

# Preprocessing Algorithm for Deciphering Historical Inscriptions Using String Metric

Loránd Lehel Tóth<sup>1</sup>, Raymond Eliza Ivan Pardede<sup>1</sup>, György András Jeney<sup>1</sup>, Ferenc Kovács<sup>2</sup>,  
Gábor Hosszú<sup>1,\*</sup>

<sup>1</sup>Department of Electron Devices, Budapest University of Technology and Economics, Budapest, Hungary.

<sup>2</sup>Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest, Hungary.

Received 25 January 2016; received in revised form 05 April 2016; accepted 17 June 2016

## Abstract

The article presents the improvements in the preprocessing part of the deciphering method (shortly preprocessing algorithm) for historical inscriptions of unknown origin. Glyphs used in historical inscriptions changed through time; therefore, various versions of the same script may contain different glyphs for each grapheme. The purpose of the preprocessing algorithm is reducing the running time of the deciphering process by filtering out the less probable interpretations of the examined inscription. However, the first version of the preprocessing algorithm leads incorrect outcome or no result in the output in certain cases. Therefore, its improved version was developed to find the most similar words in the dictionary by relaying the search conditions more accurately, but still computationally effectively. Moreover, a sophisticated similarity metric used to determine the possible meaning of the unknown inscription is introduced. The results of the evaluations are also detailed.

**Keywords:** computational paleography, rovash paleography, mathematical optimization, deciphering algorithm

## References

- [1] G. Hosszú, "The Rovas: A special script family of the central and eastern European languages," *Acta Philologica (Wydział Neofilologii Uniwersytet Warszawski, Warszawa)*, vol. 44, pp. 91-102, 2013.
- [2] L. Eikvil, *Optical Character Recognition*, Oslo: Norsk Regnesentral, 1993. (online) Access date June 21, 2015, <http://bkreaders.ru/books/OCR.pdf>.
- [3] G. Hosszú, "Mathematical statistical examinations on script relics," in *Data Mining and Analysis in the Engineering Field*, 1st ed., V. Bhatnagar, Ed. Hershey, New York: Information Science Reference, 2014, pp. 142-158.
- [4] G. Hosszú, "A novel computerized paleographical method for determining the evolution of graphemes," in *Encyclopedia of Information Science and Technology*, 3rd ed., M. Khosrow-Pour, Ed. Hershey, New York: Information Science Reference, 2015, pp. 2017-2031.
- [5] T. Hassner, M. Rehbein, P. A. Stokes, and L. Wolf, "Computation and palaeography: potentials and limits (Dagstuhl Perspectives Workshop 12382)," *Dagstuhl Reports*, vol. 2, no. 9, pp. 184-199, 2012.
- [6] B. Gottfried, M. Wegner, and M. Lawo, "Towards the interactive transcription of handwritings: anytime anywhere document analysis," *Int. J. on Document Analysis and Recognition (IJ DAR)*, vol. 18, no. 1, pp. 31-45, March 2015.
- [7] M. Panagopoulos, P. Rousopoulos, D. Arabajis, M. Exarhos, and C. Papaodysseus, "Methods and algorithms for the automatic identification of writer of ancient documents," *Proc. 1st Conf. on Computer Applications and Quantitative Methods in Archaeology Greek Chapter (CAA-GR), Rethymno, Crete, Greece, 2012*, pp. 153-158, March 2014.
- [8] E. Kavallieratou, K. Sgarbas, N. Fakotakis, and G. Kokkinakis, "Handwritten word recognition based on structural characteristics and lexical support," *Proc. Int. Conf. on Document Analysis and Recognition, IEEE Press, Aug. 2003*, vol. 1, pp. 562-566.

\* Corresponding author. E-mail address: [hosszu@eet.bme.hu](mailto:hosszu@eet.bme.hu)

Tel.: +36-1-4632724; Fax: +36-1-4632973

- [9] S. Singh, "Shape detection using gradient features for handwritten character recognition," Proc. 13th Int. Conf. on Pattern Recognition, IEEE Press, August 1996, vol. 3, pp. 145-149.
- [10] V. Märgner, H. El Abed, and M. Pechwitz, "Offline handwritten Arabic word recognition using HMM – a character based approach without explicit segmentation," Actes du 9ème Colloque International Francophone sur l'Écrit et le Document, pp. 259-264, Sept. 2006.
- [11] A. Khémiri, A. Kacem, and A. Belaïd, "Towards Arabic handwritten word recognition via probabilistic graphical models," Proc. 14<sup>th</sup> Int. Conf. on Frontiers in Handwriting Recognition, Heraklion, IEEE Press, Sept. 2014, pp. 678-683.
- [12] F. Kurniawan, A. R. Khan, and D. Mohamad, "Contour vs. non-contour based word segmentation from handwritten text lines: an experimental analysis," International Journal of Digital Content Technology and its Applications vol. 3, no. 2, pp. 127-131, Jan. 2009.
- [13] S. Gomathi Rohini, R. S. Umadevi, and S. Mohanavel, "Statistical approach for segmenting unconstrained handwritten text lines." IJCA Proc. Amrita Int. Conf. of Women in Computing (AICWIC'13). IJCA Journal, Jan. 2013, pp. AICWIC(1):21-24.
- [14] C. Chatelain, L. Heutte and T. Paquet, "A syntax-directed method for numerical field extraction using classifier combination," Proc. Ninth International Workshop on Frontiers in Handwriting Recognition (IWFHR-9), 2004, Tokyo, Japan, 26-29 Oct. 2004, pp. 93-98.
- [15] L. Heutte, A. Nosary, T. Paquet, "A multiple agent architecture for handwritten text recognition," Pattern Recognition, vol. 37, no. 4, pp. 665-674, 2004.
- [16] L. L. Tóth, R. Pardede, and G. Hosszú, "Novel algorithmic approach to deciphering rovas inscriptions," in Encyclopedia of Information Science and Technology, 3rd ed., M. Khosrow-Pour, Ed. Hershey: Information Science Reference, 2015, pp. 7222-7233.
- [17] L. L. Tóth, R. E. I. Pardede, G. A. Jeney, F. Kovács, and G. Hosszú, "Application of the cluster analysis in computational paleography," in Handbook of Research on Advanced Computational Techniques for Simulation-Based Engineering, 1st ed., P. Samui, Ed. Hershey: Engineering Science Reference, 2016, pp. 525-543.
- [18] N. A. Khan, "A shape analysis model with application to character and word recognition," Ph.D. Dissertation, Technische Universiteit Eindhoven, Eindhoven, 2000.
- [19] R. Rashli, Z. Zulkoffli, E. A. Bakar, and M. S. Soaid, "A study of 3D CAD model and feature analysis for casting object," International Journal of Engineering and Technology Innovation, vol. 2, no. 2, pp. 138-149, 2012.
- [20] G. Hosszú, Heritage of Scribes. The relation of rovas scripts to Eurasian writing systems, 2nd ed. Budapest: Rovas Foundation, 2012.
- [21] S. Theodoridis and K. Koutroumbas, Pattern recognition, 2nd ed. San Diego: Elsevier, 2003.
- [22] I. Oliver. Programming classics - implementing the world's best algorithms. Prentice Hall, 1994.