doi:10.3772/j.issn.1002-0470.2016.02.001

一种基于笔画宽度特征和半监督多示例学习的文本区域鉴别方法①

吴 锐②* 杜庆安** 张博宇* 黄庆成*

(*哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)

(**天津航天机电设备研究所 天津 300000)

摘 要 考虑到文本区域鉴别在视频文本检测中的重要作用,提出了一种基于笔画宽度特征的文本区域鉴别方法,该方法通过分析候选文本区域中笔画宽度的分布,有效地区分文本和非文本区域。此外针对笔画宽度信息提取过程中存在未知极性参数的问题,提出了一种半监督多示例学习(SS-MIL)算法,该算法可以充分利用训练样本中不完整的监督信息,提高文本区域分类器的性能。基于上述方法,实现了一个完整的视频文本检测系统,并在具有代表性的数据集上对其进行了充分的实验,实验结果表明,基于笔画宽度特征和 SS-MIL 的文本区域鉴别方法能够有效地辨别文本区域,从而使该系统检测视频文本的综合性能达到较高水平。

关键词 文本区域鉴别, 笔画宽度, 半监督学习, 多示例学习(MIL)

0 引 言

在过去的数十年里,随着视频拍摄设备的广泛普及和互联网技术的飞速发展,视频数据的数量高速增长。视频服务提供商亟需有效的方法对海量的视频数据进行管理和存储。视频中的文本内容包含着丰富的语义信息,这些信息是进行视频资料自动注释、检索、压缩的重要依据。从视频图像处理和文本检测的研究角度出发,目前已经提出了一些视频文本检测方法[1]。这些方法大致可以分为三类:基于纹理的方法[2,3],基于连通组件(connect-component)的方法[4-6]以及基于边缘的方法[7,8]。这些方法从文本区域的不同特性出发,将前景(文本)从背景中剥离出来,然后将获得的前景组合成候选的文本区域。由于背景复杂多变、光照不均以及字体字形变化等原因,准确地将文本和背景区分开仍然比较困难。在检测候选文本区域的过程中,不可避免地会产生误报。目前大多数视频文本检测方法都需要在生成候选文本区域的基础上进行文本区域鉴别,因而大多数文本检测方法都包含前景检测、候选区域生成和文本区域鉴别三个阶段。在文本区域鉴别阶段,现有的文本检测技术大多数通过检测候选文本区域的几何特性来发现上一阶段产生的误报。经常使用的几何特征包括位置、方向、长宽比以及饱和度(候选区域前景与背景面积的比值)等。这些特征往往随着应用背景的变化而变化,在具体应用中需要手动进行调整。例如,在视频文本检测中,当检测目标的位置较为确定时(视频下部的字幕区域),基于位置的判别准则是有效的。但当检测目标出现的位置具有较强随机性时(嵌入文本区域或滚动字幕),这一准则就失去了意义。同理,基于区域方向、长宽比等特征进行文本区域鉴别时都需要提供目标数据的先验知识,因而不具有普遍意义,泛化能力较差。

本文提出了一种基于笔画宽度特征的文本区域

① 国家自然科学基金(61370162,61440025)和中央高校基本科研业务费专项资金(HIT.NSRIF.2012048)资助项目。

② 男,1976 年生,博生,讲师;研究方向:文本分析,模式识别,图像处理;联系人,E-mail: simple@hit.edu.cn

(收稿日期:2015-10-28)

