

1. Record Nr.	UNINA9910145148603321
Autore	Wani M. Arif
Titolo	Machine Learning and Applications; Proceedings: International Conference on Machine Learning and Applications (6th: 2007: Cincinnati, Ohio)
Pubbl/distr/stampa	[Place of publication not identified], : IEEE Computer Society Press, 2007
ISBN	9781509089468 1509089462
Descrizione fisica	1 online resource
Disciplina	006.31
Soggetti	Machine learning
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Bibliographic Level Mode of Issuance: Monograph
Sommario/riassunto	An optical character recognition (OCR) system with a high recognition rate is challenging to develop. One of the major contributors to OCR errors is smeared characters. Several factors lead to the smearing of characters such as bad scanning quality and a poor binarization technique. Typical approaches to character segmentation falls into three major categories: image-based, recognition-based, and holistic-based. Among these approaches, the segmentation path can be linear or non-linear. Our paper proposes a non-linear approach to segment characters on grayscale document images. Our method first determines whether characters are smeared together using general character features. The correct segmentation path is found using a shortest path approach. We achieved a segmentation accuracy of 95% over a set of about 2,000 smeared characters.