

图像文本定位技术研究

黄治虎

(重庆广播电视大学技术与资源中心, 重庆 400052)

摘要: 图像文本定位是图像中文本识别的关键步骤。介绍了文本图像的分类及特征, 分析了图像文本定位常用算法的应用范围及优缺点, 提出了文本定位算法的评价方法, 阐述了文本定位的应用领域, 并对今后文本定位的研究方向进行了分析。

关键词: 图像文本定位; 图像文本识别; 图像检索

中图分类号: TP391

文献标识码: A

文章编号: 1007-9599(2013)01-0016-03

图像中的文本通常描述了图像的有效信息, 如路牌、字幕、商店招牌等。如果将图片中的文本自动地识别出来, 对图像高层语义的自动理解、索引和检索是非常有价值的。大多数图像中的文本具有相对明显的特征, 可通过计算机图像处理技术进行文本定位和分离, 再通过较为成熟的文字识别技术, 就能将图像中的文本提取出来。图像文本识别已成为数值图像处理领域的研究热点。

图像文本的识别主要包括以下部分: (1) 获取图像; (2) 对获取图像进行预处理; (3) 检测图像中包含文本的矩形区域, 进行文本定位; (4) 将文本从图像中分离出来, 并通过光学字符识别系统(OCR)进行文本识别; (5) 保存识别结果。其中第3步图像文本定位是最为关键的步骤, 成为研究的重点和难点。

1 图像中文本的分类及特征

1.1 图像中文本的分类

图像中的文本分为人工文本(ArtificialText)和场景文本(NaturalSceneText)。人工文本是指使用图像处理工具对获取的原始图像进行后期编辑, 人为加在图像上的文本, 如新闻视频中的字幕。场景文本是指在图像获取时就存在于场景中的文本, 随场景一同生成到图像中, 其本身是场景的一部分, 如广告牌, 路标, 海报等。

人工文本多为视频的字幕或图片的文字说明, 为了方便人们阅读, 人工文本与背景的对比较强, 文字颜色一致, 排列整齐, 文本内容与图片内容的相关性强。场景文本的规律性较差, 文字的方向、大小、字体没有限制, 颜色变化较大, 文本与背景的对比较跟环境有很大的关系。由于获取图像时的投影变换关系, 图像中的文字可能会发生旋转、缩放等变形, 加上拍摄角度, 光照等因素的影响, 其定位难度较人工文本大。

1.2 图像中文本的特征

(1) 文本的颜色特征。大多数情况下, 同一区域中同一行文字, 其颜色、色调、亮度等属性相同或相似, 与图片背景有较大的对比度。人工文本尤其如此。但也有不一致的情况, 如颜色渐变的艺术字, 广告中的文字等。

(2) 字符的大小。同一行文字, 字符大小基本一致。但同一图像中的文字大小, 可能差别很大, 大的可能占到图像面积的80%以上, 而小的则只有几个像素。由于太小的字符, 识别率不高, 像素低于10的字符通常被删除。

(3) 字符的边缘。由于字符与图像背景间具有一定的对比度, 因而字符与背景间都有较突出的边缘, 这为通

过边缘检测进行文本定位提供了依据。边缘包含边缘强度和边缘方向两个要素。不同的文字, 边缘的方向不同, 如英文和汉字的边缘方向集中在水平、垂直和45°方向, 而拉丁文的边缘方向集中在垂直方向。

(4) 其他特征。同一行文字中, 字符的间距相等。字符的宽和高在一定的比例范围内。文本区域具有高填充率、高频率等纹理特征。

2 图像文本定位的主要方法

2.1 基于文本边缘的方法

为方便阅读, 文本与图像背景存在较大的对比度, 从而在文本与背景间形成较强的边缘。文本边缘具有边缘强度和边缘密度两个特征。对于简单背景的图片, 边缘密度特征明显; 对于复杂背景的图片, 边缘强度特征明显。基于边缘的文本定位主要步骤包括: (1) 检测图像的边缘密度和强度; (2) 通过平滑滤波、形态学膨胀、区域合并等方法检测到可能包含文本的矩形区域, 形成候选文本区域; (3) 根据文本的一些启发性规则(如字符的宽高比, 边缘密度, 边缘方向, 填充率等)去除非文本块, 获得最终的检测结果。

