

## Abstract

Beauty is widely perceived by the human conception, and facial beauty is the most common perception of objects in human social activities. Humans investigate the concept of beauty more than a few thousand years, not only to promote continuous improvements of the aesthetic consciousness of people, but also promote the rapid development of society. However, it is still challenging to define what is a beautiful face? Recently, numerous studies in cognitive psychology found what a beautiful face was and there is a high degree of consistency over different races, ages and genders. In our work, facial beauty classification methods based on image processing and machine learning techniques have been proposed. We adopt some quantitative methods to extract facial beauty features and beauty classification, with aims to seek some guiding significance to be used in clinical cosmetic, orthopedic and prosthetic in real worlds.

While the research on facial beauty analysis via image processing and machine learning schemes is relatively small reported in the literature, it has begun to attract the attention of many researchers in recent years. By investigating the research problems on facial beauty analysis, we are like to give some objective and quantifiable descriptions. The main contributions of our work include the following four aspects:

(1) A large scale face image dataset containing both male and female are collected to establish for experimental study. These face images are mainly drawled from some popular social networking sites.

(2) To describe effective features to describe facial beauty information, the local binary pattern (LBP) descriptor and sub-block LBP (denoted by Block-LBP) are employed to extract the texture features, respectively.

(3) The scale invariant feature transform (SIFT) descriptor with spatial pyramid matching (SPM) strategy is employed to extract facial feature points, and also used to represent facial beauty information.

(4) Given a face image, we use the K-Nearest Neighbor (KNN) classifier and Support Vector Machine (SVM) classifier to perform the beauty classification task, respectively. Lastly, numerous discussions of experimental results are reported in detail.

Experimental results show that our proposed facial beauty classification approaches are effective to assess human facial beauty. In addition, we draw some

important conclusions.

**Keywords:** Facial Beauty Local Binary Patterns Scale Invariant  
Feature Transform Spatial Pyramid Matching KNN  
Support Vector Machine

# 目录

第一章 绪论.....	1
1.1 研究背景和意义 .....	1
1.2 国内外研究现状 .....	2
1.3 本文内容及组织 .....	5
1.3.1 本文主要内容.....	5
1.3.2 本文的组织.....	5
第二章 基于 LBP 的纹理特征提取.....	7
2.1 LBP 描述子基本原理.....	7
2.1.1 LBP 描述子.....	7
2.1.2 扩展的 LBP 描述子 .....	8
2.2 LBP 的发展演化 .....	10
2.2.1 LBP 统一模式 .....	10
2.2.2 旋转不变的 LBP 描述子 .....	12
2.2.3 LBP 的特点 .....	14
2.3 基于 LBP 的人脸纹理特征提取 .....	15
2.3.1 基于 LBP 人脸纹理特征提取 .....	15
2.3.2 LBP 人脸纹理特征匹配 .....	17
2.4 小结 .....	18
第三章 基于 SIFT 和 SPM 的特征提取 .....	19
3.1 SIFT 特征提取.....	19
3.1.1 构建尺度空间 .....	20
3.1.2 尺度空间极值点检测 .....	21
3.1.3 确定特征点的位置 .....	22
3.1.4 为特征点分配方向值 .....	24
3.1.5 SIFT 特征描述子.....	25
3.2 构建人脸图像的视觉词袋模型 .....	26
3.2.1 构建视觉词袋模型 .....	27
3.2.2 K-means 聚类算法.....	27
3.3 空间金字塔匹配 .....	27
3.3.1 金字塔匹配核算法 .....	28
3.3.2 空间金字塔匹配策略 .....	29

---

3.4 小结.....	30
第四章 分类器设计.....	33
4.1 K 近邻算法.....	33
4.1.1 KNN 原理.....	33
4.1.2 KNN 的优缺点.....	34
4.2 支持向量机.....	34
4.2.1 线性可分的支持向量机.....	34
4.2.2 线性不可分的支持向量机.....	36
4.2.3 非线性支持向量机.....	37
4.2.4 支持向量机的优缺点.....	39
4.3 小结.....	39
第五章 实验结果与分析.....	41
5.1 人脸美丽数据库.....	41
5.2 基于 LBP 的人脸美丽分类.....	42
5.3 基于 SIFT 和 SPM 的人脸美丽分类.....	44
5.4 实验结果分析.....	44
5.5 小结.....	45
第六章 结束语.....	47
6.1 总结.....	47
6.2 工作展望.....	48
致谢.....	49
参考文献.....	51

