

# 基于文本像素颜色聚类的场景文本检测算法

李敏, 郑建彬, 詹恩奇\*, 汪阳

武汉理工大学信息工程学院, 湖北 武汉 430070

**摘要** 提出了一种基于文本像素颜色聚类的场景文本检测方法。利用最大稳定极值区域算法提取原图像初始文本区域, 并通过笔画宽度变换算法与角度特征筛选稳定文本像素。在稳定文本像素中进行多尺度颜色聚类, 并结合支持向量机进行了字符区域验证, 最后进行文本行聚合, 实现文本检测的目标。分别在公共数据集 ICDAR2011 和 ICDAR2013 上进行测试, 算法的  $F$ -score 分别为 0.76 和 0.77, 相比其他文本检测算法, 所提算法获得了较好的检测性能。

**关键词** 图像处理; 场景文本检测; 文本像素提取; 多尺度颜色聚类; 字符验证

中图分类号 TP391.4

文献标识码 A

doi: 10.3788/LOP56.071006

## Scene Text Detection Algorithm Based on Color Clustering of Textual Pixels

Li Min, Zheng Jianbin, Zhan Enqi\*, Wang Yang

College of Information Engineering, Wuhan University of Technology, Wuhan, Hubei 430070, China

**Abstract** A scene text detection method based on color clustering of textual pixels is proposed, in which the initial textual regions of original images are extracted by the maximally extremal stable regional algorithm, and the stable textual pixels are screened out by the stroke width transform algorithm and the angle features. The multi-scale color clustering is conducted in these stable textual pixels, which is combined with a support vector machine for the realization of character regional verification. The text line aggregation is finally adopted to achieve the goal of text detection. The tests are conducted on two public datasets of ICDAR2011 and ICDAR2013, and the  $F$ -scores of this algorithm are 0.76 and 0.77, respectively. Compared with the existing text detection methods, the proposed method obtains a good text detection performance.

**Key words** image processing; scene text detection; textual pixel extraction; multi-scale color clustering; character verification

**OCIS codes** 100.2960; 110.2960; 100.2000

## 1 引言

在信息技术高速发展的时代, 图片早已成为人们进行信息交互的重要载体。日常生活中充斥着自然场景图片, 例如道路牌、广告标语等, 通常包含着重要的文字信息, 这些文字对于理解图像内容有着极大的帮助。此外, 场景文本检测也是许多其他机器视觉领域的基础, 例如图像检索、车牌识别, 以及视频复原等, 因而对其进行研究具有重要的实际意义。

自然场景文本检测被普遍认为是一项具有挑战性的任务, 主要原因如下: 1) 在自然场景图像中, 文字的大小不一、颜色多样、风格各异; 2) 图像的背景较为复杂, 存在着一些类似于文本的物体, 例如砖块、栏杆等; 3) 图像存在对比度低、阴影、光照、模糊等不利因素, 这些问题都对场景文本的准确检测带来了困难。

目前, 场景文本检测的方法主要可分为两类: 基于滑动窗口<sup>[1-2]</sup>和基于连通域分析<sup>[3-8]</sup>的方法。

收稿日期: 2018-09-03; 修回日期: 2018-10-10; 录用日期: 2018-10-30

基金项目: 国家自然科学基金(61303028)

\* E-mail: eqzhan@whut.edu.cn

基于滑动窗口的场景文本检测方法通常使用多尺度的滑动窗口来扫描原始图像,提取出候选文本区域后,利用统计机器学习的方法并结合纹理、梯度等特征进行区域验证,得到文本检测结果。Wang等<sup>[1]</sup>提出了一种端对端的场景文本检测方法,使用滑动窗口扫描图像获得候选区域后,利用训练好的卷积神经网络(CNN)验证候选区域,得到文本检测结果。Zhang等<sup>[2]</sup>对原始图像进行区域对称性分析后,得到文本置信图,提取图中置信度较高的区域作为候选文本区域,并结合CNN进行文本行验证。由于图像中文本的尺寸大小不一,该类方法通常使用多尺度的滑动窗口扫描图像,会产生数量庞大的候选区域,使得算法耗时较长。

