A Digital Palaeographic Approach towards Writer Identification in the Dead Sea Scrolls

Maruf A. Dhali[1], Sheng He[1], Mladen Popović[2], Eibert Tigchelaar[3] and Lambert Schomaker[1]
[1]*Institute of Artificial Intelligence and Cognitive Engineering (ALICE), Faculty of Mathematics and Natural Sciences, University of Groningen, PO Box 407, 9700 AK, Groningen, The Netherlands*
[2]*Qumran Institute, Faculty of Theology and Religious Studies, University of Groningen, PO Box 407, 9700 AK, Groningen, The Netherlands*
[3]*KU Leuven, Faculty of Theology and Religious Studies, Leuven, Belgium*
{*m.a.dhali, s.he, m.popovic*}*@rug.nl, eibert.tigchelaar@kuleuven.be, l.r.b.schomaker@rug.nl*

Keywords: Dead Sea Scrolls, Handwritten Document Analysis, Digital Palaeography, Writer Identification, Handwriting Recognition, Pattern Recognition, Feature Representation, Machine Learning.

Abstract: To understand the historical context of an ancient manuscript, scholars rely on the prior knowledge of writer and date of that document. In this paper, we study the Dead Sea Scrolls, a collection of ancient manuscripts with immense historical, religious, and linguistic significance, which was discovered in the mid-20th century near the Dead Sea. Most of the manuscripts of this collection have become digitally available only recently and techniques from the pattern recognition field can be applied to revise existing hypotheses on the writers and dates of these scrolls. This paper presents our ongoing work which aims to introduce digital palaeography to the field and generate fresh empirical data by means of pattern recognition and artificial intelligence. Challenges in analyzing the Dead Sea Scrolls are highlighted by a pilot experiment identifying the writers using several dedicated features. Finally, we discuss whether to use specifically-designed shape features for writer identification or to use the Deep Learning methods on a relatively limited ancient manuscript collection which is degraded over the course of time and is not labeled, as in the case of the Dead Sea Scrolls.

1 INTRODUCTION

This paper is part of a pioneering project on the Dead Sea Scrolls that is sponsored by the European Research Council (EU Horizon 2020). This multidisciplinary project brings together the natural sciences, artificial intelligence, and the humanities in order to shed new light on ancient Jewish scribal culture by investigating two aspects of the scrolls' palaeography: handwriting recognition (the typological development of writing styles) and writer identification. Recognizing the handwriting would solve the *when*, *which* and *where* questions, and identifying the writer would end up answering the *who* question. These are the four most important perspectives (figure 1) in the study of palaeography and book history (Stokes, 2015). The digitization of the Dead Sea Scrolls (DSS) has opened the door for pattern recognition to be applied in answering those four questions (4-W). We aim to bridge the gap between computational science and traditional palaeography by solving the 4-W with a potential impact on digital palaeography beyond DSS studies.

With regard to choosing the right methodology, optical character recognition (OCR) methods are not sufficient for historical manuscripts. There are modern forms of neural networks (Deep Learning) having exceptionally good results (LeCun et al., 2015) in many aspects of pattern recognition including handwritten document analysis. But these performances can only be achieved in case of millions of training examples, which are contrary to the number of documents in many historical manuscripts, especially in the DSS.

Here, we will present preliminary results of writer identification in the DSS using several hand-crafted features. Although this gives us fast results without lengthy training on the limited labelled data of the DSS, they are certainly not the best results to be expected. We consider the results as a baseline measurement for later experiments. We suggest how to improve the results by exploiting the power of parameter-heavy machine learning methods using this small dataset. In solving this, we make a three-fold proposition of advanced statistical modelling, data-augmentation, and the use of pre-trained networks.

Dhali, M., He, S., Popović, M., Tigchelaar, E. and Schomaker, L.
A Digital Palaeographic Approach towards Writer Identification in the Dead Sea Scrolls.
DOI: 10.5220/0006249706930702
In *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2017)*, pages 693-702
ISBN: 978-989-758-222-6

1.1 Dead Sea Scrolls

The DSS are a collection of ancient damaged manuscripts that were discovered in the mid-20th century in the Judaean Desert, in between Jerusalem and the Dead Sea. Most were written over a period of almost four centuries (ca. 250 BC to ca. 135 AD) (Tigchelaar, 2010; Popović, 2012; Popović, 2015) in characters commonly known as the Hebrew alphabet, which actually derives from the Aramaic script (Yardeni, 2002). The manuscripts, written by many different writers, some of whom may have written multiple manuscripts, display a broad variety and development of different styles of this Hebrew-Aramaic script. The study of ancient handwriting provides the chronological framework, but the typological sequence of writing styles has to date not been systematically assessed for the DSS.

This project carries out the first systematic assessment of the palaeographic framework of the scrolls by combining two approaches. First, we will conduct new radiocarbon (^{14}C) dating on a number of physical samples of the scrolls, kindly provided to us by the Israel Antiquities Authority (IAA). Second, we will generate for the first time quantitative data for palaeographic handwriting recognition by means of Artificial Intelligence, using the Monk system, designed by Schomaker's research group at ALICE (Schomaker, 2016; Van der Zant et al., 2008; Bulacu and Schomaker, 2007). The challenging issue of writer identification in the DSS has not been systematically dealt with before. The tools of digital palaeography enable new, significant steps forward. In this paper, we focus on this second approach of digital palaeography.

