国立大学法人
東京農工大学
Tokyo University of Agriculture and Technology

Large Improvement in Line-direction-free and Character-orientation-free On-line Handwritten Japanese Text Recognition

Yuechan Hao, Bilan Zhu and Masaki Nakagawa
Tokyo University of Agriculture and Technology, Japan
haoyuechan@yahoo.co.jp,{zhubilan,nakagawa}@cc.tuat.ac.jp

Contents

1 Background

2 Objective

3 Flow of text recognition

4 Recognition Model

5 Experiments

6 Conclusion

Background





- Development of pen-based or touch-based systems:

smart phone, tablet PC, electronic whiteboard, Anoto pen, e-pen and so on.

- Expansion of writing surface.

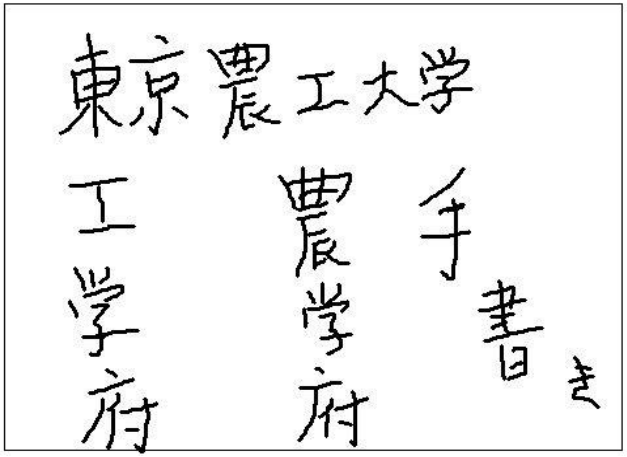
Handwritten text recognition is being sought with less constraints and allowing people to write more freely.

Objective

Recognize on-line handwritten text free from line direction and character orientation constraints.

- **In Japan and China, people write text horizontally, vertically or even diagonally on a whiteboard and tablet.**
- **Table top interfaces should allow people to write text without line direction and character orientation constraints.**
- **Recognition rates of any line direction and character orientation should be as high as that of horizontal line and normal character orientation.**

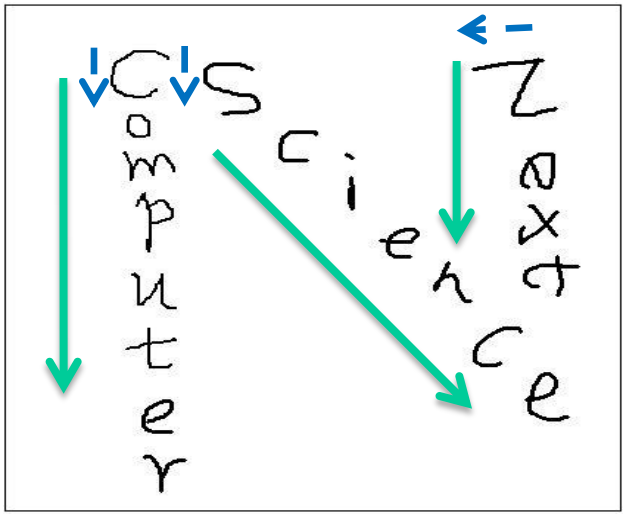
Line Direction and Charter Orientation Free Text



recognize →

東京農工大学

工学府　農学府　手書き



recognize →

Computer Science Next

→ Line direction

- - -> Character orientation

Flow of text recognition

- **Step 1 Text line segmentation**
  - Segment text lines composed of horizontal, vertical and slanted lines of text with arbitrary character orientation into text line elements.
- **Step 2 Orientation Estimation**
  - Estimate and normalize character orientation and line direction.
- **Step 3 Over-segmentation**
  - Decide segmentation points and non-segmentation points in quantized 4 directions using the two-stage classification scheme.
- **Step 4 Construction of candidate lattice**
  - Evaluate the likelihood of candidate segmentation paths.
- **Step 5 Search and recognition**
  - Optimal path can be found by the Viterbi search.

Flow of Text Recognition

(1) Text line segmentation

- Block grouping
- Pre-segmentation
- Temporal segmentation
- Temporal merge
- Spatial merge

(2) Orientation estimation

- Character orientation
  - Downward and rightward
  - Two peaks
- Quantization of line direction
  - A text line element is quantized into 4 directions
    - rightward, leftward,
    - upward and downward

