

创新性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果；也不包含为获得西安电子科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中做了明确的说明并表示了谢意。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

本人签名： 王炜

日期 2011.1.11

关于论文使用授权的说明

本人完全了解西安电子科技大学有关保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属西安电子科技大学。本人保证毕业后，发表论文或使用论文工作成果时署名单位仍是西安电子科技大学。学校有权保留送交论文的复印件，允许查阅和借阅论文；学校可以公开论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存论文。（保密的论文在解密后遵守此规定）

本学位论文属于保密，在 ____ 年解密后适用本授权书。

本人签名： 王炜

日期 2011.1.11

导师签名： 卢朝阳

日期 2011.1.11

摘要

自然场景中蕴含着丰富的文字信息,这些文字对场景的认知具有重要价值,是描述和理解场景的关键线索,在人们日常生活中起着重要的作用。作为目前制约自然场景文字识别应用的瓶颈,文本定位技术的研究是很有必要的。

本文重点研究自然场景文字检测、文字定位的方法。采用了基于角点提取的自然场景中文字定位方法和基于颜色分割聚类的自然场景文字定位方法,提出了结合两者的自然场景文字定位方法,并结合笔划、纹理周期性等特征进一步优化算法性能。

针对图象上文字区域存在的丰富的角点,本文将角点特征的检测引入文字定位,采用 SUSAN 角点提取方法定位文字。首先用 SUSAN 角点提取方法提取角点,其次使用模板分类选取有效角点,最后通过聚类 and 连通分析来定位文字。算法对复杂背景中的文字有一定的定位能力。

针对自然场景文字在颜色方面的分布特点,本文将颜色信息应用到自然场景中的文字检测。采用颜色空间分割和分层聚类的方法,首先划分颜色空间,将颜色映射为若干类,然后将每一类对应像素点分成一幅子图,通过在子图上的形态学运算和连通分析,结合多子图来得到的自然场景图象的文字位置。

针对单个特征检测方法的不足,结合角点信息和颜色信息,本文提出了角点提取和颜色聚类的文字定位方法。分析两类特征结合方法的检测结果,针对两类常见的误检,采用笔划密度分析和区域周期性的判断,再利用文字区域具有的先验知识筛选和平滑文字候选区域,提高了定位能力,能够得到规则的文字区域和比较精确的位置信息。

关键词: 自然场景 文字定位 角点检测 颜色聚类 连通分析

Abstract

In the scene of natural there are lots of graphics and images which contain a wealth of text information. This information is very important for us to describe and understand the natural scene and it plays an important role in our life. As the bottleneck of the promotion of natural scene character recognition, improvement of text localization technical is necessary.

This paper focuses on the method of natural scene text detection and localization method. Corner detection based natural scene text localization methods, color segmentation and clustering based natural scene text localization is used in this paper. A method combined the above two algorithm is also used to locate the text. Stroke and periodic of texture are join in the method to improve its performance.

Because of massive existence of corner in text area, this paper use the corner feature to the text localization and use SUSAN corner detecting algorithm to find the corner. Firstly, detecting the corner point in natural scene images, and then using the templates to classify the corner point, finally using the Euclidean distance based clustering to obtain the text area. This algorithm in complex background text has a better positioning capability.

By analysis the distribution of the colors in the natural scene image, the second location method is based on color space segmentation and color clustering. First, the color of natural scenes is divided into several categories in accordance with certain rules, each category can be conversed to a sub-image, then use the connected component analysis method to extract the text area, and finally combine the text area of the sub-image into the final result of the text area. This method is mainly for the initial segmented contingent image.

Each method has its own insignificance, so this paper find a new method of text localization that using the above two methods of text localization algorithm. This method is the combination of rough text location base on corner detection and clustering, better text location based on color segmentation and clustering, selecting based on stroke texture information and periodic of texture. Finally get the precise and smooth text region by the selecting based on prior experiment. Excise results show that this method of natural scene text detection has a better positioning accuracy and robustness.

Key words: Natural scene Text localization Corner detecting Color clustering Connected component analysis

目录

第一章 绪论	1
1.1 课题的研究目的和意义	1
1.2 自然场景中文字定位系统构成以及研究现状	3
1.2.1 自然场景文字的特点	3
1.2.2 自然场景文字定位的研究现状	5
1.3 论文结构	9
第二章 基于角点聚类的自然场景文字定位	11
2.1 引言	11
2.2 常见的角点检测方法	11
2.2.1 基于灰度图象的方法	12
2.2.2 基于二值图象的角点检测	13
2.2.3 基于轮廓曲线的角点检测	13
2.3 基于 SUSAN 角点聚类的文字定位	14
2.3.1 SUSAN 角点检测	15
2.3.2 角点分类	17
2.3.3 角点聚类	19
2.4 测试和总结	20
2.4.1 数据库	20
2.4.2 测试方法和标准	21
2.4.3 算法性能测试	23
2.4.4 测试结论	25
2.5 本章小结	25
第三章 基于颜色聚类的自然场景文字定位	27
3.1 引言	27
3.2 彩色图象颜色分割和聚类技术	27
3.2.1 直方图阈值法	27
3.2.2 特征空间聚类	29
3.2.3 基于区域的方法	30
3.2.4 基于物理模型的方法	30
3.3 基于颜色聚类的文字定位	31
3.3.1 色彩空间的选择和转换	32
3.3.2 基于二维颜色统计的二值聚类	33
3.3.3 基于 HSL 色彩空间分割的分层聚类	36
3.3.4 基于先验知识的聚类结果优化	37

3.4 测试和分析	38
3.5 本章小结	39
第四章 角点特征和颜色特征结合的文字定位	41
4.1 引言	41
4.2 两者结合的处理流程	41
4.2.1 基于角点特征的粗定位	42
4.2.2 基于颜色信息的筛选	43
4.2.3 基于笔划和周期信息的特定区域排除	44
4.2.4 基于几何特征和文字行分析的细定位	45
4.3 算法实现和评测	47
4.3.1 算法的实现和测试环境	47
4.3.2 结果分析	48
4.4 本章小结	49
第五章 总结和展望	51
5.1 总结	51
5.2 展望	51
致谢	53
参考文献	55

第一章 绪论

在人类的感知方式中，作为最直观的信息传递形式的视觉具有举足轻重的地位。数字图像处理（Digital Image Processing），它是指将图像信号转换成数字信号并利用计算机对其进行处理的过程。早期图像处理的目的是改善图像的质量，减少存储和传输的开销，图象处理主要是针对图象数据本身，如图像增强、复原、编码、压缩等，对图象内容的理解并没有发展起来。随着电子技术的进步，设备运算和处理能力的提高，图像处理技术在许多应用领域受到广泛重视并取得了重大的开拓性成就。计算机技术和人工智能、思维科学研究的迅速发展，数字图像处理也随之向更高、更深层次发展，人们已开始研究如何用计算机系统解释图像，即让计算机系统能依照人类的视觉系统的模式理解外部世界，这被称为图像理解或计算机视觉。因为人类对自己的视觉认知过程还了解还不够充分，因此图像理解是一个有待人们进一步探索的新领域。现代图像处理技术结合了相关学科最新的理论和方法，如数学形态学、人工神经网络、模糊逻辑等，这些技术的引入使图像处理更加智能化。

1.1 课题的研究目的和意义

自然场景中蕴含着丰富的文字信息，例如，道路名称、门牌号、公共汽车站牌、商店名称、广告牌和交通标识上的文字等。这些文字对场景的认知具有重要价值，是描述和理解场景的关键线索，在日常生活中起着重要的作用。图 1.1 列举了两幅包含文字信息的自然场景图象，左边是指示方向的路牌，右边是大型建筑物内的交通导向牌。如果能够利用计算机技术自动识别这些文字信息，可以对人们的生活提供极大的帮助。例如可以通过手机拍摄环境中的路标、建筑物的名称文字和门牌编号，自动检测、提取、识别这些信息，翻译显示这些信息为来旅游的外籍人士提供帮助。

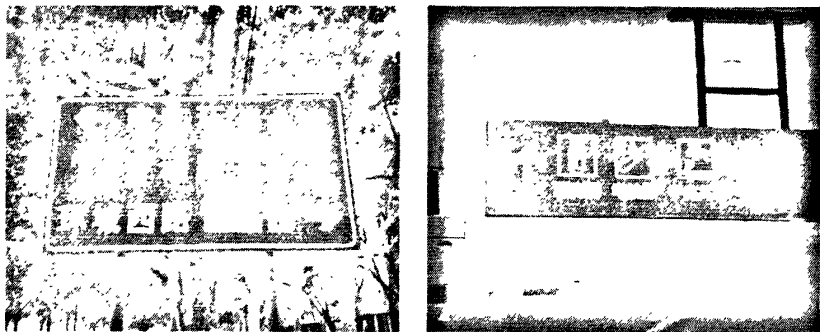


图 1.1 包含有价值信息的自然场景图象

自然场景中的文字认知是指在对二维文字或三维文字所依存的环境无限制或较少限制的认知条件下,对文字的识别与理解。作为自然场景中文字信息获取的核心技术,文字检测定位和字符识别更是吸引了众多的研究者致力于此。随着研究成果的不断出现,技术水平的不断提高,文字认知的应用领域得到极大的扩展。自然场景的文字认知技术主要有如下几个方面的应用:

- 1) 图象检索与分析系统。随着网络技术的发展,海量的图片和视频应用都面临快速检索的问题,原本依靠人工添加标签用于检索的方法已经远远无法满足需求,Google 已经开始简单的自动检索应用。而图象中包含的文字信息经过自动提取可以应对这方面的需求。
- 2) 智能人机接口和自动化。例如自动化的文字识别加上翻译系统可以让人在陌生的语言环境中获得需要的信息,提供更多的便利。智能机器人在理解场景的时候,只靠场景内的建筑、物体和人群等信息来识别是很有难度的,假使他能够识别场景中的文字,那么对场景的认知也将更加准确。
- 3) 文档处理和数字化图书馆。将纸质文字,图片等利用文字识别技术转化为电子版并加以分档归类,有利于信息的存储和检索。
- 4) 智能视频监控系统和智能交通管理。例如利用文字定位和识别,可以在监控的视频序列中检测违章的汽车并记录车牌号。

自然场景文本提取的系统是由图象采集端和处理输出端构成的,对于以算法设计和验证为目的的识别系统,本文主要关注属于处理输出端的组成部分。这部分主要由以下五大模块构成:

- 1) 数据库。数据库包括源图象和文字信息数据两部分。图象可以是静态图象,也可以是动态的图像序列或者视频,通常由数码相机,数码摄像机等设备采集得到。文字信息数据包括对图象上的文字预先人工定位和识别的结果,用来对比自动识别的结果,检测系统性能。本文所有的测试都是基于本文作者所在实验室自建的 5000 幅包含文字的自然场景图片及其人工识别结果的数据库。
- 2) 文字定位模块。从图象中检测文字,返回文字的位置,大小等信息。根据字符的样式以及结构,常见的文字定位方法是利用文字区域的最小外接矩形来标识文字的位置。本文也是采用文字区域的外接矩形来定位文字。
- 3) 文字校正模块。文字定位返回的区域图象,虽然相对整幅图象,文字区域的色彩、纹理相对简单,但是,由于文字存在三维投影变形,而且其大小、颜色都不相同,因此此时的文字块还不适合直接进行识别。对文字区域的图象进行进一步的切割,将所有非单字符的文字区域其分割为只包含单个字符的分块。将每个分块转化为白色背景和黑色字符的二值图象,修正三维投影变换并归一化文字大小以便于识别。

- 4) 文字识别模块。将上一模块得到的同一尺寸的图片上的文字识别为机器系统可以使用的编码。这一部分的技术是比较成熟的, 现阶段的 OCR 系统已经有了很高的识别率。
- 5) 文字识别性能测试模块。本模块只在工程测试阶段使用, 用来测试系统性能, 比较算法。

目前, 在自然场景文字识别的核心技术——文字定位和文字识别中, 用于文字识别的光字符识别技术(OCR)已经比较成熟, 但是他对文字的要求比较高, 对于不规则的, 存在大量噪声的文字, 通过方正阅读器自带的 OCR 模块测试, 其识别性能会急剧下降。因此, 对于自然场景的文字认知, 文字定位的性能非常重要, 但目前还没有普遍适用的定位算法, 大多算法都存在较强的样本针对性。可见, 自然场景中文字认知技术具有非常重大的理论意义和实用价值。就目前来说, 虽然复杂背景情况下的文字定位有许多大学和机构在研究, 但是其系统性能还不足以达到产品化的程度, 尤其是文字定位技术还有待进一步突破, 因此文字定位也是非常具有挑战性的。

1.2 自然场景中文字定位系统构成以及研究现状

1.2.1 自然场景文字的特点

机器要识别自然场景的文字, 一般都是通过图象采集装置得到的二维图象来分析、定位和识别文字。自然场景的文字也是人工制作的产物, 除了少数存在艺术字体的招牌, 制作时产生的大多数文字都是规则的, 表现为同文字行内字符字体相同、大小相近, 颜色较为单一。但是图片上的文字是自然环境中文字在复杂光照条件下三维到二维的映射, 因此, 规则的文字会产生一系列变化, 相对于传统的扫描文字, 自然场景下的文字的特点又有所区别, 以下从不同的几个方面介绍自然场景中的文字的特点。

- 1) 几何特征。一般来说, 同文字行的文字是沿着水平或竖直的基准线等高均匀排列。由于三维到二维投影的变换, 大多数的文字行都不是在水平或竖直方向上, 而是以一定的角度斜向排列, 同文字行的文字也不完全等大小, 越接近图象的消逝点文字相对就越小。如图 1.2a, “好家庭”文字行从左到右斜向上排列, 而且接近左方消逝点的“好”字要比“庭”字更小一些。弧形的以及不规则的文字行也大量存在, 如图 1.2b, “SKYTEAM”文字就是沿弧形结构排列的。文字的大小的分布也很分散, 大的可以占到整幅图的 50%以上, 如图 1.2d, 字符“地”几乎占满了整个图片; 小的可以不超过 5×5 大小, 如图 1.2c, 右方标牌“文明单位”下方的黑色文字小到

几乎难以辨认；这些过大的文字往往会因为提取的特征结构密度不足而产生漏检，而小文字则容易无法提取到足够特征结构也被漏检。

- 2) 纹理和结构。从频域上来说电脑文字和印刷文本文字的文字区域一般存在较多高频成分，这一点对于自然场景的文字也是一样的。在空间结构上，文字拥有较为明显和丰富的边缘和角点特征，尤其是汉字，由于其四四方方的结构和大量交叉的笔划，其角点特征更是丰富。字符的分布具有一定的周期性，表现为字符和空隙交替排列，字符中心与空隙中心各自的间距是近似相等的，如图 1.2e。
- 3) 颜色信息。一般来说，同文字行文字颜色是一致的，但实际环境中，由于不同的光照条件对色彩的影响，在光照和阴影的作用下，原本是同一颜色的同一个字符的不同部分也可能变成差异较大的颜色。如图 1.2f，原本白色的文字，在受到树叶阴影的遮盖后亮度变低，更接近灰色。过暗和过亮的环境也会对文字和背景的对比度产生影响，如图 1.2g，原本对比清晰的蓝底白字在低亮度下色彩都向灰色方向转变，可识别程度大大降低。
- 4) 文字的背景复杂，背景中包含类似文字结构的区域，对于很多常见特征，拥有近似于文字的提取结果。如图 1.2h，在用角点提取的方法分析文字区域时，树叶的边缘顶点和树枝的交叉点很容易被检测为角点，因而这类区域也大量且密集地存在角点特征，让分类器难以将他们和文字区域区分开来。又如窗户、建筑外墙的瓷砖网格等均存在丰富的边缘信息和类似于汉字的笔划结构，也很难与文字区域区分开来。

基于自然场景内文字的特点，对于自然场景中的文字识别，直接运用传统的 OCR 系统很难获得有效的识别结果，必须经过定位、提取和优化，将自然场景内的文字转化成简单的白背景上的单个单色字符才可以用 OCR 技术识别，获得良好的识别结果。



a 招牌上存在斜向文字行



b 标牌下部存在弧形英文文字



图 1.2 自然场景文字部分特点样本示意图

1.2.2 自然场景文字定位的研究现状

对于自然场景文字识别的核心技术，文字定位和文字识别，其中的光字符识别技术（OCR）已经比较成熟，但是他对文字的要求比较高，需要背景单一，字

符规整。对于复杂背景上的形状不规则的，存在大量噪声的文字，通过方正阅读器自带的 OCR 模块测试，其识别性能会急剧下降。在任意一个 PDF 文档内选取中英文混排的文字块，共 200 个字符，OCR 识别准确率可以达到 97.5%。而对 20 幅自然场景图片内的文字直接用 OCR 识别，OCR 系统无法区分文字与非文字区域，基本无法识别。由此可见，对于自然场景的文字认知，文字定位的性能非常重要。就目前来说，自然场景的文字定位还是很有技术难度的，识别技术的发展还没有达到商业化应用的程度，迫切需要进一步提高算法的性能。