## 第一章 绪论

### 1.1 研究背景和意义

美丽是人类可以广泛感知的一个概念。美能给人的心灵带来愉悦，并能强烈地激发人类的感情。爱美是人类的天性，爱美之心，人皆有之，比如，“没有什么能比一张美丽的人脸更令人赏心悦目”，“美丽的面孔激发人的感官”<sup>[1][2]</sup>。在社会生活中，美丽可以带来社会上和经济上的诸多好处。当代社会，人们对美的追求使得临床整形及其化妆品行业快速发展，这两个相关行业有着令人吃惊的庞大消费数据。由于人脸美丽吸引力在我们每天的日常生活中密切存在以及其对人们不容回避的重大影响，因此研究人脸美丽的特点或构成要素具有现实的指导意义，自然地，自古以来关于美丽和吸引力展开的相关研究受到了哲学家、心理学家及计算机学者们的重视并取得不断的发展。

美丽是令人兴奋的，但是美丽也是神秘的。在我们社会生活的经历中，一个人是否美丽，大多数人都认为是一个很抽象的判断。不同的文化有着自己评价的标准。即使是同一个民族、同一个时代、不同的个体，对人脸面孔美丽的看法也是有差异的。美丽之所以很难判断，除了文化和时代的因素外，更重要的是一个人的评价往往受如衣着、发型、社会地位、个人情感和他人评价等各种各样的因素干扰，也就是说，美是一个充满变数的概念。但是，在我们看到美的变化的一面同时，研究者们却发现了潜藏在美的评价下的规律。在物理心理学方面，达尔文派学者致力于探索脸部美是否折射出了遗传环境的影响，研究结果认为：潜藏在美的判断下面的心理机制是整个人类历史中进化的适应的结果，认为脸部的平均性、对称性以及二向异型是自然选择的结果，它们是构成了人脸美丽感知的重要因素<sup>[3][4]</sup>；同时，在认知心理学方面，研究者们推翻了长期以来所认为的“美丽是主观的”的观点，他们通过大量的实验发现人们对于什么样的脸是美丽的看法存在着高度的一致，这种高度的一致与年龄、文化、性别、种族等无关<sup>[5][6][7]</sup>。这表明，当代研究学者认为人们对于人脸美丽程度的感知确实存在着一定的客观标准，而这个客观的标准可以作为衡量美丽的主旋律，而感知者的情绪以及被感知者的社会地位等一些因素是主旋律之上的跳跃的音符。

尽管在哲学、医学及认知心理学方面对人脸美丽吸引力进行了多年的研究工作，但在计算机及信息处理学科利用图像处理、机器学习等方法来进行较客观的公正的人脸美丽评价的研究文献并不多，但是最近几年来已经开始受到不少研究者的重视。人工智能的最终目标是让计算机具有与人一样的智能，包括感知、推

理、思维、判断、识别、推广应用等各方面的能力。既然人具备着对人脸美丽吸引力的普适的认知能力，且对美的认知隐藏着客观的标准，那么，有没有可能也让计算机通过机器学习的手段来获得与人一样的智能？我们认为这就是人脸美丽分类这一研究问题的一个基本问题，这也是本文研究的主要动机。美丽的人脸上肯定有某些特征是承载着“人脸美丽的密码”。美丽的标准也并非是一种绝对抽象的东西。随着人脸图像处理、机器学习及人脸美学研究的深入，人们已经可以把美丽人脸器官的各部分比例较精确地量化出来。因此，使用图像处理及机器学习的方法来对人脸美丽程度作一个客观的评估已成为了可能。

