

复杂背景下文本检测研究

许肖, 顾磊

(南京邮电大学, 江苏 南京 210003)

摘要：在复杂背景下，图像中的文本对于整个图像的语义理解、图像检索、图像识别等应用具有非常重要的作用，有广阔的发展空间，而要获取图像中的文本就需要利用文本检测方法。在搜集整理当前文本检测研究成果的基础上，对文本检测基本方法进行了分类和探讨。分别从基于特征提取、基于机器学习以及基于这两者相结合方法等三个方面对文本检测方法进行详细的阐述；接下来介绍了两种较通用的文本检测结果的评价方法。此外还通过实例分析了各种检测方法的优点和不足，在此基础上，为文本检测的进一步发展提供了建议。

关键词：文本检测;机器学习;特征提取;评价方法

中图分类号：TP311 文献标识符：A

Survey on Text detection under complex background

XU Xiao, GU Lei

(Nanjing University of Posts and Telecommunications, Nanjing Jiangsu 210003, China)

ABSTRACT: Under complex background, the text in the image play a very important role to understanding the image semantic, image retrieval and image recognition application, and to get the text in the image you need to use text detection method. In this paper, the basic methods of text detection are classified and discussed. Illustrate the text detection method in three aspect, feature extraction, machine learning, and combination of both. We also introduces the evaluation method of text detection results. Moreover, a brief analysis is make to introduce the advantages and disadvantages of these three kinds of methods. At last, we point out the shortages of the current methods and puts forward to the development direction of the future.

KEYWORDS: text detection; Machine Learning; Feature detection; Evaluation method;

基金项目：国家自然科学基金项目，编号：61302157

作者一：许肖（1989-），女（汉族），辽宁省葫芦岛市人，硕士研究生，主要研究领域为文本检测。**作者二：**顾磊（1978-），男（汉族），江苏省南京市人，副教授，硕士生导师，主要研究领域为机器学习、模式识别。

0.引言

随着拍摄功能产品的快速发展以及价格的降低,这种数码产品已经渗透进我们的生活。智能手机,平板电脑等移动设备已经允许我们随时随地使用图片保存我们身边的文本信息。而在方便的保存信息的同时,对拍摄到的图片中文本的检测也具有重要意义及应用价值。文本检测是将图片中的文字部分区域定位出来的过程。复杂背景下的文本检测已是图像处理的一个热点和难点。复杂背景下的文本检测有着字体的变换,大小,颜色,方向变换以及类字符图案等等一些干扰,复杂的背景或者光线的变化都会导致图像中文本的失真和变形。

文本检测大致可分为预处理、特征提取、文本定位和后处理等几个步骤,其中,预处理通常是对图像进行增强,降噪等操作,如直方图均衡化用于增强,平滑滤波进行降噪等。特征提取和文本定位则是文本检测的核心部分,其目的是找到图像中文本所在的位置。后处理则是根据特定应用的需要对前面的定位结果进行后续加工,如将文本从图像背景中分离出来以便进行识别等。

本文主要阐述了文本检测的各种方法以及其评判标准。余下部分的组织结构如下:第2节主要介绍了基于特征提取和基于机器学习两种方法并举实例说明,其中基于特征提取方法为传统方式较为成熟,将其分为基于连通域,基于边缘,基于纹理三种方式分别进行阐述;第3节在前一节的基础上将两种基本方法进行了融合,介绍了一些主流和新提出的检测实例;为了得出一个算法好坏的直观结论,第4节介绍了现在通用的评判方式并结合实例对前文提到方法的优缺点进行阐述。最后一节对全文进行了总结,并提出了面临的一些问题以及未来的研究方向。

1.文本检测方法分类

1.1 基于特征提取的文本检测方法

一般来说,为了便于检测,通常需要寻找一些文本特有的特点来区别于复杂的背景,主要的特征如下:

1)几何特征。高度,宽度小于一定值的被检测图像块,很难被人眼识别,基本可以排除是文本区域的可能性。同时,若是水平方向文本,则宽度应大于高度;若是垂直方向文本,则高度大于宽度。