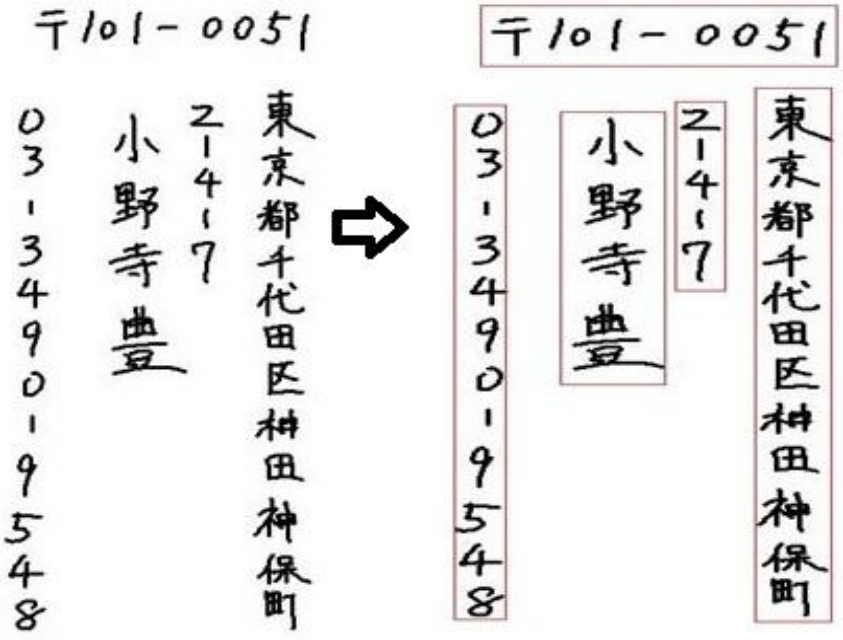


Fig.1 Examples of the segmentation result

Flow of Text Recognition

(1) Text line segmentation

- Block grouping
- Pre-segmentation
- Temporal segmentation
- Temporal merge
- Spatial merge

(2) Orientation estimation

- Character orientation
  - Downward and rightward
  - Two peaks
- Quantization of line direction
  - A text line element is quantized into 4 directions
    - rightward, leftward
    - upward and downward

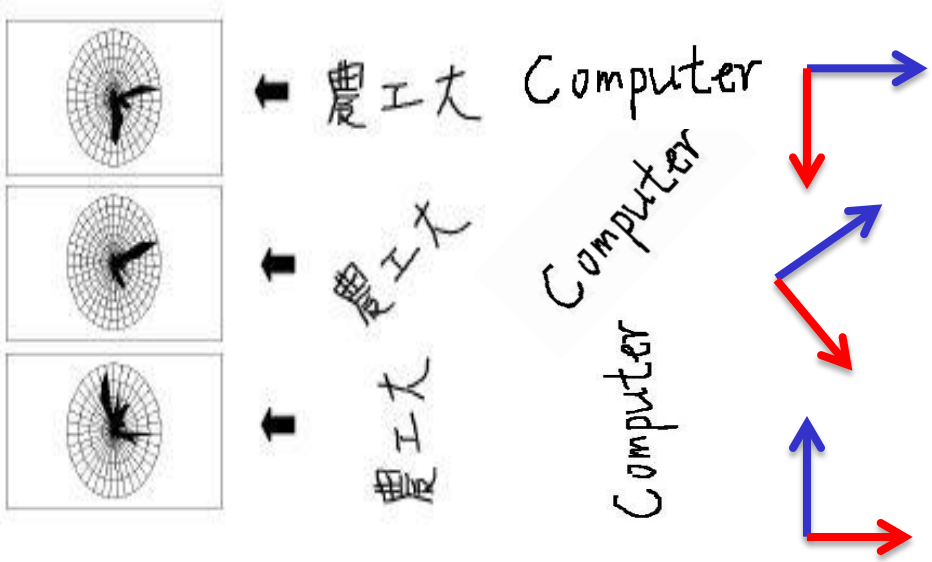


Fig.2 Two peaks in pen movement direction

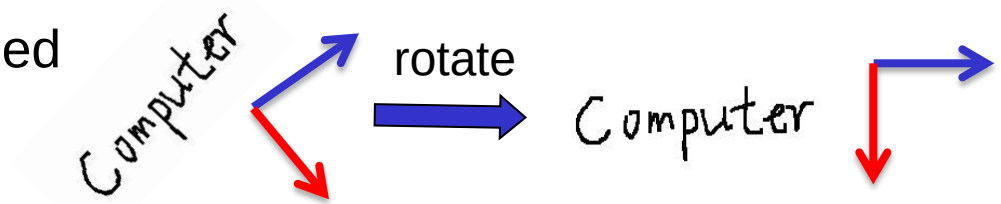


Fig.3 Normalization of character orientation

Flow of Text Recognition

(1) Text line segmentation

- Block grouping
- Pre-segmentation
- Temporal segmentation
- Temporal merge
- Spatial merge

(2) Orientation estimation

- Character orientation
  - Downward and rightward
  - Two peaks
- Quantization of line direction
  - A text line element is quantized into 4 directions
    - rightward, leftward,
    - upward and downward

