

Multimodal Computer Assisted Transcription of Handwriting Images (Duration: 6 Hrs.)

Speakers

Dr. Alejandro H. Toselli
Institut Tecnològic d'Informàtica, Universitat Politècnica de València
Camí de Vera s/n, 46071 - València, Spain
Tel +34 96 387 72 40
Fax +34 96 387 72 39
Email ahector@iti.upv.es

Dr. Moisés Pastor
Departament de Sistemes Informàtics i Computació, Universitat Politècnica de València
Camí de Vera s/n, 46022 - València, Spain
Tel +34 96 387 35 18
Fax +34 96 387 73 59
Email mpastorg@dsic.upv.es

Verónica Romero
Institut Tecnològic d'Informàtica, Universitat Politècnica de València
Camí de Vera s/n, 46071 - València, Spain
Tel +34 96 387 72 53
Fax +34 96 387 72 39
Email vromero@iti.upv.es

Alejandro H. Toselli received the M.S. degree in electrical engineering from Universidad Nacional de Tucumán (Argentina) in 1997 and the Ph.D. degree in Computer Science from Universitat Politècnica de València (Spain) in 2004. His current research interest lies in the subject of Computer Assisted and Multimodal Interaction in Pattern Recognition Systems: Handwritten Text Recognition/Transcription Applications. Dr. Alejandro is a member of the Spanish Society for Pattern Recognition and Image Analysis (AERFAI) and the International Association for Pattern Recognition (IAPR). His teaching experience consist on:

- Imparted the invited practical course “Computer Assisted Transcription of Text Images using HTK Toolkit” at the AERFAI “Summer School on New Trends in Pattern Recognition for Language Technologies” (AERFAISS’08), held in Bilbao on June 23-27, 2008.
- Experience in imparting the course “Handwritten Text Recognition” of the Official Master in Artificial Intelligence Pattern Recognition and Digital Imaging, within the Computer Science Postgraduate Programme of the Universitat Politècnica de València (2007-2009).
- Imparted the tutorial course “Interactive Multimodal Transcription of Handwritten Text Images” at the “International Conference of Document Analysis and Recognition” (ICDAR’09), held in Barcelona on July 26th, 2009.

Moisés Pastor received the M.S. degree in computer science from Universitat Politècnica de València (Spain) in 1998 and the Ph.D. degree in Computer Science from Universitat Politècnica de València (Spain) in 2007. He is professor assistant at the UPV since 2001. His current research interest lies in the areas of Pattern Recognition, Multimodal Interaction, Computer Vision and Human Language Technology..

Verónica Romero Gómez received the Master degree in computer science from the Universitat Politècnica de València, (Spain), in 2005. Since that moment she has been working on automatic handwritten text recognition and on the development of computer-assisted transcription tools for handwritten text in the "Pattern Recognition and Human Language Technology Group". In April 2006 she was granted a research fellowship by the Polytechnic University of Valencia to work towards her PhD degree. Currently, she is also collaborating on projects and teaching courses at Department of Information Systems and Computation of Polytechnic University of Valencia.

Abstract

To date, automatic Handwriting Text Recognition (HTR) systems are far from being perfect and heavy human intervention is often required to check and correct the results of such systems. This “post-editing” process is both inefficient and uncomfortable to the users. As an alternative to post-editing, a multimodal interactive approach can be adopted, where user feedback is provided by means of touch-screen pen strokes and/or more traditional keyboard and mouse operation. To this end, both on-line and off-line HTR must be combined in such a way that the two modalities help each-other in order to improve accuracy and usability. This framework is referred to as Multimodal Computer Assisted Transcription of Text Images (MM-CATTI). In this course we look at MM-CATTI from a general paradigm of Human-Interaction in Pattern Recognition and examine the important challenges and interesting research opportunities entailed the resulting framework. In particular we show how to take direct advantage of the feedback information provided by the user in each interaction step to improve raw recognition performance. We also argue that human interaction naturally entails multimodal processing and show how this fact can lead to improve overall system behavior and usability.

Topics to be Covered

I. Introduction

- a) Multimodal Interaction in Pattern Recognition (Motivation, Formal Framework and Application Examples)
- b) Interactive-Predictive Pattern Recognition and Document Image Analysis
- c) Quick survey of Handwritten Text Recognition (HTR) concepts and techniques

II. Computer-Assisted Transcription of Text Images (CATTI)

- a) Human interaction in HTR
- b) A CATTI formal framework
- c) Feedback-derived dynamic language modelling and search
- d) Performance measures and results achieved in typical applications

III. Multimodality in CATTI (MM-CATTI)

- a) Touchscreen based multimodal user correction
- b) A MM-CATTI formal framework
- c) Multimodal language modelling and search
- d) Performance measures and results achieved in typical applications

IV. Basic HTR and CATTI experiments:

- a) HTR Preprocessing
- b) Training HMMs using the “Hidden Markov Model Toolkit” (HTK)
- c) Training Language Models and Dictionaries for HTR
- d) HTR Experiments
- e) Adapting Language Models and Search for CATTI
- f) CATTI Experiments
- g) Analyzing quantitatively the CATTI performance
- h) Demonstration of a complete MM-CATTI System in a real HTR task