人脸美丽分类的目标是更好地探索美丽的本质，以及更进一步使得计算机也具有象人一样的美丽感知智能。该课题的研究涉及到认知心理学、美容医学、哲学、计算机及信息处理科学（图像处理、机器学习）等多个学科领域的交叉研究。对人脸美丽分类的机器研究除了推进人们认识人脸美的本质之外，还具有很强的实用价值。比如，可以辅助医院进行美容整形，利用检测出的数据与标准美丽模型相对比得到的结果，为整容者给出各部分改进的最佳方案，以利于整容者做出正确的决定，从而消除整容者或整容师个人对美的错误认知所带来的不良后果；研究人脸美丽吸引力的机器学习还可以为数字娱乐、智能感知、人脸图像处理等领域提供较有价值的方法和工具。如美丽自动打分、排名系统可以为以脸部吸引力作为基本要求的工作筛选出符合要求的申请者，可以为大型社交网站的用户从大量的人脸照片中搜索出美丽的人脸；人脸美丽评价也可给动画、游戏设计人员设计虚拟人提供参照，为数字娱乐行业提供客观的辅助参考数据等等<sup>[8][9][10]</sup>。

在本课题中，我们希望通过利用图像处理及机器学习方法来研究人脸美丽吸引力，能对人脸美丽这一研究问题给出客观的、可量化的描述，以推动人脸美丽研究这一问题的继续发展。

## 1.2 国内外研究现状

人脸的美丽是什么？为什么人们会被美丽吸引？人脸的美丽（Beauty）虽然容易感知，但却很难定义<sup>[11]</sup>。数个世纪以来，哲学家、人类学家、心理学家、医学家等一直在关注和研究人脸美的本质以及其构成要素，近几十年来对人脸美丽的研究迅速发展，发表的文章已数以千计<sup>[3][6]</sup>，但所取得的研究成果主要都局限在心理学及医学研究领域，并且一些研究观点还存在争议<sup>[6]</sup>。正如人脸美丽研究领域的著名学者 Rhodes 指出“人们对人脸美丽的研究并不成熟，可以说是刚刚开始”<sup>[3]</sup>。特别是利用计算机信息处理技术来研究人脸美丽的研究最近几年来才刚刚兴起。在当今时代，人类对自身的探求过程正变得越来越精妙，随着信息处理技术的发展，如何借助先进的计算机图像处理和机器学习技术来对人脸美丽进

行分析以帮助其它学科探索美丽的本质, 以及更进一步地使得计算机像人一样也具有的美丽感知智能, 是在模式识别、图像处理及机器学习领域中关于人脸分析研究中继人脸识别、人脸表情识别、年龄识别后出现的又一兴新课题<sup>[8]</sup>。

计算机科学对人脸美丽分类研究的第一大类是利用计算机图像处理技术去验证心理学、生物学所提出的一系列人脸美丽吸引力的假说, 如平均性、对称性假说等。Langlois 和 Roggman 在研究中使用的计算机数字合成技术便是这种尝试的第一步<sup>[12]</sup>。该方法把图像的像素值做了平均运算, 其研究结果也得出了平均化合成人脸的美丽程度大于原始人脸图像的结论。但是, 文献<sup>[12]</sup>中对图像只经过简单的像素值平均处理, 导致产生污点和重影, 故需要进一步的处理。后面的研究者将面部各特征的位置、大小、形状以及关键点来标注, 平均化的过程便是用这些关键点的平均值来合成人脸图像, 故这样产生的平均脸更加自然<sup>[13]</sup>。虽然关于平均脸的研究结果有些争议, 但一般认为平均脸至少可以得到中等程度以上的美丽吸引力。目前, 关于对称性的研究<sup>[14][15][16]</sup>, 争议较大, 有的研究结果认为脸的左半边和右半边完全对称使人脸看起来死板, 没有任何感情, 认为适度的不对称更有吸引力<sup>[16]</sup>; 在美丽的人脸上, 经常存在不对称<sup>[17][18][19]</sup>。

