MST在手写汉字切分中的应用

韩 勇, 须 德, 戴国忠

韩 勇1, 须 德1, 戴国忠2
1(北京交通大学 计算机与信息技术学院,北京 100044)
2(中国科学院 软件研究所 人机交互技术与智能信息处理实验室,北京 100080)作者简介: 胡文蕙(1977一术.
联系人: 韩 勇 Phn: +86-10-62645407, E-mail: hanyong_sc@hotmail.com, http://www.njtu.edu.cn

Abstract
Handwritten Chinese characters segmentation is to process strokes based on its spatial relations to form character elements for recognition. This paper introduces a method to segment Chinese characters according to the topological relations of Chinese component and minimal span tree of strokes. The experiment shows that this new method can achieve good performance. The accuracy of segmentation is over 91.6%.

Han Y, Xu D, Dai GZ. Using MST in handwritten Chinese characters segmentation. *Journal of Software*, 2006,17(3):403-409.
DOI: 10.1360/jos170403
http://www.jos.org.cn/1000-9825/17/403.htm

摘要
手写汉字切分是根据输入笔迹的空间位置关系进行汉字部件的合并切分,形成完整的汉字笔划以便进行识别处理.综合利用了汉字部件的结构位置关系和笔划的空间位置关系,根据笔划的最小生成树(minimal spanning tree,简称MST)对联机连续手写输入汉字进行切分,取得了较好的切分结果.切分的准确率超过91.6%.

References:

[1] Wu YS. Chinese Character Recognition-Principle, Method and Realization. Beijing: Higher Education Press,1992 (in Chinese).

[2] Breuel TM. Representation and metrics for off-line handwriting segmentation. In: Proc. of the 8th Int'l Workshop on Frontiers in Handwriting Recognition. 2002. 821-826. http://citeseer.ist.psu.edu/breuel02representations.html

[3] Han BX. Combination of Chinese character constituents-A latent structural unit. Journal of Chinese Information Processing, 1995,9(3):27-32 (in Chinese with English abstract).

[4] Lu Y. Segmentation of Free-format Handwritten Chinese Characters Based on Structure Features of Characters. Acta Electronica Sinica, 2000,128(5):102-104 (in Chinese with English abstract).

[5] Casey RG. A survey of methods and strategies in character segmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1996,18(7):690-706.

[6] Breuel TM. Segmentation of handprinted letter strings using a dynamic programming algorithm. In: Proc. of the 6th Int'l Conf. on Document Analysis and Recognition. 2001. 821-826. http://citeseer.ist.psu.edu/breuel01segmentation.html

[7] Punnoose J. An improved segmentation module for identification of handwritten numerals [MS. Thesis]. Massachusetts Institute of Technology, 1999.

[8] Verma B, Blumenstein M. Recent achievements in off-line handwriting recognition systems. In: Proc. of the Int'l Conf. on Computational Intelligence and Multimedia Applications. 1998. 27-33. http://citeseer.ist.psu.edu/130825.html

[9] Chen H. Segmentation and recognition of continuous handwriting Chinese text. In: Proc. of the Int'l Conf. on Computer Processing of Oriental Languages. 1997. 630-633. http://citeseer.ist.psu.edu/hong97segmentation.html

[10] Nicholas E. Recognition of handwritten mathematical expressions [MS. Thesis]. Massachusetts Institute of Technology, 1999.

[11] Zhang XW. Adaptive Character Extraction from Continuous HandWriting Chinese Text Based on Multilevel Constrains. 2003 (in Chinese with English abstract). http://iel.iscas.ac.cn/lab_achive/lab_achive_publications_2003.htm

附中文参考文献:
[1] 吴佑寿.汉字识别——原理、方法与实现.北京:高等教育出版社,1992.

[3] 韩布新.部件组合——潜在的汉字结构层次.中文信息学报,1995,9(3):27-32.

[4] 吕岳.基于汉字结构特征的自由格式手写体汉字切分.电子学报,2000,128(5):102-104.

[11] 张习文.基于多层次信息的连续手写中文的自适应分割方法.2003. http://iel.iscas.ac.cn/lab_achive/lab_achive_publications_2003.htm