EasyChair Preprint

№ 1324

Writer Characterization and Identification in Short Modern and Historical Documents: Reconsidering Paleographic Tables

Shira Faigenbaum-Golovin, David Levin, Eli Piasetzky and Israel Finkelstein

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

July 22, 2019

Writer Characterization and Identification of Short Modern and Historical Documents: Reconsidering Paleographic Tables

Shira Faigenbaum-Golovin, David Levin
School of Mathematical Sciences,
Tel Aviv University
Tel Aviv 69978
Israel
alecsan1@post.tau.ac.il,
levin@tau.ac.il

Eli Piasetzky
The Sackler School of Physics and Astronomy
Tel Aviv University
Tel Aviv 69978
Israel
eip@tauphy.tau.ac.il

Israel Finkelstein
Department of Archaeology and Ancient Near Eastern Civilizations,
Tel Aviv University,
Tel Aviv 69978
Israel
fink2@tauex.tau.ac.il

ABSTRACT

Handwriting is considered a unique "fingerprint" that characterizes a scribe (it is even used as evidence in modern forensics). In paleography (the study of ancient writing), it is presumed that each writer has a one prototype for each letter in the alphabet. Commonly, for ancient inscriptions, letters are organized into paleographic tables (where the rows are the alphabet letters, and the columns represent the examined inscriptions). These tables play a significant role in dating inscriptions based on their resemblance to columns in the table. In this paper, we argue that each scribe "fingerprint" is not represented by a single character prototype, but in fact by a distribution of characters. We introduce a framework for automatically identifying the writer style and constructing paleographic tables based on character histograms. Subsequently, we propose a method for comparing short documents utilizing letter distribution. We demonstrate the validity of the methods on two handwritten datasets: Modern and Ancient Hebrew pertaining to the First Temple period. Our methodology on the ancient dataset enables us to provide additional evidence concerning the level of literacy in the kingdom of Judah ca. 600 BCE.

CCS CONCEPTS

• **CCS → Computing methodologies → Artificial intelligence → Computer vision → Computer vision problems;** Object identification; Shape inference

KEYWORDS

Handwriting identification, hardwiring comparison, epigraphy, paleographic tables, historical documents, Hebrew ostraca, First Temple period, Iron Age, Judah

ACM Reference format:

S. Faigenbaum-Golovin, D. Levin, E. Piasetzky, I. Finkelstein. 2019. Writer Characterization and Identification in Short Modern and Historical Documents: Reconsidering Paleographic Tables. In DocEng '19: ACM Symposium on Document Engineering 2019, September 23–29, 2019, Berlin, Germany, 4 pages.

1 INTRODUCTION

Each person has an individual writing style that evolves from the tone dictated during scribal education. Later, with additional practice, individual handwriting is established. The writing style is considered to be a unique fingerprint. Accordingly, handwriting serves as significant evidence in standard forensics practice - to (manually) assess the identity of an author during trials [1].

Writing differentiation has various other targets, such as signature authentication, ancient document matching [2, 3] and identification [4, 5]. While the existing solutions successfully analyze modern writings [6-9], ancient documents raise challenges that stem from the poor level of preservation of the inscriptions (the letters are often blurred, partially erased and stained; for example, see the ink texts from Arad from ca. 600 BCE shown in Fig. 1). In addition, usually a modest number of inscriptions is available, and they are rather short. These circumstances pose a challenge for Neural Networks methods.[1] Recently, our research group suggested an algorithmic framework to compare inscriptions from the First Temple period [4]. We showed that even with limited data, a statistically significant result can be achieved.

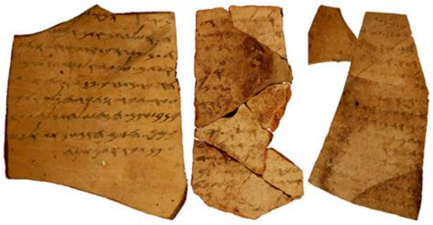


Figure 1: Ink inscriptions from Arad, ca. 600 BCE [10], from left to right: Nos. 24, 5, and 40.

