

编 号

江南大学

本科生毕业设计（论文）

题目： 蔬菜水果识别软件研究与开发

物联网工程 学 院 计 算 机 专 业

学 号 0304110118

学生姓名 张 杰

指导教师 桑庆兵 副教授

二〇一五年六月

摘 要

蔬菜水果自动识别系统在盲人购菜，外国游客菜场买菜，提高超市人员工作效率等方面有着重要作用。目前利用计算机视觉对不同种类的蔬菜水果进行分类识别仍是一项复杂的任务。本文聚焦于基于颜色，纹理特征的方法以及基于SIFT特征的分类识别方法进行研究与分析，此外，我们还使用C#开发了蔬菜水果价格查询及菜谱推荐系统，其主要内容如下：

（1）基于颜色，纹理特征的方法，我们分别采用①基于传统KMeans背景消除法；②图片背景过度删除法；③利用背景过度删除的图片和背景轻度删除的图片进行目标主体定位；④在③的基础上对已定位的图片再采用Kmeans算法进行背景删除；⑤边缘检测结合Kmeans的背景删除法共五种方法对图片进行预处理，对预处理后的图片提取颜色以及改进的和差纹理特征（ISADH, Improving Sum and Difference Histogram），最后利用径向基核函数（RBF）支持向量机进行训练分类识别。

（2）基于SIFT特征的方法，我们首先提取所有图片的SIFT特征，利用训练图片中提取的SIFT特征集构造码本字典，对于测试图片中提取的SIFT特征分别进行硬赋值（VQ），局部特征约束(LLC)，核码本(KCB),Fisher向量(FK)编码，再利用空间金字塔模型(SPM)进行池化，最后采用线性支持向量机对图片分类识别。

（3）蔬菜水果价格查询及菜谱推荐系统，采用C#语言进行WinForm编程，根据网上蔬菜水果价格网以及菜谱网的网页源代码格式进行字符串操作提取对应的蔬菜价格，相关菜谱等主要信息并在WinForm窗体中显示。

关键词：蔬菜水果识别；SVM；SIFT；空间金字塔匹配SPM；ISADH

Abstract

The System of automatic recognition about vegetables and fruits plays an important role in the field of shopping of the blind and foreign tourists and the improvement of the efficiency of supermarkets. Recognizing different kinds of vegetables and fruits with computer vision techniques is still a complex task at present. In this paper, we do analysis and research focusing on color and texture features and SIFT features. Besides, we develop a system about price checks and recipe recommendation of vegetables and fruits using C#.

(1)Methods based on color and texture:We respectively adopt five methods for image preprocessing.They are:①Picture's background subtraction based on traditional Kmeans.②Excessive picture's background subtraction;③Locating objects using excessive background subtraction and mind background subtraction;④Background subtraction using Kmeans for pictures getting form ③;⑤Combination of edge detecting and Kmeans background subtraction.We take color and ISADH texture from the pictures after preprocessing. Finally,SVM with RBF kernel is adopted for classification.

(2)Methods based on SIFT feature:First,we directly extract SIFT features from pictures.After that we construct codebooks with SIFT features extracting from training set.For SIFT features extracting form test set ,we encode them using VQ,LLC,KCB,FK respectively.Here,SPM is used to pool these features which have been encoded.In the end,linear SVM is used for classification.

(3)System of price checks and recipe recommendation of vegetables and fruits:According to the format of source codes deriving from vegetable and fruit price website and recipe recommendation website,we apply C# doing some string handling to get the main information like price of vegetables,etc.we want and show them in forms of Windows.

Keywords:Automatic recognition of vegetables and fruits;SVM;SIFT;SPM;ISADH

目 录

第一章 绪论.....	1
1.1 研究背景和意义.....	1
1.2 相关领域的研究现状.....	2
1.3 论文主要内容与组织结构.....	3
第二章 基于颜色和纹理特征的方法.....	5
2.1 引言.....	5
2.2 实验原理及重要概念解释.....	5
2.2.1 HSV 空间.....	6
2.2.2 Kmeans 算法消除背景提取颜色特征.....	6
2.2.3 结合背景过度删除图片与背景轻度删除图片的目标主体定位法.....	7
2.2.4 通过边缘检测结合 Kmeans 的方法进行背景删除.....	9
2.2.5 改进的和差直方图纹理特征 (ISADH)	9
2.2.6 线性缩放.....	10
2.2.7 核函数.....	10
2.2.8 交叉验证和网格搜索.....	11
2.3 实验方法和步骤.....	12
2.3.1 预处理.....	12
2.3.2 背景删除以及颜色特征提取.....	12
2.3.3 提取 ISADH 特征.....	13
2.3.4 线性缩放格式转化.....	13
2.3.5 对训练集交叉验证获取参数 C 和 γ ，代入获得最终实验结果.....	13
2.4 实验结果与分析.....	13
2.4.1 实验数据:.....	13
2.4.2 实验结果:	13
2.4.3 实验分析.....	16
2.5 结论.....	17
第三章 基于 SIFT 特征的方法.....	19
3.1 引言.....	19
3.2 实验原理及重要概念解释.....	19
3.2.1 主成分分析 PCA.....	19
3.2.2 kd-tree.....	21
3.2.3 Hard Assignment.....	25
3.2.4 Soft Assignment.....	25
3.2.5 KCB (Kernel Codebook)	26

3.2.6 LLC(Locality-constrained Linear Coding).....	26
3.2.7 Fisher Kernel.....	28
3.2.8 SPM.....	29
3.2.9 线性支持向量机 Linear Support Vector Machine.....	30
3.3.实验方法和步骤.....	34
3.3.1 构造 imdb.....	34
3.3.2 提取 SIFT 特征.....	34
3.3.3 构造码本字典.....	34
3.3.4 对每张图片进行编码.....	35
3.3.5 采用 SPM (Spatial Pyramid Method) 进行池化.....	35
3.3.6 采用 SVM 分类器进行分类识别.....	36
3.3.7 FKdemo 部分.....	37
3.4 实验结果与分析.....	37
3.4.1 实验配置.....	37
3.4.2 实验数据.....	37
3.4.3 实验结果.....	37
3.4.4 实验结果分析.....	38
3.5 结论.....	38
第四章 菜价查询菜谱推荐系统.....	41
4.1 主要功能.....	41
4.2 实验原理.....	41
4.2.1 模拟鼠标点击进行自动搜索.....	41
4.2.2 网页核心内容的提取.....	41
4.3 实验结果.....	41
第五章 结论与展望.....	45
5.1 结论.....	45
5.2 不足与展望.....	45
致谢.....	47
参考文献.....	48

第一章 绪论

1.1 研究背景和意义

我国是世界蔬菜水果生产大国，总产量跃居世界前列，但是因为品种结构不合理，产品商品化处理技术和设备落后，产品缺乏市场竞争力，导致出口量少，价格低，西方国家凭借专业的机器化生产设备，使其在果蔬产业占据了极大优势，对我国的果蔬的对外市场也构成了巨大挑战。虽然我国的果蔬市场中也拥有一些分类分级的操作，但大都是在人工下进行，这种方式耗时耗力，而且效率不高，同时违背了解放生产力的目标，带来了诸多不可避免的缺点，例如无统一的标准，等级判定往往是根据个人经验和瞬间判断的结果，有时还受到人的情绪和疲劳程度的影响。

识别不同种类的蔬菜和水果是超市购物中经常需要处理的任务，超市员工必须先了解蔬菜或水果的种类，然后再决定它的价格，很耗费时间。二维码的使用很好地解决了包装好的蔬菜水果的问题，但是考虑到消费者更多的时候希望自己挑选蔬菜水果，因此这些蔬菜水果必须先被称重，然后结合对应的蔬菜单价才能得到购买的蔬菜水果的最终价格。解决这一问题的常用方法是对每一类蔬菜水果查看条目代码，但这对记忆力要求较高，时常会在定价时出现错误。为了帮助超市员工解决这些问题，许多超市会给超市员工发放带有图片和编码的书本，而翻阅书本查找价格的方式又会浪费很多时间。

很多人希望出国生活，往往需要前往当地的蔬菜水果市场购买食材自己做饭做菜，但对于当地的蔬菜种类和对应的菜价不了解，害怕上当受骗，造成较大困惑，因此我们非常需要能有一款帮助我们了解当地蔬菜水果种类以及对应菜价行情的软件。

此外，对于蔬菜水果识别的软件的研究对于帮助盲人购物等也能起到较好的辅助作用。

因此研究一个智能的蔬菜水果识别系统已成为提高市场竞争力，改善人民生活条件，方便人民日常生活的必然。

计算机视觉被认为是允许计算机监测和提取图片内容和多维度数据以方便解决特定的视觉问题的研究与方法的应用，例如模式分类问题。计算机视觉主要有六大领域[1]，它们分别为感知，预处理，分割，描述，识别和解释。计算机视觉应用的例子包括工业机器人的和自动化交通工具的控制系统，对医学图片分析建模以及对地形情况进行建模[2]。此外，计算机视觉系统还被应用于农业问题例如监测农作物生长和在避雨亭[3]下的野草生长情况或者监测种植的谷物中杂草生长的情况[4]等。

特征提取是从一张图片中获取有意义的目标物体高级信息的过程[1]，特征提取也是解决这些问题或者精确描述这些问题，构造变量组合方法的一个广义词汇，在模式识别中，特征经常包含相对于灰度，纹理，形状，内容等的信息，在图像处理或者机器视觉中，最初的模式度量或者一些度量模式的子序列会被转化成新的模式特征[5]。

模式分类是使用目标高阶信息的过程，因此提取的特征被用来将目标赋值为某一类别，通过研究识别算法，它还被用来自动识别图像中的物体。

图像识别中的模式识别（Pattern Recognition），是一种从大量信息和数据出发，在专家经验和已有认识的基础上，使用来自统计，机器学习和其他领域的方法[6]，是利用计算机和数学推理的方法对形状、模式、曲线、数字、字符格式和图形自动完成识别、评价的

过程。模式识别包括两个阶段，即学习阶段和实现阶段，前者是对样本进行特征选择，寻找分类的规律，后者是根据分类规律对未知样本集进行分类和识别。

随着机器学习和模式识别技术的发展，对目标识别的速度都以及精度得到了很大提高，对于研究高识别率的蔬菜水果识别算法都提供了极大的帮助与参考。

1.2 相关领域的研究现状

而针对上述1.1中描述的问题，基于计算机视觉和图像处理技术的自动蔬菜水果识别系统是很好地替代选择。在这些方法中，算法被用来对视觉特征（纹理，形状，大小，颜色）进行编码形成特征向量，利用特征向量结合机器学习的方法对蔬菜水果图片进行分类。

水果图像的预处理和识别系统在水果包装行业已经被广泛采用[7]。主要采用统计模式识别的方法，根据水果的颜色值、平均直径(粗度)、长度、弯曲度、面积等特征，构造一个水果特征空间图，以此作为标准对水果的种类进行判断。文献[8]利用 Zemike 矩对苹果果形进行了分类。文献[9]采用遗传规划算法 对橘子、苹果、梨、香蕉和桃 5 种水果进行了分类。罗承成等提出了适合果蔬图像的包生成方法，采用改进的单色块及其邻域特征提取算法将训练样本组织成多实例包[10]。阳江平将果蔬图像采用 Gabor 滤波提取其 Gabor 特征，采用纠错支持向量机分类器进行分类[11]。

目前国外已经存在蔬菜视觉系统 VeggieVision，它是第一个拥有完整的规格并带有友好用户界面系统的超市农产品识别系统[12]。R.M.Bolle[13] 等早在 1996 年就针对果蔬设计了一套识别装置，通过巧妙的结构设计和偏振片的运用，有效的解决了光照统一性的难题，通过提取水果的颜色和纹理特征，最终能够得到 95%以上的 Top4 识别率。

METTLERTOLEDO 公司开发的 UC-GTT-P Self-service scale 已经投放到欧洲的各大超市使用。Hong 等人[14]则采用形态检测方法将核桃和榛子分为三组。Baltazar 等人[15]则结合三层贝叶斯分类器将数据融合技术用于无损的新鲜完整的番茄图片中。Pennington 等人[16]使用集成算法进行水果蔬菜的分类。Pholpho 等人[17]通过结合主成分分析，局部最小平方判别分析和弱独立类模拟建模等技术将视觉光谱学用于非青肿和青肿的龙眼分类进行分类模型的开发。Rocha 等人[18]提出使用关于自动蔬菜水果识别的不同图片特征的融合技术，采用颜色内 / 外点直方图，颜色聚合向量，和差直方图等特征作为识别特征，采用支持向量机分类器进行识别，他们的融合方法将识别任务分为多个二元分类问题，好的识别率能达到将近 97%，因为这个方法使用了大量的分类器，它结合了 $\binom{N}{2} = O(N^2)$ 个 SVM 分

类器，其中 N 是分类器个数；Arivazhagan 等人[19]则提出了基于最小距离准则的分类器进行农产品的识别，但他们在对数据集[18]进行测试时却表现了较差的性能。但是上述的技术或多或少会有一个或多个下述的缺点：（1）它们需要额外的传感器，例如气体传感器，不可见光传感器和重量传感器。（2）分类不适用于所有的水果，即它们只能识别相同类别的水果。（3）因为不同的水果图片可能会有相同的颜色和形状特征，所以识别系统的鲁棒性不强。（4）不方便使用，更多的还是局限于超市等特定地方之中，在菜市场等则不适用。

支持向量机是基于机器学习理论[20]的最先进的分类算法。相比于人工神经网络，决策树和贝叶斯网络，支持向量机拥有重要的优势：即它的高正确率，简单的数学原理和直接的几何描述。此外，它不需要大量的训练集来避免过度拟合[21]。

在本篇文章中，我们选择图像识别的方法，它只需要一个数字照相机即可。为了提高识别效果，本文提出了两种不同的方法：①基于颜色和 ISADH 纹理特征[22]的识别方法；②基于 SIFT 特征集合编码池化的识别方法，希望这两种方法可以为解决蔬菜水果分类问题提供较好的参考。

1.3 论文主要内容与组织结构

本文的主要内容与组织结构如下：

第一章绪论部分讲述研究蔬菜水果识别与菜价查询和菜谱推荐软件的意义；目前国内外的研究现状，主要的研究成果以及存在的一些弊端以及指出本文的组织结构和主要内容。

第二章讲述常规蔬菜水果分类方法：首先我们采用不同的预处理方法，包括：①常用的Kmeans消除背景算法；②采用背景过度删除算法；③利用过度删除图片和轻度删除图片对图片进行定位；④在③的基础上对已定位的图片再进行简单背景删除；⑤边缘法结合Kmeans的背景删除共五种预处理方法，在处理后的图片上提取基于目标主体的颜色和 ISADH纹理特征，之后分别对①仅适用颜色特征；②仅适用ISADH纹理特征；③颜色特征和ISADH纹理特征结合的情况采用以径向基核函数（RBF）作为核函数的支持向量机进行最终的分类识别，对比分析实验结果，获得最终结论。

第三章论述了SIFT特征法：直接对所有图片提取SIFT特征，利用文本分类的思想，将提取的SIFT特征作为我们的词汇，而每张图片则被当做为一个文本文件，对由训练集构成的所有文本文件中的词汇采用 kdtree的方法进行词本构造，对由测试集中的图片所提取的SIFT特征分别进行VQ，KCB,LLC和FK编码，再使用SPM进行池化，最终采用线性支持向量机进行处理，对比分析不同编码方式获得的结果，获得最终结论。

第四章介绍蔬菜水果价格查询和菜谱推荐软件：根据中国蔬菜水果价格网和菜谱推荐网的搜索格式进行模拟搜索，获取目标搜索网页的网页源代码，根据网页源代码格式利用C# Winform编程对字符串进行处理获取我们所需的结果并显示。

第五章总结全文并对未来进行展望。

第二章 基于颜色和纹理特征的方法

2.1 引言

基于水果蔬菜颜色，纹理，大小和形状特征的方法在蔬菜水果识别领域最为常见的方法，诸如 Shiv Ram Dubey 采用 ISADH (Improved sum and difference histogram feature descriptor) 作为特征对 15 种水果采用 SVM 进行分类识别获得了 99% 的正确率[22]，Yudong Zhang and Lenan Wu * 采用颜色，形状和纹理特征对 18 类不同水果进行分类识别获得了 88.2% 的正确识别率[23]，S.Arivazhagan¹, R.Newlin Shebiah 等人在实验中发现颜色和纹理在识别过程中是互补的，并利用颜色和纹理特征对 15 类图片进行分类识别，最终获得了 86% 的正确率[24]。

因为本次实验采用的数据中每张图片都可能含有多个物体，所以目标主体的大小和形状则很难作为判定标准，所以本文仅仅采用颜色和纹理特征的方法，因为我们只需要考虑颜色和纹理特征，所以我们只尽可能多的需要获得目标的主体部分即可，而无需获得整个目标主体。

本章采用了 (1) 传统的 Kmeans 背景删去法；(2) 对背景进行过度删除 (参见 2.2.2)；(3) 先使用 K-Means 算法进行背景的删除，结合背景过度删除与背景轻度删除的图片进行目标主体定位 (参见 2.2.3) 获得蔬菜水果位置定位，对定位的蔬菜水果直接提取 ISADH 特征；(4) 在 (3) 的基础上利用 Kmeans 方法对已定位目标主体位置的图片进行简单背景删除获得含有更少背景的目标主体，在此基础上提取蔬菜水果的 ISADH 纹理特征；(5) 通过边缘检测并结合 Kmeans 的方法进行背景删除 (参见 2.2.4) 五种预处理方法。然后再对预处理后的图片进行 ISADH 纹理特征的提取，最后结合提取的颜色特征，采用多类别支持向量机并利用交叉验证的方式对图片的进行训练，分类识别，在背景较为简单的情况下取得了较高的识别率。