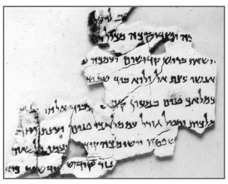

Figure 1: The four interesting questions for handwritten manuscript understanding (image from the DSS manuscript PAM 43.754, source: Brill scans).

1.2 Challenges in Digital Palaeography

In order to achieve both goals, i.e., handwriting recognition and writer identification, specific challenges at several levels of Computer Vision and Artificial Intelligence must be overcome. Initial analyses are needed for the proper extraction of characters (foreground, ink) from the background, which is mostly either animal skin or papyrus in the case of the DSS. Several image processing techniques need to be applied for optimum results of segmentation. Starting with edge detection, morphological operations, filling gaps and then finding connected components help to automatically segment the hand-written fragments. Then further processing can localize and extract the characters. Due to the difference in the textures of papyrus and animal-skin, individual measures must be taken on their distinctive periodic structures.

We have explored different feature-extraction techniques on the images of the DSS. Feature-representation maps the raw pixel intensity into a discriminant high-dimensional space (Mikolajczyk and Schmid, 2005; Li et al., 2015) in order to capture specific information of the characters which can be processed by algorithms in computers. This step is an important element in the field of computer vision and pattern recognition. There have been a lot of efforts to design discriminative and powerful features (Li et al., 2015). Though the (Deep) Learning-based feature representation may achieve better results in many cases, hand-crafted features have several advantages in the analysis of handwritten documents, especially for historical manuscripts. This is due to the amount of data in historical manuscript collections, which is usually not big enough to train deep neural networks. In contrast, the ImageNet data set (Deng et al., 2009) contains millions of samples for training the network. The challenge becomes even higher when the total number of usable pages comes to a count of hundreds in the DSS. To take the opportunities offered by the Deep Learning methods, the associated challenges need to be overcome in order to analyse the DSS.

2 DATA

2.1 Manuscript Images

We will use digital images of the DSS as our primary data. There are various sources for digital images of the DSS manuscripts. The source used in this study is kindly provided to us by Brill Publishers (Lim and Alexander, 1995). There are 2463 images in the Brill collection with varied resolutions from 600 by 600 pixels to 2800 by 3400 pixels, approximately. Another source is the high-resolution multi-spectral images of the DSS kindly provided to us by the Israel Antiquities Authority (IAA), which derive from their Leon Levy Dead Sea Scrolls Digital Library project. In this project the IAA produces multi-spectral images of scrolls fragments on both the recto and verso in 28 exposures, creating a file of 56 monochrome exposures per fragment. The system then generates a 57th file of a colour image that combines all visible wavelengths. The resolution of the files is 1,215 pixels per inch at a 1:1 ratio, capturing approximately 4 gigabytes of information per fragment (Shor et al., 2014).

The Brill images are single-layered grayscale images with 300 ppi (pixels per inch) on both axis. They have shadows and reflection from external lighting. Additionally the lighting throughout all the images is not uniform. Among the images, the ones containing several fragments are mostly not aligned in a horizontal way for text reading. This poses the issue of rotation variance in characters. Many of the images also contain paper calibration strips for scale representation and contemporary hand-written numbers. The digitization noise can also be noticed in many of the images (see figure 2).

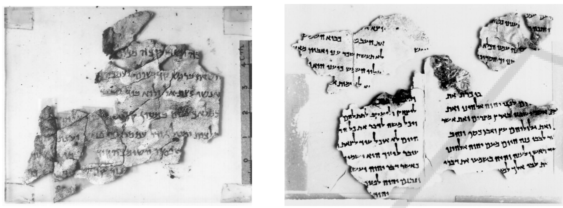

Figure 2: Two of the Brill images; PAM 40.456 (left) and PAM 40.531 (right); the images show digitisation noise, alignment issues with the small fragments, shadows near the border and lighting problems.

The IAA images are clear, properly aligned and free from the problem of lighting and shadow unlike the Brill ones. Additionally the different exposure bands of the IAA hold important underlying information regarding the fragments providing essential attributes for the scrolls. For example, one particular band provides clear information on the ink (foreground) whereas another one gives more details on the underlying leather/papyrus (background) on the retro side. Some bands are useful for the textual contents and some other bands give a better understanding on the textural properties of the scroll material. Extraction of this useful information is possible on both single images and multi-spectral-fused images. As a whole, digital image data has provided a new and broader perspective in the quantitative analysis and processing of the scrolls.

The scope of the current paper is limited to the images in the Brill collection, but in the near future we expect to publish our results on the digital data from the IAA's Leon Levy Dead Sea Scrolls Digital Library. The quality and the challenges of the Brill images can not be seen as a set-back, rather it would be a starting benchmark to the robustness of our work. Additionally, the possibility of using the IAA material will improve our results.