Handwriting fingerprint characterization is an important step towards writer authentication. In the classical study of handwriting, characterization is based on the idea that the writer can be represented by a single character for each letter in the

[1] Although it is possible to overcome the challenge by transfer learning, trained on modern writing, we choose to deal with this problem here, and in previous research [5], by utilizing classical tools form statistics.

alphabet [11]. Later, these (manually located) characters are set into paleographic tables (where the rows are letters in the alphabet and the columns are the examined inscriptions; Fig. 2). These tables usually contain script according to date, or to a corpus. The importance of these tables is not only in the fact that they enable examination of the writing style of different scribes, but also in that they provide a rare opportunity for broader analysis, e.g., inspecting the influence of scribal schools in diverse geographical locations, as well as analyzing the development of the alphabet across decades. In addition, undated inscriptions can be dated by the similarity to one of the columns in the tables (e.g., according to the stance of the letter towards the line).

Constructing paleographic tables manually may be inefficient and result in inaccurate results. In this work, we argue that the basic concept of a single character cannot represent writing style variability. Therefore, we suggest representing writing diversity via a histogram of characters with respect to the dominant prototypes. We utilize this concept to build histogram-based paleographic tables. In addition, we propose addressing the task of writer comparison through viewing the characters as data sampled from various distributions and comparing them using two-sample t-test. We verify the validity of our approach by applying the algorithm to modern texts (a number of contemporary texts written by individuals known to us). Later, we apply it to ancient texts pertaining to the First Temple period (ca. 600 BCE).

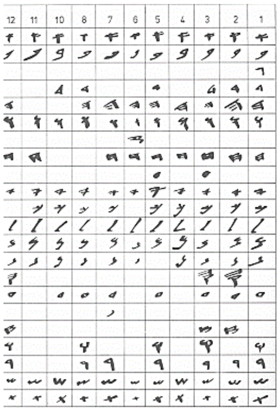

12
11
10
8
7
6
5
4
3
2
1

Figure 2: Fragment of manually created paleographic table of Arad corpus. The rows are the alphabet paleo-Hebrew letters; the columns are the examined inscriptions (from [10]).

2 PROPOSED ALGORITHM

Let us first introduce several notations which will be used in this paper. We denote alphabet signs as "letters", a (whereas in paleo-Hebrew alphabet $a \in [1,22]$). Given an inscription, the written instances of a specific letter a will be addressed as "characters," i.e. $\{l_j^a\}_j$. Thus, given a set of text t_i, their corresponding characters are $\{l_{i,j}^a\}$.

2.1 Revising Paleographic Tables

The classical paleographic tables contain information pertaining to the letters and their typical representation in the available inscriptions. In fact, the representative characters are medoids, selected manually by the paleographers. Besides the apparent problem - of choosing medoids accurately and efficiently - we raise here an additional issue: the diversity of the characters stemming from handwriting variability.

Let us consider the characters as scattered data sampled from a manifold in high dimensional space. Following cluster analysis considerations, it is clear that a single medoid does not represent the entire sampled population. The complexity of the data, reflected in the geometry of the underlying manifold, influence the number of samples needed to represent the set. Thus, the problem shifts to finding the number of representative clusters k, and the corresponding signature in the form of a character histogram. We summarize this process in Algorithm 1 below:

Algorithm 1 Handwriting Style Characterization

```
Input: Set of images of binarized, registered characters $L = \{l_{i,j}^a\} \in R^{n \times m}$ of letter type $a \in [1,22]$, from $T = \{t_j\}_{j \in J}$ texts
Output: List of the character medoids and the character histograms
Flatten character representation to vectors $\{\bar{l}_{i,j}^a\} \in R^{n \cdot m \times 1}$
$\hat{l}_{i,j}^a = \text{MDS}(\bar{l}_{i,j}^a, 1)$ % project the char. to 1-dimentional space
$k = \#$ of clusters % estimate num. of clusters via silhouette criterion
$m_k^a = medoids(\hat{l}_{i,j}^a)$
$h_j^a = hist(\hat{l}_{i,j}^a, m_k^a)$ % build a histogram of char. with chosen medoids
return $m_k^a, \{h_j^a\}$
```

First, the representative characters are found as the k-medoids of the characters data, which are projected to one-dimensional space via MDS [12]. The MDS method performs dimensional reduction, while preserving the distances between the points. The consideration behind choosing to project the data to one-dimensional space is based on the idea that the character data, can be viewed as lying approximately on a linear manifold. Thus, a single dimension, as well as linear dimensional reduction will maintain the significant properties of the data. This was empirically tested on modern data as described in the *Handwriting Identification* section. Next, we use the medoids, to construct the handwriting distribution. We apply Algorithm 1 for each letter of the alphabet and construct the histogram-based paleographic table.