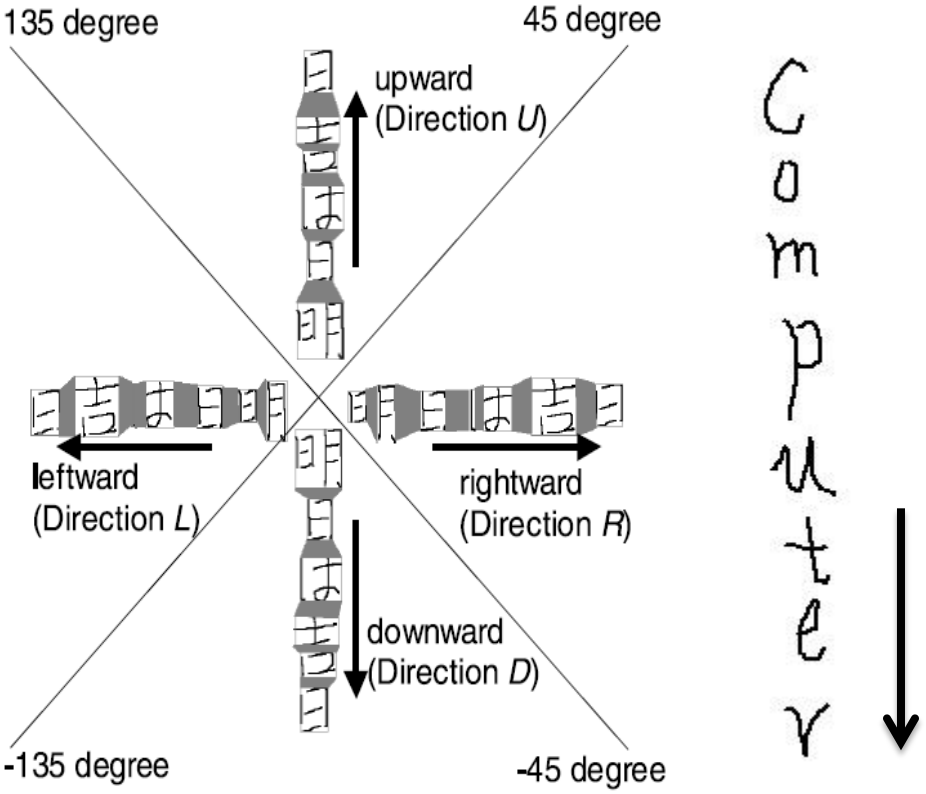
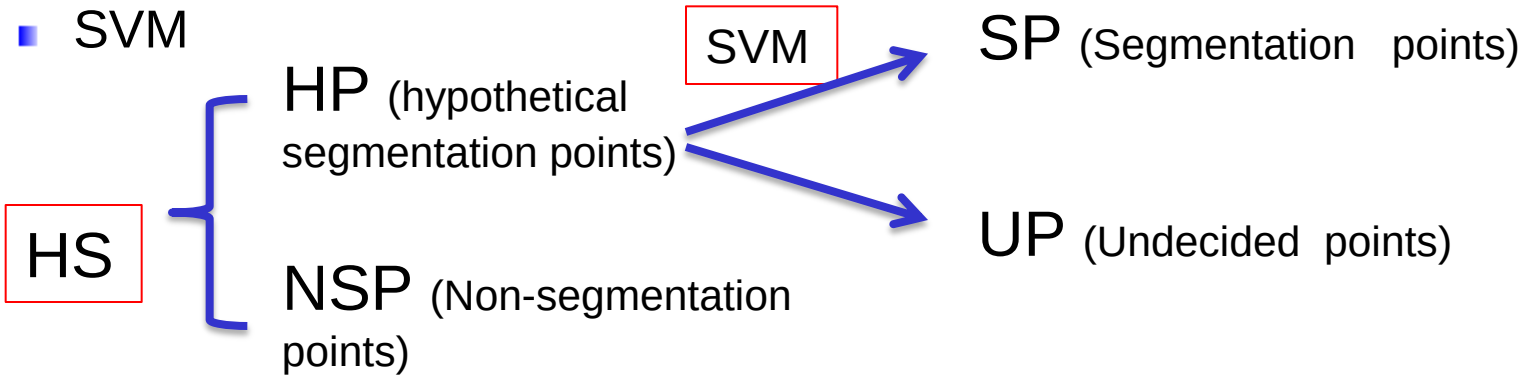


Fig.4 Quantization of line direction

Flow of Text Recognition

(3) Over-segmentation

- Two-stage scheme
  - Hypothetical segmentation (HS)
  - SVM



(4) Candidate lattice construction

- Each arc or node has a score

(5) Recognition and search

- character segmentation
- linguistic context
- character pattern structure
- character recognition
- Viterbi search into the candidate lattice

Flow of Text Recognition

(3) Over-segmentation

- Two-stage scheme
  - Hypothetical segmentation
  - SVM

(4) Candidate lattice construction

- Each arc or node has a score

(5) Recognition and search

- character segmentation
- linguistic context
- character pattern structure
- character recognition
- Viterbi search into the candidate lattice

