https://doi.org/10.2352/issn.2168-3204.2023.20.1.32

Character-based Writer Verification of Ancient Hebrew Square-script Manuscripts: On Edge-direction Feature

Tabita L. Tobing; Department of Information Security and Communication Technology, NTNU - Norwegian University of Science and Technology; Gjøvik, Norway.
Pavel Škrabánek; Institute of Automation and Computer Science, Brno University of Technology; Brno, Czech Republic
Sule Y. Yayilgan; Department of Information Security and Communication Technology, NTNU - Norwegian University of Science and Technology; Gjøvik, Norway.
Sony George; Department of Computer Science, NTNU - Norwegian University of Science and Technology; Gjøvik, Norway.
Torleif Elgvin; NLA University College; Oslo, Norway

Abstract

Handwriting significantly contributes to the task of the writer identification and verification of modern and historical documents. This work developed a writer verification system for ancient Hebrew square-script manuscripts, mainly based on the edge-direction feature. Two configurations within the proposed system are carried out, i.e., character-based edge-direction feature extraction and extraction techniques of handwriting shape representation that may drive the system performance. A classification-based verification approach, utilizing Support Vector Machine (SVM) as the classifier, is employed to evaluate the performance of the two configurations. This study has confirmed that the skeleton-based shape representation technique outperforms the edge detection technique used in the predecessor approach. Furthermore, a character-based writer verification system provides the corresponding scholars and experts with an alphabetical investigation to identify the uniqueness of each writer's handwriting.

Introduction

Handwriting has a pivotal role in the establishment of a person's identity. A key aspect of handwriting is motor skills which indirectly reflect the unique quality of an individual's muscle movement, making a person's verification and identification possible [1]. So, unsurprisingly, there is a growing body of literature on computational analysis that recognizes handwriting, especially in the writer identification and verification study [2].

Simultaneously with the existing progressive effort to digitize historical documents, researchers have shown an increased interest in writer identification and verification of historical documents. However, determining writer identification and verifying historical documents is technically challenging. To acquire a legible manuscript, one must have good contrast between ink and background, which could be better for cultural heritage documents due to the trade-off between spatial resolution and sensitivity in low-light imaging [3]. Performing binarization can help to enhance the contrast and legibility of poor-quality or noisy digital images. However, no single binarization techniques are suitable for all kinds of images of historical documents [4] due to, e.g., varying noise levels. As a result, in a writer verification system, we often must deal with noisy images that may eventually lower the system's performance. When we consider scholars with expertise in handwriting analysis, they are not as affected by noise as the computational counterpart would be. While the two different expertise are complementary, we can still improve the handwriting analysis system such that it is not as susceptible to noise.

Figure 1*. A snippet of digital image of ancient Hebrew square script document, publicly available on The Israel Museum website (http://dss.collections.imj.org.il/).*

In the writer identification and verification field, edge-based directional probability distributions, i.e., edge-direction and edge-hinge of graphemes, are considered the robust features for the writer's identification of modern and historical documents, for example, [5], [6], and [7]. These features refer to a type of feature extraction technique used in writer identification and verification, which detect the edges of handwriting strokes and analyze the directional information about an individual's handwriting style. Research has tended to focus on extracting edge-based directional features from the whole page. This approach is not well suited to the scholar or expert working on handwriting analysis of ancient historical documents, which investigate the typical shape of a particular character [8]. With this in mind, we initiated this research to develop the character-based writer verification system for ancient Hebrew square script documents. In addition to the edge-direction features implementation, we applied different techniques to extract the handwriting shape representation and to investigate its influence on the verification performance. Furthermore, it has previously been assumed by scholars in the respected study [9] and proved computationally [6] that two scribes wrote 1QIsa-a scroll (Figure l). This opens up the application of character-based writer verification of 1QIsa-a scroll of which single characters have been previously collected in the previous work [10].

This study has two primary aims: 1. To investigate the impact of different shape representation extraction techniques under the context of writer verification for 1QIsa-a scroll, and 2. To explore a classification-based approach in building a writer verification system. The approach includes robust SVM for multidimensional features classifier. In this study, we used single characters of ancient Hebrew square script instead of graphemes. This way, our research outcomes would not merely be on a writer verification system but also an alphabetical investigation to provide scholars and experts with an in-depth analysis of the uniqueness of a writer's handwriting.