基于连通域分析的方法主要利用文本之间的某些共性来提取候选文本区域,例如颜色、边缘、笔画宽度等。常用的文本区域提取方法有最大稳定极值区域(MSER)算法<sup>[3]</sup>、极值区域(ER)算法<sup>[4]</sup>、笔画变换(SWT)<sup>[5]</sup>算法和颜色聚类算法<sup>[6-8]</sup>。Yi等<sup>[6]</sup>将图像颜色直方图分析和传统 $K$ -均值( $K$ -means)聚类算法相结合,将图像进行色彩分层,从而得到候选文本区域。吴慧等<sup>[7]</sup>用 $K$ -means聚类算法将彩色图像聚成固定的5种色彩层,并通过分析候选文本块的几何特性和边缘密度验证文本,得到最终的文本检测结果。不同图像的最佳聚类数可能不同,采用固定的聚类层数无法自适应图像的变化。为了进一步合理确定聚类层数,刘美华等<sup>[8]</sup>通过多次 $K$ -

means聚类找出使色彩距离变化最小的颜色层作为聚类结果。虽然现有的色彩聚类算法能在一定程度上检测出文本区域,然而当文本与其背景颜色较为接近时,易产生欠分割现象。在这种情况下,由于随机选择初始颜色中心,影响了聚类效果,使得分层时的颜色中心往往处于背景中,导致图像分层后无法分割出文本与背景区域,召回率较低。

针对传统颜色聚类算法无法解决低对比度图像文本分割的问题,本文提出了一种新的基于文本像素多尺度颜色聚类的方法,与传统颜色聚类算法相比,该算法以稳定文本像素集为基础,而不是整个图像像素集,拟排除与文本颜色相似的背景像素的干扰。并且采用两级分组策略确定聚类层数与初始颜色中心,以提高图像分层效果。

## 2 算法流程

为了将不同质量的图像带来的影响降低到最小,算法首先对原始自然场景图像进行预处理,采用MSER算法获取初始区域后,利用笔画特征提取出稳定文本像素集,并通过两级分组策略对像素集进行层次划分。然后,设计一种基于文本像素多尺度颜色聚类的算法提取出候选文本区域,在进行字符验证时,分别提取了候选区域的几何特征、笔画特征、纹理特征与边缘特征,并使用支持向量机(SVM)进行分类,最后将字符区域聚合成文本行。所提算法流程图如图1所示。

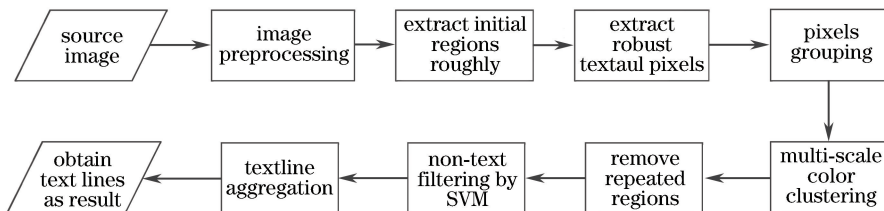


图1 所提算法流程图

Fig. 1 Flow chart of proposed method

## 3 候选区域提取

### 3.1 稳定文本像素提取

对原始图像[图2(a)],采用稳定文本像素提取处理,得到图2(c),再进行多尺度颜色聚类后得到图2(e)。图2(b)与图2(d)分别表示原始图像中所有像素与稳定文本像素在RGB颜色空间的分布情况,与图2(b)相比,图2(d)显示了文本像素在RGB空间中具有颜色一致性。

经过图像预处理后,采用MSER算法提取出初

始区域。MSER算法由于其仿射不变性在机器视觉领域被广泛使用,其借鉴了分水岭算法的思想,当使用不同的灰度阈值对图像进行二值化时得到最稳定的区域。为了检验MSER算法的有效性,从ICDAR(International Conference on Document Analysis and Recognition)2013数据集中随机抽取了50张图像,计算得到有效区域比例与文本覆盖率分别为64%与85%。实验表明:1)MSER算法是有效的,能检测出部分文本区域作为初始区域;2)该初始区域中会存在大量的非文本区域,因此必须进行有效滤除。