基于边缘的文本定位方法, 优点是速度较快, 对于文本边缘较突出, 图像背景边缘较少, 背景边缘与文本边缘的交叉连接不多的图片, 其定位准确率高, 误判率低。但对于背景边缘比较复杂、噪声较大的图片, 文本定位的结果不理想, 需要与其他方法结合起来, 进行文本的定位。

2.2 基于连通区域的方法

基于连通区域的方法主要利用同一区域文本颜色相近、边缘密度高等特性。采用自底向上的方式, 使用颜色聚类、阈值、区域生长等方法把图像分成一些小的区域, 然后根据一定的规则, 将这些小的区域连接成大的区域, 形成候选文本区域集合。再利用区域大小、宽高比、占用率、边缘强度等启发性规则来滤除非文本区域, 获得真正的文本区域。

基于连通区域方法的优点是实现比较简单, 对于文本和背景比较单一的图像, 该方法的准确率高。其缺点是对文本颜色丰富, 分辨率低, 噪声高的图像, 定位准确性差, 甚至将一个字符分成几个文本块。另外, 在滤除非文本区域时, 用到的各种阈值(如宽高比, 占用率)依赖于经验或实验, 还没有一个确定这些阈值的通用方法, 造成文本定位的精确性与图片数据库有很大的关系, 算法的通用性受到一定的限制。

2.3 基于纹理的方法

文本具有一定的风格,即具有一定的纹理,如文本区域有相同的方向;文字的字间距和行间距相同;文本在水平、垂直和45°方向的边缘较强;每个字符的大小基本一致等特性。这种文本才具有的独特纹理与图像背景的纹理具有一定的区别,从而可通过纹理分割算法将文本从图像中分离出来。常用的纹理分析方法有小波变换,傅里叶变换, Gabor 滤波, Gaussian 滤波, 空间方差等。基于纹理的方法一般将输入图像分成固定的小块,检测每小块的纹理特征,根据纹理特征确定候选文本区域,然后根据一定的算法,将相邻的文本块进行合并,得到最终的文本定位结果。

基于纹理方法的优点是稳定性高,能检测到图像背景复杂,文本与背景对比度小,噪声较大的文本。其缺点有:

(1) 纹理特征和纹理分类的计算量大,算法复杂度高,比较耗时;(2) 定位精度较低;(3) 对文字的风格、大小非常敏感,在处理“类文本(如艺术字,商标 LOGO)”时,出现错误的概率大;(4) 很难设计一个适合所有文本的纹理分类器。

2.4 基于人工智能的方法

前面三种文本定位方法都包含一个共同的阶段,即将非文本区域从候选文本区域中删除。删除非文本区域通常依赖于一个或多个阈值,而阈值则通过经验或实验得到的,因而算法的通用性不强。实际上,图像文本检测问题是典型的两分类问题,即将图像分成文本和非文本区域,因而可以使用人工智能的方法来实现分类。该方法一般分为图像预处理,特征值提取,分类器设计及训练和后处理等模块。其关键在于特征值提取和分类器的设计及训练。特征值一般使用图像的统计值,如颜色、形状、纹理、梯度等。分类器一般使用多层神经网络、支持向量机 SVM 等。

基于人工智能方法的优点是鲁棒性好,即使对噪声大和光线不均的图像,也能获得较好的结果,同时,可以通过增加学习样本来提高算法的鲁棒性和准确性。其缺点是检测结果的好坏严重依赖于特征值的提取和分类器的训练样本。当前,还没有一组特征值能够很清晰的将文本和非文本分开。在复杂背景图像中,文字的大小、字体、风格、排列方式的变化很大,很难训练出一个通用的、适合所有图像的分类器。另外,人工智能方法的时间复杂度也较大。

3 图像文本定位结果的判定

目前,专家、学者提出了各种各样的文本定位算法,但怎样客观、有效地评价、比较这些算法的优劣,即推出一个文本定位算法的评价标准显得尤为必要。文本定位算法的评价依赖于文本定位矩形框的精确度和评价算法的统一样本图像库两个前提,下面进行说明:

(1) 文本定位矩形框的精确度由标定的矩形与 Ground-truth 矩形的重合度来衡量,两者的重合度越高,文本定位精度越高。Ground-truth 矩形是指正好包含需定位文本的矩形区域,即 Ground-truth 矩形是刚好包含文字的矩形,既不偏大,也不偏小。

