Computational Qumranic Paleography

Berat Kurar-Barakat and Nachum Dershowitz

Tel Aviv University {berat,nachumd}@tauex.tau.ac.il
15.07.2024

Abstract. The discovery of the Dead Sea Scrolls over 60 years ago is widely regarded as one of the greatest archaeological breakthroughs in modern history. Modern study of the scrolls presents ongoing computational challenges, including determining the provenance of each manuscript fragment, clustering fragments based on their degree of similarity, and pairing fragments that may have originated from the same original manuscript–through analysis of paleographical and codicological features. The importance of this research lies in the need for computational methods to support and enhance the traditional practice of paleography, since acquiring paleographic skills typically requires intensive training and access to rare artifacts. Previous studies have used connected components or image patches to extract paleographical features. However, features from connected components may lack the comprehensive information of whole letters, and image patches often include background pixels that can lead to erroneous interpretations. To address this, we segment the ink regions within the fragment images and enhance Kraken's letter detection capabilities by incorporating these segmented ink regions. This allows computational algorithms to focus on individual letter shapes and pairs of letter shapes, thereby providing more precise paleographic analysis.

The image of a single fragment consists of approximately 12,000,000 pixels (3000x4000). Annotated images are essential in machine learning methods, but the annotation process for high resolution image segmentation can be particularly challenging, requiring the delineation of each individual pixel. In this research project, we utilize multi-spectral thresholding and energy minimization to automate the segmentation process. By adopting this approach, we accelerate the analysis of millions of pixels. This phase to enable highly accurate segmentation, capable of distinguishing minute elements like small pieces of rice paper, parchment, or tiny ink droplets, and even discerning holes and background from the ink.

We use energy minimization to incorporate ink regions into existing text recognition algorithms like kraken, enhancing the accuracy of detected individual letters within the fragments. Finally, we will implement letter-level and bi-letter-level paleography, leveraging accurately detected and recognized letters. This approach offers a deeper understanding of the evolution of paleographic styles, as the prototypes constructed from letters provide more descriptive insights compared to those derived from connected components or image patches.

Keywords: Dead Sea Scrolls · Paleography · Ink segmentation · Parchment segmentation · Letter detection.