Material and Methods

The flowchart of the proposed writer verification system is shown in Figure 2. The system consists of the dataset of single characters, implementation of shape representation extraction techniques, edge-direction feature extraction, classification-based scribe verification, and verification performance evaluation, which will be explained in this section.

1. Dataset: This study uses data from the author's previous work [10]: single characters correspond to Scribe 1 and Scribe 2 of the 1QIsa-a scroll. The dataset consists of 180 single characters for every 22 ancient Hebrew square-script characters, as depicted in Figure 3, to represent each scribe's handwriting.
2. Shape representation extraction: As described in [4], the proposed edge-direction feature originally extracts features from the edges of handwriting strokes or an individual handwriting shape representation. Instead of implementing only the predecessor technique, i.e., Sobel, we applied additional shape representation extraction techniques, i.e., Canny and Skeleton, as depicted in Figure 4. These varied techniques aim to comprehend better which techniques would extract the best shape representation for ancient Hebrew square scripts.
3. Feature extraction: We used the edge-direction probability distribution of a single character as the features, alternatively expressed as the probability distribution of edge fragments that form a certain angle. We also considered varying the pixel length of edge fragments extracted, i.e., 3-pixels, 4-pixels, and 5-pixels long. This

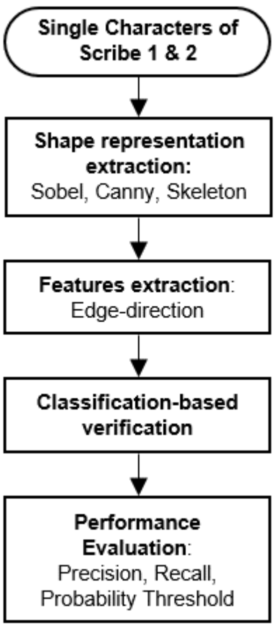

***Figure 2**. Flowchart of the proposed writer verification system.*

***Figure 3**. The 22 ancient Hebrew square-script, modified from [10].*

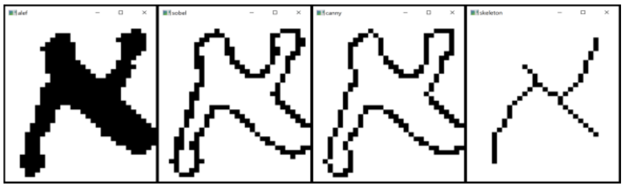

***Figure 4**. Character Alef. From left to right: original, Sobel, Canny, and Skeleton.*

directly varies the number of features used in the classification stage, i.e., 12, 16, and 20 features.

4. Classification: Support vector machine (SVM) is one of the most popular classifiers for high-dimensional and non-linearly separable data. Even so, solely relying on hyperplane-based data separation may not be effective for data with a small portion of the samples and high-level noise [11], which is most likely the case in this study. Therefore, we utilized a probabilistic SVM to solve the challenges and to get more informative output, such as a direct estimate of the probability of belonging to a specific class. To provide a reliable probabilistic SVM, we applied calibration curve representation for probability calibration using a logistic regression approach, which has shown good overall results for SVM model calibration [12]. To create a calibration curve, we first divided a dataset into two equally sized data: train data (40%) and test data (40%). The predicted data obtained from the test data was calibrated by logistic regression referring to the predicted probability from the trained data. The rest 20% of unused data were used for validation data and fed into the trained logistic regression model, which then showed on the graph of the calibration curve. The resulting values are plotted in Figure 5, where the x-axis represents the mean predicted probability, and the y-axis represents the fraction of positive

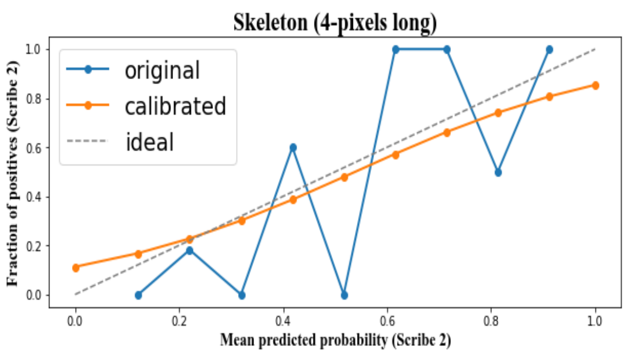

***Figure 5**. Calibration curve of SVM performance.*

output, Scribe 2 in this case. By implementing a calibration curve, we ensured that the calibrated line was close to the ideal line, representing improved reliability of the classifier's prediction in practice.