计算机科学对人类美丽研究的第二大类是人脸图像美丽分类。人脸美丽分类是研究如何让计算机具有与人类相同的或相似的对人脸美丽程度的判断能力, 它的研究基础是心理学、人类学、医学最近所提出的美丽的客观论, 即人脸美丽的判断超越种族、文化、年龄等因素, 存在着广泛的、一致的标准。这类研究涉及到人脸美丽的构成维度, 对于人脸美丽的研究起到积极的推动作用。就目前现有文献所知, 对人脸图像美丽分类的研究是从近几年开始的, 与此相关的研究报道不是很多。人脸美丽分类的初步研究始于 2001 年 Aarabi 等<sup>[7]</sup>建立的一个自动的面部美丽评分系统。他们首先定位二维的正面的中性表情的人脸照片中的脸、眼睛、眉毛和嘴等特征点, 然后用特征点之间的比例组成的向量来表示人脸, 最后采用 K 近邻变体 (K Nearest Neighbors Variant) 从一组人脸图片训练集中优化学习人脸美丽评价函数。采用的数据是 80 幅女性人脸图像, 其中 40 幅用于训练, 40 幅用于测试, 有 8 个打分者, 美丽程度使用 4 级量表。与此类似的人脸美丽系统还有 Eiseenthal<sup>[20]</sup>、Kagian<sup>[21]</sup>。Eiseenthal 等<sup>[20]</sup>学者采用了两组数据集进行实验, 都包括 92 幅人脸正面图像, 共有 28 个打分者, 打分等级 7 级, 在人脸表示上除了采用 37 个特征点距离及比率构成的特征矢量之外, 还采用灰度图像像素值, 然后对两类人脸用主元分析方法 (PCA) 降维, 最后采用支持向量机 (Support Vector Machine, SVM) 和 KNN 和线性回归方法进行预测分类, 获得与人打分的最高相关系数为 0.65, 表明特征向量的性能优于灰度图像。Kagian 等<sup>[21]</sup>采用的实验数据是 91 幅女性人脸正面图像, 测量了共 84 个特征点坐标, 3486 个特征距离, 特征距离的斜率构成 3486 维斜率向量, 采用线性回归、LS-SVM (Least Squares Support

Vector Machine)、高斯回归 (Gaussian Regression), 最终的预测结果与打分者的相关系数为 0.82。Gunes 等<sup>[22][23]</sup>让 48 个打分者对 215 幅女性图像进行评分, 打分量为 10 级, 统计分析发现所打分数围绕平均值较好的分布, 特别是对很美和最不美的人脸的打分比较一致, 他们采用黄金比例、三庭法等构成的 15 维距离向量和 13 维距离比例向量来表示人脸美丽, 用 C4.5 分类树等训练分类器, 实验显示系统预测的分数等级类似于人工的分数。

此外, 学者 Chang 等<sup>[24]</sup>人提出两类型理论来对人脸进行分类 (美丽和最不美丽两类分类问题), 两个模型分别是最美丽面孔的平均脸 (用 A 表示) 和最不美丽面孔的平均脸 (用 U 表示)。用 46 个特征点之间的距离和某些离散值构成的特征矢量代表人脸, 女性美丽模型来自 90 张女性人脸打分级别较高的人脸特征向量的平均值, 最不美丽的人脸模型则用打分级别较低的人脸特征向量的平均值, 对于给定的待分类的人脸图像, 分别计算和两个模型的相对距离来预测人脸的美丽程度。两类模型方法和平均方法的区别在于两类模型计算人脸与 A、U 之间的距离, 而不是计算和所有样本的平均脸之间的距离, 故取得了较高的分类精度。

Davis 等<sup>[25]</sup>从互联网上下载人脸图像然后进行脸部自动检测、剪裁、变换矫正等预处理工作, 再去除特征点缺失的图像, 得到 2097 幅女性人脸图像, 1523 幅男性人脸图像。然后转成灰度图并归一化处理。图像在视角、表情、光照和图像质量上都有较大的差异。打分者是由网上匿名用户打分, 10 个打分等级。最后把平均脸按照人脸吸引力打分进行演化, 研究从不美到美, 总的形状特征的变化趋势。但是在其文献中, 并没有给出人脸美丽程度的预测值。

Chen 等<sup>[26]</sup>研究人脸美丽分类的两类分类问题。他们使用了更多的数据, 一共是 23412 幅正面人脸照片, 875 幅标注为美丽。检测 68 个标记特征点的坐标构成配置矩阵, 代表人脸的几何特征。图像归一化后映射到切线空间 (tangent space) 进行统计分析: 首先测试男性平均脸和女性平均脸是否有较大得不同, 然后用 PCA 进行主要特征的描述和降维, 用 Kullback-Leibler (KL) 散度来测量美丽人脸的人群和普通人群在特征分布上的不同之处, 结果显示男性和女性的人脸服从不同的高斯分布, 但两个高斯分布有着严重的重叠, 并且男性平均脸和女性平均脸在形状上的差异要小于任何两张脸之间的差异。