2.2 实验原理及重要概念解释

本章将图片分为训练图片集以及测试图片集，对所有的图片采用 K-Means 算法进行主要背景的删除并对蔬菜水果的目标位置进行定位，然后对定位后的目标图片进行颜色和 ISADH 特征的提取，将提取的特征代入多类别支持向量机进行训练，利用训练模型对测试图片进行分类识别获得最终结果，实验的原理图可参考图片 2-1。



图 2-1 实验原理图

2.2.1 HSV空间

HSV空间中，H表示色相，色度，色彩，如红，黄，蓝等，用度数进行表示，而其中红色对应于0度，绿色对应于120度，黄色为240度等；

S表示饱和度，即色彩的深浅度，为一比例值，范围从0-1，它表示颜色的纯度和该颜色最大的纯度之间的比例，比例越大，表示色彩越纯，反之，表示色彩越灰。

V表示色调，纯度，色彩明亮程度，范围从0-1。

2.2.2 Kmeans算法消除背景提取颜色特征

在图片分类中如果直接根据灰度值均分的思想对图片进行分类，例如对灰度值为256的图片，灰度值大于等于128的为1，灰度值小于128的为0，这种方法虽然简单，但是却忽略了图片内部的特点，如果一张图片所有的灰度值都不高于128，则最终的结果将会是全为0，这会使我们损失图片所有的信息，而Kmeans算法则很好的解决了此类问题。

①Kmeans算法的基本思想为：初始随机给定K个簇中心，按照最邻近原则把待分类样本点分到各个簇。然后按平均法重新计算各个簇的质心，从而确定新的簇心。一直迭代，直到簇心的移动距离小于某个给定的值。

Kmeans算法步骤为：

(1) 从D中随机取k个元素，作为k个簇的各自的中心。

(2) 分别计算剩下的元素到k个簇中心的相异度，将这些元素分别划归到相异度最低的簇（相异度我们往往用欧几里得距离进行表示）。

(3) 根据聚类结果，重新计算k个簇各自的中心，计算方法是取簇中所有元素各自维度的算术平均数。

(4) 将D中全部元素按照新的中心重新聚类。

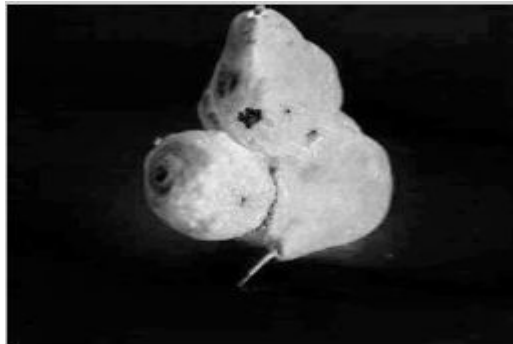
(5) 重复第4步，直到聚类结果不再变化。

(6) 将结果输出。

②采用 Kmeans 算法提取颜色特征提取，首先我们将 RGB 图像转化为 HSV 图像，并使用 S 域处理，因为 S 域中背景和主体目标灰度值差异更加明显，便于进行背景删除提取目标，如图 2-2 和图 2-3 所示 (a)，(c) 为原图，(b)，(d) 为对应的 S 域图像。



(a)



(b)

图 2-2 : (a) 梨子原图; (b) 梨子 S 域图像

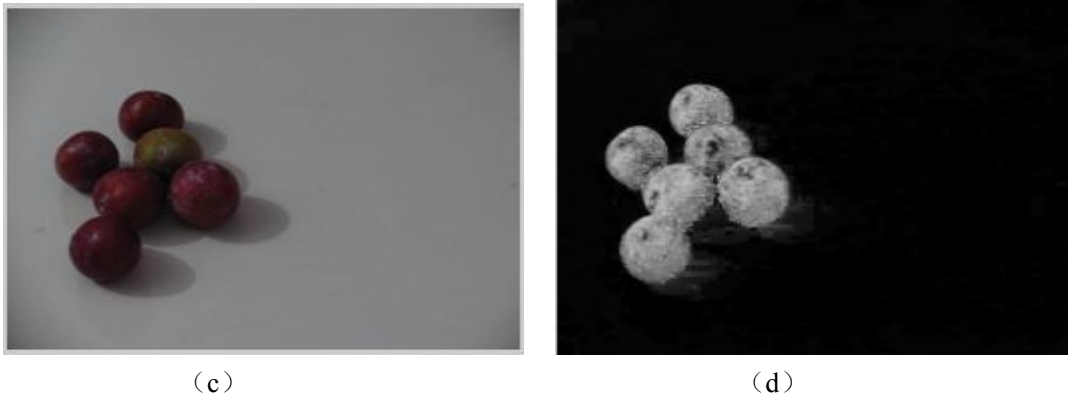


图 2-3: (c) 李子原图; (d) 李子 S 域图像

采用 Kmeans 方法获得两个聚类中心点 a_1 a_2 , 通过观察 HSV 图像的特征我们默认将 a_1 和 a_2 中较小的一个作为背景中心 (这边假设 a_1 较小), 则距离 a_1 较近的像素点都当做背景处理, 而距离 a_2 较近的点则当做目标主体处理, 但就如图 2-3 中所示, 目标主体的影子有时也会被当做主体目标 (因为它的灰度值的大小介于 a_1 和 a_2 之间), 导致我们获得的颜色不纯正 (它包含了背景的颜色), 而如果我们选择灰度值大于 a_2 的作为目标主体进行颜色提取, 虽然会损失很多目标主体的部分, 但剩余的部分都为目标主体部分, 再删除面积较小的噪声。具体可参见图 2-4, 这就是引言 (2) 中所述的过度删除情况, 我们可以直接从这上面获取对应的目标主体的颜色特征, 这边颜色特征主要通过 (R,G,B) 的形式表示, R,G,B 分别为对应的目标主体的颜色的均值。

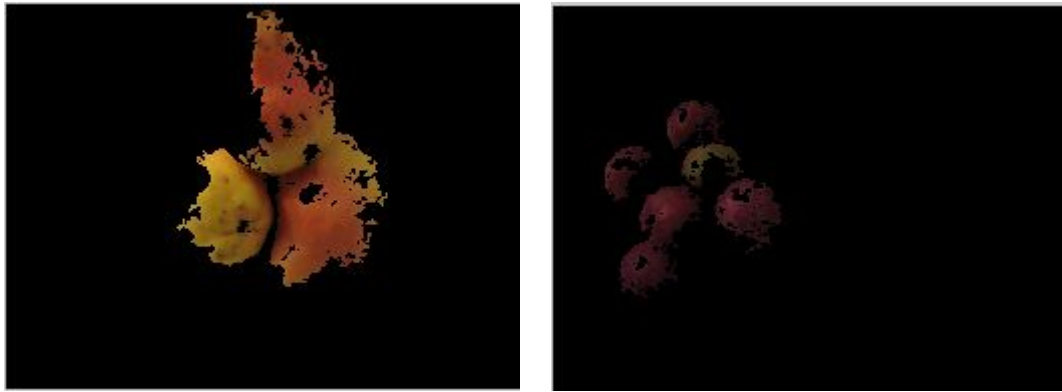


图 2-4: Kmeans 删除背景结果

2.2.3 结合背景过度删除图片与背景轻度删除图片的目标主体定位法

正如图片 2-4 所示, 将比灰度值 a_2 (这边 a_2 等价于 2.2.2 中的 a_2) 小的部分作为背景, 会损失大量信息, 为了弥补损失, 我们决定通过删除小背景 (即以小于灰度值 a_1 作为背景的情况) 的图片进行还原, 因为删去了小背景的图片必定包含了删去大背景的图片的所有元素, 而且删除大背景图片的主体部分的周围往往也是我们的目标主体被误删部分, 还原方法为:

(1) 对删除大背景图片中主体部分的周围邻近区域 (这边默认设置为 5) 进行扫描, 如图 2-5 所示, 主体部分 (红色区域) 周围为 3 的近邻区域, 如果扫描区域在小图片中同样

为主体部分（如图 2-5 中绿色的部分），我们就将其默认标记为主体部分，对大背景中的所有主体部分的像素点进行扫描，得到包含更多目标主体部分的图片 c；

（2）对上述图片 c 进行水平线和垂线扫描扫描，为了排除少数噪声带来的干扰，我们要求在同一水平线上（或者垂直线上）的所有像素点的个数超过相应个数（这边默认设置为 5），分别从上下左右四个方向扫描，扫描时如果满足上述要求，则停止，将其分别记录为上下左右四个边界的值，此时我们便完成了对于目标主体的定位，如图 2-6 所示：

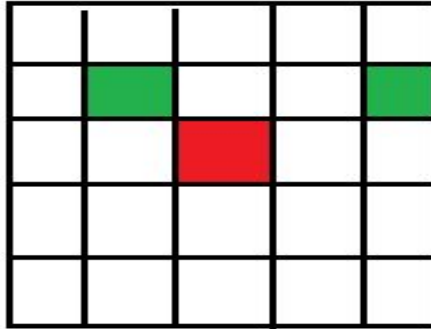


图 2-5：模拟图



图 2-6：目标主体的定位

对于初步获得的包含目标主体的图片我们可再进行简单的背景消除（我们需要保留更多的主体部分和尽量少的背景，此时我们选取 S 区域中灰度值小于 $(a_1 + a_2)/2$ 的作为背景）来减少背景带来的影响，方便后续更加准确地提取目标的纹理特征。



图 2-7：定位的主体进行再删除

2.2.4 通过边缘检测结合Kmeans的方法进行背景删除

在背景相对简单的情况下，目标物体的边缘相对较容易获得，而边缘恰恰又是背景与目标主体的分界线，利用这一特性，我们可以获得新的背景删除法，其步骤为：

- (1) 读取图片，获得图片的S区域（如图2-2,2-3右侧图片所示）；
- (2) 采用canny算子并选择合适的参数获得图片的边缘部分；
- (3) 消去零散的细小的白色噪声；
- (4) 对边缘的左右两侧分别提取图像的灰度值，用作kmeans的数据集；
- (5) 采用Kmeans算法获得两聚类中心， a_1 和 a_2 ，把灰度值小于 $\frac{a_1+a_2}{2}$ 作为背景，即赋值为0，最后我们画出删去背景后对应的原图的情况（如图2-8所示）。

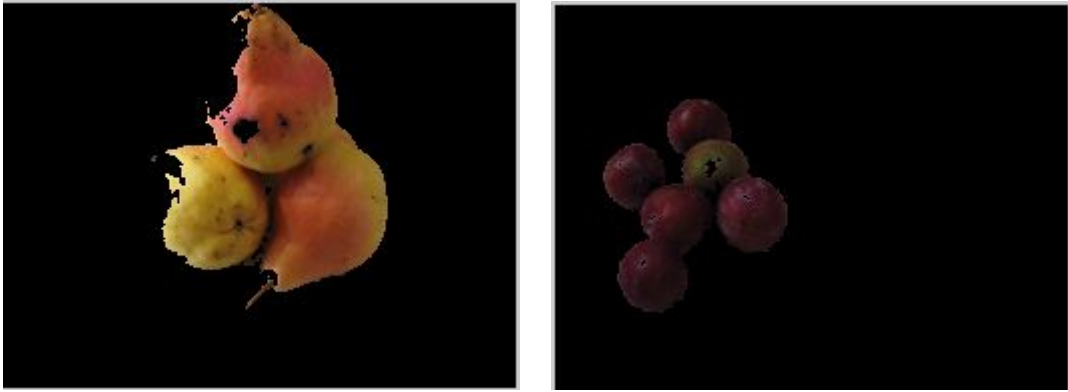


图 2-8：边缘检测结合Kmeans的方法进行背景删除法

2.2.5 改进的和差直方图纹理特征（ISADH）

User[25]已经证明了两个带有相同方差的随机变量的和与差是弱相关的并且定义了它们相关的联合概率函数的主轴，而Dubey教授[22]在此基础上又提出了图片相邻元素之间的灰度值的和差直方图纹理特征。

ISADH纹理特征算法。

- (1) 寻找图片I中每一个通道（R,G,B三个通道）的和S与差D，并采用位移量（1,0）进行平移，其中：

$$S(x, y) = I(x, y, 1) + I(x+1, y, 1)$$

$$D(x, y) = I(x, y, 1) - I(x+1, y, 1)$$

- (2) 利用位移量（0,1）寻找S的和 S_1 与差 D_1 ：

$$S_1(x, y) = S(x, y) + I(x, y+1)$$

$$D_1(x, y) = S(x, y) - S(x, y+1)$$

- (3) 利用位移量（0,1）寻找和 S_2 与差 D_2 ：

$$S_2(x, y) = D(x, y) + D(x, y+1)$$

$$D_2(x, y) = D(x, y) - D(x, y+1)$$

- (4) 通过连接 S_1, D_1, S_2 和 D_2 的直方图寻找1st通道的直方图；

- (5) 对彩色图片的2nd和3rd通道重复步骤（1）和步骤（4）；

(6) 对所有的三个通道的直方图连接, 形成最终输入图片I的ISADH纹理特征。

ISADH纹理特征依赖于相邻像素之间的灰度值, 两张同类图片的直方图可能会区别很大, 对于这些图片, 它们的ISADH的差值会很小。也就是说, 如果两张图片的差值很小, 则这两张图片属于同一类, 反之, 则属于不同类。

2.2.6 线性缩放

在实验支持向量机进行训练分别识别前, 我们需要对所有的特征进行缩放, 对数据进行缩放的目的: (1) 可以避免属性中属性数值较大的控制属性数值在较小范围内的属性; (2) 能避免在计算过程中出现的数值计算较困难的情况, 例如在线性核和多项式核计算中, 大的属性值往往会造成较大的数值计算问题。

本文我们采用线性缩放的方法将原数据缩放至区间[-1 1]中:

对于 $m \times n$ 的矩阵C, m 表示样本个数, n 表示属性个数, 我们获取每个属性中数值最大的和数值最小的元素 \max, \min , 如果最大最小值相同, 表明此属性对于我们进行分别识别没有任何帮助, 我们将此列删除, 对于最大值最小值不同的情况, 我们进行如下线性缩放操作。

$$\max_i = \max(x_{1i}, x_{2i}, \dots, x_{mi})$$

$$\min_i = \min(x_{1i}, x_{2i}, \dots, x_{mi})$$

$$x_{ij} = -1 + \frac{x_{ij} - \min_i}{\max_i - \min_i}$$

2.2.7 核函数

在直接利用颜色属性进行水果蔬菜分类时, 因为我们只取R,G,B三个通道的均值作为颜色特征表示, 而分类的种数却有15种, 所以我们倾向于使用核函数将低维空间中的点映射至高维空间中再进行线性分割。常见的核函数有:

线性核函数: $K(x_i, x_j) = x_i^T x_j$

多项式核函数: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$

径向基核函数 (RBF): $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0$

S型核函数: $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

这边 γ , r 和 d 都称作是核参数, 本章主要采用径向基核函数进行处理。

采用径向基核函数是考虑到 (1) 相比于线性核函数, 它能较好地解决属性和类标签之间的非线性情况; (2) 线性核是径向基核函数的特例, 因为带有惩罚参数C的线性核与径向基核函数在某些相同的参数 (C, γ) 下有着相同的性能[26]; (3) S型核函数在特定参数下和径向基核函数性能相同[27]。 (4) 超参数的个数会影响模型选择的复杂度, 多项式核相比于径向基核函数拥有更多的超参数; (5) 径向基核函数拥有更少的数值问题, 例如在多项式核函数中当 $(\gamma x_i^T x_j + r > 1)$ 或者 $(\gamma x_i^T x_j + r < -1)$ 并且多项式的度数较大时, 核函数的值会趋于无穷或者为0, 而在S型核函数中, 在某些参数下会出现无效的情况[25]。

2.2.8 交叉验证和网格搜索

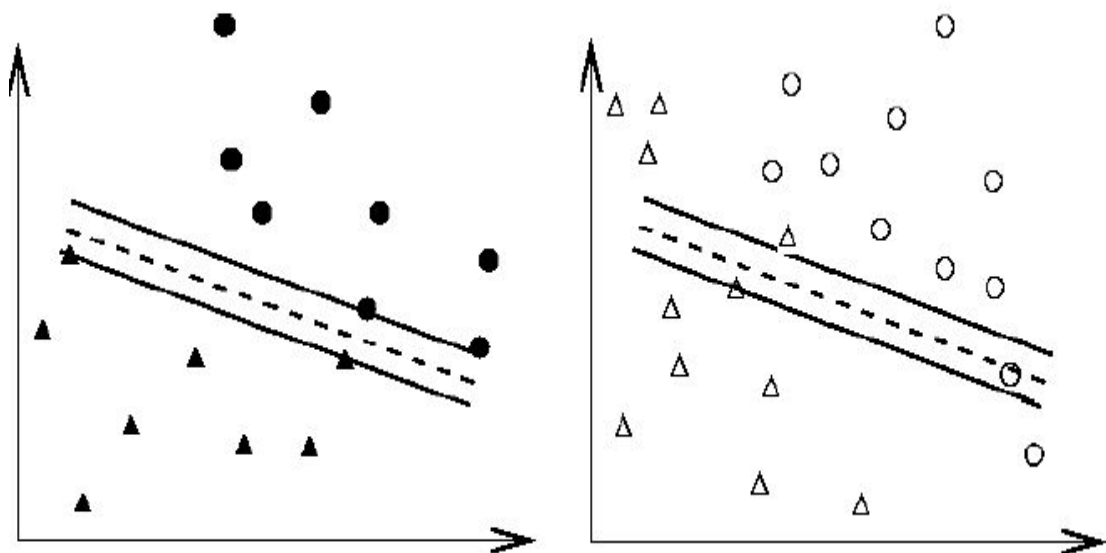
在径向基核函数中对于参数 (C, γ) 的选择会很大程度上影响实验的结果，而在给定的问题中我们很难知道参数 C 和 γ 的值，为了获取较好的参数值进行实验，我们采用交叉验证的方法。

