

结合区域颜色一致性和图割的复杂场景文本分割方法

刘晓佩

(1. 西安科技大学 通信与信息工程学院, 西安 710054; 2. 西安电子科技大学 通信工程学院, 西安 710071)

摘要: 针对复杂场景文本难以有效分割的问题, 提出一种复杂场景文本分割方法. 首先, 使用简单的线性迭代聚类(SLIC)算法将原始图像分割为若干局部区域, 并在其区域邻接图上构建图割模型; 然后, 采用高斯混合模型(GMMs)和支持向量机(SVM)后验概率模型对场景文本进行建模, 并引入每个局部区域与模型之间的匹配度用于计算似然能. 为了增强GMMs的鉴别力, 在参数学习中引入模型性能描述子, 自适应地获得模型参数. 实验结果表明, 所提出的算法能够较好地处理复杂场景文本分割问题, 文本的识别率得到了明显的提升.

关键词: 文档分析; 场景文本; 文本分割; 图割

中图分类号: TP393.1

文献标志码: A

Complex scene text segmentation method using region color consistence and graph cut

LIU Xiao-pei

(1. School of Communication and Information Engineering, Xi'an University of Science and Technology, Xi'an 710054, China; 2. School of Telecommunication Engineering, Xidian University, Xi'an 710071, China. E-mail: liuxiaopei2007@163.com)

Abstract: To solve the problem of text segmentation in complex scene images, a method of complex scene text segmentation is proposed. The original image is firstly divided into some small homogeneous regions by using the simple linear iterative clustering(SLIC) algorithm, and the graph model is constructed based on the region neighborhood connection diagram. Then, Gaussian mixture models(GMMs) and support vector machine(SVM) post probability based model are proposed to make model for foreground(text), and the degree of each region's fitness to models is introduced to calculate likelihood energy. In addition, to improve the discrimination ability of GMMs, a model performance descriptor is introduced to estimate parameters of GMMs adaptively. Experimental results show that the proposed method can deal with the problem of complex scene text segmentation efficiently, and the recognition precision rate is improved significantly.

Keywords: document analysis; scene text; text segmentation; graph cut

0 引言

自动提取自然环境中的文本信息是计算机视觉领域中一个非常具有挑战性的问题. 自然环境中的文本信息为场景的理解、环境的感知以及基于内容的图像检索等应用提供关键线索, 因此得到了研究者的广泛关注^[1-3]. 文本分割是文本信息提取过程中的一个关键技术, 其主要任务是将定位后的文本区域转换为二值文本. 当文本的颜色、光照条件及背景复杂多变时, 文本分割的性能对后续的文本正确识别具有较大的影响, 因此, 本文主要针对此问题进行研究.

现有的文本分割方法可大致分为基于阈值的分割方法、基于聚类的分割方法和基于统计模型的分割方法3种^[2]. 这些方法都有各自的优势和适用场合, 然而对于光照条件及背景复杂多变的场景文本图像, 尚有问题未得到较好的解决.

近年来, 基于图论的图割方法由于具有多特征融合能力、全局最优、数值鲁棒性强等优点在图像分割领域获得了极大的成功^[4-5]. 场景文本分割可视为一种特殊的图像分割问题, 目前已有少数的研究者注意到了这一应用热点. 文献[6]首先将视频文本图像

收稿日期: 2014-09-04; 修回日期: 2014-11-04.

基金项目: 国家自然科学基金项目(61302133); 陕西省科技研究计划工业攻关项目(2014K06-37, 2013K07-35, 2015GY023).

作者简介: 刘晓佩(1976—), 女, 讲师, 从事图像分析与识别、计算机视觉的研究.