2.2 Ground Truth

Unlike many other historical manuscripts, the DSS do not have a structured and complete dataset nor the ground truths for testing. Before diving into any sort of computer aided writer identification, the ground truths must be there to analyse the results. To establish the ground truths, we need experts in the field and also their proper access to the data. We have this two-folded advantage in our group: first through the presence of palaeographic experts and second through the Monk system which is accessible through web browsers. By integrating these, we started to label the DSS image data for ground truths.

We have proposed two different methods for labelling. The first one is to detect the region of interest in the DSS images. The second one is to create the ground truth for character-labels. Both these tasks require manual labour from experts with palaeographic background knowledge on image level to pixel level. This is the bench-mark in identifying the writers and aligning the temporal developments in script style for the DSS. In this paper, we will only use the labelled regions of interest (we call them FragmROIs, by shortening the term *fragment region of interest*) in order to build algorithms to extract features (using available methods) and identify (recognition) writers. From the DSS images, FragmROIs were selected and labelled by the palaeographic experts using the Monk system (see figure 3). Those rectangular FragmROIs could consist of the entire text on an image, or of only a section of text selected from an image. Different FragmROIs from one and the same manuscript were labelled as stemming from one writer or scribe, unless palaeographers distinguished two scribes as writers of the manuscript.

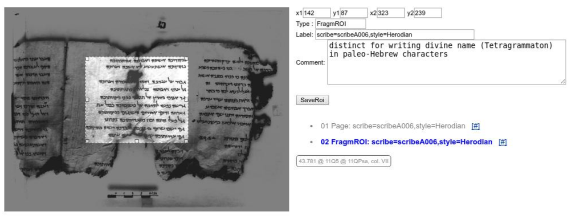

Figure 3: Using the Monk system, the palaeographic expert can select the region of interest (FragmROI), then put the associated attributes (scribe, style, comment etc.) to produce an XML file, which will later be used as labelled data. Currently the experts can only select a rectangular region, but the provision of choosing a polygonal region of interest will be added to the system in near future.

While labelling the writers, we have set up a provisional naming rule starting with the name *scribeAxxx*, where *xxx* are numerical values starting from 001. Each of the human labelled new-writers will be allocated with an individual value. The term *A* is put before the numerical values in order to preserve the tag of original labelling from the palaeographic experts. If at a later stage of our study, two of the writers are found to be the same one according to the system, then they can be referred to with the new name of *scribeBxxx* having two child-node of format *scribeAxxx*, preserving the original label.

The present pilot study is based on two distinct set of writers. The first set is a limited sample of 323 FragmROIs labelled as having been written by 13 scribes, namely the scribes of 1QIsa[a] columns 1-27, 1QIsa[a] columns 28-54, 1QS, 1QSa, 1QSb, 1QM, 1QpHab columns 1-12, 1QpHab columns 12 end-13, 4Q53, 4Q175, 11Q5, 11Q19 columns 2-5, 11Q20. We labelled them from *scribeA*001 to *scribeA*013. Distinct manuscripts were labelled as deriving from different writers, even though in several of the manuscripts of the first set palaeographers think that one and the same writer produced multiple manuscripts (Tigchelaar, 2002). To incorporate the palaeographic opinion, we then merged those 13 scribes into 7 scribes by introducing *scribeBxxx* series. Then we took the second set of 13 scribes with a limited sample of 124 FragmROIs labelled as *scribeA*014 to *scribeA*026 (the scribes of 4Q266, 4Q504, 11Q10, 1Q22, 4Q209, 4Q167, 4Q6, 4Q286, 4Q381, 4Q405, 4Q491, 4Q431, 4Q525). The main difference between these two sets are the amount of characters per scribe. The first set has a higher number of characters than the second set. Thus, for this pilot project, we have 447 FragmROIs labelled as 20 distinguishable scribes according to palaeographic opinion.

3 A PILOT PROJECT

The DSS image data has its own distinctive characteristics compared to other historical manuscript datasets. This data set has with quite a different appearance from, e.g. historical manuscripts such as the Medieval Palaeographic Scale data set (Monk, 2016) from a previous project (He et al., 2016) in three aspects: (1) the number of characters in fragments from some documents can be as low as one; (2) the ink of each character has been faded out over the course of time, making it more difficult to observe and process; (3) the large diversity and lack of uniformity among text blocks, presenting a challenge for analysis. In this section, we will present the methodology used in our pilot project in writer identification to benchmark our works in analysing the DSS.

3.1 Writer Identification

Identifying writers using computers has been done for decades (Plamondon and Lorette, 1989), which is a problem of recognizing the writer of a given document based on handwriting styles. A number of different features have been proposed and studied for writer identification on scripts from several languages including Dutch (Bulacu and Schomaker, 2007), English (Schomaker and Bulacu, 2004), Indic (Adak and Chaudhuri, 2015; Karunakara and Mallikarjunaswamy, 2011), and Arabic (Bulacu et al., 2007).

In the case of the DSS, we will be identifying the scribes behind the scrolls with Hebrew characters, and a hand-crafted feature specially for these characters is yet to be proposed and studied. Instead of designing a new feature, we initially started working with some of the existing textural-based and grapheme-based features. Textural-based features are based on the statistical information about slant and curvature of the handwritten characters, and grapheme-based features, inspired by the bag-of-words model, extract local structures and then map them into a common space (He and Schomaker, 2016). We briefly discuss the preprocessing techniques and the features used in this work in the following sections.

