doi:10.3969/j.issn.1672 - 8513.2010.03.001

● 特约稿件 ●

室外场景中标识牌文字的检测与提取技术综述

贾文静[1],曾　超[1],敖永霞[2],何祥健[1],吴　强[1]
(1.悉尼科技大学 工程与信息技术学院,澳大利亚 悉尼 2007;2.福建农林大学 计算机与信息学院, 福建 福州 350002)

摘要:室外场景中标识牌文字的检测与提取在机器视觉、辅助驾驶员系统、眼障人士辅助系统、游客帮助系统等中有着广泛的应用.近年来,针对不同应用,研究人员研制开发出许多室外场景中文字信息提取的系统与方法.对2002年以来发表在主要英文刊物和会议论文集中的方法进行综述.提出了一个分层次的系统框架结构,并按照这一结构对文献中出现的实现各主要模块的比较有代表性的方法进行了归纳和比较,旨在对该领域的研究技术水平提供一个综述,并提出尚待解决的技术问题.

关键词:文字信息提取;室外场景;综述;层次框架结构
中图分类号:TP391　　**文献标识码**:A　　**文章编号**:1672 - 8513(2010)03 - 0157 - 05

Automatic Detection and Extraction of Sign Text from Outdoor Scenes: A Contemporary Review

JIA Wen-jing[1], ZENG Chao[1], AO Yong-xia[2], HE Xiang-jian[1], WU Qiang[1]
(1. Faculty of Engineering and Information Technology, University of Technology, Sydney 2007, Australia;
2. School of Computer and Information, Fujian Agriculture and Forestry University, Fuzhou 350002, China)

Abstract: Automatically detecting and extracting sign texts from outdoor scenes has found many applications in the robot vision, driver assistant system, visually impaired assistant system, etc. In recent years, many systems and methods have been developed for sign text information extraction from outdoor scenes. This paper reviews the key techniques published in major international journals and conference proceedings since 2002. A hierarchical framework is proposed, and methods in these literatures to implement each module of the system model are reviewed and assessed. This paper aims to provide a contemporary review on the state of the art techniques on this topic and discusses the unsolved problems.

Key words: text information extraction; outdoor scene; survey; hierarchical framework

室外场景中有各种各样的标识,它们提供给人们日常生活不可或缺的信息.在这些标识中,信息量最大的一类是含有文字信息的各种标识,如含有文字的各种道路、交通、公共场所标识及商业标识,它们指示或警告人们所处周边环境的情况.

近年来,随着低价、高性能便携式数字成像设备的广泛普及,用计算机技术自动地读取这些标识中的文字信息并以一种更方便接受的形式反馈给使用者,吸引了越来越多的人们的兴趣.

自动地获取室外场景各种标识中的文字信息可以应用在很多方面,特别是当配备了其他软件,如多语言翻译软件和语音合成软件时.它可以用于各种基于视频图像信号的智能辅助系统,如智能驾驶员辅助系统[1-2]、眼障人士辅助系统[3-4]和游客辅助系统[5]等等.该系统还可以应用于其他许多需要实时读取文字标识的各种应用中.其典型的应用场景

收稿日期:2009 - 09 - 26.
基金项目:UTS ECRG 项目.
作者简介:贾文静(1976 -),女,博士,校长博士后研究员.主要研究方向:计算机视觉与图像模式识别.

是一个车载或者手持相机在移动的过程中拍摄其前方及侧前方的视频或者图像,由该系统软件对输入的每一帧图像进行分析处理,自动地提取出图像中的文字区域并用 OCR 技术将分割出的文字识别出来,并按某种优先级顺序将识别出的文字信息反馈给用户以帮助他们了解自己周边的情况.

受其广泛的应用潜力的吸引,人们对室外场景中文字信息的提取做了大量的科研工作,特别是近几年来,许多新技术被应用于或者开发以在更为复杂的背景下更为精确、快速地检测和提取文字. 本文对 2002 年以来发表在主要英文刊物和会议论文集上的室外场景中文字信息的提取方法进行综述,旨在给研究人员,特别是新进入该领域的研究人员,提供一个该领域研究最新技术水平的综合性的参考.

现有的大多数文字提取方法都可以归结为产生文字候选区域和对候选区域进行分类 2 个核心步骤,有些系统另外采用了预处理和后处理以进一步提高文字提取的精度和系统的鲁棒性. 因此,不同于现有的一些分类方法,如文献[6]中按所使用的特征进行归类的方法,本文所使用的分类方法是面向问题而不是面向方法的,并提出了一种分层次的系统模型,分别讨论实现系统的每一模块中现有的比较有代表意义的技术和方法,并对这些方法的整体性能进行了比较.