分割为若干子图, 然后对各子图分别建立图割模型, 采用 **Meanshift** 算法对前景和背景颜色进行聚类分析, 用像素颜色与前景和背景聚类中心的距离计算能量函数中似然能, 并采用 **max-flow** 算法求解模型, 最后合并子图得到分割结果. 文献 [7] 利用了种子点学习前景和背景颜色的高斯混合模型 (**GMMs**), 在此基础上建立图割模型, 采用迭代的方式获得分割结果. 这两种方法都取得了较好的结果, 但是还存在两个方面的问题: 1) 采用单像素标记的方式, 造成算法效率较低, 限制了其实用性; 2) 在构建图割模型能量泛函中的似然能时, 仅采用了颜色信息, 而在实际场景文本图像中, 背景中往往存在大量与文本颜色相似的区域, 显然仅利用颜色信息是不充分的.

针对上述问题, 本文结合区域颜色一致性和图割算法, 提出一种复杂场景文本分割新方法. 算法的基本思路是: 根据图像局部颜色一致性的特点, 首先采用简单的线性迭代聚类 (**SLIC**) 算法对原始图像进行局部区域分割, 以局部区域代替单像素点构建加权图模型, 这样可以大大减少节点的个数, 提高算法效率; 其次, 针对场景文本的复杂性, 提出分别采用 **GMMs** 和基于支持向量机 (**SVM**) 的后验概率模型对文本和背景进行建模, 将每个局部区域与模型间的匹配度引入图割模型似然能的计算. 所提出的算法融合了颜色和笔画特征, 这样可以弥补单一特征的缺点. 此外, 引入模型性能描述因子, 设计改进的期望最大化算法 (**EM**) 自适应地学习 **GMMs** 模型参数.

1 基于区域颜色一致性的图割模型

1.1 局部区域分割

为了提高图割模型的效率, 首先将原始图像预分割为一些同质区域, 然后用这些局部区域代替图像像素点构建加权图割模型, 这已成为一种有效的解决思路^[8-9]. 受这一思路启发, 同时结合文本区域通常具有相似颜色的特点, 本文采用颜色一致的局部区域代替像素点构建文本图像分割模型. 尽管场景文本图像可能存在光照不均匀、对比度较低的情况, 但本文认为只要划分的局部区域足够小, 就可以保证局部区域颜色的一致性.

现有的局部区域分割算法较多, 常用的有分水岭算法^[8]和 **Meanshift** 算法^[9]. 分水岭算法过分割严重且不能较好地保持区域边界; **Meanshift** 算法虽然能够较好地保留图像边缘, 但需要设置的参数较多, 效率较低. **SLIC** 算法由 **Radhakrishna**^[10] 提出, 是一种基于聚类的预分割算法. 该算法能够很好地保留边界, 区域一致性强、效率高, 且需要设置的参数少; 另外, 该算法利用颜色和空间信息进行聚类分析. 因此, 利用该方法能够实现本文的具有颜色一致性的局部区域

分割.

首先, 将原始彩色图像转换到 **CIE Lab** 颜色空间, 并设置 k 个初始的聚类中心, 将其均匀分布在图像上. 为了避免将聚类中心设置在图像边缘上, 将每个聚类中心调整到 3×3 邻域内梯度最小的位置上; 其次, 以 **CIE Lab** 彩色空间中颜色相似距离和空间距离的加权和为新的距离度量, 对图像像素进行聚类分析, 生成 k 个局部区域; 然后, 定位新的聚类中心, 重新进行聚类分析, 如此反复, 直到收敛到最佳位置, 即实现了区域分割. 对于如图 1(a) 所示的场景文本图像, 将其进行局部区域分割的结果如图 1(b) 所示.