3.1.1 Preprocessing

As the feature extraction technique is applied on the binarized images, first we pre-processed the DSS images. Binarizing the Dead Sea Scrolls images is quite challenging, given their diverse intensity, similarity between ink and background traces, and image quality. We first started with Sobel edge detection (Sobel, 1990) and then removed the connected objects on the border to get rid of the markings. Morphological operation was then used followed by image thresholding. We used the global Otsu threshold selection method (Otsu, 1975) as it is efficient and parameterless (see figure 4).

3.1.2 Feature Representation

Previous studies showed that the textural-based feature extraction methods perform better than grapheme-based methods (He and Schomaker, 2016; He and Schomaker, 2017). Additionally, a more powerful approach was introduced by using the spatial co-occurrence among features (Bulacu and Schomaker, 2007; Ito and Kubota, 2010; Qi et al., 2014). The

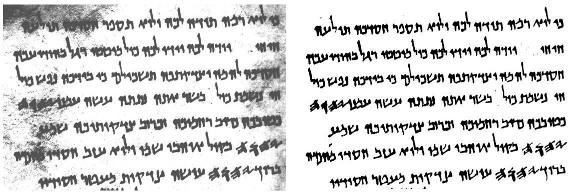

Figure 4: The left one is a FragmROI from Brill collection (PAM 43.787A) and the right one is the binarized image using the Otsu threshold selection method.

later idea has been extended in a previous work (He and Schomaker, 2017) with the introduction of the joint feature distribution principle (JFD principle). By accommodating these facts, we used eight textural-based methods (three of them following the JFD principle) and one grapheme-based method.

Hinge. The Hinge feature is the joint probability distribution of orientations of the legs of two contour fragments attached at a common-end pixel on the ink contours (Bulacu and Schomaker, 2007). Figure 5 shows two examples of the Hinge kernel on contour fragments with leg length l and the joint probability of the two orientations, α and β ($\alpha < \beta$), are quantized into a 2D histogram. Empirically we have set $l = 7$ and the number of bins of α and β is set to 23. Finally, the dimension of the feature vector is 253.

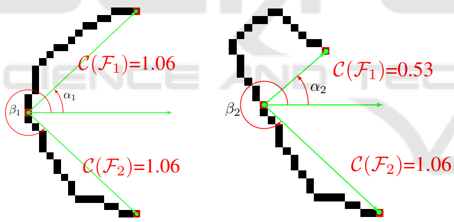

Figure 5: The two figures show two contour fragments with the same Hinge kernel (α_1=α_2 and β_1=β_2) but different fragment curvature values $\mathcal{C}(\mathcal{F}_c)$.

Co-occurrence Hinge (CoHinge). The CoHinge feature is the joint distribution of Hinge kernel following the JFD principle on two different points x_i and x_j with Manhattan distance l (figure 7) on the contours as equation 1.

$$\text{CoHinge}(x_i, x_j) = \big[\text{Hinge}(x_i), \text{Hinge}(x_j)\big] \quad (1)$$

Each Hinge kernel has two values α and β, and therefore, the CoHinge kernel has four values [$\alpha(x_i)$, $\beta(x_i)$, $\alpha(x_j)$, $\beta(x_j)$], which can be quantized into a 4D histogram. The Manhattan distance l is set to 7 based on our previous study (He and Schomaker, 2017). We set the number of bins of the angle to 10, and finally the dimension of the CoHinge feature is $10*10*10*10 = 10,000$.

Δ^nHinge. The Δ^nHinge is a rotation-invariant texture feature (He and Schomaker, 2014), computed by building a feature network with the differential operator between Hinge kernels as the kernel function K^i:

$$\begin{cases} \Delta^n\alpha(x_i) = & \frac{\Delta^{n-1}\alpha(x_i) - \Delta^{n-1}\alpha(x_i+\delta l)}{\delta l} \\ \Delta^n\beta(x_i) = & \frac{\Delta^{n-1}\beta(x_i) - \Delta^{n-1}\beta(x_i+\delta l)}{\delta l} \end{cases} \quad (2)$$

where (α, β) is the Hinge kernel and n is the order of the differential operator. Although many different features can be generated based on the feature network with different n, we work with the Δ^1Hinge feature with a feature-dimension of 780.

Quadruple Hinge (QuadHinge). QuadHinge is a powerful feature representation following the JFD principle, which incorporates the curvature information of the contour fragments in the Hinge kernel by computing a fragment curvature measurement (FCM) $\mathcal{C}(\mathcal{F}_c)$ for contour fragments (Benhamou, 2004).

Quill and QuillHinge. The Quill feature (Brink et al., 2012) is the joint probability distribution $p(\alpha, w)$ of the relation between ink direction α and the ink width w characterizing the writing material properties. The QuillHinge is an extension of the Quill and Hinge, which is the probability of $p(\alpha, \beta, w)$, resulting in a 3D histogram. We use the same parameters of the Quill and QuillHinge as the original paper (Brink et al., 2012), and the dimensions of Quill and QuillHinge are respectively 1600 and 31200.