鉴别方法。该方法根据候选文本区域内笔画宽度的分布情况来判别当前区域是否包含文本,其优势在于适用于大多数文本区域。在使用笔画宽度特征进行文本鉴别的过程中存在的一个难点是无法自动地获取文字前景与背景之间的亮度对比关系,而这一参数对于准确地提取笔画宽度信息来说至关重要。本文使用多示例学习方法(multi-instance learning, MIL)来解决这一问题。对于每一个样本,基于可能的极性参数提取笔画宽度特征,然后使用这些特征的集合来描述该样本。其中每个特征称为'示例',而特征的集合称为'示例包'。在此基础上可以使用多示例学习方法训练有效的文本区域分类器。由于在分类器训练过程中使用的训练样本集大多没有提供极性参数。本文在多示例学习方法的基础上提出了一种新的半监督多示例学习(semi-supervised multi-instance learning, SS-MIL)方法来进行文本区域分类器的训练。该方法结合多示例学习和半监督学习方法的特点,能够充分利用训练样本中不完整的监督信息,在降低学习成本的同时改进分类器的性能。

本文将上述文本区域鉴别方法与基于角点的文本区域检测方法[9]相结合,实现了完整的视频文本检测系统并在具有代表性的数据集上进行了充分的实验。实验结果表明本文提出的文本鉴别方法可以有效地辨别文本区域,使检测系统的准确率和召回率都达到了较高水平。

1 笔画宽度特征

图 1 示出了视频文本检测流程。在图中的前景检测和候选区域生成阶段,不可避免地存在误报的情况,需要采用有效的文本区域鉴别方法来排除误报的文本区域。本文提出了一种基于区域内笔画宽度的特征来实现文本区域的鉴别。该特征通过描述文本区域中笔画宽度的分布来反映区域的特性。相对于区域位置等几何特征,笔画宽度特征具有更好的泛化能力,适用于不同种类的文本。

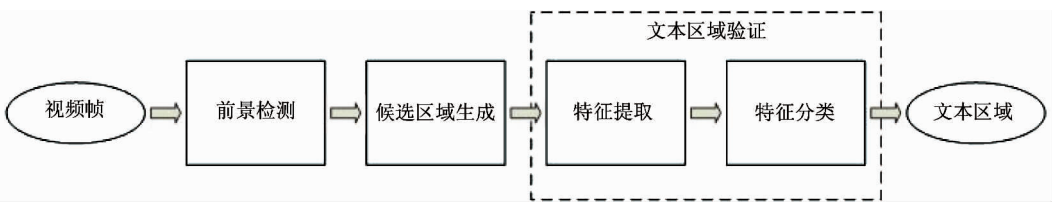
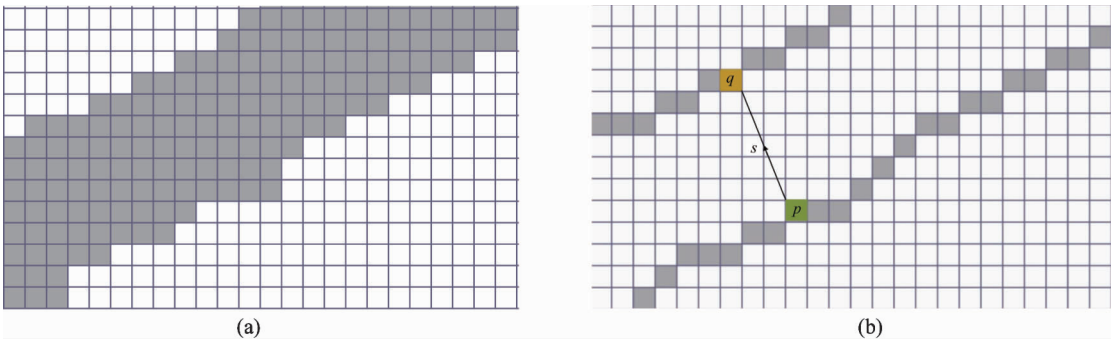


图 1 视频文本检测流程

在绝大多数语言中,字符都是由笔画构成的。为了使字符具有可识别性,笔画与背景区域之间需要明确的边界。因此笔画上的像素点都位于两个具有相反梯度方向的边缘点之间。这两个边缘点之间的距离为笔画的宽度。利用这一特性,可以使用笔画宽度变换[10]求出图像中每个像素点的笔画宽度。下面通过图 2 简要叙述笔画宽度变换的过程:

首先将所有像素点的笔画宽度值设置为 ∞,然后使用边缘检测器(本研究使用 Canny 算子)进行边缘检测。对于图像中每一个边缘点 p,记其梯度方向为 d_p。从点 p 沿其梯度方向 d_p 的反方向(假定文本的亮度低于背景亮度)发射一条射线 $s = p + n \times d_p$ 并沿该射线搜索,直到找到另外一个边缘点 q。如果点 q 的梯度方向 d_q 与点 p 的梯度方向 d_p 近似相反($|d_p - d_q| < 15°$),则射线 s 上的线段 $[p, q]$ 所经过所有像素点的笔画宽度值都设为线段 $[p,q]$ 的长度。相反,如果无法找到符合条件的点 q,则放弃射线 s,不更改 s 经过像素点的笔画宽度值。重复地对图像中的每个边缘点进行上述步骤。如果对某一个像素点发现一个比当前值更小的笔画宽度值,则将这一点的笔画宽度值更新为较小的值。当字符中出现较为复杂的情况,如笔画转弯处等,会出现明显错误的、极大的笔画宽度值。针对这种情况,算法使用中值抑制的方法来排除错误。



(a)从笔画上截取的一部分,其中每个方格表示一个像素点,灰色方格表示笔画上的点,白色方格表示背景点;(b)中灰色方格表示检测到的边缘点,从方格 p 向其梯度方向的反方向发射射线 s,到达 p 点,线段 $|p,q|$ 的长度即为射线 s 经过的像素点的笔画宽度值

图 2　笔画宽度变换示意图

实验结果表明,笔画宽度变换可以准确地提取文本图像中的笔画宽度信息。通常情况下,同一行中的文字使用的笔画宽度是大致相同的。因此,在文本区域鉴别问题中,如果一个候选区域确实包含文本内容,则其中落在字符上的像素点的数量在整个区域中应该大于一定的比例。而这些像素点的笔画宽度应该基本相同或在一个较小范围内变化。基于这一特性,本文使用笔画宽度分布直方图作为特征来描述整个区域的特性。对于一个候选的文本区域 r 而言,其笔画宽度特征的定义如下式所示:

$$hos_r = \frac{< p_1, \cdots, p_n >}{min(h,w)},\ p_i = s_i / \sum_{k=1}^{n} s_k \quad (1)$$

其中 s_k 表示候选区域中宽度值为 k 的像素点的个数,n 为使用的笔画宽度的最大值。h 和 w 分别为候选文本区域的高度和宽度,在保证候选文本区域为单行和单列的前提下,除以文字高(宽)度可以有效地消除候选区域面积的影响。当区域中不包含文本或只有少部分为文本区域的情况下,宽度信息的分布是不规律、较为杂乱的(图 3)。而当候选区域确实包含文本时,笔画宽度信息在文本的真实笔画宽度附近将出现一个较大的峰值(图 3(c))。

2　文本区域分类的多示例模型

为了能够准确地提取笔画宽度信息,需要指定候选文本区域中前景(文本)相对于背景的极性。然而在实际的应用中这一参数是难以由算法自动确定的。为了克服这一问题,本文提出了文本区域分类的多示例模型。该模型对候选文本区域依据两种可能的假设(前景亮度高于背景和前景亮度低于背景)分别进行笔画宽度变换(stroke width transform, SWT)。对于任意一个候选区域 r, 可以得到两组笔画宽度特征。其中一组能够反映 r 中真实的宽度信息分布。基于不同假设提取的笔画宽度信息如图 3 所示。