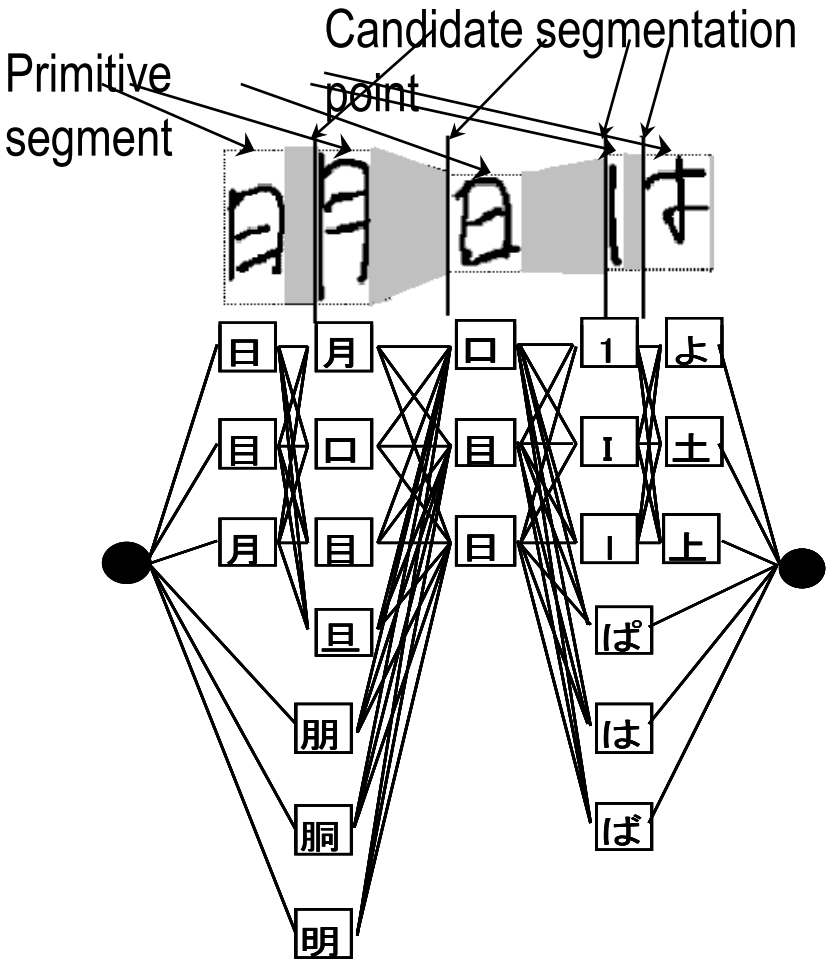


Fig.6 Segmentation and candidate lattice

Recognition Model

Extract Features

- To evaluate the score of string, we extract 6 types of features.
  - Bounding box feature $b_{i.}$
  - Inner gap feature $q_{i.}$
  - Shape feature $z_{i.}$
  - Unary position feature $p^u_{i.}$
  - Binary position feature $p^b_{i.}$
  - Between-segment gap feature g_i

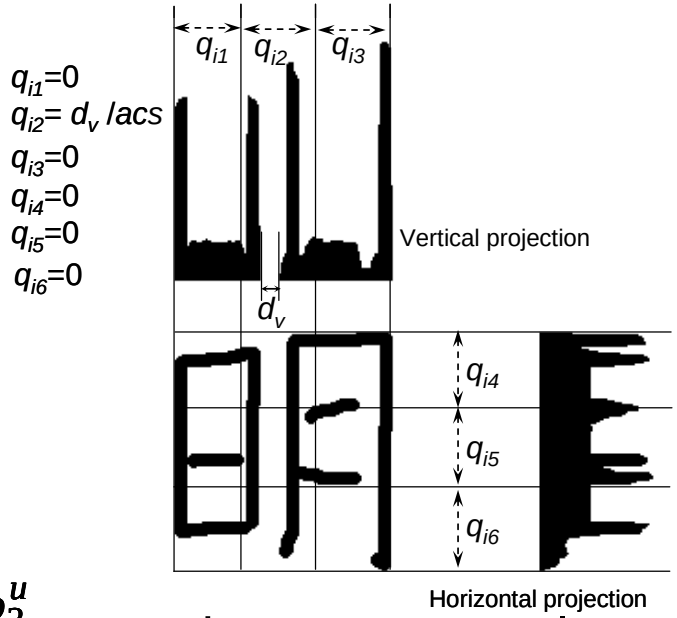


Fig.7 Feature values of character pattern inner gap

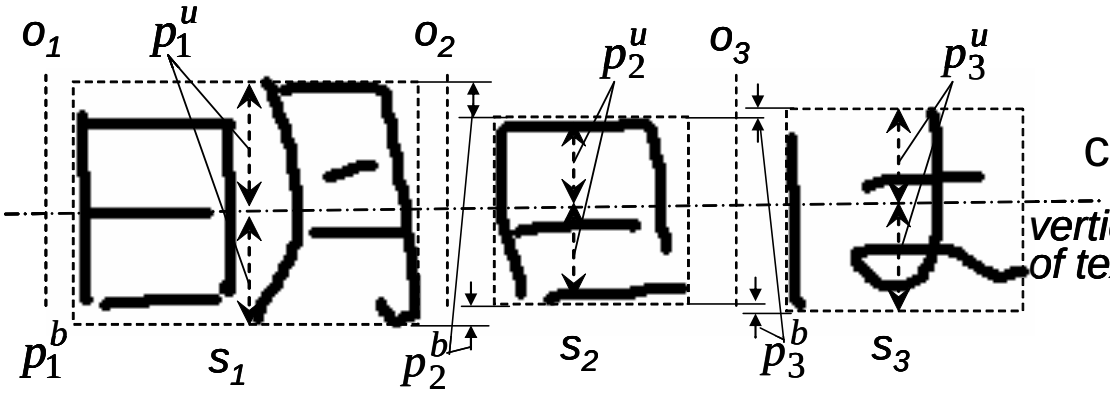


Fig.8 Some geometric features

Evaluation Function

$$f(\mathbf{X},\mathbf{C}) = \sum_{i=1}^{n} \left\{ \begin{array}{l} \sum_{h=1}^{6} \left[\lambda_{h1} + \lambda_{h2}(k_i - 1)\right] \log P_h \\ \lambda_{71} \log P(g_{j_i} \mid SP) + \lambda_{72} \sum_{j=j_i+1}^{j_i+k_i-1} \log P(g_j \mid NSP) \end{array} \right\} + n\lambda$$

