

基于视觉显著性与文字置信图的场景 文字的背景抑制方法

姜 维, 卢朝阳, 李 静, 刘晓佩, 姚 超

(西安电子科技大学综合业务网国家重点实验室, 陕西西安 710071)

摘 要: 复杂背景是影响场景文字自动定位算法性能的重要因素. 为消除背景干扰, 本文首先根据谱残差理论获取显著性区域, 然后利用基于笔画特征的点对梯度方向差比例图与基于傅里叶谱统计特征的后验概率图取得显著性区域的文字置信图, 最后借助图模型融合视觉显著性图、文字置信图与 HSI 颜色特征最终达到抑制自然图像背景并突出前景文字的效果. 在 ICDAR2011 场景文字定位竞赛数据库与实验室场景中文字数据库中的实验结果表明, 本算法可较好的抑制自然图像中复杂背景, 并有效提升算法的性能.

关键词: 场景文字; 背景抑制; 视觉显著性; 笔画特征; 傅里叶谱统计特征

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2015)01-0062-07

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2015.01.010

Visual Saliency and Text Confidence Map Based Background Suppression for Scene Text

JIANG Wei, LU Zhao-yang, LI Jing, LIU Xiao-pei, YAO Chao

(State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an, Shaanxi, 710071, China)

Abstract: The complex background is an important inference factor of scene text automatic localization algorithm. To eliminate the impact of background, this scheme firstly obtains the saliency region in saliency map according to the spectral residual theory. Then it computes text confidence map in saliency region, which is weighted sum of the fraction map of corresponding pairs based on stroke feature and the posterior probability map based on Fourier spectrum statistical feature. Finally, it combines saliency map, text confidence map and HSI color feature with the graph model to suppress the background and highlight the foreground text. The experimental results on both the ICDAR2011 scene text localization competition test dataset and laboratory Chinese dataset show that the scheme can effectively suppress the complex background and improve the scene text localization algorithm.

Key words: scene text; background suppression; visual saliency; stroke feature; Fourier spectrum statistical feature

1 引言

近年来, 信息技术虽然迅速发展, 但在目标检测与识别领域仍难以超越人类, 所以关于视觉显著性的研究一直是热点问题, 近年来受到研究者更多关注^[1,2].

场景图像中文字自动定位是计算机视觉领域中的一个重要但尚未解决的问题. 主流算法通常分为基于纹理的方法与基于区域的方法. 基于纹理的方法^[3,4]假设文字具有明显不同于背景的纹理与结构, 首先将自然场景图像分割为块状的区域, 然后使用分类器根据纹理特性去除背景, 再根据某些原则聚合文字; 基于区域的方法^[5,6]假设文字根据某种特性与周围是独立可分割的,

首先根据颜色、笔画宽度或边缘等特性进行连通域分析获取文字候选区, 再根据某些特征消除背景区域, 最后将文字候选区域聚合成文本行.

这两类算法各有优劣, 但均会受到复杂背景的干扰. 抑制自然图像的背景并突出前景文字是解决该问题的方法之一, 也是本文的研究目标. 基于此目标, 本文引入能够抑制背景并突出目标的视觉显著性理论. 目前, 基于视觉显著性的场景文字自动定位算法^[7,8]均采用自底向上的方法, 该方法需计算图像中所有像素的多种特征的显著性值并将其融合, 计算量较大. 本文选择的是基于频域的谱残差视觉显著性方法, 该方法针对图像做全局性变换, 计算量较小. 文献^[7,8]提出的是基于

视觉显著性的场景文字自动定位算法,本文目标是设计一种用于提升场景文字自动定位性能的预处理算法,二者研究目标并不尽相同.目前仅有文献[9]针对场景文字提出背景抑制算法,该算法使用图模型结合颜色、边缘和纹理信息完成场景文字背景的背景抑制.

与文献[9]相比,本文尝试利用谱残差视觉显著性,模仿人类视觉处理过程设计算法.人类的视觉处理过程分为两步^[10]:首先进行快速简单的全局并行预注意过程,快速忽略背景,获得显著目标;然后进行较慢的复杂的串行注意过程,有意识去除无效目标,突出感兴趣目标.人类感知场景文字的过程也符合以上步骤,故据此设计算法如图1所示,大致分为三步:(1)根据谱残差理论获取输入图像的显著性图,进而得到显著性区域;(2)在显著性区域中,利用笔画特征与傅里叶谱统计特征产生点对梯度方向差比例图与后验概率图,结合产生文字置信图;(3)通过图模型结合显著性图、文字置信图及 HSI 颜色特征完成针对场景文字的背景抑制.步骤(1)属于并行预注意过程,本文第2部分着重讲述该步骤;步骤(2)(3)属于串行注意过程,本文第3部分与第4部分说明这两步骤.

