

[sijang, jkim}@ai.kaist.ac.kr](mailto:{sijang, jkim}@ai.kaist.ac.kr)

Post-processing of Hangeul Recognition for Discriminating Pairs of Characters

Seung-Ick Jang^o Jin-Hyung Kim
Dept. of EECS, KAIST

가
Support Vector Machine

, Support Vector Machine

,
가

1.

가 11,172
11,172

가

가 (intra-class variation)

가 (inter-class variation)

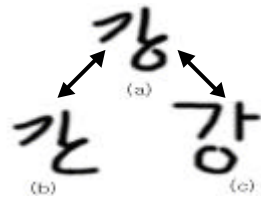
(b) 가 , (c) 가

(a) (b)

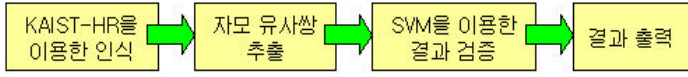
가 , (a) (c)

가 (a) (b)

가



[1]



[2]

[2] . KAIST-HR[1]

[5].

Chou가

heuristic

Support Vector Machine[9]

가

. 2

, 3

KAIST-HR

[6].

. 4

,5

3.

3.1

2.

KAIST-HR

Takahashi, Chiang

. KAIST-HR

[2,3]. Takahashi

(binary

classifier)

[2].

가

가

KAIST-HR

SERI database[7]

, 1

가

. Chiang

가

84.3%

, 2

가

91.7%

가

, 1

2

[3].

47.1%

. KAIST-HR

가

1

2

60%

가

40%

Chou,

가

[4,5,6]. Chou

[4].

가

heuristic

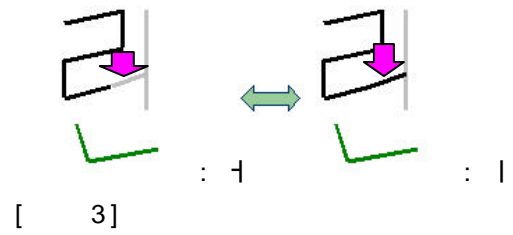
heuristic

가

가

{ ‘ ’ }

$A = \{\text{'ㅁ'}, \text{'ㅏ'}\}$, $B = \{\text{'ㅇ'}, \text{'ㅏ'}\}$
 $A - B = \{\text{'ㅁ'}\}$, $B - A = \{\text{'ㅇ'}\}$
 $A \cap B = \{\text{'ㅏ'}\}$
 $A \cup B = \{\text{'ㅁ'}, \text{'ㅇ'}, \text{'ㅏ'}\}$



[3] : A 가 B 가
 [3] : B 가 A 가
 overlapping rate

[1] SERI database 200 set, 10

3.2

KAIST-HR

label

[1]

유사쌍 누적 비율	남자 유사쌍 수	자모 유사쌍 수
10%	36	1
20%	87	3
30%	160	6
40%	258	11
50%	397	23
60%	613	39
70%	936	69
80%	1469	137
90%	2222	344
100%	2975	1031

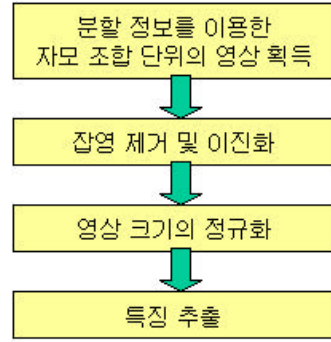
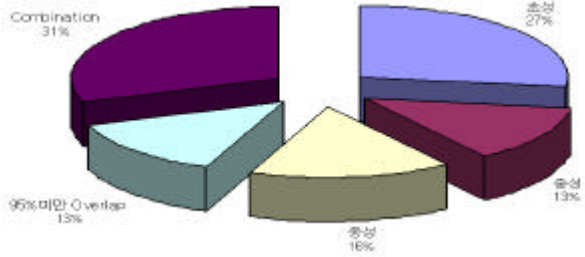
KAIST-HR
 C_1 , C_2
 S_1 , S_2
 s_i , s_j
 overlapping rate $O(C_1, C_2)$ (1)

$$O(C_1, C_2) = \frac{\min(\sum_i length(s_i), \sum_j length(s_j))}{\max(\sum_i length(s_i), \sum_j length(s_j))} \quad (1)$$

가
 가
 overlapping rate가

95%
 [2] KAIST-HR
 HR 1 2
 Overlapping rate가 95%
 56%
 가 47.1% 가
 26.4% 가 가

[2] 1 2



[5]

3.3

KAIST-HR

가 (generalization) 가

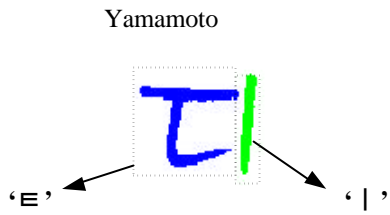
가 [4] ‘가’

‘ㅍ’, ‘ㅑ’, ‘ㅓ’, ‘ㅕ’

가

mesh feature

5] KAIST-HR



[4] ‘가’

[8].

(equalization)

30 x 30

Support Vector Machine

Joachims 가

SVM^{light}

[10].

4.

