
9,writer 9	5,45
10,writer 7	4,57
11,writer 6	4,47
12,writer 11	3,73

Figure 6.10: Results generated by HAT-2 using the cropped Arabic audience certificates from the research in [82]. The results shown in this figure are for the handwriting pairs (*writer 2* and *writer 8*).

Writer 4	
Rank, Writer, Score	
1,writer 10	20,04
2,Writer 3	10,04
3,writer 13	8,40
4,writer 8	8,24
5,writer 1	8,19
6,writer 6	7,63
7,writer 2	7,57
8,writer 5	7,18
9,writer 7	6,30
10,writer 12	6,19
11,writer 11	5,24
12,writer 9	4,97

Writer 10	
Rank, Writer, Score	
1,writer 4	20,51
2,writer 13	8,90
3,writer 8	8,60
4,Writer 3	8,13
5,writer 5	7,74
6,writer 12	7,50
7,writer 6	7,13
8,writer 1	6,82
9,writer 7	6,48
10,writer 2	6,43
11,writer 11	6,08
12,writer 9	5,69

Figure 6.11: Results generated by HAT-2 using the cropped Arabic audience certificates from the research in [82]. The results shown in this figure are for the handwriting pairs (*writer 4* and *writer 10*).

Writer 5	
Rank, Writer, Score	
1,writer 12	25,82
2,writer 7	18,28
3,writer 10	7,30
4,writer 8	6,74
5,Writer 3	6,46
6,writer 4	6,19
7,writer 13	5,71
8,writer 9	5,35
9,writer 6	4,91
10,writer 2	4,90
11,writer 1	4,28
12,writer 11	4,05

Writer 12	
Rank, Writer, Score	
1,writer 5	31,86
2,writer 7	18,42
3,writer 10	6,94
4,Writer 3	6,51
5,writer 6	5,69
6,writer 13	5,55
7,writer 4	5,40
8,writer 8	4,83
9,writer 1	4,19
10,writer 11	3,98
11,writer 9	3,53
12,writer 2	3,11

Figure 6.12: Results generated by HAT-2 using the cropped Arabic audience certificates from the research in [82]. The results shown in this figure are for the handwriting pairs (*writer 5* and *writer 12*).

6.4 HAT-2 within the Community of Manuscript Research in Humanities

Our developed software tool HAT-2 has been perceived as a useful tool and used by the community of manuscript research in Humanities. An invited talk has been given in Universität Heidelberg to present the HAT-2 due to the interest shown by the scholars in the Text-Object-Person (T-O-P) research group (`https://www.uni-heidelberg.de/forschung/profil/field_of_focus_3/forschung/`). In addition, an invited lecture will be given in Universität Basel to present our work on handwriting style analysis and the HAT-2 due to the interest shown by the scholars in the SNSF Ambizione project: “Reuniting fragments, identifying scribes and characterizing scripts: the Digital paleography of Greek and Coptic papyri” (`https://altegeschichte.philhist.unibas.ch/de/digpaleo/`).

In addition to the use cases presented in Sections 6.3.1 and 6.3.2, other scholars from manuscript research in Humanities reported an independent (without aid from experts) application of the HAT-2 to their manuscript samples such as Marco Heiles from RWTH Aachen University (Latin script), and Isabelle Marthot-Santaniello from Universität Basel (Greek script).

6.5 Conclusion

We developed an easy-to-use implementation of our proposed method as a software tool. This software tool is developed with a user-friendly GUI and it produces similarity scores in an intuitive presentation so that it can be used by the scholars from Humanities without the aid of experts from the community of computational document analysis. Our software tool has been used by scholars from Humanities for their research yielding very satisfying results for as yet two use cases from two sub-projects within the SFB 950.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

The main goal of this dissertation is to develop a computational method capable of analysing the handwriting samples of digitised manuscripts in order to generate similarity scores which can be used as a supporting information for the task of handwriting style identification.

This dissertation is a part of the Scientific Service Project Z03 “Image Processing Methods for Determining Visual Manuscript and Character Features” [3] within the Sonderforschungsbereich (SFB 950) “Manuscript Cultures in Asia, Africa and Europe” [2] at Universität Hamburg.

Requirements have been gathered and analysed from selected sub-projects within the SFB 950 with regards to the task of handwriting style identification. Then we analysed the related state-of-the-art of computational methods based on these requirements in order to find the best starting point for the development of an improved, thus novel method.

