Writer Retrieval on the Vernier Election Ballots

Marco Peer[1], Andreas Fischer[1]
Thomas Gorges[2], Mathias Seuret[2], Vincent Christlein[2]

[1] AIBEX Group, University of Fribourg, Switzerland
[2] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

December 12, 2025

Résumé. Ce rapport présente une analyse de recherche de scripteur fondée sur l'écriture manuscrite, réalisée sur 1 414 bulletins de vote issus de l'élection de Vernier du 30 novembre 2025. L'objectif est d'identifier des regroupements de bulletins susceptibles d'avoir été rédigés par une même personne, en détectant des similarités stylistiques dans le contenu manuscrit. Nous utilisons une méthodologie de reconnaissance de formes à l'état de l'art, qui commence par isoler les contenus manuscrits à partir des images numérisées, puis analyse les allographes (petites composantes de l'écriture) à l'aide de réseaux neuronaux afin de détecter des similarités indépendantes du texte. Nous présentons deux principaux résultats. Premièrement, notre analyse ne met en évidence aucun grand regroupement de bulletins rédigés dans un style similaire. Deuxièmement, nous identifions plusieurs cas de petits regroupements, généralement composés de deux bulletins, rédigés dans un style similaire. Au total, 189 des 1 414 bulletins sont inclus dans l'un de ces petits regroupements. Nous fournissons dans ce rapport une liste complète des petits regroupements afin de faciliter un examen approfondi par des experts judiciaires humains.

Abstract. This report presents a handwriting-based writer retrieval analysis conducted on 1,414 ballots from the Vernier election of November 30, 2025. The objective is to identify clusters (groups) of ballots possibly written by the same individual by detecting stylistic similarities in the handwritten content. We employ a state-of-the-art pattern recognition methodology that first isolates handwritten contents from the scanned images and then analyzes allographs (small parts of the handwriting) with neural networks to detect text-independent similarities. We report two main findings: First, our analysis does not reveal any large cluster of ballots that are written with a similar style. Secondly, we identify several cases of small clusters of typically 2 ballots that are written with a similar style. Overall, 189 out of 1,414 ballots are included in one of the small clusters. We provide a complete list of the small clusters in our report to facilitate further inspection by human forensic experts.

1 Introduction

This report presents the methodology and results of our approach to writer retrieval on ballots from the Vernier election held on November 30, 2025. The primary objective is to detect and group similar handwriting styles — that is, to *cluster* ballots likely completed by the same individual. This serves as a first step for subsequent writer-level analyses performed by human experts, which are beyond the scope of this report. Our methods are based on state-of-the-art approaches for writer identification and retrieval, previously applied in forensics [5] and historical document analysis [2, 9, 10]. Automated writer retrieval enables efficient and scalable search of documents with stylistically similar handwriting. We employ a two-stage approach: first, a rule-based method using blank ballot templates with SIFT and RANSAC to segment handwriting; second, a deep neural network that encodes handwritten content into a discriminative feature space to capture subtle writing characteristics and identify similarities and differences between writing styles.

The following sections include a short description of the dataset provided, the methodology used (consisting of handwriting segmentation and writer retrieval), and evaluation procedures, alongside a discussion of the findings and their practical implications.

2 Methodology

The following section includes the methodology of the writer retrieval, with additional preprocessing steps to segment the handwriting from the scanned ballots. The methodology for the handwriting segmentation is developed by the Pattern Recognition Lab at FAU Erlangen-Nürnberg, while the writer retrieval is performed by the AIBEX group at the University of Fribourg. However, the ballots are only processed at the University of Fribourg in Switzerland.

2.1 Data

The dataset includes 1414 RGB images of ballots scanned at a resolution of 300 dpi. The ballot is identified via a ID representing the *local* (4301, 4302, 4303, 4304) and a consecutive number per local. Each ballot contains handwritten text that is first segmented in a preprocessing step to avoid interference during the writer retrieval approach. The amount of text per sample varies, with some empty ballots, some containing only one name and some even a fully filled out ballot (e.g., 20 different names).

