

一种 stroke 滤波器文字分割算法

石振刚^{1,2} 高立群²

(沈阳理工大学信息科学与工程学院 沈阳 110168)¹ (东北大学信息学院 沈阳 110004)²

摘 要 为解决复杂背景中准确地进行文字分割的问题,提出了一种应用 stroke 滤波器进行文本分割的新方法。首先进行 stroke 滤波器的合理设计,并应用所设计的 stroke 滤波器来判别文本的彩色极性,得到初次分割的二值图。然后进行基于区域生长的文字分割。最后,应用 OCR(optical character recognition)模块提高文本分割的整体性能。将提出的算法与其他算法进行了比较,结果表明,所提算法更为有效。

关键词 文本分割,复杂背景,OCR,stroke 滤波器

中图法分类号 TP391.43 **文献标识码** A

New Algorithm for Text Segmentation Based on Stroke Filter

SHI Zhen-gang^{1,2} GAO Li-qun²

(College of Information Science and Engineering, Shenyang Ligong University, Shenyang 110168, China)¹

(College of Information Science and Engineering, Northeastern University, Shenyang 110004, China)²

Abstract In order to solve segment text accurately and robustly from a complex background, this paper proposed a new algorithm for text segmentation in images by using a stroke filter. First, we described the stroke filter briefly based on local region analysis. Second, the determination of text color polarity and local region growing procedures were performed based on the response of the stroke filter. Finally, the feedback procedure by the recognition score from an optical character recognition (OCR) module was used to improve the performance of text segmentation. The proposed algorithm was compared with other algorithms. The experimental results demonstrate that the proposed algorithm obtains satisfactory results.

Keywords Text segmentation, Complex background, Optical character recognition(OCR), Stroke filter

文本分割是指在一个书面文档或视频序列中识别具有独立意义的单元之间的边界,其分割对象可以是视频流、网络动态数据或者书面静态文本。这种预处理技术在很多领域比如信息提取、文摘生成、文本解析以及文本导航等都有着极为重要的应用。目前,在简单背景下进行文本分割已经是一项较为成熟的技术,而如何在较为复杂的背景中进行文本分割还没有一个成熟的方法。因此,在复杂背景中准确地进行文本分割就成为当前文本分割中的热点问题^[1,2]。

1 文本分割技术

文本分割主要有两种方法。第一种方法主要是根据分割目标与背景的差异进行分割,例如全局和局部的阈值分割方法及 Otsu 自动阈值分割方法等。通常,这类方法较为简单,算法运行速度快。但当分割目标与背景较为接近时,这类分割方法往往不能够进行准确的分割。第二种方法主要是根据像素的相似性进行分割,例如使用分裂合并的方法进行分割,使用边缘检测的方法进行分割及使用分水岭的方法进行分割等^[3]。但这些方法由于在进行文本分割时对待分割文本的形状作了一些人为的规定,因此在实际应用中都具有不稳定性。

上述两种文本分割方法存在的主要问题是对待分割文本的色彩、大小、字体以及复杂背景较为敏感。针对这些问题,本文在充分考虑到文本固有特性的基础上,提出了一种应用 stroke 滤波器进行文本分割的新方法。本文第 2 节介绍 stroke 滤波器的设计;第 3 节详述本文的分割算法;第 4 节给出测试手段及实验结果,并就实验结果进行讨论;最后作出总结。

2 stroke 滤波器设计

stroke 定义为用于文本分割的一段直线或一段弧,而图像中的文本是由一个或几个 stroke 所组成的。stroke 滤波器是基于局域分析法进行设计的,我们所设计的 stroke 滤波器如图 1 所示。

根据图 1 所示 stroke 滤波器来计算源图像中像素点滤波后的值。设源图像中任一点 (x, y) 为滤波器的中心点,其周围有 3 个矩形区域。图 1 中所示区域 1 表示中心区域,区域 2 和区域 3 为中心区域的两个长边侧面区域。这 3 个局部区域的方向和大小是由参数 α 和 d 决定的,其中 d 是由所要分割文本的先验知识所决定的。其他参数有如下关系: $d_1 = d_2 = d/2, w = 2d$ 。

些字符所属字体为:

$$w_k = \arg \max_{r=1, \dots, n} (p(w_r | \theta_{h_1 i_1}, \theta_{h_2 i_2}, \dots, \theta_{h_m i_m})) \quad (14)$$

通过增加用于字体识别的字符数目 m , 可以提高字体识别可信度。只有识别可信度高于预先定义的期望可信度时, 才返回字体识别结果。

4 实验与分析

本节将对本文所提算法的性能进行仿真实验和分析。实验主要分为两部分: 第一部分用本文算法对中文和英文文本进行分割, 图 2 显示了分割结果。从图中可以看到, 对于不同语言的文本, 本文算法都有较好的分割效果。这表明本文算法对于不同语言具有很好的适应性。



