苏州大学

硕士学位论文

（2006届）

彩色图像内文字的自动提取与去除的研究

The Research of Automatic Text Extraction and Removal in Color Images

研究生姓名 季丽琴

指导教师姓名 王加俊

专业名称 通信与信息系统

研究方向 多媒体通信

论文提交日期 2006年5月

彩色图像内文字的自动提取与去除的研究

中 文 摘 要

随着计算机科学、多媒体技术的飞速发展，以彩色图像为主的多媒体信息迅速成为重要的通用信息媒体。在彩色图像中，文字信息(如新闻标题、旁白等等)包含了丰富的高层语义信息，自动提取出这些文字，通过对它们的识别和分析，对于图像高层语义的索引和检索是非常有帮助的。此外，还可将提取出的文字从原图中去除，同时修复被文字所遮挡的背景区域，然后添加上多语种的文字，这对于不同语种间的图像交流和图像的再次使用也是很有意义的。

利用彩色图像文字区域与背景之间存在明显边缘轮廓的特点，本文提出了一种新的图像文字提取算法—CEMA(Color-edge detection,Morphology,logic operator“AND”)。该算法首先用垂直、水平和对角三个方向的彩色边缘检测算子从原图中提取出三幅边缘图像，然后分别对这三幅边缘图像依次运用形态学中的闭、开、水平膨胀、水平腐蚀运算，得到三幅不同的连通域图，最后，将这三幅连通域图进行逻辑与运算，去除噪声，得到最终的文字区域。实验结果证明，CEMA 算法非常有效，文字提取率高，且具有鲁棒性。

在提取出图像内的文字区域后，本文运用纹理修复技术，将提取出的文字从原图中去除，同时，修复原图中被文字所遮挡的背景区域。实

验表明，该方法能很好地去除图像内的文字信息。

关键词：彩色边缘检测　数学形态学　文字提取　图像修复

作　　者：季丽琴

指导老师：王加俊

The Research of Automatic Text Extraction and Removal in Color Images

Abstract

As the rapid development of computer science and multimedia technology, the multimedia information, mainly composed of color images, has rapidly become an important general information media. Texts in color images, such as news headlines、aside etc, usually contain much high level semantic information. So it is very helpful for indexing and retrieving images to recognize and analyze those texts automatically extracted from images. On the other hand, it is also significant for image reusing and image intercommunication among different languages to remove texts embedded in images, restore the background occupied by texts, and then adding multi-language texts to the restored color images.

Using the feature of distinct edge contour existing between the text and the background regions in color images, a novel text extraction algorithm—CEMA(Color-edge detection, Morphology, logic operator " AND") is proposed in this thesis for video images. In the CEMA algorithm, three color edge detection operators in vertical, horizontal and diagonal directions are employed separately to obtain three edge images from the original image,

then, morphological operations, such as close、open、horizontal dilation、horizontal erosion, are used to get three different images of connective regions from the above three edge images. Finally, text regions are obtained after a successive implementation of the logical "AND" operation among the aforementioned three connective images and a de-noising procedure. Experiment results illustrate the effectiveness of the proposed CEMA algorithm in its high text extraction ratio and its robustness.

After the extraction of text regions in color images, the texture-based inpainting technology is used to restore the background occupied by the texts after a removal of the texts extracted from the original image. Experiments show that this method performs well in text removing and background restoration of color images.

Key word: color edge detection, mathematic morphology, text extraction, image inpainting

Written by Liqin Ji

Supervised by Jiajun Wang

苏州大学学位论文独创性声明及使用授权的声明

学位论文独创性声明

本人郑重声明：所提交的学位论文是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含其他个人或集体已经发表或撰写过的研究成果，也不含为获得苏州大学或其它教育机构的学位证书而使用过的材料。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本人承担本声明的法律责任。

研究生签名：李加勇　日　期：2006.5

学位论文使用授权声明

苏州大学、中国科学技术信息研究所、国家图书馆、清华大学论文合作部、中国社科院文献信息情报中心有权保留本人所送交学位论文的复印件和电子文档，可以采用影印、缩印或其他复制手段保存论文。本人电子文档的内容和纸质论文的内容相一致。除在保密期内的保密论文外，允许论文被查阅和借阅，可以公布（包括刊登）论文的全部或部分内容。论文的公布（包括刊登）授权苏州大学学位办办理。

研究生签名：李加勇　日　期：2006.5

导师签名：王[illegible]　日　期：2006.5

第一章 序 言

1.1 选题背景及意义

使计算机具有人类的感知能力，能够识图认字，能与人们自然地进行信息交互，是人们长期以来的梦想。赋予计算机识图认字的智能，对信息化发展具有其特殊重要的意义。在今天，数字化图像已经无处不在，由于数据处理技术的长足进步以及大容量系统的出现，在计算机的应用方面对图像信息的需求还会不断增加，而且迫切需要一个有效的算法完成对这些数据的按内容检索，而编制索引、检索、询议以及浏览则需要有一套自动化的方法能理解数字图像的内容。当前的数字图像主要是来自于数字电影、视频会议、医学成像和用于其他用途的监视视频。对这些数字图像数据库的多媒体文档的基于内容的信息检索技术正在受到广泛的研究，并且有着巨大的商业潜力。

图像中的文字在一定程度上反映了该图像的部分重要内容，通常形成对图像内容的简练描述或说明。例如，商业广告中的文字能够提供诸如产品名称、公司名称等重要信息；新闻图片中的文字能够说明发生的事件、时间、地点等。这些对于帮助人们理解图像的内容、检索相关图像都有着重要的作用。

图像文字提取由于其广泛的应用前景，近年来越来越受到人们的重视[1,2,3,4],在许多情况下都需要提取和识别图像中的文字，例如：在新闻节目中，通过识别屏幕下方的播音员的名字，我们就能按主持人的姓名为

视频编制索引，进而为个性化的新闻检索系统服务；又比如一场篮球比赛的画面，可以通过抽取运动员的号码、姓名和球队名称对比赛画面进行注释和索引。因此，自动定位图像中的文字区域，并提取出这些文字信息，通过对它们的识别和分析，对于图像理解、检索查询是很重要的。

图像文字去除实际上是典型的图像修复(image inpainting)问题。图像修复技术[5]是当前计算机图形学和计算机视觉中的一个研究热点，在影视特技制作、多余物体剔除(如视频图像中删除部分人物、字幕、小标题等)等方面有着重大的应用价值，例如：老式的电影视频由于技术限制，字幕被固化于视频中，成为视频图像的一部分。但在实际应用中人们发现，固化的字幕严重阻碍了不同语种间的视频交流。如果将字幕从视频图中去除，并修复被字幕所遮挡的背景，然后添加上多语种的字幕，则这将对不同语种间的视频交流、对视频图像的再次使用是非常有价值的。

1.2 研究现状

1.2.1 文字提取方法

到了 90 年代，随着多媒体技术的发展以及对基于内容的多媒体检索的需求，图像文字提取逐渐成为研究热点之一。通常从图像中提取文字都需要首先定位包含文字的图像区域，但文字在字体、大小、对齐方式和排列上变化多端，文字背景复杂，图像分辨率低，而且许多应用场合还要求算法具有一定处理速度，这些都使得从图像中有效地提取文字变得非常困难。很多学者在这方面已作出了有益的探索和尝试。目前，实现彩色图像中文字提取的方法主要有四类：边缘分析法、纹理分析法、

区域分析法和学习分析法。

1.2.1.1 边缘分析法

边缘分析法是通过寻找垂直边缘来检测文字。因为文字的笔画丰富，且文字所在图像区域的边缘非常丰富，所以该方法首先检测出图像的边缘，然后通过平滑滤波等方法将边缘连接成为文字块，再使用一些启发式规则来对文字块进行进一步筛选。文献[6]就先用一个3×3的水平差分过滤器来获得垂直边界，然后用平滑过滤器来使分离的文字部分相连，并排除多余碎片，再利用一些文字行的特征(如大于70像素、45%以上的填满率、横纵比率大于0.75等)来查找文字区域。在MIPS R4499 200MHz计算机上用该方法处理一幅352×240的图像大约需0.8s。还有文献[7,8,9]也是利用图像的垂直边缘检测来定位文字区域的。

虽然通过寻找垂直边缘可以达到快速检测文字的效果，但该方法不能适应图像背景的复杂变化，检测错误率较高。

1.2.1.2 纹理分析法

纹理分析法是利用纹理特征去决定一个像素点或者像素块是否属于文字。由于字符通常由许多较细笔划组成，因此存在笔划的区域通常也是全图纹理较丰富的区域，实现对纹理的寻找即可以寻找到字符的区域。Wu等提出了一种基于K-means的算法[10,11]去识别文字像素,该方法在3个标度下使用了9个高斯二阶导数。Li Huiping等使用神经网络在Haar小波解析特征空间去抽取文字块[12]。Jain在文献[13,14]中提出一种方法，该

方法综合分析了空间差异(纹理特征)和连通区域。此外，还有通过Gabor滤波[15],空间方差分析等通过分析纹理来提取文字区域的方法。

基于纹理的方法虽具有一定的通用性，但这类方法对于文字的字体和风格比较敏感，存在着定位不准和算法复杂度高的缺点，而且计算非常耗时[12]，因此，该方法的效率比较低。

1.2.1.3 区域分析法

区域分析法是把字符作为满足特定启发式规则的单色区域来检测。假设每个字符的像素都有相似的颜色，那么用图像分割[16,17]的方法或颜色聚类[18] 的方法或连通区域分析技术[19] 即可把字符从背景中分割出来，然后再使用一些简单的启发式规则，如区域的尺寸和长宽比或者基线等来对分割到的区域进行进一步筛选即可得到字符。文献[20]中的文字定位算法就是基于连通区域分析的，需要文字或其背景是单色的。

然而，由于图像中文字并不总是单色的，故这种方法对于复杂背景图像来说，其鲁棒性较差。由此可见，基于区域的方法只适用于二值图像，不适用于彩色图像[19]。

1.2.1.4 学习分析法

对视频图像内字幕进行定位面临很多困难，如：(1)字幕的大小尺寸经常发生变化；(2)字幕字体呈现多样性；(3)视频图像内的字幕和图像背景颜色都是多变的。由此可见，视频图像内字幕的定位不能只考虑字幕本身固有特征，还应该考虑利用一种学习机制去处理这些多变因素。