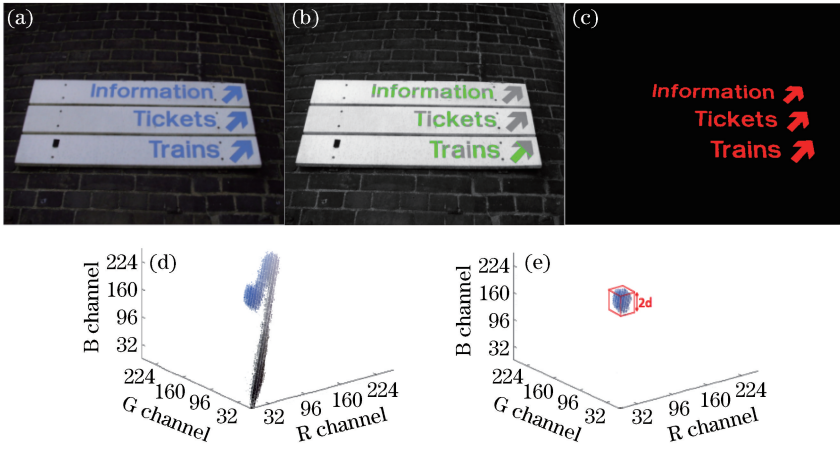


图2 颜色分层过程。(a)原始图像;(b)全部像素的RGB空间分布;(c)稳定文本像素;  
(d)仅文本像素的RGB空间分布;(e)文本像素聚类结果

Fig. 2 Color stratification process. (a) Original image; (b) distribution of all pixels in RGB space; (c) robust textual pixels; (d) distribution of only textual pixels in RGB space; (e) clustering results of textual pixels

启发式规则可以滤除大部分明显的非文本区域,保留如下规则的区域:1)宽高比在0.1~5之间;2)轮廓密度小于0.8;3)占空比大于0.2且小于0.85;4)笔画离散程度小于0.4。经过启发式规则滤除后仍会存在大量的非文本区域,继而使用分类器进行级联滤除,在4.2节中进行了详细说明。

为了进一步排除非文本区域与噪点的影响,本文利用SWT算法与角度特征来获取稳定的文本像素集,首先进行区域笔画宽度直方图统计,分别将相同笔画宽度值的像素作为一组候选稳定像素集 $P_k$ , $k$ 表示组号,并计算每个像素集的 $D_{angle}$ 特征<sup>[9]</sup>,以此来判断该像素集是否稳定,该特征通过计算字符两条平行边的平均梯度方向,能够很好地描述字符两边界平行程度,如图3所示。然后计算每个像素集的平均 $d_{angle}$ ,得到 $D_{angle}$ 。

$$D_{angle}^k = \frac{1}{N_k \cdot \pi} \sum_{j=1}^{N_k} d_{angle}(j), \quad (1)$$

式中: $N_k$ 表示第 $k$ 组像素集 $P_k$ 中的像素数; $j$ 表示该像素集中的每一个像素; $d_{angle}(\cdot)$ 表示经过该像

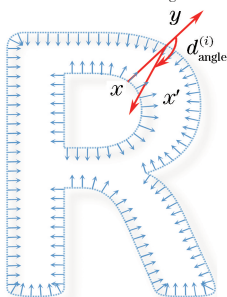


图3  $d_{angle}$ 特征示意图

Fig. 3 Schematic of  $d_{angle}$  feature

素点的射线与其梯度方向夹角的平均值,即该像素点的平均角度; $D_{angle}$ 表示该像素集的平均角度值; $D_{angle}^{min}$ 表示门限值;当 $D_{angle}^k > D_{angle}^{min}$ 时,认为该像素集是稳定的,通过实验验证,本文设置 $D_{angle}^{min} = 0.9$ 。

### 3.2 像素分组策略

在一幅自然场景图像中,文字可能具有一种或多种颜色,并且不同颜色的部分文本像素在颜色空间中距离可能会十分接近,如果直接以文本像素为单位进行聚类,会导致不同颜色文本中部分相似像素之间相互干扰,进而影响颜色分层结果。