2.2 Handwriting Identification

The uniqueness of the handwriting style can be exploited for writer identification tasks. Here, we propose a simple yet efficient writer comparison algorithm. We treat the characters as samples from an unknown distribution. Thus, given binarized and registered characters from two texts, we find their flattened vector representation and project them to one-dimensional space using the MDS method. Later, we perform a pairwise comparison, by comparing their character distributions utilizing a two-sample t-test. We test the null hypothesis H$_0$, that *two given inscriptions were written by the same author*. A corresponding p-value (P) is deduced. If $P \leq \alpha$, we reject H$_0$ and accept the competing hypothesis of two different authors; otherwise, we remain undecided. This hypothesis is tested for several different alphabet letters, and later the p-values of the various independent experiments are combined by using the Fisher method [13]. Below we summarize the details in Algorithm 2:

```
Algorithm 2 Handwriting Comparison
Input: Pair of binarized, registered characters $L = \{l^a_{i,j}\}_{a\in[1,22]} \in R^{n\times m}$ of letter type, from $T = \{t_j\}_{j\in[1,2]}$ texts, $\alpha$ significance level
Output: p-value of H0: two texts were written by the same author
for each letter type $a \in [1,22]$
    Flatten letter representation to vectors $\{\bar{l}^a_{i,j}\} \in R^{n\cdot m\times 1}$
    $\hat{l}^a_{i,j} = \mathrm{MDS}(\bar{l}^a_{i,j})$ % project the letters to one-dimensional space
    $pval_a = \mathrm{ttest2}(\{\hat{l}^a_{i,1}\}_i, \{\hat{l}^a_{i,2}\}_i)$
$pval_a = Fisher(\{pval_a\}_{a\in[1,22]})$ % combine p-values of all letters
return p-value, p-value$\le \alpha$
```

3 RESULTS AND DISCUSSION

We applied our paleographic table and writer identification methodology to two corpora: modern Hebrew and ancient handwritings - the Arad paleo-Hebrew corpus of ink inscriptions dating to ca. 600 BCE [10]. The experiment with the modern text was used for validation of the proposed framework. Thus, our study had two targets. First, we wanted to test whether performing a linear dimensional reduction to one-dimensional space would maintain the writer characterizing information. In our tests, we set the significance level at 0.15. Second, using this low dimensional representation we constructed revised paleographic tables.

3.1 Modern Hebrew Experiment

Our experiment opened with testing our framework on a set of samples collected from 18 contemporary writers of modern Hebrew [14]. Each individual filled an alphabet table consisting of ten occurrences of each of the 22 letters in the Hebrew alphabet (for additional information regarding this dataset see [4]). The collected handwritings were binarized and their characters labeled by the corresponding alphabet letter. From this raw data, a series of "simulated" inscriptions were created. Due to the need to test both same-writer and different-writer scenarios, the data for each writer was split. Furthermore, in order to imitate a common situation in the ancient corpus, where the scarcity of data is prevalent, each simulated inscription used only 4 letters (our ancient texts had 5 letters comparison in median). In total, 250 inscriptions were "simulated." The Modern Hebrew experiment resulted in 7% of FP, and 9% of FN. These results indicate the validity of our writer separation algorithmic sequence and affirm our choice to use one-dimensional data and MDS.

Subsequently, the algorithm for paleographic table creation was applied. Some of the results can be seen in Fig. 3 (the complete modern paleographic table can be found at [15]). We measure the accuracy of the k-medoids via the silhouette index, which was 0.86 in our case. As can be seen, the handwritten diversity is reflected in the number of prototypes of a letter in the table. Thus, if the handwriting is uniform, we obtain a single prior; on the other hand, for non-standardized writing – there will be multiple representative characters.