2)颜色信息。为了便于识别并保证人眼睛的视觉效果,即使图像的背景很复杂,颜色的种类很多

并且融合,但字符文本块及其周围背景的颜色却会保持相对单一。

3)边缘特性。这是文本的内在特点,即边缘部分通常比较突出。

4)字符特征。研究表明,文字笔画的宽度与其所属的文本块的宽度的比值有一定的上限,此外,有时字符的间距在同一文本中可能大致相同并遵循一定的约束。

5)纹理特征。可将文本看作是一种特殊的纹理,通过频域和时域的能量特征来确定。

目前,国内外关于复杂背景下的文本检测的方法根据上述特征,主要分为基于连通域,基于边缘,基于纹理等几种。

1.1.1. 基于连通域方法

连通域方法是图像处理技术中的一种常见方法,它是基于图片中文本的颜色灰度一致且有别于背景的颜色灰度,假设图像中字符像素具有相似的像素值,依据每个像素不同的灰度值级别,将图中所有在该灰度值的像素点连通起来,对图像进行粗切分。连通域方法中将文本区域与背景图像区分开主要运用颜色聚类法以及图像阈值分割法。颜色聚类法经常使用于彩色图像分割,通过颜色聚类可以排除掉大块的不含文本的背景区域,然后再利用连通域的特点对提取出的部分进行精确判别。颜色聚类法由于可以快速去掉大块无用区域,所以效率较高。然而去噪和滤波的过程会使边缘模糊不利于进一步的处理,同时在分割时也很难确定合适的聚类数目^[1]。图像阈值分割法是一种自适应的阈值分割法,在二值化图像方面是最具代表性的方法之一,通过计算其周围的像素灰度值的均值与方差来确定每个像素的分割阈值^[2]。Niblack 是一种非常有效的分割图像方法,其每个像素的阈值计算公式为:

$$T(i, j) = m(i, j) + k * \sigma(i, j) \quad (1)$$

在式(1)中, $k=0.2$, $m(i, j)$ 和 $\sigma(i, j)$ 分别为以 (i, j) 为中心的矩形的像素灰度值的均值和方差。但是图像阈值分割法的分割阈值不容易确定,且对噪声敏感。

近年来,一些基于连通域的方法不断被提出。例如:2009年文献[3]提出了一种基于 niblack 的检测方法,该方法表明 niblack 的输出方法是重要的,也是在文本分割中最好的阈值分割技术。2010年文献[4]提出了一种结合颜色信息与小波分解的文本检测方法,其中小波的分解是依赖于三种颜色

R, G, B 中包含的高频子带 (LH, HL 及 HH)。2013 年文献[5]提出了一种多通道连通域分割的方法, 通过颜色, RGB 通道来使用马尔科夫随机场。

1.1.2. 基于边缘的方法

边缘是图像最基本的特征之一, 图像的很大一部分信息都存储于此。边缘就是图像灰度发生剧烈变换的位置, 通常存在于文本与背景之间, 边缘检测的质量直接影响着检测系统的性能。图像依据类型可分为灰度图像边缘检测及彩色图像边缘检测两种方法。

1) 对于灰度图像, 由于图像中文字的边缘具有方向性的特点, 因此常用一阶或二阶导数来描述边缘, 检测算子则分别为一阶微分边缘检测算子和二阶微分边缘检测算子。典型的一阶微分算子有基于梯度的 Robert 算子, Sobel 算子。典型的二阶微分算子有 LOG (Laplacian of gaussian) 算子, Canny 算子。

2) 彩色图像的检测可分为标量运算法和矢量运算法两类。标量法可分为灰度法以及三通道合成法, 矢量法则有多维梯度法等。

近年来, 基于边缘的检测方法不断涌现出来, 2009 年文献[6]提出一种基于拉普拉斯算法的方法, 首先使用拉普拉斯过滤图片中的每个像素, 然后使用聚类的方法将像素分辨出是否为文本, 最后用 Sobel 算子确定文本块的边缘。2010 年文献[7]使用了基于过滤器和边缘分析的启发式规则来分类, 将视频图像分成高对比度和低对比度图像, 用于精确检测文本行的边缘。实验表明, 所提出方法比现有方法在检测率等方面有所提高。2011 年文献[8]提出了一种基于 MSER 边缘检测的方法, 再经过几何平均滤波, 最后利用笔画宽度确定文本。该算法可以较好的与可视化系统融合。

1.1.3. 基于纹理的方法

纹理是一种区域性的特征, 根据像素值的不同以及基元分布规律相互区别。可将纹理特征提取大致分为以下四种^[9]: 统计分割算法, 模型分割算法, 频谱分割算法和结构分割算法。纹理分割法可以依据图片中的不同纹理将其分割成不同的区域, 即可在图片中将具有文本纹理的区域分离出来。