因此,为了改善这种情况,本文利用同一文本行内字符颜色大致相同的先验理论,设计了两级分组策略:同一文本行内合并与文本行间合并。

1) 将含有稳定文本像素的区域提取出来,使用高度比和宽度比、角度值、区域中心距离和平均颜色差值来判断两个区域是否属于同一文本行,约束条件如表1所示。经过文本行划分后,处于同一文本

表1 文本行内合并约束条件

Table 1 Constraint conditions for merging in text lines

| Constraint        | Definition  | Threshold range |
|-------------------|---|-----------------|
| Height ratio      | $H_r = \text{height}(R_1) / \text{height}(R_2)$                                       | (0.25, 4)       |
| Width ratio       | $W_r = \text{width}(R_1) / \text{width}(R_2)$   | (0.25, 4)       |
| Angle             | $A_{angle} = \arctan(R_h / R_w)$  | [0, 30]         |
| Centroid distance | $d_c = \frac{\text{sqrt}(R_h^2 + R_w^2)}{\max[\text{width}(R_1), \text{width}(R_2)]}$ | [0, 2]          |
| Color distance    | $D_{color} = \text{sqrt} \left[ \sum_{C=(r,g,b)} (C_1 - C_2)^2 \right]$               | [0, 30]         |

行中的文本像素将被划分到一组中。

表中  $R_1$  和  $R_2$  分别表示两个不同的区域,  $R_h$  和  $R_w$  分别表示这两个区域中心点之间的竖直距离和水平距离,  $C_1$  与  $C_2$  表示这两个区域的各颜色通道值。

2) 将每个文本行作为一个整体, 比较文本行之

间的平均颜色, 当颜色距离较为接近时, 将这两个文本行的像素合并, 最终得到多组文本像素集。

图 4 对比了文本像素分组前后的颜色分层效果, 像素分组前不同文本行中部分颜色相似像素在聚类过程中发生了串扰, 影响了颜色分层结果, 经过像素分组后极大地改善了这种情况。



图 4 分组前后颜色分层结果的对比。(a)原始图像;(b)改善前的图像;(c)改善后的图像

Fig. 4 Comparison of color stratification results before and after grouping. (a) Original image; (b) image before improvement; (c) image after improvement

### 3.3 多尺度颜色聚类

图 5 为文本像素多尺度颜色聚类的算法流程。

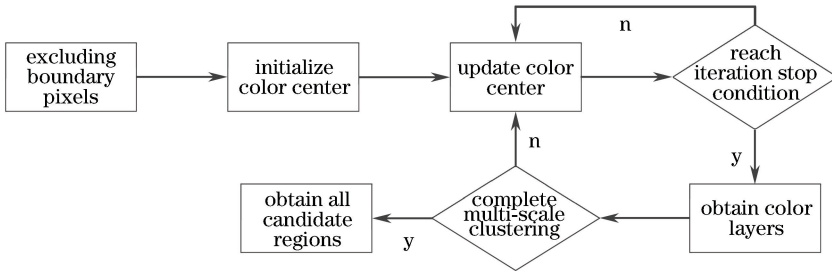


图 5 多尺度颜色聚类算法流程图

Fig. 5 Flow chart of multi-scale color clustering algorithm

输入: 原始图像和文本像素集  $P_t = \{p_k\}$ , 其中序数  $k = \{1, 2, \dots, n\}$ ,  $n$  表示组数。

输出: 候选文本区域。

1) 边缘像素移除。由于边缘像素通常颜色变化较大, 为了获取文本内的主要颜色, 需要对边缘像素进行排除。对原始图像使用 Canny 边缘检测提取出边缘像素集  $p_e$ , 移除  $P_t$  中与  $p_e$  相交的像素。

2) 确定初始颜色中心。与文献[6]和文献[7]随机选择初始颜色中心不同, 本文算法经过文本像素分组后, 可以确定文本颜色层数及其初始颜色中心。对每个文本像素集  $p_k$ , 计算初始颜色中心,

$$\{u_k^t(i) = \sum_{j=\{1,2,\dots,N_k\}} p_k^j(i)/N_k\}, i = \{r, g, b\}, \quad (2)$$

式中:  $t$  表示迭代次数, 此时  $t=1$ ;  $N_k$  表示第  $k$  组像素集  $P_k$  中的像素数;  $p_k^j$  表示  $P_k$  中的每个像素。