The focus of the state-of-the-art computational methods for writer identification was mainly on feature selection and design rather than on classifiers. Several types of features have been used by researchers to capture the individuality of handwriting. Methods using gradient-based features like SIFT descriptors demonstrated state-of-the-art results on digitised manuscripts. These features describe visual features in local regions of handwritings without the need of prior contour extraction or character segmentation which is difficult and unreliable preprocessing in digitised manuscripts of sub-projects within the SFB 950 given the typical degradation of their handwriting samples.

The handwriting samples of historical manuscripts are often sparse and without labels or even ground-truth, which do not render possible the application of learning-based methods. This is also true in the case of the selected sub-projects within the SFB 950. Therefore, we developed a learning-free method based on the Local Naïve Bayes Nearest-Neighbour (NBNN) classifier. This classifier requires dense keypoints detection algorithms such as SIFT and FAST keypoints detectors in order to provide high classification rates.

This dissertation presented a novel method for the task of handwriting style identification based on the Local NBNN classifier given small sets of unbalanced sample data. The orientations of SIFT keypoints are used to restrict the matching between descriptors to only those with similar orientation. Distances to classes are normalised by the number of keypoints for each class to cope with the prevalent problem of unbalanced data in digitised manuscripts of the selected sub-projects within the SFB 950.

The performance of our proposed method has been evaluated on several public datasets of different writing systems including musical scores and state-of-the-art results were obtained in all experiments with a fixed parameter set. This performance evaluation demonstrated the discriminative power of the proposed method w.r.t. different handwriting styles in the standard datasets. Moreover, some of these standard datasets offer handwriting styles in many different script types from a large number of writers. Therefore, the performance evaluation results also demonstrated the generality of our proposed method and the scalability to a large number of classes.

Degradation in digitised manuscripts can result e.g. from the poor preservation conditions, from the used materials, and even from the digitisation process itself. These degradation factors have a negative impact on the performance of computational methods and they cannot be always eliminated. Therefore, the proposed method has been analysed w.r.t. some of the common degradation types in digitised manuscripts in order to define the required quality of images and thus provide guidelines for the scholars on what parameters to choose according to the image quality of their handwriting samples. Images from a public historical dataset have been used in this analysis and have been selected jointly with scholars from the SFB 950. This dataset is selected to be relevant to the data used in the sub-project C08 of the SFB 950 and representative w.r.t. the typical degradation their data have.

Our analysis in this dissertation showed that having a similar degradation level in both the query and the labelled images provides higher identification rates. Nevertheless, SIFT keypoints can cope better with samples of different resolutions. On the other hand, FAST keypoints can cope better with samples of a very low contrast or a very low resolution. In addition, the impact of the relative rotations between query and labelled samples is analysed and the outcomes indicated that the typical range of rotation found in digitised manuscripts does not have any significant impact on the identification rate of the proposed method. Finally, our analysis showed that removing the elements that are not relevant to the handwriting in question from the images can improve the identification rate regardless of which keypoint detection algorithm is used.

The currently proposed methods of handwriting style analysis are beyond the reach of scholars from manuscript research in Humanities: either because of the required computational resources of the method itself or because of the lack of easy-to-use implementations. Therefore, we developed an easy-to-use software tool of our proposed method. The HAT-2 software tool is implemented with a user-friendly GUI and it produces similarity scores with an intuitive presentation so that it can be used by the scholars from Humanities without the aid of experts from the community of computational document analysis. Our software tool implementation has been made public via the website of the SFB 950 and has been used by pilot scholars from Humanities within the SFB 950 for their research yielding very satisfying results. As of today, two use cases from the sub-projects within the SFB 950 have been presented and discussed in this dissertation in order to demonstrate the applicability of the developed method as an software tool to research problems from Humanities‘ scholars.

7.2 Future Work

The performance of the proposed method can be enhanced by considering only keypoints detected on textual parts of the images. This can be accomplished either by prior detection of text regions via layout analysis or by determining the optimal Percentage of Considered Keypoints (PCK) value for each image. In addition, a possible drop in the performance can be avoided by estimation and correction of image rotation.

Furthermore, post-processing steps can be added to the proposed method in order to enhance the workflow of scholars from the Humanities for their tasks related to handwriting style analysis. For example, after the proposed method is used to rank the labelled images according to their similarity to a query image, the most visually similar local regions in the top-ranked images can be detected and located for further investigation by the scholars.

In addition to writer identification, the application range of the proposed approach may also be extended to other related problems, such as the clustering of handwriting styles, the dating of manuscripts based on handwriting style and the automatic comparison and classification of manuscript pages based on the differences in handwriting style.