① v 层交叉验证：在 v 层交叉验证中，我们首先将训练集分割为相同大小的 v 个子集，使用 $v-1$ 个子集作为训练而对剩下的一个作测试，则所有训练集的所有子集都将会被测试一次，而交叉验证的正确率就是被正确分类的数据的比率。

此外，交叉验证还可以防止过度拟合，具体情况可参见图2-9。

②网格搜索；采用网格搜索主要原因：（1）从理论上说，网格搜索做的是穷举式的参数搜索方法，更加简单可靠；（2）寻找的参数只有两个，即 (C, γ) ，在计算时间上相比于先进的方法相差不大；（3）参数 (C, γ) 之间的独立性保证我们可以采用并行编程的方法提高效率。

因为网格搜索采用的是穷举策略，效率较差，为了提高效率，我们先采用粗糙的网格搜索，在网格上确定更好的区域之后，再采用精细的网格搜索确定最终参数。



(a) 训练数据和过度拟合分类器

(b) 对测试数据使用过度拟合分类器

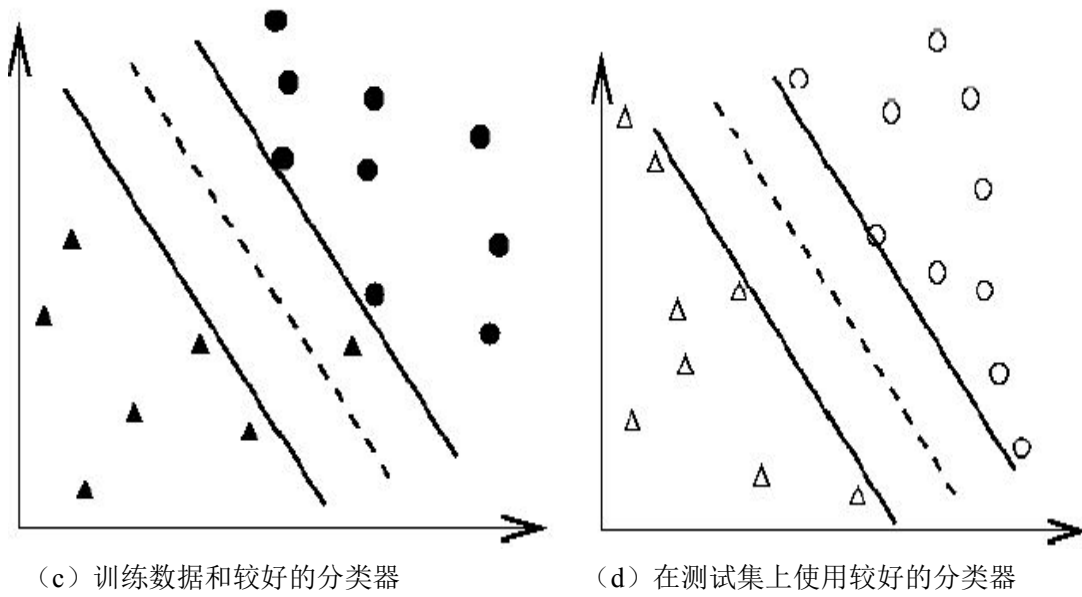


图2-9: 分类器的多种情况

2.3 实验方法和步骤

2.3.1 预处理

读入所有图片，因为读入图片较大（1024*768）并将图片转化为原图片的 $\frac{1}{4}$ 大小，并将所有图片转化为 HSV 形式，提取 S 区域；

2.3.2 背景删除以及颜色特征提取

对所有图片的 S 区域进行 Kmeans 处理，获得两个聚类中心 a_1 和 a_2 ，

(1) 针对简单的 Kmeans 背景删除法，我们直接将灰度值小于 $\frac{a_1+a_2}{2}$ 的作为背景；

(2) 我们直接将灰度值小于 $\max(a_1, a_2)$ 的作为背景，获取背景过度删除后的图片；

(3) 获取 a_1 和 a_2 二者中的较小的一个，假设 a_1 较小，我们把任何灰度值小于 a_1 的当做是背景，将其灰度值赋值为 0, 获得图片 p, 并利用背景过度删除后的图片结合图片 p 进行还原，得到还原后的图片；

(4) 在 (3) 的基础上我们重新获取聚类中心 a_3 和 a_4 ，并把灰度值小于 $\frac{a_3+a_4}{2}$ 的作为背景；

(5) 采用 canny 算子对原图片得到的 S 域进行边缘提取，获得边缘部分，把边缘部分左右两边的灰度值作为数据集，采用 Kmeans 算法计算得到聚类中心 a_5 和 a_6 ，把灰度值小于 $\frac{a_5+a_6}{2}$ 的作为背景。

因为背景过度删除后的图片必定包含目标主体的绝大部分，我们直接对这部分图片计算 R,G,B 三个通道的颜色值的和，并取它们的平均值作为图片的颜色特征；

2.3.3 提取 ISADH 特征

对 2.3.2 中的 5 种情况获得的图片分别代入函数 `isadh_hsv` 中获取 ISADH 特征。

2.3.4 线性缩放格式转化

将获得的颜色和 ISADH 特征对所有的特征按照 2.2.5 所述方式缩放至区间 $[-1,1]$ 之间，并进行标记，并转化为如下格式：

```
1 1:0.1 2:0.3 3:0.4 ...
1 1:0.3 2:0.4 3:-0.12 ...
2 1:0.4 2:0.13 3:-0.4 ...
2 1:-0.3 2:0.24 3:-0.1 ...
.....
```

其中第一个数表示所属类别，跟在后面的表示为对应特征值，例如第一行表示它属于第一类，它的第一个特征值为 0.1，第二个特征值为 0.3，第三个特征值为 0.4 依此类推，这种格式的数据我们可以直接使用 `libsvm` 中的 `windows` 下的 `svm` 进行操作。

我们也可以使用另外一种格式，分别用两个数组进行存储，`trainlabel` 为 $m \times 1$ 的向量，`traindata` 是 $m \times n$ 的矩阵，其中 `trainlabel(i)` 表示第 i 个样本属于的类别，`traindata(i,j)` 表示第 i 个样本第 j 个特征值，这样方便使用 `libsvm` 中 `matlab` 程序进行直接操作。

以上两种形式可以通过 `libsvm` 中的 `libsvmwrite` 和 `libsvmread` 实现相互转换。

2.3.5 对训练集交叉验证获取参数 C 和 γ ，代入获得最终实验结果

对于 2.3.4 中的格式 1 类型的数据我们可以直接使用 `libsvm` 中的 `windows` 下的 `svm` 的函数，在命令行格式下使用语句 `python easy.py` “训练集” “测试集” 的形式获得最终解，其中包含了交叉验证获取参数 C 和 γ 的部分，最终的结果也可以直接在 `libsvm` 中的 `windows` 中找到。或者直接使用 `libsvm` 中 `matlab` 程序直接进行求解获得最终结果。

2.4 实验结果与分析

2.4.1 实验数据:

本次实验采用的实验图片为：aagata_potato（201张），asterix_potato（182张），cashew（210张），diamond_peach（211张），fuji_apple（212张），granny_smith_apple（155张），honeydew_melon（145张），kiwi（171张），nectarine（247张），onion（75张），orange（103张），plum（264张），spanish_pear（159张），taiti_lime（106张），Watermelon（192张），共15类2633张，我们取每类图片的前2/3作为训练图片，后1/3作为测试图片。

2.4.2 实验结果:

本次实验主要对如下五类情况进行实验：

- 1) 使用常用的Kmeans算法；
- 2) 使用背景过度删除的图片；
- 3) 对图片定位后不进行任何处理；

4) 对图片定位后再进行简单背景删除;

5) 边缘法结合Kmeans的背景删除。

每类情况还分别对比了仅提取颜色特征, 仅提取ISADH特征以及采用颜色特征和ISADH特征结合的实验结果。

①常用的Kmeans算法, 如表2-1所示:

表2-1: 常用的Kmeans算法运行结果

水果种类	颜色特征	ISADH 特征	颜色特征+ISADH 特征
agata_potato	95.59%	100.00%	100.00%
asterix_potato	83.61%	95.08%	95.08%
cashew	92.86%	100.00%	100.00%
diamond_peach	70.42%	97.18%	98.59%
fuji_apple	5.63%	92.96%	95.77%
granny_smith_apple	98.08%	100.00%	100.00%
honeydew_melon	91.84%	100.00%	100.00%
kiwi	81.03%	94.83%	94.83%
nectarine	65.06%	100.00%	100.00%
onion	80.00%	92.00%	92.00%
orange	40.00%	91.43%	91.43%
plum	85.39%	97.75%	97.75%
spanish_pear	54.72%	100.00%	98.11%
taiti_lime	75.00%	100.00%	100.00%
watermelon	81.25%	96.88%	96.88%
总的评价识别率以及 对应的参数	73.45% (C=512, $\gamma=2$)	97.51% (C=128, $\gamma=0.03125$)	97.74% (C=512, $\gamma=0.0078125$)

②直接使用背景过度删除的图片的情况, 如表2-2所示:

表2-2: 背景过度删除法的运行结果

水果种类	颜色特征	ISADH 特征	颜色特征+ISADH 特征
agata_potato	95.59%	100.00%	100.00%
asterix_potato	83.61%	98.36%	100.00%
cashew	92.86%	100.00%	100.00%
diamond_peach	70.42%	97.18%	97.18%
fuji_apple	5.63%	77.46%	77.46%
granny_smith_apple	98.08%	100.00%	100.00%
honeydew_melon	91.84%	97.96%	97.96%
kiwi	81.03%	93.10%	93.10%
nectarine	65.06%	92.77%	92.77%
onion	80.00%	92.00%	92.00%
orange	40.00%	88.57%	88.57%
plum	85.39%	95.51%	95.51%
spanish_pear	54.72%	86.79%	86.79%
taiti_lime	75.00%	100.00%	100.00%
watermelon	81.25%	93.75%	93.75%
总的评价识别率以及 对应的参数	73.45% (C=512, $\gamma=2$)	94.24% (C=128, $\gamma=0.03125$)	94.35% (C=128, $\gamma=0.03125$)

③对图片定位后不进行任何处理的情况，如表 2-3 所示：

表 2-3：图片定位后不进行任何处理的运行结果

水果种类	颜色特征	ISADH 特征	颜色特征+ISADH 特征
agata_potato	95.59%	100.00%	100.00%
asterix_potato	83.61%	100.00%	100.00%
cashew	92.86%	100.00%	100.00%
diamond_peach	70.42%	98.59%	100.00%
fuji_apple	5.63%	98.59%	97.18%
granny_smith_apple	98.08%	96.15%	98.08%
honeydew_melon	91.84%	100.00%	100.00%
kiwi	81.03%	94.83%	98.28%
nectarine	65.06%	100.00%	100.00%
onion	80.00%	32.00%	52.00%
orange	40.00%	94.29%	91.43%
plum	85.39%	98.88%	98.88%
spanish_pear	54.72%	98.11%	98.11%
taiti_lime	75.00%	100.00%	100.00%
watermelon	81.25%	96.88%	96.88%
总的评价识别率以及 对应的参数	73.45% (C=512, $\gamma=2$)	96.61% (C=128, $\gamma=0.03125$)	97.40% (C=32, $\gamma=0.03125$)

④对图片定位后未再进行简单背景处理的情况，如表 2-4 所示：

表 2-4：图片定位后未再进行简单背景处理的运行情况

水果种类	颜色特征（未变）	ISADH 特征	颜色特征+ISADH 特征
agata_potato	95.59%	100.00%	100.00%
asterix_potato	83.61%	98.36%	98.36%
cashew	92.86%	100.00%	100.00%
diamond_peach	70.42%	98.59%	100.00%
fuji_apple	5.63%	92.96%	94.37%
granny_smith_apple	98.08%	98.08%	98.08%
honeydew_melon	91.84%	100.00%	100.00%
kiwi	81.03%	98.28%	98.28%
nectarine	65.06%	100.00%	100.00%
onion	80.00%	76.00%	84.00%
orange	40.00%	97.14%	97.14%
plum	85.39%	97.75%	97.75%
spanish_pear	54.72%	100.00%	100.00%
taiti_lime	75.00%	100.00%	100.00%
watermelon	81.25%	96.88%	95.31%
总的评价识别率以及 应的参数	73.45% (C=512, $\gamma=2$)	97.74% (C=128, $\gamma=0.03125$)	98.08% (C=512, $\gamma=0.0078125$)

⑤边缘法结合 Kmeans 的背景删除法的情况，如表 2-5 所示：

表 2-5: 边缘法结合 Kmeans 的背景删除法的运行情况

水果种类	颜色特征	ISADH 特征	颜色特征+ISADH 特征
agata_potato	95.59%	98.53%	98.53%
asterix_potato	83.61%	98.36%	98.36%
cashew	92.86%	100.00%	100.00%
diamond_peach	70.42%	97.18%	97.18%
fuji_apple	5.63%	90.14%	90.14%
granny_smith_apple	98.08%	100.00%	100.00%
honeydew_melon	91.84%	97.96%	97.96%
kiwi	81.03%	100.00%	100.00%
nectarine	65.06%	100.00%	100.00%
onion	80.00%	96.00%	96.00%
orange	40.00%	100.00%	100.00%
plum	85.39%	97.75%	97.75%
spanish_pear	54.72%	100.00%	100.00%
taiti_lime	75.00%	100.00%	100.00%
watermelon	81.25%	98.44%	96.88%
总的评价识别率以及对 应的参数	73.45% (C=512, $\gamma=2$)	98.19% (C=8, $\gamma=0.125$)	98.08% (C=8, $\gamma=0.125$)

2.4.3 实验分析

此处我们用①表示常用的Kmeans算法；②表示过度删除背景算法；③表示过度删除背景+图片重定位+未进行简单背景消除算法；④表示过度删除背景+图片重定位+再次进行简单背景删除算法；⑤表示边缘提取和Kmeans的组合算法；

实验中我们发现ISADH纹理特征展现了极好的性能，正确分类率达到94%以上，而仅适用颜色特征时，分类正确率仅为73.45%，颜色特征结合ISADH对于提高图片正确识别率的效果并不十分明显，仅提高了一个百分点左右，有时还会使得最终的识别率下降。

在仅提取ISADH特征的情况下，本文所采用的②③④⑤方法相比于传统的Kmeans方法，性能上相差无几，而算法⑤在最终的识别率上比上述的①②③④都略高，此外，我们在①②③④⑤中发现不同的预处理方法对于不同水果的分类性能上也略有差异，归纳主要原因如下：

(1) 背景选取不一，有的背景暗，有的背景亮。

当在背景较为明亮时，如图2-10(a)和(b)所示，一般情况下①和⑤算法能较好的进行背景与目标主体的分离；而②和④则会损失部分主体元素，但剩余的都是目标主体的主要部分；③的定位主要由②的情况决定。但是因为标准的Kmeans算法具有易陷入局部极小值的缺陷，不同的初始值会引起不同的结果，选择差别明显的实验数据往往能获得较好的结果，如图2-10中(b)和(c)所示，对同一图片进行两次实验，因为Kmeans算法固有的缺陷而出现较大的差别，主要是因为初始数据选取的差异（①选择图片所有像素点作为实验数据，⑤以边缘检测周围的像素点为实验数据），因此⑤相比于①具有更好的鲁棒性。

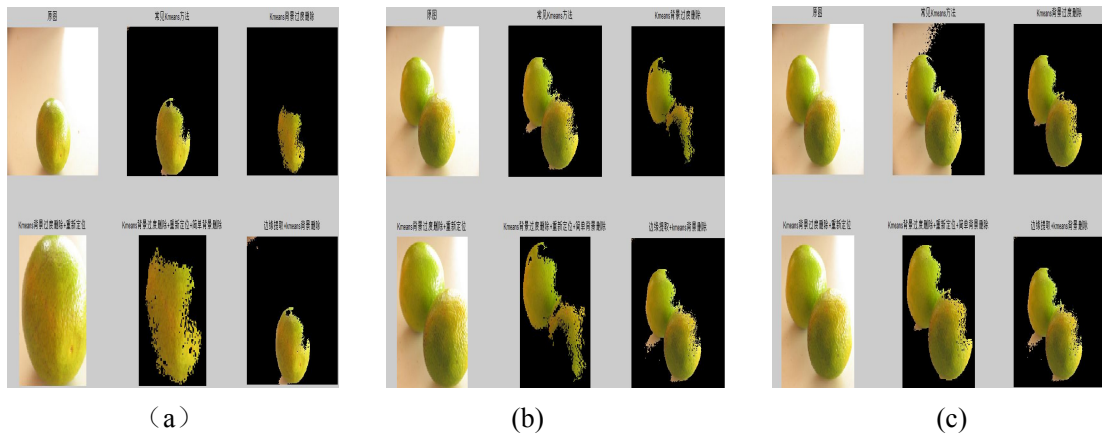


图2-10: (a) (b) (c) 背景较为明亮的Kmeans算法三种运行情况

当在背景较暗同时目标主体的影子较大时，如果2-11 (a) 和 (b) 所示，我们发现算法②和④仍能较好的提取目标主体的主要部分，而①和⑤则往往会将影子当做主体进行处理，从未引起较大误差。

