# 第二章 复杂条件下的文本图像预处理

## 2.1 引言

相对与印刷体的光学字符识别，复杂条件下的文本图像有着截然不同的性质如分辨率低，大小不一，光照不均等。这些特性影响着从文本定位，分割到识别等各个过程。在将场景条件下的文本图像输入到各个模块前，对图像进行必要的预处理对于定位和识别的正确率提高有一定的帮助。对于初始的场景文本图像进行颜色空间变换使得图像更加适合于后续的聚类和显著性检测过程。得到文本词图像后，文本行可能不是以直线方式存在的，将词图像校正到水平位置有利于后续的词图像分割。而文本图像由于相机，场景的因素存在不同的噪声，去噪对于定位和识别的过程也相当重要。二值化则贯穿在文本定位和识别的过程，好的二值化能够较大提高识别结果的正确性。

## 2.2 复杂条件下的文本图像分析

复杂条件下的文本图像分析是进行后续研究的第一步，只有深刻认识了场景文本图像的一些特点，才能够提出有效的方法。我们从两个方面来对复杂条件下的文本图像进行分析。

### 2.2.1 场景分析

场景的复杂性是造成文本定位召回率和准确率低的一个重要原因。场景文本图像的获取一般是通过智能手机或者数码相机得到的，获取的途径存在多样性造成场景的多样性。这个多样性表现在场景可以是街道（室外），教室（室内），走廊，路边，玻璃窗户，书本封面等。场景多样性会导致场景干扰的多样性，场景的干扰如窗户，玻璃，草地，砖头，天空，栅栏等。有些干扰如果有和文本相同的边缘或者纹理，就会对文本定位的性能产生很大的影响。另一个场景的问题是场景存在着光照不均的情况，这个现象在图书封面场景中经常发生，因为拍照时背景光滑反射大部分光线。场景的多样性，干扰和光照不均的情况在图中可以看到

 

### 2.2.2 场景文本分析

在上面分析了场景的特性之后，这里分析场景图像文本的特性。场景文本是本课题研究的重点对象，场景文本不同于光学字符的方面在于文本的尺寸，颜色，光照和空间位置。

文本尺寸受拍摄远近和文本本身的影响，一般商店的名称，广告的尺度都比较大，可能占据的空间会超过整幅图像一半的大小，而一些物体的说明，一些告示，书本的封面通常比较小，占据整幅图像很小的部分。

文本颜色通常跟背景的差异都比较大，这样比较容易引起人们的注意。一个文本区域的颜色一般趋向于一致，但也存在不同颜色字符构成一个文本区域的情况，另外一个字符连通域内的颜色也可能存在多个。

光照不均的问题同样存在场景文本中，场景文本如果表面反射光线比较多的话，相应的区域会呈现比较亮，如果是在室外拍摄，太阳光照，遮挡会使文本呈现明亮不均的情况。

空间位置的变化，如果拍摄场景图像时，镜头对准了文本，文本一般会出现图像的中心。但是拍摄时一般没有考虑那么多因素，文本可能出现在图像的中心，边缘，上下左右任何一个位置。场景文本变换的例子在图2.2中可以看出。

 

## 2.3 颜色空间变换

颜色空间是用来对颜色进行表示。不同颜色空间用于不同的目的如用来减少颜色的数量，适应人眼对颜色的感知。不同的颜色空间可以进行相互转换。下面对两种类型的颜色空间进行阐述并对颜色空间进行转换。

### 2.3.1 设备相关颜色空间

设备相关的颜色空间最常使用的是RGB（Red,Green and Blue）和CMYK（Cyan,Magenta,Yellow and Black），RGB应用在于显示器，采用的是加色模型，RGB空间中的所有颜色都可以看作是三个基本颜色红，绿，蓝三原色的组合而成的。大部分其他颜色空间都是从RGB空间转换而来的。CMYK应用在于打印机设备上，采用的是减色模型。CMYK中的CMY分量都是从白光中减去青、品红和黄而产生的，而K是为了打印出黑色分量而加入的。和设备相关的颜色空间与人类的感知系统有一定的差距，在面对那些需要人类的主观认识的应用中，就需要提出新的颜色空间。

### 2.3.2 设备无关的颜色空间

这类颜色空间又可以分为两种一种为CIE(International Commission on Illumination)制定标准的颜色空间如CIE XYZ，CIE Lab等，另一类为常用的HIS（Hue, Saturation and Intensity），HSV(Hue, Saturation and Value)等。在1931年，CIE通过研究人类颜色感知而提出的一种标准即为CIE XYZ。这个标准定义了称为三重刺激值的三维空间来表示颜色。随后，经过几十年的发展，CIE提出了几种比XYZ更加适合于某些目的的颜色空间，这其中经常用到的是Lab颜色空间，这一感知均匀的颜色空间能够和颜色的视觉外观相关联。HSV相比RGB更能体现出人类感知颜色的方式。H的值从小到大对应的是颜色从红到黄，绿，在回到红。S代表的是颜色的纯度，S的值从小到大对应颜色的饱和度的变化。V值代表的是颜色的亮度，从小到大对应的颜色就越来越亮。

