



(12)发明专利申请

(10)申请公布号 CN 105825216 A

(43)申请公布日 2016.08.03

(21)申请号 201610153384.X

(22)申请日 2016.03.17

(71)申请人 中国科学院信息工程研究所
地址 100093 北京市海淀区闵庄路甲89号

(72)发明人 谢洪涛 刘顺 谭建龙 戴琼

(74)专利代理机构 北京君尚知识产权代理事务
所(普通合伙) 11200

代理人 司立彬

(51)Int.Cl.
G06K 9/32(2006.01)

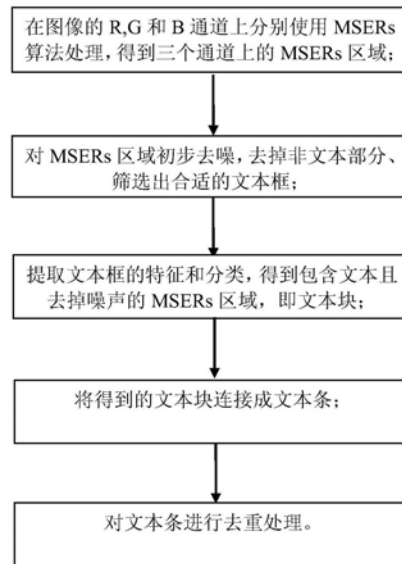
权利要求书1页 说明书4页 附图1页

(54)发明名称

一种复杂背景图像中的文本定位方法

(57)摘要

本发明公开了一种复杂背景图像中的文本定位方法。本方法为:1)在待处理的彩色图像的R、G和B三个通道上分别使用MSERs算法进行处理,得到对应三个通道上的MSERs区域;然后将得到的结果标记在该彩色图像上,得到每个MSERs区域在彩色图像上的坐标;2)进行初步去噪,去掉判断为重复的MSERs区域和非文本的MSERs区域,得到候选MSERs区域;3)从得到的候选MSERs区域中提取设定特征,然后利用一基于该设定特征训练的分类器对候选MSERs区域分类,得到包含文本的MSERs区域,即文本块;4)将得到的文本块连接成文本条;5)对得到的文本条进行去重处理。本发明大大提高了文本定位效率。



1. 一种复杂背景图像中的文本定位方法,其步骤为:

1) 在待处理的彩色图像的R、G和B三个通道上分别使用MSERs算法进行处理,得到对应三个通道上的MSERs区域;然后将得到的结果标记在该彩色图像上,得到每个MSERs区域在彩色图像上的坐标;

2) 对得到的MSERs区域进行初步去噪,去掉判断为重复的MSERs区域和非文本的MSERs区域,得到候选MSERs区域;

3) 从得到的候选MSERs区域中提取设定特征,然后利用一基于该设定特征训练的分类器对候选MSERs区域分类,得到包含文本的MSERs区域,即文本块;

4) 将得到的文本块连接成文本条;

5) 对得到的文本条进行去重处理。

2. 如权利要求1所述的方法,其特征在于,对得到的MSERs区域进行初步去噪的方法为:根据每个MSERs区域在彩色图像上的坐标计算两MSERs区域的重叠面积,将重叠面积占这两个区域合并后的面积比例超过设定比例阈值时,判定这两个MSERs区域是重复的,并去掉其中面积较小的一个MSERs区域;然后计算剩余MSERs区域内包含的小区域数量,如果该MSERs区域包含的小区域数量超过设定阈值,则将该MSERs区域剔除掉;其中将区域的面积与该区域所在MSERs区域面积的比值小于设定阈值的区域称为小区域。

3. 如权利要求1所述的方法,其特征在于,将得到的文本块连接成文本条的方法为:将两个文本块之间的距离小于设定距离阈值且相似度大于设定相似度阈值的文本块连接起来,将剩余的独立的文本块连接成文本条,得到一个或多个文本条。