1) 统计分割算法: 是利用像素及其邻域区域中灰度的分布来描述纹理的^[10], 包括基于一阶, 二阶或高阶的统计纹理描述, 有灰度共生矩阵 (gray level co-occurrence matrix, GLCM), 半方差图法等等。2) 模型分割算法: 模型分割算法主要是将

文本部分图像抽象成模型, 然后将这个模型进行特征提取, 总结估计出一些特征参数, 假设纹理就是由该特征参数分布形成的, 在以此参数为由进行图像的分割。3) 频谱分割算法: 该算法是对图像进行某种变换 (如傅里叶变换) 或用滤波器, 滤波器组对图像滤波, 得到图中每个像素变换后的结果, 然后用能量等特征值描述出像素点并提取纹理特征, 最后进行分割。4) 结构分割法: 假设图像是由多种不同类型的基元以不同的方向和数目组成的, 这需要对纹理进行精确的定义。结构分割法适用于规律性很强的纹理, 因此其使用受到了很大限制。

近年来, 基于纹理的方法处于不断的发展和革新中。2006 年文献[11]使用纹理方法来检测文本笔画。该方法首先使用小波多分辨率分解, 然后使用阈值像素标记, 最后使用纹理检测法确定文本精确位置。实验证明该方法具有较高效率。2010 年文献[12]提出了一种新的文本检测算法。在自然场景中, 我们提取笔画的纹理特征来定位文本行, 然后使用角点检测来精确定位文本。实验结果证明该方法可以准确定位文本并具有鲁棒性。2011 年文献[13]提出了一种基于频率领域的拉普拉斯方法来检测视频文本。输入的图片首先经过傅里叶-拉普拉斯滤波, 然后使用聚类方法来识别备选区域。该方法能够处理任意方向的文本, 并且不论人工场景还是自然场景都能得到较好效果。

1.2 基于机器学习的方法

随着机器学习方法的发展, 如神经网络 (Neural Networks, NN), 人工智能, 支持向量机 (support vector machine, SVM), Adaboost 等技术越来越多地被应用于文本检测中。基于机器学习的方法可以考虑实际的变化情况, 通过对实例的学习得到模型的参数, 还可通过增加学习的样本来扩充检测模式的范围, 使系统具有较高的鲁棒性。

神经网络以人脑的活动规律以及组织结构为背景, 模拟人的一部分形象思维的能力, 具有高速并行, 自适应, 非线性映射等特点, 但其存在拓扑结构难确定以及存在局部最优等问题。Adaboost 是一种基于统计学习的算法, 在学习中可不断根据正例和反例之间确定好的特征, 以及所起的作用进行权值的调整, 通过训练集训练若干不同的具有低错误率的弱分类器, 然后将弱分类器集合起来成为一个强分类器。具有算法简单, 样本易提取, 计算耗时少等优点。支持向量机是一种可训练的机器学习方法, 起源于统计学习理论, 由结构风险最小化

理论 (Structural Risk Minimization, SRM) 作为支持。通过将线性不可分的样本由非线性映射算法从低维映射到高维, 变为线性可分的样本, 再利用已知的算法来得到目标算法的全局最小值, 使得全局最优化^[14]。

基于这些常用的机器学习技术的文本检测方法近年发展迅速, 并趋于主流。例如: 2008 年文献[15]首先对初始文本边界使用滑动窗口生成一些样本, 然后将这些样本使用 SVM 分类器进行分类, 该方法能够准确的确定文本边界, 可以达到较好的召回率和精确度。2009 年文献[16]提出了完备的灰度图像的检测系统。所提出的方案使用能将小规模的异构特性融合形成较大规模的特性。基于神经网络的检测器通过学习规则来进行定位。该系统可以在复杂背景中定位到不同字体不同风格的文本。2011 年文献[17]利用基于简化的耦合脉冲神经网络方法, 其结果在检测精度上优于传统的检测方式。

2 特征提取与机器学习相结合的方法

单纯的使用基于特征提取或基于机器学习的文本检测方法往往得不到令人满意的效果。近年来, 越来越多的学者关注于将两者结合起来使用, 以提高系统的性能。

2010 年 Z. Li^[18]等人提出了一种使用关键点(key text points, KTPs)的有效提取方法。KTPs 可以通过小波变换后在高频区域获得, 纹理方向投影方法可以使定位更加精确。文本追踪可以减少背景的干扰。最后计算连通域中的 KTPs 的总数将文本提取并分割。实验证明所提出的检测和定位的方法具有鲁棒性。文本追踪显著地提高了文本检测的效率, KTPs 减少了背景变换的影响。