5. Evaluation: The final stage of the study comprised the performance evaluation with required metrics, i.e., precision, recall, and probability threshold.

Result and Discussion

In this study, we use edge-direction-feature-based classification utilizing a probabilistic SVM classification model for writer verification of historical documents. The dataset consists of 720 images of the letters Alef-s and He-s. Each image has a rectangular shape with a resolution of less than 50x50 pixels with an equally distributed representation of Scribe 1 and Scribe 2 handwritings.

To evaluate the reliability of the probability SVM classification model's prediction, Figure 5 presents how the model has been calibrated to the ideal line of mean predicted probability vs. fractions of positives. Compared to the original line, the calibrated line is much closer to the ideal line and straight, indicating that the calibrated one will estimate the probability of a sample belonging to Scribe 1 or 2 more accurately. On the contrary, with different values of mean predicted probability, the uncalibrated system may consistently perform underestimation when the line is above the ideal line or overestimate large probabilities when the line is under the ideal line.

The first result generated by the proposed verification system is reported in Table 1. The table summarizes the performance metrics scores obtained by implementing the system elements of the Skeleton-based edge-direction feature with 4-pixel-long edge fragments. The table is revealing in several ways. First, the higher the value of the probability threshold (PT), the more the number of true positive (TP) instances of Scribe 1 can be correctly predicted by the system, which means the recall score for Scribe 1's handwriting is getting higher, but the precision score is getting lower. Second, the number of true negatives (TN) instances of Scribe 2 decreases as the PT value increases. With this contrary trend of performance metric scores between Scribe 1 and Scribe 2, it is necessary to choose an optimal PT value that is equally likely to accept a false match (false positive instances) as it is to reject a true match (false negative instances).

Table 1. Scribe 1 and Scribe 2's handwriting verification performance metrics with Skeleton extraction and 4-pixel-long edge fragments.

PT	Scribe	Precision	Recall	Confusion Matrix	
0,1	S1	0.00	0.00	0	36
	S2	0.50	1.00	0	36
0,2	S1	1.00	0.22	8	28
	S2	0.56	1.00	0	36
0,3	S1	0.96	0.69	25	11
	S2	0.76	0.97	1	35
0,4	S1	0.87	0.72	26	10
	S2	0.76	0.89	4	32
0,5	S1	0.86	0.83	30	6
	S2	0.84	0.86	5	31
0,6	S1	0.82	0.89	32	4
	S2	0.88	0.81	7	29
0,7	S1	0.80	0.89	32	4
	S2	0.88	0.78	8	28
0,8	S1	0.69	0.97	35	1
	S2	0.95	0.56	16	20
0,9	S1	0.50	1.00	36	0
	S2	0.00	0.00	36	0
1	S1	0.50	1.00	36	0
	S2	0.00	0.00	36	0

Confusion	
TP	
	TN

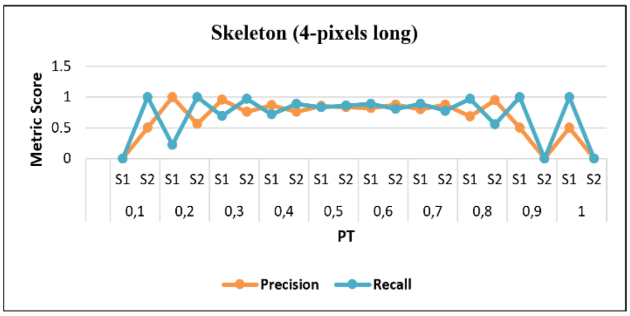

***Figure 6**. The trend of probability threshold vs. precision and recall value.*

In other words, the optimal trade-off is very much expected, as presented by the performance metrics score with a PT value of 0.5. As verified in Table 1, the verification system can obtain average precision and recall scores of 84,75% for both scribes, with only a 2% score difference for precision and 3% for recall between the two scribes. Additionally, it shows a slight score difference between precision and recall for Scribe 1 and Scribe 2, i.e., 3% and 2%, respectively. This optimum trade-off-based decision-making is used in this study to choose the best shape representation extraction technique and the probability threshold value. Furthermore, it is worthwhile noting that a PT value of 0,5 is not the universal optimum PT value of all determined pixel lengths and extraction techniques in the proposed system. And, it is most likely the case to have an optimum PT value between 0,4 and 0,7, as indicated in Figure 6.