因此,不同于此前的综述文章[7],本文是根据文字提取的框架结构以一种分层次的方式组织的. 这将更有助于那些刚刚进入这一领域的研究人员了解该领域的技术发展水平并能够理解实现这一系统的主要模块和它们之间的关联.

按如图 1 所示的系统模型,本文的结构如下. 第 1 部分主要讨论了将输入图像分解成候选区域的主要方法. 第 2 部分讨论了对文字候选区域进行分类的各种方法,并在第 3 部分中对它们的整体性能进行了比较. 最后,第 4 部分对室外场景中文字信息提取的现有技术难题进行了归纳,并建议将来的研究方向.

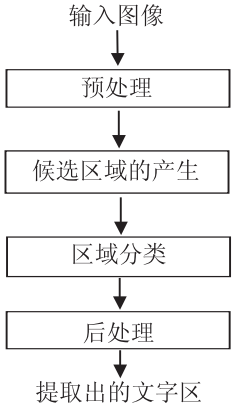

图1 分层次系统构架

1 文字候选区域的产生

文字候选区域的产生,也称为候选区域选择[8],是将输入图像分解成一系列子图像区域,以便区域分类模块对它们进行分类,相应地判决为文字区域或者非文字区域. 该步骤的实现效率和准确度对整个系统的性能关系重大. 在个别文献中提到,候选区域的产生可以是"自上而下"的也可以是"自下而上"的[9]. 在前者中,候选区域是根据图像的属性通过对图像进行分割得到的;而在后者,候选区域是由文字的某种或者某些属性由连接像素点产生的. 如在文献[9]中的自下而上的方法,候选区是用文字区的属性来定位的. 此外,值得一提的是,如后文所述,对于那些用事先训练好的分类器直接对扫描窗口所覆盖的每一个子图像区域进行分类的方法,候选区域的产生实际上是由机器对图像进行穷尽式扫描得到的,没有经过任何筛选. 图像形态学操作、图像分割技术和聚类分析技术都曾被用于从图像像素点到候选区域的生成. 对于自上而下的各种方法,所使用的文字特征主要有边界、纹理、颜色和笔划.

1.1 基于边界特征的方法

这类方法首先获得图像的二值边界图,而后经过图像形态学的膨胀和腐蚀操作,由该二值图像生成若干个连通区域,而后根据文字区的各种先验知识对所获得的连通区域进一步分析以获得最终的候选区域. 采用类似方法的文献有[8,10-11],经典的边界检测算子如 Canny 算子和高斯差分函数(DoG)等都曾见于文献中. 如在文献[8]的多分辨率文字检测工作中,由包含连续边界的最小矩形所定义的图像区域的纹理值、颜色分布和区域对比度被计算出来,根据文字区就这些特征取值的先验知识对上述区域进行过滤与合并,最终得到候选区域;在文献[10]中,每个连通区域又经过水平和垂直投影分析来进一步筛选. 在文献[11]中,通过像素级和区域级的分析,每一个连通区域被分别标记为文字区或者非文字区,该标记过程的核心是使用了用 K-SVD 算法训练出的超完备文字和字符库及对各个区域进行稀疏度测试.

边界特征在文献[12]中被进一步地复合成更为复杂的特征"杆"和"框"并作为文字区域所特有的特征标记. 并使用图形模型来描述文字与非文字的分割问题,从而进行文字区域检测.

1.2 基于纹理特征的方法

基于纹理特征的方法和基于边界特征的方法的

思路很相似,不同之处主要在于此类方法首先计算出图像的纹理特征(如像素灰度值的变化),而不单单是边界. 根据观察,由于文字区域中存在着颜色或者灰度强度值的突变,文字区域通常都有着特殊的纹理特征,同时其灰度或者颜色变化也比较大. 因此在文献[13]中,输入图像首先被分成若干个 8 ×8 像素的图像小块,根据字符属于具有某些特定纹理的通用属性的区域,使用局部的色调和饱和度的直方图信息和纹理特征和区别性训练的条件最大熵模型将上述图像块进行分类,以定位候选区域.

不同于一般的基于纹理特征的方法,在文献[14]中使用了由文字笔划产生的纹理特征而不是直接对灰度值或者颜色值进行纹理分析. 该算法首先对输入灰度图像进行 Haar 小波多尺度分解,并将在 3 个小波子带图像中的像素点逐个标记为属于背景区、过渡区和笔划区,而后在小波域使用 8 ×16 像素的滑动窗口并计算窗口内区域在 3 个小波频段的共生矩阵以描述文字笔划产生的纹理特征,并据此生成 1 个二值过滤图对图像进行二值化和标记.