庄越挺等提出一种使用SVM机制[21]来自动定位提取视频字幕的方案，即首先对每幅视频图像按照$N\times N$大小切分成若干图像子块，然后把每个子块分别人工训练标注为字幕和非字幕两类，并通过提取图像的子块特征向量来训练SVM分类器。对于测试图像，则首先将其切分成子块，然后应用训练好的SVM分类器对其进行判断，最后通过后期处理进行去噪和合成，即可得到字幕提取结果。Chen Da-tong等也使用了SVM来进行图像中文字区的分类[22]。两种方法在样本不是很多的情况下，都实现了较高精度的视频字幕定位提取。

基于学习的方法作为一种智能识别方法，其虽在相当程度上解决了许多传统方法遇到的困难，但由于其需要事先通过选取样本来对分类学习机进行训练，所以，训练样本集与测试样本集的相似程度就决定了该方法的最终识别效果。

1.2.2 文字去除方法(图像修复方法)

文字去除实质是典型的图像修复问题。图像修复技术是指针对图像中遗失或者损坏的部分，利用未被损坏的图像信息，按照一定的规则填补，使修复后的图像接近或达到原图的视觉效果。图像修复技术可以安全有效地数字化恢复损坏的艺术作品，并可去除图像中的文字或者其他不期望的物体。目前，在图像修复领域主要有基于PDE(Partial Differential Equations)的修复方法和基于纹理的修复方法两大类。

1.2.2.1 基于 PDE 的修复方法

在2000年，Bertalmio[23]等人采用高阶偏微分方程的方法进行图像修复，取得了较好的结果。在该算法中，用户需指定要修复的区域。算法将图像分为3个独立的通道，对每个通道，将待修复区域边界的等值线(isophotes)外部的信息沿轮廓法向扩散到中间待修复的像素上。该算法用二维 Laplacian[24]方法估计局部颜色的光滑度，并利用这个光滑度沿着等值线进行扩散。为保证边缘处的边界连续性，算法考虑了各向异性的扩散[25]，该方法的缺点是计算不稳定。Chan等人通过扩展基于 TV(Total Variation)模型的去噪方法，提出了基于TV模型的修复方案[26],该方法能在含有噪声的情况下有效地对图像进行修复，取得较好的结果，但该方法对参数的选择比较敏感，且运算量较大。

1.2.2.2 基于纹理的修复方法

Criminisi等人提出一种基于纹理生成的修复方法[27],该算法在待修复区域的边界通过块匹配的方式选择合适的纹理填充。这种方法计算量小，对大区域也有较好的修复效果。考虑到图像中文字区域较大以及运算量的因素，本文采用纹理修复算法来进行图像修复。

1.3 本文的工作和贡献

本文的工作主要集中在图像文字提取算法的研究和图像背景的修复算法的研究两个方面：

在图像文字的提取方面，本文在对已有的图像文字提取算法进行研

究的基础上，利用文字区域与图像背景之间存在着明显的边缘轮廓这一实际情况，提出了一种具有高文字提取率的简单算法——CEMA (Color-edge detection， Morphology，logic operator "AND")算法。该算法首先分别运用垂直、水平、对角三个方向的彩色边缘检测算子从原图中提取出三幅不同的边缘图像，并结合形态学方法形成对应的连通域图，然后将三幅不同方向的连通域图进行逻辑与运算，消除噪声，得到最终的文字区域。此后，对文字区域作进一步的处理，提取出真正的文字信息，然后将提取出的文字送入已商业化的OCR软件模块进行识别，这对基于内容的图像检索是非常有帮助的。

在图像背景的修复方面，利用Criminisi[27]等人提出的基于纹理生成的图像修复方法，并结合本文提出的文字自动提取算法，在提取出原图像内的文字之后,修复被文字所遮挡的背景，这对不同语种间的视频图像交流是很有价值的。该方法的优点是待修复区域由文字提取算法自动指定而无需用户人工参与。

第二章 相关的图像处理方法

总的来说，数字图像处理一般包括以下几项内容：点运算、几何变换、图像增强、复原、编码、重建等[28]。这些方法在工业、生物医学、军事、通信等各种领域都有着广泛的应用。

在基于彩色图像的文字自动提取与去除的研究当中，对彩色图像的处理都是属于数字图像处理的范畴。图像处理技术对于本文目标的实现很重要，所涉及的图像处理方法大致有图像颜色处理、边缘提取、二值化、形态学处理等。以下将对它们做具体介绍。

2.1 图像颜色处理[29]

2.1.1 颜色处理的理论基础

为了用计算机表示和处理颜色，必须采用定量的方法来描述颜色，即建立颜色模型。目前广泛采用的颜色模型有 3 类，即计算机颜色模型、工业颜色模型和视觉颜色模型。计算颜色模型又称为色度学颜色模型，主要应用于纯理论研究和计算推导；工业颜色模型侧重于实际应用的实现技术；视觉颜色模型用于与人直接接口的颜色描述和控制。所有颜色模型的基础都建立在色度学理论之上。

色度学的基础理论是 T.Young 在 1802 年提出的，其基本内容是：任何彩色都可以用 3 种不同的基本颜色按不同的比例混合而得到，即

$$C = aC_1 + bC_2 + cC_3 \tag{2.1}$$

其中，C_1、C_2、C_3为三原色(又称为三基色)，a、b、c为三种原色的权值(三原色得比例或浓度)，C为所合成得颜色，可为任意颜色。该理论指出：

(1)自然界的可见颜色都可以用3种原色(基色)按一定的比例混合得到；反之，任意一种颜色都可以分解为3种原色；

(2)作为原色的3种颜色应该互相独立，即其中任何一种都不能用其它两种混合得到；

(3)三原色之间的比例直接决定混合色调的饱和度；

(4)混合色的亮度等于各原色的亮度之和。

由式(2.1)可知，一幅数字图像中的每一个像素都可以用三维彩色空间(C_1,C_2,C_3)中的一个向量$[a,b,c]^T$来表示。为了讨论的方便，将C_1、C_2、C_3的系数a、b、c结合到C_1、C_2、C_3中，直接用C_1、C_2、C_3来表示颜色。三原色的色度(chromaticities)由下式定义：

$$c_i = \frac{C_i}{C_1 + C_2 + C_3} \qquad (i = 1,2,3) \tag{2.2}$$

由于$c_1 + c_2 + c_3 = 1$，三原色的色度中仅有两个是独立的，因此，三维空间(C_1,C_2,C_3)可以映射到二维平面(c_1,c_2)上，整个颜色空间都可以用表示一个色度平面的三元组(c_1,c_2,Y)来表示，其中Y由下式定义：

$$Y = C_1 + C_2 + C_3 \tag{2.3}$$

2.1.2 灰度化处理

颜色可分为黑白色和彩色。黑白颜色指颜色中不包含任何的彩色成

分，仅由黑色和白色组成。在RGB颜色模型中，如果$R=G=B$，则颜色(R,G,B)表示一种黑白颜色；其中$R=G=B$的值叫做灰度值，所以黑白颜色又叫做灰度颜色。彩色和灰度之间可以互相转化，由彩色转化为灰度的过程叫做灰度化处理。

相应地，数字图像可分为灰度图像和彩色图像。通过灰度化处理，可以使彩色图像转化成灰度图像。

灰度化就是使彩色的R、G、B分量值相等的过程。由于R、G、B的取值范围是 0～255，所以灰度的级别只有 256 级，即灰度图像仅能表现 256 种颜色(灰度)。

灰度处理的方法主要有如下 3 种：

(1)最大值法：使R、G、B的值等于 3 值中最大的一个，即

$$R=G=B=\max(R,G,B) \tag{2.4}$$

最大值法会形成亮度很高的灰度图像。

(2)平均值法：求出R、G、B值的平均值，即

$$R=G=B=(R+G+B)/3 \tag{2.5}$$

平均值法会形成较柔和的灰度图像。

(3)加权平均值法：根据重要性或其它指标给R、G、B赋予不同的权值，并使R、G、B的值加权平均，即

$$R=G=B=(W_R R+W_G G+W_B B)/3 \tag{2.6}$$

其中W_R、W_G、W_B分别为R、G、B的权值。W_R、W_G、W_B取不同的值，加权平均值法就将形成不同的灰度图像。由于人眼对绿色的敏感度最高，

对红色的敏感度次之，对蓝色的敏感度最低，因此使$W_G > W_R > W_B$将得到较合理的灰度图像。实验和理论推导证明，当$W_R = 0.30$，$W_G = 0.59$，$W_B = 0.11$时，即当

$$V_{gray} = 0.30R + 0.59G + 0.11B$$

$$R = G = B = V_{gray} \tag{2.7}$$

时，能得到最合理的灰度图像。

2.1.3 颜色量化与减色

大多数的彩色图像采集系统都采用 24 位的真彩色来存储图像，但是要在仅能显示 256 色的显示系统中显示真彩色图像时，必须使用 8 位的 256 色图像，这就需要将 24 位真彩色图像转化为 8 位彩色图像，即进行减色处理。

将 24 位真彩色图像转化为 8 位彩色图像的核心是生成一个合适的调色板，用它来显示图像时能最好的反映原图像的彩色信息。由于 8 位彩色的调色板仅能使用 256 项颜色表项，因此，将 24 位真彩色图像转化为 8 位彩色图像时，就必须从 24 位真彩色所能表现的大约 16M 种颜色中，选取最具有代表性或出现频率最高的 256 种颜色。这种从m种颜色中选取最具代表性的n种颜色($m >> n$)的操作叫做颜色量化(color quantization)。

确定了用来填写调色板的 256 种颜色(236 种是从 16M 种颜色中选出的，20 种是 Windows 保留的)后，必须将 24 位真彩色的其余颜色赋值为选定的 256 种颜色中与它最相似的颜色。两种颜色的相似程度，可用它

们在 RGB 彩色空间中的距离来表示，这一距离称为彩色距离。在 RGB 彩色空间中颜色(r_1,g_1,b_1)和(r_2,g_2,b_2)的彩色距离ΔC可定义为：

$$\Delta C=(r_1-r_2)^2+(g_1-g_2)^2+(b_1-b_2)^2 \tag{2.8}$$

考虑到人眼对红、绿、蓝色的敏感度程度不同，在实际使用时，经常给各颜色分量的增量加上一定的权值，经验的彩色距离计算公式是：

$$\Delta C=3(r_1-r_2)^2+4(g_1-g_2)^2+2(b_1-b_2)^2 \tag{2.9}$$

在实际应用中主要有 3 种算法，可用于从 24 位真彩色中获取 236 种调色板颜色，即流行色算法(popularity algorithm)、中位切分算法(median-cut algorithm)和八叉树颜色量化算法(octree color quantization algorithm)。