3) 更新颜色中心。在每个文本像素集  $p_k$  中, 判断每个像素  $p_k^j$  的颜色分量到第  $t$  次迭代颜色中心的距离是否小于  $d$ 。

$$\{|u_k^t(i) - p_k^j(i)| < d\}, i = \{r, g, b\}, j = \{1, 2, \dots, N_k\}. \quad (3)$$

将所有满足(3)式的像素保存起来, 共同表示为  $p_k^s$ , 其中  $s = \{1, 2, \dots, N_m\}$ ,  $N_m$  表示满足条件的像素总数。然后计算新的颜色中心:

$$\{u_k^{t+1}(i) = \sum_{s=\{1,2,\dots,N_m\}} p_k^s(i)/N_m\}, i = \{r, g, b\}. \quad (4)$$

最后判断是否满足迭代停止条件:

$$\{|u_k^{t+1}(i) - u_k^t(i)| < \epsilon\}, i = \{r, g, b\}, \quad (5)$$

式中  $\epsilon$  设置为 0.001, 如果不满足则继续迭代, 否则跳出迭代, 此时各组文本像素集颜色中心为  $\{u_k^{t+1}(i)\}, i = \{r, g, b\}$ 。

4) 图像颜色分层。计算图像上所有像素  $P$  到每组颜色中心的最小距离:

$$d_w = \min_{k=\{1,2,\dots,n\}} \left\{ \sqrt{\sum_{i=\{r,g,b\}} [P(i) - u_k^{i+1}(i)]^2} \right\},$$

$$w \in \{1,2,\dots,n\}. \quad (6)$$

如果  $d_w < d$ , 那么该像素被划分到第  $w$  组, 遍历完成后, 整个图像将被划分为  $n+1$  层, 然后提取出连通区域。

5) 获取所有候选区域。为了尽可能地找出所有的文本区域, 本文采用多尺度的策略,  $d$  依次从 8 增加至 72, 共有 9 个尺度, 重复步骤 2)~4)。

图 6 分别展示了部分简单背景、复杂背景, 以及对比度较低图像的颜色分层效果。图 6(a) 中的原始图像背景较为简单, 在  $d=32$  时, 文本与背景可以完全分离; 图 6(b) 中的原始图像背景复杂并且存在字符阴影, 在  $d=40$  时, 文字与阴影能被完全分离出来; 在图 6(c) 中的原始图像对比度较低, 文字与其背景的颜色相似度较高, 在  $d=24$  时可以将文本提取出来。在颜色分层时, 图像的最佳阈值很难确定, 单一阈值很可能会漏掉部分文本区域, 因此采用多尺度的方法能够可靠地找出更多的文本区域。



图 6 原图与颜色分层结果。(a)简单背景图像;(b)复杂背景图像;(c)低对比度图像

Fig. 6 Original images and color stratification results. (a) Images with simple background; (b) images with complex background; (c) images with low contrast

图 7 对比了所提算法与其他两种颜色聚类算法的分层效果。由于原图对比度较低且两种文本之间颜色较为接近, 文献[6]方法无法准确区分, 虽然文

献[10]与本文算法均能够提取出文本区域, 但是本文算法对单一尺度的聚类效果更好。



图 7 不同算法颜色分层结果的比较。(a)原始图像;(b)所提算法;(c)文献[6];(d1)文献[10]中  $d=24$  的结果;(d2)文献[10]中  $d=56$  的结果

Fig. 7 Comparison of color stratification results by different methods. (a) Original image; (b) proposed method; (c) method in Ref. [6]; (d1) result when  $d=24$  in Ref. [10]; (d2) result when  $d=56$  in Ref. [10]

## 4 字符验证

### 4.1 字符区域去重

多尺度颜色聚类会导致颜色变化稳定的区域被多次检测, 产生出大量的重复区域, 因此必须进行有效地去重处理。这些重复区域的外接矩形位置大致相同, 因此本文利用重复矩形区域的颜色变化率  $v$