2011 年 Lukas Neumann^[19]等人提出了一种端到端的实时的文本定位识别方法。在分类的第一阶段, 判断每个 ER 能否使用只有 O(1) 复杂度的新算法来得到特征值, 在第二阶段使用选出的局部最大的 ER, 最后使用计算复杂的 AdaBoost 分类器进行分类。实验最终得到了较好的效果。

2011 年 Matas^[20]等人利用文本的高阶属性提出了一种有效的文本定位识别方法。该方法证明了分组阶段在文本识别中的重要性, 并且正确精确的分组还能够弥补字符检测时产生的错误。该方法使用一种新的最大稳定极值区域(MSER)选择器, 最后使用 SVM 字符分类器来决定是否为文本区域。使用 ICDAR 的评价标准得到了较高的精确度。

2011 年 Hongwei Zhang^[21]等人提出了一种两阶段的条件随机场(CRF)迭代算法, 该算法包含使用信度传播算法(BPA)以及一个 OCR 滤波阶段。第一个 CRF 迭代阶段目的在于找到文本连通区域, 然后将无法确定是否是连通域的区域送至第二阶段迭代。第二个 CRF 迭代阶段在 OCR 的帮助下滤除非连通域区域。实验证明, 对多个文本行的提取以及区分文本与背景噪声有良好的效果。使用 ICDAR 的评判标准得到较好的效果。

2011 年 Jiangqiang Yan^[22]等人提出了一种结合使用 Gabor 滤波器和支持向量机(SVM)的新方法, 该方法基于汉字由四种笔画组成这个条件。通过使用 Gabor 滤波器提取四种笔画特征, 汉字的文本定位可以转换成纹理的分类, 这就可以使用 SVM 分类器来达到目的。所提出的方法有两个步骤, 首先, 使用不同尺度和方向的 Gabor 滤波器来获得可以代表汉字横竖撇捺的纹理图像。然后训练对应的四种 SVM 分类机来区分四种方向的纹理, 根据权重之和来确定是否为文本区域。通过与现有方法对比, 其对中文的识别能力较强

3. 评估方法与算法比较

3.1 评估方法

对各种不同的检测方法来讲, 一种客观有效的标准是十分必要的, 目前的性能评估主要分为两种:

第一种是国际会议 ICDAR 所提出的评估方法^[23]:

$$R_{tr} = \frac{b}{|T|}, P_{tr} = \frac{b}{|E|} \quad (2)$$

其中, R_{tr} 指召回率, P_{tr} 指精确度, T 为真实数据目标集合, E 为测试系统返回集合, b 为正确估计。两个检测结果的矩形框 r_1, r_2 的匹配分数 m_p 的定义如下:

$$m_p(r_1, r_2) = a(r_1 \cap r_2) / a(r_1 \cup r_2) \quad (3)$$

其中, $a(r_1 \cap r_2)$ 为相交矩形的面积, $a(r_1 \cup r_2)$ 为两矩形并集的最小外接矩形的面积, 一个矩形 r 在一个目标矩形集合 R 中的最佳匹配分数 $m(r, R)$ 的定义为 r 与 R 中所有矩形相匹配所得的

匹配分数的最大值。公式表示如 4，用匹配分数定义召回率 R_{IR} 和精确度 P_{IR} ，如 5：

$$m(r, R) = \max(m_p(r, r') \mid r' \in R) \quad (4)$$

$$p = \frac{\sum_{r \in E} m(r, T)}{|E|}, \quad r = \frac{\sum_{r' \in E} m(r', E)}{|T|} \quad (5)$$

其中， r_e 是检测出的矩形框； r_t 是真实文本矩形框。匹配面积为：

$$m_a(r, r') = \frac{2a(r \cap r')}{a(r) + a(r')} \quad (6)$$

综合考虑召回率和精确度之后的评价结果：

$$f = \frac{1}{\frac{\alpha}{p} + \frac{1-\alpha}{r}} \quad (7)$$

α 是表现 p 和 r 的重要性的加权系数。由于 ICDAR 的评估方法过于苛刻与复杂，并且现行方法大多检测出的是文本行而并非 ICDAR 中的单词，于是一些文本检测方法选择采用基于区域级评估方法进行评估^[24]，如下为评估标准：正确检测区 (TDB)：一个被检测到的区域至少包含一个正确字符。错误检测区 (FDB)：一个被检测到的区域没有包含文本。不完整检测区 (MDB)：一个被检测到的区域丢失一行文本 20% 以上的字符 (MDB 是 TDB 的子集)。

$$\begin{aligned} \text{Recall (R)} &= \text{TDA/ATB}, \\ \text{Precision (P)} &= \text{TDB/ATB}, \\ \text{F-measure (F)} &= 2 * P * R / (P + R), \\ \text{Misdetction Rate (MDR)} &= \text{MDB/TDB}. \end{aligned} \quad (8)$$