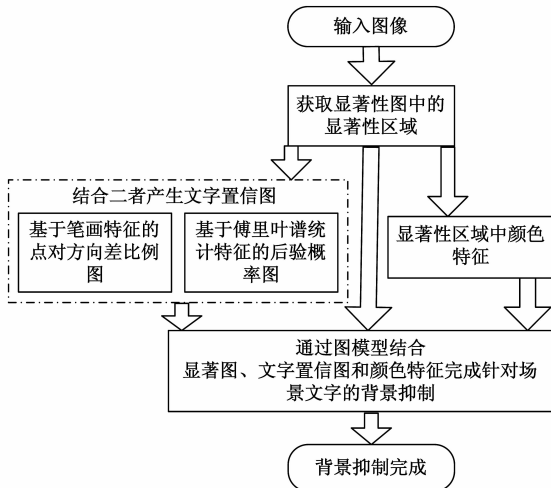


图1 算法流程图

2 谱残差视觉显著性图与显著性区域

谱残差视觉显著性算法^[10]是快速可靠且无需先验知识的显著性算法,它分为两步:首先,对图像傅里叶幅度对数谱进行卷积均值滤波;然后,从图像傅里叶幅度对数谱中减去上一步的均值滤波结果,最终得到显著性图 S . 式(1)描述谱残差视觉显著性模型的求解, $A(I)$ 表示图像的傅里叶幅度谱, $\log(A(I))$ 表示图像的傅里叶幅度对数谱, $h(I)$ 表示均值滤波.

$$S = \log(A(I)) - h(I) * \log(A(I)) \quad (1)$$

自然场景的文字多为标语、广告和指示牌等,普遍

显著而突出.文献[11]表明,人类对人脸与文字的快速准确定位能力,至少部分依赖于图像傅里叶幅度谱中的信息,而谱残差视觉显著性算法正是基于图像傅里叶幅度谱.因此,本文采用此显著性算法针对场景文字进行背景抑制,是可行且有依据的.

图2是利用谱残差视觉显著性分析得到的场景文字显著性图.其中所使用图像分别来自 ICDAR2011 场景文字定位竞赛数据库与实验室场景中文数据库.图2(a)为原图,图2(b)是对应的谱残差显著性图,图像亮度代表显著性程度.图2中,谱残差视觉显著性算法成功抑制场景图像中的非显著区域,并突出包含文字的显著区域,但同时保留了环境中其他显著元素,本文后续的文字置信图阶段可有效抑制干扰元素.



图2 谱残差视觉显著性分析样图

3 文字置信图

在显著性区域中,首先进行笔画宽度变换^[12](Stroke Width Transform)获取文字候选连通域,然后针对文字候选连通域分别获得基于笔画特征的点对梯度方向差比例图^[13]与基于傅里叶谱统计特征的后验概率图,结合二者最终得到文字置信图.笔画特征描述的是文字的细节信息,傅里叶谱统计特征描述的是文字整体情况,这两种特征符合人类对文字细节与整体两方面的认知.

3.1 基于笔画特征的点对梯度方向差比例图

文字的笔画通常是由双边缘构成的闭合完整区域,其内部灰度变化很小.笔画的双边缘上会有梯度方向相反且幅度大小相近的点对,这些点对的梯度方向差可以较好描述文字的笔画特性.据此,本文采用文献[13]的点对梯度方向差比例图.

式(2)中 $G_{\text{angle}}^{(i)}$ 表示点对 i 的梯度方向差, α 是方向

角度差的阈值^[13].式(3)中 F_{corr} 是统计满足阈值 α 的点对数量占总点对数量的比例,其值越大表明该区域是文字的可能性越大,相反则越小.

$$h(G_{\text{angle}}^{(i)}, \alpha) = \begin{cases} 1, & \text{if } G_{\text{angle}}^{(i)} \leq \alpha \\ 0, & \text{else} \end{cases} \quad (2)$$

$$F_{\text{corr}} = \frac{\sum_{i=1}^N h(G_{\text{angle}}^{(i)}, \alpha)}{N} \quad (3)$$

3.2 基于傅里叶谱统计特征的后验概率图

研究表明自然图像的傅里叶幅度频率谱的 log-log 曲线近似符合线性关系的^[14],基于此结论本文发现自然图像的傅里叶幅度频率谱的 log-log 曲线用低频与高频两段线性关系描述则更能反映曲线特性.图 3(b)中,曲线代表的是自然图像的傅里叶幅度频率谱的 log-log 曲线,高低频分割点 P 将其切分为左右两部分,左半部分是图像低频部分,右半部分代表图像高频部分,将二者分别用直线拟合比单直线拟合整条曲线更能反映曲线低高频两部分的不同特性.