正因为自然场景文字定位的价值和其高难度性，国内外许多研究者和研究机构都致力于自然场景图像中文字定位的研究，其中比较有代表性有国际文档分析和识别大会(ICDAR)组织的文字定位比赛^[1]。ICDAR 是文档识别领域最具权威的国际学术会议。与前几次的 ICDAR 相比，09 年会议发表的论文更多地偏重于有明确的应用背景，有一定限制条件，与实际应用结合紧密的方面，纯粹的自然环境或复杂背景的文字定位理论研究偏少。非英语语种的手写文字识别在 09 年的竞赛中占据了很大的比重，这与市场上掌上设备的智能化和普及化趋势是一致的。ICDAR2003 & 2005 (7th & 8th International Conference on Document Analysis and Recognition)自然场景文字识别比赛(Robust Reading Competition)分为三个分赛：文字定位(Robust Text Location)、字符识别(Character Recognition)和单词识别(Word Recognition)。表 1.1 是两次文字定位比赛前三名的结果。可以看出，虽然文字定位技术取得了一定的进步，但识别性能仍然比较低。

表 1.1 ICDAR 竞赛成绩

时间	参赛者	准确率	召回率	综合	耗时(s)
2005	Hinnerk Becker	0.62	0.67	0.62	14.4
	Alex Chen	0.60	0.60	0.58	0.35
	Qiang Zhu	0.33	0.40	0.33	1.6
2003	Ashida	0.55	0.46	0.50	8.7
	H.W.David	0.44	0.46	0.45	0.3
	Wolf Todoran	0.30	0.44	0.35	17.0

根据所使用的特征，自然场景中文本定位方法主要可以分为三种：基于连通域分析的方法、基于边缘特征的方法和基于纹理特征的方法。

1) 基于连通域分析的文本定位方法^{[2][3][4]}

基于连通域分析的方法是建立同一个文字区域或文字行的字符颜色、亮度相近并且可以与背景颜色区分的基础上的。利用不同的特征信息将图象分成若干类，在每个分类上寻找存在的连通区域，然后用先验知识来分析连通区域，得到文字区域。该类方法大致可以分为四个步骤：首先是颜色聚类，可以是在彩色空间上

的直接聚类,降低图像的彩色数量。RGB、HSV 和 YUV 及其近似变种的彩色空间都是常用彩色划分空间,而欧氏距离是最常见的划分度量方式。在灰度图上进行的二值化实际也是一种色彩分割方式;其次是连通区域生成;然后是根据先验知识去除非文字区域的连通区域;最后连通区域合并或再分割,生成最终的文字区域。由于算法实现相对简单,因此被广泛应用。

但是在颜色复杂的图像中,光照阴影等造成字符颜色的变化会对分割产生很大影响,在这种情况下该方法可能将一个字符划分到不同的连通区域中,甚至是把字符和背景划分到一个区域中;同时分割阈值的选择对性能也有很大的影响,选择自适应阈值还是固定阈值也是影响系统性能的重要因素。最后的区域合并算法也对系统性能有较大影响。

2) 基于边缘特征的文本定位方法^{[5][6][7]}

边缘检测是图像处理和计算机视觉中的基本问题,也是图像处理和计算机视觉中一个重要研究领域。自然场景文本字符和背景通常具有较高的对比度,因此文字字符和背景之间存在比较明显的边缘。基于边缘特征的方法就是根据文字区域丰富的边缘信息进行文字区域的检测。边缘检测方法很多,这其中的绝大部分可以划分为两类:基于查找和基于零穿越。常用的边缘检测算子有 Sobel 算子, Roberts 算子等。Roberts 算子特点是边缘定位准,但对噪声比较敏感。Prewitt 算子加入了平均和微分,对噪声有一定抑制作用。Sobel 算子与 Prewitt 算子类似,不同的是采用了加权的模板系数,在实际中最常用。Laplace 算子是不依赖边缘方向的二阶微分算子,是一个标量算子而不是一个向量算子,具有旋转不变性即各向同性的性质,但拉氏算子在实际应用中对噪声敏感,因此在实际中通常和高斯滤波结合起来使用。Canny 算子对信噪比与定位乘积进行测度,得到最优化逼近算子,是最优的阶梯型边缘检测算法。自然场景的彩色图象一般是要转换成灰度图象来处理,通过在灰度图上检测边缘来寻找文字区域。彩色图象的信息全部包含在彩色模型的各个分量之中,灰度图像只包含明暗信息,将彩色图象转换为灰度图,丢失了一些信息,因此张云鹤^[8]等人提出了直接针对彩色图像进行文字检测的算法,获得更好的检测效果。

基于边缘特征的文字检测方法的优点是时间复杂性比较低,可以满足实时应用的需要。但是当背景复杂时,虚警率较高。

3) 基于纹理特征的文字信息提取方法。

纹理是对图象的象素灰度级在空间上的分布模式的描述,反映物品的质地,如粗糙度、光滑性、颗粒度、随机性和规范性等。当图象中大量出现同样的或差不多的基本图象元素(模式)时,这类图象适合于用纹理分析来研究。基于纹理特征的方法将图像中的文字区域视为具有某种特殊纹理的图像区域,并应用其他的图像纹理特征,如边缘、角点等检测文字区域。

描述一块图象区域的纹理有三种主要的方法,即统计分析方法、结构分析方法和频谱分析方法^[9]。统计方法有自相关函数、纹理边缘、结构元素、灰度的空间共生概率和自回归模型。结构方法主要研究基元及其空间关系。基元一般定义为具有某种属性而彼此相邻的单元的集合,其属性包括连通区域的形状、灰度、局部一致性等;其空间关系包括基元的相邻性、在一定角度范围内的最近距离等等。根据基元间的空间联系,纹理可以分为弱纹理或强纹理。纹理也可以根据基元的空间共生频率或单位面积内的边缘数来进一步细分。基元也可以看作是灰度行程。频谱方法是建立在基础上的,根据傅立叶频谱的波峰、波谷分布等特征将图象分类。常用的参量包括峰值的数量、峰值的幅度和相位、峰值之间的距离和峰值之间的相角差等。

通常采用 Gabor 变换、小波变换和傅里叶变换等方法检测图像中文字区域的纹理特征。

由于基于纹理的方法是根据文字的特殊纹理进行定位,因此对于一些背景非常复杂,并包含与文字相似纹理特征的图象,该方法处理效果不好。另外由于计算图像纹理时需要扫描全部图像并做卷积运算,因此该方法的计算量很大,处理时间较长。

总的来说,现有的文字定位方法对所处理的文字类型具有一定的选择性,即文字定位方法对特定类型的文字具有较好的提取效果,而对其他类型的文字则效果较差。自然环境中文字的排列方式、光照不均匀、大小、排列方式、复杂背景等因素,使得场景文本定位问题非常具有挑战性。所以单独采用一种特征或方法,都不能取得令人满意的结果。现有的很多文字定位算法都是基于多方法组合而得到的。

此外,由于文字大小的变化,多尺度分析也是常用的方法。采用金字塔分解法将图象分解成多幅不通尺度的子图,在子图上采用各种方法进行定位,然后对所有子图的定位结果进行加权组合,也对定位效果的提高有一定的帮助^[10]。

由于自然场景中文字区域的复杂多变性,文字特征分布密度随文字区域的变化而改变,对于特征参数的分类,也可以由固定阈值、简单自适应阈值提高到采用较为复杂的分类器来分析特征,基于学习的文字定位正好可以满足对复杂分类器参数确定的要求。常见的有支持向量机^[11](SVM),神经网络等方法。

自然场景文字定位系统已经成为文字识别系统商业化的瓶颈,国内外各个研究机构都在寻求新的突破,开发出高准确率、高召回率、高鲁棒性的文字定位系统。多尺度分析,多方法结合以及人工智能方面成果的引入已经是文字定位方法的重要发展趋势。手机, PDA, PAD 等小型的智能设备的迅速发展也对文字定位算法提出了计算速度等方面的新要求。近几年来,该领域的文献增长十分迅速,说明自然场景文字定位技术已经引起了国内外大量研究者的研究兴趣。

1.3 论文结构

本文根据自然场景中的文字特征,以自然场景中拍摄的数字图象为研究对象,重点研究了自然场景中的文字定位的理论和方法。实验室现有的定位方法有边缘结合 RGB 颜色分割^[12]、边缘结合 Gabor 滤波^[13]、边缘结合刷子算法^[14]、神经网络^[15]等,区别于这些算法,本文实现了基于角点信息和颜色聚类的文字定位系统。该系统利用数字图像处理技术和识别技术对图像,提取纹理、颜色等方面的特征,通过数学形态学处理,利用聚类 and 连通域分析的方法分析确定文字区域,最后利用文字的几何特征等方面的先验知识分析优化定位的结果,得到自然场景的图象中文字的位置信息。

本论文各章内容安排如下:

第一章:绪论。重点介绍文字分析包括文字定位概述,结构以及发展现状。其次介绍了几种检测定位技术,如基于边缘检测的文字定位、基于连通量的文字定位,基于色彩的文字定位等,并分析了目前多特征联合检测以及加入复杂分类器的发展方向。

第二章:基于角点聚类的文字定位。首先分析了角点在文字定位中的价值,介绍了常见的角点检测理论,重点研究了 Harris 角点检测方法和 SUSAN 角点检测方法。实现了基于 SUSAN 角点提取法和角点聚类的文字定位,并采用 ICDAR2005 的算法评价方法对定位性能进行测试。

第三章:基于颜色聚类的文字定位。分析了颜色信息对于文字定位的意义,概述了常用的颜色分割及聚类方法。采用了两种有针对性的基于不同的颜色分割方法的颜色聚类算法进行文字定位,测试了算法性能并分析了算法的优缺点和适用的类型。

第四章:角点和颜色特征结合的文字定位。首先通过上两章的工作分析基于角点、颜色特征的文字定位的互补性,即两种方法结合使用的可行性。其次根据上两章的算法组合出新的算法流程。然后研究了对初步定位的结果的后处理,这些方法的目的是提高定位准确率,利于后续的处理和识别。最后通过软件实现算法并在数据库上进行了测试,给出了实验结果。

第五章:总结和展望。总结本论文所做的工作,分析不足之出并提出可能的改进方向。展望本论文相关学科未来的发展方向和应用前景。

第二章 基于角点聚类的自然场景文字定位

2.1 引言

角点是图象的一个重要的形状特征，它集中了图像上的很多重要的形状信息。在连续的图像上，角点就是目标轮廓局部曲率极大点。如图 2.1 所示，汉字和字母，尤其是汉字，其边缘可以看成是大量直线段和少量曲线段连接在一起构成的，因此存在大量的角点信息。相对应的，招牌背景上角点的密度就要低很多了。区别于常用的边缘信息，角点能够在一定程度上避免长直线和曲率较小的曲线对特征提取的干扰，可见，角点作为一个特征来分析图像，寻找文字区域是有其可行性和价值的。

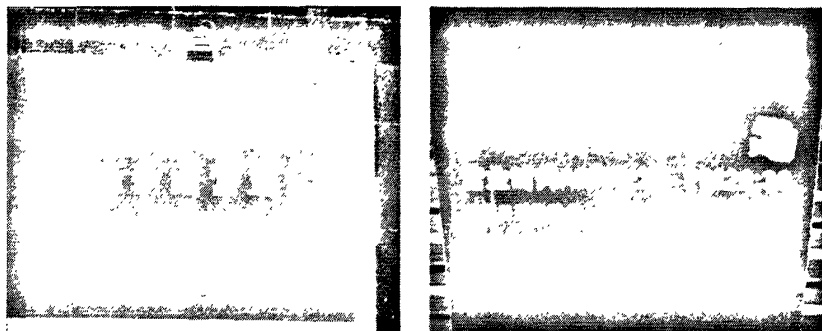


图 2.1 图象的角点特征示意图

利用角点特征来检测文字，也有如下优点：

- 1) 大部分方法检测的角点具有旋转不变性；
- 2) 受光照条件的影响较小；
- 3) 角点包含的像素点数量一般小于 5% 的图象总像素点数，利用角点进行图象处理，在保留了大量的结构信的基础上最小化了要处理的数据量。

2.2 常见的角点检测方法

对于离散的数字图像，角点在目前还没有统一的定义，不同的检测方法对角点有不同的定义。早在 1975 年，Rosenfeld 等人^[16]和 Freeman H 等人^[17]就提出通过计算角点强度 k 来检测角点，这种方法计算简单，但抗噪声性能较差，检测效果不是很理想。为了弥补这个方法的缺陷，Asada 等人^[18]提出通过对边缘采用高斯平滑，降低将局部弯曲度的突然增大点误判为角点的概率。由于角点强度 k 是通过预先确定得到还是根据曲线的弯曲度自适应调节得到，对于检测的结果影响很大。Witkin^[19]提出一种多次阈值迭代计算曲线上任意点弯曲度的自适应算法，集中曲线边缘点中满足预定条件的点，组成候选角点集合，然后增加平滑参数开始

新的循环,直到平滑参数达到预先设定的最大值为止,最后将所有候选角点集合中出现次数满足一定比例的边界点定义为角点。

现阶段已有的角点检测算法可以分为三类:基于梯度的角点检测、基于边缘特征的角点检测和基于亮度变化的角点检测^[20]。

2.2.1 基于灰度图象的方法

基于灰度的方法常见有如下几类。

- 1) 基于梯度的方法是通过计算边缘的曲率来判断角点是否存在,角点响应的大小取决于边缘强度和边缘方向的变化率,该方法对噪声比较敏感^[21]。Kchen 和 Rosenfeld^[22]给出了具体的角点检测算子 K ,通过检测算子 K 在图像某一邻域的极大值来提取角点。该算子如式 2-1 所示,它表现为水平面截线上某点 (x, y) 的曲率与该点的最大梯度的乘积。

$$K = \frac{tp^2 + rq^2 + 2spq}{p^2 + q^2} \quad (2-1)$$

- 2) 基于小波变换的方法。基于小波的方法实际上是在小波变换后的图象上提取特征。例如基于多尺度小波的角点检测,就是通过采用不同尺度的小波变换,利用各个尺度的判断准则在对应尺度上检测角点的存在。首先对图象进行小波变换,得到各个尺度上的小波分量;利用各个尺度上的小波分量在相应的尺度上提取角点并记录这些角点的位置,从大尺度到小尺度检测角点,同时剔除当前尺度与上一尺度不同的点,最后得到最终的角点分布。
- 3) 基于模板的方法。基于模板的方法可以描述为检测像素邻域点的亮度变化,统计当前点与模板内点的亮度对比,如果当前点与近邻点亮度对比足够大,那么当前点就可以定义为角点。Kitchen-Rosenfeld 算法^[22]通过模板窗口局部梯度幅值和梯度方向的变换率来度量角点响应强度,判断角点的存在性。Harris 等人则采用一个在图像上移动的高斯窗或矩形窗^[23],以通过滑动模板窗口取得的由原图像计算出的 2×2 局部结构矩阵 M 为窗口函数来计算角点响应。该方法的一个比较突出的优点是具有旋转不变性,但 Harris 方法检测得到的角点有较大的冗余。类似 Harris 算法, KLT 角点检测算法也是基于一个 $D \times D$ 的计算窗口模板计算图像计算局部结构矩阵。窗口的大小对检测结果有较大影响,太大的窗口会引起角点移动,太小的窗口则会丢失相距较近的角点。SUSAN 法也是这一类中一种常用的角点检测方法。

2.2.2 基于二值图象的角点检测

这类方法是在二值化后的图像上提取角点。刘文予等^[24]提出一种基于骨架的角点检测方法,该方法认为原始图像可以近似转化为若干个多边形,那么多边形骨架的延长线上一定存在角点,且角点所对应的骨架点的最大圆盘半径应该趋于0。那么角点就是骨架中的最大圆盘半径为0的点。由于二值图象的处理计算量并不是很大,所以算法可以保证计算的实时性。这里将基于二值图象的角点检测为一个单独的类型列出来,虽然二值图像在边缘对比度,噪声等方面不同于灰度图像,但是针对灰度图像的大多数处理方法仍然可以使用。作为灰度图到边缘轮廓图的中间步骤,专门针对二值图像的角点检测方法相对较少。