图1 局部区域分割示例

1.2 基于局部区域的图割模型

将采用 **SLIC** 方法对原始图像所分割的若干个局部区域记为 $X = \{x_s, s \in S\}$, S 为局部区域的位置点集; 定义 $G = \langle X, E \rangle$ 为建立在 X 上的加权图, E 为连接各节点的边; 设 $Z = \{z_s, s \in S\}$ 为各区域的标号集合, 其中 $z_s = \{\text{text}, \text{background}\}$. 结合经典图割方法, 定义能量泛函如下:

$$E(z) = \sum_{s \in S} U(z_s | x_s) + \lambda \sum_{(s,t) \in N} V(z_s, z_t | x_s, x_t). \quad (1)$$

其中: $U(z_s | x_s)$ 为似然能 (区域项), 表示将区域 x_s 标记为文本或背景的代价; $V(z_s, z_t | x_s, x_t)$ 为先验能 (光滑项), 表示相邻区域的边界关系, 用于估计邻接区域 x_s 和 x_t 分配为不同标号时的惩罚值; λ 为常数, 表示区域项与光滑项之间的权重因子; N 为定义在 X 上的邻域区域集合.

分割就是在加权图 G 的基础上, 通过网络流理论将标号问题的全局最优化转化为求其对应加权图的最小割问题, 进行全局最小割运算以实现能量泛函的最小化. 确定能量泛函的关键是区域项和光滑项的设计, 下面详述二者的设计过程.

2 能量函数的设计

在图割应用中, 用户需要通过人工交互在待分割图像中标记出目标和背景种子像素点集合, 并根据种子点分别对目标和背景建模, 在此基础上设计能量函数中的似然能. 本文根据文本的笔画特点自动提取前景和背景种子点, 具体的算法如下所述.

2.1 基于笔画特征的种子点提取

场景文本通常由具有相似宽度的笔画组成, 且笔画区域的颜色一致. 除此之外, 笔画边缘和背景具有较高的对比度, 因此文本笔画边缘是成对出现的, 且每个边缘对的梯度方向相反. 该特点常用于文本区域检测^[1], 本文将其用于提取文本和背景种子点. 首先采用 Canny 算法提取场景文本的边缘, 然后根据笔画边缘梯度方向将边缘像素点划分为两个点集, 将梯度幅值近似相等、方向相反的两个边缘点看作笔画点对, 连接点对即可获得文本的笔画种子点. 图 2 给出了种子点的提取过程, 图 2(b) 为采用两种灰度表示不同梯度方向的笔画边缘图. 将文本极性取反, 以类似的方式提取背景种子点. 图 2(c) 即为所提取的种子点, 其中深色线覆盖区域为文本种子点, 浅色为背景种子点.



图 2 种子点提取过程

2.2 区域代价函数计算

由于场景文本的复杂性, 本文采用两种不同的模型对文本和背景建模, 即建立文本和背景颜色的 GMMs 模型以及基于 SVM 的后验概率模型. 在两种模型中, 分别衡量每个局部区域与文本或背景的匹配度, 然后计算将该局部区域标记为文本或背景的代价.

为了表达方便, 本文将局部区域与模型间的匹配度定义为参数函数 $f: S \rightarrow R$, 表示将图像中的任意区域 $s \in S$ 映射为与前景或背景的相似程度. 这样, 设区域 x_s 与 GMMs 模型的匹配度为 $f_{\text{GMMs}}(x_s^{\text{stroke}})$, 区域 x_s 属于文本的概率置信度为 $f_{\text{SVM}}(x_s^{\text{stroke}})$, 则定义区域代价函数为

$$U(z_s = 1) = \alpha \cdot f_{\text{GMMs}}(x_s^{\text{color}})^F + (1 - \alpha) \cdot (1 - f_{\text{SVM}}(x_s^{\text{stroke}})), \quad (2)$$

$$U(z_s = 0) = \alpha \cdot f_{\text{GMMs}}(x_s^{\text{color}})^B + (1 - \alpha) \cdot f_{\text{SVM}}(x_s^{\text{stroke}}), \quad (3)$$

其中 $f_{\text{GMMs}}(x_s^{\text{color}})^F$ 和 $f_{\text{GMMs}}(x_s^{\text{color}})^B$ 分别表示 x_s 与前景和背景的匹配度. 下面详细介绍 $f_{\text{GMMs}}(x_s^{\text{color}})$

和 $f_{\text{SVM}}(x_s^{\text{stroke}})$ 的计算过程.