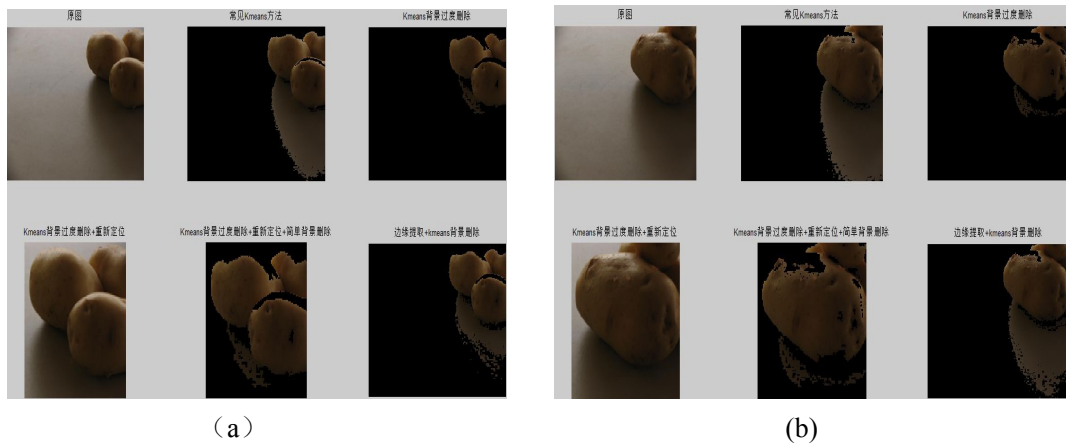


图2-11: (a) (b) 背景较暗同时目标主体的影子较大的情况

(2) 同一水果颜色差异以及颜色提取方法带来的影响。

同一种水果的颜色有很大差异，而本文采用的方法是分别提取同一水果R,G,B三个通道的颜色的均值作为特征，对于有些水果，例如orange有的部分偏黄色有的部分偏绿色，均值很难较好的描述它的颜色特性，同时颜色处理带来的问题也会很大程度上影响颜色特征与ISADH特征结合时的图片的正确分类率。

(3) 目标主体占整张图片的比例

当目标主体占整张图片的比例较大时，易获得目标主体较多的部分，纹理特征较易提取，所以识别率会相对提高。而目标主体占整张图片比例较小时，因为背景删除的原因又会损失一些目标主体，使得最终提取的特征个数较少，不具代表性。

2.5 结论

本章对比了五种预处理算法：①常用的Kmeans算法；②使用背景过度删除的图片；③对图片定位后不进行任何处理；④对图片定位后再进行简单背景删除；⑤边缘法结合

Kmeans的背景删除方法，得出以下几点结论：在简单背景下：（1）方法⑤展现了最好的识别率，在背景删除中体现了较好的鲁棒性；（2）对于目标主体较大的情况采用方法②和④进行预处理效果更好，因为此时我们能获得充足并且准确的目标主体部分以及更少的背景；（3）在目标主体占整张图片的比例不是很大而背景亮度又较好的情况下使用算法①和⑤相比于算法②和④能在背景相差不多的情况下获得更多的目标主体的部分；而当亮度较暗而且目标主体的影子也较大时采用算法①和⑤会使得删除背景的能力会大大下降，此时算法②和④效果会更好；（4）算法③的性能取决于目标主体定位的好坏程度，而背景的复杂程度也会给算法③带来不同程度的影响。

此外，ISADH纹理特征在水果分类时能展现了较好的性能，本文采用水果颜色的均值作为最终的颜色特征，却往往会因为存在同一水果含有多种颜色的情况而使得有些水果的颜色特征不具代表性。

本算法可以改进的地方（1）提出新的背景删除法，例如可以改进算法⑤，通过加上距离条件（因为目标主体距离边界区域较近），或者提出新的背景删除法以获得更好的预处理图片；（2）改进本文采用的均值表示颜色特征的方法，如采用中位数，众数等；（3）加入其它特征；（4）采用其它更加高效的机器学习算法。

第三章 基于SIFT特征的方法

3.1 引言

传统的基于水果蔬菜颜色, 纹理, 大小和形状特征的方法需要首先进行目标定位, 然后已定位的蔬菜水果进行相应特征的提取, 最后采用SVM进行训练, 分类和识别。这种方法在简单的背景下识别率较高, 但是在背景较为复杂时识别率却会大打折扣, 而我们在现实生活中对于蔬菜水果的识别时, 背景会相对复杂。

SIFT[28]特征不仅具有仿射不变性, 在图片旋转, 缩放以及光照等变化带来的影响下能保持较好的鲁棒性, 本章将文本分类的思想应用于图片分类之中[29], 采用不同的编码技术(包括VQ[29], KCB[30], LLC[31], FK[32]四种编码)结合SPM的方式[33]进行编码处理, 最后利用线性支持向量机进行训练分类识别, 在对背景较为复杂的图片的处理中获得了较为满意的识别率。

3.2 实验原理及重要概念解释

本次实验我们将文本分类技术应用于图片的分类识别问题当中。我们把从每一张图片中获得的每一个 SIFT 特征表示为一个词汇, 这样每一张图片都可表示为一篇文档, 通过训练图片获得的 SIFT 特征集合采用近似 K 均值聚类算法得到一个大小为 K 的码本字典, 然后我们把测试图片都用这些码本字典中的词汇来表示, 形成一个 K 维的向量, 第 i 维即表示第 i 个词汇在此“文档”中出现的频率, 这边我们分别采用了 Hard Assignment, KCB, LLC, FK 四种编码方式来获得词汇频率的表示, 编码完成之后, 我们再利用 SPM (Spatial Pyramid Match) 模型获取图片中的空间关系, 使用 SPM 池化之后获得每张图片的词频表示, 再使用线性的支持向量机进行最后的分类, 对于所识别的蔬菜水果的最终结果显示当地价格以及推荐与此蔬菜或水果相关的菜谱。

3.2.1 主成分分析PCA

因为本次实验过程中的SIFT特征向量拥有128维, 而对每一维进行处理都会花费较多的时间, 虽然降维意味着信息的丢失, 不过鉴于实际数据本身常常存在相关性, 能使得信息的损失尽量降低。而本文采用主成分分析法进行降维, 主成分分析通过线性变换将原始数据变换为一组各维度线性无关的数据, 可用于提取数据的主要特征分量, 常用于高维数据的降维操作。

①PCA的基本思想

降维的过程可看做是一种映射的过程, 把在高维空间中的数据映射到低维空间中, 在投影的过程中, 我们尽量使得信息最大程度的保留, PCA中利用方差对数据进行衡量, 对于数据的一个维度来说, 如果维度上数据具有很大的方差, 则说明这个维度对于数据来说具有很大的差异性, 包含了更多的信息。

此外, 如若两个维度之间无关, 即这两个维度所包含的信息没有重叠部分, 相互独立, 此时包含的信息更多, 反之, 如果两个维度的数据高度相关。如果从一个维度能推导出另外一个维度的信息, 则这两个相关的维度最多只包含一个维度的信息, 就造成了冗余, 此

时可以考虑删去其中一维，即对原数据进行降维操作。在数学上我们采用协方差的方式来衡量两个随机变量之间的相似程度，协方差为0，表示两个变量完全无关，协方差为1则表示完全相关，协方差在0-1之间则表示不同的相关程度。

因此，PCA的基本思想描述如下：

将高维空间中的点线性投影至某个低维空间中，使得投影后的点满足：

- (1) 每一个维度内的方差尽量大；
- (2) 维度之间的协方差为0，即每个维度两两正交。

假设我们拥有 $m \times n$ 的数据 X ，其中 m 表示数据的个数， n 表示数据的维数，或者特征数，

对于维度 j ，其方差为 $Var_j = \frac{1}{m} \sum_{i=1}^m (X_{ij} - \mu_j)^2$ ，而其中 μ_j 为第 j 维的期望值，可用 $\mu_j = \frac{1}{m} \sum_{i=1}^m X_{ij}$

求得，对于维度为 p 和 q 来说，它们之间的协方差为 $Cov_{pq} = \frac{1}{m} \sum_{i=1}^m (X_{ip} - \mu_p)(X_{iq} - \mu_q)$ ，对所

有数据我们可采用矩阵的形式表示，其中 $X_{ij} = X_{ij} - \mu_j$ ，经过处理之后，原来的方差和协

方差可表示为： $Var_j = \frac{1}{m} \sum_{i=1}^m X_{ij}^2$ ， $Cov_{pq} = \frac{1}{m} \sum_{i=1}^m X_{ip} X_{iq}$ ，对于所有的数据我们可得到如下表示：

$$X^T X = \begin{bmatrix} \sum_{i=1}^m X_{i1}^2 & \dots & \dots & \sum_{i=1}^m X_{i1} X_{in} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \sum_{i=1}^m X_{in} X_{i1} & \dots & \dots & \sum_{i=1}^m X_{in}^2 \end{bmatrix}$$

$C = \frac{1}{m} X^T X$ 是一个对称阵，其对角线上的元素即为每一维的方差，其它元素就是对应维度的协方差， C 即为 X 的协方差矩阵。我们令 $Y = XP$ ，其中 D 为 Y 的协方差矩阵，我们做如下推导可得：

$$\begin{aligned} D &= \frac{1}{m} Y^T Y \\ &= \frac{1}{m} (XP)^T (XP) \\ &= \frac{1}{m} P^T X^T X P \\ &= P^T \left(\frac{1}{m} X^T X \right) P \\ &= P^T C P \end{aligned}$$

由此我们只需要找到一个矩阵 P ，使得 $P^T C P$ 为1个对角矩阵，除了对角线以外的元素则均为0，我们让对角线元素按从大到小依次排列，并选取方差较大的维度。

对于 $D = P^T C P$ 的问题，我们可以采用矩阵论中的知识进行分解，即对于 C 来说，我们

可以找到单位向量 $E = (e_1, e_2, \dots, e_n)$ ，使得：

$$E^T C E = \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \dots & \\ & & & \lambda_n \end{bmatrix}$$

其中 Λ 为对角矩阵，对角元素为每个特征向量对应的特征值，特征值的大小决定了此特征对于整体的影响的大小。

②PCA的算法流程：

设有 m 条 n 维数据。

- 1) 将原始数据按列组成 n 行 m 列矩阵 X ；
- 2) 将 X 的每一行（代表一个属性字段）进行零均值化，即减去这一行的均值；

$$C = \frac{1}{m} X^T X$$

- 3) 求出协方差矩阵 C ；
- 4) 求出协方差矩阵的特征值及对应的特征向量；
- 5) 将特征向量按对应特征值大小从上到下按行排列成矩阵，取前 k 行组成矩阵 P ；
- 6) $Y = PX$ 即为降维到 k 维后的数据。

③PCA算法的分析

PCA本质上是将方差最大的方向作为主要特征，并且在各个正交方向上将数据“离相关”，也就是让它们在不同正交方向上没有相关性。

PCA也存在一些限制，例如它可以很好的解除线性相关，但是对于高阶相关性就没有办法了，对于存在高阶相关性的数据，可以考虑Kernel PCA，通过Kernel函数将非线性相关转为线性相关。另外，PCA假设数据各主特征是分布在正交方向上，如果在非正交方向上存在几个方差较大的方向，PCA的效果就大打折扣了。PCA是一种无参数技术，也就是说面对同样的数据，如果不考虑清洗，谁来做结果都一样，没有主观参数的介入，所以PCA便于通用实现，这也是本文采用PCA的最主要的原因。

3.2.2 kd-tree

kd-tree:kd树是每个节点都为 k 维点的二叉树。所有非叶子节点可以视作用一个超平面把空间分区成两个半空间。节点左边的子树代表在超平面左边的点，节点右边的子树代表在超平面右边的点。选择超平面的方法如下：每个节点都与 k 维中垂直于超平面的那一维有关。因此，如果选择按照 x 轴划分，所有 x 值小于指定值的节点都会出现在左子树，所有 x 值大于指定值的节点都会出现在右子树。这样，超平面可以用该 x 值来确定，其法矢为 x 轴的单位矢量。

① kd树的构造

变量名	数据类型	描述
Node-data	数据矢量	数据集中某个数据点，是 n 维矢量（这里也就是 k 维）
Range	空间矢量	该节点所代表的空间范围
split	整数	垂直于分割超平面的方向轴序号
Left	k-d 树	由位于该节点分割超平面左子空间内所有数据点所构成的 k-d 树
Right	k-d 树	由位于该节点分割超平面右子空间内所有数据点所构成的 k-d 树
parent	k-d 树	父节点

构建k-d树伪代码：

输入：数据点集Data-set和其所在的空间Range

输出：Kd，类型为k-d tree

1.If Data-set为空，则返回空的k-d tree

2.调用节点生成程序：

（1）确定split域：对于所有描述子数据（特征矢量），统计它们在每个维上的数据方差。以SIFT特征为例，描述子为128维，可计算128个方差。挑选出最大值，对应的维就是split域的值。数据方差大表明沿该坐标轴方向上的数据分散得比较开，在这个方向上进行数据分割有较好的分辨率；

（2）确定Node-data域：数据点集Data-set按其第split域的值排序。位于正中间的那个数据点被选为Node-data。此时新的Data-set' = Data-set \ Node-data（除去其中Node-data这一点）。

3.dataleft = {d属于Data-set' && d[split] ≤ Node-data[split]}

Left_Range = {Range && dataleft}

dataright = {d属于Data-set' && d[split] > Node-data[split]}

Right_Range = {Range && dataright}

4.left = 由（dataleft, Left_Range）建立的k-d tree，即递归调用createKDTree（dataleft, Left_Range）。并设置left的parent域为Kd；

right = 由（dataright, Right_Range）建立的k-d tree，即调用createKDTree（dataright, Right_Range）。并设置right的parent域为Kd。

② k-d树上的最邻近查找算法

输入：k-d tree类型Kd

查询数据点target

输出：最邻近数据点nearest

最邻近数据点和查询点间的距离dist

1.If Kd为NULL，则设dist为infinite并返回

2. 进行二叉查找，生成搜索路径

```

Kd_point = &Kd;                                //Kd-point中保存k-d tree根节点地址
nearest = Kd_point -> Node-data;                //初始化最近邻点
while (Kd_point)
    push (Kd_point) 到search_path中;            //search_path是一个堆栈结构，存储着搜索
索路径节点指针
    s = Kd_point -> split;                        //确定待分割的方向
    If target[s] <= Kd_point -> Node-data[s]    //target在左侧，现在需要往右侧进行搜索
        Kd_point = Kd_point -> left;
    Else                                          //target在右侧，现在需要往左侧进行搜索
        Kd_point = Kd_point -> right;
nearest = search_path中最后一个叶子节点;
Max_dist = Dist (nearest, target);              //直接取最后叶子节点作为回溯前的初始
最近邻点

```

3. //回溯查找

```

while (search_path != NULL)
    back_point = 从search_path取出一个节点指针;  //从search_path堆栈弹栈
    s = back_point -> split;                      //确定分割方向
    If Dist (target[s], back_point -> Node-data[s]) < Max_dist //判断还需进入的子空
间，如果小于目标值点与当前搜索点的距离，也就是说以目标点为圆心，当前最小距离为
半径的圆与分割线相交，则需要对分割线另外一边的点进行搜索
        If target[s] <= back_point -> Node-data[s]
            Kd_point = back_point -> right;    //如果target位于左子空间，就应进入右子
空间
        else
            Kd_point = back_point -> left;    //如果target位于右子空间，就应进入左子空
间
        将Kd_point压入search_path堆栈;
    If Dist (nearest, target) > Dist (Kd_Point -> Node-data, target)
        nearest = Kd_point -> Node-data;      //更新最近邻点
        Min_dist = Dist (Kd_point -> Node-data, target); //更新最近邻点与查询点间
的距离

```

③ k-d树性能分析

由k-d树上的最邻近查找算法可知，当目标点的邻域与分割超平面两侧的空间都产生交集时，需要查找另一侧子空间，回溯的次数大大增加。导致检索过程复杂，效率下降。一般情况下，最近邻搜索只需要检测几个叶子节点即可，如图3-1所示：

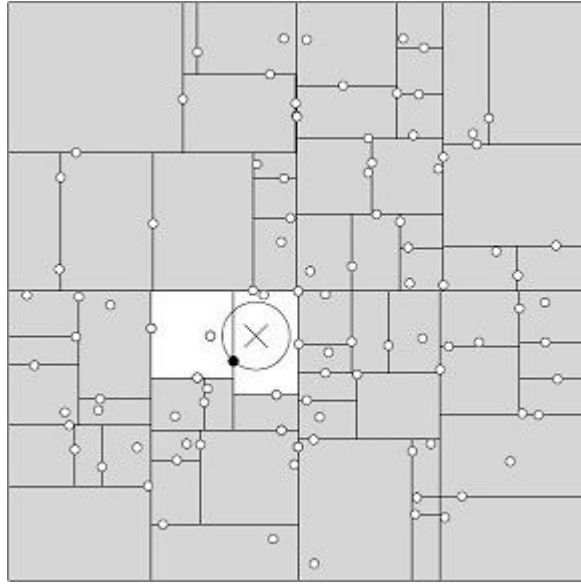


图 3-1: kd树一般情况

但是如果当实例点的分布较为糟糕时，几乎需要遍历所有的节点，如图3-2所示：

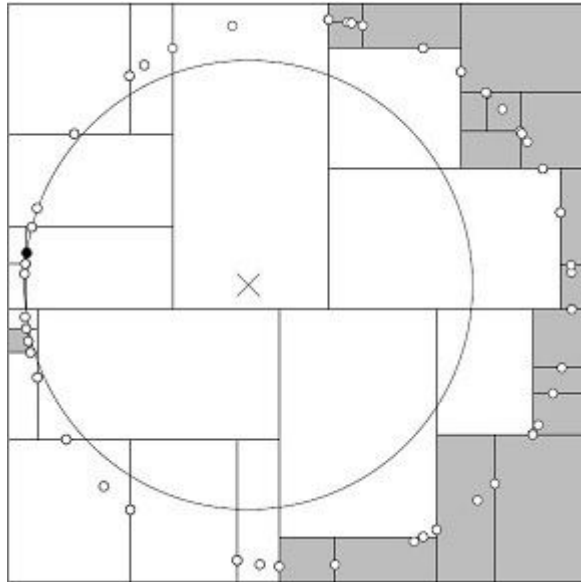


图3-2: kd树实例点分布较为糟糕的情况

研究表明 N 个节点的 k - d 树搜索过程时间复杂度最坏的情况下是 $t_{worst} = O(k \cdot N^{1-\frac{1}{k}})$ ，如果实例是随机分布的， k - d 树搜索过程的平均复杂度是 $O(\log N)$ 。 k - d 树更适用于训练实例 $N \gg 2k$ 的情况，因此空间维数不能太高，一般不大于20维，当空间维数接近实例数时，它的效率会迅速下降，几乎接近贪婪线性扫描。由于大量回溯会导致 k - d 树最近邻搜索的性能大大下降，所以我们介绍本实验采用的改进的 k - d 树最近邻搜索算法Best-Bin-First。