2.2.3 基于轮廓曲线的角点检测

1) 计算角点强度

角点强度的计算基于角点的数学定义,计算较为简单,但轮廓曲线上的毛刺等局部的极小区域不平滑部分会导致角点强度异常,效果不是很理想,很多算法都针对这个问题做了改进。肖茜等^[25]指出自适应的弯曲度测定实际上是要自适应地确定曲线段的支持区域的大小,支持区域大小的选择应该能够根据曲线的弯曲程度自适应地调整,在此支持区域上计算的曲线弯曲度才能较为准确地反映平面对象边界曲线的平滑和弯曲程度。Witkin^[19]提出一种多次阈值迭代计算曲线上任意点弯曲度的自适应算法,集中曲线边缘点中满足预定条件的点,组成候选角点集合,然后增加平滑参数开始新的循环,直到平滑参数达到预先设定的最大值为止,最后将所有候选角点集合中出现次数满足一定比例的边界点定义为角点。钟宝江等^[26]认为对于数字图象,基于像素基础的曲线是不连续的,曲线相邻点之间的距离并不均匀。Witkin 算法中认为数字化曲线上相邻两个点之间的距离是一个常数,钟宝江等人认为在实际中该假设并不成立,实际距离应该考虑曲线方向,在倾斜曲线段上相邻点的距离应该是 $\sqrt{2}$ 。基于这个发现,钟宝江等提出了基于曲线累加弦长的角点检测方法,算法的主要思想是在确定支持域时充分考虑相邻像素点之间的实际距离,并根据这一推论提出了一种新的角点强度计算公式。利用该算法可以对如尖角和圆角进行区别,检测结果具有旋转不变性。

2) 计算曲线曲率的极值

计算轮廓曲线曲率的极值也是检测角点的一类方法。计算曲线的曲率,一种方法是直接对离散的曲线进行计算,另一种方法是用连续函数分段拟合原始曲线,然后根据拟合后的分段曲线方程计算曲线曲率极值,从而得到角点的位置。Langridge^[27]使用了三次多项式的曲线来拟合离散的数据点,Mediono^[28]采用了 B

样条来拟合曲线。基于曲线拟合的方法要先计算曲线的拟合方程,因此运算量比较大。Langridge 提出根据曲线上任意点的弯曲度,结合模糊识别的方法来检测对象边界曲线的角点。角点是曲线上曲率较大的点,基于这个前提,乔宇等^[29]提出可以根据当前轮廓点前后曲线的方向的变化幅度来检测角点,该方向的度量可以用一个方向差角 d_θ 来定义,如下式 2-2。差角越大,表示曲率越大。其中基于距离误差的直线拟合可以自适应地调整拟合窗口,有效地减少了边缘噪声的干扰。

$$d_\theta = \min\{|\theta_1 - \theta_2|, 360^\circ - |\theta_1 - \theta_2|\} \quad (2-2)$$

乔宇等人除了计算角点两侧的曲线段构成向量的夹角以外,还考虑其他特征对曲线角点进行多方位的描述,同时引入模糊集合的概念,采用角点前后点组成的向量与角点的距离特征、角点前后向量夹角特征、角点的前向直线特征、角点的后向直线特征这四个特征对角点进行描述。采用这几个参数来区分角点的有效性,对真实角点的相邻点具有较强的抑制作用,检出的角点类似于人类自身期望得到的角点。不足的是该算法只是基于角点的局部细节特性,没有考虑全局特征,因此存在一定的漏检现象。结合角点细节特征和全局特征来改进算法,可能会得到更好的检测效果。

3) 多尺度角点检测

曲线上各角点均有着不同尺度的支撑域,因此很难利用事先确定的单个尺度来得到最优的角点检测记过。因此,经常采用多尺度检测结合的方法检测角点。Witkin^[19]和 Koenderink^[30]提出基于尺度空间的图像分析理论,认为多尺度曲线分析可以较好地解决该问题。在曲线尺度空间中,随着曲线尺度由小变大,一直保持较高弯曲度的点就是所要求取的角点。为了满足这个要求,Rattarangsi 等^[31]提出一种基于尺度空间理论的角点检测算法。王展等^[32]采用围线跟踪算法和二阶导数零交叉边缘检测算子得到边缘曲线,使用一组自相似二进 Gabor 小波变换的滤波器将整个频域从高频到低频分为多个子带,对两个不同尺度下的滤波器输出求差并取模,根据结果即可判定曲线上任意点是不是角点。在上面的多尺度检测中,算法仅考虑了角点的位置信息,作为补充,戚飞虎等^[33]提出将角点的位置信息和对应角点的幅度信息结合起来分析。在选定小波为高斯函数的一阶导数后,对图像轮廓的 Freeman 链码投影成的函数进行小波变换,在不同的尺度上小波变换幅值始终保持最大且位置始终是不变的点就可以认为是角点的。对于噪声,文章认为噪声的幅值只存在于有限的尺度空间上,通过对波峰幅值和出现位置的分析就能够很好地剔除伪角点。从实验结果可以看出,该算法的准确性很高。

2.3 基于 SUSAN 角点聚类的文字定位

本文采用基于灰度图的角点检测,从速度和检测质量方面综合考虑,采用

SUSAN 角点检测获取角点，经过角点筛选和聚类分析得到文字的位置信息。基于 SUSAN 角点聚类的文字定位方法主要分为三个部分：角点检测、角点筛选和角点聚类三部分。

2.3.1 SUSAN 角点检测

近年来 USAN 或 SUSAN 角点检测算法得到越来越多的关注，最小亮度变化算法(MIC)^[34]、同值分割吸收核(Universal Segment Assimilating Nucleus, USAN)算法^[35]都是基于像素邻域半径为 k 的圆形模板。SUSAN 是 Smith 和 Brady 提出的一种图像处理方法，该算法是基于像素邻域包含若干元素的近似圆形模板，对每个像素基于该模板领域的图像灰度计算角点响应函数 (CRF) 的数值。给定一个阈值，并用它来划分角点响应函数的返回值，将符合条件而且是局部极大值的点提取为角点。SUSAN 算法通常采用半径为 3 个像素的圆形模板，对于离散的数字图像，模板就是一个 7×7 的矩形模板去掉四个角各自最远端的 3 个点剩余的部分，如图 2.2 左边模板，黑色覆盖超过一半的所有方格就是常用的近似圆形模板。简化的 3×3 模板也有一定应用价值，如图 2.2 右方部分，但是比较少见。

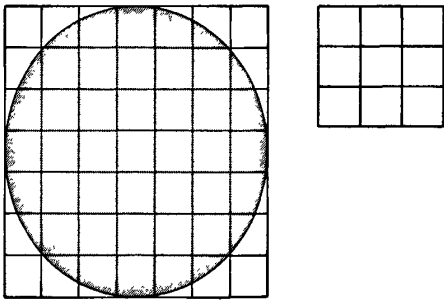


图 2.2 SUSAN 法使用的模板示意图

由本文实验和其他论文结果^{[21],[22]}可以看出，除了对单幅图像能进行角点检测以外，KLT 算法和 Harris 算法对图像序列的角点检测效果更好。KLT 算法比 Harris 算法检测角点的质量高，但 KLT 算法适用于角点数目不多且光源简单的情况，Harris 适用于角点数目较多且光源复杂的情况。而且 Harris 算法的计算公式带有一定的平滑能力，因此鲁棒性较强且对噪声不太敏感。Kitchen-Rosenfeld 算法和 USAN 算法一般来说不适合序列图像的角点跟踪，常用于单幅图像的角点检测。在单幅图像的检测中，USAN 算法的性能要优于 Kitchen-Rosenfeld 算法。

对于离散数字图像的计算，使用的模板是对圆形模板量化得到的，由于量化误差的干扰，容易导致边缘点和角点的判断混乱。对于边缘模糊的图象，使用小模板可能会丢失角点。为了解决以上两个问题，可以通过局部图象的自适应计算来确定最优的模板，杨莉等^[36]就提出乐模糊度的概念来解决这个问题。对每一个像素在计算其角点响应值之前首先测定其模糊度，若模糊度大于设定阈值，就使

用大的模板来计算；若图象清晰、模糊度小于设定阈值，则选用小的模板来计算；这种算法使得判定的准确性得到很大的提高，减少了虚警率。费旭东等人^[37]采用基于查表的方法来进行角点的快速提取，其特点是速度快，算法简单，以空间换时间的方法使其更适合于开发硬件系统。Giraudon 等^[38]采用高斯—拉普拉斯二阶微分算子来检测角点。高斯二阶微分函数与离散信号的卷积相当于高斯函数与信号的卷积再求二阶差分，其缺点是对噪声比较敏感。基于神经细胞感受野数学模型，陈燕新等^[39]提出双高斯差(Difference Of Gaussian, DOG)模型来检测角点。文章认为高斯二阶微分函数是DOG函数在其两个高斯函数相互逼近时的一个极端形式，DOG函数与信号的卷积相当于两个高斯函数与信号的卷积结果之差，具有较强的抗噪声能力。综上所述，这些角点都是通过模板与图像进行卷积运算来检测角点，所以也应该属于模板类的方法。

如图 2.3 所示为 SUSAN 圆形模板与图象中目标区域的 5 种几何位置关系，对于图像中非纹理区域的任一点，在以它为中心的模板窗中存在一块亮度与其相同的区域，这块区域即为 SUSAN 的 USAN (Univalue Segment Assimilating Nucleus) 区域，它包含了图像结构的重要信息。由图可知，当模板中心像素点位于区域内部时，USAN 的面积最大，当该像素点位于区域边界时，其面积为最大的一半，当该像素点为角点时，USAN 区域面积约为最大的 1/4。

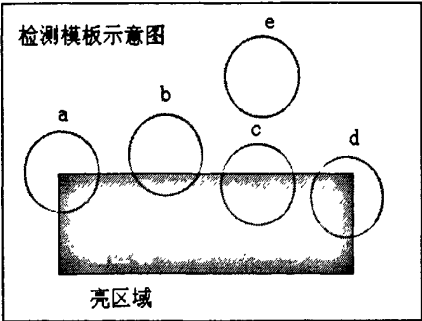


图 2.3 SUSAN 角点检测原理简化示意图

如上图所示，根据计算以检测点为中心的近似圆形模板内的亮度分布，可以确定当前点是否为角点。

基于以上原理，SUSAN 法检测角点的算法如下。

- 1) 选取图象上任意一点 r_0 为待检测点，以 r_0 为中心使用模板， r 为模板内其他点， I 函数表示对应点的灰度值， t 为预先设定或训练的阈值，那么计算模板内其他点和中心点的灰度差异函数可以用如下公式 2-2 或 2-3 来计算，其中 2-2 式是带有一定的平滑的计算式，计算量相对比较大，2-3 式是基于图 2.3 的简单数量统计算法，有较快的计算速度；

$$c(\vec{r}, \vec{r}_0) = e^{-\frac{(I(\vec{r}) - I(\vec{r}_0))^2}{t}} \tag{2-2}$$

$$c(\vec{r}, \vec{r}_0) = \begin{cases} 1 & \text{if } |I(\vec{r}) - I(\vec{r}_0)| \leq t \\ 0 & \text{if } |I(\vec{r}) - I(\vec{r}_0)| > t \end{cases} \quad (2-3)$$

- 2) 计算总的灰度差异函数值，即对所有 r 对应的上一步的计算结果求和，计算方法如下式 2-4；

$$n(\vec{r}_0) = \sum_{\vec{r}} c(\vec{r}, \vec{r}_0) \quad (2-4)$$

- 3) 计算角点响应函数，对应公式如下 2-5 式；

$$R(\vec{r}_0) = \begin{cases} g - n(\vec{r}_0) & \text{if } n(\vec{r}_0) < g \\ 0 & \text{otherwise} \end{cases} \quad (2-5)$$

其中， g 为角点响应函数阈值。

- 4) 判断存在非零角点响应的点是否局部最大值，是局部最大值的保留，否则去掉；局部最大值的比较范围大小和模板相同。
- 5) 遍历整幅图像，重复以上 1 到 4 步骤，找到图像上所有的角点。

经过以上处理，可以得到一幅角点特征图，如图 2.4。可以看出，部分角点是在边缘区域垂直于边缘走向的阶跃断点上，并不是真正本文所需的角点，因此，下一节就此问题做角点分类处理。

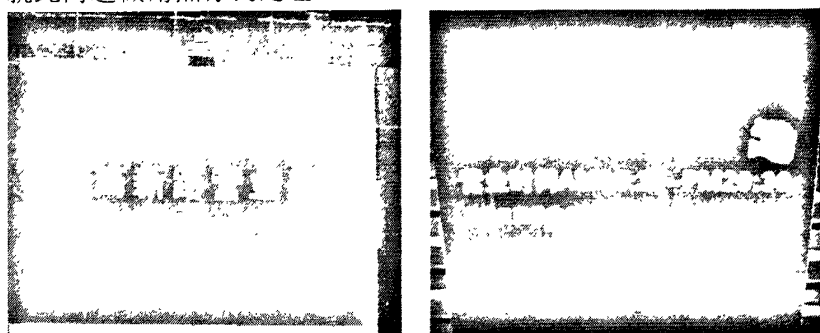


图 2.4 图象的角点特征示意图

2.2.2 角点分类

在自然场景中遇到的角点类型非常多，例如下面图 2.5 所示的几种角点。SUSAN 算法一般能够有效和快速地检测“L”、“T”、“Y”、“X”等类型角点，但是由于算法只考虑 USAN 区域的面积大小而没有考虑到 USAN 区域的形状，因此对于一些形状更为复杂的角点 SUSAN 不能有效地检测，同时如图 2.4d 的边缘点也有可能被当作角点检测出来。

针对上述情况，基于二值化后的模板区域灰度图，本文选择 2 个 SUSAN 模板圆环区域计算二值跳变数来解决问题。由于上文选 7×7 SUSAN 模板，那么这里的模板同样基于 7×7 模板来构建。如图 2.6 所示，其一是与中心像素最近的 8 个

像素组成一个圆环，其二是外层的 16 个像素组成一个模板。以当前点为圆心，检查 2 个圆环上像素的灰度跳变情况，像素值从 1 到 0 或从 0 到 1 算一次跳变。根据跳变情况将所有像素点分为 3 类：

- 1) 变化次数为 0 的区域内部点。这类点只可能有少量存在，可以排除。
- 2) 变化次数为 2 的边缘点或者“L”型角点，应用 SUSAN 算法的角点响阈值应来区分角点和边缘点。
- 3) 变化次数在 2 以上的复杂形状角点，在角点提取时可能由于面积原因已被排除，留下的都是角点，因此这里不做处理。

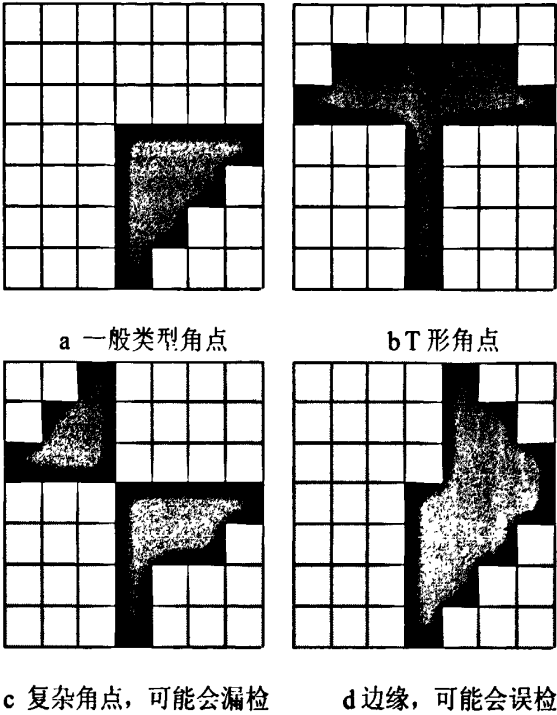


图 2.5 角点类型示意图

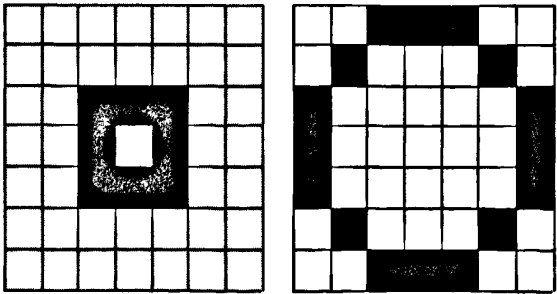


图 2.6 跳变数检测环形模板

除此之外，可以用计算重心来排除其他伪角点。计算 USAN 区域重心，判断 USAN 区域的重心和模板中心的连线所经过的像素点是否都属于 USAN 区域，然后计算重心和模板中心的距离，实际所需要的角点重心应该和模板中心有一定的距离，如果距离较小则不是正确的角点。判断 USAN 区域的重心和模板中心的连线所经过的像素点是否都属于 USAN 区域，这一方法基于 USAN 区域的一致性，

