

APPROACHES TO HANDWRITTEN/MACHINE PRINTED DISCRIMINATION PROBLEM

Maria-Fiorina Popa, graduate student, Dept. of Math,
and Comp. Science, Bucharest University,

Doru-Cristian Balcan, graduate student, Dept. of Math,
and Comp. Science, Bucharest University.

Abstract

Automatic document processing tasks aim getting correct identification and interpretation of the elements (e.g. text, pictures, graphs) currently found in document images. The research reported in the paper explores new approaches in extracting text, the methodology being based on page segmentation and discrimination between handwritten (HW) and machine-printed (MP) text. The work consists of using several performance measures taking into account a series of regularities of shape and dimensions, as for instance: self-correlation, symmetry, aspect-ratio. For the HW/MP discrimination problem a heuristic procedure and a Bayesian approach were considered and studied. Conclusions and the experimental results are shown in the final section of the paper.