进行去重。对于一个区域的外接矩形, 寻找与其左上角顶点和右下角顶点坐标相近的所有其他矩形, 保留颜色变化率最小的区域, 去除其他重复区域, 计算公式为

$$v = F_{\text{norm}}[B_{\text{std}}(R, G, B)], \quad (7)$$

式中:  $F_{\text{norm}}(\cdot)$  表示求 L2 范数;  $B_{\text{std}}(\cdot)$  表示求标准差。

## 4.2 字符验证

与非文字相比,文字不仅具有一定的几何特征和笔画特征,通常还具备显著的边缘特征和纹理特征。因此本文选取以下特征组成特征向量。

1) 字符长宽比  $A_r$ 。字符一般具有相对规则的形状,其长宽比也较为稳定。

$$A_r = \min(w, h) / \max(w, h), \quad (8)$$

式中: $w$ 、 $h$  分别表示区域外接矩形的宽与高。

2) 字符占空比  $O_r$ 。字符区域在其所处的外接矩形中一般不会太满也不会太空。

$$O_r = |C_n| / \text{Area}(C_n), \quad (9)$$

式中: $|C_n|$  表示字符连通域的像素数; $\text{Area}(C_n)$  表示连通域外接矩形的面积。

3) 对比度  $C_g$ 。文字通常具有一定的对比度以方便人眼识别。

$$C_g = |H_{\text{mg}}(C_f) - H_{\text{mg}}(C_b)|, \quad (10)$$

式中: $H_{\text{mg}}(\cdot)$  表示区域中各点的灰度平均值; $C_f$  和  $C_b$  分别表示前景区域与背景区域。

4) 笔画面积比  $\zeta$ 。由于同一字符区域中的笔画宽度较为稳定,因此,本文使用文献[11]提出的 SSP 特征计算笔画面积比。

$$\zeta = 2 \sum_{z \in S} w_z d_z / A, \quad (11)$$

式中: $w_z$  表示归一化权重; $d_z$  表示像素点距离笔画边缘的距离; $A$  表示前景像素点个数; $w_z = 3/N_z$ ,  $d_z = S_z/2$ ,  $A = S_w \times S_l$ 。其中, $N_z$  表示像素点  $z$  处  $3 \times 3$  领域内笔画支持像素的个数, $S_w$  和  $S_l$  分别表示该笔画的宽度和长度。

除了几何特征与笔画特征之外,本文还提取了梯度方向直方图(HOG)与多块局部二值模式(MB-LBP)特征。HOG 特征可以有效地描述图像中的轮廓与梯度方向信息,MB-LBP 特征是 LBP 特征的一种改进形式,能够有效地弥补 LBP 特征易受噪声影响的缺点。

为了进行字符分类,本文使用 SVM 进行模型训练与分类,在训练集中提取了正样本与负样本,计算上述特征形成特征向量,进行交叉验证后得到最优参数,训练出分类模型。利用该模型对候选区域进行分类,得到文本区域。

## 4.3 文本行聚合

经过字符验证后,文本区域被提取出来,利用文本之间的长宽比、区域面积比、角度、区域中心距离和颜色相似度这 5 个约束条件进行区域合并。具体步骤如下。

1) 将所有文本区域按照中心的横坐标进行排序。

2) 在每次合并时,判断未被标记的文本区域与已标记的文本区域是否满足约束条件限制,如果满足条件,将该文本区域设置为已标记状态,加入文本行集合,直至水平方向上所有区域都被标记。

3) 最后对每个文本行求取最小外接矩形,作为整个文本定位算法的最终结果。

## 5 实验结果与分析

### 5.1 实验数据库和评价指标

为了验证算法的有效性,分别在 ICDAR2011 和 ICDAR2013 公共数据库上进行了测试,其均为国际文本检测大赛的专用数据库。ICDAR2011 数据库中共有 484 幅图像,包含 229 幅训练集图像和 255 幅测试集图像。ICDAR2013 数据集中共有 462 幅图像,包含 229 幅训练集图像和 233 幅测试集图像。这些图片均取自于自然场景,且尺寸各异。评价指标采用文献[12]提出的标检测算法性能的方法,该算法使用召回率  $R$ 、准确率  $P$  和平均性能参数  $F\text{-score}(f)$  表示算法的性能,计算方法分别为

$$R = \frac{\sum_{q=1}^{|T|} \text{match}_T(T_q)}{|T|}, \quad (12)$$

$$P = \frac{\sum_{q=1}^{|E|} \text{match}_E(E_q)}{|E|}, \quad (13)$$

$$f = \frac{1}{\alpha/R + (1-\alpha)/P}, \quad (14)$$

式中: $T$  表示真实数据集中所包含的检测目标的总数; $E$  表示算法所检测到的目标的总数; $\text{match}(\cdot)$  判断目标矩形框是否匹配,若匹配为 1,否则为 0; $q$  表示序号; $\alpha$  的取值为 0.5。