The current implementation of the software tool HAT-2 can be improved by providing interaction-based functionalities, like text region selection. Furthermore, visualising the detected keypoints can be helpful in selecting the optimal PCK value for FAST keypoints. In addition, the software tool can be re-implemented as a web application in order to provide a platform-independent implementation.

The research group in the iXMan Lab in the Department of Informatics, Universität Hamburg, developed an integration architecture and re-implemented HAT-2 as a web application with the feature of text region selection through an interaction-based functionality. Furthermore, this group is currently working on the visualisation of the detected keypoints on handwritten manuscript samples.

List of Publications

[1] H. Mohammed. Handwriting Analysis Tool v2.0 (HAT-2). Available at https://www.manuscript-cultures.uni-hamburg.de/hat.html. Accessed: 01.10.2018.

[2] H. Mohammed. Research plan: Computational analysis of writing style in digital manuscripts. Available at http://u-pat.org/ICDAR2017/img/Leaflets-DC_ICDAR2017.pdf, 2017. Accessed: 01.10.2018.

[3] H. Mohammed, V. Märgner, T. Konidaris, and H. S. Stiehl. Normalised local naïve bayes nearest-neighbour classifier for offline writer identification. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 01, pages 1013–1018. IEEE, Nov 2017.

[4] H. Mohammed, V. Märgner, and H. S. Stiehl. Writer identification for historical manuscripts: Analysis and optimisation of a classifier as an easy-to-use tool for scholars from the humanities. In *2018 16th Interna-*

tional Conference on Frontiers in Handwriting Recognition (ICFHR), pages 534–539. IEEE, 2018.

[5] K. Terzić, H. Mohammed, and J.M.H. du Buf. Shape detection with nearest neighbour contour fragments. In Mark W. Jones Xianghua Xie and Gary K. L. Tam, editors, *Proceedings of the British Machine Vision Conference (BMVC)*, pages 59.1–59.12. BMVA Press, Sep 2015.

Bibliography

[1] "DFG research group 963: Manuscript cultures in Asia and Africa." Available at https://www.manuscript-cultures.uni-hamburg.de/archive/index_e.html, 2011. Accessed: 30.10.2018.

[2] "Sonderforschungsbereich (SFB 950): Manuscript cultures in Asia, Africa and Europe." Available at https://www.manuscript-cultures.uni-hamburg.de/index_e.html, 2011. Accessed: 30.10.2018.

[3] "Z03 project: Methods of image processing for the determination of visual manuscript and character features." Available at http://www.manuscript-cultures.uni-hamburg.de/Projekte_p2.html#Z03. Accessed: 01.10.2018.

[4] "International conference on document analysis and recognition (ICDAR)." Available at http://u-pat.org/ICDAR2017/index.php, 2017. Accessed: 30.10.2018.

[5] O. Hilton, "Some basic rules for the identification of hand-writing," *Medicine, science, and the law*, vol. 3, no. 2, pp. 107–117, 1963.

[6] M. Bulacu and L. Schomaker, "Text-independent writer identification and verification using textural and allographic features," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 29, no. 4, pp. 701–717, 2007.

[7] E. Dalton and N. R. Howe, "Style-based retrieval for ancient syriac manuscripts," in *Proceedings of the 2011 Workshop on Historical Document Imaging and Processing*, pp. 1–5, ACM, 2011.

[8] H. Mohammed, "Research plan: Computational analysis of writing style in digital manuscripts." Available at http://u-pat.org/ICDAR2017/img/Leaflets-DC_ICDAR2017.pdf, 2017. Accessed: 01.10.2018.

[9] T. Konidaris, H. S. Stiehl, V. Märgner, and H. Mohammed, "Requirements analysis report," Nov 2016. Internal Report within The SFB 950.

[10] "C08 sub-project: East Frankish manuscripts containing collections of formulae." Available at https://www.manuscript-cultures.uni-hamburg.de/Projekte_p2.html#C08, 2015.

[11] "C04 sub-project: Reading, memorizing and recording: Manuscripts in alevi village communities in Anatolia." Available at https://www.manuscript-cultures.uni-hamburg.de/Projekte_p2.html#C04, 2015. Accessed: 01.10.2018.

[12] "B05 sub-project: The handling of qur'an manuscripts in Islamic-Arabic culture using the example of small and rolling Koran." Available at `https://www.manuscript-cultures.uni-hamburg.de/Projekte_p2.html#B05`, 2015. Accessed: 01.10.2018.