文献[15]中提出的方法是 2 种基于图像灰度值分析的启发式算法的混合. 在第 1 种算法中,输入的图像首先用图像像素灰度值的中值进行二值化操作,并随后提取出连通区域及其包含相应区域的最小矩形. 随后,根据各连通矩形区域的尺寸、位置和长宽比去除不满足先验知识的区域,如长线条和小的噪点. 最后相邻的区域进行合并生成了文字候选区域. 在第 2 个基于分裂 - 合并的算法中,输入图像首先按一定的准则从整张图像开始进行分裂直到分裂后生成的区域的最大、最小灰度值之差小于某个阈值. 接下来,相邻的 2 个或者多个区域又进行合并,直到合并后的新区域的最大、最小灰度值之差大于上述阈值. 对最后得到的区域进行二值化和形态学腐蚀操作,生成又一文字候选区域. 最终的定位结果为上述 2 种算法的结果的组合.

1.3　根据颜色信息

除了基于灰度值的边界和纹理特征,颜色信息也被越来越多的人所使用,这是基于大部分路牌中文字和背景的颜色是均匀分布的这一假设进行的. 如在文献[16]中,使用了广义学习矢量量化算法将在 LUV 空间中具有相似颜色的像素点归成一组从而实现对图像的分割,继而对各个分割出的区域的空间分布进行分析筛选从而得到可能包含文字的候选区域. 在文献[17]中,输入图像在经过一个对称邻域滤波器的保边平滑后,使用一个分层次的连通元算法同时考虑像素级的连接和区域级的连接将相似的像素合标记成属于同一区域. 在文献[18]中,首先分别在图像的色调和饱和度分量上使用了基于聚类分析的图像分割,而后将每个候选文字区域的尺寸归一化成 64 ×64 像素,并从归一化后的区域中提取小波特征矢量输入给神经网络进行分类.

在文献[19]中,这种颜色信息被以另外一种形式加以使用. 根据观察发现,由于“渗色”效应的存在,文字和其相邻背景之间存在着颜色的过渡,即位于文字边缘的像素的强度值按对数规律变化. 从而对图像进行分析并生成颜色的过渡图,进而生成连通区域并将其作为文字区域的一个重要特征. 而后根据先验知识对这些过渡图的形状进行分析、变形,就获得了文字候选区域.

1.4　根据文字的笔划特征

笔划是文字最重要的特征之一[14]. 在文献[20 - 21]中,使用了基于文字笔划特征的局部和全局的约束条件来定义子图像候选区域. 从局部看,文字区域中有很多似笔划的结构;从全局看,这样的似笔划结构在图像中具有特定的空间分布. 据此设计出了一种基于局部空间分析和空间相似度 CCA 的笔划过滤器以产生文字候选区域. 文献[9]中设计了一个快速简捷有效的算法用于检测字符的笔划,并定义了 2 个特征,即近似连续笔划宽度和局部对比度,从而先定位字符的笔划再用其进行文字定位.

2　区域分类

将输入图像分解成候选区域后,文字的检测与提取就变成了一个图像区域的分类问题,即根据先验知识或者用训练好的分类器将各个候选区域分类成文字区域或者非文字区域. 这类方法也包含下文中将直接用滑动窗口对图像进行穷尽式扫描,而后对各窗口所覆盖区域逐一进行分类的方法.

2.1　基于候选区域的方法

2.1.1　基于启发式搜索

在这类方法中,往往并没有复杂的分类器. 对候选区域的筛选是通过应用经验性的法则来决定保留文字区域或者滤除非文字区域的. 例如,在文献[11]中,使用了版面分析将可能被误检的区域滤除. 尺寸相似的水平相邻的候选区域被合并成为“行”. 对于短行,只有当边界点所占的比率足够大时,才会被保留. 在文献[22]中,3 条经验性的法则及相应的阈值被用于验证连接区域是否为定位的文字.

2.1.2　基于分类器

这类方法通常先提取出候选区域的某些特征,而后将其输入事先训练好的分类器,进行判决. 在文献[17]中,SVM 分类器被用来对候选区域进行分

类,所使用的特征是一种2维形状的描述特征.同样地使用SVM分类器,文献[16]中使用的是小波系数直方图的值和颜色变化值作为表示文字的特征,文献[10]中使用的特征是基于归一化的灰度值和连续梯度方差的,同时使用了颜色分布和几何学的先验知识来最终确定最后结果.