2.2 Handwriting Segmentation

To provide suitable input data for the writer identification task, handwriting must be isolated from the printed content in each scanned ballot. For each scan, a processing pipeline comprising four stages is applied:

Template Selection A set of blank templates is defined. For each template, the printed foreground is detected using Sauvola thresholding [13], and the resulting mask is dilated to gain robustness to minor misalignments. A masked similarity score between the scanned ballot and the template is computed. Mean squared error (MSE) of the intensities is used, where a lower MSE indicates higher similarity, and the template with the lowest MSE is selected.

Scan Alignment The identified template and the scan are converted to grayscale. Keypoints in both images are detected with SIFT (Scale-Invariant Feature Transform) [6]. The corresponding descriptors are matched using a brute-force matcher with Lowe's ratio test. A projective homography is then estimated with USAC-MAGSAC [12, 1, 3] and the scan is warped to align with the template.

Handwriting Isolation A Gaussian blur is applied to the aligned scan to reduce noise. The brightness of the scan is normalized to the template by matching the mean intensity over template background pixels. Printed content on the template is separated from the background using Sauvola thresholding [13], and foreground masks with Sauvola thresholding are computed for both the normalized scan and the template. An exclusion mask is obtained by dilating the template foreground to protect the printed content. Handwriting is segmented as scan foreground pixels that lie outside this mask.

Handwritten Text Refinement Connected-component analysis and morphological operations are applied to remove small noise and to connect fragmented strokes. The final output is an image that contains the handwriting on a white background.

2.3 Writer Identification

Preprocessing First, the segmented ballots are binarized via thresholding since the background (e.g., the ballot template) is already removed. This ensures that the pen has no influence on the handwriting. In case there are problems with the segmentation (e.g., the wrong template is matched or the segmented image contains artefacts of the template), the images are manually cropped to only contain the handwriting of the ballot.

Handwriting Similarity The methodology for estimating handwriting similarity is based on the approach proposed in [9]. The estimation of the similarity is based on extracting 32×32 patches at the contour of the handwriting, which allows to observe the handwriting in a text-independent way since it only sees *allographs* (stroke-like buildings which form the actual characters). Furthermore, it utilizes a deep convolutional neural network to extract feature vectors from handwriting patches. The network is trained using a triplet loss function, which encourages patches from the same writer to have similar feature representations while pushing apart patches from different writers. During training, triplets of patches are formed, consisting of an anchor patch, a positive patch (from the same writer), and a negative patch (from a different writer). The network learns to minimize the distance between the anchor and positive patches while maximizing the distance to the negative patch in the feature space. The network is trained in a supervised manner on two publicly available datasets: CVL [4] (English, German), and Firemaker [14] (Dutch). While the CVL dataset is a popular benchmark dataset for writer identification, the Firemaker dataset is chosen because each writer contributed a page free-forged handwriting as well as one page of block capital handwriting. To further improve the performance, we apply kRNN-reranking as described in [9] with $k = 3$. We train five different neural networks on different train/validation splits to ensure robust page descriptors. They are concatenated and then jointly whitened via PCA to a dimension of 512.

Clustering To group ballots according to handwriting similarity, we examine the sets of documents present within the top-10 retrieval results of the first initial suspicions provided and determine a similarity threshold of 0.8 for the clustering. Afterwards, for each document, the cluster candidates are calculated by thresholding the similarity matrix. Subsequently, we conduct a manual postprocessing step and filter out obvious false positives, particularly regarding clusters of size two (e.g., due to very short text, identical text or empty ballots).

Additionally, we experiment with traditional clustering techniques, including HDBSCAN [8] and hierarchical clustering based on cosine distance, while also incorporating global handwriting features. However, the results obtained from these methods are not reported in this work. Their performance is highly sensitive to hyperparameter choices, and the resulting cluster structures lacked stability and statistical reliability. As a consequence, these methods do not produce clusters that are sufficiently consistent or meaningful for this application.

