**移动应用后端服务框架的设计和实现**

**鞠强**

# 摘 要

随着移动互联网技术的崛起，继SaaS（软件即服务：Software as a Service）、IaaS（基础设施即服务：Infrastructure as a Service）和PaaS（平台即服务：Platform as a Service）之后，BaaS（后端即服务：Backend as a Service）生态系统正从一个小众垂直领域迅速成为非常重要的行业环节。本文从理论和实践两个角度出发，研究native的移动开发（以IOS、Android为例）过程中存在的主要问题，学习后端即服务的产业现状和实现技术，并且设计和实现一个简单的移动应用后端服务框架。

本文首先较为详尽地介绍了在本系统中运用到的知识背景。接着对Android和IOS应用开发中数据存储传输方面存在的共同问题进行探讨，对其解决之道“后端即服务生态系统”的现状进行研究。而后对设计和实现一个简单的移动应用后端服务框架的可行性进行了分析，同时简要介绍了系统中所用到的开发语言Ruby和数据库Mongodb的特点以及使用方法。对系统的设计思想、设计目标与系统的整体结构进行了明确的规划。最后对框架的设计与实现作了较为详细的讲解。

系统涉及的主要技术有Ruby on Rails、Mongodb、Java + Android SDK以及Objective-c + Cocoa SDK。其主要功能有：移动应用数据远程存储和本地缓存服务，一站式的移动应用后台搭建，消息推送服务，简单的统计服务。

论文在撰写过程中，力求将理论与实践应用相结合，提高系统的可用性和通用性，希望通过本文的论述能够更充分地体现出后端即服务的设计思想以及上述理论和技术在本系统中的应用。

**关键字**：移动、框架、后端即服务

# Abstract

As the rapid development of mobile Internet, following the SaaS (Software as a Service), IaaS (Infrastructure as a Service) and PaaS (Platform as a Service) BaaS (Backend as a service) ecosystem from a vertical niche areas quickly become a very important industry links. From both theoretical and practical perspectives, we will study the main problems in native mobile development (IOS, Android, for example) process, learning the backend that service the industry status quo and implementation techniques, and design and implement a simple back-end services framework for mobile application.

We will firstly introduce the background knowledge in more detail used in this article. Then we will discuss common problems to Android and IOS application development in data storage transmission to study the status of the "(Backend as a service" of its solution. After that，we’d like to analyze the feasibility of the design and implementation of a simple mobile application framework for back-end services, while brief introduction to the system used in the development of language Ruby and database Mongodb characteristics and usage. Clear planning system design thinking, design goals and overall system structure. Finally, the design of the framework and made ​​more detailed explanation.

The system involves technologies are Ruby on Rails, Mongodb, Java + Android SDK, and Objective-c + Cocoa SDK. Its main function: remote data storage and local caching services of mobile applications, building a one-stop mobile application background, message push service, simple statistical services.

We strive to make the system be the combination of theory and practical and has high usability and versatility. I hope to be able to more fully reflect the the backend that service design thinking and the theory and technology through the discussion of this paper in the system application

**Keywords**: mobile, framework, backend as a service.

# 目 录

摘 要 1

Abstract 2

目 录 3

第一章 绪 论 4

1.1 课题背景 4

1.2 文档图像二值化的发展 5

1.3 文档图像二值化的应用领域 5

1.4 文档图像二值化的技术难点 5

1.5 本文的组织结构和创新点 5

第二章 技术背景及相关知识简介 6

2.1 开发环境 6

2.1.1 C++语言简介 6

2.1.2 Visual C++ 6.0简介 6

2.2 相关知识介绍[7] 7

2.2.1 调色板 7

2.2.2 GDI位图 8

2.2.3 设备无关位图DIB 8

第三章 文档图像二值化方法 8

3.1 文档图像二值化的过程 8

3.2 文档图像二值化方法概述 9

3.3 现有的二值化方法分析 9

3.3.1 全局二值化方法 9

3.3.2 局部自适应二值化方法 13

3.4 二值化方法小结 16

第四章 两种新的文档图像二值化方法 16

4.1 本章概述 16

4.2 基于背景表面的文本图象二值化方法 17

4.2.1 粗略的估计前景区域 18

4.2.2 获取背景区域[2] 19

4.2.3 提取文字部分 19

4.2.4 后期处理[27] 20

4.2.5 实验结果 20

4.3 Otsu算法和Bernsen算法相结合的二值化方法 21

4.3.1 超限邻域平均法[28] 23

4.3.2 使用Otsu算法进行全局二值化 24

4.3.3 使用Bernsen算法进行局部二值化 24

4.3.4 后期处理 25

4.3.5 实验结果 25

第五章 几种图像二值化方法的比较 27

第六章 结论 29

6.1 两种新的文本图像二值化方法总结 29

6.2 文档图像二值化的研究现状及存在的问题 29

6.3 文档图像二值化的发展趋势 30

致 谢 30

参考文献 31

# 第一章 绪 论

## 1.1 课题背景

据《中国互联网络发展状况统计报告》显示，到2012年12月底，中国互联网用户规模达到5.64亿，其中，移动互联网用户达到4.2亿。截止2012年底，中国移动互联网用户数量为4.2亿，年增长率为18.1%，增幅超过整体互联网用户增长幅度。在所有互联网用户里，使用移动设备接入网络的用户比例由69.3%提升到74.5%。移动终端无疑已成为互联网入口的第一大终端。

随着移动互联网技术的发展和移动应用开发技术的成熟，移动应用开发过程中分工细化的趋势越来越明显，移动应用后端即服务生态系统正从一个小众垂直领域迅速成为非常重要的行业环节。

## 1.2 移动应用开发技术的发展

为了顺应移动互联网的快速发展，移动开发领域也经历了一代又一代的尝试。这些尝试包括最初个人英雄主义的本地应用开发、Flash、html5、widget、移动中间件等。

## 1.3 开发者面临的问题

经过了上述的实验和考证后，人们发现移动开发难度远比PC开发的难度高很多，《神庙逃亡》、《愤怒的小鸟》等神话毕竟是少数，小团队和独立开发者越来越难在移动应用生态系统中获得一席之地。从WAP、原生开发等各式各样的移动开发方式和平台在移动领域无攻而返时，人们开始思考到底什么样的开发方式更能高效、快捷适应移动产业的需要。