### 2.3.3 颜色空间的变换

在场景文本的定位和识别中，最经常用到的颜色空间变换是将RGB空间转换到CIE Lab空间比如在计算视觉显著性和进行颜色聚类的时候。Lab空间最接近于人类的视觉，它使用L部分来调整亮度对比度并用a，b部分来进行精确的颜色平衡更正。RGB空间无法直接转换到Lab空间，需要通过CIE XYZ来进行中间转换，RGB到XYZ空间的转换如下式



XYZ空间转换到Lab空间



其中

## 2.4 文本图像倾斜校正

文档图像的倾斜校正一般是在文本定位后进行。定位后得到词图像如果没有处于水平直线的位置，在后续采用投影图像进行分割的时候，容易造成切分出错。所以在把词图像放到分割模块前，先对词图像进行倾斜校正。对词图像进行倾斜校正的关键地方是找出词图像的倾斜基线。倾斜基线的查找方式有两种，分别是Hough变换和Randon变换。

### 2.4.1 Hough变换

Hough变换的主要思想是通过变换空间域来分析。图像像素点看成是X,Y平面上的点，那么平面上的两个点可以构成一条直线，直线有斜率和截距，利用斜率和截距作为新空间的坐标。根据 知道X,Y 平面上一点对应ρ,θ平面的一条正弦曲线，这样X,Y平面就可以转换成ρ,θ平面进行分析。在ρ,θ平面设置一些很小的累加器单元，在这个累加器单元里的ρ,θ点比较接近，相当于变换到X,Y平面上的一条直线。Hough变换了解决噪声，不均匀照明而产生的边缘间断。在实验中，先对图像进行二值化，然后计算图像的canny边缘，图像经Hough后得到的是一个参数空间的矩阵，矩阵的每一个值是对应ρ,θ坐标处累加器的值，在得到Hough变换的矩阵后，接着检测矩阵中符合要求有可能是直线的参数点。得到峰值后就要确定每个峰值所确定的线段，遍历所有线段并显示出来，找到其中最长的一条。用这一条确定斜率，由斜率得到角度，然后进行旋转就可。Hough变换倾斜校正的效果如图



### 2.4.2 Randon变换

Randon变换最早是从1917年提出来的，被广泛的应用在断层扫描中，用来从物体的各个方向的投影的数据中重构物体的图像。给定一系列的角度对一副图像进行Randon变换可以看作是在某些给定的角度上计算图像的投影。投影的结果是每个方向上像素亮度值的和即线积分。Randon变换可用下式表示



其中 为变换后的图像，为原始图像。类似于Hough变换，Randon变换将图像从笛卡尔坐标映射到极坐标。Randon变换检测直线的原理是当图像中存在一条直线时，投影的光束沿着直线方向进行投影，这时计算的线积分最大。所以当沿着不同方向进行投影时，就会得到一系列和方向对应的投影值，投影值最大的地方最有可能存在直线。在实验中，对于输入的文本词图像先灰度化在二值化，同样的计算二值图像的canny边缘。对于得到的边缘图像，投影直线角度从0到179变化，每个角度计算一次投影值。取最大投影值的角度为倾斜角度，即是词图像直线基线。



## 2.5 文本图像二值化

文本图像的二值化是一个精简信息的过程，在整个文本定位和识别系统占据重要地位。文本图像的二值化是将文本区域赋值为1，而将其余区域赋值为0。在分割时，不需要太多的文本区域内部的信息，只需要字符间的相对位置信息，所以一般分割的算法都是在二值化后的图像进行的。提取单个字符的全局特征时，字符内部局部的，细节的信息不是很关键，只需要保留有字符的整体，形状的特征即可，所以一般也是在二值化后进行提取。

二值化算法包括全局二值化和局部二值化, 全局二值化具有速度快但效果相对差的特点, 局部二值化算法具有速度慢效果好的特点。

### 2.5.1 全局阈值

全局阈值不考虑灰度图像的局部特性，整个图像都采用同一个阈值，像素灰度大于某个阈值的元素赋值为1，小于某个像素的为0。全局阈值的设定可以通过人工经验也可以通过迭代法进行计算。迭代法计算全局阈值，主要思想是先求图像的全局平均值作为初始阈值，然后利用这一阈值将图像分为两部分，求每部分的均值，同样利用这一均值分割图像，不断迭代。