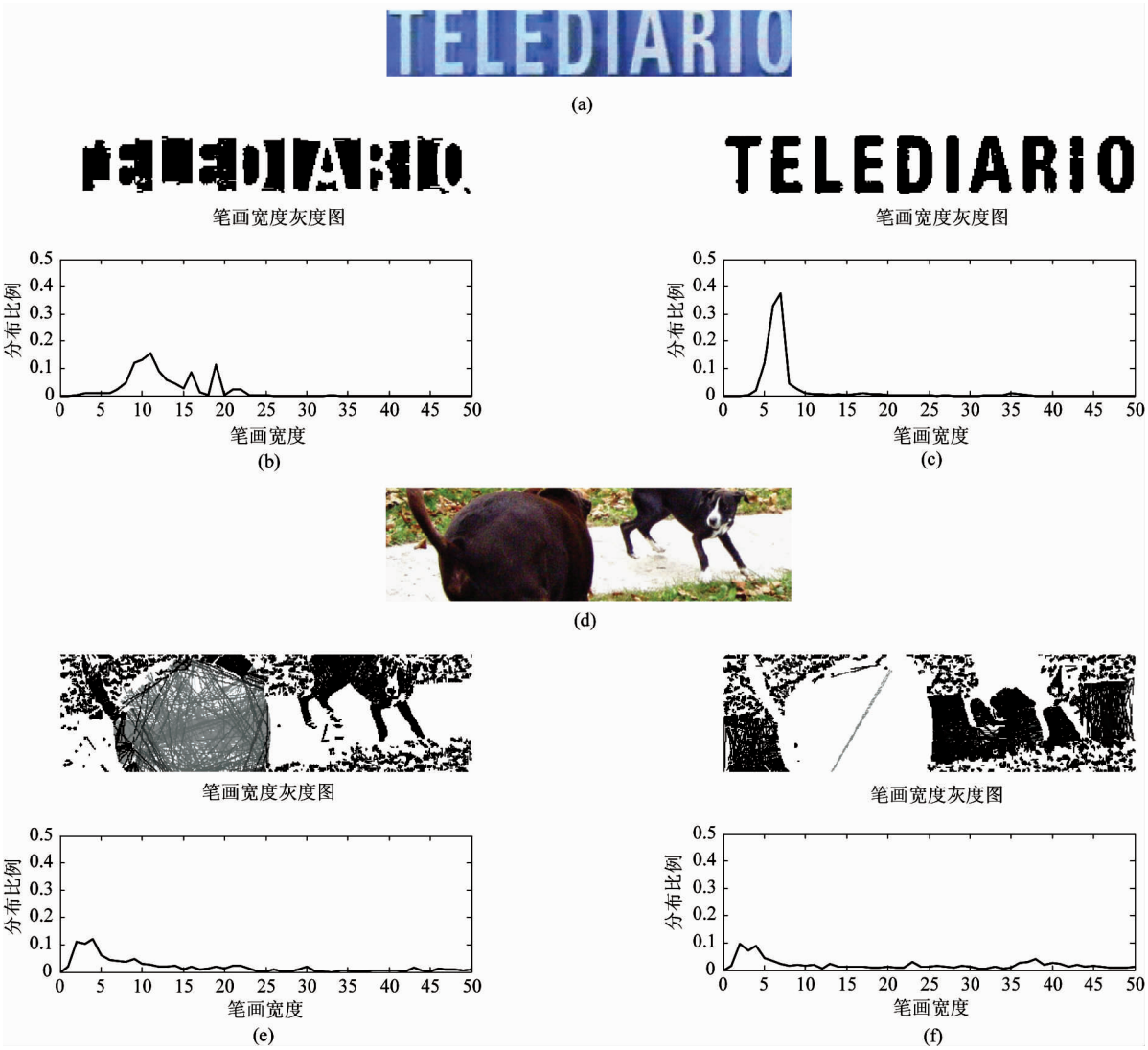
因此,本文使用集合 $hos_r = \{hos_r^d, hos_r^l\}$ 来描述区域 r 的笔画宽度特征。在多示例学习中,将该集合称为一个“示例包”,其中的每个特征称为一个“示例”。其中 hos_r^d 表示在假定文本区域的亮度低于背景时提取的笔画宽度特征,hos_r^l 则是在假定文本区域的亮度高于背景时提取的笔画宽度特征。假定已经获得了一个有效的基于 hos 特征的分类器 $p = F(hos)$, 则该分类器称为示例级的分类器,其中 p 表示分类器 F 对 hos 特征的概率输出。