回顾这几年来对人脸美丽分类的图像处理和机器学习方法, 我们总结出相关研究具有以下几个特点:

(1) 实验中所采用的数据规模比较小, 大部分研究采用的人脸图像数据只有几十幅到几百幅, 尽管在文献<sup>[26]</sup>中采用了高达 16202 幅人脸图像数据, 但他们仅仅研究最美和最不美这两个简单二分类问题。

(2) 图像预处理和人脸的检测及特征点的定位从早期的手工检测方法发展到自动检测方法;



(3) 几何特征是目前学者应用的最多的特征，主要包括特征点坐标、距离、距离比例率、三角网格边长等，而纹理特征中主要使用色度、饱和度和光滑度、特征脸等特征，直方图等；

(4) 分类方法主要有 K 近邻 (KNN)、支持向量机 (SVM)、C4.5 决策树、神经网络等分类器；

(5) 普遍预测评分等级不是很多，最高的为 10 级，在预测等级不多的情况下，分类性能还比较好，但随着评分等级增加，分类性能将下降。

### 1.3 本文内容及组织

#### 1.3.1 本文主要内容 本文主要研究了人脸美丽的特征提取及分类方法，主

要研究内容包括：

(1) 创建人脸美丽数据库：为了给人脸美丽分类给出一个客观的评价，我们首先在互联网上搜集了较大规模的男性人脸美丽图像库和女性人脸美丽图像库。

(2) 基于 LBP 的人脸纹理特征的提取：先前的文献中的人脸美丽特征主要是面部器官之间的距离比率、距离等等，这就需要对各个器官进行准确定位，这需要很大的工作量，并且存在人为的主观因素，为此，采用局部二值模式 (LBP) 和分块的 LBP (Block-LBP) 提取了整幅图像的纹理特征，可以更鲁棒的描述人脸的美丽特征。

(3) 基于 SIFT 和 SPM 的人脸特征提取：人脸图像的关键点对描述人脸有着非常重要的作用，为此，利用尺度不变特征变换 (SIFT) 来提取人脸美丽的特征点，接着对其用词袋 (Bag of Word, BoW) 模型把每个特征点量化为视觉词，进而求得整幅图像视觉词的统计直方图，来描述人脸美丽特征。此外，为了加入视觉词的空间信息，我们采用空间金字塔匹配策略。

(4) 分类器的设计：提出使用 K 近邻 (KNN) 和支持向量机 (SVM) 作为图像美丽的分类器，然后对分类结果进行客观分析及评价。

#### 1.3.2 本文的组织 本文的论文

结构安排如下：

第一章是绪论，简要概述了人脸美丽分类的历史背景和研究意义，并对当前的一些研究现状进行了详细介绍。

第二章详细介绍人脸图像的纹理特征的提取方法，即利用局部二值模式 (LBP) 描述子来提取人脸图像的纹理特征，同时，为了利用图像的空间信息，

我们采用了分块的 LBP (Block-LBP) 特征来描述人脸的美丽特征。

第三章介绍人脸图像美丽特征点的提取方法, 即利用尺度不变特征变换 (SIFT) 方法来提取人脸图像的美丽特征点, 进而用视觉词袋 (BoV) 模型来表示人脸图像, 同时, 为了加入 SIFT 特征是描述子的位置信息, 我们采用空间金字塔匹配 (SPM) 策略来表示人脸美丽特征, 以便后续的分类。

第四章介绍人脸美丽分类的分类器设计, 采用 K 近邻 (KNN) 分类器和支持向量机 (SVM) 分类器对提取的人脸特征进行分类, 输出量化的分类结果。

第五章介绍人脸美丽分类的实验结果, 首先在互联网上搜集了较大规模的男性人脸美丽图像库和女性美丽图像库, 接着用 KNN 和 SVM 对提取的特征分类, 并客观地分析和评价实验结果。