### 5.2 结果分析

表 2 和表 3 分别展示了在 ICDAR2011 和 ICDAR2013 数据集上,本文算法与其他算法的性能比较。从表中可以看出,相比于其他方法,本文算法虽然准确率不高,但是在召回率和平均性能参数上均优于其他算法,主要原因如下:

1) 在计算初始颜色中心时,本文算法使用的是文本像素,而不是在整个图像像素集上,因此颜色中心会更加接近于真实文本。

2) 在进行颜色分层时,采用了多尺度的策略,逐步增加颜色中心的聚类范围,能够更多地提取出真实文本区域。

3) 本文充分利用了初始区域,将初始区域也加入到最终的候选区域中,提高了召回率。

表 2 不同算法在 ICDAR2011 数据库上的性能比较  
Table 2 Performance comparison of different methods on ICDAR2011 database

| Method              | <i>R</i>    | <i>P</i>    | <i>f</i>    |
|---------------------|-------------|-------------|-------------|
| Proposed method     | <b>0.71</b> | 0.82        | <b>0.76</b> |
| Method in Ref. [10] | 0.68        | 0.82        | 0.75        |
| Method in Ref. [13] | 0.68        | <b>0.86</b> | <b>0.76</b> |
| Method in Ref. [14] | 0.69        | 0.81        | 0.75        |
| Method in Ref. [15] | 0.65        | 0.84        | 0.73        |

表 3 不同算法在 ICDAR2013 数据库上的性能比较  
Table 3 Performance comparison of different methods on ICDAR2013 database

| Method              | <i>R</i>    | <i>P</i>    | <i>f</i>    |
|---------------------|-------------|-------------|-------------|
| Proposed method     | <b>0.73</b> | 0.82        | <b>0.77</b> |
| Method in Ref. [10] | 0.70        | 0.84        | 0.76        |
| Method in Ref. [16] | 0.65        | 0.84        | 0.73        |
| Method in Ref. [17] | 0.68        | 0.79        | 0.73        |
| Method in Ref. [18] | 0.63        | <b>0.85</b> | 0.72        |

所提算法对简单背景、复杂背景,以及对对比度较低的图像的检测情况如图 8 所示。



图 8 部分文本检测结果

Fig. 8 Partial text detection results

图 9 为部分检测失败的图像,当文字受强光照

影响或被栏杆等物体遮挡时,算法无法准确定位。

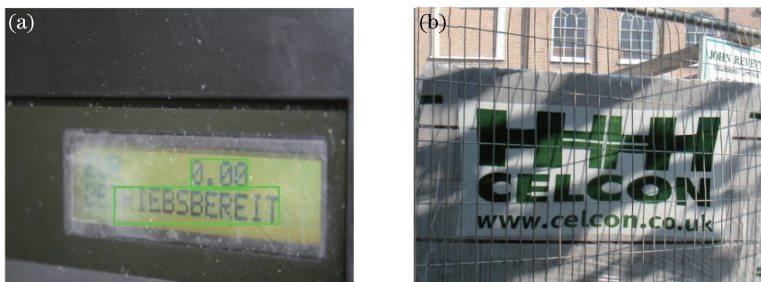


图 9 部分检测失败图像。(a)受光照影响;(b)文字被遮挡

Fig. 9 Partial images under detection failure. (a) Influence of illumination; (b) text occlusion

## 6 结 论

针对传统颜色聚类算法对低对比度自然场景图像检测不佳的问题,提出了一种基于文本像素颜色聚类的场景文本检测方法。该方法以稳定文本像素为聚类基础,减少了颜色相似的背景像素的影响,提