2.2.1 基于区域的 GMMs 模型

颜色是区分文本和背景的一个主要的特征, 本文提取 RGB 颜色特征对前景和背景建模. 利用前景和背景种子点分别训练一个 GMMs 模型, 即前景模型为

$$\text{Foreground} = \{G_{F1}, G_{F2}, \dots, G_{FM}\},$$

背景模型为

$$\text{Background} = \{G_{B1}, G_{B2}, \dots, G_{BN}\}.$$

为了更准确地描述局部区域, 采用局部自适应高斯分布描述每个局部区域^[9], 即 $G_s = \{\mu_s, \Sigma_s\}$. 这样, 任意局部区域 x_s 与前景和背景的匹配度可分别定义为

$$f_{\text{GMMs}}(x_s)^F = \exp(-\min_{i \in [1, M]} (\text{dis}_F(G_s, G_{Fi}))), \quad (4)$$

$$f_{\text{GMMs}}(x_s)^B = \exp(-\min_{i \in [1, M]} (\text{dis}_B(G_s, G_{Bi}))). \quad (5)$$

上述两个高斯分布间的距离可通过 KL 散度计算. 设 G_i 和 G_j 分别具有高斯参数集 $\{\mu_i, \Sigma_i\}$ 和 $\{\mu_j, \Sigma_j\}$, 则可采用如下形式计算两者之间的距离:

$$\text{dis}(G_i, G_j) = \frac{1}{2}(\text{tr}(\Sigma_i^{-1}\Sigma_j + \Sigma_j^{-1}\Sigma_i) + (\mu_i - \mu_j)^T \times (\Sigma_i^{-1} + \Sigma_j^{-1})(\mu_i - \mu_j) - 2r)^{\frac{1}{2}}. \quad (6)$$

其中: tr 为矩阵的迹运算; r 为高斯特征空间的维数, 即颜色空间的通道数.

2.2.2 基于 SVM 的后验概率模型

区域代价衡量的是将某一个节点设置为前景或背景的代价, 如果一个节点属于前景的概率较大, 则其代价较小, 反之亦然. 由于 SVM 分类器具有良好的泛化能力, 本文选择 SVM 分类器来估计每一个局部区域属于前景的概率 (匹配度). 但是标准 SVM 分类器的输出结果为样本的类别判决, 不能给出样本对类别的隶属度, 即样本的后验概率. 为了解决这个问题, 本文采用双参数 Sigmoid 模型将 SVM 的决策函数输出映射为待分类样本的后验概率^[13]. Sigmoid 转换模型定义为

$$P(z_s = \text{text} | f_s) = \frac{1}{1 + \exp(Af_s + B)}. \quad (7)$$

其中: $f_s = f(x_s)$ 为 SVM 分类器的决策函数, 参数 A 和 B 通过最大似然算法 (ML) 在训练集上训练获得.

为了估计每个局部区域的后验概率, 本文提取笔画特征训练 SVM. 由前面的分析可知, 文本的笔画边缘通常是闭合的, 且由边缘梯度近似相等、方向近似相反的点对组成, 因此文本通常具有较低边缘点对梯度方向差. 本文将笔画边缘梯度方向差的均值、方差及笔画能量作为笔画特征.

1) 均值. 该特征表达笔画边缘梯度方向差的平均值, 其值越小表示为文本的可能性越大, 具体定义如下:

$$\mu_{\text{angle}} = \frac{1}{N} \sum_{i=1}^N G_{\text{angle}}^i. \quad (8)$$

其中: N 表示边缘点对的个数; $G_{\text{angle}}^i = |\theta_i - \theta_{i,\text{pair}} - \pi|$, 表示点对 i 的相位差; θ_i 和 $\theta_{i,\text{pair}}$ 分别表示点对 i 的边缘梯度方向.