在上述条件下,可以基于示例级分类器构建包级的分类器。候选区域的类别标签 l_r 可以用式

$$l_r = \mathrm{sgn}(\max(F(hos_r^d),\ F(hos_r^l)) + k) \quad (2)$$

计算,其中 k 为常数偏移量。

以式(2)为依据,对于一个需要进行鉴别的文本区域,只要其中基于不同假设提取的两组特征中有一组具有符合要求的笔画宽度分布,就认为该区域通过了基于笔画宽度特征的文本区域验证。



(a)检测到的包含文本的候选文本区域;(b)基于使用笔画宽度变换在假定前景亮度高于背景的条件下获取的笔画宽度信息绘制的图像和小于 50 的笔画宽度的分布,其中像素点的灰度值设为检测到的笔画宽度;(c)基于相反假设获取的笔画宽度信息和笔画宽度分布;(d)在 Pascal 数据集中得到的不包含文本的候选文本区域;(e)和(f)分别为基于不同假设获取的笔画宽度灰度图和相应的笔画宽度分布

图 3　在两种不同条件下进行笔画宽度变换的结果

3　文本区域分类器训练

本节给出了文本区域分类的多示例模型。为了获得一个有效的基于 *hos* 特征的文本区域分类器,需要提供足够的训练样本。由于使用笔划宽度变换(SWT)算法获取笔画宽度信息需要前景与背景之间的极性参数来判断搜索的方向,而现有的数据集中提供的监督信息往往只标注当前样本属于文本区域还是非文本区域,不提供极性参数,因而对训练样本中的每一个文本区域进行手工标注需要的人力消耗较大,这给分类器训练任务带来了困难。

上述问题可以使用多示例学习(MIL)方法来解决。对本文中的训练问题而言,文本区域分类器的学习是一个特殊的多示例学习问题,每个示例包中有且仅有两个示例。

在多示例学习方法的基础上,考虑到多示例学习算法的学习效果与监督学习效果相差较大,希望通过引入少量具有完整监督信息的训练样本来提高分类器训练的效果。本文提出了一种半监督多示例学习方法来训练的文本区域分类器。采用这一方法的主要目的是在降低消耗的同时充分利用样本中不

完整的监督信息。该学习算法的具体流程见算法1。

该方法首先根据实验数据的实际标注信息,将样本分为正例集合 P、反例集合 N 和无标签集合 U。正例集合 P 中的样本为真实的文本区域,而且极性参数是已知的。反例集合 N 中样本为非文本区域,其中的两组特征都不反映真实文本区域中的笔画宽度,不需要提供极性参数。集合 U 中的样本同样是真实的文本区域,但其中的极性参数并未提供。在算法的最初阶段进行有监督的学习,使用集合 P 和 N 中的示例进行分类器的训练。然后使用得到的分类器参数对集合 U 中的示例进行标注。进而使用所有样本再次进行分类器的训练。算法循环地执行上述步骤,直到集合 U 中示例的标签不再变化或达到预设的迭代次数为止。