升了算法对低对比度场景图像的稳健性。采用两级分组策略对文本像素进行分组,并通过多尺度颜色聚类提升了图像分层的效果。经过字符区域验证后,所提算法在 ICDAR2011 和 ICDAR2013 数据集上的 *F*-score 分别达到 0.76 与 0.77,证明了算法的有效性。通过实验分析,算法对低对比度与背景复

杂的图像具有一定的优势,但是对光照较强、文本被栏杆遮挡等图片不能准确定位,可作为进一步的研究方向。

## 参 考 文 献

- [1] Wang T, Wu D J, Coates A, *et al.* End-to-end text recognition with convolutional neural networks[C] // Proceedings of the 21<sup>st</sup> International Conference on Pattern Recognition, 2012: 3304-3308.
- [2] Zhang Z, Shen W, Yao C, *et al.* Symmetry-based text line detection in natural scenes[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015: 2558-2567.
- [3] Matas J, Chum O, Urban M, *et al.* Robust wide-baseline stereo from maximally stable extremal regions[J]. *Image and Vision Computing*, 2004, 22(10): 761-767.
- [4] Sung M C, Jun B, Cho H, *et al.* Scene text detection with robust character candidate extraction method [C] // Proceedings of 13<sup>th</sup> International Conference on Document Analysis and Recognition, 2015: 426-430.
- [5] Epshtein B, Ofek E, Wexler Y. Detecting text in natural scenes with stroke width transform [C] // Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010: 2963-2970.
- [6] Yi C C, Tian Y L. Text string detection from natural scenes by structure-based partition and grouping[J]. *IEEE Transactions on Image Processing*, 2011, 20(9): 2594-2605.
- [7] Wu H, Zhao Y Q, Li X H, *et al.* Text detection from natural scene images based on color clustering [J]. *Journal of Central South University (Science and Technology)*, 2015, 46(6): 2098-2103.  
吴慧, 赵于前, 李香花, 等. 基于色彩聚类的自然场景文本检测[J]. *中南大学学报(自然科学版)*, 2015, 46(6): 2098-2103.
- [8] Liu M H, Qu X L, Zhou X F, *et al.* Text localization based on color distance minimization and maximum color difference [J]. *Journal of Optoelectronics • Laser*, 2015, 26(10): 2008-2015.  
刘美华, 屈喜龙, 周细凤, 等. 基于色彩距离最小化和最大色彩差的文本定位[J]. *光电子 • 激光*, 2015, 26(10): 2008-2015.
- [9] Zhang J, Kasturi R. A novel text detection system based on character and link energies [J]. *IEEE Transactions on Image Processing*, 2014, 23(9): 4187-4198.
- [10] Wu H, Zou B J, Zhao Y Q, *et al.* Natural scene text detection by multi-scale adaptive color clustering and non-text filtering [J]. *Neurocomputing*, 2016, 214: 1011-1025.
- [11] Neumann L, Matas J. Efficient Scene text localization and recognition with local character refinement [C] // Proceedings of 13<sup>th</sup> International Conference on Document Analysis and Recognition, 2015: 746-750.
- [12] Wolf C, Jolion J M. Object count/area graphs for the evaluation of object detection and segmentation algorithms [J]. *International Journal of Document Analysis and Recognition*, 2006, 8(4): 280-296.
- [13] Yin X C, Yin X W, Huang K Z, *et al.* Robust text detection in natural scene images [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(5): 970-983.
- [14] Koo H I, Kim D H. Scene text detection via connected component clustering and non-text filtering [J]. *IEEE Transactions on Image Processing*, 2013, 22(6): 2296-2305.
- [15] Yu C, Song Y H, Zhang Y L. Scene text localization using edge analysis and feature pool [J]. *Neurocomputing*, 2016, 175: 652-661.
- [16] Yin X C, Pei W Y, Zhang J, *et al.* Multi-orientation scene text detection with adaptive clustering [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1930-1937.
- [17] Bai B, Yin F, Liu C L. Scene text localization using gradient local correlation [C] // Proceedings of 12<sup>th</sup> International Conference on Document Analysis and Recognition, 2013: 1380-1384.
- [18] Shi C Z, Wang C H, Xiao B H, *et al.* Scene text recognition using part-based tree-structured character detection [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2013: 2961-2968.