2.1.3.1 流行色算法

流行色算法的基本思路是：对彩色图像中所有彩色出现的次数做统计分析，创建一个数组用于表示颜色和颜色出现频率的统计直方图。按出现递减的次序对该直方图数组排序后，直方图中的前 236 种颜色就是图像中出现次数最多(频率最大)的 236 种颜色，将它们作为调色板的颜色。该算法用统计直方图来分析颜色出现的频率，因此又称为彩色直方图统计算法。图像中其它的颜色采用在 RGB 颜色空间中的最小距离原则映射到与其邻近的 256 种调色板颜色上。

流行色算法的实现较简单，对颜色数量较小的图像可以产生较好的结果。但是该算法存在的主要缺陷是，图像中一些出现频率较低，但对人眼的视觉效果明显的信息将丢失。

2.1.3.2 中位切分算法

中位切分算法的基本思路是：在RGB彩色空间中，R、G、B三基色对应于空间的3个坐标轴，将每一坐标轴都量化为0～255。0对应于最暗(黑)，255对应于最亮(白)，这样就形成了一个边长为256的彩色立方体，所有可能的颜色都对应于立方体内的一个点：将彩色立方体切分成236个小立方体，每个立方体中都包含相同数量的在图像中出现的颜色点；取出每个小立方体的中心点，则这些点所表示的颜色就是所需要的最能代表图像颜色特征的236种颜色。

中位切分算法是 Paul Heckbert 在 20 世纪 80 年代初期提出来的，现被广泛应用于图像处理领域。该算法的缺点是涉及复杂的排序工作，而且内存开销较大。

2.1.3.3 八叉树颜色量化算法

1988 年，奥地利的 M.Gervautz 和 W.Purgathofer 发表了一篇题为“A Simple Method for Color Quantization：Octree Quantization”的论文，提出了一种新的采用八叉树数据结构的颜色量化算法，一般称为八叉树颜色量化算法。该算法的效率比中位切分算法高，而且内存开销小。

八叉树颜色量化算法的思路是：将图像中使用的RGB颜色值分布到层状的八叉树中。八叉树的深度可达9层，即根节点加上分别表示8位的R、G、B值的每一位的8层节点。RGB值中更重要的位放在八叉树的较上层，较低的节点层则对应于较不重要的RGB值的位(右边的位)，因此为了提高效率和节省内存，可以去掉最低部的2～3层，这样不会对结果有太大

的影响。叶节点编码存储像素的个数和R、G、B颜色分量的值；中间的节点组成了从最顶层到叶节点的路径。这是一种高效的存储方式，既可以存储图像中出现的颜色和其出现的次数

，也不会浪费内存去存储图像中不出现的颜色。

扫描图像的所有像素，每遇到一种新的颜色就将它放入八叉树中，并创建一个叶节点。图像扫描完后，如果叶节点的数量大于调色板所需的颜色数(236)时，对所有叶节点按其编码内像素点的个数进行排序，找到像素点个数之和为最小的那些叶节点，把它们合并成一个新的节点，并将其转化成上一层节点的叶节点，在其中存储颜色及其出现的次数。八叉树减少叶节点的示意图见图 2.1。这样，减少叶节点的数量，直到叶节点的数量等于或小于调色板所需的颜色数。如果叶节点的数量小于或等于调色板所需的颜色数(236)，则可以遍历八叉树，将叶节点的颜色填入调色板的颜色表中。

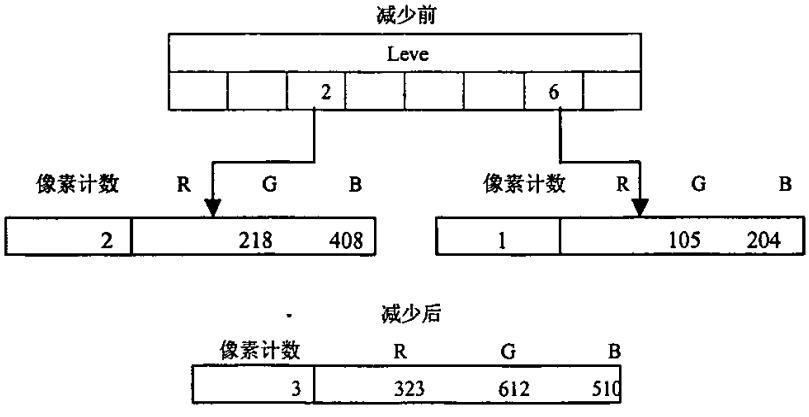
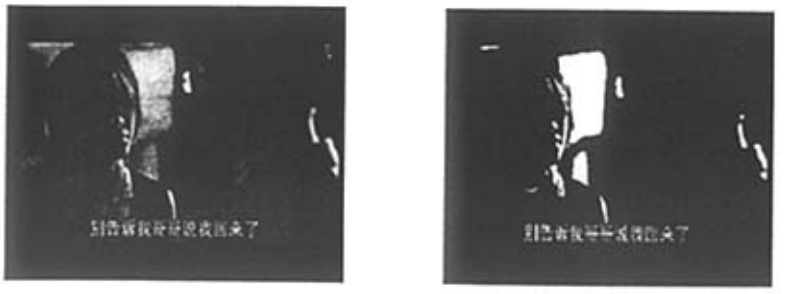


图 2.1 八叉树减少叶节点示意图

在本文的彩色图像文字提取与去除中，采用的是八叉树颜色量化算

法实现将提取出的文字区域的颜色减少为指定的数目(比如减少为只有两种颜色的图像，即为 1 位二值图像)。图 2.2(b)为 24 位真彩色图像转化为 1 位二值图像的结果。



(a) 输入的 24 位真彩色图像　　(b) 减色后的 1 位图像

图 2.2　八叉树颜色量化算法实现颜色的减色

2.2　边缘检测

图像边缘是图像的最基本特征。所谓边缘是指其周围像素灰度有阶跃变化或屋顶变化的那些像素的集合。边缘广泛存在于物体与背景间、物体与物体之间。因此，它是图像分割所依赖的重要特征[30]。经典的边缘检测方法是考察图像的每个像素在某个邻域内灰度的变化，利用边缘邻近一阶或二阶方向导数变化规律，用简单的方法来检测边缘。如果一个像素落在图像中某一个物体的边界上，那么它的邻域将成为一个灰度级的变化带。对这种变化最有用的两个特征是灰度的变化率和方向，它们分别以梯度向量的幅度和方向来表示。边缘检测算子检查每个像素的邻域，并对灰度变化率进行量化，也包括方向的确定。大多数使用基于方向导数掩模求卷积的方法。下面介绍几种常用的边缘检测算子[30]。

2.2.1 Roberts 边缘检测算子

Roberts边缘算子采用的是对角方向相邻的两个像素之差。从图像处理的实际效果来看，边缘定位准，对噪声敏感。Roberts边缘检测算子是一种利用局部差分算子寻找边缘的算子，它由下式给出。

$$g(x,y)=\sqrt{[f(x,y)-f(x+1,y+1)]^2+[f(x+1,y)-f(x,y+1)]^2} \tag{2.10}$$

其中$f(x,y)$是具有整数像素坐标的输入图像，$g(x,y)$表示处理后(x,y)点的灰度值，平方根运算使该处理类似于在人类视觉系统中发生的过程。该算法的算子如下：

0	1
-1	0

(a)

1	0
0	-1

(b)

图2.3 Roberts边缘检测算子

2.2.2 Sobel 算子

图2.4所示的两个卷积核形成了Sobel边缘算子，图像中的每个点都用这两个核做卷积，一个核对通常的垂直边缘响应最大，而另一个对水平边缘响应最大。两个卷积的最大值作为该点的输出位。

-1	-2	-1
0	0	0
1	2	1

(a) 水平边缘Sobel算子

-1	0	1
-2	0	2
-1	0	1

(b)垂直边缘Sobel算子

图2.4 Sobel边缘检测算子

2.2.3 拉普拉斯算子

拉普拉斯算子是对二维函数进行运算的二阶导数算子。由于拉普拉斯算子是一个二阶导数，它将在边缘处产生一个陡峭的零交叉。通常使用的拉普拉斯算子如图2.5所示。

0	-1	0
-1	4	-1
0	-1	0

(a)

-1	-1	-1
-1	8	-1
-1	-1	-1

(b)

图2.5　拉普拉斯边缘检测算子

对于彩色图像的边缘检测来说，要求有高精度的图像边缘定位、抑制噪声的能力。要达到这个要求，就必须综合考虑色彩的全部信息。在后面的相关章节中，本文将讨论如何将上述经典的灰度图像边缘检测算子推广到彩色图像中，设计出适合彩色图像文字提取的彩色边缘检测算子。

2.3 二值化

在对彩色图像内的文字进行提取时，最终往往是对黑白二值图像进行操作。灰度图像转换为黑白二值图像的操作称为二值化，或阈值化。最简单的二值化可以通过设定阈值实现。如果图像中某像素的灰度值小于该阈值，则将该像素的值设为0，否则设为255。它的表达式如下:

$$f(x,y)=\begin{cases}0 & x<T\\ 255 & x>T\end{cases} \tag{2.11}$$

其中T为指定的阈值。由公式(2.11)可以看到，二值化的关键是阈值T的选择。习惯上常用三元函数的形式表示阈值T：

$$T = T(f(x,y), N(x,y), (x,y)) \tag{2.12}$$

其中(x,y)是像素坐标，$f(x,y)$是坐标为(x,y)的像素的灰度值，$N(x,y)$表示坐标为(x,y)的像素的局部灰度特征。

根据彩色图像文字提取的特点，本文采用整体阈值二值化，即用最小误差方法求得分割阈值，该算法简单，时间开销小，可以较好地分离背景和文字。在后面的相关章节中将具体介绍这一方法。

2.4 数学形态学

数学形态学是一门建立在严格的数学理论基础之上的学科。它主要以积分几何、几何代数及拓扑论为理论基础，此外还涉及随机集论、现代概率论、近世代数、图论等一系列数学分支[30,31]。它以图像的形态特征作为研究对象，它的主要内容是设计一整套概念、变换和算法用以描述图像的基本特征和基本结构，即描述一幅图像中元素与元素、部分与部分之间的关系。数学形态学的理论虽然很复杂，但它的基本思想却是简单而完美的。它最基本的思想是把图像看成是点的集合，用结构元素(structuring element)对其进行移位、交、并等集合运算就构成形态学的各种处理算法。其中结构元素也是由点的集合构成的，如线性、圆形、方形，它相当于一种“探针”，在图像中不断地移动结构元素，便可以考察各个部分的关系。不同的点的集合形成不同性质的结构元素。由于不同