Triple Chain Code. The triple chain code feature (Siddiqi and Vincent, 2010) is based on the chain code on a pixel of the writing contours, which is the one of eight directions where the next pixel is on, denoted from 1 to 8.

$$\text{TCC}(x_i, x_{i+l}, x_{i+2l}) = [\text{CC}(x_i), \text{CC}(x_{i+l}), \text{CC}(x_{i+2l})] \quad (3)$$

where $\text{CC}(x_i) \in \{1, 2, \cdots, 8\}$ is the chain code value on position x_i, and l is the Manhattan distance along the writing contours. We take the same value of $l = 7$, similar as the CoHinge feature. The feature dimension is 512.

Cloud Of Line Distribution (COLD). COLD is a curvature-free feature designed with the fact that writing contours can be approximated by a set of line segments obtained by the sequential polygonization algorithm (Siddiqi and Vincent, 2010) and the lengths and orientations of these straight lines can capture the handwriting styles. The high ordered curvature points on the writing contours are obtained using the method (Prasad et al., 2011), denoted by $\mathcal{P} = \{p_i(x_i, y_i), i = 0, 1, 2, \cdots, n\}$, where (x_i, y_i) is the coordinate of the point p_i (see figure 6). The line segments can be obtained between any pair of the dominant points (p_i, p_{i+k}), where k is the parameter which denotes the distance on the dominant sequence $\mathcal{P}$. Each line can be measured by a pair (θ, ρ) in the polar coordinate space, where θ is the line orientation and ρ is the line length. All the lines in a given handwritten document can form a distribution in the polar coordinate space and can be quantized into a log-polar histogram inspired by the Shape Context (Belongie et al., 2002). The features obtained with $k = 1, 2, 3$ in the log-polar space with the radius 7 and the angular intervals 12 are concatenated into one feature vector with the dimension: $7 * 12 * 3 = 252$.

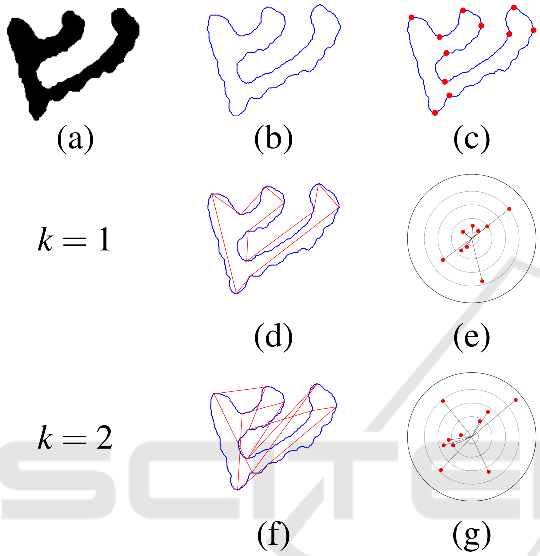

Figure 6: Illustration of the process of the COLD construction on the *Shin* character: (a) The given binarized connected component; (b) The contour extracted from the binarized image (a); (c) Detected dominant points (red points); (d) Line segments (red lines) obtained between pair dominant points when $k = 1$; (e) The distribution of lines from (d) in the polar coordinate space; (f) Line segments when $k = 2$ (Note that some long lines are not shown in order to make the figure more clear); (g) The distribution of lines from (f) in the polar coordinate space.

Junction Features. Junclets (He et al., 2015), a grapheme-based feature, is the stroke-length distribution in every directions from 0 to 2π around a reference point (see Figure 7) inside the ink trace. When the center point lies on the junction points, such as the fork points and high curvature points on the skeleton line of the ink strokes, the corresponding feature is the junction feature, which contain the junction information around the joint point. We have taken the stroke length distribution in 120 directions equidistantly sampled from 0 to 2π and the feature dimension of each junction is 120.

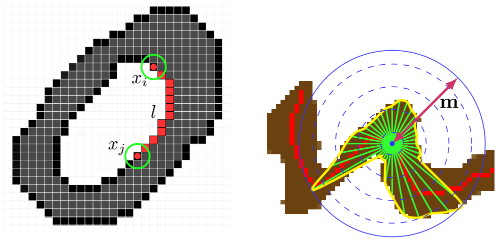

Figure 7: *Left:* Co-occurrence patterns on ink contours. *Right:* An illustration of the stroke-length distribution on a reference point (the blue point in the center). The green rays are the partial length in each direction, and the yellow curve is the distribution of the partial length in the polar space. The red line is the skeleton line of the stroke ink. **m** is the maximum measurable stroke length (He and Schomaker, 2017).

3.1.3 Identification Methodology

Writer identification is simply answering the *who* question. For a query document $Q_{scribeA_x}^{script_i}$, where $script_i$ is the script of the hand-written manuscript and $scribeA_x$ is the writer which we want to identify, all the documents in the database $dss_{scribeA_i}^{script_i} \in \mathcal{DSS}^{script_i}$ are sorted according to the feature distance between $Q_{scribeA_x}^{script_i}$ and $dss_{scribeA_i}^{script_i}$ to produce a hit-list where the writer of the top document is assigned to $scribeA_x$. Here $scribeA_i$ is the label of all the writers and for our case $script_i$ is a single script of Hebrew. The nearest neighbour classification method is performed using the leave-one-out (Brink et al., 2012; Siddiqi and Vincent, 2010) strategy. We take the query document out and sort the remaining documents according to their distance function to an output hit-list. For the distance function of the feature vectors, we have taken the χ^2 (chi-squared) distance for its better performance (Bulacu and Schomaker, 2007).