当Html5、Widget横空出世火速增援的时候，我们似乎看到了移动蓝海的救星，可是好景不长，Facebook放弃Html5另觅他路，让Html5的弱点暴露无疑，具备一定开发能力的厂商也发现Html5并非为移动开发而生，好多在移动中实现问题还有待解决，Widget也远没有如传说中的那样趟平整个移动领域，移动中间件则经过多年的锤炼一直拥有良好的口碑和解决问题的能力。可是需要学习第三方的语言和开发原理才能应用自如，学习曲线和难度之大使得开发者敬而远之，并且市场上主流的移动中间件产品在技术上采取全面封锁措施，对开发者采取封闭状态，同时更多的软件厂商也不愿把关系产品命脉的赌注，全部押给并不开放的移动中间件的手里，这也是目前移动中间件所面临的四面高歌却孤立冷清的局面。

## 1.4 “后端即服务”所提供的服务和解决方案

后端即服务的移动解决方案，让繁琐的移动开发过程更简单快捷，也使得移动开发领域分工更加明确，让android/IOS等移动开发者无需掌握任何一门服务器开发语言(JAVA、PHP)，就可以按照提供的android/IOS/WP三个平台的SDK内置的预定义功能接口快速开发移动应用。解决当前原生开发难以解决、耗时耗力的数据存储、文件存储、消息PUSH等工作。

## 1.5 本文的组织结构和创新点

本文共分为六章，具体内容结构安排如下:

第一章绪论部分。介绍论文的研究背景和研究意义，阐述了论文的大体结构。

第二章简单介绍了毕业设计的技术背景及相关知识。

第三章介绍当前后端即服务的发展情况。

第四章设计和实现了一个简单的移动应用后端服务框架

第五章讨论了设计框架的通用性和可用性

第六章总结。

# 第二章 技术背景及相关知识简介

## 2.1 开发环境

* 服务端使用Ruby和Node.js
* Android端使用Java和Android SDK
* IOS端使用Objective-c和Cocoa SDK
* 使用Mongodb进行数据存储
* 使用NetBeans、Xcode、Eclipse作为集成开发环境
* Windows和Mac下进行开发，服务端部署在Ubuntu的Linux上

### Ruby语言简介

Ruby，一种为简单快捷的[面向对象](http://baike.baidu.com/view/125370.htm)编程（[面向对象程序设计](http://baike.baidu.com/view/249254.htm)）而创的[脚本语言](http://baike.baidu.com/view/76320.htm)，在20世纪90年代由日本人[松本行弘](http://baike.baidu.com/view/1438214.htm)（まつもとゆきひろ/Yukihiro Matsumoto）开发，遵守GPL协议和Ruby License。它的灵感与特性来自于 Perl、Smalltalk、Eiffel、Ada 以及 Lisp 语言。由 Ruby 语言本身还发展出了JRuby（Java 平台）、IronRuby（.NET 平台）等其他平台的 Ruby 语言替代品。Ruby的作者于1993年2月24日开始编写Ruby，直至1995年12月才正式[公开发布](http://baike.baidu.com/view/602345.htm)于fj（[新闻组](http://baike.baidu.com/view/834.htm)）。因为Perl发音与6月诞生石pearl（珍珠）相同，因此Ruby以7月诞生石ruby（红宝石）命名。

在开发层面，Ruby具有以下优点：

* 语法简单
* 普通的[面向对象](http://baike.baidu.com/view/125370.htm)功能(类,方法调用等)
* 特殊的[面向对象](http://baike.baidu.com/view/125370.htm)功能(Mixin,特殊方法等)
* [操作符重载](http://baike.baidu.com/view/1033032.htm)
* 错误处理功能
* [迭代器](http://baike.baidu.com/view/925158.htm)和[闭包](http://baike.baidu.com/view/648413.htm)
* [垃圾回收](http://baike.baidu.com/view/159846.htm)
* 动态载入(取决于系统架构)
* 可移植性高.不仅可以运行在多数UNIX上,还可以运行在DOS,Windows,Mac,BeOS等平台上
* 适合于快速开发，一般开发效率是JAVA的5倍

### 2.1.2 Mongodb简介

MongoDB是一个基于分布式文件存储的数据库。由C++语言编写。旨在为WEB应用提供可扩展的高性能数据存储解决方案。

MongoDB是一个介于[关系数据库](http://baike.baidu.com/view/68348.htm" \t "_blank)和非关系数据库之间的产品，是非关系数据库当中功能最丰富，最像关系数据库的。他支持的数据结构非常松散，是类似json的bson格式，因此可以存储比较复杂的数据类型。Mongo最大的特点是他支持的查询语言非常强大，其语法有点类似于面向对象的查询语言，几乎可以实现类似关系数据库单表查询的绝大部分功能，而且还支持对数据建立[索引](http://baike.baidu.com/view/262241.htm" \t "_blank)

MongoDB具有以下特点：

* 面向集合存储，易存储对象类型的数据。
* 模式自由。
* 支持动态查询。
* 支持完全索引，包含内部对象。
* 支持查询。
* 支持复制和故障恢复。
* 使用高效的二进制数据存储，包括大型对象（如视频等）。
* 自动处理碎片，以支持云计算层次的扩展性。
* 支持[RUBY](http://baike.baidu.com/view/45135.htm)，[PYTHON](http://baike.baidu.com/view/21087.htm)，[JAVA](http://baike.baidu.com/view/29.htm)，[C++](http://baike.baidu.com/view/824.htm)，[PHP](http://baike.baidu.com/view/99.htm),[C#](http://baike.baidu.com/view/6590.htm)等多种语言。
* 文件存储格式为BSON（一种JSON的扩展）。
* 可通过[网络](http://baike.baidu.com/view/3487.htm)访问。

### 2.1.3 Android SDK

Android开发工具包，为Android手机开发人员提供手机开发的Java接口，从2009年的1.5版本至今，Android SDK已经发布了十数个版本。

### 2.1.4 Cocoa SDK

Cocoa是苹果公司为Mac OS X所创建的原生面向对象的[编程环境](http://baike.baidu.com/view/1405990.htm)，是Mac OS X上五大API之一（其它四个是Carbon、POSIX、X11和Java）。