的结构元素可以用来检测图像不同侧面的特征，如目标的大小、形状、连通性和方向等，因此可以将结构元素理解成为观察图像的手段或角度。数学形态学将集合形状化为代数的形式，通过一系列变换达到处理的目的[30]。

数学形态学的基本运算有4个:腐蚀(erosion)、膨胀(dilation)、开(open)和闭(close)。基于这些基本运算还可以推导和组合成各种数学形态学实用算法。下面分别介绍这几种形态学基本运算。

2.4.1 腐蚀

把结构元素B平移a后得到Ba，若Ba包含于X，记下这个a点，所有满足上述条件的a点组成的集合称做X被B腐蚀(Erosion)的结果。用公式表示为:$E(X)=\{a|Ba\subset X\}=X\Theta B$,如图2.6所示。

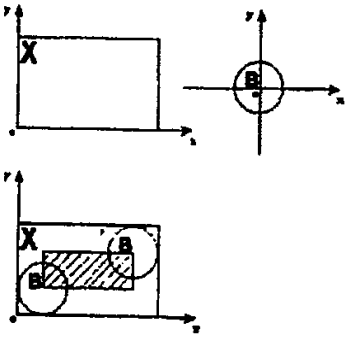


图2.6 腐蚀的示意图

图2.6中X是被处理的对象，B是结构元素。不难知道，对于任意一个在阴影部分的点a,Ba包含于X，所以X被B腐蚀的结果就是那个阴影部分。阴影部分在X的范围之内，且比X小，就像X被剥掉了一层似的，这就是

为什么叫腐蚀的原因。

下面具体看一下腐蚀运算操作的过程。在图2.7中，左边是被处理的图像X，中间是结构元素B,标有origin的点是中心点，即当前处理元素的位置。腐蚀的方法是，拿B的中心点和X上的点一个一个地比对，如果B上的所有点都在X的范围内，则该点保留，否则将该点去掉。右边是腐蚀后的结果。

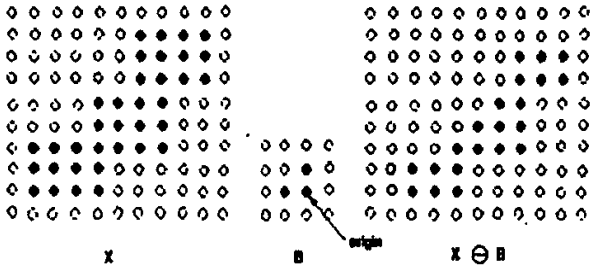


图2.7 腐蚀运算

如果结构元素B是关于原点对称的，即B的对称集B^V=B，那么X被B腐蚀的结果和X被B^V腐蚀的结果是一样的。如果B不是对称的，那么X被B腐蚀的结果和X被B^V腐蚀的结果将不同，如图2.8所示。

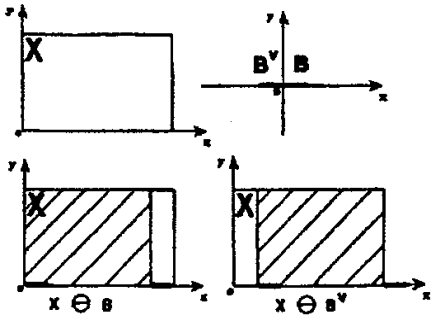


图2.8 结构元素非对称时，腐蚀的结果不同

2.4.2　膨胀

膨胀(dilation)可以看成是腐蚀的对偶运算，其定义是:结构元素B平移a后得到Ba，若Ba击中X，记下这个a点。所有满足上述条件的a点组成的集合称做X被B膨胀的结果。用公式表示为:D(X)= {a|Ba ↑ X} =X⊕B,如图2.9所示。图2.9中X是被处理的对象，B是结构元素，不难知道，对于任意一个在阴影部分的点a,Ba击中X,所以X被B膨胀的结果就是那个阴影部分。阴影部分包括X的所有范围，就像X膨胀了一圈似的，这就是为什么叫膨胀的原因。

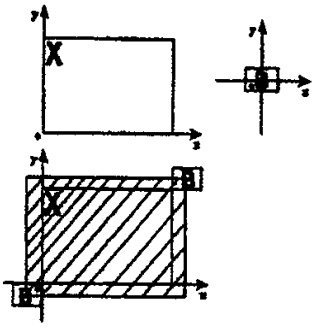


图2.9　膨胀的示意图

下面具体看一下膨胀运算的操作过程。在图2.10中，左边是被处理的图像X，中间是结构元素B。膨胀的方法是，拿B的中心点和X上的点及X周围的点一个一个地比对，如果B上有一个点落在X的范围内 ,则该点就为黑。右边是膨胀后的结果。可以看出，它包括X的所有范围，就像X膨胀了一圈似的。

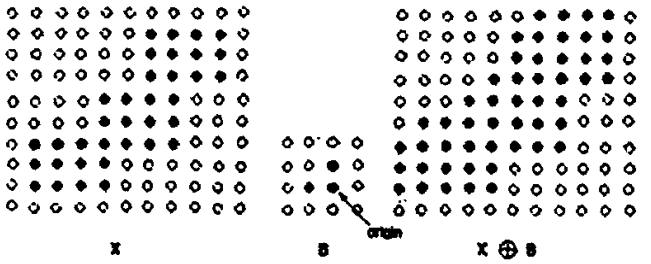


图2.10 膨胀运算

同样，如果B不是对称的，X被B膨胀的结果和X被B^V膨胀的结果不同。腐蚀运算和膨胀运算互为对偶的，用公式表示为$(X\Theta B)^C=(X^C\oplus B)$，即X被B腐蚀后的补集等于X的补集被B膨胀。

2.4.3 开

先腐蚀后膨胀称为开(open),即OPEN(X)=D(E(X))。下面具体看一下开运算操作的过程。如图2.11所示。

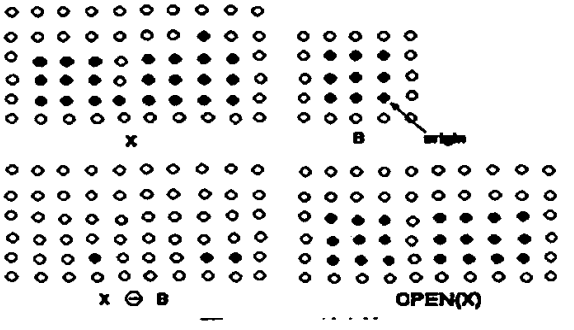


图2.11 开运算

在图2.11上面的两幅图中，左边是被处理的图像X，右边是结构元素B,下面的两幅图中左边是腐蚀后的结果:右边是在此基础上膨胀的结果。可以看到:原图经过开运算后，一些孤立的小点被去掉了。一般来说，开运算能够去除孤立的小点，毛刺和小桥(即连通两块区域的小点)，有对图像进行平滑的作用[30]。这就是开运算的作用。

2.4.4　闭

先膨胀后腐蚀称为闭(close)，即CLOSE(X)=E(D(X))。下面具体看一下闭运算的操作过程。如图2.12所示。

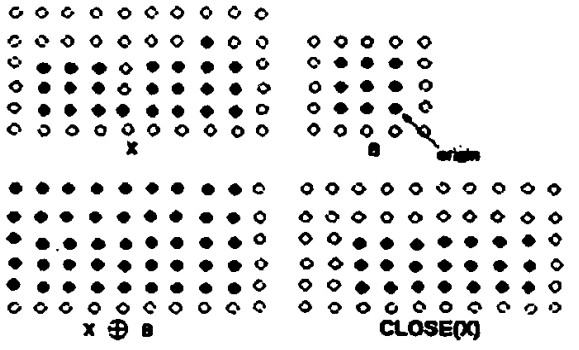


图2.12　闭运算

在图2.12上面的两幅图中，左边是被处理的图像X，右边是结构元素B,下面的两幅图中左边是膨胀后的结果，右边是在此基础上腐蚀的结果。可以看到:原图经过闭运算后，断裂的地方被弥合了。一般来说，闭运算能够填平小孔，弥合小裂缝，这就是闭运算的作用[30]。开和闭也是对偶运算，用公式表示为$(OPEN(X))^C=CLOSE((X^C))$，或者$(CLOSE(X))^C=OPEN((X^C))$。即X开运算的补集等于X的补集的闭运算，或者X闭运算的补集等于X的补集的开运算。

第三章 文字区域的提取

对于彩色图像中的文字来说，其字符排列有序，字体基本相同，最重要的是字符本身与背景之间存在着明显的边缘轮廓，因此在提取文字时，如能有效利用文字的上述特点，对于最终从彩色图像中提取出文字是十分重要的。本文提出用垂直、水平、对角方向的彩色边缘检测算子来检测出不同方向的文字边缘，并结合逻辑与运算和相关的图像处理方法，很好地完成了文字区域的提取。

3.1 文字区域提取算法—CEMA

通过观察可知，图像中的文字与背景之间存在着明显的边缘轮廓，且字符排列有序。基于以上特点，本文提出了基于彩色边缘检测、形态学和逻辑与运算的文字区域提取算法—CEMA(Color-edge detection, Morphology, logic operator "AND")。该算法的流程图如下所示。

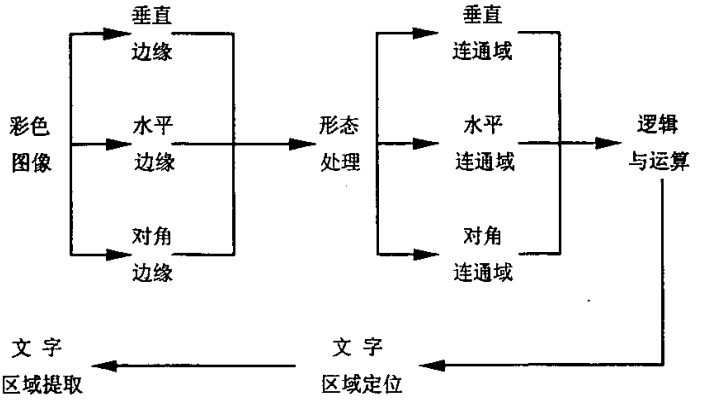


图 3.1 文字区域提取算法—CEMA 的流程图

3.2 垂直、水平、对角方向的彩色边缘检测

由于彩色图像中的文字与背景有较强的对比度，表现为在文字与背景的交界处，存在十分明显的高频区域，因此可以用提取边缘的方法来估计出文字可能存在的区域。文献[12]将小波分析推广到二维情形，即通过多分辨率分析和 Mallat 塔式分解方法，得到文字在垂直、水平以及对角方向的分量图。在分析了上述方法的基础上，本文提出用三个简单的不同方向的彩色边缘检测算子来代替小波变换，如图 3.2 所示。