4. 如权利要求3所述的方法,其特征在于,所述距离阈值设为该代理处理彩色图像的宽度的8%。

5. 如权利要求1所述的方法,其特征在于,对得到的文本条进行去重处理的方法为:采用随机森林分类器对得到的每一文本条进行投票,如果投票率小于设定阈值,则将该文本条作为非文本剔除掉;对于剩余文本条,计算两文本条交叉部分面积占这两文本条合并部分面积的比例,如果该比例超过设定比值,则将面积较小的文本条删除。

6. 如权利要求5所述的方法,其特征在于,对得到的文本条进行去重处理的方法为:如果一个文本条包含若干小的文本条,则将这些小的文本条删除。

7. 如权利要求1~6任一所述的方法,其特征在于,所述设定特征为Hog特征,所述分类器为SVM分类器。

一种复杂背景图像中的文本定位方法

技术领域

[0001] 本发明属于图像处理的范畴,是一种基于MSERs复杂背景图像中的文本定位方法。

背景技术

[0002] 图片中的文本往往包含有价值的信息,并且在很多基于内容的图片和图像应用中被开发,像基于内容的网络图片查找,图像信息检索,和自动文本分析和识别。由于背景,文本方向、颜色、大小,字体的复杂性,在文本被识别和检索前需要鲁棒地检测到。许多研究者把光学字符识别(OCR)看作一个已经解决的问题,而对于图像中文本的检测和识别这个领域还有很大的研究空间。国际上对于图像中文本检测和识别的研究已经进行了很多年,有很多相关的国际竞赛。Robust Reading竞赛(2003,2005,2011,2013),Camera-based Document Analysis And Recognition(2005-2013,一年两次)。

[0003] 场景文本检测的方法大致分为三类:基于滑动窗的方法,基于成分连接的方法,混合方法。

[0004] 本文中使用的Maximally Stable Extremal Regions(MSERs,最大稳定极值区域)可以归为基于成分连接的方法。MSERs(Maximally Stable Extremal Regions,最大稳定极值区域)该算法最早是由Matas等人于2002年提出,它是基于分水岭的概念,最早用于斑点检测,后来越来越多的研究者使用它来进行图像中的文本检测。MSER的基本原理是对一幅灰度图像(灰度值为0~255)取阈值进行二值化处理,阈值从0到255依次递增。阈值的递增类似于分水岭算法中的水面的上升,随着水面的上升,有一些较矮的丘陵会被淹没,如果从天空往下看,则大地分为陆地和水域两个部分,这类似于二值图像。在得到的所有二值图像中,图像中的某些连通区域变化很小,甚至没有变化,则该区域就被称为最大稳定极值区域。数学原理如下:

[0005] 图像I作为一个映射 $I: D \subset \mathbb{Z}^2 \rightarrow S$,极值区域可以定义在图片上,如果:

[0006] 1.S是完全有序的 $S = \{1, 2, \dots, 255\}$

[0007] 2.定义一个连接关系

[0008] 区域Q是D的连续子集。

[0009] 区域边缘: $\partial Q = \{q \in D \setminus Q : \exists p \in Q : qAp\}$

[0010] 极值区域(ERs):对 $Q \subset D$,对所有 $p \in Q, q \in \partial Q : I(p) > I(q)$ (或者 $p \in Q, q \in \partial Q : I(p) < I(q)$)。

[0011] 最大稳定极值区域(MSERs):对于一组内嵌的极值区域 $Q_1, \dots, Q_{i-1}, Q_i, \dots$ ($Q_i \subset Q_{i+1}$),对于 $q(i) = |Q_{i+\Delta} - Q_{i-\Delta}| / Q_i$ 当且仅当在i处取到局部最小值时,就称 Q_i 是最稳定的,其中, $h \in S$ 。