(2) 统一的样本图像库。如果没有统一的样本图像库,而是不同的算法在各自的样本库上测试和评价,这样的测试结果缺乏可信度,参考价值小。在没有公认的图像样本库前,很多算法都使用到 ICDAR (International Conference on Document Analysis and Recognition) 2003 年推出的样本图像库。该图像库主要是通过数码相机拍摄得到的场景图像,以门牌、广告、指示标识为主。图像中的英文在字体、颜色、大小、对比度等方面存在多种变化,图像背景具有不同的复杂度。字符的大小变化极大,小的高度小于 10 个像素,大的高度大于 800 个像素。图像库分为样例库、训练库、测试库 3 个部分,分别包含 20、258、400 张图像。

有了这两个前提,就可以用文本定位矩形的查准率和查全率来评价文本定位算法。一般来说,文本定位算法输出的文本区域集合中包括文本区域和非文本区域。查准率是指文本区域集合中的文本区域数与所有文本区域数(即文本区域数与非文本区域数的和)的比值,公式为:查准率=文本区域数/所有文本区域总数。查准率越高,算法把图像背景误认为文本的概率就越低,文本定位的效果越好。查全率是指正确定位的矩形数与测试图像中所有文字矩形数的比值(即正确定位到的文本区域数与未定位到的文本区域数的和),公式为:查全率=正确定位的文本区域数/图像中所有文本区域总数。查全率越高,算法检测出图像中所有文本区域的概率越高,漏检文本的概率越低。

4 图像文本定位的应用

通过图像文本定位可以分割出图像中的文本区域,再通过文本识别技术就可获得文本的内容,而图像中的文字通常能表达图像内容,因此,图像文本定位具有广泛的用途,主要有以下几个方面:

(1) 图像、视频检索。随着监控系统、影视摄影、照相和网络技术的飞速发展,越来越多的信息以图像的形式存储、传播。怎样快速地检索到需要的图像,成为急需解决的问题,基于图像文本识别的图像内容检索成为重要的解决方案之一。

(2) 网络图像过滤。随着网络技术的飞速发展和网络普及率的不断提高,通过网络传播的信息爆炸式增长,其中不乏存在一些不良、甚至有害的信息。通过网络不良信息过滤系统可以阻止以文字方式传播的信息。为了逃避检查,一些不良信息被嵌入到图片中,或者直接用图片来传播文本信息。通过图像文本识别技术,就可将这些图片中的文本提取出来,从而实现不良网络图片的过滤。

(3) 智能交通:文本定位技术广泛地应用于智能交通中的车辆牌照、路标等信息的收集,提高交通的信息化和智能化管理水平。

(4) 邮件分拣:当前,我国的邮件分拣工作基本上由人工完成,这不仅费时费力,而且常会因分拣人员疲劳而出现分拣错误。若能利用计算机自动地定位邮件上的信息,进行自动分拣,既能把人员从简单繁重的机械劳动中解放出来并获得高效的工作效率,又能 24 小时不间断工作,并且不会因人员疲劳而增加出错概率。

(下转第39页)

4.2.2 车辆定位位置显示方式

不同车辆实时定位位置在电子地图上以图符形式显示,在系统参数设置中可以设置各类车辆图符颜色,同时显示各个车辆车牌号码,点击某个车标,可在新窗口中显示该车辆详细信息,如SIM卡号、IP地址、车辆牌号、车辆品牌、车辆型号、车辆颜色、车辆类别、车辆类型、管理员、所属部门、备注等。

4.3 车辆实时表格监测

提供以表格形式实时显示所选车辆运行状态信息,包括:车辆牌号、连续在线时间、车速、方向、状态(开、停、离线),其中,方向用箭头方式表示。实时表格所显示的车辆可按车辆类别选择,同时提供选择“全部车辆”功能。

4.4 车辆轨迹回放

监控中心对车辆历史行驶状态进行详细记录,在地图上可以描绘出车辆的历史行程轨迹,可以作为车辆管理、监控的依据。具体操作,选择车辆、指定某段历史时间、选择轨迹回放速度,点击“轨迹回放”按钮,在电子地图上以图形化显示车辆选择时间段内的行驶轨迹,在轨迹回放过程中,可以选择“暂停”按钮暂停回放,选择“开始”按钮继续轨迹回放。