第六章是总结和展望。总结本课题的研究结果, 并分析存在的缺点和不足。继而指出今后可能的研究方向, 以克服所存在的不足, 完善整个科研课题。

## 第二章 基于 LBP 的纹理特征提取

人类视觉系统认识图像是靠图像具有的某些特征。图像特征是指图像的原始特性、属性，包括图像的亮度、边缘的轮廓、纹理、直方图、颜色等。利用计算机系统识别图像，必须找出一些特征提取算法，来分析图像的特征，然后将特征用数学建模的方式表示出来，使得计算机也能理解这些特征。

纹理特征反映了物体表面亮度和颜色的某种变化，同时这些变化与物体本身的属性有关。举一个形象的例子，例如同一个种类的木材具有相同的或相似的纹理，通常仅仅通过识别木纹的种类来识别木材的树种和材质等。纹理在图像处理中起着非常重要的作用，并广泛运用于卫星遥感图像分析、气象云图分析、生物组织和细胞的研究等应用领域。这些年来，基于局部二值模式（LBP）的纹理描述方法受到广泛的关注，并且在人脸识别中获得巨大成功<sup>[27][28][29]</sup>。LBP 是 1996 年由 Ojala 等<sup>[30]</sup>人提出来，LBP 是一种计算简单、非参数化的局部纹理模式描述算子，它对单调的灰度变化具有不变性，从而对光照具有一定程度的鲁棒性，因此，LBP 是一种有效的纹理描述算子。一般来说，LBP 描述图像纹理特征是基于纹理像素间的相互关联性，并利用结构化方法思想分析固定窗口特征，然后利用统计法作整体特征提取。由于具有这些特点，LBP 算子能够很好地提取人脸面部的纹理关联特征。

本章重点介绍了 LBP 的基本理论及其发展，然后利用 LBP 描述子和分块的 LBP（Block-LBP）提取人脸美丽的纹理特征。

### 2.1 LBP 描述子基本原理

本节主要介绍 LBP 描述子的基本概念和原理。

#### 2.1.1 LBP 描述子

LBP 描述子最早用于人脸识别中。LBP 算子是一种非参数化的图像局部空间结构表述算子，是一种有效的纹理描述算子。LBP 算子是通过重叠选取  $3 \times 3$  像素的子区域窗口，然后对所有非边界像素进行 LBP 编码操作。通常采用的编码方式是以窗口中心点的灰度值作为阈值，然后对窗口内其他像素作二值化处理，并将周围 8 个像素的灰度值与中心像素比较，如果大于中心像素则标记为 1，否则标为 0。这样  $3 \times 3$  的窗口周围的 8 个值以一定的顺序组成一个 8 位的二进制数，然

后根据像素的不同位置进行加权求和，进而得到一个以十进制数来表示该窗口中心点的 LBP 值，该 LBP 值表示了该中心像素点的细节纹理信息，详细的计算流程如图 2.1 所示。

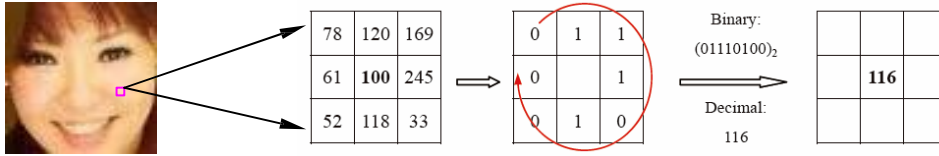


图 2.1 LBP 描述子编码形式

正式地，LBP 值可由如下式子求得：

$$LBP(x_c, y_c) = \sum_{n=0}^7 f(g_n - g_c) 2^n \quad (2-1)$$

其中， $g_c$  表示窗口中心像素的纹理灰度值， $g_n$  表示中心像素周围的八个邻域点的灰度值，函数  $f(x)$  可表示为：

$$f(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (2-2)$$

因此，在图 2.1 中，我们有下面的关系：

LBP 模式 = 01110100

LBP 值 =  $2^6 + 2^5 + 2^4 + 2^2 = 116$

### 2.1.2 扩展的 LBP 描述子

基本的 LBP 描述子的最大缺陷在于它仅仅覆盖了一个固定半径范围内的小区域，显然，这不能满足不同尺度、不同频率纹理的要求。为了适应于不同尺度的纹理特征，同时也达到灰度不变和旋转不变性的要求，Ojala 等<sup>[31]</sup>人随后对基本的 LBP 算子进行了一定程度的改进和发展，他们将  $3 \times 3$  邻域区域扩展到任意邻域区域，并采用圆形邻域替代了原始的正方形邻域，改进后的 LBP 算子允许在半径为  $R$  的圆形邻域内采样任意多个像素点。