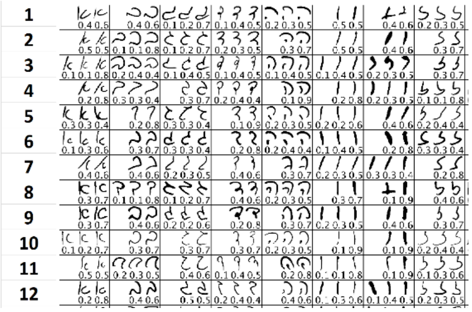

1
2
3
4
5
6
7
8
9
10
11
12

Figure 3: Automatically created paleographic table of modern Hebrew handwriting. Each cell contains the representative characters as well as its weight in the characterization.

3.2 Ancient Hebrew Experiment

We utilized our algorithmic framework to analyze 18 texts[2] (e.g., Fig. 1) from the Arad fortress, located in arid southern Judah, on the border of the kingdom with Edom [10]. The inscriptions contain military commands regarding the movement of troops and provisions (wine, oil, and flour) set against the background of the stormy events of the final years before the fall of Judah (ca. 600 BCE).

Due to the low level of preservation of the inscriptions, we had to perform handwriting segmentation prior to performing other tasks. After examining existing algorithms [16], we decided to use a stroke reconstruction algorithm [17] to create character binarization. We reconstructed seven letters – *alep*, *he*, *waw*, *yod*, *lamed*, *shin* and *taw* – since they were the most prominent in our database. Thus, 427 characters were restored (the database can be found in [18]). For additional detailed information regarding this dataset, see Table S3 in [4].

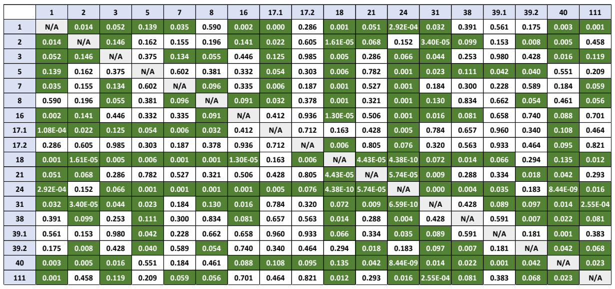
The handwriting comparison results are summarized in the Table in Fig. 4. The inscription numbers head the rows and columns of the table, with the intersection cells providing the comparisons P. The cells with P ≤ 0.15 are marked in green, indicating that the two texts are considered to be written by different authors. We reiterate that when P > 0.15 we cannot claim that they were written by a single author.

Subsequently, we can estimate the minimal number of writers in the tested inscriptions - as the size of the largest clique in a graph - where the nodes are the inscriptions, and there exists a vertex between two nodes if P ≤ 0.15. From the table, we see that there are seven cliques of size four, 12 cliques of size five, eight cliques of size six, and even one clique of size seven[3]. This result sheds additional light on the literacy level in Judah ca. 600. While

[2] Inscriptions 1, 2, 3, 5, 7, 8, 16, 17, 18, 21, 24, 31, 38, 39, 40, and 111, which are relatively legible, and have sufficient numbers of characters for examination, were tested. Two of the inscriptions (Nos. 17 and 39) are inscribed on both sides of the sherd, bringing the number of texts under investigation to 18.

[3] The largest pairwise distinct writer groups: (2,21,31,39.2,40); (1,2,3,18,31,40); (1,2,16,18,31,40); (1,2,18,21,31,40); (1,3,7,18,24,111); (3,7,8,18,24,111); (1,16,18,24,31,40); (1,18,21,24,31,40); (3,8,18,24,31,111); (1,3,18,24,31,40,111). A complete list of all the cliques can be found in [15].

in a previous analysis [4] it was shown that at least four hands can be detected by the algorithm (and six hands when adding analysis of the texts in the tested inscriptions), here we prove that in this modest set of inscriptions the lower bound for the number of scribes is seven). This significantly increases our knowledge about the literacy rates in Judah ca. 600 BCE.