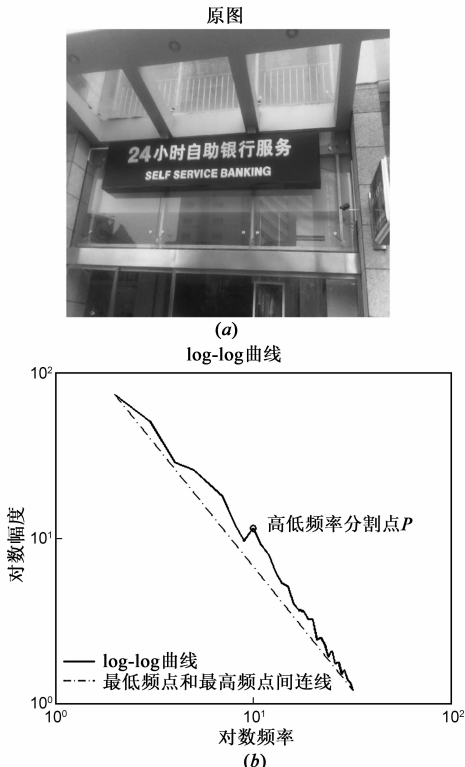


图3 自然图像傅里叶谱log-log曲线图

求解高低频分割点 P 首先需找到傅里叶幅度频率谱的 log-log 曲线中频率最低点与频率最高点,根据两点确定一条直线;然后,遍历傅里叶幅度频率谱 log-log 曲线中所有点,寻找距离该直线最远的点,即高低频分割点 P .

获得高低频分割点后,分别用直线拟合傅里叶幅度频率谱 log-log 曲线的低频与高频部分以及整条曲

线,可知低频部分拟合直线斜率 k_1 、高频部分拟合直线斜率 k_2 与整条曲线拟合直线斜率 k_3 ,三者组成傅里叶谱统计特征.为了验证该特征是否可用于区分文字与背景,本文采集 1159 个文字样本与 1485 个背景样本,文字样本来自 ICDAR2011 场景文字定位竞赛数据库和实验室场景中文字数据库中的文字区域,背景样本来自以上二者的背景区域.本文分别对文字和背景样本求斜率 k_1, k_2 与 k_3 ,最终绘图得到图 4.从图 4 可看出,“+”代表的文字与“○”代表的背景是基本可分的.因此本文采用朴素贝叶斯方法计算文字候选区的后验概率图,特征 $f = (k_1, k_2, k_3)$,类别 $c \in \{\text{text}, \text{bg}\}$,其中 text 表示文字类别, bg 表示背景类别.基于傅里叶谱统计特征的后验概率如式(4)所示,其中 $p(\cdot)$ 代表对应的概率分布.先验概率 $p(c)$ 从训练样本计算得出.

傅里叶谱统计特征的文字与背景样本分布图

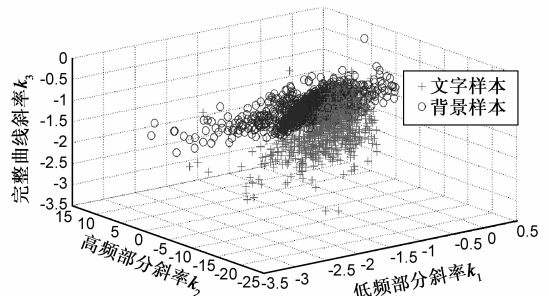


图4 傅里叶谱统计特征的文字与背景样本分布图

$$p(\text{text} | f) = \frac{p(\text{text}) \prod_i^3 p(k_i | \text{text})}{\sum_{c \in \{\text{text}, \text{bg}\}} p(c) \prod_i^3 p(k_i | c)} \quad (4)$$

依据式(4)求出特征 f 对应的属于文字类别的后验概率,再将其按照连通域映射到原图中得到基于傅里叶谱统计特征的后验概率图.

