ABSTRACT

This paper presents two new preprocessing techniques for cursive script recognition. Enhanced algorithms for core-region detection and effective uniform slant angle estimation are proposed. Reference lines composed of core-region are usually obtained as the ones surrounding highest density peaks, but are strongly affected by the presence of long horizontal strokes and erratic characters in the word. Therefore, it caused confusion with the actual core-region and leads to decisive errors in normalizing the word. To overcome this problem in core-region detection quantile is introduced to make resulting process robust. On the other hand, research community has introduced computationally heavy approaches to remove slant in cursive script. Therefore, a simple formalized and effective method is presented for the detection and removal of slant angle for offline cursive handwritten words to avoid heavy experimental efforts. Additionally, already not-slanted words are not affected negatively by applying this algorithm. The core-region detection is based on statistical features, while slant angle estimation is based on structure features of the word image. The algorithms are tested on IAM benchmark database of cursive handwritten words. Promising results for core-region detection, slant angle estimation/removal are reported and compared with widely applied Bozinovic and Srihari method (BSM).