[13] "C06 sub-project: East Frankish manuscripts containing collections of formulae." Available at `https://www.manuscript-cultures.uni-hamburg.de/Projekte_p2.html#C06`, 2015. Accessed: 01.10.2018.

[14] S. McCann and D. G. Lowe, "Local Naïve Bayes Nearest Neighbor for image classification," *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3650–3656, June 2012.

[15] R. Plamondon and G. Lorette, "Automatic signature verification and writer identificationthe state of the art," *Pattern recognition*, vol. 22, no. 2, pp. 107–131, 1989.

[16] M. Sreeraj and S. M. Idicula, "A survey on writer identification schemes," *International Journal of Computer Applications*, vol. 26, no. 2, pp. 23–33, 2011.

[17] S. M. Awaida and S. A. Mahmoud, "State of the art in off-line writer identification of handwritten text and survey of writer identification of Arabic text," *Educational Research and Reviews*, vol. 7, no. 20, p. 445, 2012.

[18] A. A. Ahmed, M. S. Al-Tamimi, O. I. Al-Sanjary, and G. Sulong, "Classification of Arabic writer based on clustering techniques," in *International Conference of Reliable Information and Communication Technology*, pp. 48–58, Springer, 2017.

[19] S. He, P. Samara, J. Burgers, and L. Schomaker, "Historical manuscript dating based on temporal pattern codebook," *Computer Vision and Image Understanding*, vol. 152, pp. 167–175, 2016.

[20] K. Adam, S. Al-Maadeed, and A. Bouridane, "based classification of Arabic scripts style in ancient Arabic manuscripts: Preliminary results," in *1st International Workshop on Arabic Script Analysis and Recognition (ASAR)*, pp. 95–98, IEEE, 2017.

[21] I. Siddiqi, F. Cloppet, and N. Vincent, "Contour based features for the classification of ancient manuscripts," in *Conference of the International Graphonomics Society*, pp. 226–229, 2009.

[22] G. Joutel, V. Eglin, S. Bres, and H. Emptoz, "Curvelets based feature extraction of handwritten shapes for ancient manuscripts classification," in *Document Recognition and Retrieval XIV*, vol. 6500, p. 65000D, International Society for Optics and Photonics, 2007.

[23] A. Brink, J. Smit, M. Bulacu, and L. Schomaker, "Writer identification using directional ink-trace width measurements," *Pattern Recognition*, vol. 45, no. 1, pp. 162–171, 2012.

[24] Z. A. Daniels and H. S. B., "Discriminating features for writer identification," in *12th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1385–1389, IEEE, 2013.

[25] A. Nicolaou, A. D. Bagdanov, M. Liwicki, and D. Karatzas, "Sparse radial sampling lbp for writer identification," in *13th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 716–720, IEEE, 2015.

[26] I. Dinstein and Y. Shapira, "Ancient Hebraic handwriting identification with run-length histograms.," *IEEE Transactions on Systems, Man and Cybernetics.*, vol. 12, no. 3, pp. 405–409, 1982.

[27] C. Djeddi, I. Siddiqi, L. Souici-Meslati, and A. Ennaji, "Text-independent writer recognition using multi-script handwritten texts," *Pattern Recognition Letters*, vol. 34, no. 10, pp. 1196–1202, 2013.

[28] S. He and L. Schomaker, "General pattern run-length transform for writer identification," in *Document Analysis Systems (DAS), 12th edition workshop*, 2016.

[29] D. Fecker, A. Asi, V. Märgner, J. El-Sana, and T. Fingscheidt, "Writer identification for historical Arabic documents," in *2014 22nd International Conference on Pattern Recognition (ICPR)*, pp. 3050–3055, IEEE, 2014.

[30] F. Wahlberg, "Gaussian process classification as metric learning for forensic writer identification," in *Document Analysis Systems (DAS), 13th edition workshop*, 2018.

[31] S. He, M. Wiering, and L. Schomaker, "Junction detection in handwritten documents and its application to writer identification," *Pattern Recognition*, vol. 48, no. 12, pp. 4036–4048, 2015.

[32] P. Dondi, A. Danai, L. Lombardi, M. Malagodi, and M. Licchelli, "Handwriting identification of short historical manuscripts," in *Document Analysis Systems (DAS), 13th edition workshop*, 2018.

[33] V. Christlein, M. Gropp, S. Fiel, and A. Maier, "Unsupervised feature learning for writer identification and writer retrieval," *arXiv preprint arXiv:1705.09369*, 2017.