④ Best-Bin-First方法

BBF的查询思路就是将“查询路径”上的节点进行排序，如按各自分割超平面（称为Bin）与查询点的距离排序。回溯检查总是从优先级最高的（Best Bin）的树节点开始。另外BBF还设置了一个运行超时限制，当优先级队列中的所有节点都经过检查或者超出时间

限制时，算法返回当前找到的最好结果作为近似的最近邻,具体代码可以参见[24]。

3.2.3 Hard Assignment

每次从码本字典中获取与当前词汇距离最近的词汇近似表示此词汇，此方法最大的优点是简单易懂，但也往往会出现两种缺陷：码字的不确定性和码字的说服力。如图 2-3 所示，对于图中 h, i 两点之间的正方形若按照传统硬赋值进行操作则仅会选择 i 和 h 中的一个作为正方形的值，但是它距离 h 和 i 的距离相差不大，换句话说就是 h, i 与正方形都相关，这就是码字不确定性；而至于码字可行性，如图中距离 c 点的菱形，它距离 c 点相对较远，即相关性不大，但传统的硬赋值方法还是会将其赋值为 c ，这就不具有说服力。

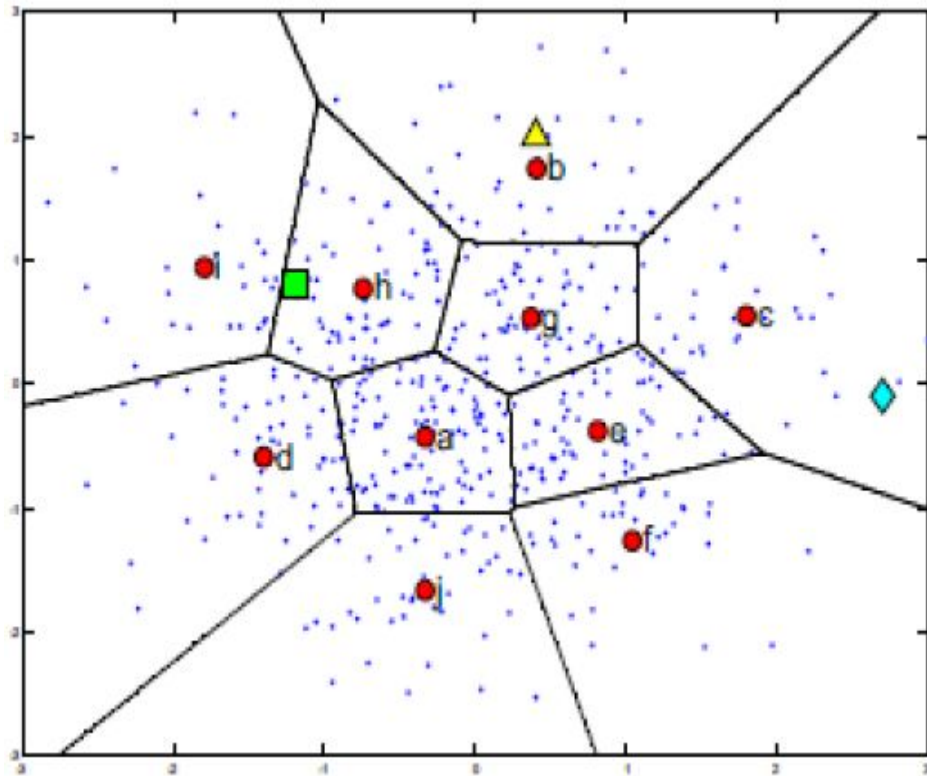


图3-3: Hard Assignment

KCB,LLC 的基础为软赋值 Soft Assignment。

3.2.4 Soft Assignment

传统的矢量化方法是将图片中的高维的区域描述算子映射到词汇表中的单个视觉词汇，但往往因为图片噪声，光照的影响，非仿射变化等都会产生特征描述子的可变性，这种方法便会带来很大的误差。所以我们创新性地采用描述子空间软赋值和图像空间软赋值的方法(即对于测试的任意一个描述算子都用 r 个最近邻的词向量进行表示，软赋值和硬赋值的对比如图 3-4 所示，其中 A,B,C,D,E 表示聚类中心，而 1,2,3,4 表示特征，在硬赋值中，特征 3 和 4 将永远都不会被匹配，因为它们会分别被分到 A 和 C，尽管它们在空间上很邻近，而在软赋值中，特征 3,4 将会以不同的权值被赋值为 A,B,C，这样就可以较好的进行匹配。利用加权的视觉词汇的组合来描述图片。实验结果也显示，描述算子空间软赋值相

比于硬赋值，尽管软赋值会占用更多的存储空间，需要更多的计算，但是却大大提高了分类的性能。

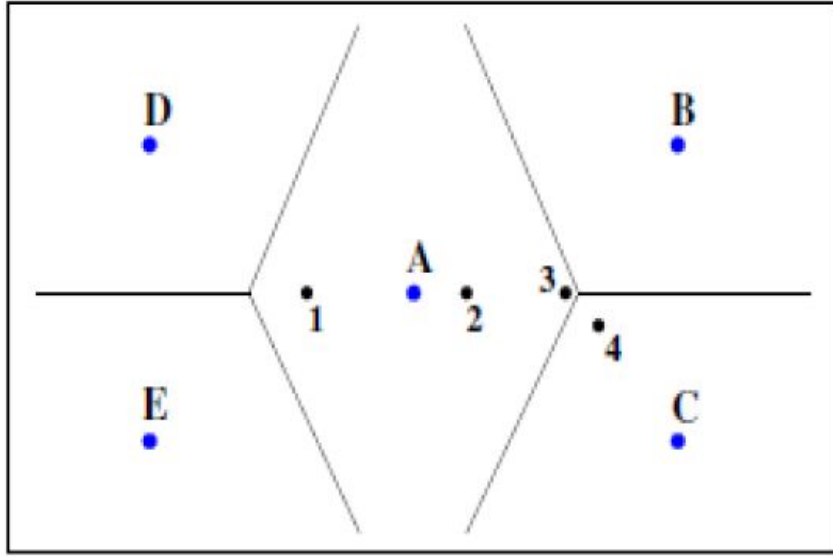


图 3-4: Soft Assignment

3.2.5 KCB (Kernel Codebook)

硬赋值 Hard Assignment 拥有两个缺陷：码字的不确定性和码字的说服力。上述的采用软赋值的方法可以较好地解决此类问题，而此处的 KCB 方法则是在软赋值的基础上提出的一种新的加权表示方式，即利用核密度函数 $K_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{x^2}{2\sigma^2})$ 对图片特征进行加权，其中 x 表示特征与码本字典中词汇之间的距离，最后我们再把最近的 K 个词汇通过核密度函数计算得到的值分别赋予码本字典中的相应词汇，核密度函数加权的方式不仅保留了特征之间的模糊关系，同时也提高了模型的鲁棒性。

3.2.6 LLC(Locality-constrained Linear Coding)

相比于传统的基于特征包 (BoF) 的 SPM 方法需要非线性分类器获得较好的分类性能，且需要额外的计算代价。这边提出的 LLC 方法，它可看做是 LCC 的快速实现，通过利用区域限制将每个描述子投影到它的局部坐标系统中。对比三种区域约束线性编码 (1) VQ (即硬赋值) (2) ScSPM (3) LLC。

我们假设：

X:从一张图片中提取的 D 维的局部描述子， $X=[x_1, x_2, \dots, x_N] \in R^{D \times N}$

B:带有 M 个词汇的码本， $B=[b_1, b_2, \dots, b_M] \in R^{D \times M}$

①使用 VQ 编码描述:是求解满足 (1) 式的约束最小平方问题：

$$\begin{aligned} \arg \min_C \sum_{i=1}^N \|x_i - Bc_i\|^2 \\ s.t. \|c_i\|_0 = 1, \|c_i\|_1 = 1, c_i \geq 0, \forall i \end{aligned} \quad (1)$$

其中 $C=[c_1, c_2, \dots, c_N]$ 是 X 的码字集合， $\|c_i\|_0 = 1$ 表示在所有码字 c_i 中仅有一个非 0 元素，

非负的 l^1 约束 $\|c_i\|_1=1, c_i \geq 0$ 表示对于 x 的编码的权重是1，也就是只从码本中选取一个码本词汇进行表示。

②ScSPM 编码描述:为了减少(1)带来的VQ编码带来的量化损失，我们将(1)中的基数约束 $\|c_i\|_0=1$ 使用稀疏规则化的项进行松弛，这边我们用 c_i 的 l^1 范数，同时对每个局部描述 x_i 进行编码，因此问题转化为了一个标准的稀疏编码(SC)问题：

$$\arg \min_C \sum_{i=1}^N \|x_i - Bc_i\|^2 + \lambda \|c_i\|_1 \quad (2)$$

稀疏规则化项在此的重要性表现为：码本 B 通常是过完备的，也就是说 $M>D$ ，因此 l^1 规则化对于保证欠定系统具有唯一解是必要的；稀疏能保证已知的表示获得局部描述子的显著模式；稀疏编码可以获得比VQ更少的量化错误。

③LLC 编码描述：在LCC中已经提到，局部就必将会产生稀疏性，LLC方法使用局部约束替代了稀疏约束，LLC使用如下准则：

$$\begin{aligned} \arg \min_C \sum_{i=1}^N \|x_i - Bc_i\|^2 + \lambda \|d_i \otimes c_i\|^2 \\ s.t. 1^T c_i = 1, \forall i \end{aligned} \quad (3)$$

其中 \otimes 表示点乘， $d_i = \exp(\frac{\text{dist}(x_i, B)}{\sigma})$ ， $\text{dist}(x_i, B) = [\text{dist}(x_i, b_1), \dots, \text{dist}(x_i, b_M)]^T$ ， $\text{dist}(x_i, b_j)$ 表示 x_i 与 b_j 之间的欧几里得距离， σ 是用来调整权值下降速度的局部参数，通常我们还会对 d_i 进行规范化至区间 $(0,1]$ 之间，通过减去 $\max(\text{dist}(x_i, B))$ 从 $\text{dist}(x_i, B)$ ，约束 $1^T c_i = 1$ 保持着LLC编码的平移不变要求，注意等式(3)中的LLC编码在 l^0 范数上是不稀疏的，但是在解只有极少重要的值的意义上它是稀疏的，而对于(3)，我们可以获得如下的解析解：

$$\begin{aligned} \tilde{c}_i &= (C_i + \lambda \text{diag}(d)) \setminus 1 \\ c_i &= \tilde{c}_i / 1^T \tilde{c}_i \end{aligned} \quad (4)$$

通过比较我们可以发现LLC具有明显的三大优势(1)LLC相比于VQ具有更好的重构性，VQ编码忽略了不同基之间的关系使得对于相似的描述算子的编码可能不同，而在LLC中每个描述算子都由多个基进行表示，通过分享基可获得相似的描述算子之间的关系，从而能更好的进行重构；(2)局部平滑稀疏，SC与LLC都可以获得相对较好的重构性，但因为SC码本的过完备性，对于相同的包，SC可能会选取完全不同的基，因为会损失码本之间的相关性。LLC在此情况下确保了相似包能获得相似的码字，保证了码本之间的相关性。(3)LLC可获得解析解，从而大大降低了计算的代价，提高了计算速度。(三种编码方式的比较也可参考图3-5)

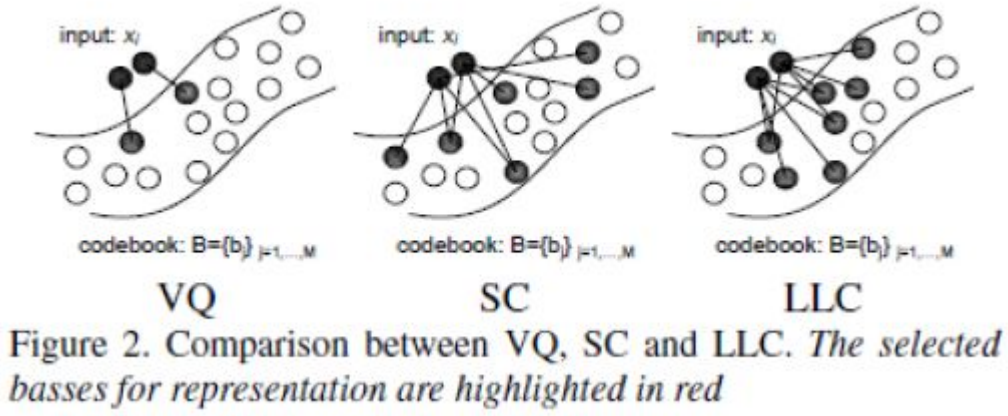


图 3-5: VQ, SC, LLC 的情况

3.2.7 Fisher Kernel

① Fisher Kernel 简介

Fisher Kernel: 我们让 $X = \{x_t, t=1, \dots, T\}$ 为 T 个局部描述子, $x_t \in \mathcal{X}$, 假设 u_λ 为在 \mathcal{X} 中元素的生成过程的概率密度函数, 其中 $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_M]^T \in R^M$, 表示 u_λ 中带有 M 个参数的向量, 在统计学中, 计分函数由模型上的数据对数似然函数的梯度给出:

$$G_\lambda^X = \nabla_\lambda \log u_\lambda(X) \quad (5)$$

这个梯度描述了单个参数对于生成过程的贡献, 换句话说, 它描述了生成模型参数 u_λ 需要如何进行调整以便更好的拟合数据 X 。其中 $G_\lambda^X \in R^M$, 因此 G_λ^X 的维度仅仅依赖于参数在 λ 中的参数 M 的个数而不是样本大小 T 。

从信息集合理论中, 一个贡献的参数族 $\nu = \{u_\lambda, \lambda \in \Lambda\}$ 可以被认为是一个由 FIM $F_\lambda \in R^{M \times M}$ 给出的带有局部度量标准的黎曼流型 M_Λ :

$$F_\lambda = E_{x \sim u_\lambda} [G_\lambda^X G_\lambda^{X'}] \quad (6)$$

对此, Jaakkola 和 Haussler 提出使用 FK 来评价两个样本 X 和 Y 的相似度, 其中 FK 被定义为:

$$K_{FK}(X, Y) = G_\lambda^{X'} F_\lambda^{-1} G_\lambda^Y \quad (7)$$

因为 F_λ 是正半定的, 所以它存在矩阵形式, 使用 Cholesky 分解, $F_\lambda^{-1} = L_\lambda^T L_\lambda$, 则 (7) 中的 FK 式子可以重新表示为

$$K_{FK}(X, Y) = \varphi_\lambda^{X'} \varphi_\lambda^Y \quad (8)$$

其中

$$\varphi_\lambda^X = L_\lambda G_\lambda^X = L_\lambda \nabla_\lambda \log u_\lambda(X) \quad (9)$$

我们就称这个关于 X 的规范化的梯度向量为 FV, FV 的维度等于梯度向量 G_λ^X , 使用 K_{FK} 作为核的非线性的学习机等价于使用 φ_λ^X 作为特征向量的线性核学习机。而使用 FK 作为核的一个明显优势也就是它可以使用线性分类器高效地进行学习。

② 改进的 Fisher Kernel:

因为由 Jaakkola 和 Haussler 提出的 FK 的方法不受每个视觉词的出现次数的限制, 而且它还能将描述算子之间的额外分布信息进行编码, 扩展了 BOV 方法 (传统的 BOV 方法会受限对于多通道的特征提取时的高计算代价, 非线性 SVM 学习的计算复杂度在 $O(N^2)$ 和 $O(N^3)$ 之间, 不适用于 N 较大的情况)。所以采用 FK 方法主要优势为: 它仅仅使用 SIFT 描述算子和计算代价较小的线性分类器。本实验使用的 FK 方法又在传统 FK 方法上进行了 3 点改进

(1) 采用 L2 进行规范化, 使用 $\frac{K(X,Y)}{K(X,X)K(Y,Y)}$ 替代原先的 $K(X,Y) = G_\lambda^x F_\lambda^{-1} G_\lambda^y$;

(2) 幂规范化, 实践经验显示, 高斯个数增长时, Fisher 向量会变得越加稀疏, 而 L2 距离在稀疏向量的相似性上是一个较差的测量指标, 针对此问题, 可采用两种改进方法, 一是使用在稀疏向量上鲁棒性更好的核替代点积运算, 但这往往需要付出非线性分类的计算代价。另一种则是“unsparsify”表示以保证点积的相似度。本实验采用函数 $f(z) = \text{sign}(z) \sqrt{|z|}$ 用以“unsparsify”原始表示; (3) 空间金字塔, 利用空间金字塔获得图片间空间信息 (关于 SPM 的详细信息具体参考下方), 本实验还将 FK 方法和 SPM 方法结合。

3.2.8 SPM

① 原始的 SPM 简介

SPM 是一种基于 implicit correspondence 对无序, 不等长的向量集合的高效核函数, 其基本思想是将无序的特征集映射到多解析直方图上, 然后对不同解析度上的直方图进行加权, 最后图片匹配时做交集计算。金字塔匹配在特征个数上计算是线性的, 它对于不同解析度上的直方图寻找匹配, 而且对于额外的不匹配特征不进行惩罚, 所以聚类的鲁棒性强, 同时金字塔核是正定的。 $K_\Delta(\psi(y), \psi(z)) = \sum_{i=0}^L w_i N_i$, 其中 w_i 表示在第 i 层时的权值, N_i 表示在第 i 层时新增加的特征匹配个数, $N_i = I(H_i(y) - H_i(z)) - I(H_{i-1}(y) - H_{i-1}(z))$;

$I(A, B) = \sum_{j=0}^r \min(A^{(j)}, B^{(j)})$, $I(A, B)$ 表示 A, B 的特征匹配个数, 此操作可参考图 3-6。