	1	2	3	5	7	8	16	17.1	17.2	18	21	24	31	38	39.1	39.2	40	111
1	N/A	0.014	0.052	0.139	0.035	0.590	0.002	0.000	0.286	0.001	0.051	2.92E-04	0.032	0.391	0.561	0.175	0.003	0.001
2	0.014	N/A	0.146	0.162	0.155	0.196	0.141	0.022	0.605	1.61E-05	0.068	0.152	3.40E-05	0.099	0.153	0.008	0.005	0.458
3	0.052	0.146	N/A	0.375	0.134	0.055	0.446	0.125	0.985	0.005	0.286	0.066	0.044	0.253	0.980	0.428	0.016	0.119
5	0.139	0.162	0.375	N/A	0.602	0.381	0.332	0.054	0.303	0.006	0.782	0.001	0.023	0.111	0.042	0.040	0.551	0.209
7	0.035	0.155	0.134	0.602	N/A	0.096	0.335	0.006	0.187	0.001	0.527	0.001	0.184	0.300	0.228	0.589	0.184	0.059
8	0.590	0.196	0.055	0.381	0.096	N/A	0.091	0.032	0.378	0.001	0.321	0.001	0.130	0.834	0.662	0.054	0.461	0.056
16	0.002	0.141	0.446	0.332	0.335	0.091	N/A	0.412	0.936	1.30E-05	0.506	0.001	0.016	0.081	0.658	0.740	0.088	0.701
17.1	1.08E-04	0.022	0.125	0.054	0.006	0.032	0.412	N/A	0.712	0.163	0.428	0.005	0.784	0.657	0.960	0.340	0.108	0.464
17.2	0.286	0.605	0.985	0.303	0.187	0.378	0.936	0.712	N/A	0.006	0.805	0.076	0.320	0.563	0.933	0.464	0.095	0.821
18	0.001	1.61E-05	0.005	0.006	0.001	0.001	1.30E-05	0.163	0.006	N/A	4.43E-05	4.38E-10	0.072	0.014	0.066	0.294	0.135	0.012
21	0.051	0.068	0.286	0.782	0.527	0.321	0.506	0.428	0.805	4.43E-05	N/A	5.74E-05	0.009	0.288	0.334	0.018	0.042	0.293
24	2.92E-04	0.152	0.066	0.001	0.001	0.001	0.001	0.005	0.076	4.38E-10	5.74E-05	N/A	0.000	0.004	0.035	0.183	8.44E-09	0.016
31	0.032	3.40E-05	0.044	0.023	0.184	0.130	0.016	0.784	0.320	0.072	0.009	6.59E-10	N/A	0.428	0.089	0.097	0.014	2.55E-04
38	0.391	0.099	0.253	0.111	0.300	0.834	0.081	0.657	0.563	0.014	0.288	0.004	0.428	N/A	0.591	0.007	0.022	0.081
39.1	0.561	0.153	0.980	0.042	0.228	0.662	0.658	0.960	0.933	0.066	0.334	0.035	0.089	0.591	N/A	0.181	0.001	0.383
39.2	0.175	0.008	0.428	0.040	0.589	0.054	0.740	0.340	0.464	0.294	0.018	0.183	0.097	0.007	0.181	N/A	0.042	0.068
40	0.003	0.005	0.016	0.551	0.184	0.461	0.088	0.108	0.095	0.135	0.042	8.44E-09	0.014	0.022	0.001	0.042	N/A	0.023
111	0.001	0.458	0.119	0.209	0.059	0.056	0.701	0.464	0.821	0.012	0.293	0.016	2.55E-04	0.081	0.383	0.068	0.023	N/A

Figure 4: Arad handwriting comparison table. Green boxes mark the comparison that rejects the single writer H_0.

Later, we created an automatic histogram-based, paleographic table for the 18 Arad inscriptions, based on the seven reconstructed letters. The table can be seen in Fig. 5. The silhouette index of the k-medoids was 0.96.