2.2 基于扫描窗口的方法

在这类方法中,候选区域的产生是通过穷尽式扫描得到的.由于扫描是穷尽式的,待分类的子图像区域的数目远远超过其他方法所产生的子图像候选区域.这就要求分类器对各个区域的分类效率足够高、分类速度足够快.近年来,基于Boosting和SVM的分类器因其快速、高效地分类而被广泛地用于许多视觉相关的目标检测系统中.就所使用的特征而言,上文提及的边界和纹理特征因其能快速运算而受青睐.

2.2.1 基于分类器的方法

在文献[3]中,扫描窗口是一个长宽比固定为2:1、尺寸可变的窗口,分类器是用AdaBoost算法训练出的由若干个弱分类器级联而成的.弱分类器基于3类若干个渐趋复杂的特征:基于图像水平和垂直方向梯度模值的均值和方差,基于像素灰度值、像素灰度值的梯度值和梯度方向的直方图和基于边界和边界连接的结果.基于AdaBoost算法的分类器也被应用在文献[23]的研究工作中.不同之处在于,它所使用的候选特征基于梯度方向直方图(HOG)和多尺度局部二值模式(msLBP).类似地使用级联分类器的思路也见于文献[6]中.作者针对文字区域提出了12个揭示其内在特性的特征,12级分别基于这些特征的弱分类器被级联在一起生成最终的分类器.在文献[22]中,基准窗口是包含32×32像素或者其整数倍的"文字段",从每一个文字段中提取出均差、标准差和HOG作为特征库,并使用了2个弱分类器,即线性判别式和基于高斯假设的对数似然率进行评估.最后不同尺度上的检测结果根据边界密度准则合并在一起,从而从含有文字行的图像中提取出相互重叠的文字段.此外,SVM因其完备的数学模型和丰富的开放代码资源,也被广泛应用于模式分类中.在文献[20]中,每一个15×15像素的扫描区域(子图像候选区域)被基于径向基函数的SVM分类器用来进行分类.

2.2.2 基于聚类分析的方法

在文献[24]中,对子图像区域的分类是通过在多通道小波空间中的特征聚类分析实现的,输入图像的每一颜色通道首先被变换到小波系数域,而后一个8×8像素的扫描窗口扫描该小波域中的每个位置,在每一个位置处,计算出特征值并将其输入k-means算法进行聚类分析.

2.2.3 基于变换阈分析的方法

寻找一种最适合文字检测的频带首见于文献[25],在这篇文献中,作者使用了基于空间频率的改进DCT的特征生成文字候选区域.为了获得更高的精度,结合Fisher判别分析法和Otsu求最佳阈值的思想提出了一种无监督求最佳阈值法.输入图像被划分成16×16像素的小块.对每一块进行DCT变换,计算出修正的基于DCT变换的特征值,并用Otsu求阈值方法的思想确定将区域分为文字区或者非文字区的最佳阈值.

3 性能比较

提到性能的比较,目前广泛使用的衡量标准是检测率和精准率.其中,检测率定义为检测出的目标占所有目标总数的比率,精准率定义为目标占全部检测为目标的比率.此外,为了便于比较系统的性能,很多方法在2003ICDAR数据集[26]上进行了测试,其他的使用了自己创建的数据集进行的测试.本文中提到的一些方法的性能如表1所示,其他方法有的没有明确给出检测率,有的使用了其他方式进行评估.

表1 文字信息检测的性能比较

文献	图像尺寸(数据集)	检测率/%	精准率/%	单帧处理时间/s
[10]	720×480(文字行高度:10~150像素)	0.972536	0.976241	0.64
[11]	(ICDAR 2003数据集)	75.23	67.64	
[13]	1024×768	94		
[14]	384×288	91.1	88.9	
[15]	(ICDAR 2003数据集)	64.3	56.3	
[16]	640×480到1024×768	86.2		
[20]	720×480(文字行高度:10~150像素)	94.9	98.8	0.196
[21]	720×480(文字高度:10~250像素)	97	42.4	0.466
[22]	(ICDAR 2003数据集)	79.2	37.3	<2
[23]	(ICDAR 2003数据集)	68	67	1.5
[25]	300×260到800×600	74	52	

4 结语

由表1可以看出,尽管大量的科学研究,复杂背

景下室外场景中任意文字的提取仍然是一个难题.许多问题还未解决,亟待进一步的研究,特别是下面几个方面带来的挑战:文字的多样性,特别是在字符的语种、字体、大小、颜色、排列等等方面;复杂的背景;难以预测不均匀的光照;快速处理.