0	0	0
-3	3	0
0	0	0

(a)垂直方向

0	-3	0
0	3	0
0	0	0

(b)水平方向

-3	0	0
0	3	0
0	0	0

(c)对角方向

图 3.2　三个不同方向的边缘检测算子

这三个检测算子分别作用于彩色图像的红、绿、蓝三个分量上来提取边缘，以像素点(i,j)为例，定义垂直方向彩色边缘检测(其它方向的检测算子雷同)如下：

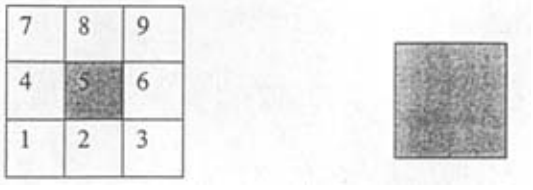
$$R_V(i,j)=\sum_{k=1}^{9}e_V(k)*R(k) \tag{3.1}$$

$$G_V(i,j)=\sum_{k=1}^{9}e_V(k)*G(k) \tag{3.2}$$

$$B_V(i,j)=\sum_{k=1}^{9}e_V(k)*B(k) \tag{3.3}$$

其中，$R_V(x,y)$、$G_V(x,y)$、$B_V(x,y)$分别是利用垂直检测算子在像素点(i,j)处得到的红、绿、蓝分量；$R(k)$、$G(k)$和$B(k)$分别是在像素点(i,j)处及它

的八邻域内读取到的红、绿、蓝分量；$e_v(k)(k=1,..9)$是垂直边缘检测算子中的数值。图3.3给出了检测算子扫描像素点(i,j)周围的像素的顺序。



(a)检测算子扫描像素的顺序 (b)当前像素点(i,j)

图 3.3 检测算子扫描像素的顺序

通过以上垂直、水平和对角方向算子的检测，得到相对应的彩色边缘图像$E_V(x,y)$、$E_H(x,y)$、$E_D(x,y)$，从而避免运用复杂的小波变换来提取文字边缘特征，可见本文方法的可行性和简单性。

3.3 灰度化处理和二值化

通过上述的边缘检测后，得到的是彩色的边缘图像，因此必须对这些彩色边缘图像进行灰度化处理，这样才能进行二值化。于是，本文采用第二章介绍的灰度化处理中的第三种方法，即加权平均值法来对彩色边缘图像进行灰度化。实验证明，该法取得了较好的效果。

对于文字提取来说，边缘图像的二值化是很重要的问题，二值化处理的效果将直接影响到文字提取的准确度。从第二章的分析可知，二值化的关键在于阈值的选择，因此本文针对不同的边缘图像采用整体阈值二值化，即用最小误差法求分割阈值。

该方法的基本思想是找到一个阈值，使按这个阈值划分目标和背景的错误分割概率最小。假设图像中的目标像素点的灰度服从正态分布，

密度函数为$P_1(x)$，均值和方差为μ_1和σ_1^2，设背景点的灰度也服从正态分布，密度为$P_2(x)$，均值和方差为μ_2和σ_2^2(图3.4显示了正态分布的目标和背景的灰度分布曲线)。

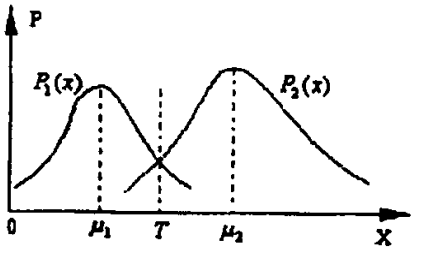


图 3.4 目标和背景灰度为正态分布下的最优阈值确定

设目标的像素点数占图像总点数的百分比为Q,背景点占$(1-Q)$，则混合概率密度为：

$$P(X)=QP_1(x)+(1-Q)P_2(x)$$

$$=\frac{Q}{\sqrt{2\pi}\sigma_1}\exp\left[-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right]+\frac{1-Q}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right] \tag{3.4}$$

可定义阈值T,使得所有灰度值小于T的像素可以被认为是目标点，而所有灰度值大于T的像素点可以被认为是背景点。此时，将目标点误判为背景点的概率为：

$$E_1(T)=\int_T^{\infty}P_1(x)dX \tag{3.5}$$

把背景点误判为目标点的概率为：

$$E_2(T)=\int_{-\infty}^{T}P_2(x)dX \tag{3.6}$$

则总的误判概率为：

$$E(T)=QE_1(T)+(1-Q)E_2(T) \tag{3.7}$$

为了找到一个阈值T使得上述的误判概率为最小，必须将$E(T)$对T求微

分，并令其结果等于零，即令$\frac{\partial E(T)}{\partial T}=0$,则有：

$$QP_1(T)=(1-Q)P_2(T) \tag{3.8}$$

在式(3.8)中代入$P_1(T)$、$P_2(T)$的表达式，变形可得：

$$\frac{Q}{\sqrt{2\pi}\sigma_1}\exp\left[-\frac{(T-\mu_1)^2}{2\sigma_1^2}\right]=\frac{1-Q}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{(T-\mu_2)^2}{2\sigma_2^2}\right] \tag{3.9}$$

两边取对数，化简可得：

$$\frac{(T-\mu_1)^2}{2\sigma_1^2}-\frac{(T-\mu_2)^2}{2\sigma_2^2}=\ln\frac{Q\sigma_2}{(1-Q)\sigma_1} \tag{3.10}$$

当$\sigma_1^2=\sigma_2^2=\sigma^2$时，由式(3.10)可得：

$$T=\frac{\mu_1+\mu_2}{2}+\frac{\sigma^2}{\mu_2-\mu_1}\ln\frac{Q}{1-Q} \tag{3.11}$$

若$Q=1/2$，则公式(3.11)变形为：

$$T=\frac{\mu_1+\mu_2}{2} \tag{3.12}$$

这表示当目标点和背景点各占图像的一半时，且其具有相同的分布时，图像的最佳分割阈值就是目标和背景灰度分布中心的平均值。

基于上述思想，迭代求图像最佳分割阈值的算法描述如下：

(1)求出图像中的最小、最大灰度值Z_1和Z_k，阈值初值为：

$$T^0=(Z_1+Z_k)/2 \tag{3.13}$$

式中k为迭代次数，其初值为0；

(2)根据阈值T^k将图像分割成目标和背景两部分，使得所有灰度值小于T^k的像素可以被认为是目标点，而所有灰度值大于T^k的像素点被认为是背景点。分别求出目标和背景的平均灰度值Z_O和Z_B：

$$Z_O = \frac{\sum_{Z(x,y)<T^k} Z(x,y)\times N(x,y)}{\sum_{Z(x,y)<T^k} N(x,y)} \tag{3.14}$$

$$Z_B = \frac{\sum_{Z(x,y)>T^k} Z(x,y)\times N(x,y)}{\sum_{Z(x,y)>T^k} N(x,y)} \tag{3.15}$$

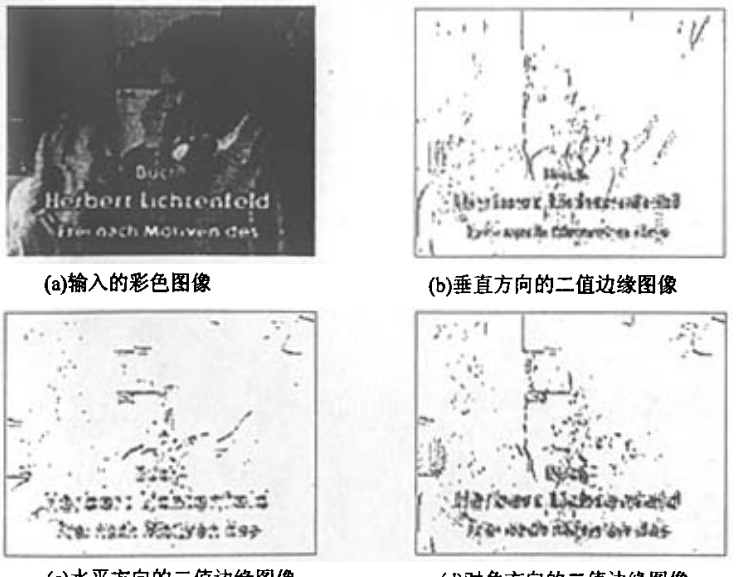
式(3.15)中$Z(x,y)$是图像上点(x,y)的灰度值，$N(x,y)$是点(x,y)的权重系数，一般取$N(x,y)=1.0$。

(3)迭代，求出新的阈值：

$$T^{k+1} = (Z_O + Z_B)/2 \tag{3.16}$$

(4)如果$T^{k+1}=T^k$或$k>100$，则结束；否则令$k=k+1$，转到第二步。

图 3.5 为输入的彩色图像在经过了彩色边缘检测、灰度化处理之后，采用最小误差法迭代求出的阈值对垂直、水平和对角边缘图像进行二值化处理的结果。



(a)输入的彩色图像　(b)垂直方向的二值边缘图像

(c)水平方向的二值边缘图像　(d)对角方向的二值边缘图像

图 3.5 不同方向的二值边缘图像

3.4 形态处理

形态学可将图像信号与其几何形状联系起来，利用一定形态的结构元素度量和提取图像中的对应形状和结构，所以本文采用形态学处理来提取文字在彩色图像中对应的形状。

由第二章知道，形态学最基本的操作有腐蚀、膨胀、开、闭。其中，膨胀具有扩大目标区域的作用，腐蚀具有收缩目标区域的作用，开运算可删除目标区域中的小分支，闭运算可填补目标区域中的空洞。基于这四个运算的特点，本文对上述三幅二值边缘图像的形态处理流程为：

(1) 一次闭：采用 8 连通结构元素($B=\begin{Bmatrix}1,1,1\\1,1,1\\1,1,1\end{Bmatrix}$)填补边缘图像中文字区域的空洞；