算法1　半监督多示例学习算法

输入: 正例集合 P(带有示例标签),正例集合 U(带有包标签),反例集合 N

输出: 示例分类器 F,其参数集合为 θ

1:基于 P 和 N 训练分类器 F,求解参数 $\bar{\theta}$

2:利用 $\bar{\theta}$ 计算集合 U 中每个样本中两个示例的标签 $\bar{l}$

3:基于 P,U 和 N 重新训练分类器,求解 $\hat{\theta}$

4:重新计算集合 U 中示例的标签 $\hat{l}$

5:若 $\hat{l} \neq \bar{l}$

a. 令 $\bar{l} = \hat{l}$

b. 基于 P,U 和 N 训练分类器,更新 $\hat{\theta}$

c. 更新集合 中示例的标签 $\hat{l} = F(\hat{\theta})$

6:返回 $\theta = \hat{\theta}$

算法1不但利用了集合 P 和集合 N 中具有确定性的监督信息,还利用了集合 U 中不完整的监督信息,从而有利于提高文本区域鉴别的准确率。

值得注意的是,本文提出的半监督多示例学习方法并不局限于某些特定的分类方法,能够配合不同的分类器使用。当训练样本中同时存在有监督样本、无监督样本和半监督样本时,使用本文提出的方法可以充分利用样本中不完整的监督信息,提高分类器的性能。

4　实　验

4.1　实验数据

使用的测试数据集包括 Hua 等[11]收集的微软通用测试集(microsoft common test set,MCTS)。该数据集包括45帧包含文本内容的视频图像,其中包含的文本区域都进行了详细的标注,包括文字内容、位置、对比度等信息。此外,为了使实验具有更强的说服力,本文收集了一组新的视频文本数据。这些数据的来源包括新闻、体育、演讲、电影以及卡通等不同类型的视频片段。其中包含457帧图像,每一帧都包含有一个或多个文本区域,总的文本区域数量为1633个。本文的余下部分中将 MSTS 数据集记为'MS',本文收集的数据记为'PIC'。

为了对文本区域鉴别方法进行测试,首先要将文本区域从图像中提取出来。对于正例样本,可以比较容易地依据标记信息从图像中提取子图像并根据监督信息对这些图像进行归类。实验使用 MS 数据集中包含的152个文本区域作为监督信息完整的正例集合 P,PIC 数据集中包含的1633个文本区域作为半监督正例集合 U。为了获取反例样本,本文使用一种改进的基于角点的视频文本检测算法[12]进行候选文本区域检测。该方法首先将视频帧投影到尺度空间,然后在不同的尺度下进行角点检测并生成候选文本区域,最后将不同尺度下得到的候选文本区域合并。

具体地,本文使用 Pascal VOC 数据集[13]中的图像样本来生成反例。首先从该数据集中人工选择一组不包含文本内容的图片(247张),然后使用上述检测算法进行文本区域检测。由于上述图像样本中不包含任何文本内容,算法得到714个不包含文本内容的候选区域作为反例集合 N。使用有效的文本检测方法来获取反例的好处在于得到的区域边缘密度较大,与文本区域的相似程度更高。

4.2　文本区域鉴别

对4.1节中得到的文本区域样本,使用 SWT 算法从上述样本中提取笔画宽度信息(基于两种可能的极性参数)。然后分别统计笔画宽度信息的分布

并利用区域宽度与高度的最小值对该分布进行归一化,避免文本区域尺度对特征的影响。

实验考察指标为文本区域分类的准确率。准确率的计算采用交叉验证的方式:对于每次实验,将正例和反例样本随机地分成相等的两部分。其中一部分作为训练样本,另一部分作为测试样本。最终的准确率为将上述随机过程重复十次的平均值。本文首先使用带示例标签的正例集合 P 和反例集合 N 进行有监督的分类器训练。然后使用 EM-DD[14] 方法(改进的 MIL 方法)基于全部正例 U,P 和反例集合 N 进行多示例学习。最后使用本文提出的 SS-MIL 算法使用全部正例 U,P 和反例集合 N 训练分类器。基于 SS-MIL 算法的分类器训练过程如算法 1所述。上述三种学习模式下,均使用基于径向基函数(RBF)核的支持向量机(SVM)分类器[15] 作为示例级的分类器,分类器的具体实现基于 Libsvm[16]。三种不同学习模式得到的 SVM 分类器在相同的测试数据集上进行测试,得到的结果如表 1 所示。