苹果的[面向对象](http://baike.baidu.com/view/125370.htm)开发框架，用来生成 Mac OS X 的应用程序。主要的开发语言为 Objective-c, 一个c 的超集。 Cocoa 开始于1989年9月上市的NeXTSTEP 1.0，当时没有Foundation框架，只有动态[运行库](http://baike.baidu.com/view/1032404.htm)， 称为 kit, 最重要的是AppKit. 1993 年 NeXTSTEP 3.1 被移植到了 Intel, Sparc, HP 的平台上，Foundation 首次被加入，同时Sun 和 NeXT 合作开发OpenStep 也可以运行在Windows 系统上。

Cocoa 应用程序一般在苹果公司的开发工具Xcode（前身为Project Builder）和Interface Builder上用Objective-C写成。

# 第三章 移动开发后端即服务现状

## 3.1 移动开发分工细化催生后端即服务

开发一个需要联网的移动客户端应用是复杂的。开发者不仅需要掌握一门客户端语言（java、objective-c等），了解相应开发工具包（Android SDK、Cocoa SDK等）还需考虑服务器编程和数据库技术；除了需要关注用户交互、产品逻辑，还需分神于数据存储、网络交互、推送服务、数据统计和用户行为分析。随着移动开发分工的细化，以方便移动客户端开发部署为目的的移动应用后端即服务概念应运而生。

## 3.2 当前后端即服务的服务内容概述

当前已知已有二十多家国外公司以及几家国内公司提供移动应用后端服务，综合看来，提供的服务主要包括以下方面：

* 数据存储
* 推送服务
* 地理位置
* 社交整合
* 统计和用户行为分析

## 3.3 分析几个典型的移动后端服务厂商

### 3.3.1 StackMob

StackMob是最早进入移动应用后端服务领域的厂商，成立于2010年。

**（1） 平均灰度法[8]**

该算法的思想是：求整个图像各象素灰度的和，除以象素数，求出图像的平均灰度，并以这个灰度值作为阈值，把灰度图像转变成二值图像。该方法简单、易于理解、执行速度快，但是处理效果往往不是太好。

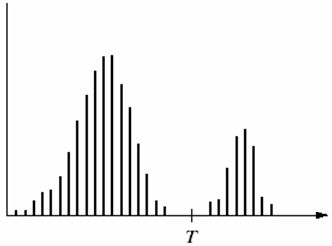
**（2） p-tile法[9]**

早在1962年，Doyle提出p-tile法(也称为p--分位数法)，可以说是最古老的一种阈值选取方法。比如，一幅文本图像，其中文字约占图像的75%,那么参照直方图选择阈值，使得直方图的75%的像素在阈值的一边。该方法使目标或背景的像素比例等于其先验概率来设定阈值，简单高效，但是对于先验概率难于估计的图像却无能为力。

**（3） 双峰法[10]**

1966年Prewitt等人提出的[11]。这是较为经典的阈值选取法。该方法是在文档图像的灰度直方图（如下图）基础上使用的，如果直方图中有两座明显的山峰（分别对应于背景部分和前景文字部分），那么可以选择这两座山峰中最低谷处作为阈值。

双峰法是通过查找双峰直方图的谷点确定分割阈值，因此对于目标与背景的灰度级有明显差别的图像, 当分割阈值位于谷底时, 图像分割可取得很好的效果。但是对于灰度变化不明显具有两个山峰的文档图像，这种方法是不适用的，即当前后景的对比较为强烈时，分割效果较好；否则基本无效。



像素个数

0

255 灰度值

双峰法的思想在以后的研究中常被人借用，但这种方法有相当大的局限性。譬如，对于灰度直方图呈单峰，双峰不明显，双峰间低谷比较宽广等情况，就很难有所作为了。

**（4） 迭代法[12]**

迭代法是基于逼近的思想，其步骤如下：

1. 求出图象的最大灰度值和最小灰度值，分别记为和，令初始阈值；
2. 根据阈值将图象分割为前景和背景，分别求出两者的平均灰度值和；
3. 求出新阈值；
4. 若=+1，则所得即为阈值；否则转2，迭代计算。

使用迭代法所得到的二值化图像效果良好。基于迭代的阈值能区分出图像的前景和背景的主要区域所在，但在图像的细微处还没有很好的区分度。但有时微小数据的变化却会引起分割效果的巨大改变，原因还有待进一步研究。

经实验比较，对于直方图双峰明显，谷底较深的图像，迭代方法可以较快地获得满意结果。但是对于直方图双峰不明显，或图像目标和背景比例差异悬殊，迭代法所选取的阈值不如最大类间方差法。

**（5） Otsu方法[13，14]**

Otsu(大津法)由大津于1979年提出，所以称为Otsu方法。它是在最小二乘法原理基础上推导出来的，这种方法的基本思想是将直方图在某一阈值处分割成两组，一组对应于背景部分，一组对应于前景文字部分，当被分成的两组的组内方差最小，组间方差最大时，决定阈值。这种基于两组间最佳分类而决定阈值的方法也称为最大类间方差法。

具体实现方法是如下，对图像Image，记为前景与背景的分割阈值，前景点数占图像比例为，平均灰度为；背景点数占图像比例为，平均灰度为。图像的总平均灰度为：



从最小灰度值到最大灰度值遍历，当使得下面的公式：



最大时即为分割的最佳阈值，公式（3）实际上就是类间方差值，阈值分割出的前景和背景两部分构成了整幅图像，而前景取值，概率为，背景取值，概率为，总均值为，根据方差的定义即得该式。因方差是灰度分布均匀性的一种度量，方差值越大，说明构成图像的两部分差别越大，当部分目标错分为背景或部分背景错分为目标都会导致两部分差别变小，因此使类间方差最大的分割意味着错分概率最小。