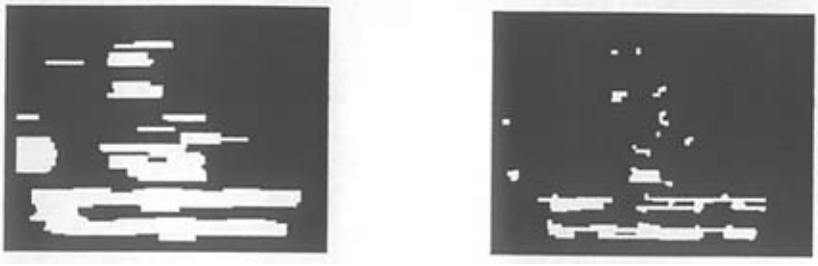
(2) 一次开：采用 8 连通结构元素($B=\begin{Bmatrix}1,1,1\\1,1,1\\1,1,1\end{Bmatrix}$)删除文字区域中的小分支；

(3) 六次水平膨胀：因为图像中文字一般是水平方向分布，所以为了有效地形成候选文字连通域，本文采用水平结构元素 B= {1,1,1,1,1,1}；

(4) 三次水平腐蚀：因膨胀后文字连通域明显大于实际文字区域的大小，所以需要进行水平方向的腐蚀，以确保形成的候选文字连通域与实际的文字区域大小相差无几。因此，本文采用水平结构元素 B= {1,1,1,1,1,1}；

实验发现，若水平结构元素的尺寸太大(比如B= {1,1,1,1,1,1,1,1})，会导

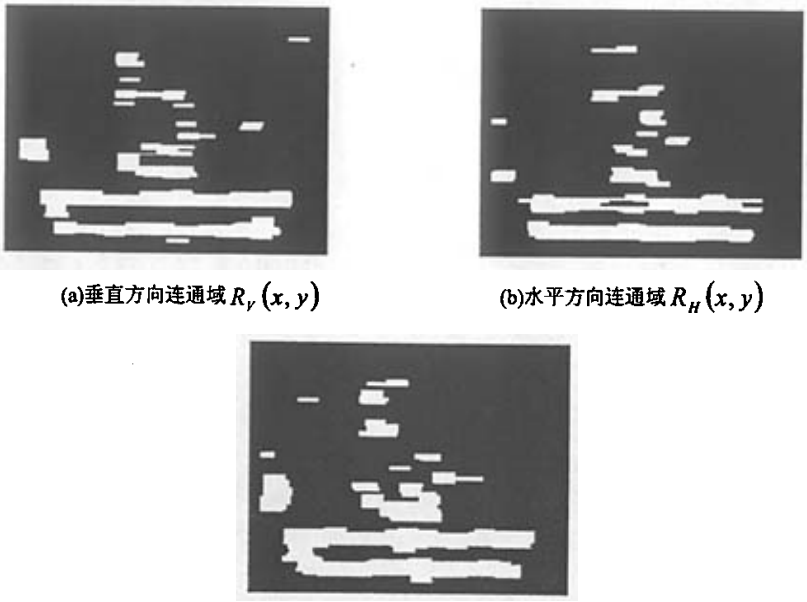
致无效的膨胀重叠现象，增大计算量，而且形成的对角方向的候选文字连通域(如图3.6(a))与实际的文字区域大小(如图3.5(a))之间存在较大的误差。若水平结构元素尺寸太小(比如B= {1,1,1})，将不能有效地形成水平方向的候选文字连通域，从而会导致文字的漏检(如图3.6(b))。



(a)结构元素过大形成的对角方向连通域 (b)结构元素过小形成的水平方向连通域

图 3.6 结构元素尺寸过大或过小导致的结果

所以，结构元素 B 的选择对于候选文字连通域的形成与文字区域的提取至关重要。实验证明，本文采用的结构元素很好地形成了各个方向的候选文字连通域$R_V(x,y)$、$R_H(x,y)$、$R_D(x,y)$(如图 3.7)。



(a)垂直方向连通域$R_V(x,y)$ (b)水平方向连通域$R_H(x,y)$

(c)对角方向连通域$R_D(x,y)$

图 3.7 本文采用的结构元素形成的各个方向的候选文字连通域

3.5 逻辑与运算定位文字区域

为了定位最终的文字区域，本文提出将上述三幅候选文字连通域图进行逻辑与运算，因为此运算综合了垂直、水平和对角方向的文字信息，能有力地保证文字区域存在的准确性和精确性。实验证明，经过此运算后，去掉了很大部分噪声区域，得到较精确的连通域图$R_I(x,y)$,用公式表示为:

$$R_I(x,y)=R_V(x,y)\cap R_H(x,y)\cap R_D(x,y) \tag{3.17}$$

但是，$R_I(x,y)$中仍可能存在一些虚假的不含文字的连通区域，所以需对所有连通区域做进一步分析。和文献[32]中采用了七个判定规则来判断所有连通区域相比，本文在进行逻辑与运算之后，仅需运用一个判定规则就可得到确定的只有文字区域的连通域图$R_{II}(x,y)$。该规则为：采用递归算法[33]来递归统计各连通区域的白色像素总数(记为 PixelNum)，若PixelNum<areapixel(经实验，本文将 areapixel 定为图像高度×图像宽度/150),则将虚假的不含文字的区域删除。实验证明，本文的文字区域定位方法简单且非常有效。其实验结果($R_I(x,y)$，$R_{II}(x,y)$)如图 3.8 所示。

(a)基于与运算得到的$R_I(x,y)$ (b)得到的最终的文字区域$R_{II}(x,y)$

图 3.8 基于逻辑与运算及判定规则得到的结果

3.6 文字区域的提取

3.6.1 外接矩形的形成

为了最终提取出定位好的文字区域，本文采用递归算法[33]求出文字区域的外接矩形，此算法的大致思想为：首先定义一个全局 Crect 类型数组变量 textLocation[n](n 为图像中的文字区域数)，top、left、bottom、right 分别为 textLocation[n]的四个成员变量，代表外接矩形的左上角和右下角的坐标。若像素点(x,y)为白色，则扫描点(x,y)的 4 邻域像素点，若 4 邻域中仍存在白色像素点，则调整全局变量 textLocation[n]结构中的 top、left、bottom、right 的大小，然后再进行递归调用。算法结束返回的 textLocation[n]即为所求的外接矩形，设此函数为 TextLocate(x,y,rect),调整全局变量 Crect 结构的函数为 RecursiveRect(x,y,rect)，则此递归算法的形式化描述如下：

```
TextLocate(x,y,rect)
  begin
      if 像素点(x,y)是黑色的，算法结束;
      else
        begin
        if 像素点(x,y)是白色的，then
            begin
            RecursiveRect(x,y,rect);
             end;
        end;
end;
```

最终利用求出的外接矩形(如图 3.9 所示)来提取出图像中的文字区域。

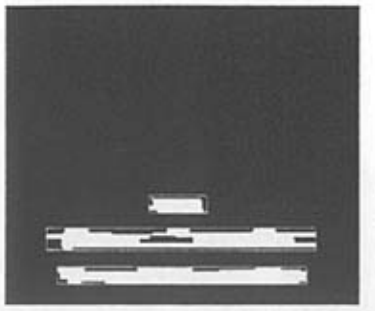


图 3.9 在文字区域周围形成的外接矩形

3.6.2 实验结果及分析

本文的彩色图像内文字的自动提取是在 Windows 环境下用 Visual C++6.0 实现的，以下实验所用的视频图像均来自网站 MoCA(Movie Content Analysis)[34]。图 3.10(a)和(b)是原彩色图像，图 3.10(c)和(d)是本文方法的文字提取结果。为了与文献[34]的方法作进一步的对比，我们也给出了其方法所得到的文字提取结果，如图 3.10(e)和(f)所示。

(a) 原图

(b) 原图

(c)本文方法提取的文字区域

(d)本文方法提取的文字区域

(e)文献[34]提取的文字区域

(f)文献[34]提取的文字区域

图 3.10　彩色图像内文字提取的结果及比较

从以上这些结果，可以明显看出，本文提出的文字提取方法—CEMA不但简单，而且对于文字区域的定位很精确。从图 3.10(e)、(f)中可看出，文献[34]提取出的文字区域中有许多是非文字区域，而原图中本是文字的，却没有提取出来。可见，本文提出的方法要优于文献[34]所采用的方法。

以上都是针对文字是水平方向排列及字符颜色单一的情况，为了验证本文方法的鲁棒性，本文对文字倾斜方向排列和文字有多种颜色的图像进行了测试，如图 3.11 所示。

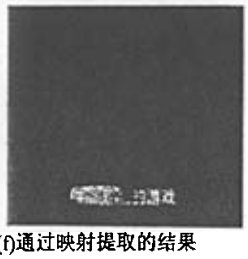
(a)文字倾斜方排列

(b)通过外接矩形提取的结果

(c)通过映射提取的结果

(c)多种颜色的文字

(d)通过外接矩形提取的结果

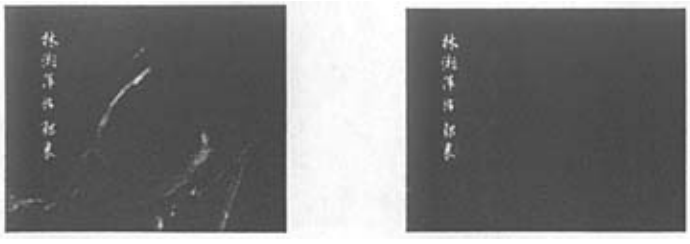


(f)通过映射提取的结果

图 3.11　不同情况下的文字提取结果

通过这些图可以看出：对于倾斜文字而言，映射(把得到的白色文字区域直接对应到原图中的位置，从而得到相应的彩色的文字区域)提取的结果要优于外接矩形提取的结果，因为图 3.11(b)中的非文字部分要多于图 3.12(c)；但文字水平方向排列时结果却恰恰相反，外接矩形精确地提取出文字区域，无一漏检，而通过映射提取出的文字区域有较大的缺损。所以当文字水平排列时，用外接矩形法提取文字区域要优于用映射法提取，而当文字倾斜时，映射法要优于外接矩形法，不过一般情况下，图像中(尤其在视频帧内)的文字都是水平方向排列的，所以比较通用的方法就是外接矩形法提取文字区域。

除了上述两种情况外，还有一种特殊情况，即文字垂直方向排列。此时如果再用 CEMA 算法中的水平膨胀和水平腐蚀处理显然不合适，所以，本文只需将 CEMA 算法中的水平膨胀和水平腐蚀分别改为垂直膨胀和垂直腐蚀处理就可，即将水平结构元素改为垂直结构元素。如图 3.12 所示。



