

Segmentation and Recognition of Touching Chinese Characters

Liang Xu (PhD student--3rd year) Advisor: Dr. Cheng-Lin Liu

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences

Introduction

•Touching characters are often met in offline Chinese handwriting. Recognition of touching characters is very difficult and has not been solved very well.

•We apply an over-segmentation strategy to attack this problem.

•There are mainly two types of touching between two characters: single-connected pattern and double-connected pattern. Such as Fig.1 and Fig.2.

Touching Pattern Separation

Finished work: Single-connected pattern separation. Visibility-based Foreground analysis.

i.e., Contour, Skeleton analysis. Such as Fig.3.







Fig.3: (From left to right) Character pair with a touching pattern, outer contour, skeleton, and separation points.





Fig.1: (From left to right) Single-connected pattern, its separation of touching portion, and corresponding text.







Fig.2: (From left to right) Double-connected pattern, its separation of touching portion, and corresponding text.

Issues

•Correct *separation* of the touching pattern.

Traditional methods are not fully using the background and foreground information.

Recognition of separated patterns.

Traditional geometric context doesn't consider the confidence of separation point.

Ongoing work:

 Foreground + Background-based separation of singleconnected pattern.

Utilize background information, e.g. skeleton.

Separation of double-connected pattern.

Learning-based Separation Points Filtering. Such as Fig.4. Structure features of separation point and its neighborhood. Support vector machine (SVM).







Fig.4: Separation points filtering.

Touching Characters Recognition

Flowchart



Ongoing work:

Recognition of separated patterns. Candidate pattern graph, such as Fig.5. Geometric context: Confidence of separation point.



Fig.5: Candidate pattern graph, corresponding to Fig.4(b).

References

[1] Liang Xu, Fei Yin, Qiu-Feng Wang, Cheng-Lin Liu. Touching Character Separation in Chinese Handwriting Using Visibility-Based Foreground Analysis, ICDAR2011, accepted.

[2] Liang Xu, Fei Yin, Cheng-Lin Liu. Touching Character Splitting of Chinese Handwriting Using Contour Analysis and DTW, Proc. 2010 Chinese Conference on Pattern Recognition (CCPR), Chongqing, China, Oct. 2010, pp.814-818.

[3] Qiu-Feng Wang, Fei Yin, Cheng-Lin Liu. Integrating Language Model in Handwriting Chinese Text Recognition, Proc. 10th ICDAR, pp.1036-1040, 2009.

[4] Fei Yin, Qiu-Feng Wang, Cheng-Lin Liu. Integrating Geometric Context for Text Alignment of Handwritten Chinese Documents, Proc. 12th ICFHR, 2010.