

Contents

1	Introduction	1
1.1	Organization of the Monograph	1
1.2	Notation	3
1.3	State of Art	4
1.4	Research Issues and Challenges	5
1.5	Figures	5
1.6	MATLAB OCR Toolbox	5
	References	6
2	Optical Character Recognition Systems	9
2.1	Introduction	9
2.2	Optical Character Recognition Systems: Background and History	12
2.3	Techniques of Optical Character Recognition Systems	15
2.3.1	Optical Scanning	15
2.3.2	Location Segmentation	17
2.3.3	Pre-processing	17
2.3.4	Segmentation	22
2.3.5	Representation	23
2.3.6	Feature Extraction	28
2.3.7	Training and Recognition	29
2.3.8	Post-processing	34
2.4	Applications of Optical Character Recognition Systems	35
2.5	Status of Optical Character Recognition Systems	37
2.6	Future of Optical Character Recognition Systems	40
	References	40
3	Soft Computing Techniques for Optical Character Recognition Systems	43
3.1	Introduction	43
3.2	Soft Computing Constituents	46
3.2.1	Fuzzy Sets	46

3.2.2	Artificial Neural Networks	48
3.2.3	Genetic Algorithms	50
3.2.4	Rough Sets	53
3.3	Hough Transform for Fuzzy Feature Extraction	55
3.4	Genetic Algorithms for Feature Selection	56
3.5	Rough Fuzzy Multilayer Perceptron	59
3.6	Fuzzy and Fuzzy Rough Support Vector Machines	66
3.7	Hierarchical Fuzzy Bidirectional Recurrent Neural Networks . . .	73
3.8	Fuzzy Markov Random Fields	78
3.9	Other Soft Computing Techniques	82
	References	82
4	Optical Character Recognition Systems for English Language	85
4.1	Introduction	85
4.2	English Language Script and Experimental Dataset	87
4.3	Challenges of Optical Character Recognition Systems for English Language	88
4.4	Data Acquisition	90
4.5	Data Pre-processing	90
4.5.1	Binarization	90
4.5.2	Noise Removal	91
4.5.3	Skew Detection and Correction	91
4.5.4	Character Segmentation	91
4.5.5	Thinning	92
4.6	Feature Extraction	92
4.7	Feature Based Classification: Sate of Art.	94
4.7.1	Feature Based Classification Through Fuzzy Multilayer Perceptron	95
4.7.2	Feature Based Classification Through Rough Fuzzy Multilayer Perceptron	95
4.7.3	Feature Based Classification Through Fuzzy and Fuzzy Rough Support Vector Machines	96
4.8	Experimental Results	96
4.8.1	Fuzzy Multilayer Perceptron	96
4.8.2	Rough Fuzzy Multilayer Perceptron	100
4.8.3	Fuzzy and Fuzzy Rough Support Vector Machines	100
4.9	Further Discussions	105
	References	106
5	Optical Character Recognition Systems for French Language	109
5.1	Introduction	109
5.2	French Language Script and Experimental Dataset	111
5.3	Challenges of Optical Character Recognition Systems for French Language	113
5.4	Data Acquisition	114
5.5	Data Pre-processing	115

5.5.1	Text Region Extraction.	115
5.5.2	Skew Detection and Correction	116
5.5.3	Binarization	117
5.5.4	Noise Removal	118
5.5.5	Character Segmentation	118
5.5.6	Thinning	120
5.6	Feature Extraction Through Fuzzy Hough Transform	120
5.7	Feature Based Classification: Sate of Art.	122
5.7.1	Feature Based Classification Through Rough Fuzzy Multilayer Perceptron	123
5.7.2	Feature Based Classification Through Fuzzy and Fuzzy Rough Support Vector Machines	123
5.7.3	Feature Based Classification Through Hierarchical Fuzzy Bidirectional Recurrent Neural Networks	124
5.8	Experimental Results	124
5.8.1	Rough Fuzzy Multilayer Perceptron	124
5.8.2	Fuzzy and Fuzzy Rough Support Vector Machines	127
5.8.3	Hierarchical Fuzzy Bidirectional Recurrent Neural Networks	129
5.9	Further Discussions	132
	References	135
6	Optical Character Recognition Systems for German Language	137
6.1	Introduction	137
6.2	German Language Script and Experimental Dataset	139
6.3	Challenges of Optical Character Recognition Systems for German Language	140
6.4	Data Acquisition.	141
6.5	Data Pre-processing	142
6.5.1	Text Region Extraction.	142
6.5.2	Skew Detection and Correction	143
6.5.3	Binarization	144
6.5.4	Noise Removal	145
6.5.5	Character Segmentation	145
6.5.6	Thinning	146
6.6	Feature Selection Through Genetic Algorithms.	148
6.7	Feature Based Classification: Sate of Art.	150
6.7.1	Feature Based Classification Through Rough Fuzzy Multilayer Perceptron.	151
6.7.2	Feature Based Classification Through Fuzzy and Fuzzy Rough Support Vector Machines	152
6.7.3	Feature Based Classification Through Hierarchical Fuzzy Bidirectional Recurrent Neural Networks	152
6.8	Experimental Results	152
6.8.1	Rough Fuzzy Multilayer Perceptron	153
6.8.2	Fuzzy and Fuzzy Rough Support Vector Machines	155

