



# Modern vs Diplomatic Transcripts for Historical Handwritten Text Recognition

Verónica Romero<sup>1</sup>(✉), Alejandro H. Toselli<sup>1</sup>, Enrique Vidal<sup>1</sup>,  
Joan Andreu Sánchez<sup>1</sup>, Carlos Alonso<sup>2</sup>, and Lourdes Marqués<sup>2</sup>

<sup>1</sup> PRHLT, Universitat Politècnica de València,  
Camí de Vera s/n, 46022 València, Spain

{vromero, ahector, evidal, jandreu}@prhlt.upv.es

<sup>2</sup> Instituto Andaluz de Patrimonio Histórico, Centro de Arqueología Subacuática,  
Consejería de Cultura, Junta de Andalucía,  
Avda. Duque de Nájera, 3, 11002 Cádiz, Spain

{carlos.alonso.v, lourdes.marquez.carmona}@juntadeandalucia.es

<http://www.prhlt.upv.es>

**Abstract.** The transcription of handwritten documents is useful to make their contents accessible to the general public. However, so far automatic transcription of historical documents has mostly focused on producing diplomatic transcripts, even if such transcripts are often only understandable by experts. Main difficulties come from the heavy use of extremely abridged and tangled abbreviations and archaic or outdated word forms. Here we study different approaches to train optical models which allow to recognize historic document images containing archaic and abbreviated handwritten text and produce modernized transcripts with expanded abbreviations. Experiments comparing the performance of the different approaches proposed are carried out on a document collection related with Spanish naval commerce during the XV–XIX centuries, which includes extremely difficult handwritten text images.

**Keywords:** Handwritten text recognition · Historical documents · Modern transcripts

## 1 Introduction

In the last years, there is an increasing interest to digitally preserve and provide access to handwritten historical documents residing in libraries, museums, and archives. Such documents are a unique public asset, forming the collective and evolving memory of our societies. Many of these historical documents have an

---

Work partially supported by the BBVA Foundation through the 2017–2018 Digital Humanities research grant “Carabela”, by Miniterio de Ciencia/AEI/FEDER/EU through the MIRANDA-DocTIUM project (RTI2018-095645-B-C22), and by EU JPICH project “HOME – History Of Medieval Europe” (Spanish PEICTI Ref. PCI2018-093122).

© Springer Nature Switzerland AG 2019

M. Cristani et al. (Eds.): ICIAP 2019 Workshops, LNCS 11808, pp. 103–114, 2019.

[https://doi.org/10.1007/978-3-030-30754-7\\_11](https://doi.org/10.1007/978-3-030-30754-7_11)