

# Interactive Multimodal Transcription of Handwritten Text Images

Enrique Vidal  
Alejandro Tosselli

## Overview

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 behaviour and usability.

## Intended Learning Outcomes

This tutorial is intended to introduce students and newcomers to modern approaches to one of the major topics of ICDAR research; namely Handwritten Text Recognition. Moreover, the tutorial aims at motivating and explaining how this important ICDAR topic can be advantageously addressed under the emerging paradigm of Multimodal Interaction based on Pattern Recognition technology. More specifically, attendants of this course will:

- a) acquire basic knowledge about the Interactive-Predictive, Multimodal-Interaction framework in Pattern Recognition and Image Processing,
- b) learn how these concepts are used in the field of HTR,
- c) conduct basic HTR experiments using the "Hidden Markov Model Toolkit" (HTK) and other experiments and user tests related to the MM-CATTI approach taught in the theory part of the course,
- d) analyze quantitatively the MM-CATTI performance, through the comparison between the interactive correction-costs using MM-CATTI and those required by post-editing the conventional, noninteractive HTR output,
- e) get hands-on experience of using a complete multimodal interactive system for HTR.

## Outline

### I. Introduction

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

- II. Computer-Assisted Transcription of Text Images (CATTI)
  - II.a) Human interaction in HTR
  - II.b) A CATTI formal framework
  - II.c) Feedback-derived dynamic language modelling and search
  - II.d) Performance measures and results achieved in typical applications
- III. Multimodality in CATTI (MM-CATTI)
  - III.a) Touchscreen based multimodal user correction
  - III.b) A MM-CATTI formal framework
  - III.c) Multimodal language modelling and search
  - III.d) Performance measures and results achieved in typical applications
- IV. Basic HTR and CATTI experiments:
  - IV.a) HTR Preprocessing
  - IV.b) Training HMMs using the "Hidden Markov Model Toolkit" (HTK)
  - IV.c) Training Language Models and Dictionaries for HTR
  - IV.d) HTR Experiments
  - IV.e) Adapting Language Models and Search for CATTI
  - IV.f) CATTI Experiments
  - IV.g) Analyzing quantitatively the CATTI performance
  - IV.h) Demonstration of a complete MM-CATTI System in a real HTR task