3.2 算法比较

通过给每个参赛者的可执行软件输入测试数据，生成一个原始结果文献，再用评价算法将该结果与数据集的真实参考数据进行比较，得到使用 ICDAR 评价方法的结果如表 1，使用基于区域级评价方法的结果如表 2：

表 1 使用 ICDAR2003 方法得出结果

Method	P(precision)	R(recall)
Kim ^[25]	0.56	0.64
Choi ^[26]	0.56	0.69
Pan ^[27]	0.68	0.75

表 2 使用区域级评价方法结果

Method	R	P	F
Edge1 ^[28]	0.53	0.61	0.57
Edge2 ^[29]	0.67	0.33	0.44
Color ^[30]	0.60	0.44	0.45
Gradient ^[31]	0.52	0.83	0.64

从上述表 1 表 2 中可以看出：1) 基于连通域的方法对字体不敏感。当文本条件比较好时，能够准确定位文本区域取得比较好的效果。但当背景与字符边缘模糊或图像分辨率较低时，得不到理想结果，且很多地方需依赖经验。2) 边缘检测有三种主要误差：对正确边缘的漏检，误将噪声等非边缘判断为边缘，边缘定位偏差。即使使用预处理技术事先进行平滑处理，但是也会因为平滑范围较大而导致边缘模糊化，定位不准确。尽管如此，基于边缘的方法其优势在于检测速度较快，对于背景不复杂或笔画边缘突出的图像，检测效果较好。3) 纹理方法具有较好的鲁棒性，可以很好的处理噪声，识别出文字与背景对比度较小的图片，但是识别不到文本的精确位置和大小，并且对字符数量较少的文本区域难以定位，计算较复杂。4) 通过对实例的学习得到模型的参数，可通过增加学习的样本来扩充检测模式的范围，使系统具有较高的鲁棒性。但算法的性能容易受制于分类器的性能和推广能力，所以需要与基于特征提取方法相结合，可以提高系统性能。

4. 结束语

本文详细介绍了用于复杂背景下文本检测的主流方法，分析了各个方法的利弊，举出了该方法的常用算法，举例一些典型的综合各个方法的检测案例，最后介绍了几种评价检测优劣的方式。目前许多检测方式已趋于成熟，综合多种方法对一些特定的场景也有较好的定位效果。但是对于背景多变，文本与背景融合等较复杂情况的检测效果还不理想。随着人工智能等其他学科的发展与学科间的融合，将改进系统性能，使其能应用于实际的检测当中。因此下一步的工作将以以下几点为核心：

1) 由于背景较为复杂，使用单一的方式进行文本检测往往达不到理想的效果，需将多种方法结合起来。如何将多种算法合理的融合在一起一同进行文本检测是需要进一步研究的。2) 涌现出许多新方法。如特征选择，特征提取，新型机器学习方法等，将其融合应用于文本检测中，创造出更好的

检测效果。3) 一些好的图像分割的方法, 如水平集方法, 方法型模型等。可将其移植应用于复杂背景下的文本检测, 发挥其优势。4) 现在很多算法复杂, 只存在于理论并不适用于应用在实际生活中, 如何将其广泛的使用在生活中, 使我们的生活更加便利是未来需要探讨钻研的。

参考文献:

- [1]张赫.基于颜色聚类 and 连通域分析的视频文本提取方法[D].北京:北京邮电大学.2010.
- [2]秘新浩. 基于连通域的文本定位方法研究[D].浙江: 浙江工业大学.2007.4
- [3]S.Farid. Application of Niblack's Method on Images [C]. International Conference on Emerging Technologies, 2009
- [4]Palaiahnakote Shivakumara.New Wavelet and Color Features for Text Detection in Video[M]. International Conference on Pattern Recognition,2010.
- [5]Xiaobing Wang. Natural scene text detection with multi-channel connected component segmentation[M]. 12th International Conference on Document Analysis and Recognition,2013
- [6]Trung Quy Phan. Laplacian Method for Video Text Detection [C]. 10th International Conference on Document Analysis and Recognition, 2009
- [7]Palaiahnakote Shivakumara. Accurate video text detection through classification of low and high contrast images [J]. Pattern Recognition, 2010(43):2165-2185
- [8]Huizhong Chen, Sam S. Robust text detection in natural images with edge-enhanced maximally stable extremal regions [C]. 18th IEEE International Conference on Image Processing, 2011
- [9]余鹏, 封举富.基于高斯混合模型的纹理图像分割[J].中国图象图形学报.2005, 10(3):5-6
- [10]LIU Li, KUANG Gang-yao. Overview of Image Textural Feature Extraction Methods. Journal of Image and Graphics [J].2009, 14(4):2-6.
- [11]Caifeng Zhu. Text Detection in Images Using Texture Feature from Strokes [C]. 7th Pacific Rim Conference on Multimedia. 2006:296-300.
- [12]Xiaodong Huang. Automatic detection and localization of natural scene text in video [C]. 20th International Conference on Pattern Recognition.2010:3216-3219
- [13]Palaiahnakote Shivakumara. A Laplacian Approach to Multi-Oriented Text Detection in Video [J]. IEEE transactions on pattern analysis and machine intelligence. 2011(33):412-418
- [14]Lienhart R, Axel W. Localizing and Segmenting Text in Images and Videos [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(4): 256-268.
- [15]Cheolkon Jung, Qifeng Liu. SVM output score based text line refinement for accurate text localization [C]. International Conference on Intelligent Computing and Integrated Systems, 2008.
- [16]Shehzad Muhammad Hanif. Text Detection and Localization in Complex Scene Images using Constrained AdaBoost Algorithm[C]. 2009 10th International Conference on Document Analysis and Recognition, 2009
- [17]ZHANG Xin, SUN Fuchun. Pulse Coupled Neural Network Edge-Based Algorithm for Image Text Locating [J]. Tsinghua Science and Technology, 2011, 16(1):22-30
- [18]Z. Li G. Liu X. Effective and efficient video text extraction using key text points[C]. IET Image Processing Received, 2010
- [19]Luka's Neumann. Real-Time Scene Text Localization and Recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2012.
- [20]Matas. Text Localization in Real-world Images using Efficiently Pruned Exhaustive Search [C]. International Conference on Document Analysis and Recognition, 2011.
- [21]Hongwei Zhang. An improved scene text extraction method using Conditional Random Field and Optical Character Recognition [C]. International Conference on Document Analysis and Recognition, 2011.
- [22]Jianqiang Yan, Jie Li. Chinese text location under complex background using Gabor filter and SVM [J]. Neurocomputing, 2011.
- [23]S.M.Lucas, A. Panaretos, L. ICDAR 2003 robust reading competitions [C]. In Proceedings of International Conference on Document Analysis and Recognition, 2003:682-687.
- [24] D CHEN, J M ODOBEZ, J P THIRAN. A localization/verification scheme for finding text in images and video frames based on contrast independent features and machine learning [J]. Image Communication. vol. 19, pp. 205-217, 2004.
- [25]J. Kim, S. Park, and S. Kim. Text locating from natural scene images using image intensities. In Proceedings of eighth International Conference on Document Analysis and Recognition, volume 2, pages 655-659, 2005.
- [26]Y. Choi. Scene text extraction in natural images using

hierarchical feature combining and verification. In the 2nd KAIST-Tsinghua Joint Workshop on Pattern Recognition, pages 76–102, 2003.

[27]Wumo Pan. Text Detection from Scene Images Using Sparse Representation[C]. Pattern Recognition, 2008

[28]C LIU, C WANG, R DAI. Text detection in images based on unsupervised classification of edge-based features [C]. Document Analysis and Recognition, pp. 610-614, 2005.

[29]M CAI, J SONG, M R LYU. A new approach for video text detection [C]. Image Processing, 2002 International Conference on .pp. 117-120, 2002.

[30]V Y MARIANO, R KASTURI. Locating uniform-colored text in video frames [C]. Pattern Recognition, 15th International Conference on. pp. 539-542, 2000.

[31]E.K. Wong and M. Chen, “A New Robust Algorithm for Video Text Extraction,” Pattern Recognition, vol. 36, pp. 1397-1406, 2003.