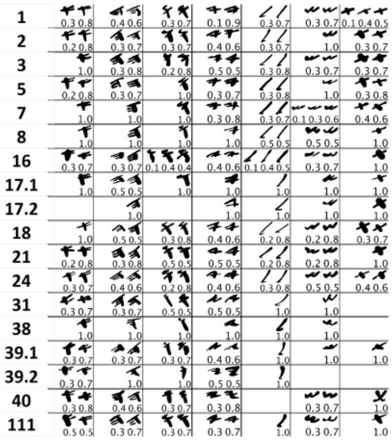

1 0.3 0.8 0.4 0.6 0.3 0.7 0.1 0.9 0.3 0.7 0.3 0.7 0.1 0.4 0.5
2 0.2 0.8 0.3 0.7 0.3 0.7 0.4 0.6 0.3 0.7 1.0 0.3 0.7
3 1.0 0.3 0.8 0.2 0.8 0.5 0.5 0.3 0.8 0.3 0.7 0.3 0.7
5 0.2 0.8 0.3 0.7 1.0 0.3 0.7 0.3 0.8 1.0 0.3 0.8
7 1.0 1.0 1.0 0.3 0.8 0.3 0.7 0.1 0.3 0.6 0.4 0.6
8 1.0 1.0 1.0 1.0 0.5 0.5 0.5 0.5 1.0
16 0.3 0.7 0.3 0.7 0.1 0.4 0.4 0.4 0.6 0.1 0.4 0.5 0.3 0.7 1.0
17.1 1.0 0.5 0.5 1.0 1.0 1.0 1.0 1.0
17.2 1.0 1.0 1.0 1.0 1.0
18 1.0 0.5 0.5 0.3 0.8 0.4 0.6 0.2 0.8 0.2 0.8 0.3 0.7
21 0.2 0.8 0.3 0.8 0.5 0.5 0.5 0.5 0.2 0.8 0.2 0.8 1.0
24 0.3 0.7 0.4 0.6 0.2 0.8 0.4 0.6 0.3 0.8 0.5 0.5 0.4 0.6
31 0.3 0.7 0.3 0.7 0.5 0.5 0.5 0.5 1.0 1.0
38 1.0 1.0 1.0 1.0 1.0 1.0
39.1 0.3 0.7 0.3 0.7 0.3 0.7 0.4 0.6 1.0 1.0 1.0
39.2 0.3 0.7 1.0 1.0 0.5 0.5 1.0
40 0.3 0.8 0.4 0.6 0.3 0.7 0.3 0.8 0.3 0.7 1.0
111 0.5 0.5 0.3 0.7 0.3 0.7 0.3 0.7 1.0 0.3 0.7 1.0

Figure 5: Automatically created paleographic table of the Arad inscriptions. Each cell contains the representative characters as well as their weight in the characterization.

4 CONCLUSIONS

In this paper, we proposed two algorithms: one for writer identification and another for generating histogram-based paleographic table. The constructed paleographic table offers a more accurate and better representation of the individual style of a scribe. In most cases, we see that handwriting diversity is expressed in the table by multiple prototypes. The automatization of the table creation process provides an efficient and simple way to address the paleographic task.

In addition, the suggested writer differentiation algorithm, simple yet efficient, shows good results on modern data. The results of applying the algorithm on ancient inscriptions from Arad, indicated that *at least seven hands* wrote the 18 inspected inscriptions (as compare to previous analysis result in [4]). This algorithmic evidence increases the knowledge regarding the level of literacy in Judah ca. 600 BCE, and provides a possible stage setting for a compilation of biblical texts in that time.

ACKNOWLEDGMENTS

This study was supported by a generous donation from Mr. Jacques Chahine, made through the French Friends of Tel Aviv University. We were also funded by the ISF grant 2062/18 and the Dan David Foundation. This research was made possible by the dedicated reconstruction work of Ms. Ma'ayan Mor. The kind assistance of Ms. Myrna Pollak is greatly appreciated. Additionally, we would like to thank Barak Sober, Arie Shaus and Eythan Levy for fruitful discussions and comments. Ostracon images are courtesy of the Institute of Archaeology, Tel Aviv University (photographer: Michael Cordonsky) and of the Israel Antiquities Authority. Paleographic tables and facsimiles are courtesy of the Israel Exploration Society [10].