发明内容

[0012] 本发明的主要目的在于提供一种复杂背景图像中的文本定位方法。

[0013] 本发明的技术方案为：

[0014] 一种复杂背景图像中的文本定位方法，其步骤为：

[0015] 1)在待处理的彩色图像的R、G和B三个通道上分别使用MSERs算法进行处理，得到对应三个通道上的MSERs区域；然后将得到的结果标记在该彩色图像上，得到每个MSERs区域在彩色图像上的坐标；

[0016] 2)对得到的MSERs区域进行初步去噪，去掉判断为重复的MSERs区域和非文本的MSERs区域，得到候选MSERs区域；

[0017] 3)从得到的候选MSERs区域中提取设定特征，然后利用一基于该设定特征训练的分类器对候选MSERs区域分类，得到包含文本的MSERs区域，即文本块；

[0018] 4)将得到的文本块连接成文本条；

[0019] 5)对得到的文本条进行去重处理。

[0020] 进一步的，对得到的MSERs区域进行初步去噪的方法为：根据每个MSERs区域在彩色图像上的坐标计算两MSERs区域的重叠面积，将重叠面积占这两个区域合并后的面积比例超过设定比例阈值时，判定这两个MSERs区域是重复的，并去掉其中面积较小的一个MSERs区域；然后计算剩余MSERs区域内包含的小区域数量，如果该MSERs区域包含的小区域数量超过设定阈值，则将该MSERs区域剔除掉；其中将区域的面积与该区域所在MSERs区域面积的比值小于设定阈值的区域称为小区域。

[0021] 进一步的，将得到的文本块连接成文本条的方法为：将两个文本块之间的距离小于设定距离阈值且相似度大于设定相似度阈值的文本块连接起来，将剩余的独立的文本块连接成文本条，得到一个或多个文本条。

[0022] 进一步的，所述距离阈值设为该代理处理彩色图像的宽度的8%。

[0023] 进一步的，对得到的文本条进行去重处理的方法为：采用随机森林分类器对得到的每一文本条进行投票，如果投票率小于设定阈值，则将该文本条作为非文本剔除掉；对于剩余文本条，计算两文本条交叉部分面积占这两文本条合并部分面积的比例，如果该比例超过设定比值，则将面积较小的文本条删除。

[0024] 进一步的，对得到的文本条进行去重处理的方法为：如果一个文本条包含若干小的文本条，则将这些小的文本条删除。

[0025] 进一步的，所述设定特征为Hog特征，所述分类器为SVM分类器。

[0026] 与现有技术相比，本发明的积极效果为：

[0027] 在我们自己的图像库上进行实验对比，我们方法F-measure达到85%，性能优于已有的state-of-art方法的75.5%，大大提高了文本定位效率。

附图说明

[0028] 附图为本发明的方法流程图。

具体实施方式

[0029] 为使本发明的目的、技术方案和优点更加清楚，下面结合附图和具体实施例对本发明做进一步说明。

[0030] MSERs方法往往使用在灰度图上,但是在灰度图上由于在彩色图转化为灰度图的过程中一些信息的丢失,造成一些必要的信息定位不到,所以我们把基于灰度图的MSERs算法扩展到了R,G,B三通道,对于一张彩色图像,我们在这张彩色图像的R,G和B三个通道上分别使用MSERs算法进行处理,在三通道上各设置最小为3的像素变化值和最小区域为30像素的阈值,这样得到三个通道上的MSERs区域,再把得到的结果标记在原来的彩色图像上,得到每个MSERs区域在彩色图像上的坐标,去掉中心坐标和宽高完全一致的区域,最后综合起来进行处理。

[0031] 由于得到大量的MSERs区域中有很多的噪声即非文本区域,像树叶、人、建筑等等,所以要对得到的MSERs区域要进行去噪声和筛选,即去掉非文本部分、筛选出合适的文本框。对得到的MSERs区域处理主要分为四部分:初步去噪→候选区域筛选→字符块连接→文本去重。

[0032] 初步去噪:

