

# Contents

## Part I Recognition of Indic Scripts

|   |     |
|---|-----|
| <b>Building Data Sets for Indian Language OCR Research</b> . . . . .                        | 3   |
| C.V. Jawahar, Anand Kumar, A. Phaneendra, and K.J. Jinesh                                   |     |
| <b>On OCR of Major Indian Scripts: Bangla and Devanagari</b> . . . . .                      | 27  |
| B.B. Chaudhuri  |     |
| <b>A Complete Machine-Printed Gurmukhi OCR System</b> . . . . .                             | 43  |
| G.S. Lehal  |     |
| <b>Progress in Gujarati Document Processing and Character Recognition</b>                   | 73  |
| Jignesh Dholakia, Atul Negi and S. Rama Mohan   |     |
| <b>Design of a Bilingual Kannada–English OCR</b> . . . . .                                  | 97  |
| R.S. Umesh, Peeta Basa Pati and A.G. Ramakrishnan   |     |
| <b>Recognition of Malayalam Documents</b> . . . . .   | 125 |
| N.V. Neeba, Anoop Namboodiri, C.V. Jawahar, and P.J. Narayanan                              |     |
| <b>A Complete OCR System for Tamil Magazine Documents</b> . . . . .                         | 147 |
| Aparna Kokku and Srinivasa Chakravarthy   |     |
| <b>Experiments on Urdu Text Recognition</b> . . . . .                                       | 163 |
| Omar Mukhtar, Srirangaraj Setlur, and Venu Govindaraju                                      |     |
| <b>The BBN Byblos Hindi OCR System</b> . . . . .  | 173 |
| Prem Natarajan, Ehry MacRostie, and Michael Decerbo   |     |
| <b>Generalization of Hindi OCR Using Adaptive Segmentation and<br/>Font Files</b> . . . . . | 181 |
| Mudit Agrawal, Huanfeng Ma, and David Doermann  |     |
| <b>Online Handwriting Recognition for Indic Scripts</b> . . . . .                           | 209 |
| A. Bharath and Sriganesh Madhvanath   |     |

## **Part II Retrieval of Indic Documents**

|   |            |
|---|------------|
| <b>Enhancing Access to Primary Cultural Heritage Materials of India . . .</b>                 | <b>237</b> |
| Peter M. Scharf and Malcolm Hyman   |            |
| <b>Digital Image Enhancement of Indic Historical Manuscripts . . . . .</b>                    | <b>249</b> |
| Zhixin Shi, Srirangaraj Setlur and Venu Govindaraju   |            |
| <b>GFG-Based Compression and Retrieval of Document Images in<br/>Indian Scripts . . . . .</b> | <b>269</b> |
| Gaurav Harit, Santanu Chaudhury and Ritu Garg   |            |
| <b>Word Spotting for Indic Documents to Facilitate Retrieval . . . . .</b>                    | <b>285</b> |
| Anurag Bhardwaj, Srirangaraj Setlur, and Venu Govindaraju                                     |            |
| <b>Indian Language Information Retrieval . . . . .</b>  | <b>301</b> |
| Prasenjit Majumder and Mandar Mitra   |            |
| <b>Colour Plates . . . . .</b>  | <b>315</b> |
| <b>Index . . . . .</b>  | <b>321</b> |