

## Guide to OCR for Indic Scripts

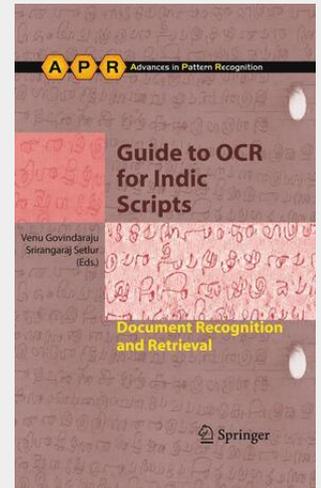
Document Recognition and Retrieval

of researchers, scholars, and students use and depend upon Internet-based content and computational resources.

The chapters in this book describe a critically important area of investigation—addressing conversion of Indic script into machine-readable form. Rough estimates have it that currently more than a billion people use Indic scripts. Collectively, Indic historic and cultural documents contain a vast richness of human knowledge and experience. The state-of-the-art research described in this book demonstrates the multiple values associated with these activities. Technically, the problems associated with Indic script recognition are every day and will contribute to and inform related v vi Foreword script recognition efforts. The work also has enormous consequence for enriching and enabling the study of Indic cultural heritage materials and the historic record of its people. This in turn broadens the intellectual context for domain scholars focusing on other societies, ancient and modern. Digital character recognition has brought about another milestone in collective communication by bringing in, text into an interactive digital realm.

In doing so, the information has gained additional functionalities which expand our abilities to connect, combine, contextualize, share, and collaboratively pursue knowledge making. High-quality Internet content continues to grow in an explosive fashion. In the new global cyber environment, the functionalities and applications of digital information continue to transform knowledge into new understandings of human experience and the world in which we live. The possibilities for the future are limited only by available research resources and capabilities and the imagination and creativity of those who use them. Arlington, Virginia Stephen M.

Optical Character Recognition (OCR) is a key enabling technology critical to creating indexed, digital library content, and it is especially valuable for Indic scripts, for which there has been very little digital access. Indic scripts, the ancient Brahmi scripts prevalent in the Indian subcontinent, present some challenges for OCR that are different from those faced with Latin and Oriental scripts. But properly utilized, OCR will help to make Indic digital archives practically accessible to researchers and lay users alike by creating searchable indexes and machine-readable text repositories. This unique guide/reference is the very first comprehensive book on the subject of OCR for Indic scripts, providing an overview of the state-of-the-art research in this field as well as other issues related to facilitating query and retrieval of Indic documents from digital libraries. All major research groups working in this area are represented in this book, which is divided into sections on recognition of Indic scripts and retrieval of Indic documents. Topics and features: - Contains contributions from the leading researchers in the field - Discusses data set creation for OCR development - Describes OCR systems that cover eight different scripts: Bangla, Devanagari, Gurmukhi, Gujarati, Kannada, Malayalam, Tamil, and Urdu (Perso-Arabic) - Explores the challenges of Indic script handwriting recognition in the online domain - Examines the development of handwriting-based text input systems - Describes ongoing work to increase access to Indian cultural heritage materials - Provides a section on the enhancement of text and images obtained from historical Indic palm leaf manuscripts - Investigates different techniques for word spotting in Indic scripts - Reviews mono-lingual and cross-lingual information retrieval in Indic languages This is an excellent reference for researchers and graduate students studying OCR technology and methodologies. This volume will contribute to opening up the rich Indian cultural heritage embodied in millions of ancient and contemporary documents spanning topics such as science, literature, medicine, astronomy, mathematics and philosophy. Venu Govindaraju FIEEE FIAPR, is a Distinguished Professor of Computer Science and Engineering at the University at Buffalo. He has over 20 years of research experience in pattern recognition, information retrieval and biometrics. His seminal work on handwriting recognition was at the core of the first handwritten address interpretation system used by



**160,49 €**

149,99 € (zzgl. MwSt.)

Lieferfrist: bis zu 10 Tage

**Artikelnummer:** 9781447125181

**Medium:** Buch

**ISBN:** 978-1-4471-2518-1

**Verlag:** Springer

**Erscheinungstermin:** 14.03.2012

**Sprache(n):** Englisch

**Auflage:** 2010

**Serie:** Advances in Computer Vision and Pattern Recognition

**Produktform:** Kartoniert

**Gewicht:** 528 g

**Seiten:** 325

**Format (B x H):** 155 x 235 mm