3 Qualitative Analysis

3.1 Retrieval Results

We provide the retrieval results, including the corresponding images, both as a separate download and in the appendix. Retrieval is performed using a leave-one-out strategy, following standard practice: each ballot is used once as a query (marked as “Q”), and all remaining documents are ranked according to the cosine similarity of their feature descriptors. The retrieval output includes a similarity score for each match (ranging from 0 to 1), although this score is not directly interpretable. For each query, we report the five nearest neighbours in the dataset.

Initial Findings We manually inspect the retrieval results of the first 20 documents prior to conducting clustering analyses. In several cases, the most similar documents are visually highly consistent with the query and exhibit similarity scores above 0.8. Examples include:

– `4301-001`, `4301-002`

- 4301-005, 4301-006, 4301-346
- 4301-009, 4301-010, 4301-307
- 4301-013 – 4301-016
- 4301-017, 4301-019

3.2 Clusters of Similar Handwritings

The clustering and pattern analysis of the ranked retrieval lists reveal several groups of ballots with similar handwriting. Table 1 summarizes the number of clusters by cluster size. A full list of clusters, including ballot identifiers, is provided in the appendix. Additionally, a two-dimensional UMAP [7] visualization of the page descriptors is supplied separately, with cluster assignments highlighted. In total, we identify 79 clusters, the majority of which (≈ 70.8%) consist of only

Table 1. Number of clusters per cluster size found in the ballots.

Cluster Size	Number of Clusters
2	56
3	18
4	3
5	1
6	1
Sum	79

two samples. Approximately 13.3% (189) of the 1414 samples are assigned to at least one cluster, indicating that these ballots share a sufficiently similar handwriting style to be linked by our method. No ballot is assigned to more than one cluster. Manual inspection further confirms that several of the discovered clusters align with items already flagged in the initial list of suspicions. Overall, our method does not reveal large clusters of similar handwriting, e.g., a large set of many documents that could plausibly originate from a single writer. Instead, the findings suggest that while multiple handwriting styles recur across the dataset, each appears only in a small number of ballots (no more than six), relative to the size of the collection.

4 Limitations

A key limitation of our approach lies in the limited amount of handwriting available on certain ballots. As demonstrated in a previous study [11], current methods struggle when only small amounts of text are present - such as a single line or a short phrase - since the model has fewer stylistic cues to learn from.

A second limitation concerns potential bias when comparing samples that contain identical text, such as candidate names or predefined ballot entries, even if the method used in this report is text-independent. When two samples contain the same written content, their handwriting may appear more similar than it truly is, simply because writing the same sequence of characters contains similar allographs and stroke patterns. This effect can artificially inflate similarity scores and may obscure genuine differences in writing style. Moreover, the form of the handwriting - whether block script, cursive, or a mixed style - can also be influenced by the specific text being written, further complicating the comparison.

Finally, because our analysis is performed in a retrieval setting aimed at distinguishing between unknown writers, the resulting clusters and similarity scores are not interpretable as forensic evidence. The retrieval process is designed to rank samples based on relative similarity, not to provide probabilistic or legally meaningful conclusions about authorship. Additionally, writers who contribute only a single document introduce noise into the ranking: their samples appear in the retrieval results even though they are not relevant for identifying clusters of multiple documents from the same writer. These "singletons" can distract from relevant similarities and reduce the overall clarity of the rankings.

5 Conclusion

In summary, our approach is able to 1) successfully segment handwriting in the ballots presented and 2) identify small clusters of ballots with similar handwriting, suggesting that only six or fewer documents share a similar handwriting style within the collection. However, the limited amount of handwriting on many ballots and the sensitivity of similarity measures to repeated text constrain the robustness of the clustering results.