在有些存在较强噪声的实际图像中很有必要。

通过以上处理，可以去除大量边缘上出现的虚假角点，如图 2.7 所示。

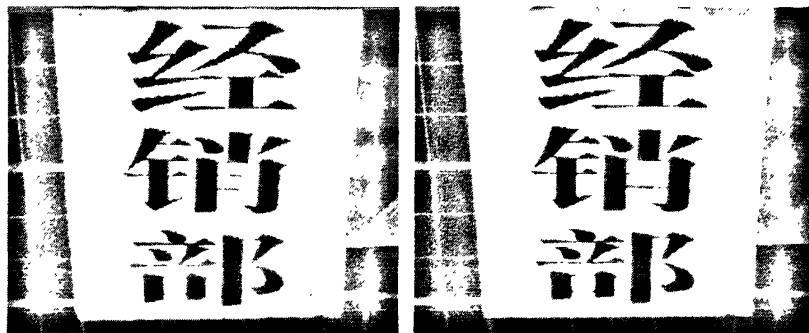


图 2.7 角点筛选前后对比

2.3.3 角点聚类

检测出来的角点只是图象的特征，还需要将其进一步处理得到文字区域。通常情况下，文本区域的角点分布比较密集，因此采用判断点之间的欧式距离来确定点所属的分类，包含一个分类所有角点的矩形框就是初步的文字框。对于任意两个角点，若他们在图象上的坐标是 (x_1, y_1) 和 (x_2, y_2) ，计算距离 d 如 2-7 式。设定阈值 T ，若两点距离 $d < T$ ，则认为两点属于同一类，否则不属于同一类。

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2-6)$$

具体算法流程如下：

- 1) 选取一个初始点作为第一个点集，其包含的点只有一个；
 - 2) 遍历下一个点，如果这个点和某一点集内其中一点距离小于阈值 T 的，那么将其加入这个点集；如果这个点和现有点集内所有点的距离大于阈值 T ，那么创建一个新的点集，此点集包含这个点；
 - 3) 重复步骤 2，直到遍历所有角点；
 - 4) 此时，所有的角点将被分到若干点集，去掉点数少于 4 的点集；
 - 5) 取其中一个点集，以任意两点为对角顶点生成一个矩形，判断下一个点和矩形框的位置关系，若点在框内，继续判断下一个点；否则向点的方向拉伸矩形，使这个点成为矩形的顶点；
 - 6) 遍历点集内其余点，重复步骤 5；
 - 7) 对每一个点集重复以上两步，得到的对应矩形框记为初步的文字框。
- 初步聚类得到的文本区域如图 2.8。

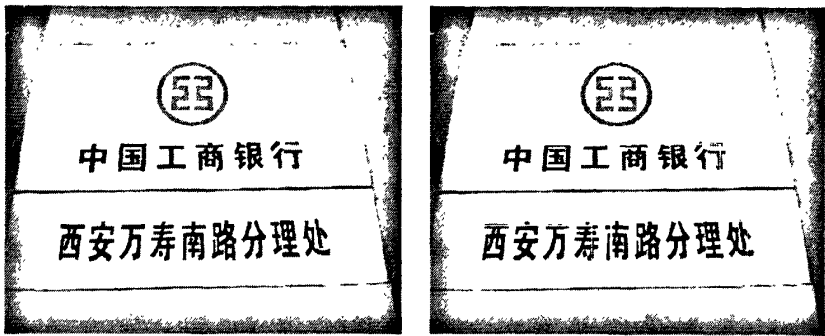


图 2.8 聚类得到的初步文本框示意图

2.4 测试和总结

2.4.1 数据库

本论文采用的数据库由实验室同学共同采集和制作。包括自然场景图片库和对应的标定信息库构成。图片库由西安市区内各种包含文字的自然场景图片数据构成，采集工具为手机，数码相机。图片数据库共包括 5000 张图片，大小均为 1600pix×1200pix，内容有交通标识、商店标牌、楼宇标识、广告牌以及其他包含文字的场景，拍摄的角度和拍摄时的场景光照具有较好的多样性，数据库具有较广泛的涵盖面。标定信息库的内容是与每张图片一一对应的文字信息记录表。表格记录的信息有如下几类，保存为 Excel 格式。

- 1) 对应自然场景图片编号；
- 2) 图片内文字行外接四边形的四个顶点坐标以及所包含字符的内容；
- 3) 文字行内每个词组的位置，内容；
- 4) 词组内每个字符的位置、内容和颜色。

图 2.9 为数据库中的样本在标定软件中标定结束后的显示示意图。

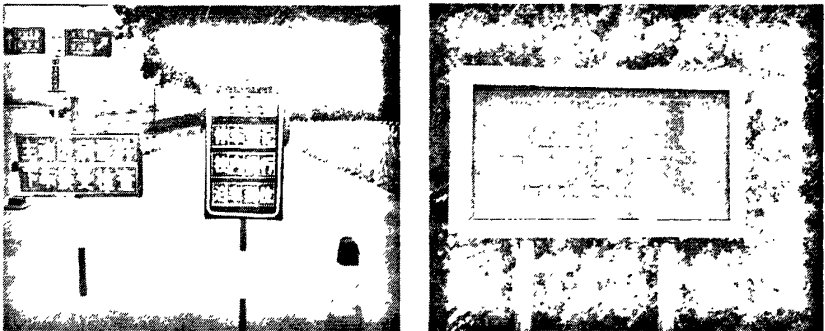


图 2.9 已标定的自然场景图像样本

表 2.1 为对应的 Excel 表格的部分节选，其记录信息与上图的方框对应，并包括文字和颜色信息。此表对应上图右图。从 A 到 D 的坐标表示方框从左上到右上逆时针排列的四个顶点，坐标信息从左到右依次为当前方框所在的蓝色框序号、

绿色框顺序号、红色框顺序号、点的横坐标和纵坐标。表 2.1a、2.1b 和 2.1c 分别是 从三层框中各选一个对应的文字和位置信息。

表 2.1 数据库自然场景图像上的文字信息表的部分节选

表 a 第一层蓝色框信息

1A 坐标	1B 坐标	1C 坐标	1D 坐标
(1,0,0,190,322)	(1,0,0,190,804)	(1,0,0,1246,804)	(1,0,0,1246,322)

表 b 第二层“学生公寓”所在文字行绿色框信息

2A 坐标	2B 坐标	2C 坐标	2D 坐标
(1,1,0,270,358)	(1,1,0,270,546)	(1,1,0,1172,546)	(1,1,0,1172,358)
第二层含义			
[学生公寓 students' Apartment]			

表 c 第三层字符“学”的位置以及文字信息

3A 坐标	3B 坐标	3C 坐标	3D 坐标
(1, 1, 1, 300, 382)	(1, 1, 1, 300, 474)	(1, 1, 1, 390, 474)	(1, 1, 1, 390, 382)
第三层含义	第三层尺寸	颜色	
[学]	8280	{白色}	

2. 4. 2 测试方法和标准

本文采用 ICDAR2005^[1]提供的开放图片集和测试方法对本文的文字信息提取方法，利用实验室现有的数据库选取部分样本对性能进行测试。本节给出文字定位中文字区域位置的描述方法和文字定位的评价指标：准确率、召回率和综合性能。

这里准确率 p 和召回率 r 的定义借鉴了信息检索系统中使用的对应概念。在信息系统中通常使用精确率和召回率用来量化信息检索系统的性能。对于给定的查询问题，设 T 是真实目标的数据集合， E 为估值，即查询系统返回的数据集合，若估值集合中正确的估值数量为 c ，那么准确率则定义为正确的估值数量除以全部估值数量，如公式 2-7；召回率则定义为正确的估值数量除以全部真实目标的数量，如公式 2-8；其中 $|R|$ 表示集合 R 内的元素个数。

$$p = \frac{c}{|E|}$$

(2-7)

$$r = \frac{c}{|T|}$$

(2-8)

用图像中文字区域最小外接四边形作为文字区域，将上述定义拓展到文字定位算法的性能的评测上。

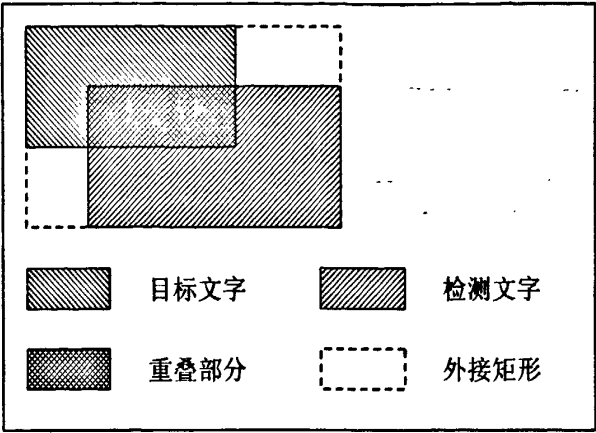


图 2.10 准确率和召回率计算示意图

如图 2.10 所示，考察目标文字区域外接四边形与算法返回文字区域外接四边形的匹配程度，对于分属于目标区域和算法返回区域的两个四边形框，定义参数 m_p 为两个四边形框的重叠的面积除以包含这两个矩形区域的最小外接矩形的面积的商， m_p 取值在 $[0, 1]$ 区间。当 m_p 取 0 时两四边形框无重叠， m_p 取 1 时两框重合。因此，矩形区域 r 与一个矩形区域集合的最佳匹配为 r 与集合 R 中的所有矩形区域的最佳匹配，如公式 2-9。

$$m(r, R) = \max(m_p(r, r') \mid r' \in R) \tag{2-9}$$

设 T 是图像中目标文字区域的外接矩形集合， E 是文字定位算法返回的文字区域的外接矩形集合，定义文字定位的准确率 p 和召回率 r 如下式 2-10，2-11，其中 $|R|$ 表示集合内的元素个数。

$$p = \frac{\sum_{r_E \in E} m(r_E, T)}{|E|} \tag{2-10}$$

$$r = \frac{\sum_{r_T \in T} m(r_T, E)}{|T|} \tag{2-11}$$

准确率和召回率以式 2-12 的方式结合起来形成一个综合性能指标 f ，其中 α 是加权系数，一般取 0.5，即准确率和召回率加权系数相同。

$$f = \frac{1}{\frac{\alpha}{p} + \frac{1-\alpha}{r}} \tag{2-12}$$

精确率描述定位的准确程度，越高则说明定位返回结果中正确的越多；召回率则表示了实际文字区域的检出率，越高则表示漏掉文字区域越少；综合评价指标是对精确率和召回率两者的综合。

运算速度的测试通过提取系统时间，计算相应函数的执行时间，精确到毫秒，有关的软硬件条件如下：

CPU: AMD Athlon X2 5000+ 2.6Ghz.
操作系统: Windows XP Professional SP3.
软件平台: Windows Visual Studio 2008.

2.4.3 算法性能测试

对于评价算法的速度,准确率,召回率等几大指标,这里首先测试速度指标。分别测试使用 2-2 式和 2-3 式时计算所耗费的时间。表 2.2 为两种方法运算 100 幅图片的时间均值。图 2.11 是不同的算子得到的角点图。

表 2.2 采用两种不同算子的系统运算耗时对比表

方法	式 2-2 算子	式 2-3 算子
时间(s)	126ms	941ms

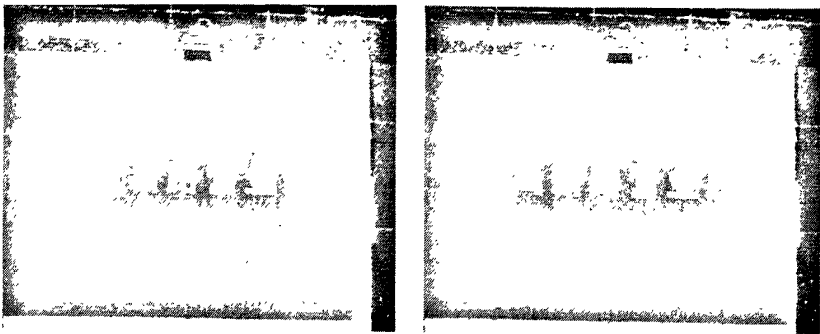


图 2.11 采用不同算子的执行结果对比

可见,非平滑算子在拥有较高的速度,比较适用于实时的对速度要求比较高的场合,而平滑算子检出的角点更为有效,可以去除一部分局部噪声引起的误检。

在角点提取过程中,两个阈值 t 和 g 是比较关键的,对提取性能有较大影响,其中 t 代表相似度阈值,即以 t 为界限,区分模板内其他点是与当前点属于同一类别(灰度值近似)的点还是不属于当前点, t 的大小决定角点的提取能力。阈值 g 表示以当前的为中心的模板内同类点的个数上限, g 越小,提取出来的角点越尖锐。由于提取的是文字角点,需要的角点都是位于直角甚至锐角尖端的角点,因此 g 阈值应该在全模板像素数的四分之一左右。采用不同的阈值 t 和 g ,其中 t 选用[10, 30]区间内以 4 为单位的 6 个值, g 选取从 9 到 6 的 4 个值,测试阈值改变对结果的影响。由于没有合适的角点真值作对比,角点的检出效果无法通过直接对比检出角点和需要的文字角点得出。本文检测角点是为了得到一个可以更好地区分文字与非文字区域的特征。因此,这里以检测到的角点落入文字区域的比例和文字区域内角点密度来确定角点检测方法的性能,从而确定阈值。角点密度是指文本区域内角点数量和区域面积之比乘以 100 得到的参数。测试结果如下表。2.3a 是

阈值与文字区角点所占比例关系表，2.3b 是阈值与文字区角点密度关系表。

表 2.3a 阈值与文字区角点所占比例关系表

阈值 g	阈值 t					
	10	14	18	22	26	30
5	---	---	40%	41%	43%	---
6	---	---	46%	47%	47%	---
7	----	46%	50%	54%	56%	53%
8	50%	54%	60%	62%	68%	65%
9	---	52%	57%	60%	61%	---
10	---	---	58%	57%	64%	---

表 2.3b 阈值与文字区角点密度关系表

阈值 g	阈值 t					
	10	14	18	22	26	30
5	---	---	4.3	4.1	2.8	---
6	---	---	5.7	4.7	3.3	---
7	----	8.9	8.1	7.2	4.1	1.8
8	11.3	9.6	9.3	7.7	4.6	2.4
9	---	11.7	9.9	8.1	7.4	---
10	---	---	11.3	8.5	7.9	---

由表中数据可以看出，选取阈值 t 为 22 和阈值 g 为 8，可以兼顾角点特征的区分性，又能使角点保持足够的密度，取得比较好的效果。

根据以上提取模板的尺寸，角点的距离均大于模板半径 4。聚类阈值 T 的选择关系到检测字符的大小，根据统计分析得到，文字区域的单个字符面积在 1000 到 3000 之间，10000 附近分布较广，此外还有面积在 30000 以上的较大字符。对于前两类字符，经过下采样处理后，其对应高或宽是 25 和 50。对于这两个尺寸的字符，取聚类距离从 8 到 16 测试。由于聚类后可以得到文字的外接矩形框，因此用召回率来度量聚类性能如下表。

表 2.4 阈值与性能关系表

阈值 T	8	12	14	16
性能	50.7%	51.8%	57.4%	58.6%

根据上表，12 选取 16 为阈值。但这个阈值对于大字符来说，太小了，因此后续的改进要考虑这个问题。

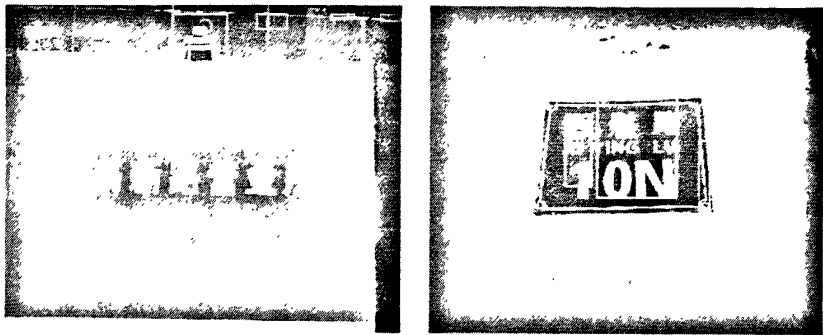


图 2.12 部分定位结果示例

2.4.4 测试结论

由上文测试可见，角点信息是可以用来定位文字的。但是，角点信息也有其局限性，如图 2.11 所示，角点的分布并不完全集中在文字区域。其次，角点计算对不同对比度水平的图象通用性有一定局限，同一阈值下，低对比度图象避免漏检角点和高对比度图象检出过多角点难以兼顾。对于角点的聚类，属于同一个文字的角点的欧氏距离分布和文字的大小有很大的关系，比较大文字上的角点过于稀疏而导致角点无法正确聚类，如图 2.13。这是一个需要注意的问题。

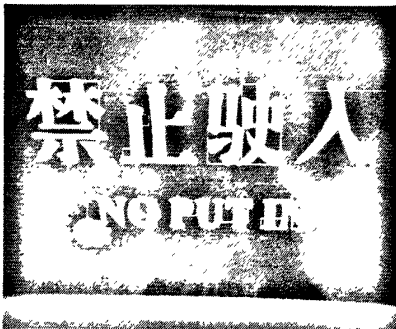


图 2.13 大字聚类分裂的样本

2.5 本章小结

总体来看，虽然角点信息可以用来定位文字，但是由于背景复杂多样，非文本区域也存在大量的角点，因此单纯的基于角点的文字定位，性能还是比较低的，无法满足文字定位的要求，因此有必要通过结合其他方法来提高总体性能。角点是一种描述空间和结构特征的一种特征信息，因此可以选用不同于角点这种类型的特征来定位文字，下一章的基于颜色的分析就是从这个角度出发，通过对颜色信息的应用来改进定位能力。