3.3 文字置信图

基于笔画特征的点对梯度方向差比例图描述的是文字的细节信息,基于傅里叶谱统计特征的后验概率图描述的是文字的整体情况.综合考虑文字细节与整体情况,本文对二者加权求和得到文字置信图,权重分别是 $w_1 = 0.5$ 与 $w_2 = 0.5$.权重是在训练集中以 0.01 为步长,遍历 $[0, 1]$ 区间,取性能最优的权重得到.

文字置信图是基于笔画宽度连通域计算得到的,图 5 中文字置信图阶段的效果图,图 5(b)是基于笔画特征的点对梯度方向差比例图,图 5(e)是剔除比例值小于 0.75 的像素的结果,图中树叶与草丛等杂乱区域因此被较好去除;图 5(c)是基于傅里叶谱统计特征的后验概率图,图 5(f)是剔除后验概率值小于 0.75 的像素的结果,图中既去除部分杂乱区域又消除类文字的墙壁纹理区域;图 5(d)是二者结合的文字置信图,图 5

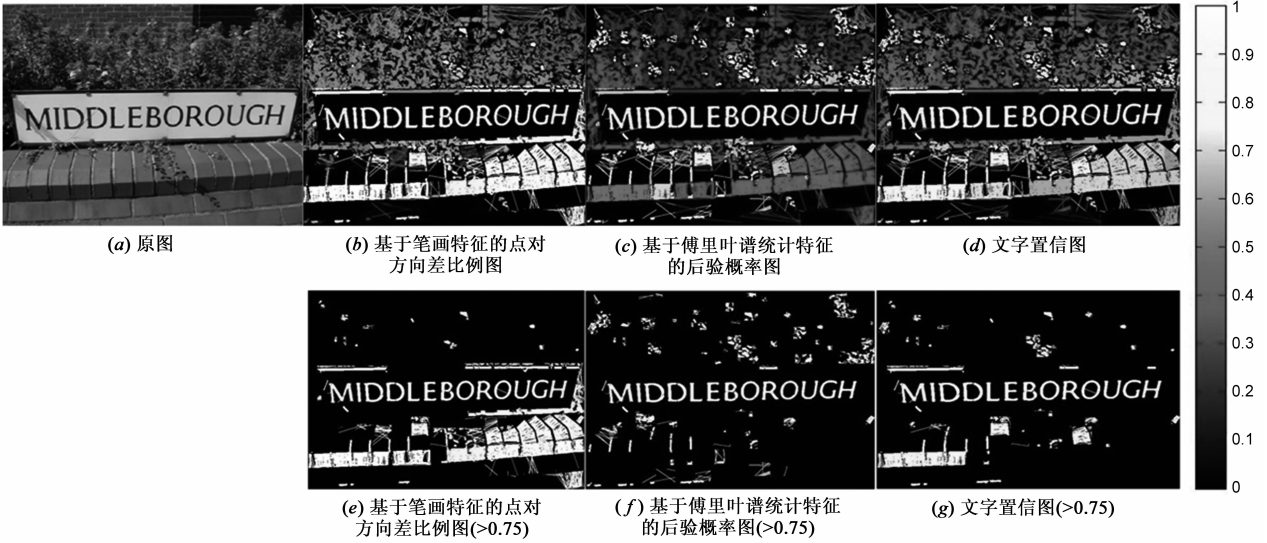


图5 文字置信图阶段各效果图

(g)是保留图 5(d)中文字置信度大于 0.75 的像素的效果图,该图很好结合基于笔画特征的点对梯度方向差比例图与基于傅里叶谱统计特征的后验概率图,起到了背景抑制效果。

4 图模型

本文算法尝试用无向图模型 $G = \{V, E\}$ 集合显著性图、文字置信图与 HSI 颜色特征。 V 是图模型中的节点,代表图像像素; E 是图模型中的无向边,由像素的标准 8 邻域构成。利用图模型融合显著性图、文字置信图与颜色特征是一个类似图像分割问题的像素标注过程,标注结果是针对场景文字的背景抑制二值图,其中文字像素被标注为 1,背景像素被标注为 0。