图 2 中英文文本分割结果

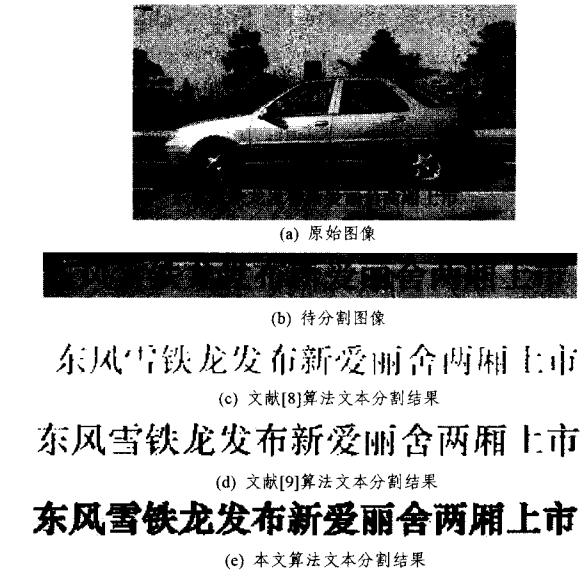


图 3 不同算法文本分割结果比较

在实验的第二部分, 对本文所提算法与文献[8,9]所提算法的分割效果从主观和客观上进行了比较。主观上, 图 3 显示了各种算法的分割结果, 从中可以看到本文的分割效果是最好的。客观上, 选取分割准确率(计算公式见式(15))与算法执行速度这两个指标进行比照, 结果如表 1 所列。从表 1 中可以看出, 本文所提算法的分割准确率均高于所对比的 3

种文本分割算法。但从执行速度上来看, 本文算法较文献[8]的 Otsu 方法慢, 与文献[9]的 Song 方法相当。主要原因在于本文算法为提高文本分割的准确率执行了将 OCR 模块作为最后修正分割结果的附加模块, 这也是本文算法在今后所要改进和提高的地方。

$$\text{分割准确率} = \frac{\text{文本中正确分割的字符数}}{\text{文本中的字符总数}} \quad (15)$$

表 1 不同算法文本分割准确率与执行速度比较

方法	分割准确率	算法执行时间
文献[8]算法	86.3%	9.8ms
文献[9]算法	90.5%	13.3ms
本文算法	96.8%	13.7ms

结束语 本文提出了一种应用 stroke 滤波器进行文本分割的新方法。该算法通过合理地构造 stroke 滤波器, 使其在进行文本分割时能够更好地反映出待分割文本的本质特征。为消除 stroke 滤波器分割文本时所产生的过分割现象, 本文算法对通过 stroke 滤波器所分割的文本执行了局部区域成长程序。最后, 通过执行 OCR 附加修正模块, 使得复杂背景下文本的分割准确率达到了 97% 左右。

但是, 由于本文所提算法具有较高的计算复杂度, 因此算法执行速度较其他常规方法略慢。为解决此问题, 下一步将深入研究在保证文本分割准确率的前提下, 改进算法有关模块, 以降低算法计算复杂度。另外, 在未来的工作中, 将考虑如何把基于 stroke 滤波器的文本分割技术应用于文档摘要及信息检索等系统中。

参考文献

[1] Dimitrova N, Zhang H J, Shahraray B, et al. Applications of video content analysis and retrieval[J]. IEEE Multimedia, 2002, 9 (3): 43-55

[2] Wang Y, Liu Y, Huang J C. Multimedia content analysis using both audio and visual clues[J]. IEEE Signal Process, 2000, 17 (6): 12-36

[3] Jung K, Kim K I, Jain A K. Text information extraction in images and video: a survey[J]. Pattern Recognition, 2004, 37 (5): 977-997

[4] Tseng Y H, Lee H J. Recognition - based handwritten chinese character segmentation using a probabilistic Viterbi algorithm [J]. Pattern Recognition, 1999, 20(9): 791-806

[5] 陈东, 刘希玉. 基于支持向量基的条码分类研究[J]. 山东师范大学学报: 自然科学版, 2007, 22(4): 24-26

[6] 钟将, 温罗生, 冯永, 等. 基于近似支持向量机的 Web 文本分类研究[J]. 计算机科学, 2008, 35(3): 167-169

[7] Tao D, Li X, Wu X, et al. General tensor discriminant analysis and Gabor features for gait recognition[J]. IEEE Trans. Pattern Analysis, 2007, 29(10): 1700-1715

[8] Otsu N. A threshold selection method from gray-scale histogram [J]. IEEE Trans. Syst. Man Cybern, 1979, 9: 62-66

[9] Song J, Cai M, Lyu M R. A robust statistic method for classifying color polarity of video text[C]//Proceedings of the IEEE International Conference on Acoustics, 2003: 581-584