(e)文字垂直排列 (f)通过外接矩形提取的结果

图 3.12 垂直文字的提取结果

为了尽可能包罗文字提取可能出现的各种情况，本文对100幅背景的颜色和底图非常复杂的图像进行了测试，其中有的同一行字符中有多种

颜色，文字有中文也有英文，字体多样，文字尺寸有大有小。

在实验之前我们已经手工统计了这100幅图像中所包含的文字总数，实验结果的比较如表1所示。

表1 电影、广告中文本提取结果

	含文字的图像数目	正确提取的文本区域数	提取出的所有区域数	文本提取正确率(%)
电影类	60	106	113	93.8
广告类	40	70	78	89.7

从表中可看出，本文的CEMA文字提取算法取得了较高的文字提取率。因此，基于彩色边缘检测和形态学且运用逻辑与运算的文字提取方法非常有效、简单且具有好的鲁棒性。

第四章　图像修复

4.1　文字区域的后处理

纹理修复的目标区域并非是整个文字区域，而只是文字区域中的文字像素，不包括文字区域中的背景。一般情况下，要区分文字像素还是背景区域的前提是在文字像素的周围存在一些高亮度的像素。如图 4.1 所示。

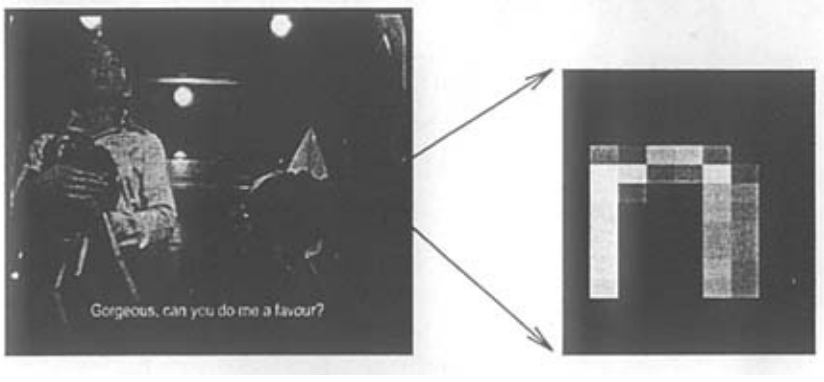


图 4.1　在文字像素周围存在高亮度像素的例子

从上图中可以看出，在文字像素“n”的边缘存在着一些黑色像素，这些像素突出了“n”的存在，起到了区分像素“n”和与“n”有相似颜色的背景区域的作用。因此，在去除文字像素的时候，除了要精确地去除这些文字像素外，还要把文字的边缘黑色像素考虑进去，即也要将它们去除。因此，本文在采用八叉树颜色量化算法(见第二章)对提取出的文字区域(如图4.2(b))进行二值化处理后(如图4.2(c))，运用形态处理的方法，

即采用3×3的结构元素B=$\begin{Bmatrix}1,1,1\\1,1,1\\1,1,1\end{Bmatrix}$，膨胀已提取出来的文字像素，直到完全包含文字的所有边缘像素(如图4.2(d))为止(根据实验，只需执行一次上述的膨胀就可包含文字的所有边缘像素)，这样才能保证可以精确地去除图像中的所有文字信息，完全地修复被文字遮挡的背景区域。

(a)原图

(b)提取出的文字区域

(c)二值化处理

(d)膨胀，形成的较大的待修复区域

(e)纯绿色的待修复区域

图 4.2 文字区域的后处理

4.2 基于纹理块匹配的图像修复技术

图像修复技术(inpainting)是当前计算机图形学和计算机视觉中的一个研究热点，在文物保护、影视特技制作、虚拟现实、多余物体剔除(如图像中删除部分文字、小标题、人物等)等方面有着重大的应用价值。

Bertalmio 等人在 2000 年提出一种基于高阶偏微分方程的算法[5]来修复用户指定的区域(如文字区域)；Chan 等人在 2001 年提出了基于 Total Variationa 模型[26]和基于曲率的扩散模型(curvature-driven diffusion)[35]的修复算法。这些算法主要是利用边界信息向待修复区域内进行各向异性的迭代扩散，其缺点是计算量大，在修复区域较小时有较好的效果。Chang Woo Lee[36,37]等人提出利用相邻帧的信息来填充待修复帧，这对于持续多帧不变的字幕显然不合适。Criminisi[27]等人提出基于纹理生成的修复方法，在待修复区域的的边界通过块匹配的方式选择合适的纹理填充。这种方法计算量小，对较大区域有好的修复效果。考虑到图像中文字区域较大，以及计算量的因素，本文采用纹理修复算法进行图像修复。

在修复图像前，本文将形成的待修复区域映射到原彩色图像中，并将所要修复的区域用一种颜色表示(文中采用纯绿色，如图 4.2(e))，以示区别。然后，使用 Criminisi[27]等人提出的纹理修复算法进行图像修复。

如图 4.3 所示，设整幅图像为 I，待修复的目标区域表示为 Ω，区域的边界为 $\delta\Omega$，Φ 为视频中的非修复区域，Ψ_p 是以像素点 p 为中心点的待修复正方形模板。

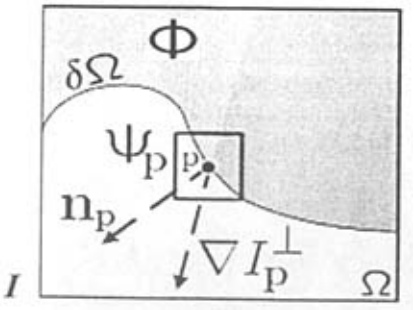


图 4.3　纹理修复示意图

首先，修复算法需要计算出Ψ_p的优先级$P(p)$，其定义如下：

$$P(p)=C(p)*D(p) \tag{4.1}$$

其中，$C(p)$称为待修复模板Ψ_p的置信度，$D(p)$为Ψ_p的数据信息项，代表p处的等值线与待修复区域边界$\partial\Omega$碰撞的强度。$C(p)$的定义如下所示：

$$C(p)=\frac{\sum_{q\in\Psi_p\cap\bar{\Omega}}C(q)}{|\Psi_p|} \tag{4.2}$$

其中，q为模板Ψ_p中无需修复的像素，$|\Psi_p|$是模板Ψ_p的面积。并规定$C(p)$函数的初始值为：

$$C(p)=\begin{cases}0 & \forall p\in\Omega\\ 1 & \forall p\in I-\Omega\end{cases} \tag{4.3}$$

$D(p)$的定义如下：

$$D(p)=\frac{|\nabla I_p^{\perp}\cdot n_p|}{\alpha} \tag{4.4}$$

$\nabla I_p^{\perp}$是p处的等值线(包括方向和强度)，n_p是点p在待修复区域边界$\partial\Omega$的法向量。此后，依次求出以待修复区域边界上的各像素点为中心的模板的优先级，从而得到优先级最大、最先修复的模板$\Psi_{\hat{p}}$。用公式表示为：

$$\Psi_{\hat{p}}\mid\hat{p}=\arg\max_{p\in\partial\Omega}P(p) \tag{4.5}$$

然后，从非修复区域Φ中找出与待修复模板$\Psi_{\hat{p}}$最相似的匹配模板$\Psi_{\hat{q}}$，用

公式表示为：

$$\Psi_{\hat{q}} \mid \hat{q} = \arg\min_{\Psi_q \in \Phi} d\left(\Psi_{\hat{p}}, \Psi_q\right) \tag{4.6}$$

其中，距离$d\left(\Psi_{\hat{p}}, \Psi_q\right)$定义为两个模板中无需修复的像素之间的差的绝对值之和。最后，将匹配模板$\Psi_{\hat{q}}$覆盖待修复模板$\Psi_{\hat{p}}$的区域，从而实现了图像的一次修复。为了完全地修复图像，需要不断重复以上操作，直至图像中不存在待修复区域(文中表示为绿色区域)。

4.3 实验结果与比较

基于上述技术，本文在Windows2000环境下使用Visual C++6.0实现了图像内文字的去除，以图4.2(a)为例，修复过程中得到的结果如图4.4所示。

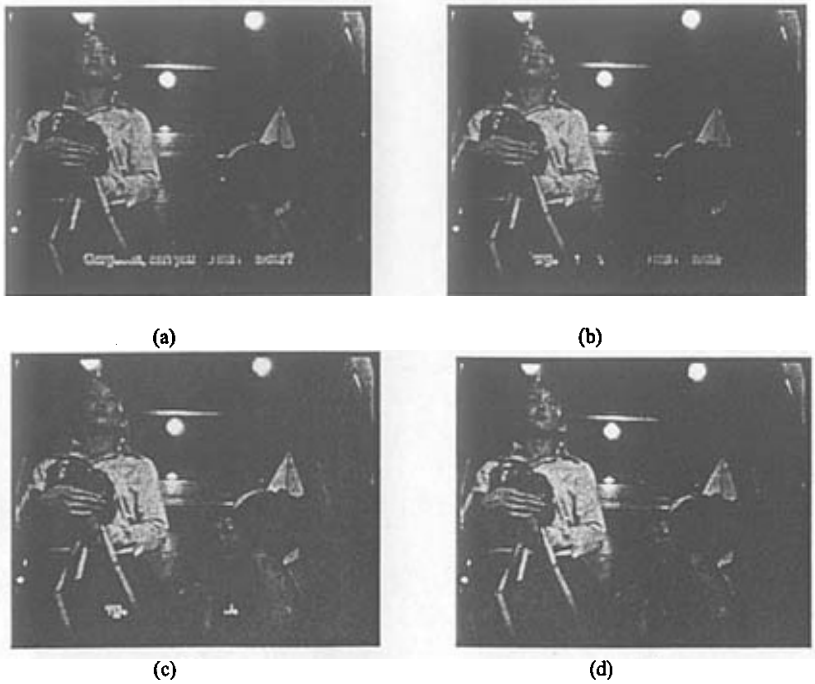


图4.4 修复过程中的结果

通过目视图4.4(d)，可以清楚地看出，原来图像中被文字遮挡的背景

都被完全地修复了，特别是图中被文字所遮挡的属于包带的那部分区域(土黄色)被完好地修复了。因此，本文方法很好地完成了图像修复的目标。图4.5给出了Chang Woo Lee方法和本文方法的一个比较。