式(5)中,图模型代价函数 $E(A)$ 由前半部分的一元代价函数与后半部分的二元代价函数组成^[15]。 $A = (A_1, \dots, A_p, \dots, A_{|P|})$ 是一组二进制向量,是像素集合 P 对应的标注值, $|P|$ 是像素集合中像素的数目, N 是像素的标准 8 邻域集合。式(6)中, S 是像素的显著性值, T 是像素的文字置信值, β 与 α 分别是二者的权重, σ 是精度系数。式(7)中, H_p 与 H_q 代表邻域像素的 HSI 空间的色度。 α 、 β 与 σ 通过交叉验证分别设定是 0.4、

0.6 与 0.25。

$$E(A) = \sum_{p=1}^P U_p(A) + \sum_{|p,q| \in N} B_{|p,q|}(A) \quad (5)$$

$$U(A) = \begin{cases} \beta \cdot \exp(-\frac{S^2}{2 \cdot \sigma^2}) + \alpha \cdot (1 - T), & A = 1 \\ \beta \cdot \exp(-\frac{(1-S)^2}{2 \cdot \sigma^2}) + \alpha \cdot T, & A = 0 \end{cases} \quad (6)$$

$$B_{|p,q|}(A) = \sum_{|p,q| \in N} \sum_{p \in P} \|H_p - H_q\|^2 \quad (7)$$

式(5)前半部分描述像素本身的特性,在本文算法中是显著性图与文字置信图;后半部分对应的二元代价函数描述的是当前像素与 8 邻域之间在 HSI 颜色空间中是否具有相同的色度,若色度差别较小则二元代价函数较小。本文的图模型求解采用文献[16]的最大流/最小割算法。

图 6 是算法阶段效果图,图 6(b)是显著性图,图 6(c)是图 6(b)二值化后得到的显著性区域内的文字置信图,图 6(d)是最终的针对场景文字的背景抑制效果图,可看出本文算法在经过前文提到的三个步骤去除场景中杂乱区域与类文字的墙壁纹理区域,达到了针对场景文字的背景抑制的效果。

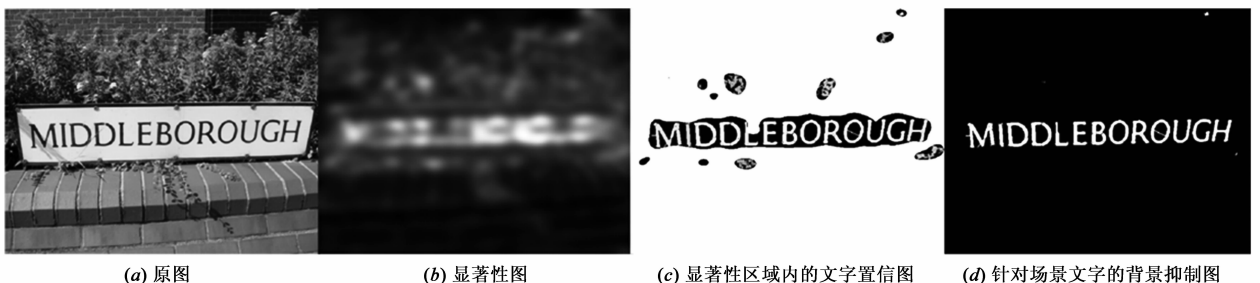


图6 算法阶段效果图

5 实验与分析

5.1 评价标准

场景文字自动定位算法通常采用 ICDAR2011 评价标准,但本文算法是定位算法的预处理部分,直接采用该评价标准并不合适.而至今唯有文献[9]与本文研究问题类似,故采用该文献提出的评价标准,如式(8)~(11)所示.

$$PCR = N_{\text{text}}/N \quad (8)$$

$$RR = DTB/TB \quad (9)$$

$$RP = DTB/(DTB + DNB) \quad (10)$$

$$RF = 2 \times RP \times RR / (RP + RR) \quad (11)$$

文献[9]的评价标准分为像素级的评价标准与区域级的评价标准.式(8)是像素级的评价标准,其中 PCR 代表的是文字像素的检测率, N_{text} 是检测到文字像素的数量, N 是检测到所有像素的数量. PCR 表明检测到的区域中文字像素含量的高低,优秀的背景抑制算法的 PCR 应具有较高水平.式(9)~(11)是区域级的评价标准.式(9)中 DTB 是检测到的文字区域个数, TB 是真实文字区域的个数, RR (Region Recall) 代表检测到的文字区域召回率,因为背景抑制算法是场景文字自动定位算法预处理阶段,所以较高的 RR 是十分必要的.式(10)中 DNB 是检测到区域中的非文字区域的个数, RP (Region Precision) 代表检测到的文字区域准确率.式(11)中 RF 代表综合 RR 和 RP 的算法整体性能.

