A Web-Based Demo to Interactive Multimodal Transcription of Historic Text Images

Verónica Romero, Luis A. Leiva, Vicent Alabau,
Alejandro H. Toselli, and Enrique Vidal

ITI- Instituto Tecnológico de Informática,
Universidad Politécnica de Valencia, Spain
{vromero,luileito,valabau,ahector,evidal}@iti.upv.es

Abstract. Paleography experts spend many hours transcribing historic documents, and state-of-the-art handwritten text recognition systems are not suitable for performing this task automatically. In this paper we present the modifications on a previously developed interactive framework for transcription of handwritten text. This system, rather than full automation, aimed at assisting the user with the recognition-transcription process.

Key words: Handwritten recognition, Interactive framework, Web, HCI.

1 Introduction

Nowadays, there is an increasing number of on-line digital libraries publishing a large quantity of digitized legacy documents. These documents need to be transcribed in order to provide historians and other researchers new ways of indexing, consulting and querying them. Up-to-date Handwritten Text Recognition systems (HTR) cannot replace the experts on this task, since there are no perfect accuracy solutions. Therefore, once the full recognition process of one document has finished, the human expert revision is required to produce a quality transcription. Such post-edition solution is rather inefficient and uncomfortable for the user.

As an alternative to post-editing, a multimodal interactive approach is proposed in this work. The user feedback allows to improve the system accuracy [1,2], while multimodality increases system ergonomics and user acceptability.

2 Review of the MM-CATTI Framework

In the MM-CATTI framework, the user is involved in the transcription process since she is responsible of validating and/or correcting the HTR output [3]. The protocol that rules this process, is formulated in the following steps:

- The HTR system proposes a full transcription a handwritten text line image.
- The user validates the longest prefix and amends the first error in the suffix.
- A new extended prefix is produced based on the previous validated prefix.
- Using this new prefix, the system suggests a suitable suffix.
- These previous steps are iterated until a perfect transcription is obtained.

M. Agosti et al. (Eds.): ECDL 2009, LNCS 5714, pp. 459–460, 2009.

3 Demo Description

The demo presented in this paper is a web-based demo. First, a series of available documents is shown. After selecting the document, the user must choose a page of the document to work with. Finally, the user must transcribe the handwritten text images line by line, making corrections with pen strokes and also using the keyboard. If pen strokes were available, the MM-CATTI server uses an on-line HTR feedback subsystem to decode them. Then, taking into account the decoded word and the off-line models, the MM-CATTI server responds with a suitable continuation to the prefix validated by the user.

The web-based demo proposed in this paper differs from the demo presented in [4] mainly in the client-server communication. In the previous one, the communication was made asynchronously via Ajax and PHP. On the contrary, the new approach communicates much faster through sockets. Furthermore, it allows a more flexible architecture featuring multiple user connections and server load balancing.

4 Evaluation Results

Several experiments were carried out on a corpus corresponding to a historic handwriting document identified as "Cristo Salvador" [2,5]. This document was kindly provided by the Biblioteca Valenciana Digital (BIVALDI). The estimated human effort to produce error-free transcription using MM-CATTI is reduced by a 15% on average, with respect to the classical HTR system. Therefore, from every 100 words misrecognized by a conventional HTR system, a human post-editor will have to correct all the 100 erroneous words, while a MM-CATTI user would correct only 85 - the other 15 would be automatically corrected by the system.

Acknowledgment

This work has been supported by the EC (FEDER), the Spanish MEC under grant TIN2006-15694-C02-01 and the research programme Consolider Ingenio 2010 MIPRCV (CSD2007-00018) and by the UPV (FPI fellowship 2006-04).

References

1. Toselli, A.H., et al.: Computer assisted transcription of handwritten text. In: Proc. of ICDAR 2007, pp. 944–948. IEEE Computer Society, Los Alamitos (2007)
2. Romero, V., Toselli, A.H., Rodríguez, L., Vidal, E.: Computer assisted transcription for ancient text images. In: Kamel, M.S., Campilho, A. (eds.) ICIAR 2007. LNCS, vol. 4633, pp. 1182–1193. Springer, Heidelberg (2007)
3. Toselli, A.H., et al.: Computer assisted transcription of text images and multimodal interaction. In: Popescu-Belis, A., Stiefelhagen, R. (eds.) MLMI 2008. LNCS, vol. 5237, pp. 296–308. Springer, Heidelberg (2008)
4. Romero, V.: et al.: Interactive multimodal transcription of text images using a web-based demo system. In: Proc. of the IUI, Florida, pp. 477–478 (2009)
5. Romero, V., others: Improvements in the computer assisted transciption system of handwritten text images. In: Proc. of the PRIS 2008, pp. 103–112 (2008)