- P_h, h=1,…,6, stand for $P(C_i|C_{i-2}C_{i-1})$, $P(b_i|C_i)$, $P(q_i|C_i)$, $P(z_i|C_i)$, $P(p^u{}_i|C_i)$ and $P(p^b{}_i|C_{i-1}C_i)$, respectively.
  - b_i, q_i, z_i, $p^u{}_i$, $p^b{}_i$: extracted features from the candidate patterns and between-candidate patterns compatibilities.
- The weighting parametersλ_{h1}, λ_{h2} (h=1~7) and λ are learnt using a genetic algorithm.
- The optimal path can be found by the Viterbi search.

Demonstration

ル(F) 実行(R) 検索(S) 認識(C) ツール(T) 一筆戻す(Z) WriteDb1 OpenDb1 CreatLatticeFile NextStroke

検索

分割結果表示

Button2

認識結果表示

認識

CreateLattice

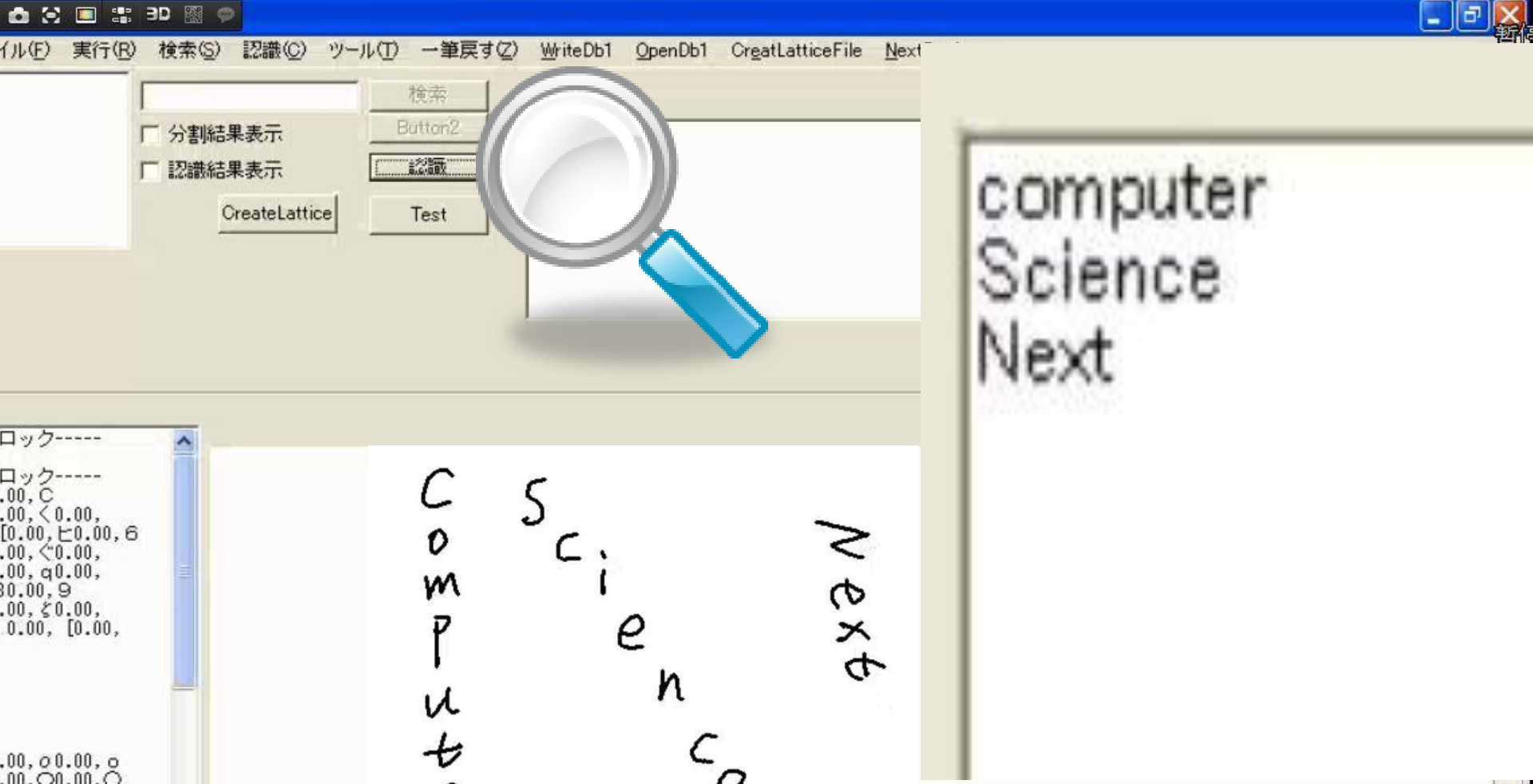
Test

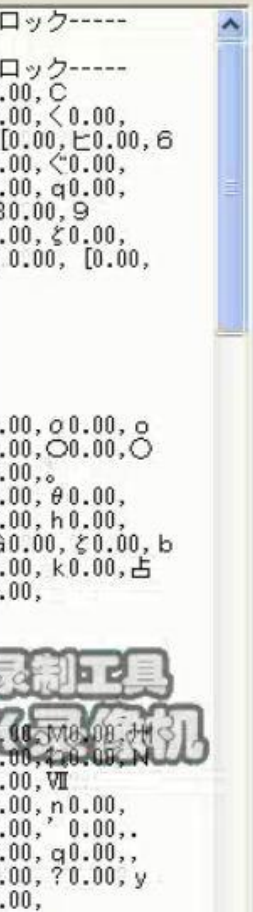
ック-----

ック-----
0,C
0,く0.00,ヒ
0,σ0.00,6
0, [0.00,
0,ε0.00,8
0,9 0.00,々
0,G0.00,t
0,

0,O0.00,O
0,〇0.00,〇
0,
0,◇0.00,
0, ‘0.00,占
0,ど0.00,片
0,反0.00,旨
0,

0,W0.00,れ
0,
0,;0.00,’
0,↑0.00,
.00,『0.00,芹

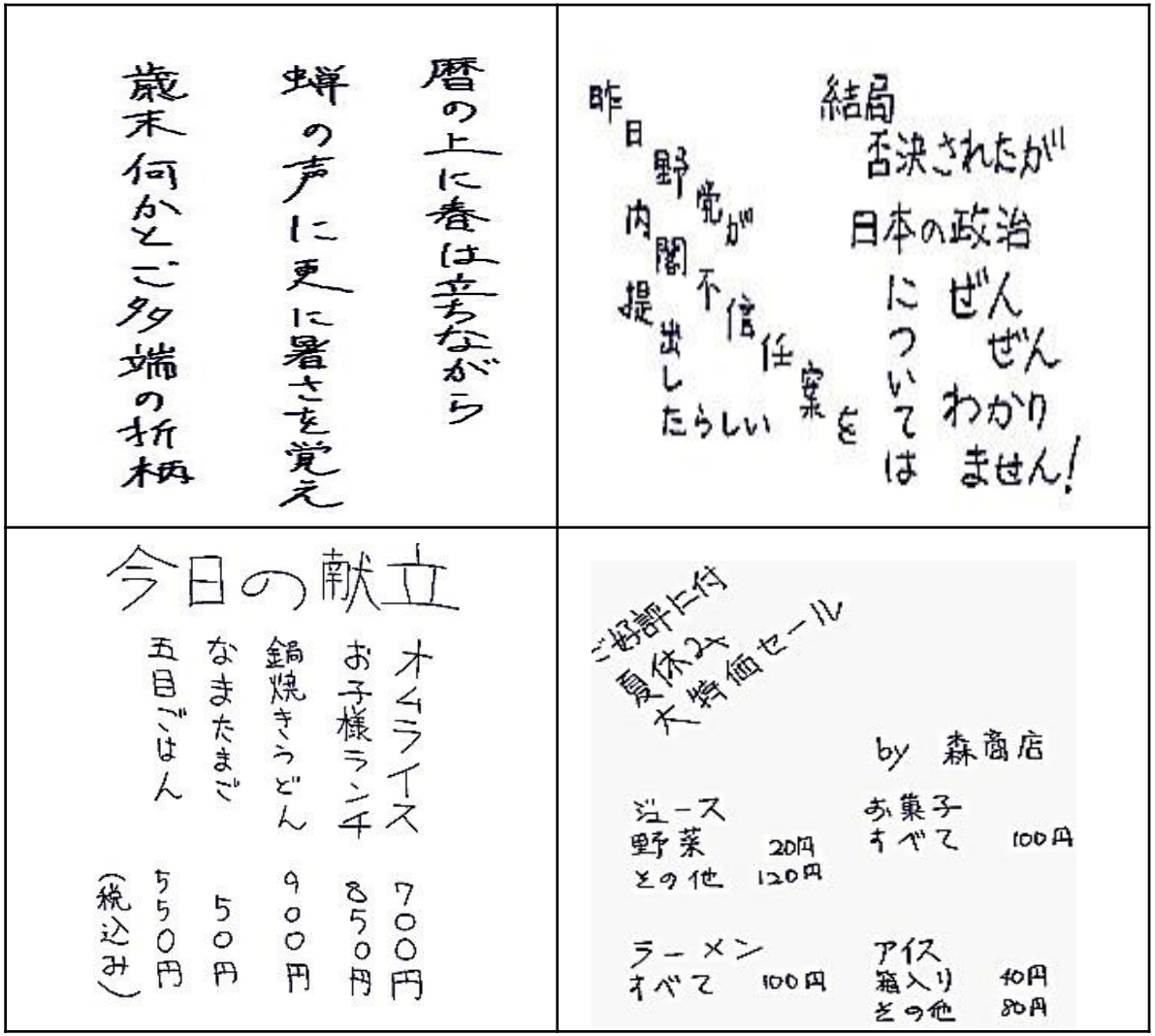

ル(F) 実行(R) 検索(S) 認識(C) ツール(T) 一筆戻す(Z) WriteDb1 OpenDb1 CreatLatticeFile Next
検索
Button2
分割結果表示
認識結果表示
認識
CreateLattice
Test
computer
Science
Next


ロック-----
ロック-----

Computer
Science
Next

Experiments

Examples of Kondate-databse



Examples of HANDS-Kondate_t_bf-2001-11.