6.8.3	Hierarchical Fuzzy Bidirectional Recurrent Neural Networks	161
6.9	Further Discussions	162
	References	163
7	Optical Character Recognition Systems for Latin Language	165
7.1	Introduction	165
7.2	Latin Language Script and Experimental Dataset	167
7.3	Challenges of Optical Character Recognition Systems for Latin Language	168
7.4	Data Acquisition	170
7.5	Data Pre-processing	170
7.5.1	Text Region Extraction	170
7.5.2	Skew Detection and Correction	171
7.5.3	Binarization	172
7.5.4	Noise Removal	173
7.5.5	Character Segmentation	173
7.5.6	Thinning	174
7.6	Feature Selection Through Genetic Algorithms	175
7.7	Feature Based Classification: Sate of Art	178
7.7.1	Feature Based Classification Through Rough Fuzzy Multilayer Perceptron	178
7.7.2	Feature Based Classification Through Fuzzy and Fuzzy Rough Support Vector Machines	179
7.7.3	Feature Based Classification Through Hierarchical Fuzzy Rough Bidirectional Recurrent Neural Networks	179
7.8	Experimental Results	180
7.8.1	Rough Fuzzy Multilayer Perceptron	180
7.8.2	Fuzzy and Fuzzy Rough Support Vector Machines	183
7.8.3	Hierarchical Fuzzy Rough Bidirectional Recurrent Neural Networks	186
7.9	Further Discussions	188
	References	190
8	Optical Character Recognition Systems for Hindi Language	193
8.1	Introduction	193
8.2	Hindi Language Script and Experimental Dataset	196
8.3	Challenges of Optical Character Recognition Systems for Hindi Language	197
8.4	Data Acquisition	200
8.5	Data Pre-processing	200
8.5.1	Binarization	200
8.5.2	Noise Removal	201
8.5.3	Skew Detection and Correction	201
8.5.4	Character Segmentation	201
8.5.5	Thinning	202

8.6	Feature Extraction Through Hough Transform	202
8.7	Feature Based Classification: Sate of Art.	204
8.7.1	Feature Based Classification Through Rough Fuzzy Multilayer Perceptron.	205
8.7.2	Feature Based Classification Through Fuzzy and Fuzzy Rough Support Vector Machines	205
8.7.3	Feature Based Classification Through Fuzzy Markov Random Fields	206
8.8	Experimental Results	206
8.8.1	Rough Fuzzy Multilayer Perceptron	206
8.8.2	Fuzzy and Fuzzy Rough Support Vector Machines	208
8.8.3	Fuzzy Markov Random Fields	208
8.9	Further Discussions	209
	References	215
9	Optical Character Recognition Systems for Gujrati Language	217
9.1	Introduction	217
9.2	Gujrati Language Script and Experimental Dataset.	219
9.3	Challenges of Optical Character Recognition Systems for Gujrati Language	220
9.4	Data Acquisition.	224
9.5	Data Pre-processing	224
9.5.1	Binarization	224
9.5.2	Noise Removal	225
9.5.3	Skew Detection and Correction	225
9.5.4	Character Segmentation	225
9.5.5	Thinning	225
9.6	Feature Selection Through Genetic Algorithms.	226
9.7	Feature Based Classification: Sate of Art.	228
9.7.1	Feature Based Classification Through Rough Fuzzy Multilayer Perceptron.	229
9.7.2	Feature Based Classification Through Fuzzy and Fuzzy Rough Support Vector Machines	230
9.7.3	Feature Based Classification Through Fuzzy Markov Random Fields	230
9.8	Experimental Results	231
9.8.1	Rough Fuzzy Multilayer Perceptron	231
9.8.2	Fuzzy and Fuzzy Rough Support Vector Machines	231
9.8.3	Fuzzy Markov Random Fields	235
9.9	Further Discussions	236
	References	238
10	Summary and Future Research.	241
10.1	Summary	241
10.2	Future Research	243
	References	244
	Index.	247

Optical Character Recognition Systems for Different
Languages with Soft Computing

Chaudhuri, A.; Mandaviya, K.; Badelia, P.; K Ghosh, S.

2017, XIX, 248 p. 95 illus., Hardcover

ISBN: 978-3-319-50251-9