In an attempt to investigate whether or not different pixel lengths of edge fragments and different shape representation extraction techniques will influence the precision and recall scores, Table 2 and Table 3 highlight the precision and recall scores with different combinations of edge fragment pixel lengths and shape representation extraction techniques. It should be noted that the color range of blue and orange on the tables helps visualize the precision and recall score levels. The blueish

***Table 2**. Classification-based verification results of the letter Alef-s with three different shape representation extraction techniques.*

LETTER ALEF	SOBEL		CANNY		SKELETON	
3-pixels long	Precision	Recall	Precision	Recall	Precision	Recall
Scribe 1	0.71	0.69	1.00	0.53	0.83	0.81
Scribe 2	0.70	0.72	0.68	1.00	0.81	0.83
4-pixels long	Precision	Recall	Precision	Recall	Precision	Recall
Scribe 1	0.65	0.67	1.00	0.03	0.86	0.83
Scribe 2	0.66	0.64	0.51	1.00	0.84	0.86
5-pixels long	Precision	Recall	Precision	Recall	Precision	Recall
Scribe 1	0.00	0.00	0.65	0.83	0.78	0.81
Scribe 2	0.50	1.00	0.77	0.56	0.80	0.78

***Table 3.** Classification-based verification results of the letter He-s with three different shape representation extraction techniques.*

LETTER HE	SOBEL		CANNY		SKELETON	
3-pixels long	Precision	Recall	Precision	Recall	Precision	Recall
Scribe 1	0.78	0.69	0.68	0.64	0.72	0.78
Scribe 2	0.73	0.81	0.66	0.69	0.76	0.69
4-pixels long	Precision	Recall	Precision	Recall	Precision	Recall
Scribe 1	0.76	0.69	0.80	0.56	0.74	0.69
Scribe 2	0.72	0.78	0.66	0.86	0.71	0.75
5-pixels long	Precision	Recall	Precision	Recall	Precision	Recall
Scribe 1	0.67	0.78	0.52	0.86	0.88	0.83
Scribe 2	0.73	0.61	0.58	0.19	0.84	0.89

color indicates a high score, while the orangish color indicates a low score. Tables 2 and 3 show no significant correlation between precision/ recall score and edge fragment pixel length is identified. The small resolution of each character can justify this apparent lack of correlation; as stated previously, it is not enough to emphasize the difference in probability distribution despite pixel length variation. Interestingly, for different pixel lengths, letters (Alef-s and He-s), and scribes, the Skeleton technique provides a significantly higher and more consistent precision-recall score than Sobel and Canny. This suggests Skeleton might be the best extraction technique for character-based writer verification of ancient Hebrew square script for all corresponding Hebrew characters.

Conclusion and Future Work

This study investigated the impact of different shape representation extraction techniques and explored a classification-based approach in improving a writer verification system's reliability for a historical document. The skeleton technique has emerged as a reliable element to enhance the performance of edge-based directional features for writer verification of an ancient Hebrew square-script manuscript. This study also shows that adjusting the probability threshold significantly shifts the proposed system into a reliable writer verification system, allowing the system to verify whether a given character belongs to the identified writer.

A further study could assess the performance of writer verification using the remaining letters of ancient Hebrew square script and investigate whether the characters could give different performance scores. Such investigation could give scholars a new comprehension of which characters have the best distinctive stroke, allowing better verification performance. Also, implementing other performance evaluation metrics used for writer verification, i.e., FAR (False Acceptance Rate), FRR (False Rejection Rate), and ERR (Equal Error Rate), would make improvements in the system's accuracy and reliability, as well as user experience and satisfaction, especially for scholar and expert in the field of historical document analysis.

References

[1] R. V. Yampolskiy and V. Govindaraju, "Behavioural biometrics: a survey and classification," *Int. J. Biom.*, vol. 1, no. 1, p. 81, 2008, doi: 10.1504/IJBM.2008.018665.

[2] A. Rehman, S. Naz, and M. I. Razzak, "Writer identification using machine learning approaches: a comprehensive review," *Multimed. Tools Appl.*, vol. 78, no. 8, pp. 10889–10931, Apr. 2019, doi: 10.1007/s11042-018-6577-1.

[3] T. R. Peery and D. W. Messinger, "Spatial resolution as a trade-space for low-light imaging of sensitive cultural heritage documents," *J. Cult. Herit.*, vol. 45, pp. 81–90, Sep. 2020, doi: 10.1016/j.culher.2020.02.014.