参考文献:

[1] WU W, CHEN X, YANG Y. Incremental detection of text on road signs from video with application to a driving assistant system[C]//Proceeding of the 12th Annual ACM International Conference on Multimedia, 2004:852 - 859.

[2] KASTRINAKI V, ZERVAKIS M, KALAITZAKIS K. A survey of video processing techniques for traffic applications[J]. Image and Vision Computing, 2003, 21(4):359 - 381.

[3] CHEN X, YUILLE A L. Detecting and reading text in natural scenes[C]// Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004,2:366 - 373.

[4] CHEN X, YANG J, ZHANG J, et al. Automatic detection and recognition of signs from natural scenes[J]. Image Processing, IEEE Transactions on,2004,13(1):87 - 99.

[5] YANG J, CHEN X, ZHANG J, et al. Automatic detection and translation of text from natural scenes[C]//Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing,2002,2:2101 - 2104.

[6] ZHU K, QI F, JIANG R, et al. Automatic character detection and segmentation in natural scene images[J]. Journal of Zhejiang University (Science),2007,8(1):63 - 71.

[7] JUNG K, KIM K, JAIN A K. Text information extraction in images and video: a survey[J]. Pattern Recognition, 2004,37:977 - 997.

[8] CHEN X, YANG J, WAIBEL A. Automatic detection of signs with affine transformation[C]//Proceedings of the sixth IEEE Workshops on Applications of Computer Vision,2002:32 - 36.

[9] SUBRAMANIAN K, NATARAJAN P, DECERBO M, et al. Character - stroke detection for text - localization and extraction[C]//Proceedings of Ninth International Conference on Document Analysis and Recognition,2007,1:33 - 37.

[10] JUNG C, LIU Q, KIM J. Accurate text localization in images based on SVM output scores[J]. Image and Vision Computing,2009,27:1295 - 1301.

[11] PAN W, BUI T D, SUEN C Y. Text detection from scene images using sparse representation[C]//19th International Conference on Pattern Recognition,2008:1 - 5.

[12] SHEN H, COUGHLAN J. Finding text in natural scenes by figure - ground segmentation[C]//18th International Conference on Pattern Recognition,2006,4:113 - 118.

[13] SILAPACHOTE P, WEINMAN J, HANSON A, et al. Automatic sign detection and recognition in natural scenes[C]// Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,2005.

[14] ZHU C, WANG W, NING Q. Text detection in images using texture feature from strokes[J]. Lecture Notes in Computer Science, 2006,4261:295 - 301.

[15] KIM J, PARK S, KIM S. Text locating from natural scene images using image intensities[C]//Proceedings of the 2005 Eighth International Conference on Document Analysis and Recognition,2005,2:655 - 659.

[16] YE Q, JIAO J, HUANG J, et al. Text detection and restoration in natural scene images[J]. Journal of Visual Communication & Image Representation, 2007,18:504 - 513.

[17] HARITAOGLU E D, HARITAOGLU I. Real time image enhancement and segmentation for sign/text detection [J]. Proceedings of the International Conference on Image Processing, 2003,3:993 - 996.

[18] PARK J, PARK S. Detection of text region and segmentation from natural scene images[J]. Lecture Notes in Computer Science,2005,3804:666 - 671.

[19] KIM W, KIM C. A new approach for overlay text detection and extraction from complex video scene[J]. IEEE Transactions on Image Processing,2009,18:401 - 411.

[20] JUNG C, LIU Q, KIM J. A stroke filter and its application to text localization[J]. Pattern Recognition Letters, 2009,30:114 - 122.

[21] LIU Q, JUNG C, KIM S, et al. Stroke filter for text localization in video images[J]. IEEE International Conference on Image Processing,2006:1473 - 1476.

[22] HANIF S M, PREVOST L, NEGRO P A. A cascade detector for text detection in natural scene images[C]//19th International Conference on Pattern Recognition,2008:1 - 4.

[23] PAN Y, HOU X, LIU C. A robust system to detect and localize texts in natural scene images[C]// The Eighth IAPR Workshop on Document Analysis Systems,2008:35 - 42.

[24] SAOI T, GOTO H, KOBAYASHI H. Text detection in color scene images based on unsupervised clustering of multi - channel wavelet features[C]//Proceedings of the 2005 Eighth International Conference on Document Analysis and Recognition,2005,2:690 - 694.

[25] GOTO H. Redefining the DCT - based feature for scene text detection[J]. International Journal on Document Analysis and Recognition, 2008,11:1 - 8.

[26] 2003 ICDAR Text Location Contest trial test database[EB/OL]. http://algoval.essex.ac.uk/icdar/datasets.html.

(责任编辑　庄红林)