Evaluation(1)

- **Compare the performance of the proposed method with our initial attempt in [1]. (whole page recognition rate test)**

Table 2 Improved aspects compared with the initial attempt.

	Initial Attempt in [1]	Method proposed here
Line segmentation	Off-strokes distance and Directional change	Text line grouping method
Over-segmentation	Overlap between the stroke	Two-stage segmentation
Recognition model	Not use the term related to the number of primitive segments	Insensitive to the number of segmented characters

[1]: Motoki Onuma, Akihito Kitadai, Bilan Zhu and Masaki Nakagawa: An On-line Handwritten Japanese Text Recognition System Free from Line Direction and Character Orientation Constraints, IEICE Trans. Inf&Syst., Vol.E88-D, No.8, pp.1823-1830, August 2005

Result

Table 3 Recognition rate on mixture of vertical, horizontal and skewed text lines.

Page No.	Character recognition rate	
	Initial attempt in [1]	Method proposed here
4	90.90%	98.48%
16	68.32%	93.60%
18	60.51%	91.14%
24	62.66%	87.89%
27	62.10%	83.15%

[1]: Motoki Onuma, Akihito Kitadai, Bilan Zhu and Masaki Nakagawa: An On-line Handwritten Japanese Text Recognition System Free from Line Direction and Character Orientation Constraints, IEICE Trans. Inf&Syst., Vol.E88-D, No.8, pp.1823-1830, August 2005

Evaluation(2)

- **Compare the performance of the proposed method with our second attempt in [8]. (text lines with 4 directions' test)**

Table 4 Improved aspects compared with the second attempt.

	Second attempt in [8]	Method proposed here
Over-segmentation	One-stage segmentation	Two-stage segmentation
Recognition model	Not use the term related to the number of primitive segments	Insensitive to the number of segmented characters

[8]:Zhu B. and Nakagawa M., Segmentation of on-line freely written Japanese text using SVM for improving text recognition, IEICE Trans. Inf. & Sys., E91-D(1), 105-113 (2008).

Result

Table 5 Results of the 4 directional testing sets compared with the second attempt.

Line Direction	Method			
	Second attempt in [8]		Method proposed here	
	f	*Cr*	*f*	*Cr*
Rightward	0.9660	73.61%	0.9856	92.15%
Leftward	0.9838	80.75%	0.9711	89.24%
Downward	0.9897	80.23%	0.9872	90.78%
Upward	0.9647	75.83%	0.9712	90.25%

f : segmentation measure, *Cr*: character recognition rate.

[8]:Zhu B. and Nakagawa M., “Segmentation of on-line freely written Japanese text using SVM for improving text recognition,” IEICE Trans. Inf. & Sys., E91-D(1), 105-113 (2008).

Conclusion

- We presented significant improvement in line-direction-free and character-orientation-free on-line handwritten Japanese text recognition.
- Recognition rate of vertical and skewed text is now almost comparable with horizontal text.
- Recognition speed is quick enough for practical use.

Finish

Thank you for your attention.

Preparation

- Trained the character recognition engine and the four quadratic discriminant function (QDF) classifiers on $P(b_i|C_i)$, $P(q_i|C_i)$, $P(p^u{}_i|C_i)$ and $P(p^b{}_i|C_i)$ using the database: Nakayosi.
- Prepared the tri-gram table from the ASAHI newspaper and the NIKKEI newspaper.
- Applied the method to the database HANDS-Kondate_t_bf-2001-11collected from 100 persons which stores on-line text of various character orientations and line directions.
  - 75 persons' text lines for learning candidate segmentation point probability and the weighting parameters.
  - The remaining 25 persons' text lines for evaluating the performance.

Statistics of training/testing text lines

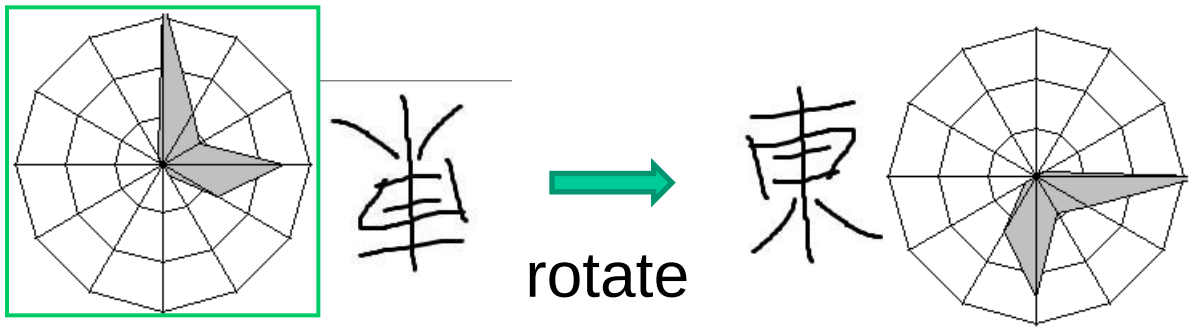
Table 1 Training and testing sets for each direction.