表 1 文本区域分类精确度

学习算法	训练数据	准确率(%)
监督学习	P,N	83.32
EM-DD[14]	P,U,N	74.5
SS-MIL(本文)	P,U,N	91.72

实验结果证明,通过引入包含不完整监督信息的样本,使用本文提出的 SS-MIL 算法可以提高分类器的识别准确率。而 EM-DD 方法训练的分类器由于无法利用示例级标签包含的信息,分类器的准确率较低。

4.3 文本区域检测

为了验证高性能的文本鉴别方法对整个文本检测系统性能的影响。基于本文文本鉴别方法,本小节实现了一个完整的视频文本检测系统。为了保证实验结果的准确性,本研究在 MS 数据集和 PIC 数据集上分别进行了文本区域检测算法性能的系统测试。

在实验过程中,为了保证实验结果的可靠性,将训练文本区域分类器的样本与测试样本分开。首先从 PIC 数据集中选出 138 帧图像,其中包含 524 个文本区域。然后从中随机地选择 200 个文本区域并为其手动增加对比度参数,即示例级的标签信息。这些样本作为训练样本的正例。从 Pascal VOC 数据集中获取的 714 个不含文本的区域全部作为反例。分类器训练过程与 4.2 节相同。相应地,测试数据包含 MS 数据集中的 45 帧图像以及 PIC 数据集中剩余的 319 帧图像。

在候选文本区域检测阶段,使用文献[12]中提出的多尺度视频文本检测方法来获取候选文本区域。实验中使用的滑动窗口大小的变化范围为 10 到 40,步进值为 5。角点强度阈值 t 设为 0.3(正规化到 0 至 1 之间),形态学操作的参数 o = 15。在区域融合阶段,将重合区域大于0.80的区域融合成为一个区域。表 2 分别对‘MS’和‘PIC’两组数据集统计了在进行文本区域鉴别之前该方法在文本区域检测任务中的性能。

在文本区域鉴别阶段每个候选文本区域最终的标签由式(2)决定。为了提高算法的效率,除了使用笔画宽度特征外,本文还使用区域大小和饱和度来过滤过小、明显错误的候选区域。区域大小的阈值设定为 1000(像素点),饱和度的大小设置为0.6。本文将提出的方法与另外两种典型方法进行了对比试验,文本区域检测的召回率和精确度如表 3 所示。

表 2 文本区域检测召回率

实验数据	MS	PIC
召回率	93.42%	98.2%
准确率	71.4%	52.1%

表 3 中的结果显示,本文中提出的算法可以有效地检测不同类别视频帧中的文本区域。算法的准确率优于对比方法,召回率也达到较高水平,其综合性能(F 值)优于同类方法。结合表 2 和表 3 的结果来看,本文提出的文本区域鉴别方法显著地提高了检测的准确率,从而提升了检测系统的整体性能。

表 3　基于 SS-MIL 的文本检测系统在测试数据上的结果

	召回率(%)	准确率(%)	F 值(%)
文献[17]	92.15	87.72	89.88
文献[9]	89.21	88.47	88.83
本文	94.86	86.13	90.28

5　结 论

本文针对视频文本检测问题,提出了一种基于笔画宽度特征的方法来实现更有效的文本区域鉴别。实验结果证明,该特征可以更有效地反映文本区域的特性,因此比目前大多数方法采用的几何特征具有更好的普适性和鲁棒性。此外,本文提出一种新的半监督多示例学习算法来解决文本区域分类器训练过程中监督信息不完整的问题。该方法可以有效地利用训练样本中不完整的监督信息,在降低训练成本的同时提高分类器的性能。本文最终将上述方法与一种具有较高召回率的文本检测方法相结合,实现了一个完整的视频文本检测系统。实验结果表明,该系统可以有效地检测视频中的文本区域,这一结果有力地证明了本文提出的文本区域鉴别方法的有效性。

参考文献

[1] Sharma N, Pal U, Blumenstein M. Recent advances in video based document processing: a review. In: IAPR International Workshop on Document Analysis Systems, Gold Coast, Australia, 2012. 63-68