[4] R. D. Lins, R. Bernardino, R. da S. Barboza, and R. C. De Oliveira, "Using Paper Texture for Choosing a Suitable Algorithm for Scanned Document Image Binarization," *J. Imaging*, vol. 8, no. 10, p. 272, Oct. 2022, doi: 10.3390/jimaging8100272.

[5] M. Bulacu, L. Schomaker, and L. Vuurpijl, "Writer identification using edge-based directional features," in *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.*, Edinburgh, UK: IEEE Comput. Soc, 2003, pp. 937–941. doi: 10.1109/ICDAR.2003.1227797.

[6] M. Popović, M. A. Dhali, and L. Schomaker, "Artificial intelligence based writer identification generates new evidence for the unknown scribes of the Dead Sea Scrolls exemplified by the Great Isaiah Scroll (1QIsaa)," *PLOS ONE*, vol. 16, no. 4, p. e0249769, Apr. 2021, doi: 10.1371/journal.pone.0249769.

[7] S. Al-Ma'adeed, E. Mohammed, D. Al Kassis, and F. Al-Muslih, "Writer identification using edge-based directional probability distribution features for arabic words," in *2008 IEEE/ACS International Conference on Computer Systems and Applications*, Doha, Qatar: IEEE, Mar. 2008, pp. 582–590. doi: 10.1109/AICCSA.2008.4493590.

[8] A. Shaus and E. Turkel, "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*, Kyoto Japan: ACM, Nov. 2017, pp. 13–18. doi: 10.1145/3151509.3151511.

[9] E. Tov, *Scribal Practices and Approaches Reflected in the Texts Found in the Judean Desert*. 2004.

[10] T. L. Tobing, S. Y. Yayilgan, S. George, and T. Elgvin, "Isolated Handwritten Character Recognition of Ancient Hebrew Manuscripts," *Arch. Conf.*, vol. 19, no. 1, pp. 35–39, Jun. 2022, doi: 10.2352/issn.2168-3204.2022.19.1.8.

[11] H.-X. Li, J.-L. Yang, G. Zhang, and B. Fan, "Probabilistic support vector machines for classification of noise affected data," *Inf. Sci.*, vol. 221, pp. 60–71, Feb. 2013, doi: 10.1016/j.ins.2012.09.041.

[12] P. Xu, F. Davoine, H. Zha, and T. Denœux, "Evidential calibration of binary SVM classifiers," *Int. J. Approx. Reason.*, vol. 72, pp. 55–70, May 2016, doi: 10.1016/j.ijar.2015.05.002.

Author Biography

Tabita L. Tobing received her BSc in Physics from Universitas Indonesia (2016) and MSc in Electrical Engineering from Institut Teknologi Bandung (2021). She is currently a PhD student at NTNU under The Lying Pen of Scribes project; her research field is computational analysis for ancient Hebrew manuscripts using machine learning.

Pavel Škrabánek received a Ph.D. in Technical Cybernetics in 2010 from the University of Pardubice (UPce), Czech Republic. He joined the Faculty of Electrical Engineering and Informatics, UPce as an Assistant in 2007 and 2011, he became an Assistant Professor there. Since 2018, he has worked as an Assistant Professor at the Institute of Automation and Computer Science, Brno University of Technology, Czech Republic. His research includes, among other,s computer vision and machine learning.

Sule Y. Yayilgan is a Professor at the Department of Information Security and Communication Technology (IIK), NTNU. She belongs to the Center for Cyber Information Security (ccis.no) and is leading the MR PET: Multidisciplinary Research Group on Privacy and data protEcTion. Her research interests include Artificial Intelligence and Cybersecurity in various application fields.

Sony George is currently Associate Professor at The Colourlab, NTNU, since 2017. Before joining NTNU, he worked as a researcher at Gjøvik University College, Norway. Sony obtained a PhD in Photonics from the Cochin University of Science and Technology, India, in 2012. His research interests are color imaging, spectral image processing, computer vision, etc.

Torleif Elgvin is a Professor Emeritus of theology, religion, and philosophy at NLA University College. He holds a theological degree from MF Norwegian School of Theology, Religion, and Society. He received his Doctorate on the Dead Sea Scrolls at the Hebrew University of Jerusalem in 1998. He was head of the Nordic network for Dead Sea Researchers from 2003 to 2007, and since 1992 has participated in the publishing team for the Dead Sea Scrolls. His research interests are Dead Sea Scrolls and biblical studies.