由于直接应用大津法计算量较大，因此在实现时通常采用了其等价的公式



这种方法，在直方图中具有两个波峰（存在一个波谷）的时候，将取波谷处灰度值。即使不存在波谷，也可以求出最佳分界阈值，所以这种方法经常被采用。实验表明，大津法选取出来的阈值非常理想，对各种情况的表现都较为良好。虽然它在很多情况下都不是最佳的分割，但分割质量通常都有一定的保障，可以说是最稳定的分割，并且由于它计算简单、稳定有效，所以大津法是一种较为通用的图像二值化算法[15]。

**（6） 最大熵法**

Pun使用Shannon关于熵的概念，在假设图像仅完全由其灰度分布直方图表示的基础上，定义了图像的熵，并用这一量度来实现目标和背景的分离。

1985年Kapur等人提出了最大熵阈值选取方法[16]。其基本思想是根据目标和背景两个概率分布来定义目标熵和背景熵,准则函数为目标熵和背景熵之和, 图像分割所选取的最佳阈值，应该使得分割后的图像所保留的信息量最大，即熵最大。目标和背景像素越均匀, 分割效果越好, 其熵值越大,所以最佳分割阈值选取在目标熵和背景熵之和的最大值处。

1989年Abutaleb[17]考虑到灰度的空间分布，将一维最大熵方法推广到二维的最大熵，但是还是依据计算判决函数极值来确定最佳阈值这一研究模式。

最大熵方法对不同目标大小和信噪比（SNR）的图像均能产生很好的分割效果。它受目标大小的影响小，可以用于小目标分割，但是它运算量大，运算速度比较慢[18]。

**（7） 最小误差法[19]**

1986年Kittler和Illingworth提出了最小误差法[20]。最小误差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。若图中只有目标物和背景，设图像中感兴趣的目标像素点的灰度和背景点的灰度都呈正态分布，它们的灰度级分布概率密度分别为,；灰度均值分别为,；灰度的方差为，。设目标像素点数占全图像素点数的百分比为，则背景点数占全图像素点数的百分比为。这样图像总的灰度级概率密度分布可用公式（5）来表示:



其中



把目标点错划为背景点的概率为：



把背景点错划为目标点的概率为：



总的错误概率：

为了使总的误差最小，令



带入公式（5）则有



解公式（8），求得即为最佳阈值。

最小误差法受目标大小和噪声影响小，对于小目标图像，例如目标与背景大小之比低于1：100的图像也有很好的处理效果。

### 3.3.2 局部自适应二值化方法

当图像中有阴影、光照不均匀、各处的对比度不同、突发噪声、背景灰度变化等，如果只用一个固定的全局阈值对整幅图像进行二值化处理，由于不能兼顾图像各处的情况而使效果受到影响。一种解决办法就是用与象素位置相关的一组阈值来对图像各部分分别进行分割。这种与坐标相关的阈值也叫动态阈值，或自适应阈值法。这类算法的时间复杂性和空间复杂性比较大，但是抗噪能力强，对一些用全局阈值不易分割的图像有较好的处理效果。以下就介绍几种局部自适应二值化方法。

**（1） 分块均值法**

将图像分为若干个方块，对任一块计算平均灰度值，将方块中的每一个像素的灰度值与平均灰度进行比较。若大于平均灰度值，则将该像素的灰度值置为255；反之则置为0。该方法简单、易于理解，但对于含有噪声等质量较低图像无法得到好的效果。

**（2） Bernsen方法[21]**

该方法的思想是：对每一个象素，取阈值为:



其中和是以为中心的正方形邻域中灰度级最低和最高的象素的灰度值。但如果对比度，那么这整个邻域中只包含一类，要么是字符，要么是背景。Trier和Taxt的研究表明，和是比较好的选择。

Bernsen算法是一种典型的局部阈值法，但局部阈值法也存在以下问题和缺点:实现速度较全局阈值法慢，不能保证字符笔划连通性，容易出现伪影现象(即在背景区域受噪声干扰得到笔划结果)等[22]。这是因为Bernsen算法以局部窗口内最大、最小值作为考察点的邻域，当考察窗内无目标点时，个别噪声点将引起阈值的突变，背景灰度的非均匀性也将影响局部阈值的变化，当考察窗内均为目标点时，局部阈值被拉伸，这样势必使得宏观上本应同类的部分象素:目标(或背景)被强行二值化为背景(或目标)，从而出现所谓的笔划断裂及伪影现象。

但是Bernsen方法可以与其他方法搭配且鲁棒性不错。

**（3） Niblack方法[23]**

这个方法的思想是根据局部平均值和局部标准差在图像中变动阈值。点处的阈值这样来计算：



其中和依次是的局部邻域的样本平均值和标准差。邻域的大小应当足够小以便保存局部细节，但同时也要足够大来抑制噪声。通常可以选择使用邻域。*k*值是用来调整决定多大的字符目标边界被作为给定目标的一部分。给出了分离的很好的字符目标。

但是，Niblack方法也有它不足的方面，那就是会产生连续大块的非目标黑色区域，噪声仍会遗留在处理后的图像中。我们通过后面的实验结果可以看到这一点。

**（4） Eikvil方法[24]**

该方法使用了两个同心的窗口，一个小窗口S和一个大的同心窗口L，S位于L的中心。让两个窗口沿着图像滑动，滑动的步长为S的边长（通常S取，L取）。使用Otsu方法对所有L内的象素计算阈值来把象素分成两类（文本类和背景类）。分别计算两类像素的灰度均值和。若（通常取15），则对S内的像素用阈值进行二值化；反之，则说明窗口L的像素可能同属于文字类或背景类，此时，比较和，并计算窗口S内的像素灰度均值，比较与、，把S内的像素归为与它相近的一类进行二值化。

**（5） Sauvola和Pietikainen方法[25]**

Sauvola和Pietikainen于2000年提出了一种针对文档图像二值化的方法。该方法的思想是：假设前景文字部分的像素灰度值接近0，背景像素的灰度值接近255，将文档图像划分为大小相等的矩形窗口，可以取大约10到20个像素宽，对每个小窗口用公式（11）计算阈值：



