# Abstract


## Autonomously Normalized Horizontal Differentials as Features for HMM-Based Omni Font-Written OCR Systems for Cursively Scripted Languages

Automatic font-written Optical Character Recognition (OCR) is highly desirable for numerous modern information technology (IT) applications. Reliable font-written OCR's for Latin scripts are readily in use since long. For cursively scripted languages, that are the mother tongues of over one fourth of the world population, such OCR's are however not available at a robust and reliable performance. In this regard, the main challenge is the mandatory connectivity of characters/ligatures (i.e. graphemes) that has to be resolved simultaneously upon the recognition of these graphemes. Among the various approaches tried over decades, Hidden Markov Models (HMM)-based OCR's seem to be the most promising as they capitalize on the ability of HMM decoders to achieve segmentation and recognition simultaneously similar to the widely used HMM-based automatic speech recognition (ASR). Unlike ASR's, what is missing in HMM-based OCR's is the definition of a rigorously founded features vector capable to robustly achieving minimal “font type/size-independent” (omnifont) word error rates comparable to those realized with Latin scripts. Here comes the contribution of this paper that introduces such a sound features vector design, and experimentally shows its superiority in this regard.