SERI

database SERI database 1000 set

set

520

가

SERI database 80 set



(a) 수직 수평에 대한 투영

(b) 표본화

(c) 결과

[6] Yamamoto

40,580

KAIST-HR

[4]

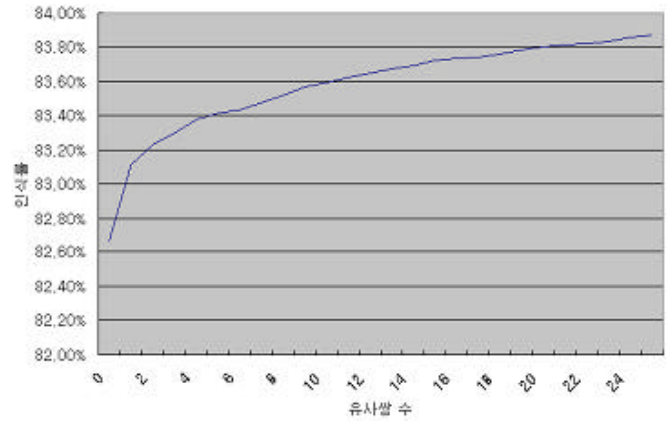
8,000
1,000
SERI database 120 set, 62,398

가 25

87%

KAIST-HR

가



유사쌍수	0	5	10	15	20	25
인식률	82.66%	83.41%	83.60%	83.72%	83.80%	83.87%
오류감소율	0.00%	4.33%	5.42%	6.11%	6.57%	6.98%

[3]

유사쌍	전체수	정답수	오류정정	오류발생	오류감소
ㅁ↔ㅇ	5522	4946	346	64	49.00%
ㄸ↔ㅍ	473	375	77	6	72.40%
ㅈ↔ㅊ	1251	1162	63	19	49.40%
ㅇ↔ㅎ	392	314	55	8	60.30%
ㄷ↔ㅌ	1104	1048	27	4	41.10%
ㄹ↔ㄺ	366	305	27	14	21.30%
ㄱ↔ㅅ	428	353	34	7	36.00%
ㅁ↔ㅂ	742	674	34	8	38.20%
ㅇ↔ㄷ	233	184	33	2	63.30%
ㄷ↔ㄴ	586	557	20	3	58.60%
ㄹ↔ㄴ	498	445	27	7	37.70%
ㄹ↔ㅎ	169	135	20	6	41.20%
ㅇ↔ㅂ	156	130	16	0	61.50%
ㅁ↔ㅍ	242	213	12	3	31.00%
ㄱ↔ㅈ	207	178	20	0	69.00%
ㄷ↔ㄹ	338	300	11	3	21.10%
ㅍ↔ㅌ	61	43	13	8	27.80%
ㄴ↔ㄷ	398	371	13	3	37.00%
ㄷ↔ㅈ	122	99	16	0	69.60%
ㅎ↔ㅌ	84	68	14	3	68.80%
ㅁ↔ㄱ	280	263	11	4	41.20%
ㅇ↔ㄴ	150	135	11	5	40.00%
ㅅ↔ㅈ	435	399	12	4	22.20%
ㅎ↔ㅂ	126	104	17	2	68.20%
ㄹ↔ㅂ	76	63	9	1	61.50%
계	14439	12864	938	184	47.90%

[3] . [3]

가

‘ㅁ’ ‘ㅇ’ 5,522
KAIST-HR 4,946
64 가 가 346
, 49.0%

[4]

62,398 KAIST-HR

82.66%

5

가

가 25

가

25 가

[7]

7] (a) KAIST-HR

(b) (c)

가



(a)



(b)



(c)

[7]

5.

Support Vector Machine

KAIST-HR

. Support

Vector Machine

mesh feature

가

25

89.1%

94.3%

47.9%

가

6.

[1] H.Y. Kim, Representation and parameter estimation of hierarchical random graph and its application to

handwritten Hangeul recognition. Ph.D. Dissertation. KAIST (1999)

[2] H. Takahashi, T.D. Grffin, Recognition enhancement by linear tournament verification, Proc. Second International Conference on Document Analysis and Recognition, (1993) pp. 585-588

[3] C.C. Chiang, S.S. Yu, A method for improving the machine recognition of confusing Chinese characters. Proc. Thirteenth International Conference on Pattern Recognition, (1996) pp. C79-83

[4] K.S. Chou, K.C. Fan, C.K. Lin, A knowledge based approach to the recognition of on-line confusing Chinese characters, Proc. Fourth International Workshop of Frontiers of Handwriting Recognition, (1994) pp. 185-194

[5] J.O. Kwon, B.K. Sin, J. Kim, Recognition of on-line cursive Korean characters combining statistical and structural method, Pattern recognition 30 (8) (1997) pp. 1255-1263

[6] I.J. Kim, Handwritten Chinese Character Recognition using Statistical Structure Modeling and Stroke Weighting. Ph.D. Dissertation. KAIST (2001)

[7] D.-I. Kim, S.-W. Lee, An automatic evaluation of handwriting qualities for offline handwritten Hangeul character database KU-1, Proc. of the 25th Korea Information Science Society Conference 25 (1) (1998) pp. 707-709

[8] H. Yamada, K. Yamamoto and T. Saito, A Nonlinear Normalization method for Handwritten Kanji Character Recognition-Line Density Equalization, Pattern Recognition, vol. 23, no. 9 (1990) pp.1023-1029

[9] Vladimir N. Vapnik, The Nature of Statistical Learning Theory. Springer, (1995)

[10] http://ais.gmd.de/~thorsten/svm_light/