这里，为了便于介绍 LBP 算子，我们先做一些基本的定义。假定一个局部区域的纹理分布为局部区域内像素灰度值的联合概率密度分布，定义如下：

$$T = t(g_c, g_0, K, g_p) \quad (2-3)$$

上式中  $g_c$  表示局部区域的中心像素点的灰度值， $g_p$  ( $p = 0, 1, K, P-1$ ) 对应

于中心点周围的等距分布的  $P$  个点。由于通常不能保证所有像素点都刚好落在整数点上,因此采用双线性插值算法对没有完全落在像素位置上的点计算其灰度值。对一副数字图像,邻域内的  $g_p$  点的坐标可以表示为

$$(x_p, y_p) = (x_c + R \cos(2\pi p / P), y_c + R \sin(2\pi p / P)) \quad \text{式(2-4)}$$

这里  $(x_c, y_c)$  表示中心像素的坐标。在不损失任何图像信息的假设下,如果将中心像素点的灰度值从其对应的邻域像素的灰度值中减去,则这个局部区域的纹理可以用中心像素点和中心像素点与周围像素值的差的联合分布来表示,如:

$$T = t(g_c, g_0 - g_c, K, g_{P-1} - g_c) \quad \text{式(2-5)}$$

进一步地,若假设中心点像素  $g_c$  与周围点像素的差值  $g_p - g_c (p = 0, 1, K, P-1)$  独立于中心点则上式可以改写为:

$$T H t(g)(g - g_c, K, g_{P-1} - g_c) \quad \text{式(2-6)}$$

然而,在实际应用中上述独立性假设并不总是成立,因为数字图像中像素灰度值的表示范围有限(0 到 255 之间的整数),从而那些较大或较小的像素灰度值会明显减小差值的取值范围,因此这种独立性的假设以及基于此假设的一些推论有可能导致信息的丢失。然而,通常允许损失少量的信息能够使局部区域纹理的描述对于图像灰度范围内的平移具有不变的特性。另外,又由于仅仅是描述了整个图像的亮度的分布情况,而跟图像的局部纹理无关,它也不能为纹理分析提供任何有价值的信息,因而可以忽略不计。故式(2-6)可改写为下面的公式:

$$T H t(g - g_c, K, g_{P-1} - g_c) \quad \text{式(2-7)}$$

这一  $P$  维的差值分布函数记录了该邻域中每个像素的纹理分布情况。对于相对平滑、灰度变化相对缓慢的图像区域,这些差值较小或接近于 0;而对于单个像素点,各个方向的差值都会比较大;在图像区域的边缘处,某些方向的值比较大,也有某些方向的值比较小。

由于上式定义中差值的符号与平均亮度的变化无关,因此,差值的联合分布具有灰度平移不变性,也就是邻域中所有  $P$  个像素同时加上或者同时减去某个值,其表面的纹理特征不变。但是当对所有像素的灰度值进行同时放大或者缩小相同倍数后,其表征的纹理特征发生变化。故为了达到对尺度变化不变的目的,他们只考虑了差值的符号:

$$T H t(f(g - g_c), K, f(g_{P-1} - g_c)) \quad \text{式(2-8)}$$

其中:

$$f(x) = \begin{cases} 1 & x \geq 0, \\ 0 & x < 0 \end{cases} \quad \text{式(2-9)}$$

通过计算上式，将得到一个 8 位的二进制数，再对像素按不同位置用  $2^p$  进行加权求和，这样就得到了一个仅与邻域周围像素点相关的并且唯一的 LBP 纹理值，我们称这个值为二值模式。这个值描述了以  $(x_c, y_c)$  为中心的局部区域的纹理，也可以用公式表示为：

$$LBP(x_c, y_c) = \sum_{p=0}^{P-1} f(g_p - g_c) \cdot 2^p \quad \text{式(2-10)}$$

实际上，式（2-10）代表的差值的符号将转化为一个二进制数，进而再转化为一个十进制的离散的 LBP 值，或者也可以说转化为了一种 LBP 模式。这个局部区域的纹理，可以用这个 LBP 纹理值或 LBP 模式近似地描述为：