2) 方差.

$$\sigma_{G_{\text{angle}}} = \frac{1}{N} \sqrt{\sum_{i=1}^N (G_{\text{angle}}^i - \mu_{G_{\text{angle}}})^2}. \quad (9)$$

3) 笔画能量.

$$E = \frac{1}{2} \cdot \left(\frac{\pi - G_{\text{angle}}}{\pi} + \frac{\sum_{i=1}^N h(G_{\text{angle}}^i, t)}{N} \right). \quad (10)$$

其中

$$h(G_{\text{angle}}^i, t) = \begin{cases} 1, & G_{\text{angle}}^i \geq t; \\ 0, & \text{else,} \end{cases}$$

t 为阈值. 可见 E 越大, 笔画边缘越相似.

2.3 邻域代价函数

邻域代价函数衡量相邻像素的一致性, 如果相邻像素相差越大, 则标记为相同类别的代价越大, 反之亦然. 本文在局部区域上构建模型, 定义改进的邻域代价函数如下:

$$V(z_s, z_t) = \sum_{s, t \in N} (\text{dist}(G_s, G_t) + (E_s - E_t)^2 \cdot \delta(z_s, z_t)). \quad (11)$$

其中

$$\delta(z_s, z_t) = \begin{cases} 1, & z_s = z_t; \\ 0, & z_s \neq z_t. \end{cases}$$

E_s 和 E_t 分别为局部区域 x_s 和 x_t 的笔画能量特征.

3 GMMs 参数估计

采用 GMMs 模型对前景和背景的颜色进行建模, 其模型参数为 $\theta = \{\mu, \Sigma, \omega\}$. 为了对 GMMs 模型进行参数学习, EM 算法是一种常用的选择, 该算法通常收敛到局部最优解. 在估计 GMMs 模型参数时, 一般的思路是依据经验指定模型成分个数. 然而, 对于复杂的场景文本图像, 该参数设定为固定值会造成模型对图像特征的描述不够准确. 此外, 在传统的 EM 框架下, 前景和背景的 GMMs 模型是在对应样本上分别进行训练的, 没有考虑不同类别样本间的相似性. 为了提高模型的区分力, 本文训练前景或背景颜色模型时, 引入不同类别样本对当前训练模型的拟合误差. 受文献 [12] 的启发, 定义用于描述模型区分力的参数为

$$d_f = \sum_{n=1}^{N_b} G_F(x_n^b; \theta^f) / \sum_{n=1}^{N_f} G_F(x_n^f; \theta^f). \quad (12)$$

其中: x_n^f ($n = 1, 2, \dots, N_f$) 为前景像素训练集, x_n^b ($n = 1, 2, \dots, N_b$) 为背景像素训练集, G_F 为前景 GMMs

模型, d_f 为背景种子对前景模型的拟合误差与前景种子的拟合误差之比. 很明显, d_f 越小表明前景模型的区分力越强. 将 d_f 引入模型参数学习中, 则本文设计的 EM 算法具体步骤如下.

Step 1 输入训练数据 $\{y_1, y_2, \dots, y_N\}$, 设置迭代终止条件, 模型成分个数初始值 $n_c = 1$, 最大迭代次数 K .

Step 2 采用 K 均值聚类算法将训练数据分为 n_c 类, 计算模型初始参数 $\theta\{0\}$.

1) 设置内部迭代次数 $t = 1$.

2) E 步: 根据模型参数, 计算模型分量训练数据 y_i 的响应度

$$\alpha_{ik} = \frac{\omega_k(t) P(y_i | \theta_k(t))}{\sum_{k=1}^{n_c} P(y_i | \theta_k(t))}. \quad (13)$$

其中: $i = 1, 2, \dots, N; k = 1, 2, \dots, n_c$.