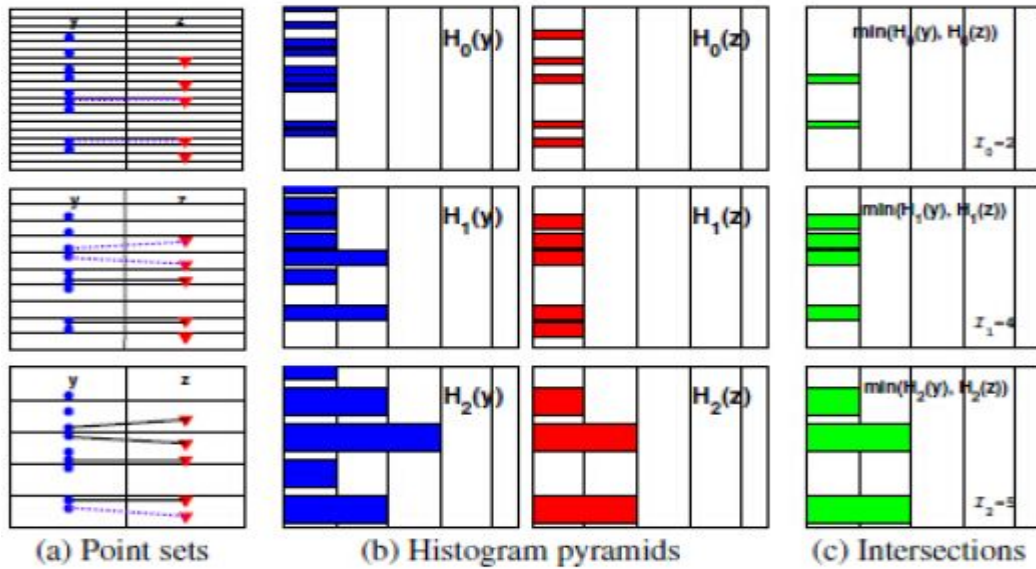


图 3-6: 两个 1 维的特征集被用来构造两个直方图金字塔，每一行对应一个金字塔层，在 (a) 中，集合 y 在左边，集合 z 在右边，轻的点线表示的是 bin 边界，加粗的虚线表示了在此层形成的匹配对，加粗的实线表示在此解析层之前已经形成的匹配对。在 (b) 中显示的是多解析直方图，水平轴是 bin 个数。在 (c) 中，是 (b) 中的金字塔交集。

② 本实验采用的 SPM

原 SPM 方法丢弃了所有的空间信息，所以本实验在原 SPM 的基础上采用了另外的正交方法，将所有特征向量量化至 M 个离散的类中，并假设只有相同类型(即同一个离散类中)的特征才可以互相匹配，每个通道(channel)给出两个二维向量的集合，最终的核的计算结果则是这些分散通道核的总和。核函数并未改变，只是增加了新的规则，即只有相同类型(即同一个离散类中)的特征才可以互相匹配，这种方法在视觉词汇上保持了连续性，此方法简单并且不需要构造明显的目标模型，使用全局 cues 作为物体表示的间接信息，能在无序的图片表示上获得了极大的改善。

其中关于支持向量机 SVM。

3.2.9 线性支持向量机 Linear Support Vector Machine

线性支持向量机算法是较好的有监督学习算法,首先我们可以看图 3-7,假设中间那条线是 $y=f(x)$,考虑上面 3 个点 A、B 和 C。从图中我们可以确定 A 是 \times 类别的, B 我们也能够确定是属于 \times 类的,然而 C 因为离分割线较近,我们较难确定其属于哪一类。所以我们更应该关心靠近中间分割线的点,让他们尽可能地远离中间分割线以便于我们更好的进行分类,而忽略那些离中间分割线较远的点,即我们不关心已经确定远离的点,因为那样的话,要使得一部分点靠近中间线来换取另外一部分点更加远离中间线。

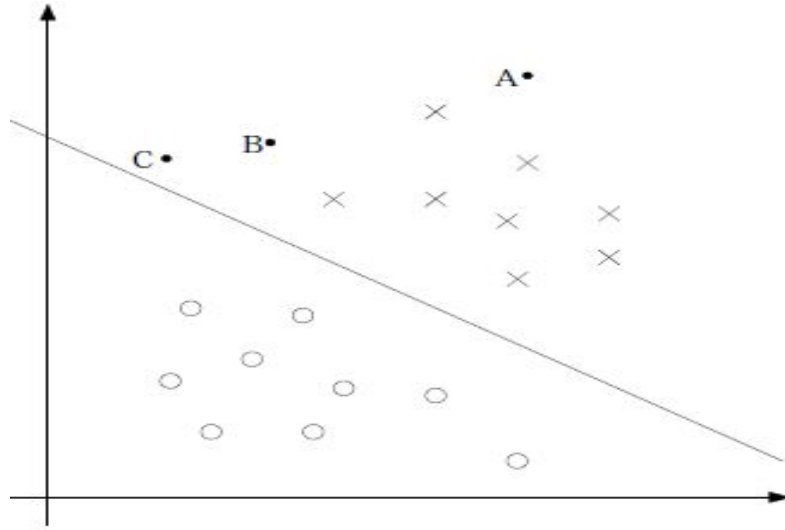


图 3-7: 分类点的分布情况

而 SVM 的基本思想可参见图 3-8, 我们希望能够使得黑 O 和白 O 两类能够都尽量远离中间的分割线, 获得最大的间隔 **Margin**。我们将每个训练数据都用 $\{x_i, y_i\}, i=1, \dots, l, y_i \in \{-1, 1\}, x_i \in R^d$, 假设我们有一分割超平面可以直接分割正负样本, 而在分割超平面上的点满足 $w \bullet x + b = 0$, w 与分割超平面垂直, $|b|/\|w\|$ 即为从超平面到原点的垂直距离, 其中 $\|w\|$ 是 w 的欧几里得范数, $d_+(d_-)$ 分别表示正负样本点到分割超平面的最短距离, 则 $\text{Margin} = d_+ + d_-$, 而对于线性支持向量机来说, 我们需要找到带有最大 **Margin** 的分割超平面, 为了解决这个问题, 我们假设所有训练数据满足如下约束:

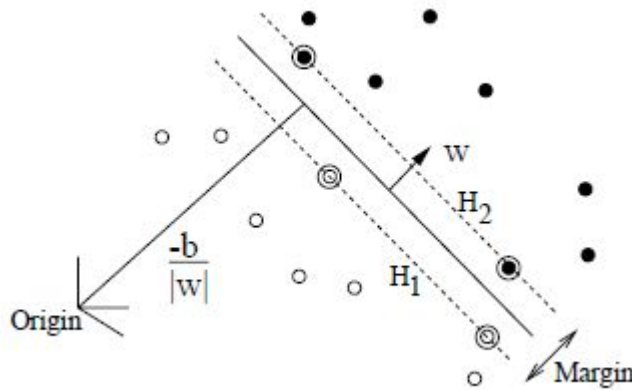


图 3-8: 最大 Margin 的分割超平面

$$w \bullet x + b \geq +1, y_i = +1 \quad (10)$$

$$w \bullet x + b \leq -1, y_i = -1 \quad (11)$$

而 (10) 和 (11) 又可以用如下不等式表示:

$$y_i(w \bullet x + b) - 1 \geq 0, \forall i \quad (12)$$

现在我们满足求等式 (10) 的训练样本点, 这些点在超平面 H_1 上: $x_i \bullet w + b = 1$, 这个超平面与原点 **Origin** 的垂直距离为 $|1-b|/\|w\|$, 相似的超平面 H_2 到原点 **Origin** 的距离为 $|-1-b|/\|w\|$, 所以 $d_+ = d_- = 1/\|w\|$, 而 H_1 与 H_2 之间的距离即为 $2/\|w\|$, 这边 H_1 与 H_2 是平行

的而且没有训练点落在二者中间，而我们需要的是尽可能使得 $2/\|w\|$ 变大。而这恰巧等价于在满足约束条件（12）的前提下求 $\|w\|^2$ 的最小值，而在 H_1 与 H_2 上的点则称作是支持向量，因为它们的变化会改变最终的解，目前我们的主要工作就是求解如下最优问题：

$$\begin{aligned} \min & \frac{1}{2} \|w\|^2 \\ \text{s.t.} & 1 - y_i(w \bullet x + b) \leq 0, \forall i \end{aligned} \quad (13)$$

这边引入 $\frac{1}{2}$ 的目的是方便后面求解，因为目标函数与约束条件满足凸性，所以可以通过构造 Lagrange Dual 函数，并可通过求解 Lagrange Dual 问题获得最优解，把问题转化为 Lagrange Dual 的主要目的则是为了①约束条件（12）可以被 Lagrange 乘子替代，另外一点就是在新的形式中，训练数据只会以向量内积的形式存在，这可以使我们将问题泛化到非线性的例子中（即非线性核函数，但因为本次实验并未使用到非线性核函数，所以不提，有兴趣可以参考[34]），我们引入 Lagrange 乘子 $\alpha_i (\alpha_i \geq 0)$ 于是我们得到：

$$L_p = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i y_i (x_i \bullet w + b) + \sum_{i=1}^l \alpha_i \quad (14)$$

此时我们需要求解这个二次规划问题。我们可以通过对偶问题的几何描述得到， L_p 的最大值和 L_p 的最小值会在相同的 w, b 和 α 上取得。

最小化关于 w 和 b 的 L_p 函数，同时需要 L_p 中所有关于 α_i 的导数为 0，这些都满足约束 $\alpha_i \geq 0$ ，分别对 w 和 b 求导并另导数为 0 我们得到：

$$w = \sum_i \alpha_i y_i x_i \quad (15)$$

$$\sum_i \alpha_i y_i = 0 \quad (16)$$

代入（4）可得到：

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \bullet x_j \quad (17)$$

支持向量训练等价于求 L_D 在满足约束条件（16）和 $\alpha_i \geq 0$ 下关于 α_i 的最大值，在求得的解中，如果 $\alpha_i > 0$ ，那么对应的点则成为支持向量，这些向量在 H_1 或者 H_2 上，其余的训练点则对应 $\alpha_i = 0$ ，这些都可以用 KKT 条件来描述，对于上述问题，对应的 KKT 条件为：

$$\frac{\partial}{\partial w} L_p = w_v - \sum_i \alpha_i y_i x_{iv} = 0, \alpha = 1 \dots d \quad (18)$$

$$\frac{\partial}{\partial b} L_p = -\sum_i \alpha_i y_i = 0 \quad (19)$$

$$y_i(x_i \bullet w + b) - 1 \geq 0, i = 1 \dots l \quad (20)$$

$$\alpha_i \geq 0, \forall i \quad (21)$$

$$\alpha_i(y_i(x_i \bullet w + b) - 1) = 0, \forall i \quad (22)$$

在凸优化问题中，KKT 条件和 w, b, α 是原问题的解二者互为充分必要条件，因此求解

SVM 问题就等价于寻找 KKT 条件的解，这样我们即易求得 w 和 b (b 一般取所有满足等式 (22) 的均值)。

刚刚是在理想情况下能恰好进行完全分类，但是正如图 3-9 所示：

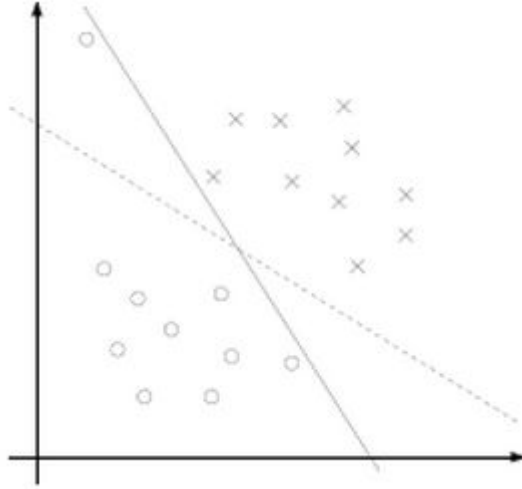


图 3-9：理想情况

我们会因为一个奇异点（可能来自噪声等）而使得我们的分割超平面产生较大的变化，或者有时也会出现线性不可分的情况，这往往就会造成过拟合的情况，过拟合的定义为：给定一个假设空间 H ，一个假设 h 属于 H ，如果存在其他的假设 h' 属于 H ，使得在训练样例上 h 的错误率比 h' 小，但在整个实例分布上 h' 比 h 的错误率小。

针对上述问题，我们采用的方法是引入新变量 ζ_i 对不等式 (10) 和 (11) 进行松弛处理，于是我们得到：

$$w \bullet x + b \geq +1 - \zeta_i, y_i = +1 \quad (23)$$

$$w \bullet x + b \geq +1 + \zeta_i, y_i = -1 \quad (24)$$

$$\zeta_i \geq 0, \forall i \quad (25)$$

所以我们的问题转变为：

$$\begin{aligned} \min & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \zeta_i \\ \text{s.t.} & y_i(w \bullet x_i + b) \geq 1 - \zeta_i, i = 1, \dots, m \\ & \zeta_i \geq 0, i = 1, \dots, m \end{aligned} \quad (26)$$

这边 C 是惩罚函数（主要由用户选择），用来控制 ζ_i 带来的误差，同样对于 (26) 我们也可以采用 KKT 条件进行求解，最终我们只需求解下列情况即可：

$$\frac{\partial}{\partial w} L_P = w_v - \sum_i \alpha_i y_i x_{iv} = 0, \alpha = 1 \dots d \quad (27)$$

$$\frac{\partial}{\partial b} L_P = -\sum_i \alpha_i y_i = 0 \quad (28)$$

$$\frac{\partial}{\partial \zeta_i} L_P = C - \alpha_i - \mu_i \quad (29)$$

$$y_i(x_i \bullet w + b) - 1 + \zeta_i \geq 0, i = 1 \dots l \quad (30)$$

$$\zeta_i \geq 0 \quad (31)$$

$$\alpha_i \geq 0 \quad (32)$$

$$\mu_i \geq 0 \quad (33)$$

$$\alpha_i \{y_i(x_i \bullet w + b) - 1 + \zeta_i\} = 0 \quad (34)$$

$$\mu_i \zeta_i = 0 \quad (35)$$

通过求解获得的 w, b ，对于测试数据，我们仅需求解 $w \bullet x + b$ 的值即可，通过值的正负性来判断它的最终分类情况。

3.3.实验方法和步骤

3.3.1-3.3.7 是蔬菜水果分类识别部分，其中 3.3.1-3.3.6 部分主要是 Vqdemo, Kcbdemo, LLCdemo 的实验步骤部分, 3.3.7 部分指出 FKdemo 与前三种方法的不同之处, 3.3.8 是根据前面的识别结果给出当日的菜价以及相关菜谱的推荐部分。

3.3.1 构造imdb

采用 PreparePic 函数对训练图片和测试图片进行预处理，将训练图片的索引值，测试图片的索引值，训练图片和测试图片的标签（即所属类别）分别进行存储，用于后续处理。

3.3.2 提取SIFT特征

SIFT (Scale-invariant feature transform) 特征是由 David G. Lowe 提出的具有尺度不变性的特征描述子，它通过求解一张图片中的兴趣点及其有关的尺度和方向描述子得到。SIFT 特征除了拥有尺度不变特征，在图片旋转，图像亮度和拍摄视角改变的情况下利用 SIFT 特征进行图片匹配也能得到较好的效果。

SIFT 特征的提取过程为：

- a. 对图片进行灰度化，再利用归一化操作去除因光照而使图像灰度值发生 scale 变化带来的影响；
 - b. 使用 Gaussian 函数构建尺度空间 (Scale-Space)，并建立图像金字塔，利用图像金字塔获得 DoG (Difference of Gaussian)，对于 DoG 进行局部极大极小值检测，移除靠近边界的关键点，利用阈值强度移除边界关键点；
 - c. 方向的计算；
 - d. 计算获得 128 维的 SIFT 特征（选取 128 维是由实验结果而得）
- 这样每一张图片就可以表示为由多个 128 维的 SIFT 特征组成的集合。

3.3.3 构造码本字典

对训练集中的图片提取 SIFT 特征，获得所有这些图片的 SIFT 特征集合，因每张图片都会产生大量的 SIFT 特征，所以我们从中随机选取 100000 个特征采用近似最近邻 K-means 算法（先从所有的图片特征中随机取 4000 个构造初始字典中的词汇，构造 Best Bin First kd

树，在 Best Bin First kd 树上进行相应的查找，删去不符合条件的字典中的词，这边假设删去了 m 个不符合条件的词，重新进行随机选择删去单词个数的特征，即 m 个特征，作为字典中新单词，再构造 kdtree,迭代指定次数结束）进行码本字典的构造，为了提高后续码本字典的查找速度这边我们将码本字典用 Best Bin First kdtree 表示。

3.3.4 对每张图片进行编码

① VQ编码

通过 kdtree 获得字典中距离图片中每个“单词”最近的单词在字典中的序号（例如图片中某一个单词 i 距离字典中第 j 个单词最近，我们就把这两个单词近似认为是同一个单词），最终我们即可获得一个 $n*1$ 的数组 codeids，第 i 维的值即表示图片中第 i 个单词在字典中对应的单词的序号，对字典中每个单词在图片中出现的次数进行整理，我们可获得一个 $n*1$ 的向量 code，第 i 维表示图片中第 i 个单词出现的次数，最后通过 l1 范数进行标准化处理得到 $n*1$ 维向量 pcode，第 i 维表示单词 i 在图片中出现的频率。

② LLC编码

通过 Best Bin First kdtree 获得字典中距离图片中每个“单词”最近的 K 个码本字典中的单词，按照递增顺序排列（本次实验中 K 取 5，并且设置了最大迭代次数为 500 来获得近似最优解），最终得到 $K*m$ 的 ix 矩阵（第 i 列表示字典中距离图片中第 i 个单词距离最近的 5 个码本字典中的单词），然后通过 LLC 近似编码获得最终的编码表示，与 VQ 编码不同的是这边我们在寻找最近的 K 个单词时采用的 Best Bin First 搜索的方法，最终我们得到 n 维的向量 encoding，类似于 VQ 编码中得到的 pcode，因为图片中每一个单词都由字典中的若干个“单词”表示，所以 encoding 中所有元素的和不为 1，但是第 i 个元素也可以近似表示为单词 i 在图片中出现的频率。

