

Contents

0 Introduction	1
1 Chinese Characters: Properties and Problems	3
1.1 History	3
1.11 The "liu shu": Six Classes of Characters	3
1.12 Styles: From Oracle Bones to Movable Types	7
1.13 Applications: The "Latin of East Asia"	10
1.14 Character Simplification	12
1.2 Modern Printed Characters	14
1.21 Font Styles	14
1.22 Font Sizes	16
1.23 Punctuation and Layout	17
1.3 Character Structure	18
1.31 Strokes	19
1.32 Components	20
1.33 Composition Structure	21
1.4 Chinese Characters in the Computer	22
1.41 Processing Problems	23
1.42 Input Coding	26
1.43 On-line Character Recognition	27
1.44 Internal and Transmission Coding	27
1.45 Requirements to Recognition Systems	28
2 Input and Preprocessing: Setting the Stage	31
2.1 Optical Input	31
2.11 Resolution Requirements	31
2.12 Camera Input	32
2.13 Standardizing Recording Conditions	33
2.14 Scanner Input	34
2.15 Digitizing Devices	34
2.16 Frame Preprocessing	35
2.2 Picture Segmentation	35
2.21 The Fixed Distance Approach	36
2.22 An Improved Approach	36
2.23 Check on Correct Segmentation	39

2.3	Size Transformation	41
2.31	Individual Size Determination	43
2.32	Frame-global Size Determination	44
2.33	Bi-Linear Interpolation	44
2.34	Cubic Spline Interpolation	46
2.35	Size Normalization on Binary Patterns	47
2.4	Binarization	50
2.5	Edge Smoothing on Binary Patterns	53
3	Feature Extraction	55
3.1	Principles of Feature Extraction	55
3.2	Useful Tools	56
3.21	Blackness	56
3.22	Stroke Width	56
3.23	Projection Profiles	58
3.24	Transitions	59
3.3	Some Feature Algorithms	60
3.31	Pattern Match	60
3.32	Peripheral Features	61
3.33	Stroke Density	65
3.34	Local Direction Contributivity	67
3.35	Stroke Proportion	69
3.36	Black Jump Distribution in Balanced Subfields	71
3.37	Total Stroke Length	75
3.4	Combination of Features	76
3.41	Principles of Feature Combination	76
3.42	Successful Combination	77
3.5	Structural Analysis	82
3.51	Thinning Algorithm	83
3.52	Feature Point Detection	84
3.53	Stroke Tracking and Stroke Segmentation	85
3.54	Combination of Segments	87
3.55	Stroke Synthesis	89
3.56	Structural Description	90
4	Classification	94
4.1	Principles of Classification	94
4.2	Classification Tools	96
4.21	Special Discriminant Function	96
4.22	Similarity	96
4.23	Classification Behavior	97
4.3	Combining Distance Measures	99

4.4 Hierarchical Classification	101
4.5 The Overlap Problem	102
4.6 Dynamic Classification	106
4.7 Plausibility Checks	108
4.71 Absolute Plausibility Checks	109
4.72 Character Frequency	111
4.73 Pattern Match	111
4.74 Discrimination Procedure	113
4.8 Learning Mechanisms	114
5 The TECHIS System: Implementation and Results	116
5.1 Hardware and Software Conditions	116
5.2 Program Implementation	117
5.3 Database on Chinese Characters	118
5.31 Reference Patterns	118
5.32 Frequency Data	119
5.33 Linguistic Data: Characters	121
5.34 Linguistic Data: Words and Phrases	122
5.4 Tests on Features	123
5.41 Feature Files	123
5.42 Test Data	124
5.43 Results for the BJD-BS Feature	125
5.44 Results of Feature Combination	127
References	130
List of Figures	137
Index	139