Multimodal Computer-Assisted Transcription of Ancient Documents

by Verónica Romero, Alejandro Héctor Toselli, Enrqiue Vidal, Elsa Cubel and Joan Andreu Sánchez

A multimodal interactive approach for transcription of ancient documents is proposed. In this approach, user's feedback directly facilitates improvements to system accuracy while multimodality increases system ergonomy and user acceptability.

Huge historical document collections residing in libraries, museums and archives are currently being digitalized for preservation purposes and to make them available worldwide through large on-line digital libraries.The number of these on-line digital libraries is dramatically increasing. This is in part due to the ever-decreasing costs of digital storage devices and to recent advances in image digitalization and processing technologies. Thanks to these advances, hundreds of terabytes worth of ancient document digital images have been collected and can be easly made available to the historians and the public alike. However, such huge amounts of data are only of very limited use in their present raw digital image form and current efforts focus on technologies aimed at reducing the human effort required for the annotation of the raw images with informative content. In the case of text images, which are among the most numerous and interesting, the most informative annotation level is their (palaeographic) transcription into an adequate textual electronic format that would provide new ways of indexing, consulting and querying these documents.

Figure 1: Using the MM-CATTI system with a touch-screen.

Given the kind of (typically handwritten) text images involved in ancient (or even more recent historic) documents, state-of-the-art Handwritten Text Recognition (HTR) technology is very far from offering useful solutions to the transcription problem and heavy human intervention is often required to check and correct the results. This post-editing process is both inefficient and inconvenient to the user.

As an alternative to fully manual transcription and post-editing, a multimodal interactive approach is proposed here where user feedback is provided by means of touch-screen pen strokes and/or more traditional keyboard and mouse operation. User's feedback directly facilitates improvements in system accuracy, while multimodality increases system ergonomy and user acceptability. Multimodal interaction is approached in such a way that both the main and the feedback data streams help each other to optimize overall performance and usability.

In this new multimodal interactive approach for transcription of text images (Multimodal Computer-Assisted Transcription of Text Images - MM-CATTI), the user is directly involved in the transcription process, in which following a preset protocol, he/she validates and/or corrects the HTR output during the process. The protocol that rules this interaction process is formulated in the following steps: The HTR system proposes a full transcription of the input handwritten text line image. Then, the user validates the longest prefix of the transcription which is error-free and enters some on-line touch-screen pen-strokes and/or some amendment keystrokes to correct the first error in the suffix. An on-line HTR feedback subsystem (or HFR) is used to decode this input. In this way, a new extended consolidated prefix is produced based on the previous validated prefix, the on-line decoding word and the keystroke amendments. Using this new prefix, the HTR suggests a suitable continuation of it. These previous steps are iterated until a final, perfect transcription is produced.

MM-CATTI is shown to work quite well by an implemented Web-based Demo (http://cat.iti.upv.es/iht/). Figure 1 shows a user interacting with the MM-CATTI system by means of a touch-screen. The on-line form of such MM-CATTI system allows collaborative tasks with thousands of users across the globe to be carried out, thus reducing notably the overall image recognition process. Since the users operate within a web browser window, the system also provides cross-platform compatibility and requires no disk space on the client machine.

To test the effectiveness of the MM-CATTI approach, experiments were carried out on several corpora corresponding to different handwritten text transcription tasks. From the reported results on these corpora and assuming for simplicity that the cost of correcting an on-line decoding error is just similar to that of another on-line touch-screen

interaction, the estimated human effort to produce error-free transcription using MM-CATTI is reduced by as much as 15% on average, regarding to the classical HTR system. In other words, 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 are corrected automatically by the MM-CATTI system.

This multimodal interactive paradigm has been proposed by the Pattern Recognition and Human Language Technology (PRHLT) group from the Universidad Politécnica de Valencia (UPV). PRHLT is currently involved in the use of interactive techniques for machine translation and transcription through the large-budget MIPRCV Spanish project (http://miprcv.iti.es). The MIPRCV project is leaded by the PRHLT research group. MIPRCV establishes a five-year research programme to develop pattern recognition approaches that explicitly deal with the challenges and opportunities entailed by the human-interaction paradigm.

Please contact:
Joan Andreu Sánchez
Universidad Politécnica de Valencia, Spain
Tel: +34 963877253
E-mail: jandreu@dsic.upv.es