4.5 报警条件设置

车辆报警可以有诸多条件,可以根据需要进行规范设置,如下述:

超速报警:即所选类型车辆或单车,超过设定车速后报警。该类报警需要选择车辆类型和限制车速值。

越界报警:即所选类型车辆或单车,驶入或驶出划定区域后报警。该类报警需要选择车辆类型、划定区域(可为矩形、正方形、点圆、线型及查询对象范围)、选择驶入或驶出方向来设置报警。

4.6 下达指令管理

下达指令管理包括预设指令内容、编辑发送指令、发

送指令、查询已发送指令信息等功能。

预设指令内容:提供预先定义一些常用指令内容,便于快速发送指令。

编辑发送指令:选择预设指令后编辑或直接编辑要发送的指令内容。

发送指令:编辑要发送的指令内容后,选择发送车辆,向其发送指令。

查询已发送指令:监控中心向车载终端发送的指令信息会详细记录,提供根据时间段、车辆牌号条件,查询发送指令内容。

5 应用效果

基于车辆跟踪管理信息系统的应用在各个领域中均有应用,包括医疗卫生部门、物流运输部门、出租车运管部门等等,应用系统运行良好,效果理想。不仅为企业生产运营管理、安全管理、指挥调度等提供强有力的支持保障服务,同时提高了生产、经营、管理、决策效率和水平,为企业的安全运营和有效管理提供了高科技的信息化手段和工具。

该系统目前在机场领域的应用还可谓凤毛麟角,但潜力巨大并逐渐受到考虑和重视,因为它利用空间地理数据服务于生产运行,凭借高效的管理减少了大量安全事件的发生,节省了大量的人力和运营成本,相信在不久的将来基于多模卫星定位技术的实现将会更高效地服务于各行各业!

参考文献:

[1]钱文序.煤炭企业管理信息系统的设计与实现[J].计算机光盘软件与应用,2012,22:208-209.

[2]王硕,高玉珊.企业人事信息管理系统的设计与实现[J].计算机光盘软件与应用,2012,23:189-190

[作者简介]陈君(1985-),男,汉,首都机场股份有限公司,助理工程师,研究方向:地理信息技术。

参考文献:

[1]KeechulJung,KwangInKim,AnilK.Jani.TextInformation ExtractioninImagesandVideo:ASurvey[J].PatternRecognition,2004,37(5):977-997.

[2]XiaoqingLiu,JagathSamarabandu.MultiscaleEdge-basedTextExtractionfromComplexImages.ICME2006:1721-1724.

[4]M.EmreCelebi.EffectiveInitializationofk-meansforColorQuantization.ICIP2009:1649-1652.

[5]K.I.Kim,K.Jung,S.H.Park,H.J.Kim.SupportVectorMachine-basedTextDetectioninDigitalVideo.PatternRecognition,2001,34(2):527-529.

[作者简介]黄治虎(1974-),男,四川中江人,硕士,副教授,主要研究方向为远程教育,教育信息化,网络技术。

[资助基金]重庆市教育委员会科学技术研究项目“面向网络图像过滤的图像文本提取研究”(基金号:kj121606)

(上接第17页)

5 结束语

由于图像文本定位具有广泛的用途,其已成为计算机领域的研究热点。通过研究者坚持不懈的努力,该领域的研究已取得一些进展,如自动视频检索系统,名片识别系统等。由于人工文本具有文本排列规则,与背景的对比度高,图像的质量较好等特点,其文本定位相对容易。对于场景文本,由于其文本大小、字体颜色、字体种类、排列方式没有限制,加上图片获取时易受设备、光照等因素的影响,定位较人工文本大得多。

使用如边缘、区域、纹理、人工智能等单一技术的文本定位方法,对复杂背景的图像,很难达到满意的文本定位结果。为提高文本定位算法的性能和通用性,往往需要多种技术融合,才能获得较好的结果。随着神经网络、支持向量机等人工智能、机器学习技术的发展,怎样使用机器学习技术来提升复杂背景图片的文本定位结果,成为今后图像文本定位算法的研究重点。