4 RESULTS

In this section we present the performance of writer identification based on the features and methodology explained in section 3.1. 447 FragmROIs were used for the pilot test. In the first set we took 323 FragmROIs with 13 writers labelled from *scribeA*001 to *scribeA*013 having 74, 33, 14, 13, 26, 37, 58, 3, 25, 24, 4, 10 and 2 FragmROIs respectively. The first set consists of writers with a large number of characters in their corresponding FragmROIs.

We first calculated the feature vectors for all the FragmROIs. Then we performed the writer identification using the methodology explained in 3.1.3. We produce the output hit-list of all the FragmROIs sorted out in accordance with their distance to the input FragmROI. The top-n performance is calculated when the query FragmROI is recognized as the writer of the FragmROI on the top n of the hit-list. For example, the top-10 hit-list signifies the overall percentage of finding the same writer as input within the first ten candidates (shortest distanced) of the output hit-list. Similarly the top-1 means the top-most candidate in the output hit-list corresponds to the same writer as the input. The performance of top-1 and top-10 hit-list for first set is presented in Table 1. According to the majority palaeographic opinion, *scribeA*001 and *scribeA*002 are the same scribe, and so also *scribeA*003, 004, 005, 008, 009 and *scribeA*010, 012, which are then labelled as *scribeB*001, *B*002 and *B*003 respectively. The result is presented in Table 2 for these seven scribes. Then we took the second set of 13 scribes with 124 FragmROIs. The amount of text is lower in this set than in the first one. The result is shown in Table 3. Finally, we took all the scribes together for testing. Table 4 presents the result of these 20 scribes together (i.e., *scribeB*001, *B*002, *B*003, *A*006, *A*007, *A*011, *A*013, *A*014 to *A*026). We briefly discuss the results and our propositions in the next section (Section 5).

Table 1: The top-1 and top-10 performance (in percentage) of writer identification for 13 scribes from *scribeA*001 to *scribeA*013.

Feature	**Top-1**	**Top-10**
Hinge	87.61	**97.83**
CoHinge	81.11	95.97
Δ^1Hinge	79.87	94.73
QuadHinge	**89.47**	96.59
Quill	80.80	93.80
QuillHinge	76.78	89.78
TripleChainCode	84.82	96.59
COLD	82.35	94.42
Junclet	81.42	95.04

Table 2: The top-1 and top-10 performance (in percentage) of writer identification for 7 scribes: *scribeB*001, *B*002, *A*006, *A*007, *B*003, *A*011 and *A*012 .

Feature	**Top-1**	**Top-10**
Hinge	92.26	**98.76**
CoHinge	**93.80**	97.52
Δ^1Hinge	90.71	96.28
QuadHinge	93.50	96.90
Quill	88.54	96.28
QuillHinge	88.85	96.90
TripleChainCode	92.26	98.14
COLD	91.33	96.28
Junclet	56.03	88.85

Table 3: The top-1 and top-10 performance (in percentage) of writer identification for another 13 scribes from *scribeA*014 to *scribeA*026, with limited text fragments.

Feature	**Top-1**	**Top-10**
Hinge	61.90	**92.06**
CoHinge	62.69	89.68
Δ^1Hinge	43.65	85.71
QuadHinge	**63.49**	90.47
Quill	48.38	89.51
QuillHinge	45.16	74.19
TripleChainCode	61.90	88.09
COLD	58.87	88.71
Junclet	31.45	76.74

Table 4: The top-1 and top-10 performance (in percentage) of writer identification for all 20 scribes.

Feature	**Top-1**	**Top-10**
Hinge	78.30	**94.40**
CoHinge	79.19	89.93
Δ^1Hinge	68.23	85.48
QuadHinge	**79.64**	89.04
Quill	71.58	86.57
QuillHinge	69.57	82.10
TripleChainCode	79.19	91.28
COLD	76.95	88.37
Junclet	40.71	72.93

5 DISCUSSIONS

5.1 Performance Evaluation

We presented the results of writer identification on a limited number of scribes. Of the shape based methods, the QuadHinge performs the best in the top-1 hit-lists for three out of four cases (only for the case of seven scribes in Table 2, CoHinge performs better with a small difference of 0.30% than QuadHinge), whereas the Hinge feature gives better result in all the top-10 hit-lists. The reason for this performance can be deduced from the design criteria of the features themselves. Hinge feature takes into account the joint probability distribution of the orientations of legs of two contour fragments from a common end pixel on ink contours, which proves to be a strong identical property for individual scribes of these ancient manuscripts. Additionally, the incorporation of FCM to the Hinge feature following the JFD principle gives the QuadHinge feature a boosted performance.