[0033] 根据前面得到的区域坐标和宽高信息计算区域是否重叠,如果两个区域重叠在一起且重叠面积超过两个区域面积并的80%,就可以认为这两个区域是重复的。对于重叠区域,去掉其中面积较小的一个,最终去掉一部分重合区域;根据文字的形态特征,文本区域里经常也会包含一些小的区域(将区域的面积与该区域所在MSERs区域面积的比值小于设定阈值的区域称为小区域),只是包含的区域数量往往不会太大,是小于某一个数值的。根据语言中单词(字母)所含有的细小结构不多的特点,我们设置一个阈值为10,如果一个MSERs内包含的较小的MSER区域数量超过这个阈值,则把这个MSERs区域认为成不良区域而剔除掉。通过以上两种操作可以删除一部分重复和简单的非文本区域。

[0034] 候选区域筛选:

[0035] 在这一部分训练了一个SVM(Support Vector Machine,支持向量机)分类器,分类特征是提取的文本块的Hog特征。首先把图像中去噪剩下的MSERs区域规格化为 24×32 大小,提取其hog(Histogram of Oriented Gradient,方向梯度直方图)特征,其中每个MSER区域图像中的块(block)大小为 8×8 ,而每个块中的细胞(cell)大小为 4×4 ,每个细胞中取9个特征。再通过SVM分类器进行文本块(character or word)或噪声的二类分类,由此分类后得到图像中的包含文本的MSERs区域同时去掉作为噪声的MSERs区域。这些包含文本的MSERs区域即可看作文本块。

[0036] 文本块连接:

[0037] 这一步主要是把上一步得到的文本块(character or word)连接成文本条(text)。根据文本块的高和宽,相互之间的距离等关系,把相临近的(在这里两个文本块之间的距离阈值设为输入图像宽度的8%,在这个距离范围内则认为两个文本块是相临近的)、高度相似的(一个文本块高度高于另一个高度的0.5倍小于2倍)文本块连接起来,于是得到一个或多个文本行。

[0038] 文本条去重:

[0039] 在这一部分,一个随机森林分类器被用来筛选那些通过了第一层的文本行。作为一个成熟的分类器算法,随机森林分类器是一个由弱分类器组合成强分类器的分类器,其输出结果由包含的决策树投票决定,其输出的类别概率即是这一类别投票所占的比率。在这一阶段,一系列文本行水平的基于区分文本和非文本的特征被挖掘出来训练随机森林分

类器。这些特征的提取步骤如下所示：

[0040] 1)所有的文本行被归一化为26*122像素。对于彩色图,R,G和B每个颜色通道中的从第2行到第25行和从第2列到第121列每个像素点的梯度都要计算,然后取三通道中的梯度值最大的一个作为彩色图像素的梯度。经过计算以后,可以得到一个包含像素梯度的二维向量。

[0041] 2)这些梯度值通过L2范数进行规则化

[0042] $v \rightarrow v / \sqrt{\|v\|_2^2 + \varepsilon^2}$ 其中 ε 为梯度向量,而是一个很小的值(这里 ε 被设置为0.1)。

[0043] 3)进行简单的池化操作。所有相邻接的2*2块被提取出来,这些块里面的最大值和最小值被挑选出来组成输入向量。

[0044] 4)最后,总共1440个特征从每个文本行中提取出来。

[0045] 每个候选文本行是文本的概率等于随机森林中决策树投票为正的比率,这个比率为一个double类型。如果一个文本行候选作为文本的概率小于一个阈值(本文设为0.3)将被作为非文本而剔除掉。如果两个文本行重叠部分占合并部分面积的90%,这两个文本行就可以认为是重复的文本行。如果两个文本行是重复的,面积比较大的那个会被留下而面积小的会被作为非文本删除。如果一个文本行包含一些小的文本行(面积小于此文本行的一半),那么这些小的文本行也会被删除。当每个文本候选都执行完上面提到的这些步骤,筛选剩下的即为我们想要的文本结果。