其中是局部邻域的样本平均值，是标准偏差的动力固定在128，是一个正常数（通常取0.5），该方法对于文档图像的二值化有比较好的处理效果。

**（6） White和Rohrer动态阈值算法[26]**

在这个方法中，一个偏移变动平均值（*Biased Running Average*）被用作每一个象素点的阈值。点的水平变动平均值*H* 为：



其中是输入图像，加权因子是一个的一个非线性函数，且对所有变量有。

点的垂直变动平均值*V* 为：



其中加权因子*g*是一个的一个非线性函数，对所有变量也有。 为了使用需二值化的象素的两边信息，用了大量的前瞻象素（*Look Ahead Pixel*）。处的阈值为：



其中*b*是偏移函数。使用和依次作为*x*和*y*方向的前瞻象素。非线性函数和*g*由White的查找表提供。偏移函数*b*则需要对每一类图像个别地调整。

## 3.4 二值化方法小结

综上所述，全局二值化方法和局部自适应二值化方法都存在本身的优点和缺点。全局二值化方法执行速度快、算法相对简单，但是对于光照不均、具有明显噪声、明暗对比度较差的图像，执行效果较差。因为它仅仅考虑了图像的灰度信息，而忽略了图像中的空间信息；局部自适应二值化方法对质量低下的图像有较好的适应性，但是会出现笔画断裂、伪影等情况，而且时间复杂度和空间复杂度均比较高。

# 第四章 设计和实现一个移动应用后端服务框架

## 4.1 本章概述

本文根据文本图像特点，提出了两种新的文本图像二值化的方法。一种基于背景表面的文本图象二值化方法，首先使用Otsu算法估计前景区域，通过插值得到文档图像的背景表面，根据背景图像和源图像提取文字部分，进行二值化处理，最后经过后期处理得到质量较好的二值化图像。另一种是整体与局部相结合的二值化方法，其中全局二值化方法使用的是Otsu算法，局部二值化方法使用的是Bernsen算法。该方法具有全局二值化方法的优点——保持较快的运算速度，同时也保持了局部二值化方法的优点——对于光照不均匀、含有噪声等低质量的图像有比较好的处理效果。该方法能够有效避免断笔、伪影等现象。

## 4.2 基于背景表面的文本图象二值化方法

该方法可以分为以下几个步骤：

* 使用Otsu算法估计前景区域；
* 通过插值获得文档图像的背景表面；
* 提取文字部分；
* 后期处理。

首先我们使用一种全局二值化方法——Otsu算法将前景区域和背景区域大致分离开，由于Otsu算法能够比较稳定的找出阈值，因此能较好的得到估计的前景区域，但由于这种全局二值化方法对输人图像有噪声或不均匀光照等情况处理能力较差，可能有部分噪声区域或背景区域也被选入到前景区域。

将经过第一步处理得到的二值化图像的前景区域对应到源图像中相同的区域，在源图像中通过用邻域背景像素的插值填充前景区域来得到背景表面图像。然后使用均值滤波器对背景表面图像进行滤波。

当背景图像与源图像对应像素的灰度之差大于某个固定的值时，把该像素标记为文字像素，从而获得文字部分，通过这一步的处理，虽然可以有效的较少噪声区域的面积，但是同时笔划的边缘部分有可能被划入背景区域，所以还要进一步处理。

后期处理主要做的工作是减少噪声、保证字符笔划连通性等，以得到质量较高的二值化图像。

图表1显示了该方法对文档图像二值化的处理过程的大体流程。

最终结果

二值化图像

源文本图像

Otsu方法

估计前景区域

插值

得到背景图像

设置阈值

提取文字部分

后期处理

二值化图像

背景灰度图像

均值滤波器

背景灰度图象

二值化图像

图表 1 基于背景表面的文本图象二值化方法处理流程图

### 4.2.1 粗略的估计前景区域

我们这一阶段的目标是获得粗略的前景区域，使用Otsu算法将前景区域和背景区域大致分离开，当然如果源图像中存在噪声，那么大部分噪声也将被选入前景区域。这一阶段的输入是源图像，输出为二值化图像。

### 4.2.2 获取背景区域[2]

我们由和来得到背景图像，对于在图像中像素灰度值为255的点（白色像素），的值与相等，对于剩余的像素（即在中灰度值为0的像素），的值由在中其邻域内背景像素的插值计算得到，即



其中为所选窗口的长和宽，是该窗口内像素的灰度值大于的像素的个数。实验中的取值为，。

为了获得更好的效果，使用均值滤波器对进行平滑处理得到。均值滤波器的思想是若以象素为中心的屏蔽窗口内平均灰度值为时，无条件作处理，由用户给定，且取值越大，噪声减少越明显。但平均是以图像的模糊为代价的。均值滤波器运行效率较高且平滑效果较好。我们在实验中使用。

### 4.2.3 提取文字部分

最终的阈值d是由源图像和背景图像决定的。文字部分是由背景图像和源图像之差大于阈值d的像素组成的，即



前景区域和背景区域之间的平均距离由下面的公式(8)求得



阈值d由公式(9)获得



根据大量的实验，建议公式中各个参数的取值为， ，。接下来的工作就是根据公式(7)对文档图像进行二值化处理，得到二值化图像。

### 4.2.4 后期处理[27]

为了获得质量较高的二值化文档图像，减少噪声、保证字符笔划连通性等，我们需要对得到的二值化图像进行后期处理，进行以下三步滤波处理：

首先，减少噪声。检测二值化图像中的每一个前景像素，是以前景像素为中心像素的窗口内背景像素的个数，如果则将该中心像素变为背景像素，其中

然后，在文字部分填充裂痕、缺失。检测二值化图像中的每一个背景像素，是以背景像素为中心像素的窗口内前景像素的个数，，是窗口内所有前景像素横坐标和纵坐标的平均值，如果、且则将该中心像素变为前景像素，其中，。后两个约束条件是为了防止笔划宽度过宽。

最后一步是为了提高文字笔划的质量。检测二值化图像中的每一个背景像素，是以背景像素为中心像素的窗口内前景像素的个数，如果则将该中心像素变为前景像素。

经过以上三步后期处理，得到最终的二值化图像。

### 4.2.5 实验结果

图表2是使用基于背景表面的文本图像二值化方法对图像进行二值化处理每一步的图像处理结果。通过观察，可以看出该方法尽量消除背景区域被二值化为笔划的伪影现象、较好地保留了有效的笔划像素、抗噪声能力比较强，另外该算法的实现速度也较快。

|  |  |
| --- | --- |
| wenben  a | otsu  b |
| background  c | character  d |
| postprocessing  e |  |