3) M 步: 重新估计模型参数

$$\omega_k(t+1) = \frac{1}{N} \sum_{i=1}^N \alpha_{ik}, \quad (14)$$

$$\mu_k(t+1) = \sum_{i=1}^N \alpha_{ik} y_i / \sum_{i=1}^N \alpha_{ik}, \quad (15)$$

$$\Sigma_k(t+1) = \frac{\left(\sum_{i=1}^N \alpha_{ik} (y_i - \mu_k(t+1))(y_i - \mu_k(t+1))^T \right)}{\sum_{i=1}^N \alpha_{ik}}. \quad (16)$$

4) 重复 2)、3) 步, $t = t + 1$, 直到收敛.

Step 3 若 $|d_f(K+1) - d_f(K)| \leq \varepsilon$ 或 $n_c \geq K$, 则停止迭代; 否则, 转入 Step 2, $n_c = n_c + 1$.

上述算法中, $P(\cdot)$ 是标准多元高斯概率密度函数. 采用上述算法可以自动获得模型成分个数及模型参数.

4 实验结果及分析

实验中采用 100 幅场景文本图像^[14]对算法进行测试, 其中包括非均匀光照、复杂背景等多种类型的质量下降图像. 评价标准采用分割准确率和识别率两个性能指标, 其定义形式为

$$\text{CPR} = N_c / N, \quad (17)$$

$$\text{PPR} = M_p / M. \quad (18)$$

其中: CPR 和 PPR 分别表示识别率和分割正确率, N_c 表示正确识别的字符个数, N 表示测试数据集中所包含的字符个数, M_p 表示正确分割的像素个数, M 表示总的像素个数.

4.1 区域分割结果对算法性能的影响

本文采用 SLIC 算法进行局部区域分割, 实验中将其结果与采用分水岭和均值漂移算法的区域分割结果进行比较, 结果如图 3 所示.

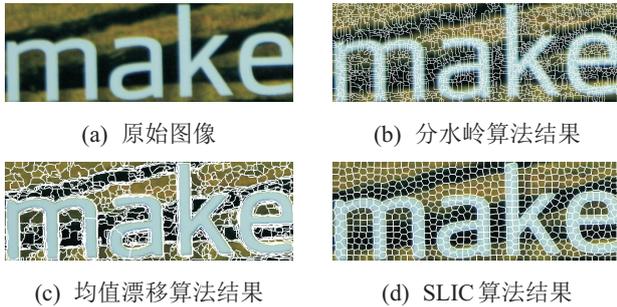


图 3 采用不同方法的区域分割结果比较

从图 3 可以看出: 分水岭算法的过分割严重, 且局部区域一致性较差, 这样势必会增加模型建立的复杂性和分割结果的错误; 均值漂移算法的区域一致性较好, 但是在背景复杂时, 过分割严重; 与前两者相比, 采用 SLIC 算法可获得大小基本一致的局部区域, 且区域一致性较好. 另外, 从运行效率来看, 三者相差不大. 总的说来, 本文采用 SLIC 算法进行局部区域分割具有较大的优势.

4.2 GMMs 模型成分个数对算法性能的影响

为了验证本文采用自适应学习 GMMs 模型成分个数的优越性, 实验中将本文算法与采用固定 GMMs 成分个数的算法结果进行比较. 根据经验设置前景和背景颜色 GMMs 模型的成分个数分别为 3 和

5, 比较结果如表 1 所示.

表 1 采用不同 GMMs 成分个数结果对比

| GMMs 成分个数 | CPR /% | PPR /% |
|-----------|--------|--------|
| 3 | 60.89 | 81.77 |
| 5 | 61.56 | 83.41 |
| 自适应选择 | 62.35 | 88.91 |

从表 1 可以看出, 采用固定成分个数的 GMMs 模型不能较好地描述所分割对象的特征; 而本文所采用的自适应学习方法可以根据每一幅图像的特点, 自适应地选择 GMMs 模型成分个数, 能够更好地描述文字和背景颜色, 因此具有最优的分割结果.