5.2 实验结果与分析

本文实验图像来自 ICDAR2011 场景文字定位竞赛数据库和实验室场景中文数据库. ICDAR2011 场景文字定位竞赛数据库(后文简称英文数据库)是目前英语文字定位算法的主要测试数据库,包含 255 幅自然场景下英文环境测试图像;实验室场景中文数据库(后文简称中文数据库)是本实验室建立的自然场景下中文环境图像数据库,拥有 5000 副各种场景的汉字图像.

本节在上述数据库中设计四个实验分别用于选择显著性算法、验证文字置信图对背景抑制算法的影响、背景抑制算法在英文数据库的有效性与背景抑制算法在中文数据库的有效性.实验在 Intel E7400/2G RAM, MatlabR2010a 混合编程情况下完成,实验中单幅图像均按照文献[9]保持长宽比归一化高度为 300.

5.2.1 显著性算法比较与选择

文献[17]比较四种显著性算法,认为 Torralba 的显著性算法最适合用于场景文字检测.本文基于场景文字的背景抑制的目的,将谱残差算法与 Torralba 的算法产生的显著图进行二值化按照 5.1 节评价标准进行比较如表 1,其中 T 代表每幅图像的平均处理时间.

从表 1 可看出两种算法的 PCR 、 RP 相差不大,谱

残差算法略低 1%,但在 RR 远高于 Torralba 算法,且速度比其快 10 倍以上.场景文字的背景抑制算法作为文字定位算法的预处理阶段,速度也是一个重要衡量指标.所以,本文选择了谱残差视觉显著性算法.

表 1 显著性算法比较

方法	算法性能				
	PCR	RP	RR	RF	T
Torralba 算法	18.7%	18.3%	63.7%	28.0%	1.30s
谱残差算法	17.6%	17.4%	83.2%	28.8%	0.09s

5.2.2 文字置信图对背景抑制算法的影响

本次实验在英文数据库中分别获取显著性区域后的结果(算法 1)、显著性图与基于笔画特征的点对梯度方向差比例图通过图模型结合的结果(算法 2)、显著性图与基于傅里叶谱统计特征的后验概率图通过图模型结合的结果(算法 3)以及前文所述的背景抑制算法结果(本文算法)按照文献[9]评价标准进行性能评估如表 2 所示.相对于算法 1,本文算法在评价标准上全面胜出;相对于算法 2 与算法 3,本文算法在 PCR 与 RP 下降了 1% 左右,而在 RR 与 RF 提升 8% - 9%.综上,使用结合两种特征的文字置信图是必要的.

表 2 文字置信图对算法的影响

方法	算法性能			
	PCR	RP	RR	RF
算法 1	54.5%	83.8%	36.7%	66.0%
算法 2	86.8%	92.3%	83.1%	87.5%
算法 3	87.0%	92.6%	84.8%	88.6%
本文算法	85.9%	91.8%	93.6%	92.7%

5.2.3 背景抑制算法在英文数据库的有效性

为评价本文算法在英文环境下的有效性,本次实验在英文数据库中采用文献[9]评价标准,进行 4 种算法性能比较,如表 3 所示.表中的文献[18]与文献[19]算法是文献中定位算法的起始阶段,文献[9]因与本文研究问题相近,本文采用其完整算法比较.从表 1 可看出,本文算法在像素级与区域级两方面的性能均优于文献[18]与文献[19]的预处理算法;本文算法相比文献[9],在像素级表现较好,在区域级的准确率 RP 略低,召回率 RR 更好,综合性能 RF 略好.

表 3 ICDAR2011 场景文字定位竞赛数据库中算法性能比较

方法	算法性能				
	PCR	RP	RR	RF	T
文献[18]算法	17.7%	60.5%	75.3%	67.1%	< 1s
文献[19]算法	17.4%	80.4%	72.5%	76.2%	> 1s
文献[9]算法	59.2%	95.6%	88.7%	92.0%	1s
本文算法	85.9%	91.8%	93.6%	92.7%	0.91s