The directional measurement of the ink-trace width makes the Quill feature, which is quite informative on quill-based medieval scripts, a weak candidate for the DSS. This is due to the uniformity of the ink-trace in these documents coming from a probably fairly blunt tip of the ancient writing equipment. Consequently, the QuillHinge fails to provide a higher performance in this test set. The Δ^1Hinge has a limited performance, indicating that on Hebrew characters, loss of the angle with respect to the horizontal removes too much of the writer-specific information.

The grapheme-based feature, Junclets, gives lower performance than the cross-script writer identification (He et al., 2015) due to the lower variability in the stroke-length distribution in every direction around a reference point inside the ink of the DSS' Hebrew characters.

5.2 Propositions

The challenges in analysing the DSS are unique and unprecedented. Using the dedicated features (in 3.1.2), we found fast results without lengthy training on the limited labelled data of the DSS. But they are certainly not the best results to be expected. Especially when the amount of data is small with large variability (as in Table 3), the performance becomes lower. To overcome this situation, we need to consider a pragmatic approach incorporating several propositions.

1) Statistical modelling can be used in the case of the DSS, where the sample size is low and there are differences in the scholarly opinion of writers as well. We can use the differences in writing attributes of a set of different manuscripts to build a population model. A writer model can be built using the query manuscript. The classification is then carried out by evaluating the similarity of a further manuscript sample with respect to the models. We can build our provisional model, similar to the work of speaker identification (Leuzzi et al., 2016), as follows:

$$\Lambda(d(W_i,W_j)) = \frac{p_b(d(W_i,W_j))}{p_w(d(W_i,W_j))} \quad (4)$$

Here, $d(W_i,W_j)$ is the distance computed from W_i, the query writer to W_j, the suspected writer. Λ denotes the likelihood ratio over $d(W_i,W_j)$. The distribution of distances between the suspected writer and the population is denoted by $p_b(d(W_i,W_j))$, which can be referred as the between-group distance among the writers. $p_w(d(W_i,W_j))$ is the distribution of distances taken within different instances of the suspected writer (within-group distance). The collection of statistical models (Fisher, 1925), analysis of variance (ANOVA), can be used to analyse the within-group and between-group variances of the writers.

2) Another possibility is transfer learning (Long and Wang, 2015). It starts with the use of pre-trained networks on massive not-labelled handwriting collections. Such networks are trained to reconstruct images over (via) a very limited number of values (hidden units). After training, such a network implicitly *knows* a lot about historical handwritings in general. In a second stage, such a network is then applied to the DSS, using those *hidden unit vectors* as feature descriptors.

3) Data augmentation can be utilized in the processing of the DSS. If there is a believable random transformation of the DSS' text patterns, i.e., one that remains legible by humans, then for each natural sample of a character, a number of *N* derived random versions of it may be added to the training set, effectively enlarging the amount of labelled data. Known already in the nineties (Baird, 1992) this was later made popular in handwriting recognition later by the use of hidden Markov models (Varga and Bunke, 2003; Ha and Bunke, 1997).

5.3 Conclusions

In this paper, we have introduced digital palaeography of the DSS by presenting a pilot project, which is part of a pioneering multi-disciplinary project that brings together the natural sciences, artificial intelligence, and the humanities. By introducing the rule to establish ground-truths, we performed writer-identification tests using dedicated features on provisionally labelled data. The varying performance of results for different sets of writers led us to the propositions of statistical modelling, transfer learning, and data augmentation for this largely diverse collection of manuscripts.

We consider the results of this paper as a baseline measurement for our later experiments. We will combine both the aspect of specifically-designed shape features and the Deep Learning methods to produce fresh empirical data for the study of the DSS. Additionally, we will conduct new radiocarbon (^{14}C) dating on a number of physical samples of the scrolls. The outcome of ^{14}C dating will then be subjected to Bayesian statistics methods in combination with the results from temporal alignment using pattern recognition to reach more accurate and precise dating of the DSS.

ACKNOWLEDGEMENTS

The authors would like to thank Ruwan van der Iest (research assistant for the ERC project at the Qumran Institute) for his valuable inputs in labelling the regions of interest through the Monk-system.

This work has been supported by an ERC Starting Grant of the European Research Council (EU Horizon 2020): *The Hands that Wrote the Bible: Digital Palaeography and Scribal Culture of the DSS* (HandsandBible # 640497). Additional support comes from NWO (Netherlands Organisation for Scientific Research) and FWO (the Research Foundation Flanders): *Models of Textual Communities and Digital Palaeography of the DSS* (# 326-25-001).

REFERENCES

Adak, C. and Chaudhuri, B. B. (2015). Writer identification from offline isolated bangla characters and numerals. In *ICDAR*, pages 486–490. IEEE.

Baird, H. S. (1992). Document image defect models. In *Structured Document Image Analysis*, pages 546–556. Springer.

Belongie, S., Malik, J., and Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE PAMI*, 24(4):509–522.

Benhamou, S. (2004). How to reliably estimate the tortuosity of an animal's path: straightness, sinuosity, or fractal dimension? *Theoretical Biology*, 229(2).

Brink, A., Smit, J., Bulacu, M., and Schomaker, L. (2012). Writer identification using directional ink-trace width measurements. *PR*, 45(1):162–171.