References

1. Barath, D., Matas, J., Noskova, J.: Magsac: Marginalizing sample consensus. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)
2. Christlein, V., Gropp, M., Fiel, S., Maier, A.K.: Unsupervised feature learning for writer identification and writer retrieval. In: 14th IAPR International Conference on Document Analysis and Recognition, ICDAR 2017, Kyoto, Japan, November 9-15, 2017. pp. 991–997. IEEE (2017). https://doi.org/10.1109/ICDAR.2017.165, https://doi.org/10.1109/ICDAR.2017.165
3. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (Jun 1981). https://doi.org/10.1145/358669.358692, https://doi.org/10.1145/358669.358692
4. Kleber, F., Fiel, S., Diem, M., Sablatnig, R.: Cvl-database: An off-line database for writer retrieval, writer identification and word spotting. In: 12th International Conference on Document Analysis and Recognition, ICDAR 2013, Washington, DC, USA, August 25-28, 2013. pp. 560–564. IEEE Computer Society (2013). https://doi.org/10.1109/ICDAR.2013.117, https://doi.org/10.1109/ICDAR.2013.117

5. Koepf, M., Kleber, F., Sablatnig, R.: Writer identification and writer retrieval using vision transformer for forensic documents. In: Uchida, S., Smith, E.H.B., Eglin, V. (eds.) Document Analysis Systems - 15th IAPR International Workshop, DAS 2022, La Rochelle, France, May 22-25, 2022, Proceedings. Lecture Notes in Computer Science, vol. 13237, pp. 352–366. Springer (2022), https://doi.org/10.1007/978-3-031-06555-2_24
6. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the seventh IEEE international conference on computer vision. vol. 2, pp. 1150–1157. Ieee (1999)
7. McInnes, L., Healy, J.: UMAP: uniform manifold approximation and projection for dimension reduction. CoRR **abs/1802.03426** (2018), http://arxiv.org/abs/1802.03426
8. McInnes, L., Healy, J., Astels, S.: hdbscan: Hierarchical density based clustering. J. Open Source Softw. **2**(11), 205 (2017). https://doi.org/10.21105/JOSS.00205, https://doi.org/10.21105/joss.00205
9. Peer, M., Kleber, F., Sablatnig, R.: Towards writer retrieval for historical datasets. In: Fink, G.A., Jain, R., Kise, K., Zanibbi, R. (eds.) Document Analysis and Recognition - ICDAR 2023 - 17th International Conference, San José, CA, USA, August 21-26, 2023, Proceedings, Part I. Lecture Notes in Computer Science, vol. 14187, pp. 411–427. Springer (2023), https://doi.org/10.1007/978-3-031-41676-7_24
10. Peer, M., Kleber, F., Sablatnig, R.: SAGHOG: self-supervised autoencoder for generating HOG features for writer retrieval. In: Smith, E.H.B., Liwicki, M., Peng, L. (eds.) Document Analysis and Recognition - ICDAR 2024 - 18th International Conference, Athens, Greece, August 30 - September 4, 2024, Proceedings, Part II. Lecture Notes in Computer Science, vol. 14805, pp. 121–138. Springer (2024). https://doi.org/10.1007/978-3-031-70536-6_8, https://doi.org/10.1007/978-3-031-70536-6_8
11. Peer, M., Sablatnig, R., Kleber, F.: Towards the influence of text quantity on writer retrieval. In: Yin, X., Karatzas, D., Lopresti, D. (eds.) Document Analysis and Recognition - ICDAR 2025 - 19th International Conference, Wuhan, China, September 16-21, 2025, Proceedings, Part II. Lecture Notes in Computer Science, vol. 16024, pp. 129–145. Springer (2025), https://doi.org/10.1007/978-3-032-04617-8_8
12. Raguram, R., Chum, O., Pollefeys, M., Matas, J., Frahm, J.M.: Usac: A universal framework for random sample consensus. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(8), 2022–2038 (2013). https://doi.org/10.1109/TPAMI.2012.257
13. Sauvola, J., Pietikäinen, M.: Adaptive document image binarization. Pattern Recognition **33**(2), 225–236 (2000). https://doi.org/https://doi.org/10.1016/S0031-3203(99)00055-2, https://www.sciencedirect.com/science/article/pii/S0031320399000552
14. Schomaker, L., Vuurpijl, L.: Firemaker image collection for benchmarking forensic writer identification using image-based pattern recognition (2000). https://doi.org/10.5281/ZENODO.1194612, https://zenodo.org/record/1194612