第三章 基于颜色聚类的自然场景文字定位

3.1 引言

颜色特征是在图象中应用最为广泛的视觉特征之一，主要原因在于颜色往往和图象中所包含的物体或场景十分相关。此外，与其他的视觉特征相比，颜色特征对图象本身的尺寸、方向、视角的依赖性较小，从而具有较高的鲁棒性。

由于自然场景中的文本通常具有相同的颜色或灰度，同时与背景具有较高的对比度，所以采用本文中采用颜色聚类的方式进行文本定位。利用同区域文字颜色近似，通过色彩空间的分割将图象分成若干个图层，提取连通区域并分析区域内的颜色分布以及区域外接框的几何特征来定位文字。将色彩空间分割到几类颜色，并计算每一类包含的连通分量，也就是实现颜色聚类的过程。

一般来说，基于颜色聚类的文字定位需要结局以下几个方面的问题。

- 1) 颜色空间的选择和此空间内颜色的表达方式；
- 2) 聚类核心数或分割分类数量的选择；
- 3) 颜色距离的度量形式；

3.2 彩色图象颜色分割和聚类技术

彩色图象的分割是基于颜色的相似性的大小来将颜色映射到若干个分类，从而将图象分割成若干部分。图象分割问题实际上就是基于颜色和空间特征的分类问题，同样可以分为有监督和无监督两大类。Power 等将有监督学习算法用于不同的 RGB、HIS 以及混合颜色空间等不同的颜色空间进行图象分割^[40]。有监督算法包括最大似然、决策树、K2 最近邻、神经网络等。Pal 等对彩色图象分割进行了简要介绍^[41]，并将彩色图象于多光谱图象作类比，认为彩色图象可以看成是多光谱图象的一个特例，因此适用于多光谱图象的分割方法都可以用于彩色图象分割。此外，将灰度看成是一维空间的色彩模式，将其推广到多维空间，那么，灰度图象分割技术也可以扩展到彩色图象。彩色图象分割方法可以是把灰度图象分割方法直接应用于颜色空间获得分割结果，也可以应用于每个颜色分量上，通过一定的方式对其结果进行组合获取最后的分割结果^[42]。以下是常见的几类文字分割聚类技术。

3.2.1 直方图阈值法

直方图阈值法是灰度图象上广泛使用的一种分割方法。对于灰度图象来说，

前景目标同区域内部的近邻像素或背景区域内部的近邻像素具有近似的灰度值,同时前景目标区域和背景区域上的像素灰度差异较大,这种特征反映在灰度直方图上,表现为不同目标或背景区域的灰度值分布在直方图上对应不同的波峰。那么在灰度分割时,可以将阈值选取在两个波峰之间的最低波谷最低点或是波峰中点附近的波谷。基于上述分析可见,直方图阈值法是可以用来区分图象的前景和背景的。郭国栋等通过计算图象数据在多维特征空间的概率密度函数,采用基于直方图熵的方法来确定多维颜色分布的波峰和波谷。Ohlander 等提出了一种多维直方图阈值化方案^[43],在 RGB、YIQ 和 HSI 空色彩间的 R、G、B、Y、I、Q、H、S 和 I 这 9 个分量上计算阈值用于图象分割。对于每个区域,先计算 R、G、B、Y、I、Q、H、S、I 这 9 个分量各自的强度分布直方图,根据每个直方图峰值可分离的程度选出具有最佳可分性的直方图,在这个直方图上选取一个阈值来在对应分量上进行颜色的分割映射,从而将图象分割成若干个新的子图。对于新分裂出的子图重复上述的计算过程直到每个区域的颜色满足一定的相似性,不能再进行分裂为止。

对于彩色图象,由于彩色图象采用的颜色模式一般来说有 3 个颜色分量,因此其直方图对应的是一个 3 维分布。许多算法没有考虑 3 个颜色分量的信息的联系,而是将 R、G、B 三个分量分开计算,在每个颜色分量的灰度图上进行直方图阈值化。由于忽略了 3 个颜色分量间的相关性,分割算法难以达到理想的效果。理想的分割一个是用若干个相交的三维曲面来将所有三维分量遍历组合的三维向量分为数类,但是这种方法计算量大,分类困难。投影是常见的简化高维空间计算的一种方法,这种解决方法也可以用到多维色彩空间直方图阈值的计算上。通过一定的投影变换降低颜色维数,即将 3 维颜色投影到二维空间,如果能找到一个准则,使得投影到降维后空间上的点能够很好地分开,就能同时满足简化运算和同时利用 3 个颜色分量信息的要求。Celenk 利用一条直线来分隔颜色空间,这条直线由 Fisher 线性判别来确定。在 CIE3 维颜色空间中,首先计算 CIE 空间上的 H、S、I 这 3 个分量的一维直方图,然后结合三个直方图来估计像素点在三维颜色空间分布的圆形柱面体单元的形状,并将每个柱面单元内的像素点投影到 Fisher 线性判别方法所确定的直线上,最后利用一维直方图阈值化得到的阈值来对柱面单元所确定的区域进行分割。Underwood 等把 3 维颜色空间(X, Y, I)投影到 3 个 2 维平面(X-Y, X-I, Y-I)^[45],对 3 个二维平面分别计算分隔线,结合 3 个平面的结果来分割三维颜色空间。

直方图阈值法的优点是不需要先验信息,且计算量较小,但它的缺点也是和明显的:

- 1) 基于颜色分割,不采用其他方法,不能保证得到的前景区域是完整的;
- 2) 在复杂图象中,一个或多个分量的直方图可能不存在明显的波峰和波谷,

这使得它无法满足直方图阈值化操作的前提条件;

- 3) 当颜色映射到 3 个直方图的不同位置时, 颜色信息会发散;
- 4) 局部的颜色分布是存在相关性的, 这类方法没有加以考虑。

3.2.2 特征空间聚类

是一种无监督的统计方法, 它是通过迭代地执行分类算法来提取各类的特征值, 其中模糊 C 均值(FCM)、K 均值等是最常用的分类方法^[46], 其优点是不需要训练样本。对于彩色图象, 颜色空间本身就可以作为聚类算法所需要特征空间, 相对于其他特征空间, 颜色空间比较直观。由于颜色空间一般是三维空间, 因此其计算也相对比较简单。基于颜色空间的特征空间聚类, 最大的问题是需先进行, 确定了可分性和聚类中心数量才能有效地进行分割聚类, 因此需要引入可分性度量指标。对于模糊 C 均值算法, 目前已有不少分类指标, 如基于隶属度矩阵的分割系数(partition coefficient, PC)和分割熵(partition entropy, PE), 还有考虑数据点紧密度-分离度(compactness-separation)的 Xie-Beni 指标和 Fukayama-sugno 指标, 以及将紧密度-分离度和隶属度矩阵特性进行综合的 SC 指标。

Lim 等把直方图阈值化和模糊 C 均值相结合用于彩色图象分割, 处理流程分为寻找聚类中心和模糊 C 均值聚类两个阶段。第一阶段是通过色彩空间进行分区, 判断小区域内是否存在聚类中心, 将存在聚类中心的小区域作为初始聚类中心。具体方法是先利用尺度空间滤波器来确定各小区域划分尺度, 然后把颜色空间按照该尺度分为若干个子空间, 当某个子空间包含的像素点数超过某一阈值时, 就被认为是一个有效的聚类, 然后利用该子空间直方图的峰值所在的位置确定聚类中心。接下来的阶段则是根据在寻找聚类中心阶段所确定的聚类中心, 利用模糊 C 均值算法来对第一阶段无效聚类中的像素点进行分类。这种方法首先确定聚类的有效性; 通过预先计算数目和中心点, 解决了模糊 C 均值聚类计算量大的问题。但是由于直方图有可能并不是只有一个明显峰值, 因此大多数直方图方法具有的缺点, 这种方法也同样存在。Carevic 等提出了基于加权的空间的相关信息和图象的二阶统计信息结合的特征量的聚类方法, 其不仅使得聚类在空间上是连续的, 还可以通过调整加权因子来提高聚类性能。

模糊 C 均值还有一个问题就是计算量较大, 为了提高模糊 C 均值的运算速度, 刘健庄使用金字塔分解来处理图象数据集, 利用均值将图象做 1/2 下采样, 即将每一层图象的任一像素点四邻域的颜色求平均值作为上层对应像素点的颜色值, 在经过多层金字塔分解处理之后, 处理塔顶的图象, 这种方法是牺牲了分辨率来换取较高的运算速度。由于图象分辨率的降低, 导致了图象有用信息的丢失。林开颜等利用分层减法聚类方法实现颜色聚类, 首先把图象数据分成一定数量的色彩

相近的子集,将子集颜色分布的中心作为初始的聚类中心,然后利用子集中心点和子集内的颜色分布密度来进行模糊聚类。其聚类数目可以利用聚类有效性分析指标来快速确定。由于该方法显著减少了聚类样本数量,且每个样本层的聚类计算也比较量小,因而可以大幅提高模糊 C 均值算法的计算速度。

传统的模糊 C 均值算法每次迭代都要对所有聚类中心同时进行优化,对此,陈铁岐等提出了一种新的模糊聚类算法,此算法的隶属度函数考虑了与其他类的关系,可以使聚类中心通过逐一迭代优化产生,有效地减少了计算量。Michael 等在分析向量量化和聚类关系的基础上,采用最小平方和准则,以竞争学习优化的权向量作为聚类中心,实现了不同颜色空间变换基础上的彩色图象分割,并讨论了不同颜色空间对于图象分割结果的影响。

总的来说,虽然聚类分析不需要训练集,不考虑空间信息,但需要事先确定分类个数,且初始参数对分类结果影响较大;另一方面,对噪声也比较敏感。

3.2.3 基于区域的方法

基于区域的方法是区域生长、区域分裂、区域合并以及这三类方法的有机结合。区域生长是基于区域像素的相似性来实现的,它具有相似性质的像素集合起来构成区域,例如灰度差小于某一阈值的像素点就可以通过生长来聚为一类;而区域分裂技术则是将种子区域割出已被标记的感兴趣的物体^[47]。这种标记选取可以说是分水岭算法的一个难点,选取到不恰当的标记会导致图象过分割。因此,使用这类方法,首先要做的是选取标记,数学形态学在这方面有着广泛的应用。现阶段已有不少标记提取方法,Shafarenko 等通过对原图象进行闭运算来对伪谷底进行填充,以获得更有价值的局部极值点作为标记。Shiji 等也是利用闭运算填充后的图象统计直方图来计算标记,得到初步标记的图象,然后在标记图象的梯度和原图特征值变化均比较缓慢的地方选取标记。Whatmough 等则先对经过开、闭运算的结果图象做二值标记预处理,并将其作为初始分割步骤,然后将所得的分割结果作为分水岭算法的种子^[48]。Lezoray 等将初步标记经过贝叶斯分类器分类,从结果中选取标记。

3.2.4 基于物理模型的方法

在现有的彩色图象分割方法中,区域的获取是基于颜色的相似性的,对于有强光反射区域和阴影区域的彩色图象,由于光照和阴影对颜色的改变,导致了目标表面颜色不一致,分割还是有困难的。利用 HSI、HSL 等模型将亮度分离出来,可以在一定程度上解决这个问题,但由于低饱和度时色调会不稳定,这个方法不能完全解决问题。因此,许多研究机构提出了基于物理模型的分割方法以解决这

个问题，该类方法是通过结合物体表面的物理模型来计算彩色信息。由于这类方法一般需要较多约束条件，难以建立合适的模型来拟合现实，同时模型参数也难以计算，因而目前应用范围有限，有待算法的进一步发展。

3.3 基于颜色聚类的文字定位

颜色聚类，直观的讲就是通过将颜色从某一色彩空间映射到规定或计算得到的若干类上。对于文字定位，通过颜色聚类，可以简化背景，获得更好的前景与背景的区分度，使得文字区域和非文字区域更加易于区分。整幅图片颜色的复杂程度很高，因此聚类颜色除了可以直接在整幅图片上进行之外，也可以在初步分割的局部图象上进行。如图 3.1a，对于整幅图片来说，由于文字颜色的多样性，直接颜色聚类，很难有效地利用颜色信息，而对于包含文字的局部图片来说，由于单个文字的颜色相对单一的，可以用来分割文字与非文字区域，如图 3.1b 所示。

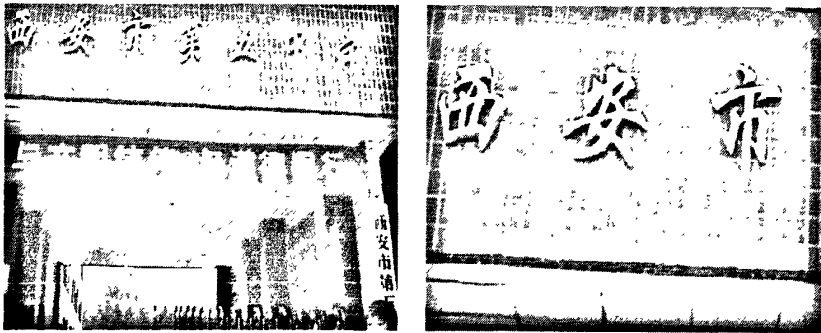


图 3.1 图片样本和包含文字的局部图象

基于颜色聚类的文字定位具体的基本流程如下图 3.2。颜色聚类的方法，可以是只适应的，也可以是基于先验知识的，本文中的颜色划分方法，本文采用了两种不同的算法，具体算法见 3.3.2、3.3.3 中颜色分割部分。

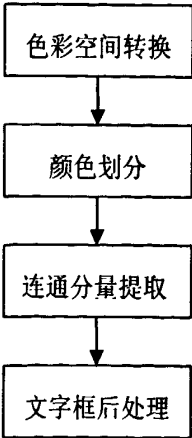


图 3.2 基于颜色聚类的文字定位

3.3.1 色彩空间的选择和转换

最近一百多年来,为满足各种不同用途的需求,人们已经开发了许多不同名称的颜色空间,尽管几乎所有的颜色空间都是从 RGB 颜色空间导出的,但随着科学和技术的进步,人们还在继续开发形形色色的颜色空间。为满足不同的应用需求,需要在各种不同的颜色空间之间进行转换。计算机绘图用的颜色空间包括 HSV(hue, saturation and value), HSL/HLS(hue, saturation and lightness), HSI(hue, saturation and intensity), HSB(hue, saturation and brightness)等。这些颜色空间都是从 RGB 变换来的、与设备相关的颜色空间,特点是指定颜色方式非常直观,很容易选择所需要的色调。同时可以把亮度从颜色信息中分离出来,这对于减弱光照对颜色分割的影响是很有意义的。

最常见的显示用颜色空间 RGB 和印刷用颜色空间 CMYK 分别是加法原色和减法原色模型,以原色组合的方式定义颜色。HSL 和 HSV 是两种常见的颜色空间,区别于前二者,HSV 和 HSL 则是以人类更熟悉的方式描述了颜色的信息,它们尝试定义比 RGB 更接近人的感知的颜色描述方法,并尽可能保持在计算上简单。其中 HSL 的 H, S, L 三个分量表示 hue (色相)、saturation (饱和度)、lightness (亮度)。色调是视觉系统对一个区域呈现的颜色的感觉,即对可见物体辐射或发射的光波波长的感觉。色调是最容易把颜色区分开的属性,可以用红、橙、黄、绿、青、蓝、紫等术语来刻画。饱和度用来表示颜色的纯洁性,当一种颜色掺入其他光成分越多时,就说该颜色越不饱和。

HSL 的优点是它对称于亮与暗,这意味着在 HSL 中,饱和度分量总是从完全饱和色变化到等价的灰色,而在 HSV 中,在极大值 V 的时候,饱和度从全饱和色变化到白色,这可以被认为是反直觉的;其次在 HSL 中,亮度跨越从黑色起通过选择的色相到白色的完整范围,而在 HSV 中, V 分量只分布在一半行程上,从黑到选择的色相。因此,对于自然场景中的文字分析,选取更接近人的视觉感知模式的 HSL 模型。

图 3.3 是 HSL 模型示意图。从下到上亮度增强,亮度的变化范围是 0 到 1 之间;水平切面的半径代表饱和度,饱和度的取值范围也是 0 到 1;颜色空间某一点所在水平切面上该点与坐标中心连线的极坐标值就是色相,不同的色相值代表不同的颜色类型,取值从 0 到 360。