图表 2 使用基于背景表面的文本图像二值化方法对图像进行二值化处理。a为源灰度图像，b为使用Otsu算法得到的前景图像，c为背景表面图像，d为提取文字部分得到的二值化图像，e为经过后期处理后的最终二值化图像。原图像的大小为：282\*139。

## 4.3 Otsu算法和Bernsen算法相结合的二值化方法

Otsu算法是全局二值化方法，全局二值化方法主要适用于前景与背景在灰度级上有比较明显的差异的图像。但当图像质量低时，如存在光照明暗不均匀、文字区域和背景区域对比度低、图像中存在噪声等现象，使用全局方法无法得到好的处理效果。全局二值化方法的优点是处理速度快。Otsu算法的基本思想是将直方图在某一阈值处分割成两组，一组对应于背景部分，一组对应于前景文字部分，当被分成的两组的组内方差最小，组间方差最大时，决定阈值。

Bernsen算法是局部二值化方法。局部二值化方法以象素的邻域信息为基础来计算每一个象素的阈值。如果图像中的一个象素的灰度级高于其区域内定义的阈值时，把该象素标记为背景，否则为前景字符。局部自适应二值化方法被广泛的应用于文档图像的分析中，因为对于质量低的图像使用它可以得到比较好的处理效果。局部二值化法方法的缺点是执行速度不高。Bernsen算法以局部窗口内最大、最小值的均值作为对应考察点的阈值，当窗口内无目标点时，个别噪声点将引起阈值的突变。另外，如果背景部分光照明暗变化很明显的话也将影响局部阈值的变化。当考察窗内均为文字部分时，局部阈值被拉伸，这样会使得本应同类的前景像素被强行二值化为背景像素，所以可见局部窗口大小的选取和笔划宽度密切相关，如果能大于笔划宽度，就能避免窗口完全是目标的情况出现。当然当考察窗内均为背景部分时，局部阈值也会被拉伸，这样会使得本应同类的背景像素被强行二值化为前景像素，从而出现笔划断裂及伪影现象。所以窗口大小的选取在Bernsen算法中是很关键的。窗口选取的过小，容易造成断笔；窗口选取的过大，又会影响运算速度和二值化的效果。

但是Bernsen算法可以与其他二值化方法搭配使用，并且鲁棒性不错。

因此，为了能较好地保持文字笔划信息、减少噪声，又有较高的执行效率，我们将Otsu算法和Bernsen算法相结合，来对文档图像进行二值化处理，扬长避短，弥补各自的不足，发挥各自的优点，希望得到质量较高的二值化图像，同时有较高的执行效率。

该方法可以分为以下几个步骤：

* 使用超限邻域平均法对源图像进行预处理；
* 使用Otsu算法进行全局二值化；
* 使用Bernsen算法进行局部二值化；
* 后期处理。

使用超限邻域平均法对低质量的文档图像进行预处理，可以有效地减少图像中的噪声、平滑背景区域、同时增强背景区域与文字部分的对比度，为以后更有效地进行二值化处理。

对平滑后的图像使用Otsu算法进行全局二值化，根据公式（4）得到的阈值求得两个新的阈值、，将图像中部分像素按照这两个阈值进行二值化。

对于第二步中未被处理的像素使用合适的窗口大小根据Bernsen算法进行局部二值化处理。

后期处理主要做的工作是减少噪声、保证字符笔划连通性等，以得到质量较高的二值化图像。

图表3显示了Otsu算法和Bernsen算法相结合的二值化方法的处理流程。

最终二值化图像



源文本图像



Otsu算法

全局二值化

Bernsen算法

局部二值化

超限邻域平均法

平滑图像

后期处理

灰度图像

二值化图像

图表 3 Otsu算法和Bernsen算法相结合的二值化方法的处理流程图

### 4.3.1 超限邻域平均法[28]

普通的平滑滤波器平滑图像时，使边缘处的灰度值趋向均匀,造成了边缘模糊,为了减少模糊效应, 又能有效的去除噪声，我们使用超限邻域平均法对文档图像进行平滑处理。阈值的邻域平均法以某个灰度值作为阈值,如果某个像素的灰度值大于其邻域像素的平均值,并且该像素的灰度值与其邻域像素平均值之差超过阈值,使用平均灰度值置换这个像素的灰度值,进行平均处理,可去除噪声；否则还保留原值,不进行平均处理,从而减少模糊。

如果某个像素的灰度值大于其邻域像素的平均值，且达到了一定水平，则判断该像素为噪声，继而用邻域像素的均值取代这一像素值，它的数学表达式：



在实验中，阈值。

在操作中对窗口的大小及门限的选择要慎重，太大，噪声消除不干净；太小，易使图像模糊。在实际应用中我们一般用3\*3窗口。这种算法对抑制椒盐噪声比较有效，同时也能较好的保护仅有微小灰度差的图像细节。

### 4.3.2 使用Otsu算法进行全局二值化

使用Otsu算法根据公式（4）得到全局二值化的阈值，从而得到两个新的阈值和，对平滑后的图像中像素灰度值满足和的像素使用公式（11）进行二值化处理。



其中实验中取值。

### 4.3.3 使用Bernsen算法进行局部二值化

对于中剩余的像素，即灰度值满足  的像素使用Bernsen算法进行局部二值化处理，使用公式（9）获取局部阈值。我们在实验中使用的是7\*7大小的窗口。

经过全局二值化和局部二值化处理后，两部分的像素点结合起来，得到二值化文档图像。

### 4.3.4 后期处理

后期处理与基于背景表面的文本图像二值化方法中的后期处理方法类似。

首先，在文字部分填充裂痕、缺失。检测二值化图像中的每一个背景像素，是以背景像素为中心像素的窗口内前景像素的个数，，是窗口内所有前景像素横坐标和纵坐标的平均值，如果、且则将该中心像素变为前景像素，其中，。后两个约束条件是为了防止笔划宽度过宽。