[34] C. Adak, S. Marinai, B. Chaudhuri, and M. Blumenstein, "Offline bengali writer verification by PDF-CNN and Siamese net," in *Document Analysis Systems (DAS), 13th edition workshop*, 2018.

[35] R. A. Huber and A. M. Headrick, *Handwriting identification: facts and fundamentals*. CRC press, 1999.

[36] R. Jain and D. Doermann, "Writer identification using an alphabet of contour gradient descriptors," in *12th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 550–554, IEEE, 2013.

[37] B. Arazi, "Handwriting identification by means of run-length measurements," *IEEE Transactions on Systems, Man and Cybernetics.*, no. 12, pp. 878–881, 1977.

[38] B. Arazi, "Automatic handwriting identification based on the external properties of the samples," *IEEE Transactions on Systems, Man and Cybernetics.*, no. 4, pp. 635–642, 1983.

[39] M. Bulacu, L. Schomaker, and A. Brink, "Text-independent writer identification and verification on offline Arabic handwriting," in *9th International Conference on Document Analysis and Recognition (ICDAR)*, vol. 2, pp. 769–773, IEEE, 2007.

[40] M. Crosier and L. D. Griffin, "Texture classification with a dictionary of basic image features," in *2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–7, IEEE, 2008.

[41] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2005)*, vol. 1, pp. 886–893, IEEE, 2005.

[42] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[43] D. Fecker, A. Asi, W. Pantke, V. Märgner, J. El-Sana, and T. Fingscheidt, "Document writer analysis with rejection for historical Arabic manuscripts," in *14th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pp. 743–748, IEEE, 2014.

[44] Y. Xiong, Y. Wen, P. Wang, and Y. Lu, "Text-independent writer identification using SIFT descriptor and contour-directional feature," in *13th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 91–95, IEEE, 2015.

[45] A. J. Newell and L. D. Griffin, "Writer identification using oriented basic image features and the delta encoding," *Pattern Recognition*, vol. 47, no. 6, pp. 2255–2265, 2014.

[46] L. Schomaker, M. Bulacu, and K. Franke, "Automatic writer identification using fragmented connected-component contours," in *9th International Workshop on Frontiers in Handwriting Recognition*, pp. 185–190, IEEE, 2004.

[47] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 509–522, Apr. 2002.

[48] M. Jehanzeb, G. B. Sulong, and I. Siddiqi, "Improving codebook-based writer recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 27, no. 06, p. 1353003, 2013.

[49] M. Abdi and M. Khemakhem, "A model-based approach to offline text-independent Arabic writer identification and verification," *Pattern Recognition*, vol. 48, no. 5, pp. 1890–1903, 2015.

[50] A. Gordo, A. Fornés, and E. Valveny, "Writer identification in handwritten musical scores with bags of notes," *Pattern Recognition*, vol. 46, no. 5, pp. 1337–1345, 2013.

[51] G. R. Ball, S. N. Srihari, and R. Stritmatter, "Writer verification of historical documents among cohort writers," in *International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pp. 314–319, IEEE, 2010.

[52] H. Arora, S. Lee, S. N. Srihari, and S. Cha, "Individuality of handwriting," *Journal of forensic science*, vol. 47, no. 4, pp. 1–17, 2002.

[53] V. Christlein and A. Maier, "Encoding CNN activations for writer recognition," in *Document Analysis Systems (DAS), 13th edition workshop*, 2018.

[54] O. Boiman, E. Shechtman, and M. Irani, "In defense of nearest-neighbor based image classification," *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[55] C. Adak and B. B. Chaudhuri, "Writer identification from offline isolated Bangla characters and numerals," in *13th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 486–490, IEEE, 2015.

[56] S. Fiel and R. Sablatnig, "Writer identification and retrieval using a convolutional neural network," in *International Conference on Computer Analysis of Images and Patterns*, pp. 26–37, Springer, 2015.

[57] R. Timofte, T. Tuytelaars, and L. Van Gool, "Naïve bayes image classification: beyond nearest neighbors," in *Asian Conference on Computer Vision*, pp. 689–703, Springer, 2012.

[58] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 1, pp. 105–119, 2010.

[59] T. Tuytelaars, M. Fritz, K. Saenko, and T. Darrell, "The NBNN kernel," in *IEEE International Conference on Computer Vision (ICCV)*, pp. 1824–1831, IEEE, 2011.

[60] R. Behmo, P. Marcombes, A. Dalalyan, and V. Prinet, "Towards optimal naïve bayes nearest neighbor," in *European Conference on Computer Vision*, pp. 171–184, Springer, 2010.