4.3 与其他算法的比较

采用本文所提出的算法对场景文本图像进行分割, 并将分割结果与其他 3 种算法的分割结果进行对比. 算法 1 为基于经典图割模型的分割算法^[4], 采用灰度直方图对前景和背景进行建模; 算法 2^[7]是目前最好的复杂场景分割算法, 该算法采用单像素标记的方式对像素进行标记, 并使用 GMMs 模型对前景和背景颜色分别建模, 模型成分的个数均设置为 5; 算法 3 采用分水岭算法进行区域分割, 其余步骤和本文算法相同. 另外, 前景和背景模型的训练集均使用本文方法获得.

实验中选取不同类型的场景文本图像, 其中包含复杂背景、对比度低和光照不均匀等情况. 将对比较算法用于分割图 4(a) 所示的场景文本图像, 结果如图 4(b)~图 4(e) 所示.



图 4 算法分割结果比较

从实验结果可以看出, 本文算法在分割时能够兼顾区域一致性和准确性, 具有最优的性能. 算法 1 采用灰度直方图对前景和背景建模, 在对比度较低、光照变化较大时, 错误分割严重; 算法 2 能够较好地处理复杂背景下的文本分割, 但是所使用的 RGB 颜色特征对光照变化敏感, 当光照不均匀或对比度较低时, 分割效果不理想; 算法 3 采用分水岭算法进行区域分割, 因为该算法获得的局部区域一致性较差, 所以区域标记错误较多.

为了比较算法的适应性, 本文在实验数据集上对

所有对比算法的性能进行测试, 并通过识别率、分割准确率及耗时 3 个方面进行定量评价, 评价结果如表 2 所示.

表 2 与其他算法的比较结果

| 方法 | CPR /% | PPR /% | 耗时 /s |
|------|--------|--------|-------|
| 算法 1 | 51.20 | 78.29 | 10.8 |
| 算法 2 | 59.38 | 83.22 | 28.5 |
| 算法 3 | 58.67 | 80.62 | 5.7 |
| 本文算法 | 62.35 | 88.91 | 3.5 |

实验中, 选取商业 OCR 软件 Tesseract^[16]为识别软件, 该软件识别图片中文字信息具有较高的准确率.

将4种算法的分割结果分别送入Tesseract软件进行识别,本文算法分割结果的文本识别率最高.这主要是因为其他3种算法不能准确分割文本和背景像素,导致白色背景中存在较多的噪声(即非文字像素),这对正确识别文本造成了较大的干扰,导致识别准确率较低.

从算法效率来看,算法1和算法2耗时较长,这主要是因为二者均采用单像素标记的方式;本文算法以局部区域代替单像素点构建图模型,大大降低了节点的个数,节省了时间开销.

另外,将4种算法应用于30幅不同类型的场景文本图像,将其分割准确率绘制成曲线,如图5所示.总体来说,本文算法相比其他算法具有更好的性能.

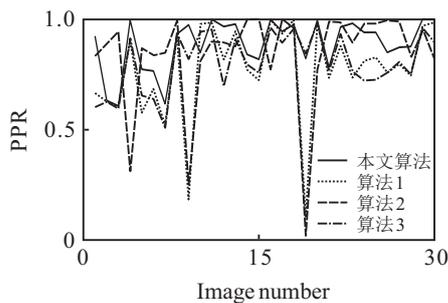


图5 4种分割算法性能比较

5 结论

本文针对复杂场景文本图像分割困难的问题,在交互式图割和局部区域一致性基础上,提出了一种场景文本分割方法.所提出的算法在局部区域上建立图模型,结合GMMs混合模型及基于笔画特征的文字后验概率构建邻域连接权系数.此外,根据文本笔画特点,自动获取前景和背景种子点,分别对前景和背景颜色建立其高斯混合模型,并自适应获得GMMs成分个数.实验结果表明,本文算法能够较好地实现场景文本的分割,提高了后续文本识别的准确率.