③ KCB编码

通过 Best Bin First kdtree 搜索获得字典中距离图片中每个“单词”最近的 K 个码本字典中单词的序号，并按照升序排列（本次实验中 K 取 5，并且设置了最大迭代次数为 500 来获得近似最优解），最终得到一个 $K*m$ 的 ix 矩阵（第 i 列表示距离图片中第 i 个 SIFT 特征距离最近的 K 个码本字典中的单词的序号），以及 $K*m$ 的 distsq 矩阵（第 i 列表示字典中与图片中第 i 个单词距离最近的 5 个单词的距离， m 表示此张图片由 m 个 SIFT 特征

组成），然后用高斯函数 $\frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-dist}{2\sigma^2}}$ 进行编码，然后再对每一列进行 l1 规范化操作得到 $K*m$ 的 kerDists，然后再用 vl_binsum 得到 $n*1$ 的 encoding，这边的 encoding 的解释类似于 LLC 方法中的 encoding，即第 i 个元素近似表示为单词 i 在图片中出现的频率（类比于词频的理解）。

3.3.5 采用SPM（Spatial Pyramid Method）进行池化

本次实验对每张测试图片提取 8 层特征，分别为图 3-10 中的 1,2,3,4,5,6,7,8，其中 1,2,3,4 将整张图片四等分，5,6,7 则将整张图片三等分，8 则是整张图片。将每一部分作为一个整体，对每一部分我们都进行相同的编码（VQ,LLC,KCB 等）获得 n 维的 encoding（解释同上），把这 8 部分组合在一起即可获得 $8n$ 的 pcode，再对 pcode 使用 l_1 规范化，然后使用 vl_homkernmap 函数对 pcode 进行 chi2 approximation 得到最终的 $(3*8n)$ 维的池特征向量 pcode。

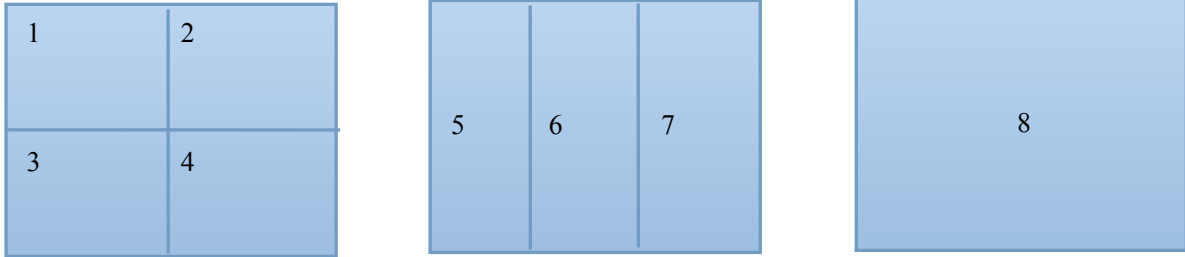


图 3-10: 8 层特征

3.3.6 采用 SVM 分类器进行分类识别

此处主要采用 SVM 对于提取的特征进行训练构造模型，然后对于测试图片集进行测试，获得较好的参数 C （通过计算 maxap，最高的正确率获得），用于最后的分类识别。

因为本次实验中码本字典数为 4000（相当于特征数为 4000），而分类的个数不多，仅有几十类，所以此次实验采用线性 SVM 分类器。

①计算核矩阵：如果用所有图片的 SIFT 特征直接形成矩阵，则矩阵会很大，较难存储，实验中采用 chunking method 解决了此问题,将每 100 张图片存于 1 个 chunk 中。而核矩阵 K 在这边则采用了分块矩阵的乘法进行运算，最终存于矩阵 K 中。

②获得分类超平面的 w 和 b 系数：通过 getlmbGT 获得训练图片的标签 labels_train，labels_train 是一个 $a*1$ 的 cell， a 表示一共有 a 类图片， $a(i)$ 表示第 i 类所包含的图片索引值。利用 labels_train 和 K 矩阵进行 svmtrain 获得 $a*1$ 的 cell 的 libsvm，libsvm $\{i\}$ 中存放了第 i 类与其它类分类的支持向量 SVs 和相应的系数 sv_coef。在利用支持向量与对应的支持向量系数通过函数 train(trainvecs, labels_train)计算分割超平面并获得分类超平面的系数 w, b 。

③使用线性 SVM 分类器获得测试结果：使用 SVM 进行测试，显示使用 svmtest 函数计算每一张测试图片(每一张图片用一个向量 x 表示) $scoretest=wx+b$ 的值，并将每一列按照降序排序，排第一的也就是 scoretest $\{i\}$ 中值最大的所对应的索引号 j ，我们把第 i 张图片归为第 j 类，然后用 est_label 存储所有图片的分类结果。（其中 est_label $\{i\}=j$,表明第 i 张测试图片属于第 j 类），这样我们就得到了所有测试图片的分类情况。

④计算分类正确率：使用 getlmbGT 函数获得测试图片的实际情况，也就是实际情况中测试图片所属的类，并存于 gt 中，gt 是一个 $n*m$ 的矩阵， n 是分类的个数， m 是测试图片个数，每一类存储的都是 1，-1,gt $\{i\}$ 表示第 i 类的情况，其中 1 表示属于第 i 类，而 -1 表示不属于这一类。sortidx=sort(scoremat(c_i ,:)); 对第 c_i 类的值（也就是 $wx+b$ ）排序，得到的值存于 sortidx，sortidx 是一个 $1*m$ 的向量，sortidx $\{i\}$,测试索引值为 sortidx $\{i\}$ 的测试图片第 i 大可能是属于第 c_i 类图片的。然后利用计算 AP $\{i\}$,第 i 类图片的测试准确率。

⑤利用最优参数进行分类:本实验中利用不同的惩罚参数 c 进行实验,比较每个 c 对应的实验结果,取使得平均正确率 $\text{mean}\{AP\}$ 最大的 c 用做最后的 test,得到最终的结果。

3.3.7 FKdemo 部分

FKdemo 与前面三种方法在实现过程中主要有以下两处不同:

①同样对训练图片采用函数 `vl_phow` 提取 dense sift 特征,但随后则采用主成分分析法,将 128 维的 dense sift 特征降维成 128×80 的 `low_proj`, `low_proj{:,j}` 表示的是第 j 大的特征值所对应的特征向量,训练图片获得的 128 维 dense sift 算子乘上 `low_proj` 即可获得 80 维的 dense sift 特征(即用 80 维近似表示 128 维),再使用高斯混合模型进行训练获得 80×256 的码本字典 `codebook`。

②对于获得的 $80 \times n$ (n 表示特征个数)的 `feats` 进行 Fisher 编码获得 40960×1 的 Fisher 向量 `code` 而不是像原来的 `vq` 等编码一样在每个区域提取 BOV 直方图,最后用 `pcode` 存储 8 个 channel 的 `code` (同上 `vq`, `llc`, `kcb`),然后使用 Hellinger Kernel 核 `pcode=sign(pcode) .* sqrt(abs(pcode))` 处理,使用 l_2 范数规范化得到最终的池特征向量 `pcode`。

3.4 实验结果与分析

3.4.1 实验配置

1.在运行本实验代码前我们必须安装如下两个库:

①VLFeat library, 可从网站<http://www.vlfeat.org>上下载获得;

②LIBSVM library, 可以从网站<http://www.csie.ntu.edu.tw/~cjlin/libsvm>上下载;

2.本次实验主要采用的是 matlab2012a, 并需要安装 VS2008 C++ 的 SDK 作为编译器,因为本实验中为了提高程序的运行速度,采用了 C++ 和 matlab 混合编程的方式,所以需要在程序运行之前,先在 matlab 上安装好 C++ 编译器。

3.4.2 实验数据

本次实验采用的实验图片为: `agata_potato` (201张), `asterix_potato` (182张), `cashew` (210张), `diamond_peach` (211张), `fuji_apple` (212张), `granny_smith_apple` (155张), `honeydew_melon` (145张), `kiwi` (171张), `nectarine` (247张), `onion` (75张), `orange` (103张), `plum` (264张), `spanish_pear` (159张), `taiti_lime` (106张), `Watermelon` (192张), 共15类2633张,我们取每类图片的前2/3作为训练图片,后1/3作为测试图片。

3.4.3 实验结果

如表3-1所示VQ,KCB,LLC是码本字典大小为4000的情况下的结果,FK是码本字典大小为256情况下的实验结果:

表3-1: VQ,KCB,LLC,FK的运行结果

水果种类	VQ	KCB	LLC	FK
agata_potato	58.82%	95.59%	100.00%	4.41%
asterix_potato	98.36%	98.36%	96.72%	0.00%
cashew	100.00%	100.00%	100.00%	15.71%
diamond_peach	98.59%	100.00%	100.00%	14.08%
fuji_apple	100.00%	100.00%	98.59%	9.86%
granny_smith_apple	100.00%	100.00%	100.00%	5.77%
honeydew_melon	100.00%	100.00%	93.88%	6.12%
kiwi	94.83%	98.28%	98.28%	3.45%
nectarine	97.59%	98.80%	98.80%	21.69%
onion	100.00%	96.00%	100.00%	0.00%
orange	94.29%	94.29%	100.00%	2.86%
plum	100.00%	100.00%	100.00%	15.73%
spanish_pear	92.45%	96.23%	94.34%	0.00%
taiti_lime	100.00%	100.00%	100.00%	0.00%
watermelon	82.81%	82.81%	82.81%	10.94%
总的评价识别率以及对应的参数	94.12% (C=8)	97.51% (C=7.8)	97.51% (C=5.4)	8.94% (C=2)

3.4.4 实验结果分析

通过表 3-1 我们发现 (1) VQ 编码, KCB 编码, LLC 编码都展示除了极好的分类性能; (2) 在相同码本大小 $K=4000$ 的情况下, KCB 除了在 onion 的分类性能上比 VQ 编码略差之外, 其余水果的分类性能都大于或等于 VQ 编码的分类性能; (3) LLC 编码除了在水果 asterix_potato, fuji_apple, honeydew_melon 的分类性能上比 VQ 编码略差之外, 在其余水果的分类性能都大于或等于 VQ 编码的分类性能; (4) FK 编码表现出了极差的分类性能。究其原因, 可归结如下:

KCB 编码和 LLC 编码利用了特征点周围相邻点之间的关系, 具有更好的鲁棒性, 较好的摆脱了 VQ 编码中存在的码本不确定性和码本说服性问题。同时 VQ, LLC 和 KCB 编码结合 SPM 较好的结合了图片的空间特性。

FK 编码采用的码本字典较小, 仅为 256, 对其最终分类结果造成了较大影响。

3.5 结论

本章使用 SIFT 算子作为主要特征并结合文本分类的思想进行图片分类, 分类过程中分别对测试集中每张图片的 SIFT 特征进行 VQ, LLC, KCB, FK 编码, 再使用 SPM 进行池化, 最后利用线性支持向量机进行分类。

LLC 和 KCB 算法利用了特征点与周围点的关系进行编码, 解决了 VQ 编码中出现的码本不确定性和码本说服性问题, 获得了更好的分类性能, 编码完成后再结合 SPM 进行池化, 充分利用图片中各特征之间的空间关系。

Fisher 向量因其自身构造的问题展示了较差的性能。

本算法可改进之处（1）使用大小更合适的码本字典；（2）在 SIFT 特征基础上加上新的特征一起进行编码，例如蔬菜水果的颜色等特征；（3）采用其它机器学习方法进行分类识别。

第四章 菜价查询菜谱推荐系统

4.1 主要功能

本软件主要实现两大功能：

- (1) 通过用户需要查找对应地区的蔬菜水果最近一周的价格并予以显示；
- (2) 通过用户查找的蔬菜水果推荐相关的菜谱；
- (3) 双击菜谱能跳转到相应的菜谱的制作网页。

4.2 实验原理

4.2.1 模拟鼠标点击进行自动搜索

通过对应目标网址的搜索特点如菜谱网“<http://www.douguo.com/search/recipe/>”每次搜索不同菜谱时只需在后面加上菜名即可，同样中国水果蔬菜网中水果蔬菜的搜索仅需在如下字符串“http://nc.mofcom.gov.cn/channel/gxdj/jghq/jg_list.shtml?”后面添加相应的搜索信息即可，如我们需要搜索关键词，则只需在后面添加&keyword=“关键词”，然后通过WebClient.DownloadData(url)读取对应网页内容，即模拟鼠标点击的功能。

4.2.2 网页核心内容的提取

①提取蔬菜价格信息：根据中国蔬菜网上的网页源代码的格式进行目标元素提取，因为所有菜的信息都在<td>和</td>中间，我们使用string.IndexOf()函数进行位置提取，然后对其中的字符串进行简单的处理，如寻找中文信息时，我们可以利用中文字符大于127的信息。

②提取菜谱的信息：根据菜谱网网页源代码的格式进行目标信息的提取，因为所有菜谱信息都在字符串“<http://www.douguo.com/cookbook/>”之后，我们使用string.IndexOf()函数对菜谱信息位置进行定位，而菜的主料前都在字符串“<p class=\"pbm\">主料：<”之后，我们同样使用string.IndexOf()函数进行位置定位，然后在使用字符串的信息进行最终信息获取，如中文字子都大于127的信息等。