另一个比较重要的全局阈值的二值化算法是Otsu算法【】。这一算法假设图像存在两个类的像素即前景像素和背景像素，通过不断搜索阈值从1到最大亮度，分割图像计算每个类的类间方差和两个类的类间方差，使类间方差最大的阈值就是最后二值化所有的阈值。还一些算法将重点放在改善图像的直方图特性上，如【】采用非线性对比度的方式增强图像的对比度，这样一来一些原本混入到背景的前景像素就被映射到前景上来了。

### 2.5.2 局部阈值

局部阈值的运算时间相对较长，这一类型的方法阈值的确定考虑的是像素某个邻域。一般的局部阈值算法的流程是先将图像进行分块，在每一块内部，可以使用简单的迭代或者最大类间差的来确定局部块的阈值。局部阈值的方法对最终的效果有较大提高，相应的处理时间就变的比较长了。

另一种比较常见的局部二值化算法为Bernsen 算法，这一算法对于不均匀的光照尤其是阴影去除具有很好的效果。这一算法的主要流程是首先计算当前像素P的窗口大小为W的所有像素的最大值M和最小值N，并计算两者的均值T。设定一个阈值S，若M-N大于S的话，当前像素P的阈值就为T，否则说明这一窗口的像素的灰度级别相差比较小，要判断窗口所处区域。设定第二个阈值T1，如果窗口的T值比这一阈值大，那么就判定窗口处在前景区域，当前的像素值P可以设定为1，否则的话为0。在实验中S设为15， T1设为20。

在实验中，对一张文本词图像进行5种二值化方法比较，从图中可以看出局部阈值的方法要优于全局阈值。

## 2.6 文本图像去噪

噪声存在场景文本图像的各个部分，其有可能是拍摄时引入，也有可能是在对图像进行处理的过程中引入的。这里的去噪一般是在词图像分割之前和对单个的字符进行处理。分割之前是为了不让噪声影响切分，而对单个字符的去噪是为了在特征提取阶段不生成无效的表征。下面从两个方面对文本图像中的噪声进行说明分别为噪声分类和去噪方法。

### 2.6.1 噪声分类

噪声的幅度的变化范围可以从良好光照条件下拍摄的图像上无法觉察的斑点到几乎布满整幅图像的射电天文学噪声。这里只研究简单加性噪声模型即



其中 为噪声图像， 为原始图像， 为随机噪声。

其中图像中噪声的典型的模型是高斯模型。高斯模型的噪声独立于每个像素和像素亮度，符合是零均值的高斯分布如下式。



另一个比较常见的噪声模型是脉冲噪声模型，这一噪声模型仅有两个可能的值，并且噪声在图像中出现的概率不能超过0.1。脉冲噪声一般是由相机的传感器单元异常，存储单元失败或者是图像数字化和传输错误引起的。

### 2.6.2 去噪方法

每类噪声都有相对应于各个噪声的去噪方法，在场景文本图像中，比较常见的噪声是高斯噪声和脉冲噪声。统计排序滤波器比较适用于处理脉冲噪声的。它的原理是在一个窗口中对所有的像素按照灰度值大小进行排序，然后把窗口中间的像素替换为排序序列的某一顺序值。依据要替代窗口中像素值的像素所处的位置可以分为中值滤波器，最大值滤波器和最小值滤波器。中值滤波器能够有效的去除椒盐噪声。

对于高斯噪声而言，一般是采用线性滤波器进行滤除。这些滤波器首先定义一个窗口大小的模板，模板上设定好掩模系数，然后将模板不断的滑过窗口进行卷积运算得到滤波后的图像。

在实验中，文本图像的去噪主要是对裁剪后的单个字符进行的，主要采用是中值滤波器和时域高斯低通滤波器，其中中值滤波器的窗口大小设置为5，同样高斯滤波器的窗口也设置为 5。





## 2.7 本章小结

本章首先对复杂条件下的文本图像从场景和文本两个进行分析，得出复杂条件下的文本图像的难点和需要重点关注的问题如干扰，光照不均，倾斜等，另外也说明了预处理步骤对于文本图像后续的定位和识别相当重要。这对这些重点问题本章从颜色空间，倾斜校正，去噪和二值化四个方面来对文本图像进行预处理。这些预处理步骤贯穿于整个定位和识别的过程，首先颜色空间的变换是对整幅自然场景文本图像进行处理，颜色空间的变换有利于聚类和视觉显著性提取等过程。接着在得到文本词图像后，对于文本词倾斜的情况进行校正，分别采用了两种方式来进行。Randon变换的步骤较少也能够取得不错的效果。将文本词图像校正后，就需要进行二值化精简信息为分割，提取特征做准备，本章探讨了5种常用的二值化方法并进行了比较。分割得到字符图像后或者在词图像分割前有必要对图像进行简单的去噪，本章仅分析两种常见噪声椒盐噪声和高斯噪声，并用简单的方法进行处理，实验表明去噪对图像质量有一定的改善。