参考文献(References)

[1] Sharma N, Pal U, Blumenstein M. Recent advances in video based document processing: A review[C]. Proc of the 10th IAPR Int Workshop on Document Analysis Systems. Queensland: IEEE Press, 2012: 63-68.

[2] Zhang H G, Zhao K L, Song Y Z, et al. Text extraction from natural scene image: A survey[J]. Neurocomputing, 2013, 122: 310-323.

[3] 姜维, 卢朝阳, 李静, 等. 基于视觉显著性和提升框架的场景文字背景抑制方法[J]. 电子与信息学报, 2014, 36(3): 617-623.
(Jiang W, Lu Z Y, Li J, et al. Visual Saliency and Boosting Based Background Suppression for Scene Text[J]. J of Electronics & Information Technology, 2014, 36(3): 617-623.)

[4] Boykov Y, Funka-Lea G. Graph cuts and efficient $N-D$ image segmentation[J]. Int J of Computer Vision, 2006, 70(2): 109-131.

[5] 刘松涛, 殷福亮. 基于图割的图像分割方法及其新进展[J]. 自动化学报, 2012, 38(6): 911-922.
(Liu S T, Yin F L. The basic principle and its new advances of image segmentation methods based on graph cuts[J]. Acta Automatica Sinica, 2012, 38(6): 911-922.)

[6] Shi C, Xiao B, Wang C, et al. Adaptive graph cut based binarization of video text images[C]. The 10th IAPR Int Workshop on Document Analysis Systems. Queensland: IEEE Press, 2012: 58-62.

[7] Mishra A, Alahari K, Jawahar C V. An MRF model for binarization of natural scene text[C]. Proc of the 11th Int Conf on Document Analysis and Recognition. Beijing: IEEE Press, 2011: 11-16.

[8] 徐胜军, 韩九强, 刘光辉, 等. 基于局部空间自适应MRF模型的图像分割[J]. 控制与决策, 2013, 28(6): 889-893.
(Xu S J, Han J Q, Liu G H, et al. Image segmentation based on local spatial adaptive Markov random field model[J]. Control and Design, 2013, 28(6): 889-893.)

[9] 韩守东, 赵勇, 陶文兵, 等. 基于高斯超像素的快速 Graph Cuts 图像分割方法[J]. 自动化学报, 2011, 37(1): 11-20.
(Han S D, Zhao Y, Tao W B, et al. Gaussian super-pixel based fast image segmentation using graph cuts[J]. Acta Automatica Sinica, 2011, 37(1): 11-20.)

[10] Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state-of-the-art superpixel methods[J]. IEEE Trans on PAMI, 2012, 34(11): 2274-2282.

[11] Zhang J, Rangachar K. Character energy and link energy-based text extraction in scene images[C]. Computer Vision-ACCV 2010. New Zealand: IEEE Press, 2011: 308-320.

[12] Wang J. Discriminative gaussian mixtures for interactive image segmentation[C]. Int Conf on Acoustics, Speech and Signal Processing. Hawaii: IEEE Press, 2007: 601-604.

[13] Lin H T, Lin C J, Weng R C. A note on Platt's probabilistic outputs for support vector machines[J]. Machine Learning, 2007, 68(3): 267-276.

[14] Robust word recognition dataset[DB/OL]. (2011-09-18) [2014-09-04]. <http://lgoval.essex.ac.uk/icdar/RobustWord.html>.

[15] Gatos B, Ntirogiannis K, Pratikakis I. ICDAR 2009 document image binarization contest[C]. Proc of the 10th Int Conf on Document Analysis and Recognition. Catalonia: IEEE Press, 2009: 1375-1382.

[16] Tesseract OCR[DB/OL]. (2013-04-25)[2014-09-04]. <http://code.google.com/p/tesseract-ocr/downloads/list>.