然后，减少噪声。检测二值化图像中的每一个前景像素，是以前景像素为中心像素的窗口内背景像素的个数，如果则将该中心像素变为背景像素，其中

经过后期处理后，得到最终的二值化文档图像。

### 4.3.5 实验结果

图表4 d是使用Otsu算法和Bernsen算法相结合的二值化方法对图像进行二值化处理的结果。通过与图表4 b（只使用Bernsen算法进行二值化的处理结果）和图表4 c（只使用Bernsen算法进行二值化的处理结果）的对比，可以看出结合后的方法在消除伪影方面较Bernsen算法有很大的改进，在去除噪声和保持有效笔划方面较Otsu方法有很大的改进。另外该算法的实现速度也比只使用Bernsen算法要快得多。

|  |  |
| --- | --- |
| 1232  a | bernser1  b |
| otsu1  c | otsubernser1  d |

图表 4 a为原灰度图像，b为只使用Bernsen算法处理得到的二值化图像，c为只使用Otsu算法处理得到的二值化图像，d为使用Otsu算法和Bernsen算法相结合的方法得到的二值化图像。原图像的大小为：396\*407。

通过这组图像的对比，我们可以看出，结合后的方法较只使用Bernsen算法有了很大的改进，但是与只使用Otsu算法的区别不是特别明显，所以我们通过局部放大来观察结合后的算法较Otsu算法在消除断笔和伪影现象以及减少噪声方面的优势：

|  |  |
| --- | --- |
| otsuduanbi  a | obduanbi  b |
| otsuzaosheng  c | obzaosheng  d |
| otsuweiying  e | obweiying  f |

图表 5 a为经过Otsu算法得到的二值化图像的文字部分出现断笔现象，b为使用Otsu算法和Bernsen算法相结合的方法得到的二值化图像相应区域削除了断笔现象，c为经过Otsu算法得到的二值化图像中存在的噪声，d为使用Otsu算法和Bernsen算法相结合的方法得到的二值化图像相应区域削除了该噪声，e为经过Otsu算法得到的二值化图像存在伪影现象，f为使用Otsu算法和Bernsen算法相结合的方法得到的二值化图像相应区域削除了伪影现象。

通过几幅放大后的图像的对比，我们可以看出结合后的二值化方法较单一的Otsu算法有效地消除了断笔和伪影现象，同时减少了噪声。和Bernsen算法相比，综合算法在大大提高运算速度的同时，明显地减少了伪影，断笔现象也大大减少。综合算法在三种方法中，实验效果最佳。

# 第五章 通用性和可用性讨论

这一章我们对比几种常用的文本图像二值化方法和本文中提到的两种二值化方法，比较它们的处理效果。

|  |  |
| --- | --- |
| wenben  a | niblak  b |
| niblack40  c | otsu  d |
| bernser  e | obgaidong  f |
| postprocessing  g |  |

图表 6 a为源图像，b为经过Niblack算法处理得到的二值化图像，其中窗口大小为w=15，c为经过Niblack算法处理得到的二值化图像，其中窗口大小为w=40，d为经过Otsu算法处理得到的二值化图像，e经过Bernsen算法处理得到的二值化图像，f经过Otsu算法和Bernsen算法相结合的方法处理得到的二值化图像，g经过基于背景表面的文本图像二值化方法处理得到的二值化图像。

通过图表6各种二值化方法对相同文档图像的处理效果的比较，我们可以看出：Niblack局部二值化方法受窗口大小的影响比较大，当窗口较小时会在背景区域产生连续大块的非目标黑色区域；当使用较大的窗口时非目标黑色区域减少了，但是对于光照明暗的变化变得不那么有效了，并且会出现不一致的笔划宽度，噪声仍会遗留在处理后的图像中。使用全局Otsu算法得到的结果中明显有断笔现象。而Bernsen算法的结果中则伪影现象比较严重，并且也存在断笔现象。Otsu算法和Bernsen算法相结合的二值化方法有效地消除了断笔和伪影现象，同时减少了噪声，与Bernsen算法相比，综合算法也大大提高了运算速度。基于背景表面的文本图象二值化方法尽量消除背景区域被二值化为笔划的伪影现象、较好地保留了有效的笔划像素、抗噪声能力比较强，实现速度也较快。

# 第六章 结论

## 6.1 两种新的文本图像二值化方法总结

本文通过对以往常用的图像二值化方法的研究，以及对大量文献的学习，根据文档图像自身的特点，结合背景表面方法的处理流程，提出了一种新的基于背景表面的文本图像的二值化方法。该方法能够消除伪影现象、较好地保留了有效的笔划像素、抗噪声能力比较强，实现速度也较快。另外，本文又综合了全局二值化算法与局部二值化算法的思想提出了一种新的综合文档图像二值化算法。该方法有效地消除了断笔和伪影现象，同时减少了噪声，与Bernsen算法相比，综合算法也大大提高了运算速度。

## 6.2 文档图像二值化的研究现状及存在的问题

从二十世纪六十年代初期，人们提出了第一种图像二值化算法，迄今为止，人们已经提出了上千种算法。如此繁多的算法，是人们在不同的研究和工程应用中提出的。因此很难找到一种较好的标准，将所有的图像二值化算法进行合理的归类。本报告主要根据阈值所作用的范围将文档图像二值化算法大体分为两类：全局二值化和局部自适应二值化来介绍的。

但是到目前为止还不存在一个文档图像二值化的通用方法，也不存在一个判断分割是否成功的客观标准。阈值选的过低，会造成字符的笔画断裂；选的过高，图像中的背景不能被去除。这些将严重影响后期的文字识别系统性能。怎样才能兼顾算法的效果与执行的速度两方面，是一个值得深入研究的领域，这些都可以作为我们以后研究的方向。

## 6.3 文档图像二值化的发展趋势

我们必须建立行之有效的一些文档图像二值化质量评价的方法，结合原有的二值化方法，发展更多处理效果良好的新方法。近年来已有一些新的研究手段被引入到阈值选取中，比如人工智能[29]、神经网络[30]、数学形态学[31]、小波分析与变换[32]等等，伴随着文档图像越来越多的进入人们的生活，文档图像二值化会有更广阔的发展空间。

# 致 谢

经过了大概三个月的时间，通过对二值化技术的学习、参考相关书籍和网上信息、资料的查找，以及在老师和同学们的指导和帮助下，我得以顺利的完成了本科毕业设计。在此表示我由衷的谢意。