Category	Line Direction							
	Rightward		***Leftward***		***Downward***		***Upward***	
Patterns	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Text lines	10518	3623	130	36	6836	2032	72	17
English letters	4123	1868	6	0	2993	1231	4	0
Numerals	17147	5618	52	7	7782	2569	4	1
Kanas	40502	13927	573	178	31851	10352	262	45
Chinese characters	37495	12986	331	114	26680	8730	143	36
Other characters	11994	3614	67	10	7636	2258	27	6
Nsp	100743	107151	899	273	70106	23108	368	71
Nnsp	300753	34390	2874	1079	214508	72306	11235	260
Nac	11	11	8	9	11	12	6	5
Nal	1483.48	1520.52	13.72	12.36	1025.89	1005.6	5.87	3.52

Nsp is the number of true segmentation points, *Nnsp* is that of true non-segmentation points, *Nac* is the average number of characters in a text line and *Nal* is the average number of characters.

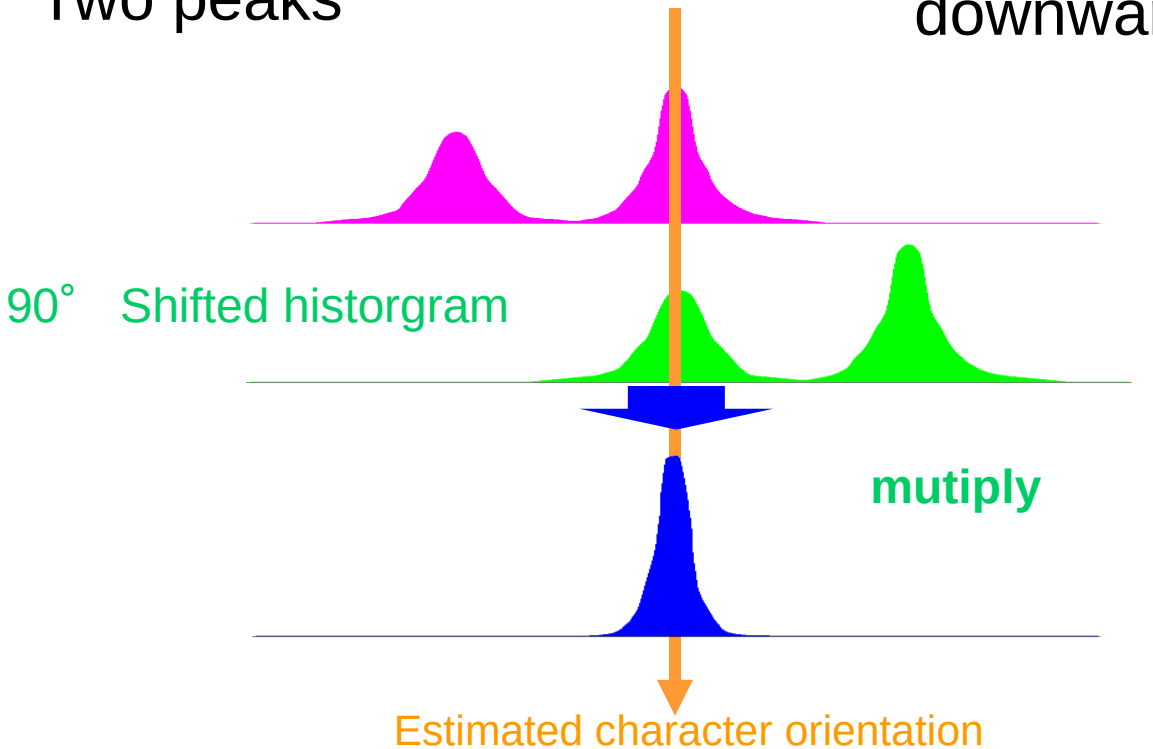
complementary

Character orientation



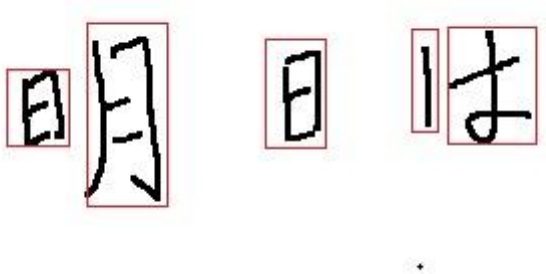
Two peaks

downward

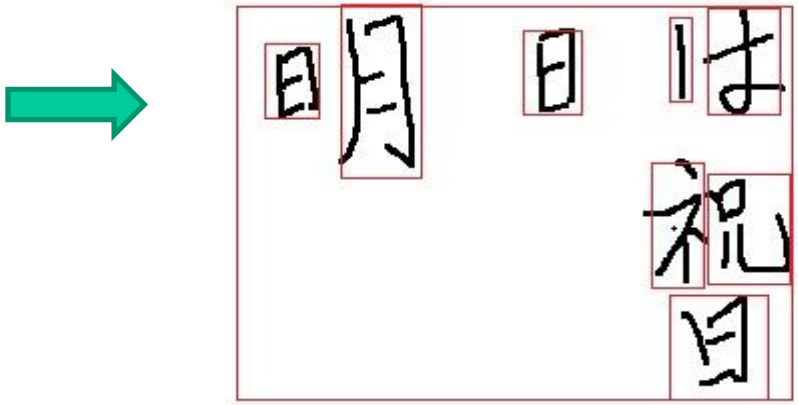


complementary

Block-grouping



pre-segmentation



temporal-segmentation