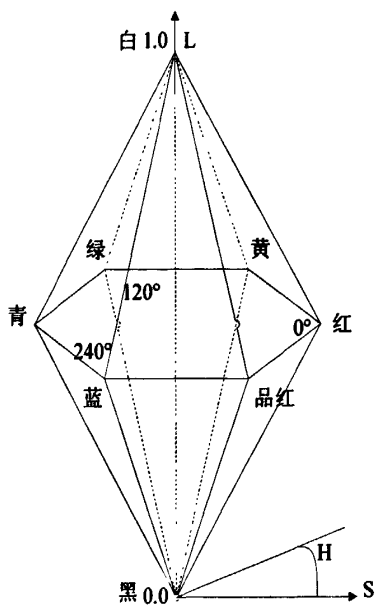


图 3.3 HSL 色彩模型示意图

从 RGB 到 HSL 的色彩空间转换依照以下 3 个公式。其中 RGB 的三分量使用 $[0, 1]$ 区间的取值方式。

$$h = \begin{cases} 0 & \text{if } \max = \min \\ 60 \times \frac{g-b}{\max-\min} + 0 & \text{if } \max = r \text{ and } g \geq b \\ 60 \times \frac{g-b}{\max-\min} + 360 & \text{if } \max = r \text{ and } g < b \\ 60 \times \frac{b-r}{\max-\min} + 120 & \text{if } \max = g \\ 60 \times \frac{r-g}{\max-\min} + 240 & \text{if } \max = b \end{cases} \quad (3-1)$$

$$l = \frac{1}{2}(\max + \min) \quad (3-2)$$

$$s = \begin{cases} 0 & \text{if } l = 0 \text{ or } \max = \min \\ \frac{\max - \min}{\max + \min} & \text{if } 0 < l \leq \frac{1}{2} \\ \frac{\max - \min}{2 - \max - \min} & \text{if } L > \frac{1}{2} \end{cases} \quad (3-3)$$

3.3.2 基于二维颜色统计的二值聚类

投影是常见的简化高维空间计算的一种方法，这种解决方法也可以用到多维色彩空间直方图阈值的计算上。对于 HSL 模式的彩色，三维到二维的降维，投影可以投影到 H-S、H-L 和 S-L 这三个平面上。由于光照对颜色的影响，因此只选择

H-S 平面上的投影, 即只使用 H-S 平面的二维颜色信息进行颜色的聚类分割处理。

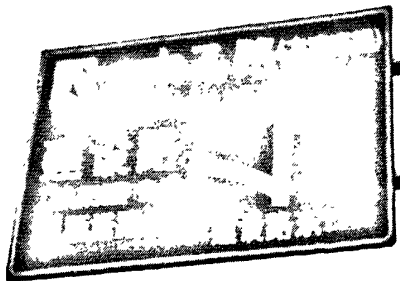
基于二维颜色统计的二值聚类是建立对在 HSL 模式下的 H 和 S 两个颜色分量的分布统计的基础之上的。类似于基于灰度直方图统计的灰度图象二值分割, 这个方法实际上是将这个方法推广到二维空间, 以两个二维峰值的等距离点构成的分割线将 H, S 分量构成的二维颜色面分割成两部分。计算二维平面上的两个聚类中心, 通过对所有样本点计算它在二维平面上与两个中心的欧式距离, 并将其归入距离更近的中心所属的一类中。

由于不能事先确定哪个聚类中心是属于文字区域的, 哪个聚类中心是属于背景区域的, 因此将所有像素点区分为三大类, 和其中一个中心距离小于某一阈值的一类标记为 1, 和另一个中心距离小于某一阈值的一类标记为 2, 第三类为剩余像素区域。通过对“2”类和“1”类的连通区域分析来找出文字区域以及他的外接矩形。最后通过先验知识来确定需要保留的区域和需要去掉的区域。

文字及背景有以下特点使其适合于这种方法。

- 1) 图象颜色较全图来说相对简单;
- 2) 文字颜色差异较小;
- 3) 文字和背景对比更加清楚。

由于上述的特点, 文字区域会在颜色二维分布上形成一个峰值区域; 背景区域若颜色较为单一, 则也会产生一个分布密度较大峰值区域; 背景颜色复杂, 那么其二维颜色分布将会比较分散, 平坦。因此, 对于局部的图象的 H 和 S 颜色分量的二维分布, 存在一个或两个峰值区域。以背景和文字区域二维颜色分布中心作为两个聚类中心, 利用欧氏距离可以将图象划分为两个区域。图 3.4 为文字块样本图, H、S 分量图以及对应的 H, S 分量直方图。第一行为原图, 第二行是 H、S 分量图, 其中 H、S 分量均线性映射到适合灰度图显示的空间, 左图是 H 分量图, 右图是 S 分量图, 第三行为第二行两图对应的统计直方图。



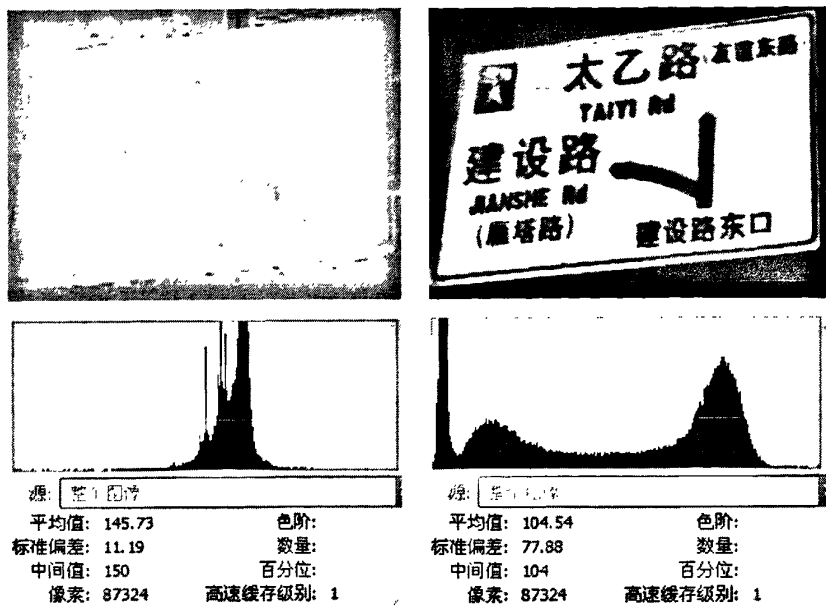


图 3.4 图象 H、S 分量图以及对应统计直方图

对应上述理论，使用算法流程如下：

- 1) 将图象转化到 HSL 空间；
- 2) 计算 H 分量各个灰度级的分布，寻找分布最大值对应的 H 值 H_0 ，设定阈值 D ，查找 S 分量值落在 $[H_0-D, H_0+D]$ 范围内的点对应的 S 分量的分布最大值对应的 S 值 S_0 ，将 (H_0, S_0) 作为第一个聚类中心；
- 3) 在 $[H_0-D, H_0+D]$ 区间之外寻找 H 分量分布最大值对应的 H 值，记为 H_1 ，查找 S 分量值落在 $[H_1-D, H_1+D]$ 范围内的点对应的 S 分量的分布最大值对应的 S 值 S_1 ，将 (H_1, S_1) 作为第二个聚类中心；
- 4) 分别计算候选区域每个点的 (H, S) 分量对与上面的两个聚类核心 (H_0, S_0) ， (H_1, S_1) 的欧氏距离，以阈值 T 为分界，分别构建出两幅二值图象；
- 5) 在两幅二值图上分别寻找前景点构成的连通分量；
- 6) 根据连通分量得到初步的文字框。

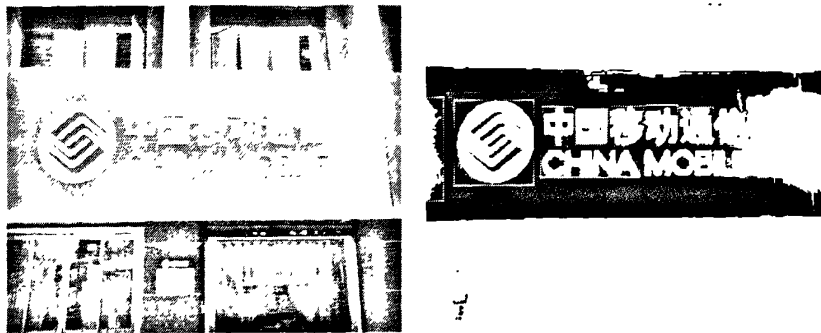


图 3.5 二值颜色聚类结果示意

3.3.3 基于 HSL 色彩空间分割的分层聚类

以上的方法针对颜色简单的图象，如果背景比例较高且存在多峰值，那么可以采用多层色彩分割寻找连通分量的方法 3 来定位文字。图 3.6a 是 HSL 色彩空间的 H-S 分量极坐标分布，3.6b 是 H-S 直角坐标分布。左图极半径为 S，角度为 H，右图横坐标位 H，纵坐标为 S。由图可以看出，常见的颜色可以通过固定的阈值分割成比较少的几种颜色，将颜色归类后可以通过分析属于每一类色彩区间的像素点的分布区域来定位文字。

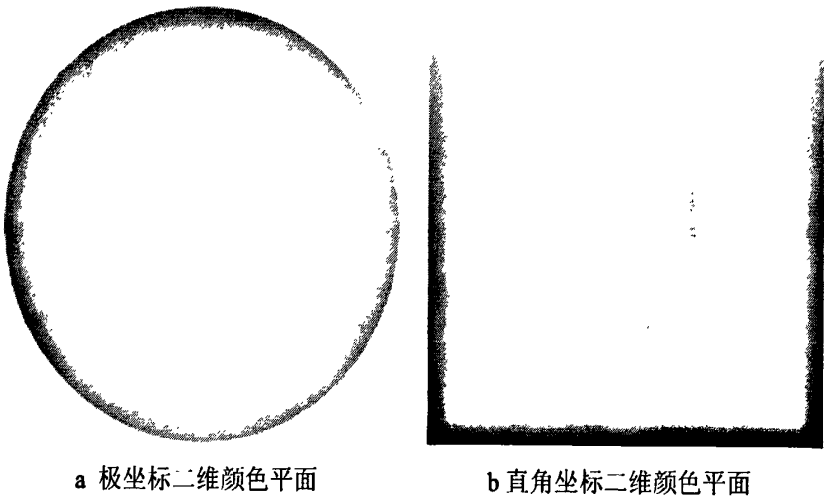


图 3.6 HSL 色彩空间 H-S 分量分布图

根据颜色转换公式 3-1、3-2 和 3-3，当 L 很小和很大时，H 和 S 分量对颜色视觉感受的影响较小，因此这颜色应提前予以区分。其中 H、L 分量取值均在[0, 1]区间，而 S 分量取值在 0 到 360 之间，对应上左图的圆的一周角度。

根据常见的颜色类型以及图 3.4，本文将 HSL 色彩空间的颜色按照如下规则划分为 8 类。

- 1) 将 L 分量小于 0.1 的区域划分到黑色区域，L 分量大于 0.9 的区域划分到白色区域，其余待选；
- 2) 上一步留下来的区域中 S 分量小于 0.1 的加入黑色区域，S 分量大于 0.9 的加入白色区域，其余颜色待选；
- 3) 按下表 3.2 将剩余颜色部分按 H 分量分开；

表 3.1 H 分量颜色划分表

H 分量范围	对应颜色
330-360, 0-30	红
30-90	黄
90-150	绿
150-210	青

210-270	蓝
270-330	品红

由此，颜色被归类成黑、白、红、黄、绿、青、蓝和品红 8 大类。将每一类颜色所包括的像素点构成一幅子图，可以得到 8 幅二值图象。用数学形态学处理子图，通过在图像上进行连通分量分析，得到各个子图上带有颜色信息的连通区域，可以作为初步的文字区域。

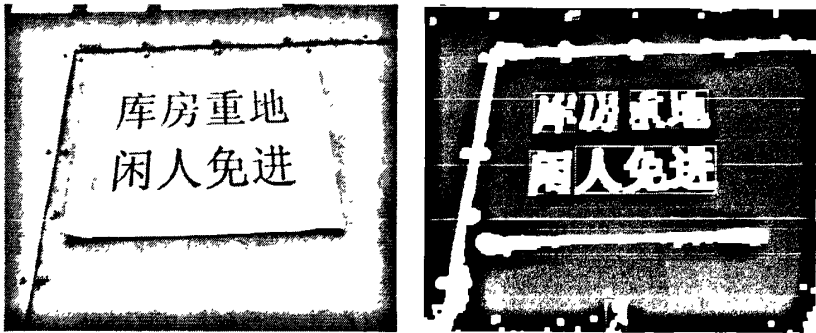


图 3.7 左图的黑色子图上处理结果示意

3.3.4 基于先验知识的聚类结果优化

由于以上两个方法得到的区域都是比较粗略的，在复杂背景的干扰下，存在很多错检，因此在整体的算法流程中必须利用先验知识来筛选文字区域。

以经过上述方法流程得到外接框作为定位区域，根据数据库标定的真值做统计得到的先验知识，默认宽度和高度中较大的为宽度，如果连通域宽度和高度之比大于 10，则认为不是文字区。同时宽度和高度有一个小于 4 个像素宽的区域，也加以排除。如果连通域内白点像素和黑点像素的比率等满足公式 3-4 所示条件，则认为该连通域为备选文字区域，否则丢弃。备选文字区域通过以下条件进行筛选。通过统计数据库部分样本中文字区域二值化后的前景和背景像素数， $a=0.3$ ， $b=0.7$ 。但是由于自动处理得到的备选区域往往不像真值那样紧贴文字，存在更多背景成分，因此适当放宽限制条件，取 $a=0.2$ ， $b=0.8$ 。

$$\frac{Sum(p > T)}{Sum(p)} \in (a,b) \tag{3-4}$$

连通域提取结果分析分析。通过文字区域连通域提取，得到代表了文字区域的字符位置的外接矩形框。同一文字行或列的字符应该在字符高度、宽度、字符间距等方面存在一定的相关性，所以这里可以用矩形框的分布来判断备选区域内是否包含文字。文字区域向矩形框较长的一条边的前景点投影密度曲线有明显的波峰和波谷，波峰对应字符的笔画，波谷对应字符间隙，而非文字区域相应的投

影曲线相对平滑，没有明显的波峰或波峰的数量较少。文字框的初步筛选主要是利用文字框内部黑白两色的比例来区分文字和非文字，一般来说文字区域的黑白两色所占的比例是比较均衡的，白色（代表文字）比例过高的可能是一个封闭的色块，而过低则有可能是由类似于窗框等的边框连通出的空心区域。因此要通过这个方法来去掉一部分。

3.4 测试和分析

对两个方法采用同样的样本进行测试，测试样本分全图组和部分区域组。全图组就是采用全幅图象来直接颜色分割区域，而部分组使用包含文字的局部图片来测试，两种样本示例如图 3.8，第一排是全图样本，第二排是局部图样本。局部图象是在对应自然场景图象中切割出来的文字所占面积比例较高的部分图象，相对整个图象，颜色的文字一般是同种颜色，背景的颜色种类比较少，颜色分布也较为简单。使用局部图象的目的是为了说明颜色聚类适用的场合。测试性能的评价方法和第二章最终测试方法相同，也是利用 ICDAR2005 的方法，以召回率为标准测试性能。



图 3.8 测试样本示例

测试结果如下表。

表 3.2 测试结果对比表

	全图组	部分组
投影聚类	3.2%	80%
平面划分	12%	82%

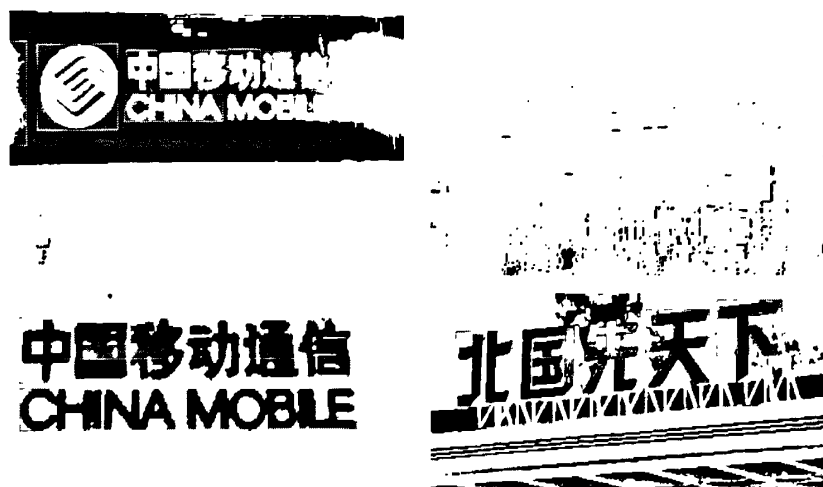


图 3.9 图 3.8 图象对应的颜色分割区域聚类结果

图 3.9 是对应图 3.8 的处理后的结果, 由图可以看出, 直接利用颜色信息在全图上做分割是比较困难的, 背景和文字的区别性并不是很好, 左图由于存在大块的单色区域, 在投影聚类时还可以将文字区域分开, 而对于右图这种背景非常复杂的图形甚至完全不能使用。相对的, 局部图象由于文字颜色的单色分量分布较高, 且背景颜色相对集中, 颜色区域的分割也更加有效, 定位处理结果提高了很多。