REFERENCES

[1] U. Sharma, and V. Rana. 2017. Handwriting Comparison in Court Trials: a Need for Forensic Upgradation. *International Journal of Advanced Research in Computer Science*, 8. DOI: http://dx.doi.org/10.26483/ijarcs.v8i9.5077

[2] L. Wolf, N. Dershowitz, L. Potikha, T. German, R. Shweka, and Y. Choueka. 2011. Automatic Palaeographic Exploration of Genizah Manuscripts, 3, Books on Demand (BoD).

[3] G. Levi, P. Nisnevich, A. Ben-Shalom, N. Dershowitz and L. Wolf. 2018. A Method for Segmentation, Matching and Alignment of Dead Sea Scrolls. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*.DOI: https://doi.org/10.1109/WACV.2018.00029.

[4] S. Faigenbaum-Golovin, A. Shaus, B. Sober, D. Levin, N. Na'aman, B. Sass, E. Turkel, E. Piasetzky, and I. Finkelstein. 2016. Algorithmic handwriting analysis of Judah's military correspondence sheds light on composition of biblical texts. *Proceedings of the National Academy of Sciences*, 113, 4664-4669. DOI: https://doi.org/10.1073/pnas.1522200113.

[5] A. Shaus, and E. Turkel. 2017. Writer Identification in Modern and Historical Documents via Binary Pixel Patterns, Kolmogorov--Smirnov Test and Fisher's Method. *Electronic Imaging*, 2017, 203-211. DOI: https://doi.org/10.2352/ISSN.2470-1173.2017.14.HVEI-144.

[6] P. Dondi, A. Danani, L. Lombardi, M. Malagodi, and M. Licchelli. 2018. Handwriting identification of short historical manuscripts. In *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*. DOI: https://doi.org/10.1109/DAS.2018.45.

[7] A. Abdalhaleem, B. K. Barakat, and J. El-Sana. 2018. Case Study: Fine Writing Style Classification Using Siamese Neural Network. In *2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR)*. DOI: https://doi.org/10.1109/ASAR.2018.8480212.

[8] A. A. Brink, J. Smit, M. L. Bulacu, and L. R. B. Schomaker. 2012. Writer identification using directional ink-trace width measurements. *Pattern Recognition*, 45, 162-171. DOI: https://doi.org/10.1016/j.patcog.2011.07.005.

[9] S. He, and L. Schomaker. 2017. Writer identification using curvature-free features. *Pattern Recognition*, vol. 63, pp. 451-464. DOI: https://doi.org/10.1016/j.patcog.2016.09.044.

[10] Y. Aharoni, and J. Naveh. 1981. Arad inscriptions, Israel exploration society.

[11] A. Shaus, and E. Turkel. 2017. Towards letter shape prior and paleographic tables estimation in Hebrew First Temple period ostraca. In *Proceedings of the 4th International Workshop on Historical Document Imaging and Processing*. DOI: https://doi.org/10.1145/3151509.3151511.

[12] T. F. Cox, and M. A. A. Cox. 2000. Multidimensional scaling, Chapman and hall/CRC.

[13] R. A. Fisher. 1925. Statistical methods for research workers Oliver and Boyd. *Edinburgh, Scotland*, 6.

[14] Modern Hebrew Dataset. (Online): www-nuclear.tau.ac.il/ eip/ostraca/DataSets/Modern_Hebrew.zip. Last accessed: 27/5/2019.

[15] Supplemental Material. (Online): https://docs.wixstatic.com/ugd/b9b8a4_cd710d52d3a942d49f0c108b5d126e32.pdf. Last accessed: 27/5/2019.

[16] S. Faigenbaum, A. Shaus, B. Sober, E. Turkel, and E. Piasetzky, "Evaluating glyph binarizations based on their properties," in *Proceedings of the 2013 ACM symposium on Document engineering*, 2013.

[17] B. Sober, and D. Levin. 2017. Computer aided restoration of handwritten character strokes. *Computer-Aided Design*, 89, 12-24. DOI: https://doi.org/10.1016/j.cad.2017.04.005.

[18] Ancient Hebrew Dataset. (Online): www-nuclear.tau.ac.il/eip/ostraca/DataSets/Arad_Ancient_Hebrew.zip. Last accessed: 27/5/2019.