[61] P. Domingos and M. Pazzani, "On the optimality of the simple bayesian classifier under zero-one loss," *Machine learning*, vol. 29, no. 2-3, pp. 103–130, 1997.

[62] Z. Wang, Y. Hu, and L.-T. Chia, "Image-to-class distance metric learning for image classification," in *European Conference on Computer Vision*, pp. 706–719, Springer, 2010.

[63] T. Lindeberg, "Scale-space theory: A basic tool for analysing structures at different scales," *Journal of Applied Statistics*, vol. 21, pp. 224–270, 1994.

[64] M. Brown and D. Lowe, "Invariant features from interest point groups," in *Proc. BMVC*, pp. 23.1–23.10, 2002. doi:10.5244/C.16.23.

[65] C. Harris and M. Stephens, "A combined corner and edge detector.," in *Alvey vision conference*, vol. 15, pp. 10–5244, Citeseer, 1988.

[66] C. Djeddi, S. Al-Maadeed, A. Gattal, I. Siddiqi, A. Ennaji, and H. El Abed, "ICFHR2016 competition on multi-script writer demographics classification using "QUWI" database,"

[67] e-codices Virtual Manuscript Library of Switzerland, "St. Gallen, stiftsbibliothek." Available at `http://www.e-codices.ch`. Accessed: 01.10.2018.

[68] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration.," in *VISAPP (1)*, pp. 331–340, 2009.

[69] A. Fornes, A. Dutta, A. Gordo, and J. Llados, "The ICDAR 2011 music scores competition: Staff removal and writer identification," in *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1511–1515, IEEE, 2011.

[70] H. Mohammed, V. Märgner, T. Konidaris, and H. S. Stiehl, "Normalised local naïve bayes nearest-neighbour classifier for offline writer identification," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 01, pp. 1013–1018, IEEE, Nov 2017.

[71] G. Louloudis, N. Stamatopoulos, and B. Gatos, "ICDAR 2011 writer identification contest," in *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1475–1479, IEEE, 2011.

[72] G. Louloudis, B. Gatos, and N. Stamatopoulos, "ICFHR 2012 competition on writer identification challenge 1: Latin/greek documents," in *International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pp. 829–834, IEEE, 2012.

[73] G. Louloudis, B. Gatos, N. Stamatopoulos, and A. Papandreou, "ICDAR 2013 competition on writer identification," in *12th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1397–1401, IEEE, 2013.

[74] F. Kleber, S. Fiel, M. Diem, and R. Sablatnig, "CVL-database: An off-line database for writer retrieval, writer identification and word spotting," in *12th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 560–564, IEEE, 2013.

[75] S. Fiel, F. Kleber, M. Diem, V. Christlein, G. Louloudis, S. Nikos, and B. Gatos, "ICDAR2017 competition on historical document writer identification (Historical-WI)," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1377–1382, IEEE, 2017.

[76] A. Fornés, A. Dutta, A. Gordo, and J. Lladós, "CVC-MUSCIMA: a ground truth of handwritten music score images for writer identification and staff removal," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 15, pp. 243–251, Sep 2012.

[77] S. A. Maadeed, W. Ayouby, A. Hassane, and J. M. Aljaam, "QUWI: An Arabic and English handwriting dataset for offline writer identification," in *2012 International Conference on Frontiers in Handwriting Recognition*, pp. 746–751, Sept 2012.

[78] V. Christlein, D. Bernecker, and E. Angelopoulou, "Writer identification using VLAD encoded contour-zernike moments," in *13th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 906–910, IEEE, 2015.

[79] G. Simone, M. Pedersen, and J. Y. Hardeberg, "Measuring perceptual contrast in digital images," *Journal of Visual Communication and Image Representation*, vol. 23, no. 3, pp. 491–506, 2012.

[80] H. Mohammed, “Handwriting Analysis Tool v2.0 (HAT-2).” Available at `https://www.manuscript-cultures.uni-hamburg.de/hat.html`. Accessed: 01.10.2018.

[81] H. Mohammed, V. Märgner, and H. S. Stiehl, “Writer identification for historical manuscripts: Analysis and optimisation of a classifier as an easy-to-use tool for scholars from the humanities,” in *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pp. 534–539, IEEE, 2018.

[82] T. Seidensticker, “Audience certificates in Arabic manuscripts - the genre and a case study,” vol. 8, pp. 75–91, manuscript cultures, 2015.