3.5 本章小结

综上所述, 颜色信息也是可以用来提取文字的。相对于角点信息, 颜色信息更适合局部的颜色较为简单的图象的处理, 因此, 可以将全局的角点特征和局部的颜色特征结合起来, 以获得更加良好的定位结果。根据二、三章的测试, 第四章结合角点与颜色信息, 并用笔划结构和周期纹理检测做辅助来定位文字。

第四章 角点特征和颜色特征结合的文字定位

4.1 引言

角点特征主要描述的是纹理和结构的特征，这些都是图象空间结构上的特点。而颜色所包含的信息和角点特征所包含的信息是不同的，基于颜色分割的区域分割也是常见的有效的文字定位算法。根据上两章的测试分析，在角点检测的基础上加入色彩分割来进一步筛选定位结果是有一定价值的。

- 1) 文字上存在丰富的角点，但是角点信息同样广泛存在于许多非文字区域。角点区域即图象上线段交叉的区域，这种结构在自然环境中普遍存在，例如交叉的树枝，窗格上的交叉点，建筑物外墙纹理的交叉点等。因此，需要用不同于角点这种类型的特征对基于角点的定位作补充。
- 2) 图象的颜色包含丰富的信息，但全局的色彩分割效果并不是很好。通过角点聚类得到的文字区域排除了大部分背景，内部包含的颜色成分比整幅图像要简单，更易于利用色彩聚类来进一步处理。

4.2 两者结合的处理流程

结合角点聚类和颜色聚类的文字定位，是在一种角点聚类和颜色聚类前后结合的处理方法，其流程如下图 4.1。对于这两者结合的方法不能很好区分的树丛，周期性的纹理等类型的误检结果。分别采用笔划信息的检测和傅里叶谱周期性的检测来改善这些问题。最后根据文字框几何尺寸等先验信息筛选，合并文字框，从而得到最终结果。

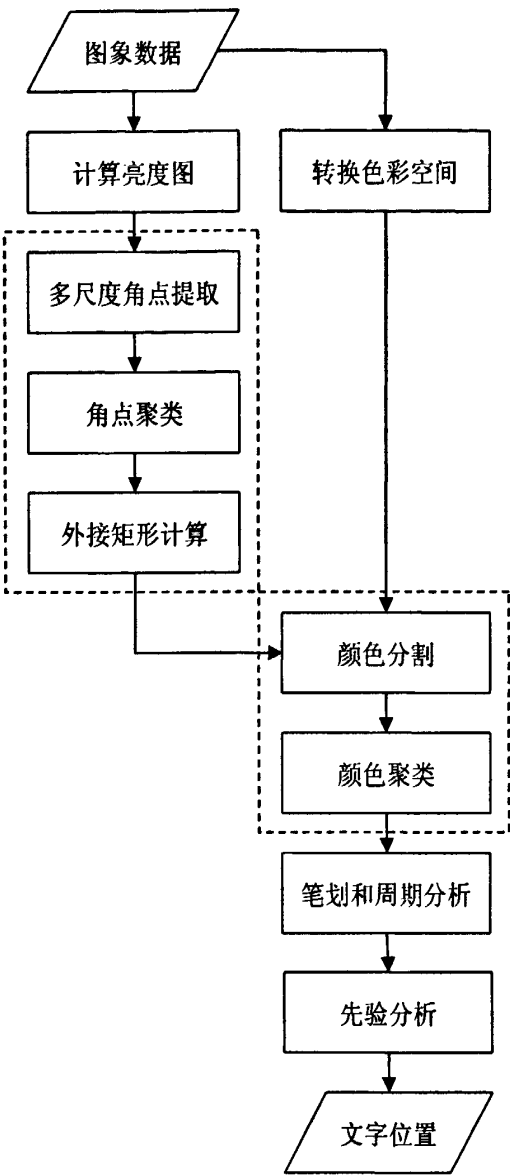


图 4.1 结合角点聚类和颜色聚类的文字定位流程

4.2.1 基于角点特征的粗定位

由第二章的测试可以看出，角点方法对不同尺度的文字敏感程度不同，聚类阈值只能针对处于特定大小范围内的文字，因此可以用金字塔分解的方法来解决，但是由于金字塔方法会增大处理的数据量，因此只做一层下采样子图。在较低分辨率的图像上定位包含大尺寸字符的文字区域，提高文字定位方法对字符尺寸变化的适应能力。

将原始图像 G 作为金字塔第 0 层，对原分辨率子图像 G 进行双线性插值采样得到第 1 层子图像 G_1 。对图像进行金字塔分解用双线性插值的方法，该方法缩放

后的图像质量高，不会出现像素值不连续的情况。具体方法为：首先采用向后映射法，即通过输出图像 G_1 像素位置 (i, j) 来计算其在 G 图上对应的像素位置 $(u + \alpha, v + \beta)$ ，其中 i, j, u, v 为非负整数， α, β 为浮点数，且 $0 \leq \alpha \leq 1, 0 \leq \beta \leq 1$ 。然后根据输入图像 $(u + \alpha, v + \beta)$ 处个相邻像素 $G(u, v)$ 、 $G(u+1, v)$ 、 $G(u, v+1)$ 和 $G(u+1, v+1)$ 的灰度值计算该位置的灰度值作为输出图像 G_1 在位置 (i, j) 处的灰度值。计算公式为 4-1.

$$G_1(i, j) = G(u + \alpha, v + \beta) = (1 - \alpha)(1 - \beta)G(u, v) + \alpha(1 - \beta)G(u + 1, v) + (1 - \alpha)\beta G(u, v + 1) + \alpha\beta G(u + 1, v + 1) \tag{4-1}$$

采用第二章的 SUSAN 角点提取方法和角点聚类来进行定位，根据第二章的测试结果，选取阈值 $t=22$ ，阈值 $g=6$ ，用如下图流程。

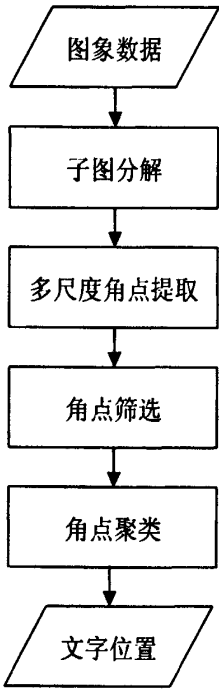


图 4.2 基于角点聚类的粗定位

4.2.2 基于颜色信息的筛选

采用第三章的二维投影后的二值聚类提取方法和角点聚类来进行定位，根据第三章的测试结果，用如下图流程。

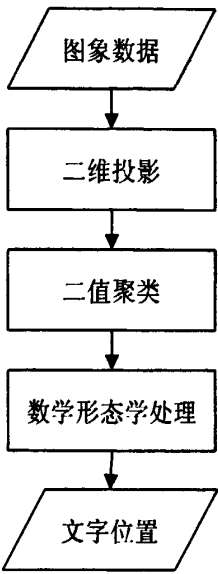


图 4.3 基于颜色聚类的文字细定位

4.2.3 基于笔划和周期信息的特定区域排除

文字的笔划一般具有下述特性，文字笔划可以视为小于某一宽度上限的双边结构而且文字笔划的延伸方向以 0 、 $\pi/4$ 、 $\pi/2$ 和 $3\pi/4$ 这四个方向为主^[49]。而树丛等角点丰富，色彩单比较集中区域，在利用以上使用的方法时并不能很好的区分。因此，基于文字笔画的上述特性，本文将它应用到自然场景文字定位中基于笔划的区域筛选步骤中。文字笔画提取过程是在灰度图上进行，考虑到二值化时并不能确定白色还是黑色是文字区域，因此在白底黑字的反向文字和对应的黑底白字的正向文字图象上分别采用以下处理。下面以正向文字为例说明笔画提取的主要步骤。

1) 在正向文字中，方向笔画强度的计算公式为：

$$DE_d^+(p) = \max_{i=1}^{W-1} \{ \min \{ f_d(p-i), f_d(p+W-i) \} \} - f(p) \tag{4-2}$$

$$DE_d(p) = \begin{cases} DE_d^+(p) & \text{if } DE_d^+(p) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4-3}$$

式中， $d=0, 1, 2, 3$ 分别代表笔画最常见的 0 、 $\pi/4$ 、 $\pi/2$ 和 $3\pi/4$ 这四个方向， W 为笔画宽度上限； f_d 函数表示 d 方向上与点 p 距离为 i 的点像素灰度值。笔画特征图定义为四个方向笔画特征的最大值，用公式表示为：

$$DE_w(p) = \max_{d=0}^3 \{ DE_d(p) \} \tag{4-4}$$

2) 对笔画特征图的二值化，采用 Otsu 方法对笔画特征图二值化。定义窗口的笔画密度公式为：

$$Den = \frac{m}{w \times h} \quad (4-5)$$

式中, Den 为笔画密度, m 为窗口中笔画点的个数, w 为窗口宽度, h 为窗口高度。显然, 文字区域一般都具有很高的笔画密度, 由此使用笔画密度分析的方法来滤除大部分非文字区域。

对于存在角点且纹理周期性很强的区域, 可以用测量傅里叶谱来排除。周期性纹理和非周期区域纹理在傅里叶变换后的图象基于极坐标的投影分布上有明显区别。将傅里叶变换结果的极坐标表示记做 $S(\theta, r)$, 对于每一个极角 θ , 存在一个随极半径 r 变化的一维函数 $S_\theta(r)$; 同样对于每一个半径 r , 也存在一个随极角 θ 变化的一维函数 $S_r(\theta)$ 。沿着等极角和等极半径方向分别计算以上两个函数, 得到新的度量函数 $S(\theta)$ 和 $S(r)$, 定义如下:

$$S(r) = \sum_{\theta=0}^{\pi} S_\theta(r) \quad (4-6)$$

$$S(\theta) = \sum_{r=0}^R S_r(\theta) \quad (4-7)$$

对于度量函数 $S(\theta)$ 和 $S(r)$, 周期图象会有如下图类型的分布类型。即左图 $S(r)$ 对应图, 右图为 $S(\theta)$ 对应图。 $S(r)$ 以 0 为峰值随半径增大而逐渐减小, $S(\theta)$ 在 0 和 π 附近存在较大峰值, 其他角度值均比较小。可以依据以上描述来筛选待选区域, 去除存在角点且纹理周期性很强的区域。

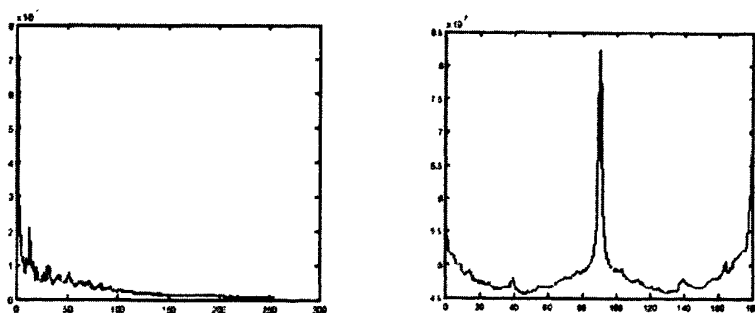


图 4.4 傅里叶变换谱分析示意图

4.2.4 基于几何特征和文字行分析的细定位

以上流程得到的备选文字区域还可以通过以下三个条件进行筛选。

- 1) 颜色空间可分。字符和背景具有一定的区分度, 即文字区域内字符和背景在颜色空间上可分。由于被选文字区域是经过上两个方法初步筛选出来的, 因此属于字符的像素点应该在整個区域中占一定的比例。通过统计备选文字区域的灰度直方图, 用 OSTU 二值化方法^[44], 求出自适应二值化

阈值 T 。如果阈值 T 不满足式 3-4 的条件, 那么认为该区域不包含字符, 是非文字区域, 将该区域从备选区域中删除。式 3-4 中 $\text{Sum}(p>T)$ 表示备选区域内大于阈值 T 的象素个数, $\text{Sum}(p)$ 表示备选区域内总的象素个数。通过统计数据库部分样本中文字区域二值化后的前景和背景像素数, 取 $a=0.3$, $b=0.7$ 。但是由于自动处理得到的备选区域往往不像真值那样紧贴文字, 因此适当放宽限制条件, 取 $a=0.2$, $b=0.8$ 。

- 2) 连通域分布。通过文字区域连通域分析, 得到代表了文字区域的字符位置的外接矩形框。同一文字行或列的字符应该在字符高度、宽度、字符间距等方面存在一定的相关性, 所以这里可以用矩形框的分布来判断备选区域内是否包含文字。因为文字区域的矩形框通常刚好包围一个或多个字符, 首先对矩形框宽度 width 、高度 height 、分层处理后每层图上文字区域内前景象素和背景象素比率 r 等条件所示限制, 如果矩形框不满足条件, 则认为该矩形框内不包含字符。限制条件如式 4-8、4-9 和 4-10 所示, 其中此处以高和宽数值较大的作为宽度, Height 为待选框所属的第一次角点聚类得到大框的高度。

$$r \in (0.2, 0.8) \quad (4-8)$$

$$\text{width} \% \text{height} < 0.3 \times \text{height} \quad \text{or} \quad \text{width} \% \text{height} > 0.7 \times \text{height} \quad (4-9)$$

$$\text{height} > 0.5 \times \text{Height} \quad (4-10)$$

- 3) 投影分析^[5]。文字区域向矩形框较长的一条边的前景点投影密度曲线有明显的波峰和波谷, 波峰对应字符的笔画, 波谷对应字符间隙, 而非文字区域相应的投影曲线相对平滑, 没有明显的波峰或波峰的数量较少。这里采用投影曲线的波峰数量 Num 和曲线方差 Var 作为判断标准, 分别对它们做条件限制。因为在试验中发现: 曲线方差大小往往与投影区域高度的平方有正比关系, 所以这里对曲线方差做了相应的调整, 如公式 4-11、4-12 所示。文中如果 Var 小于 0.05 或 Num 小于 5, 则认为该区域不包含文字。这里, $\text{Sum}(i)$ 表示投影曲线在点 i 的值; Width , Height , Left , Right 分别为备选文字区域的宽度、高度、左侧和右侧横坐标。

$$\text{Var} = \frac{1}{\text{Width} \times \text{Height}^2} \sum_{i=\text{Left}}^{i<\text{Right}} (\text{Sum}(i) - \text{mean})^2 \quad (4-11)$$

$$\text{mean} = \frac{1}{\text{Width}} \sum_{i=\text{Left}}^{i<\text{Right}} \text{Sum}(i) \quad (4-12)$$

由于不同子图得到的文字矩形区域经常相互交错、覆盖, 需要采用区域合并的方法, 合并重叠的文字区域。合并条件如 4-13 所示, 如果两个相交文字区域的尺度和比例满足条件, 则这两个区域合并为一个区域。公式中的各参量值, 本文

取 $p=0.8$, $q=0.2$, $s=10$, 其中 $A(R)$ 函数表示矩形区域 R 的面积。

$$\begin{cases} \frac{A(R_1 \cap R_2)}{\min\{A(R_1), A(R_2)\}} > p \\ \frac{A(R_1 \cap R_2)}{\max\{A(R_1), A(R_2)\}} > q \\ \frac{\max\{A(R_1), A(R_2)\}}{\min\{A(R_1), A(R_2)\}} < s \end{cases}$$

(4-13)

4.3 算法实现和评测

4.3.1 算法的实现和测试环境

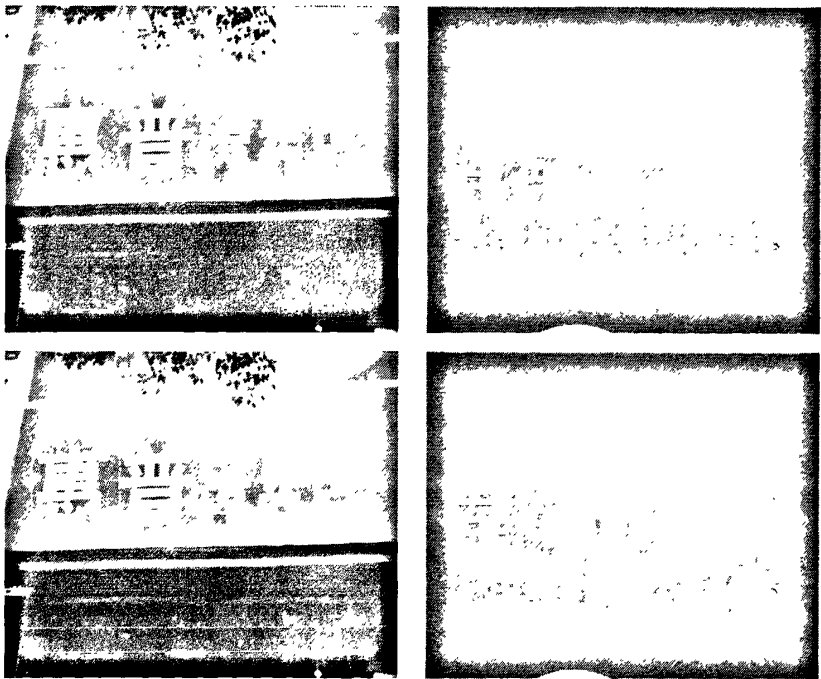
利用 Windows 平台下的 Microsoft Visual Studio 2008 软件, 基于 C++语言设计实现以上算法流程。