与文献[18,19]相比,本文考虑视觉显著性信息(谱残差视觉显著性图)、笔画细节信息(基于笔画特征的梯度点对梯度方向差比例图)与文字整体信息(基于傅里叶谱统计特征的后验概率图),因此算法针对场景文字进行背景抑制性能会优于二者.与文献[9]相比,本文算法较优的像素级评价结果说明本文算法的背景抑制结果中非文字像素更少,获得的文字区域更准确.区域级的召回率 RR 更好,表明本文算法获得的文字区域更完整,遗漏的文字区域更少;区域级的准确率 RP 较低,是因为本文算法在权衡区域级的召回率 RR 和准确率 RP 时优先考虑召回率 RR .由于本文算法是场景文字自动定位算法的预处理阶段,更高的召回率 RR 可以提高场景文字自动定位算法的整体性能.区域级的综合

性能 RF 较高,说明本文算法在区域级评价中整体性能优于文献[9]算法.综合以上两个级别的评价标准看,本文算法相比文献[9]具有更好的性能.这得益于本文算法对于谱残差视觉显著性理论的应用以及选择更具判别性的特征.

表3中4种算法速度比较,本文算法最快,略快于文献[9]算法.本文算法采用的谱残差显著性模型快速可靠,使本文算法只需关注显著性区域,因此可用于提高算法效率.本文算法运算效率取决于待处理图像面积,而谱残差显著性模型获取显著性区域忽略背景,可减少待处理图像面积,提高算法效率.此外,从图7可看出本文算法使用图模型实际做了文字预分割,减少文字识别阶段工作量,提高了场景文字识别算法性能与效率.



图7 场景文字背景抑制效果图

5.2.4 背景抑制算法在中文数据库的有效性

为验证本文算法对含有中文文字的自然场景图像的有效性,本次实验在中文数据库中随机抽取150副图像进行测试.本文采用算法对自然场景下中文文字定位的性能提升效果作为其有效性的依据.本文选择的场景文字自动定位算法是以边缘获取文字候选区域,根据文字简单特征(长宽比,文字在候选区域外接矩形中占有率以及文字候选区域实际大小)设定阈值验证文字候选区域.本文算法则作为预处理阶段,本实验的算法均采用ICDAR2011评价标准^[20].从表4可看出,背景抑制后的文字定位算法相比原先算法的准确率 P 大幅提升,本文算法的背景抑制效果使文字定位结果更加准确,大量背景区域被剔除;召回率 R 有所下降,是因为文字定位采用的阈值参数是针对原先算法设置,为保证实验的公正性对背景抑制后的文字定位算法并未调整优化;综合性能 F 得到较大的提升,它综合考虑了准确率 P 与召回率 R ,是一个整体算法性能指标,该指标的较大提升表明本文算法可针对文字较好的抑制自然场景中背景区域,较大改善场景文字自动定位算法的性能.

果图,其中(a)行是场景文字图像,(b)行是场景文字背景抑制效果图.效果图中高亮度区域是文字候选区,较暗区域是背景.因为本文算法只是场景文字自动定位算法的预处理阶段,所以在图7中依然有一些虚警率,这在自然场景文字自动定位算法的后续阶段是比较容易消除的.图7的部分结果也说明本文算法可检测非水平文字的,对于待检测文字的方向是鲁棒的.

表4 实验室场景中文数据库中算法实验结果

方法	算法性能		
	P	R	F
文字定位算法(%)	49.0%	81.1%	61.1%
背景抑制后的文字定位算法(%)	66.7%	73.6%	70.0%

6 结论与展望

本文首先获取显著性区域,并取得显著性区域的文字置信图,再根据图模型将视觉显著性图、文字置信图与颜色特征结合,最终达到抑制自然场景文字背景的结果.实验证明,本文算法同时适用于自然场景下中英文环境的背景抑制,可有效提高场景文字自动定位算法的性能,并且对文字排布方向是鲁棒的.在未来工