(a)原图

(b) Chang Woo Lee方法修复的结果

(c)本文方法修复的结果

图4.5 两种方法的修复结果

从上图中可看出，本文方法与Chang Woo Lee方法的修复结果相差无几。但是，Chang是利用相邻帧的信息来填充待修复帧，这对于持续多帧

不变的文字显然不合适，而且计算量大。而本文提出先用基于CEMA的算法检测出图像内的文字信息，然后结合算法简单、计算量小的纹理修复方法，很好地去除了文字信息，恢复了背景。所以本文方法要优于Chang Woo Lee提出的方法。

同时，为了体现本文方法的鲁棒性，我们也对图像内的倾斜文字做了一个实验。如图4.6所示。

(a)原彩色图像(未加入文字)

(b)人工加入倾斜文字后的图像

(c)修复(去除文字)后的图像

图4.6 倾斜文字的去除

通过目视，发现修复后的图像(图4.6(c))与未加入文字的原有图像(图4.6(a))基本一致。因此，本文方法较好地完成了图像修复的目标。

结 论

本文详细阐述了文字提取与文字去除的处理流程：彩色边缘检测、二值化、形态处理、文字区域提取、图像修复；并解释了各个环节采取的策略。本文主要完成的工作包括：

(1)文字提取—在分析了文献[12]的基础上，利用彩色图像文字区域与背景对比度较大、存在明显的边缘轮廓的特点，本文提出了一个新的文字提取算法—CEMA 算法。该算法分别运用垂直、水平和对角三个方向的彩色边缘检测算子从原图中提取出相对应的边缘图像，然后对上述三幅边缘图像依次运用形态学中的闭、开、水平膨胀、水平腐蚀运算，得到三幅不同的连通域图，最后，将这三幅连通域图进行逻辑与运算，去除噪声，得到最终的文字区域。实验结果证明，CEMA 算法简单有效，文字提取率高，且具有鲁棒性。该算法适用于静态视频帧字幕提取、广告信息提取等多种应用领域。

(2)文字去除(图像修复)—在提取出图像内的文字区域后，本文运用计算量小、而且算法简单的纹理修复技术，将提取出的文字从原图中去除，同时，修复原图中被文字所遮挡的背景区域。实验表明，该方法能很好地去除图像内的文字信息。

由于作者学识、技能和时间有限，本文实现的文字提取与文字去除(图像修复)还有待于进一步的提高和完善。进一步的工作可以从以下几个方面考虑：

(1)由于文字信息提取时需要进行图像的二值化，所以对阈值的依赖非常强，阈值的确定直接影响到文字信息的提取，特别是当同一行文字区域中字符颜色不同时，对阈值的要求就更加高。因此，自适应的、有效的、智能的阈值选取方法需要进一步研究。

(2)本文处理的彩色图像都是静止的，即图像文字和图像背景都是静止的，所以可以考虑如何将本文方法推广到静止的文字覆在运动的背景上或者运动的文字覆在静止和运动的背景上，因为以上几种情况经常出现在视频图像中。

(3)进一步改进图像修复方法，可将纹理修复方法和其它修复算法综合起来，以便更快、更好地修复图像或者视频帧。

(4)文字提取与去除系统的实用化还需要进一步的研究，如多种方法的综合运用、合适的计算复杂度、系统处理过程的自动化程度等。

参考文献

[1] A.K.Jain,Bin Yu. Automatic Text Location in Images and Video Frames[J]. Pattern Recognition,1998,31(12):2055-2076.

[2] C.Strothopoulos,N.Papamarkos,A.E.Atsalakis.Text extraction in complex color documents[J]. Pattern Recognition ,2002,35:1743-1758.

[3] Edward K.Wong,Chen Minya. A new robust algorithm for video text extraction[J]. Pattern Recognition,2003,36:1397-1406.

[4] Keechul Jung,Kwang In Kim,Anil K.Jain. Text information extraction in images and video: a Survey[J]. Pattern Recognition,2004,37:977-997.

[5] Bertaimin M, Sapiro G, Caselles V, et al. Image inpainting [A]. In: Proceedings of SIGGRAPH 2000[C],New Orleans,USA, 2000:417- 424.

[6] Li Hui-ping, Kia O,Doermann D. Text enhancement in digital Videos[A]. In:Proceedings of SPIE99-Document Recognition And Retrieval[C], SanJose, CA,USA,January,1999:1-8.

[7] Hori O. A video text extraction method for character recogniton[A]. In:Proceedings of 5th International Conference Document Analysis and Recognition(ICDAR1999)[C],Bangalore,India,1999:25-28.

[8] Wu V,Manmatha R,Riseman E. TextFinder:An automatic system to detect and recognize text in images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1999,21(11):1224-1229.

[9] Hua Xian-sheng,Chen Xiang-rong,Liu Wen-yin,et al. Automatic location of text in Video frames[A]. In:3rd International Workshop on Multimedia Information Retrieval(MIR2001)[C],Ottawa,Canada,2001:126-129.

[10] Wu V,Manmatha R,Riseman E. Finding text in images[A].In:Proceedings of 2nd ACM International Conference Digital Libraries[C], Philadelphia, PA, USA,1997:23-26.

[11] Wu V,Manmatha R. Document image clean up and binarization [A]. In:Proceedings of SPIE Symosium on Electronic Imaging 1998[C], SanJose, CA,USA,January,1998:263-273.

[12] Li Hui-ping,Doermann D,Kia O. Automatic text detection and tracking in digital Video[R].LAMP Technology Report 028, Maryland University, USA,1998.

[13] Zhong Y,Karu K,Jain A K. Locating text in complex color images[J]. Pattern Recognition,1995,28(10):1523-1536.

[14] Jain A.K,Yu B. Automatic text location in images and Video frames[J]. Pattern Recognition,1998,31(12):2055-2076.

[15] Jain A K,Sushil Bhattacharjee. Text Segmentation using gabor filters for automatic document processing[J]. Machine Vision and Applications, 1992, 5(3):169-184.

[16] Lienhart R, Effelsberg W. Automatic text segmentation and text recognition for video indexing[J]. Multimedia System ,2000,8(2):69-81.

[17] Wernicke A. Text localization and text segmentation in images,videos and web pages[D]. M.S. thesis,University of Mannheim,Mannheim, Germany, Mar.2000.

[18] Smith M A,Kanade T. Video skimming for quick browsing based on audio and image characterization[R]. Technology Report CMU-CS- 95-186, Carnegie Mellon University,Pittsburgh,PA, USA,July 1995.

[19] Tan C L. Text extraction using Pyramid[J]. Pattern Recognition, 1998, 31(1):63-72.

[20] Lienhart R. Automatic text recognition for video indexing [A]. In:Proceedings of ACM Multimedia 96[C],Boston,MA,USA,1996:11-20.

[21] 庄越挺,刘骏伟,吴飞等. 基于支持向量机的视频字幕自动定位与提取[J]. 计算机辅助设计与图形学学报,2002,14(8):750-7 53.

[22] Chen Da-tong,Bourland Herve,Thiran Jean-Philippe. Text identification in complex background using SVM[A]. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition 2001[C], Kauai Marriott, Hawaaii,USA,2001,2:621-626.

[23] Bertalmio M,Bertozzi A L,Sapiro G,et al. Navier stokes, fluid dynamics,and image and video inpainting[A]. In:Pro ceedings of the International Conference on Computer Vision and Pattern Recognition [C], Kauai,HI,2001,1:355-362.

[24] Gomes J,Velho L. Image processing for computer graphics [M]. New

York:Springer-Verlag,1997.

[25] Perona P,Malik J. Scale-space and edge detection using anisotropic diffusion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1990,12(7):629-639.

[26] Chan T, Shen J. Mathematical Models for Local Deterministic Inpaintings [R].TR00-11,Department of Mathematics,UniveRsity of California-Los Angeies,Los Angeles,California,USA,2000.

[27] Criminisi A,Perez P,Toyama K. Object removal by exemplarbased inpainting[A].In:Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern RecogNition[C], Madison, Wisconsin, 2003,2:721-728.

[28] 郎锐. 数字图像处理学. 北京：希望电子出版社，2003.

[29] 周长发. 精通 Visual C++图像处理编程. 北京：电子工业出版社，2004.

[30] 何斌,马天予,王运坚等. Visual C++数字图像处理. 北京：人民邮电出版社,2001.

[31] 崔屹. 图像处理与分析. 北京：科学出版社，2000.

[32] 张引,潘云鹤. 面向彩色图像和视频的文本提取新方法[J]. 计算机辅助设计与图形学学报,2002,14(1):36-40.

[33] 徐慧, Visual C++数字图像实用工程案例精选,北京：人民邮电出版社,2004.

[34] Keechul Jung,Jung Hyun Han. Hybrid approach to efficient text extraction in complex color images[J]. Pattern Reco gnition Letters, 2004, 25:679-699.

[35] Chan T,Shen J. Non-Texture Inpainting by Curvature-Driven Diffusion [R], TR00-35,Department of Mathematics,Univers ity of California-Los Angeles,Los Angeles. California, USA,2000.

[36] Chang Woo Lee, etc. Spatiotemporal Restoration of Regions Occluded by Text in Video Sequences[C]. In:Proceedings of IDEAL 2003, Springer-Verlag, Berlin, Heidelberg.

[37] Chang Woo Lee,Keechul Jung,Hang Joon Kim. Automatic text detection and removal in video sequences[J]. Pattern Recognition Letters,2003, 24: 2607-2623.

攻读硕士学位期间公开发表的论文

1. 季丽琴,王加俊.视频图像内文字的自动提取新方法，苏州大学学报(自然科学版),已录用.
2. 王加俊,季丽琴.视频字幕的自动检测与去除，已投《中国图象图形学报》.

致 谢

我的毕业论文是在导师王加俊教授的指导下完成的。从论文的选题到实验模拟到论文的完稿，王老师都给予了精心的指导。回顾这三年来的学习和研究工作，无论在生活上还是学习上导师都给了我大量的支持和帮助，特别是导师勤勉的作风和严谨的治学态度，使我终生受益。在此，谨向王老师表示衷心的感谢。

感谢浙江大学CAD&CG国家重点实验室的汤锋曾经给予过的帮助。另外要感谢研究生学习期间对我的学习和生活上给予过帮助的身边的同学们。他们是：金永明、张庆峰、徐俊、伊怀峰、苟博等人。

此外我还得到了有关院系领导和老师的关怀和帮助，包括黄贤武老师、杨德生老师、季锦诚老师，在此向他们表示感谢。

最后我要感谢我的父母和亲人对我自始至终的支持，无论是精神上还是物质上我都得到他们给予的支持和帮助，在此向他们表示崇高的谢意。