在 5000 幅图象的数据库上进行测试。选取前 3200 幅, 以 400 幅为一组统计定位性能, 结果如下表:

表 4.1 测试结果

算法性能指标	准确率	召回率	总体性能
数据	56.5%	62.3%	59.3%

以下是图 4.4 部分测试结果样本和真值的对比。



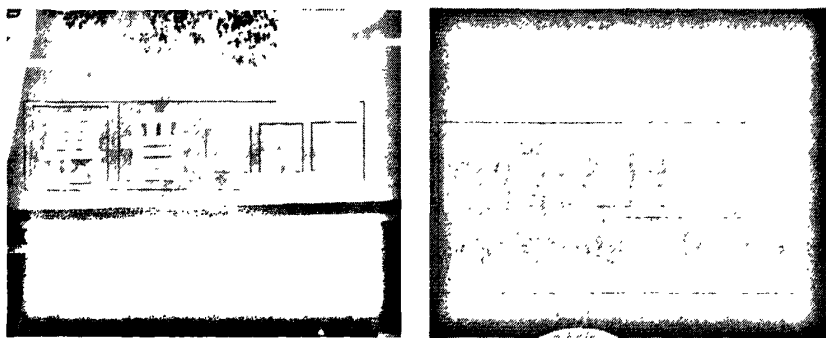


图 4.4 部分测试结果图

4.3.2 结果分析

可见，本文采用的利用角点和颜色信息来定位文字是可行的，但是仍有一定的局限性，对于含有丰富角点的非文字区域等还是不能做到准确区分。引入笔划密度的检测可以较好的去除树丛这种类型的误检；引入周期纹理检测来针对性的处理窗格，建筑外墙瓷砖网格等周期性较强的部分，由于没有找到合适的波形对比方法，对检测效果的帮助不是很大。

造成错误的主要因素是文字检测方法在对文字区域进行分类时存在的漏检和误检，使得文字定位性能下降。通过分析产生错误的样本图像特征，漏检和误检产生的原因主要有以下几点。

- 1) 低对比度文字产生的漏检。这种情况下，第一步的角点提取算法无法正确提取角点，如图 4.5。降低角点区分阈值 t 可以提高低对比度图象的角点检出能力，但正常图象会因此检出过多角点，可见，需要通过图像预处理来改变这类图象的对比度，使他能够适应针对大多数正常对比度图象的角点区分阈值。

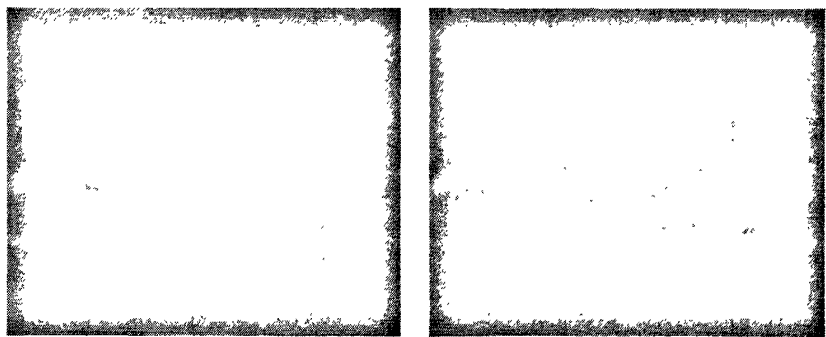


图 4.5 低对比度无法提取角点的图象

- 2) 含有类似文字结构的区域产生的误检测，降低了准确率。对于含有类似文字结构的区域，虽然本文加入了一定的判断，但由于背景的复杂多变，这些有针对性的检测不能涵盖所有的误检类型，而且还会引入少量漏检。其结果依然有不足的地方，还需要有针对性地加以改进。

4.4 本章小结

基于角点聚类 and 颜色聚类结合的文字定位方法是可行的, 该方法对自然环境的文字定位检测具有较强的适应能力, 可以处理多种类型图像中的文字。但是由于自然场景中文字及其背景的复杂性, 本文方法仍然存在误检和漏检, 其结果依然不够理想, 还需要有针对性地加以改进。

第五章 总结和展望

5.1 总结

自然场景中蕴含着丰富的文字信息这些文字对场景的认知具有重要价值,是描述和理解场景的关键线索,在人们的日常生活中起着重要的作用。本文针对自然场景中文字信息提取技术进行了一系列深入研究,重点研究自然场景中文字检测、文字定位的方法。采用了基于角点提取的自然场景中文字定位方法和基于颜色分割聚类的自然场景中文字定位方法,提出了结合两者的自然场景中文字定位方法,并结合笔划、纹理周期等特征进一步优化算法性能。

本论文主要从以下几个方面进行了深入地研究与探讨:

- 1) 针对图象上文字区域存在的丰富的角点,将角点特征的检测引入文字定位,对角点的提取算法进行了研究。采用了经过角点分类处理改进了的 SUSAN 角点提取方法,并通过聚类 and 连通分析来定位文字。实验结果表明,基于角点的文字定位对复杂背景中的文字检测有一定的效果。
- 2) 针对自然场景中文本在颜色方面的分布特点,将颜色信息应用到自然场景中的文字检测,采用自适应颜色空间分割和分层聚类的方法,实现了基于颜色聚类的自然场景文字定位。
- 3) 结合角点信息和颜色信息,提出了基于角点提取和颜色聚类的文字定位方法。通过对文字检测结果的分析,针对两类常见的误检,结合笔划的提取和待选区域周期性的判断,利用文字区域的特殊属性筛选和平滑文字候选区域,定位效果良好,能够得到规则的文字区域和比较精确的位置信息。

5.2 展望

本论文的研究只是对自然场景中文字定位的一个初步探索,仍然存在很多不完善的地方和没有解决的问题有待进一步深入研究。文字定位问题实际上可以分为两个部分,即特征提取和分类。

首先、文字信息提取的本质是图像分割问题,需要有能够区分文字区域与非文字区域的有效特征才能正确提取文字区域。本论文和实验室以前的工作^[1-4]虽然在这方面进行了研究,但是所选择的特征仍然是比较传统的边缘、角点、颜色等简单分类信息,这些信息均存在可分性不足的问题,需要进一步对文字属性特征的选择以及特征选择的评价方法作深入研究。

其次、本文文字区域中的文字特征的分类还是基于低维统计阈值的简单分类,尤其是当图像色彩十分复杂时,这种分类方式往往难以很好的区分文字与非文字区域,存在很大的局限性。低维不可分的或可分性较差的特征往往在高维空间更容易区分,利用多个超平面可以获得良好的分类结果。因此,引入机器学习是很有必要性的。

再次、现阶段大多数的文字定位是独立于是别的。既然定位是为了更好的识别,那么将文字定位与识别结合起来,可能会取得更好的效果,引入笔划信息也是这方面的一个尝试。

文字定位技术目前还有很大的不足,其性能还达不到商业化标准。现阶段的很多算法都遇到了瓶颈,期待进一步的突破。随着人类对自身认知过程的深入了解和人工智能的进一步发展,可以期待,未来文字识别将在这些方面取得重大突破。

致谢

在论文即将完成之际，首先向我的导师卢朝阳教授致以最诚挚的谢意！

卢老师有着渊博的学识、敏锐的洞察力、严谨的治学态度、突出的学术成就。在做课题的过程中给了我许多有益的建议以及方向性的指导。他平易近人，关心学生，耐心地解答我的问题，对我的一些错误及时指正并且宽容地对待。同时卢老师在生活上和思想上也对我非常关心，在我课题进展不顺利时给我鼓励，让我重拾信心。卢老师认真的工作态度，一丝不苟的科研精神是我学习的榜样，他对我的言传身教使我受用一生，在此深深的感谢卢老师对我各方面的关心与帮助！

同时感谢实验室李静老师！李老师平易近人，爱护学生，在工作、学习上给了我很多的建议，在平时算法研究以及程序设计遇到的问题给了我许多帮助，所有这些都让我受益匪浅！

感谢我的师姐刘晓佩博士，她的带领为我的工作打下了良好的基础；在课题研究上也给了我很多帮助，使我能够尽快的掌握数字图像处理，尤其是文字提取方面的知识和研究方法。感谢课题组中的我的同年级同学汪文芳，谭福秀，师妹葛巧瑞，与他们的讨论拓宽了我的研究思路，他们在实验上和编程过程中对我的帮助也使我收获颇丰。感谢与我同一实验室的同学柴进，郁映卓，李乔，孙华凯，詹小孟，王小明，张宇等同学对我生活上和学习上的帮助！感谢实验室其他同学在各个方面对我研究思路的开阔和启迪。

深深地感谢我的父母！他们在我成长过程中付出了大量的心血和辛勤的操劳。在我获得成绩的时候给我善意的提醒，在我灰心的时候又给我理解与鼓励，他们深深的期待和多年来在生活上无微不至的照顾都化为我前进的动力。

最后，再次向所有支持和帮助过我的老师，同学，朋友和家人们表示衷心的感谢！

参考文献

- [1] Lucas S M. ICDAR 2005 text locating competition results [C]. International Conference on Document Analysis and Recognition, 2005.
- [2] Kim H K. Efficient automatic text location method and content-based indexing and structuring of video database [J]. Journal of Visual Communication and Image Representation, 1996, 7(4): 336-344.
- [3] Lienhart R, Stuber F. Automatic text recognition in digital videos: proc. of SPIE [C], 1996: 180-188.
- [4] Jain A K, YU B. Automatic text location in images and video frames [J]. Pattern Recognition, 1998, 31(12): 2055-2076.
- [5] Smith M A, Kanade T. Video skimming for quick browsing based on audio and image characterization, Technical Report CMU-CS-95-186 [R].
- [6] Yassin M Y H, Karam L J. Morphological text extraction from images [J]. IEEE Transactions on Image Processing, 2000, 9(11): 1978-1983.
- [7] Chen Datong, Shearer K, Boulard H. Text enhancement with asymmetric filter for video OCR: Proc. of International Conference on Image Analysis and Processing [C]. [S. l.]: [s. n.], 2001: 192-197.
- [8] 张引, 潘云鹤. 复杂背景下文字提取的彩色边缘检测算子设计[J]. 软件学报, 2001, 12(08): 1229-1235.
- [9] Sing-Tze Bow. Pattern recognition and image preprocessing [M]. M Dekker, New York, 1992.
- [10] 欧文武, 朱军民, 刘昌平. 自然场景文字定位[J]. 中文信息学报, 2004, 18 (5): 42 -63.
- [11] 许剑峰, 黎绍发. 基于颜色边缘与 SVM 的图像文字定位[J]. 计算机应用研究, 2006, 3: 154-157.
- [12] 李昭早. 自然场景文本区域定位. 西安电子科技大学硕士学位论文. 2006, 1.
- [13] 孟天. 自然场景文本区域定位. 西安电子科技大学硕士学位论文. 2007, 1.
- [14] 严由辉. 基于手机设备的自然场景文本分析. 西安电子科技大学硕士学位论文. 2009, 1.
- [15] 曹璿. 自然场景中汉字识别关键技术研究. 西安电子科技大学硕士学位论文. 2010, 1.
- [16] Rosenfeld A, Weszka J S. An Improved Method of Angel Detection on Digital

- Curves [J]. IEEE Trans. Computers, 1975, C-24(9): 940-941.
- [17] Freeman H, Davis L S. A Corner Finding Algorithm for Chain Coded Curves [J]. IEEE Trans. Computers, 1977, 26(3): 297-303.
- [18] Asada H, Brady M. The Curvature Primal Sketch [J]. IEEE Trans. PAMI, 1986, 8(1): 2-14.
- [19] Witkin A P. Scale Space Filtering [C]. Karlsruhe: Proceedings of the 8th International Joint Conference on Artificial Intelligence, 1983: 1019-1021.
- [20] 赵文彬, 张艳宁. 角点检测技术综述[J]. 计算机应用与研究, 2006, 23(10): 17-19.
- [21] Kitchen L, Rosenfeld A. Gray Level Corner Detection [J]. Pattern Recognition Letters, 1982, 3(1): 95-102.
- [22] Kitchen L, Rosenfeld A. Analysis of Gray Level Corner Detection [J]. Pattern Recognition Letters, 1999, 20(2): 149-162.
- [23] Harris C, Stephens M. A Combined Corner and Edge Detector [C]. Manchester: Proceedings of the 4th Alvey Vision Conference, 1988: 147-151.
- [24] 刘文予, 朱光喜. 二值图像角点检测的形态骨架法[J]. 信号处理, 2000, 16(3): 276-280.
- [25] 肖茜, 鲁宏伟. 基于高斯平滑的自适应角点检测[J]. 计算机辅助设计与图形学学报, 2003, 15(11): 1358-1361.
- [26] 钟宝江, 廖文和. 基于精化曲线累加弦长的角点检测技术[J]. 计算机辅助设计与图形学学报, 2004, 16(7): 939-943.
- [27] Langridge D J. Curve Encoding and the Detection of Discontinuities [J]. CVGIP, 1982, 20(1): 58-71.
- [28] Mediono G, Yasumoto Y. Corner Detection and Curve Representation Using Cubic B-splines [J]. Computer Vision, Graphics and Image Processing, 1987, 39(3): 267-278.
- [29] 乔宇, 黄席樾, 柴毅. 基于自适应直线拟合地角点检测[J]. 重庆大学学报, 2003, 26(2): 29-31.
- [30] Koenderink J J. The Structure of Image [J]. Biological Cybernetics, 1984, 50(5): 363-370.
- [31] Rattarangsi A, Roland T. Scale-based Detection of Corners of Planar Curves [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992, 14(4): 430-449.
- [32] 王展, 黄埔堪, 万建伟. 基于多尺度小波变换的二维图像角点检测技术[J]. 国防科技大学学报, 1999, 21(2): 46-49.

- [33] 戚飞虎, 刘健峰. 一种有效的不变性角点检测方法[J]. 上海交通大学学报, 1995, 29(6): 112-116.
- [34] Miroslav T, Mark H. Fast Corner Detection [J]. Image and Vision Computing, 1998, 16(1): 75-87.
- [35] Smith S M, Brady J M. SUSAN: A New Approach to Low Level Image Processing [J]. Int. Journal of Computer Vision, 1997, 23(1): 45-78.
- [36] 杨莉, 初秀琴, 李玉山. 最小亮度变化角点自适应检测算法研究[J]. 西安电子科技大学学报, 2003, 30(4): 530-533.
- [37] 费旭东, 荆仁杰. 基于知识的快速角点提取[J]. 计算机学报, 1994, 17(1): 30-36.
- [38] Giraudon G, Derich R. On Corner and Vertex Detection [J]. Computer Vision and Pattern Recognition, 1991, 3(6): 650-655.
- [39] 陈燕新, 戚飞虎. 一种新的提取轮廓特征点的方法[J]. 红外与毫米波学报, 1998, 17(3): 171-176.
- [40] Power W, Clist R. Comparison of supervised learning techniques applied to color segmentation of fruit image [A]. Proceeding of SPIE, Intelligent Robots and Computer Vision, XV: Algorithms, Techniques, Active Vision, and Material Handling [C]. Boston, MA, USA, 1996: 370-381.
- [41] Pal N R, Pal S K. A review on image segmentation techniques [J]. Pattern Recognition, 1993, 26(9): 1277-1294.
- [42] Yang C K, Tsai W H. Reduction of color space dimensionality by moment-p reserving thresholding and its application for edge detection in color images [J]. Pattern Recognition Letters, 1996, 17(5): 481-490.
- [43] Ohlander R, Price K, Reddy D R. Picture segmentation using a recursive region splitting method [J]. Computer Graphics and Image Processing, 1978, 8(3): 313-333.
- [44] Celenk M. A color clustering technique for image segmentation [J]. Computer Vision, Graphics, and Image Processing, 1990, 52(2): 145-170.
- [45] Underwood S A, Aggarwal J K. Interactive computer analysis of aerial color infrared photographs [J]. Computer Graphics and Image Processing, 1977, 6(1): 1-24.
- [46] 林开颜, 吴军辉, 徐立鸿. 彩色图像分割方法综述[J]. 中国图象图形学报, 2005, 1, 10(1): 1-10.
- [47] 罗希平, 田捷, 诸葛婴. 图像分割方法综述[J]. 模式识别与人工智能, 1999, 12(3): 300-312.

- [48] Whatmough R J. Automatic threshold selection from a histogram using the exponential hull [J]. *Graphical Models and Image Processing*, 1991, (53): 592-600.
- [49] 程豪, 黄磊, 刘昌平, 谭怒涛. 基于笔画和 Adaboost 的两层视频文字定位算法[J]. *自动化学报*, 2008, 10, 34(10): 1312-1318.