图7是本文算法在前文提到的两个数据库中的效

作中,本文算法将扩展为场景文字自动定位识别系统。

参考文献

- [1] 冯松鹤,等.一种融合图学习与区域显著性分析的图像检索算法[J].电子学报,2011,39(10):2288-2294.
Feng S H, et al. Combining graph learning and region saliency analysis for content based image retrieval [J]. Acta Electronica Sinica, 2011, 39(10): 2288 - 2294. (in Chinese)
- [2] 钱晓亮,等.一种基于加权稀疏编码的频域视觉显著性检测算法[J].电子学报,2013,41(6):1159-1165.
Qiao X L, et al. A spectral algorithm based on weighted sparse coding for visual saliency detection [J]. Acta Electronica Sinica, 2013, 41(6): 1159 - 1165. (in Chinese)
- [3] 田破荒,等.基于文字穿越线和笔画连通性的视频文字提取方法[J].电子学报,2009,37(1):72-78.
Tian P H, et al. An approach for video-text extraction based on text traversing line and stroke connectivity [J]. Acta Electronica Sinica, 2009, 37(1): 72 - 78. (in Chinese)
- [4] Bai B, Yin F, Liu C L. Scene text localization using gradient local correlation [A]. International Conference on Document Analysis and Recognition [C]. Washington DC: IEEE, 2013. 1412 - 1416.
- [5] Yao C, et al. Detecting texts of arbitrary orientations in natural images [A]. IEEE Conference on Computer Vision and Pattern Recognition, Providence [C]. USA: IEEE, 2012. 1083 - 1090.
- [6] 姜维,卢朝阳,李静,等.基于角点类别特征和边缘幅值方向梯度直方图统计特征的复杂场景文字定位算法[J].吉林大学学报(工学版),2013,43(1):250-255.
Jiang W, Lu Z Y, Li J, et al. Text localization algorithm in complex scene based on corner-type feature and histogram of oriented gradients of edge magnitude statistical feature [J]. Journal of Jilin University Engineering and Technology Edition, 2013, 43(1): 250 - 255. (in Chinese)
- [7] Meng Q, Song Y H. Text detection in natural scenes with salient region [A]. International Workshop on Document Analysis Systems [C]. Gold Coast: IEEE, 2012. 384 - 388.
- [8] Sun Q Y, Yue L, Sun S L. A visual attention based approach to text extraction [A]. International Conference on Pattern Recognition [C]. Istanbul: IEEE, 2010. 3991 - 3995.
- [9] Shi C Z, Xiao B H, Wang C H, et al. Graph-based background suppression for scene text detection [A]. International IAPR Workshop on Document Analysis Systems [C]. Queensland: IEEE, 2012. 210 - 214.
- [10] Hou X D and Zhang L Q. Saliency detection: a spectral residual approach [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Minneapolis: IEEE, 2007. 1 - 8.
- [11] Honey C, Kirchner H, Van Rullen R. Faces in the cloud: Fourier power spectrum biases ultrarapid face detection [J]. Journal of Vision, 2008, 8(12): 1 - 13.
- [12] Epshtein B, Ofek E, Wexler Y. Detecting text in natural scenes with stroke width transform [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. San Francisco: IEEE, 2010. 2963 - 2970.
- [13] Zhang J, Kasturi R. Character energy and link energy-based text extraction in scene images [A]. Asian Conference on Computer Vision, Queenstown [C]. Berlin Heidelberg: Springer, 2011. 308 - 320.
- [14] Ruderman D. The statistics of natural images [J]. Network: Computation in Neural Systems, 1994, 5(4): 517 - 548.
- [15] Boykov Y, Jolly P. Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images [A]. IEEE International Conference on Computer Vision [C]. Vancouver: IEEE, 2001. 105 - 112.
- [16] Boykov Y, Kolmogorov V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision [J]. IEEE Transaction Pattern Analysis and Machine Intelligence, 2004, 26(9): 1124 - 1137.
- [17] Shahab A, Shafait F, Dengel A. How salient is Scene text [A]. International Workshop on Document Analysis Systems [C]. Gold Coast: IEEE, 2012. 317 - 321.
- [18] Lyu M R, Song J Q, Cai M. A comprehensive method for multilingual video text detection, localization, and extraction [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(2): 243 - 255.
- [19] Liu C M, Wang C H, Dai R W. Text detection in images based on unsupervised classification of edge-based features [A]. International Conference on Document Analysis and Recognition [C]. Seoul: IEEE, 2005. 610 - 614.
- [20] Shahab A, Shafait F, Dengel A. Dengel ICDAR 2011 robust reading competition challenge 2: Reading text in scene images [A]. International Conference on Document Analysis and Recognition [C]. Beijing: IEEE, 2011. 1491 - 1496.

作者简介



姜维 男, 1981年12月出生, 河南郑州人. 现西安电子科技大学博士研究生, 研究方向为自然环境文字分析与识别.
E-mail: jwmianzu@gmail.com

卢朝阳 男, 1963年出生, 陕西华县人. 现为西安电子科技大学通信工程学院教授、博士生导师. 研究方向为图像分析与图像理解.
E-mail: zhylu@xidian.edu.cn