Bulacu, M. and Schomaker, L. (2007). Text-independent writer identification and verification using textural and allographic features. *IEEE PAMI*, 29(4):701–717.

Bulacu, M., Schomaker, L., and Brink, A. (2007). Text-independent writer identification and verification on offline arabic handwriting. In *ICDAR*, volume 2, pages 769–773. IEEE.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR*.

Fisher, R. A. (1925). *Statistical methods for research workers*. Genesis Publishing Pvt Ltd.

Ha, T. M. and Bunke, H. (1997). Off-line, handwritten numeral recognition by perturbation method. *IEEE PAMI*, 19(5):535–539.

He, S., Samara, P., Burgers, J., and Schomaker, L. (2016). A multiple-label guided clustering algorithm for historical document dating and localization. *IEEE Transactions on Image Processing*, 25(11):5252–5265.

He, S. and Schomaker, L. (2014). Delta-n hinge: Rotation-invariant features for writer identification. In *ICPR*, pages 2023–2028.

He, S. and Schomaker, L. (2016). Writer identification using curvature-free features. *PR*.

He, S. and Schomaker, L. (2017). Beyond ocr: Multi-faceted understanding of handwritten document characteristics. *PR*, 63:321–333.

He, S., Wiering, M., and Schomaker, L. (2015). Junction detection in handwritten documents and its application to writer identification. *PR*, 48(12):4036–4048.

Ito, S. and Kubota, S. (2010). Object classification using heterogeneous co-occurrence features. In *European Conference on Computer Vision*, pages 701–714. Springer.

Karunakara, K. and Mallikarjunaswamy, B. (2011). Writer identification based on offline handwritten document images in kannada language using empirical mode decomposition method. *Writer*, 30(6).

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Leuzzi, F., Tessitore, G., Delfino, S., Fusco, C., Gneo, M., and Zambonini, G. (2016). A statistical approach to speaker identification in forensic phonetics field.

Li, Y., Wang, S., Tian, Q., and Ding, X. (2015). Feature representation for statistical-learning-based object detection: A review. *PR*, 48(11):3542–3559.

Lim, T. and Alexander, P. (1995). Volume 1. In *The Dead Sea Scrolls Electronic Library*. Brill.

Long, M. and Wang, J. (2015). Learning transferable features with deep adaptation networks. *CoRR, abs/1502.02791*, 1:2.

Mikolajczyk, K. and Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE PAMI*, 27(10):1615–1630.

Monk (2016). Medieval palaeographic scale data set (online collection).

Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27.

Plamondon, R. and Lorette, G. (1989). Automatic signature verification and writer identification - the state of the art. *Pattern recognition*, 22(2):107–131.

Popović, M. (2012). Qumran as scroll storehouse in times of crisis? a comparative perspective on judaean desert manuscript collections 1. *Journal for the Study of Judaism*, 43(4-5):551–594.

Popović, M. (2015). The ancient 'library' of qumran between urban and rural culture. In *The Dead Sea Scrolls at Qumran and the Concept of a Library*, pages 155–167. Brill.

Prasad, D. K., Quek, C., Leung, M. K., and Cho, S.-Y. (2011). A parameter independent line fitting method. In *ACPR*, pages 441–445.

Qi, X., Xiao, R., Li, C.-G., Qiao, Y., Guo, J., and Tang, X. (2014). Pairwise rotation invariant co-occurrence local binary pattern. *IEEE PAMI*, 36(11):2199–2213.

Schomaker, L. (2016). Design considerations for a large-scale image-based text search engine in historical manuscript collections. *it-Information Technology*, 58(2):80–88.

Schomaker, L. and Bulacu, M. (2004). Automatic writer identification using connected-component contours and edge-based features of uppercase western script. *IEEE PAMI*, 26(6):787–798.

Shor, P., Manfredi, M., Bearman, G. H., Marengo, E., Boydston, K., and Christens-Barry, W. A. (2014). The leon levy dead sea scrolls digital library: The digitization project of the dead sea scrolls. *Journal of Eastern Mediterranean Archaeology and Heritage Studies*, 2(2):71–89.

Siddiqi, I. and Vincent, N. (2010). Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features. *PR*, 43(11):3853–3865.

Sobel, I. (1990). An isotropic 3×3 image gradient operator. *Machine Vision for three-demensional Sciences*.

Stokes, P. A. (2015). Digital approaches to paleography and book history: some challenges, present and future. *Frontiers in Digital Humanities*, 2:5.

Tigchelaar, E. (2002). In search of the scribe of 1qs. In *Emanuel*, pages 339–352. Brill.

Tigchelaar, E. (2010). Dead sea scrolls. In *The Eerdmans Dictionary of Early Judaism*, pages 163–180. Eerdmans.

Van der Zant, T., Schomaker, L., and Haak, K. (2008). Handwritten-word spotting using biologically inspired features. *IEEE PAMI*, 30(11):1945–1957.

Varga, T. and Bunke, H. (2003). Effects of training set expansion in handwriting recognition using synthetic data. In *Proc. 11th Conf. of the Int. Graphonomics Society*, pages 200–203. Citeseer.

Yardeni, A. (2002). *The book of Hebrew script: history, palaeography, script styles, calligraphy & design*. Oak Knoll Pr.