[2] Ye Q, Huang Q, Gao W, et al. Fast and robust text detection in images and video frames. *Image and Vision Computing*, 2005, 23(6): 565-576

[3] Qian X, Wang H, Hou X. Video text detection and localization in intra-frames of H. 264/AVC compressed video. *Multimedia tools and applications*, 2014, 70(3): 1487-1502

[4] Koo H I, Kim D H. Scene text detection via connected component clustering and nontext filtering. *IEEE Transactions on Image Processing*, 2013, 22(6): 2296-2305

[5] Yi C, Tian Y. Text string detection from natural scenes by structure-based partition and grouping. *IEEE Transactions on Image Processing*, 2011, 20(9): 2594-2605

[6] Chen H, Tsai S S, Schroth G, et al. Robust text detection in natural images with edge-enhanced maximally stable extremal regions. In: Proceedings of the IEEE International Conference on Image Processing, Brussels, Belgium, 2011. 2609-2612

[7] Shivakumara P, Sreedhar R P, Trung Q P, et al. Multi-oriented Video Scene Text Detection Through Bayesian Classification and Boundary Growing. *IEEE Transactions on Circuits and Systems for Video Technology*, 2012, 22(8): 1227-1235

[8] Sharma N, Shivakumara P, Pal U, et al. A new method for arbitrarily-oriented text detection in video. In: IAPR International Workshop on Document Analysis Systems, Gold Coast, Australia, 2012. 74-78

[9] Zhao X, Lin K H, Fu Y, et al. Text from corners: a novel approach to detect text and caption in videos. *IEEE Transactions on Image Processing*, 2011, 20(3): 790-799

[10] Epshtein B, Ofek E, Wexler Y. Detecting text in natural scenes with stroke width transform. In: IEEE Conference on Computer Vision and Pattern Recognition, Providence, USA, 2012. 2963-2970

[11] Hua X S, Wenyin L, Zhang H J. Automatic performance evaluation for video text detection. In: International Conference on Document Analysis and Recognition, Seattle, USA, 2001. 545-550

[12] Zhang B, Liu J F, Tang X L. Multi-scale video text detection based on corner and stroke width verification. In: Visual Communications and Image Processing, Kuching, Malaysia, 2013. 1-6

[13] Everingham M, Van Gool L, Williams C K I, et al. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 2010, 88(2): 303-338

[14] Zhang Q, Goldman S A. EM-DD: An improved multiple-instance learning technique. In: Advances in neural information processing systems, 2006. 1073-1080

[15] 张学工. 关于统计学习理论与支持向量机. 自动化学报, 2000,(01): 36-46

[16] Chang C C, Lin C J. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2011, 2(3): 27

[17] Kim K I, Jung K, Kim J H. Texture-based approach for text detection in images using support vector machines

and continuously adaptive mean shift algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(12): 1631-1639

A text region identification method based on stroke width features and semi-supervised multi-instance learning

Wu Rui*, Du Qingan**, Zhang Boyu*, Huang Qingcheng*

(*Department of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001)

(**Tianjin Institute of Aerospace Electrical Equipment, Tianjin 300000)

Abstract

In consideration of the importance of text region identification to video text detection, a new text region identification method based on stroke width features was proposed. The proposed method can effectively distinguish text regions form non-text regions by analyzing the distribution of the stroke width information in candidate text regions. Moreover, a new semi-supervised multi-instance semi-supervised learning (SS-MIL) algorithm was given to solve the problem that the polar parameter is uncertain in the process of extracting stroke width feature information. The proposed SS-MIL algorithm can improve the performance of region classifier by utilizing incomplete sample labels in training data. A complete video text detection system was implemented based on the proposed methods, and it was tested thoroghty by using the typical data sets such as MCTS. The results showed that the text region identification based on stroke width features and SS-MIL was effective, so the video text detection system achieved the higher overall performance in video test detection.

Key words: text region verification, stroke width, semi-supervised learning, multi-instance learning (MIL)