4.3 实验结果

主界面部分，提供查询蔬菜水果等产品的列表，如图4-1所示：

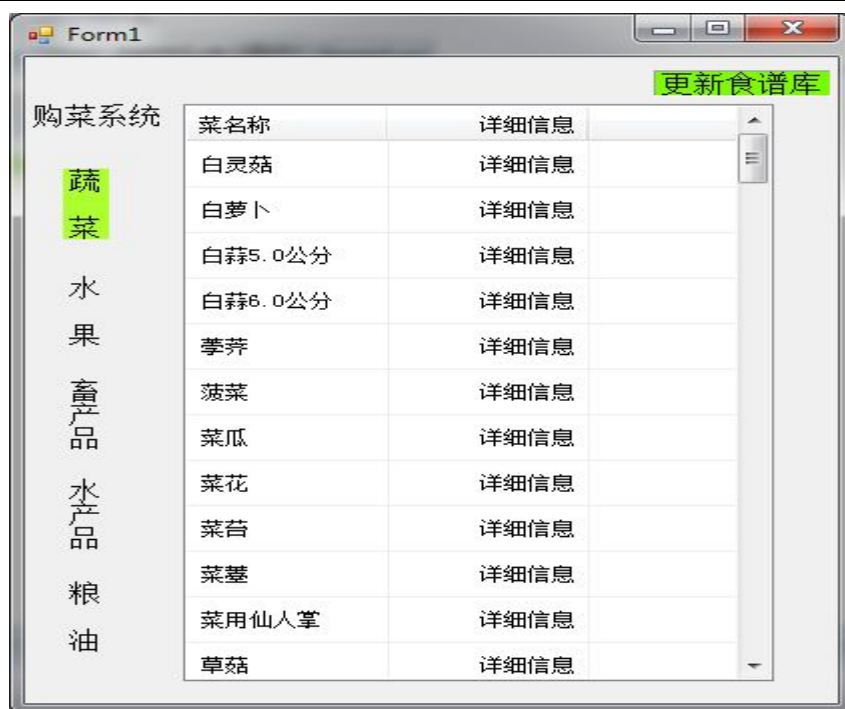


图 4-1：主界面

双击获得对应蔬菜价格及菜谱推荐部分，如图4-2所示，在左上角的部分可以选择不同省市的情况，单机搜索即可获得不同区域的菜价，如图4-3所示：

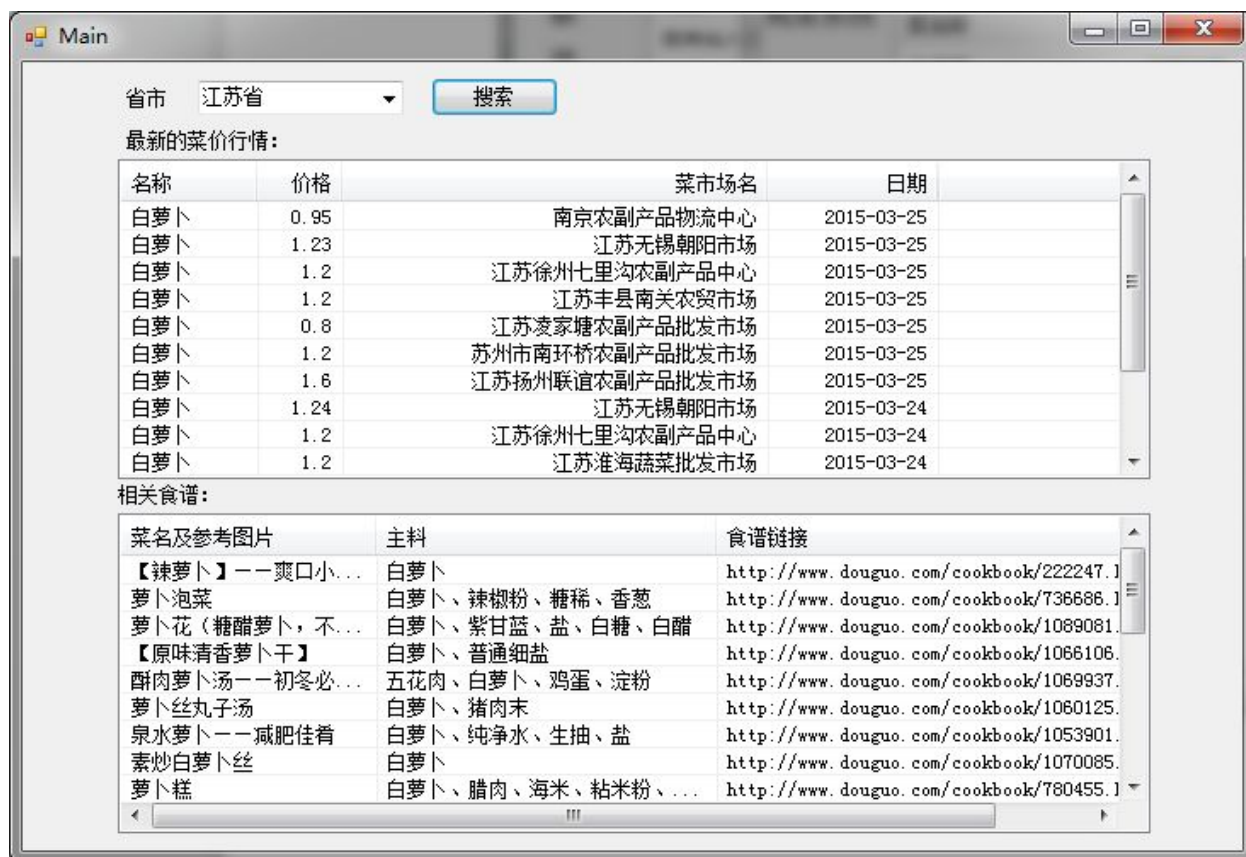


图 4-2：蔬菜价格及菜谱推荐部分

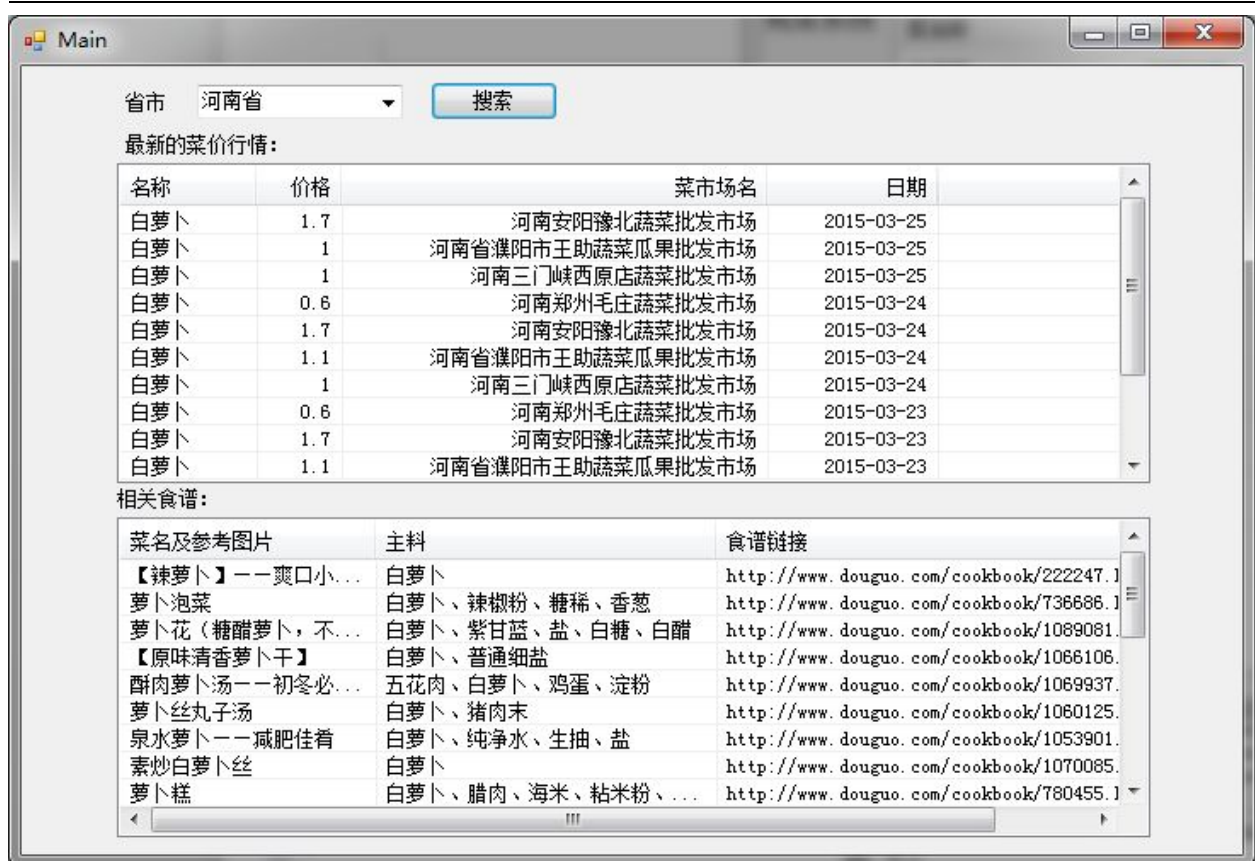


图 4-3: 修改省份重新搜索

第五章 结论与展望

5.1 结论

本文采用两种蔬菜水果识别算法对蔬菜水果进行分类识别，同时开发了蔬菜价格查询以及菜谱推荐的软件。其主要内容包括：

基于颜色，纹理特征的方法：对比了五种简单背景下的图片背景删除法：①基于KMeans方法消除图片背景；②采用背景过度删除；③利用过度删除图片和轻度删除图片对图片进行定位；④在③的基础上对已定位的图片再采用Kmeans算法进行简单背景删除；⑤边缘法结合Kmeans的背景删除法。然后对预处理的图片提取颜色以及ISADH纹理特征，最后采用交叉验证获得最优参数进行最终分类。对实验结果的分析中我们得到如下结论：（1）算法⑤在对背景删除目标提取上展现了极好的性能以及鲁棒性；（2）ISADH纹理特征表现了较好的分类性能；（3）颜色均值作为颜色特征在水果蔬菜识别中会因同一蔬菜或水果的颜色多样以及不同水果可能会是同一颜色等情况而使得最终正确分类率会降低；

SIFT 特征方法：相比于传统的以颜色，大小，形状，纹理等作为主要特征的算法，我们采用简单的 SIFT 特征结合文本分类的思想，对测试集中提取的 SIFT 特征分别使用 VQ,LLC,KCB,FK 进行编码，并利用 SPM 进行池化，最终采用线性支持向量机进行分类的方法，在背景较为复杂的情况下获得了 80%以上的识别率，在简单背景下的识别率则高达 94%以上，本次实验采用的模型对于光照的影响，图片的大小，尺度的变换等都具有很好的鲁棒性。其中 LLC 和 KCB 很好地利用了特征点的与周围特征点之间的关系，相比于 VQ 具有更高的识别率。

蔬菜水果价格查询及菜谱推荐系统：用 C#语言进行 WinForm 编程，根据网上蔬菜水果价格网以及菜谱网的网页源代码格式进行字符串操作获取对应的蔬菜价格，相关菜谱等主要信息并在 WinForm 窗体中显示，对人们了解当地菜价行情等能带来较大帮助。

5.2 不足与展望

本实验不足之处①颜色和ISADH纹理方法：去颜色均值作为特征性能较差；②SIFT特征方法：为了提高蔬菜水果的识别速率我们需要将SVM的训练模型进行保存，这会耗费较大的存储空间，不适用于手机等设备使用，而且此模型在背景较为复杂的情况下识别率不高，较难直接使用。

可改进的地方：①在背景删除目标提取时利用将距离等元素考虑进去（边缘处存在目标主体部分）；②对于颜色特征表示可以采用众数，中位数，灰度值个数排行前几的灰度值等；③可在现有的模型上添加其它特征，诸如：颜色，纹理，形状，大小等，结合SIFT算子一起作为训练的特征；④研究比SIFT特征更好的特征表示算法进行特征提取；⑤采用其它更好的分类算法进行分类识别；⑥选用更多更全的样本进行训练，采用合适大小的码本字典以及惩罚参数C，对测试集进行最终分类识别；⑦为解决训练模型存储问题，可将训练模型存于网上服务器中，将照片上传至网上，通过网上服务器计算获得最终的识别结果并在手机端显示。⑧采用其它运行时间更短，识别率更好的机器学习算法。

展望：希望在不久的将来即可人们利用手机拍照获取图片，并能在较短的时间内获得

正确的识别结果，同时显示出当地的蔬菜价格及推荐的菜谱的手机app。

致谢

感谢桑老师在论文选题以及修改中给予学生的指导与教诲，感谢南大张利军老师提供的方案指导与悉心指导，感谢印度教授 Dubey 提供的水果图片资源与交流，同时感谢舍友们在论文写作过程中对于我的文章格式修改等的知道和帮助。

参考文献

- [1] E.Umbaugh, S., Computer Vision and Image Processing: A Practical Approach using CVIP tools.[J] first ed. 1998: Prentice Hall Professional Technical Reference.
- [2] WikiPedia. Computer Vision. 2007 [cited 2008 11/11/2008]; Available from: http://en.wikipedia.org/wiki/Computer_vision.
- [3] A.R.M. Syafiudin, C.C.T., O.M. Isa, APPLICATION OF COMPUTER VISION SYSTEM TO MONITOR CROPS GROWTH AND WEEDS UNDER RAIN SHELTER,[J] in International Symposium on Greenhouses, Environmental Controls and In-house Mechanization for Crop Production in the Tropics and Sub-Tropics, I.H.R. R. Kamaruddin, N.R. Abdul Hamid Editor. 2006, ISHS Acta Horticulturae:Malaysia.
- [4] Pajares, G.T., A. Burgosartiz, X.-P. Ribeiro, A., Design of a computer vision system for a differential spraying operation in precision agriculture using hebbian learning.[J] Computer Vision, IET, 2007. 1(3-4): p. 93-99.
- [5] Feature Extraction. [cited 3rd April 2007]; Available from: <http://www.photonics.com/directory/dictionary/lookup.asp?url=lookup&entrynum=1837&letter=f>.
- [6] Pattern recognition.[M] [cited 4th November 2008]; Available from:<http://encyclopedia.thefreedictionary.com/Pattern%20recognition>.
- [7] 林洁 基于 BP 算法的三分类设计[J]-广西轻工业 2007,23(9)
- [8] 赵玖玲 基于改进 BP 网络的指纹分类[J]-仪器仪表学报 2004,25(z1)
- [9] 周建平. 陶春辉. 吕文正. 何拥华. 顾春华. ZHOU Jian-ping.TAO Chun-hui.LU Wen-zheng.HE Yong-hua.GU Chunhua 基于 BP 网络对模拟声呐信号分类[J]-海洋学研究 2007,25(2)
- [10] 罗承成, 李书琴, 唐晶磊. 基于多示例学习的超市农产品图像识别[J]. 计算机应用, 2012,32(6):1560-1562
- [11] 阳江平. 基于计算机视觉的果蔬识别方法研究[D]. 大连: 大连理工大学, 2011
- [12] Bolle, R.M.; Connell, J.H.; Haas, N.; Mohan, R.; Taubin, G. VeggieVision: A Produce Recognition System.[J] In Proceedings 3rd IEEE Workshop on Applications of Computer Vision,[J] WACV'96, Sarasota, FL, USA, 2-4 December 1996, pp. 244-251.
- [13] Bolle R.M. Connell J.H. Haas. N.et al.Veggievision:a Produce recognition system.[J] In:WACV,Sarasota,USA. 1996:1-8.
- [14] Hong, S.G.; Maccaroni, M.; Figuli, P.J.; Pryor, B.M.; Belisario, A. Polyphasic Classification of alternaria isolated from hazelnut and walnut fruit [J].in Europe. Mycol. Res. 2006, 110, 1290-1300.
- [15] Baltazar, A.; Aranda, J.I.; González-Aguilar, G. Bayesian classification of ripening stages of tomato fruit using acoustic impact and colorimeter sensor data[J]. Comput. Electron. Agric. 2008, 60,113-121.
- [16] Pennington, J.A.T.; Fisher, R.A. Classification of fruits and vegetables[J]. J. Food Compos. Anal.2009, 22, S23-S31.
- [17] Pholpho, T.; Pathaveerat, S.; Sirisomboon, P. Classification of longan fruit bruising using visible spectroscopy[J]. J. Food Eng. 2011, 104, 169-172.

- [18] A. Rocha, D. C. Hauagge, J. Wainer, and S. Goldenstein, “Automatic fruit and vegetable classification from images,” [J] Elsevier COMPAG, vol. 70, no. 1, pp. 96 – 104, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S016816990900180X>
- [19] S. Arivazhagan, R. N. Shebiah, S. S. Nidhyanandhan, and L. Ganesan, “Fruit recognition using color and texture features,” [J] CIS Journal, 2010.
- [20] Patil, N.S.; Shelokar, P.S.; Jayaraman, V.K.; Kulkarni, B.D. Regression models using pattern search assisted least square support vector machines. [J] Chem. Eng. Res. Des. 2005, 83, 1030–1037.
- [21] Li, D.; Yang, W.; Wang, S. Classification of foreign fibers in cotton lint using machine vision and multi-class support vector machine. [J] Comput. Electron. Agric. 2010, 74, 274–279.
- [22] Shiv Ram Dubey, A S Jalal a* Robust Approach for Fruit and Vegetable Classification [J]
- [23] Yudong Zhang and Lenan Wu *. Classification of Fruits Using Computer Vision and a Multiclass Support Vector Machine. [J] Sensors 2012, 12, 12489-12505
- [24] S.Arivazhagan¹, R.Newlin Shebiah¹, S.Selva Nidhyanandhan¹, L.Ganesan² Fruit Recognition using Color and Texture Features. [J] Journal of Emerging Trends in Computing and Information Sciences. VOL. 1, NO. 2, Oct 2010
- [25] Vapnik. The Nature of Statistical Learning Theory. [J] Springer-Verlag, New York, NY, 1995.
- [26] S.S. Keerthi and C.-J. Lin. Asymptotic behaviors of support vector machines with Gaussian kernel. [J] Neural Computation, 15(7):1667–1689, 2003.
- [27] H.-T. Lin and C.-J. Lin. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. [J] Technical report, Department of Computer Science, National Taiwan University, 2003. URL <http://www.csie.ntu.edu.tw/~cjlin/papers/tanh.pdf>.
- [28] D. Lowe. Distinctive image features from scale-invariant keypoints. [J] IJCV, 60(2):91-110, 2004.
- [29] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos [J] In Proc. ICCV, 2003.
- [30] J. C. van Gemert, J. M. Geusebroek, C. J. Veenman, and A. W. M. Smeulders. Kernel codebooks for scene categorization [J] In Proc. ECCV, 2008.
- [31] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification [J] In Proc. CVPR, 2010.
- [32] F. Perronnin, J. Sánchez, and T. Mensink. Improving the fisher kernel for large-scale image classification [J] In Proc. ECCV, 2010.
- [33] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories [J] In Proc. CVPR, 2006.
- [34] Usama Fayyad. A Tutorial on Support Vector Machines for Pattern Recognition. [J] Data Mining and Knowledge Discovery, 2, 121–167 (1998)
- [35] Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. [J] In: CVPR. (2006) 2169-2178
- [36] Aizerman, M.A., Braverman, E.M. and Rozoner, L.I. Theoretical foundations of the potential function method in pattern recognition learning. [J] Automation and Remote Control,

25:821–837, 1964.

- [37] Yudong Zhang and Lenan Wu *.Classification of Fruits Using Computer Vision and a Multiclass Support Vector Machine [J]
- [38] Bolle, R.M.; Connell, J.H.; Haas, N.; Mohan, R.; Taubin, G. VeggieVision: A Produce Recognition System. [J] In Proceedings 3rd IEEE Workshop on Applications of Computer Vision, WACV'96, Sarasota, FL, USA, 2–4 December 1996, pp. 244–251.
- [39] Woo Chaw Seng, Seyed Hadi Mirisae. A New Method for Fruits Recognition System[J]
- [40] Fabio Augusto Faria, Jefersson Alex dos Santos, Anderson Rocha and Ricardo da S. Torres. Automatic Classifier Fusion for Produce Recognition[J]
- [41] S.Arivazhagan¹, R.Newlin Shebiah¹, S.Selva Nidhyanandhan¹, L.Ganesan². Fruit Recognition using Color and Texture Features[J]
- [42] Anup Vibhute S K Bodhe. Applications of Image Processing in Agriculture:A Survey International Journal of Computer Applications[J] (0975 – 8887) Volume 52– No.2, August 2012
- [43] Chih-Chung Chang and Chih-Jen Lin <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [44] Md.Towhid Chowdhury¹, Md.Shariful Alam², Muhammad Asiful Hasan³, Md.Imran Khan⁴. Vegetables detection from the glossary shop for the blind.[J]
- [45] <http://www.cnblogs.com/eyesjzwang/articles/2429382.html>
- [46] <http://zh.wikipedia.org/wiki/K-d%E6%A0%91>
- [47] k-d tree 的优化查找算法 BBF <http://www.cnblogs.com/eyesjzwang/articles/2437706.html>
- [48] KD 树核心思想简介
<http://www.cnblogs.com/snake-hand/archive/2012/08/13/2636236.html>
- [49] K. Chatfield, V. Lempitsky, A. Vedaldi, A. Zisserman. The devil is in the details: an evaluation of recent feature encoding methods. [J] British Machine Vision Conference, 2011
- [50] <http://zhouyichu.com/%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0/PCA-Tutorial.Html>
- [51] <http://blog.codinglabs.org/articles/pca-tutorial.html>
- [52] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin A Practical Guide to Support Vector Classification[J]
- [53] Unser M.Sum and Difference Histograms for Texture Classification[J]. IEEE Transaction and Pattern Analysis and Machine Intelligence;1986 vol.8,no.1,pp.118-125
- [54] Seng, W.C.; Mirisae, S.H. A New Method for Fruits Recognition System. [J] In Proceedings of International Conference on Electrical Engineering and Informatics, ICEEI'09, Selangor, Malaysia, 5–7 August 2009; Volume 1, pp. 130–134.
- [55] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases[J] In Proc. CVPR, 2008.