首先，我要感谢山东大学的指导老师阎中敏，感谢阎老师在本课题的开题、设计和撰写论文期间，给予我的宝贵意见和指导，以及为我提供的良好的实践环境。阎老师渊博的学识、严谨的治学态度、平易近人的风格给我留下了极其深刻的印象。她在我遇到困难的时候鼓励我、教导我要保持良好的心态去克服困难，以积极的心态和严谨的态度面对以后的研究，她的教导使我受益终生。

同时我要感谢中科院自动化所的肖柏华老师和我的导师戴汝为院士，他们在百忙之中抽出时间对我在毕业设计中遇到的难题进行细致的分析与指导，并且在指导的过程中悉心负责，在毕业设计的开始阶段，他们为我的课题选择进行了严谨的分析，为我的毕业设计指明了方向，引导我的设计思路，关注设计的进度，改进程序的功能，使我无论从理论研究还是从实践动手两个方面都受益匪浅。他们渊博的专业知识、严谨的教学态度、敬业的工作态度都给予了我极大的鼓励。

其次，我要感谢我的同学，在我开发过程中，遇到不少难题，是他们不厌其烦地为我解答，给我提供的帮助。

最后，我还要感谢所有指导、帮助我完成毕业设计的专业课老师、教务员老师、辅导员老师，他们的帮助，使我的毕业设计得以顺利完成。

# 参考文献

[1] N. George, “Twenty years of document image analysis in PAMI”, IEEE Trans. On Pattern Analysis and Machine Intelligence, Page(s): 38-62, 2000

[2] B. Gatos, I. Pratikakis, S.J. Perantonis，Adaptive degraded document image binarization，Pattern Recognition, Vol.39, pp317-327.

[3] 刘平，图像分割阈值选取技术综述，中科院成都计算所，2004.2.26

[4] 胡家忠，计算机文字识别技米[M]，北京气象出版社，1994.12一13.

[5] 章毓晋，图像分割评价技术分类和比较，中国图象图形学报，1(2):151一158. 1996.

[6] Barroso J,Rafael A,et a1.Number plate reading using computer vision.Proc. EEEInternationarSymposiumo nIn dustrialE lectionics,Portuga1,1997

[7] 何斌等，Visual C++ 数字图像处理(第二版)，人民邮电出版社2004，人民邮电出版社

[8] 夏良正，数字图像处理，东南大学出版社，南京，1999

[9] Doyle W. Operations Useful for Similarity-invariant Patern Recognition. JACM. 9(2):259-267. 1962.

[10] Lee S , Chung S. A Comparative Performance Study of Several Global Thresholding Techniques for Segmentation[J ] . Computer Vision , Graphics , and Image Processing , 1990 , 52 : 171 - 190.

[11] Prewit J. M. S, Mendelsohn M. L. The Analysis of Cell Images. Ann N Y Acad Sci. 128:1035-1053. 1966.

[12] 于新文，几种图像分割算法在棉铃虫图像处理中的应用，中国农业大学学报 ，2001，6

[13] 张旭丽，两种基于空间域聚类分析的彩色图像分割方法比较，贵州工业大学学报（自然科学版）, 第30卷，第2 期（总第109期），第75－79页，2001.4

[14] N. Otsu ,A threshold selection method from gray-level histograms ,IEEE Trans. Systems Man Cybernet. 9(1) (1979) 62-66.

[15] Otsu N. Discriminant and least square threshold selection. In: Proc 4IJCPR, 1978:592-596.

[16] Kapur J N , Sahoo P K, Wong A K C. A new method for gray-level picture thresholding using the entropy of the histogram. Computer Vision , Graphics and Image Processing , 1985 , 29 (2) : 273～285

[17] Abutaleb A. S. Thresholding of Gray-level Pictures Using Two-dimension Entropy. CVGIP. 47:22-32. 1989.

[18] PK.Sahoo, S.Soltani, A.K.C. Wong, and Y.C.Chen, "A Survey of Thresholding Techniques", Computer Graphics ,Vision and Image Processing 1988 (41),pp.233-260

[19] 阎之烨，基于计算机视觉的苹果颜色分级系统的研究，南京农业大学，2003-6-1

[20] Kittler J, Illingworth J. Minimum error thresholding. Pattern Recognition, 1986; 19(1): 41-47.

[21] J. Bernsen, Dynamic Thresholding of Grey-level Images, *Proceedings of the Eighth International Conference on Pattern Recognition*, pp. 1251-1255, Paris, France,1986

[22] Qivind D, Torfinn T. Evaluation of binazimtion methods for document images[J] . IEEE Trans PAMI,1995,17(3):312-315

[23] W. Niblack ,A Introduction to Digital Image Processing , Prentice Hall , Englewood Cliffs ,NJ ,1986 pp.115-116.

[24] L. Eikvil, T. Taxt and K. Moen, A Fast Adaptive Method for Binarization of Document Images, *Proceedings of the First International Conference on Document Analysis and Recognition*, pp. 435-443, 1991

[25] J. Sauvola ,M. Pietikainen ,Adaptive document image binarization ,Pattern Recognition 33 (2000) 255-236.

[26] J.M. White and G.D. Rohrer, Image Thresholding for Optical Character Recognition and Other Applications Requiring Character Image Extraction , *IBM J. Research and Development*, vol. 27, no. 4, pp. 400-411, 1989.7

[27]R.J. Schilling, Fundamentals of Robotics Analysis and Control, Prentice-Hall, Englewood Cliffs, NJ, 1990.

[28] 杨淑莹,VC++图像处理程序设计,清华大学出版社/北方交通大学出版社,2003.11

[29] 梁栋，李新华，一种基于人工智能的阈值自动选取方法，微电子学与计算机，1999.1

[30] 叶锋，基于神经网络的综合集成车牌识别技术的研究及其应用，2002.4 p22-26

[31] 彭进业等，分形维数在灰度图像二值化中的应用，小型微型计算机, 2001, 22(8): 961-963.

[32] 汪巧萍，一种基于小波分析的纹理背景下的物体分割算法，数据采集与处理，1998，3