$$T H t(LBP(x_c, y_c)) \quad \text{式(2-11)}$$

根据 LBP 描述子的定义可以明显地得出，LBP 描述算子对于单调的灰度变化具有很好的鲁棒性，只要各像素的位置保持不变，所得的 LBP 纹理值或模式也是不变的。通常采用符号  $LBP^R$  来表示在半径为  $R$  的圆形邻域内有  $P$  个像素点  $g_p$  ( $p = 0, 1, \dots, P-1$ ) 的 LBP 描述子。常见的 LBP 描述子有  $LBP_8^1$ 、 $LBP_{16}^2$ 、 $LBP_8^2$  等，如下图所示：

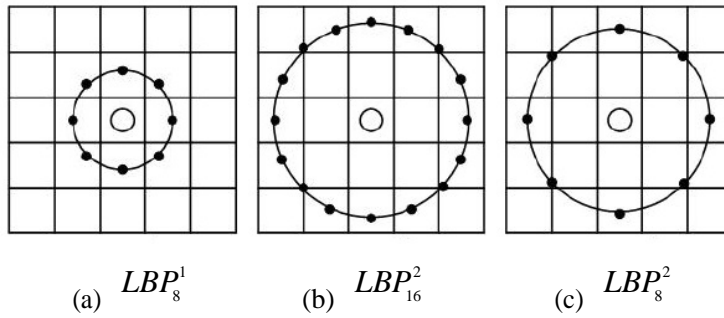


图 2.2 几种常见的 LBP 描述子

## 2.2 LBP 的发展演化

本小节主要详细介绍 LBP 统一模式、旋转不变的 LBP 描述子以及其所具有的一些重要特点。

### 2.2.1 LBP 统一模式

通过观察 LBP 描述子的定义我们可以发现，一个 LBP 描述子可以产生许多

不同的二进制模式，对于 LBP 将会产生  $2^P$  种模式。很显然，随着对邻域内采样点数的增加，二进制模式的种类是迅速增加的。如  $3 \times 3$  邻域内的 8 个采样点，则得到  $2^8$  种二进制模式； $5 \times 5$  邻域内的 20 个采样点，将有  $2^{20}$  种二进制模式；如果  $7 \times 7$  的邻域内有 36 个采样点，则二进制模式的种类可达到  $2^{36}$ 。显然，如此多的二值纹理模式，不管对于纹理特征的提取还是对于纹理的识别、分类及特征的存取都是极其不利的。在实际的应用中，不仅要求所采用的描述子尽量简单，同时也要求计算的速度足够快、数据的存储量尽量小。然而随着模式种类的增多，计算量和数据存储量也会跟着急剧增加。更重要的，模式种类的过多对于纹理特征的表达也是非常不好的。例如，将 LBP 描述子用于纹理分类或人脸识别时，通常采用 LBP 的统计直方图来表达图像的纹理信息，然而过多的模式种类将直接导致数据量过大，并且统计直方图过于稀疏。因此，我们需要对基本的 LBP 模式来进行降维处理，使得在数据量减少的情况下也能最好的表述图像的纹理信息。

幸运的是，为了能够化解二进制的模式过多的问题，并提高其统计特性，Ojala 等<sup>[31]</sup>人提出了使用一种“统一模式”（Uniform Pattern）来对 LBP 描述子的二进制模式种类来做降维。他们认为，在实际的数字图像中，绝大多数的 LBP 二值的模式至多包括两次从 1 到 0 的跳变，或者从 0 到 1 的跳变。因此，Ojala 他们将“等价模式”做了这样的定义：如果某个局部二进制模式所对应的二进制循环数从 1 到 0 或者从 0 到 1 至多有两次的跳变时，则该局部二进制模式所对应的二进制串就称为一个统一模式类。例如 00000000、11111111、10001111 都是统一模式类的例子。其实，检验某种二值模式是不是统一模式的最简单方法是：将其和其移动一位后的二进制模式串按位相减然后求绝对值的和，有如下定义：

$$U(G_p) = |f(g_{p-1} \oplus g_c) \oplus f(g_0 \oplus g_c)| + \sum_{p=1}^{P-1} |f(g_p \oplus g_c) \oplus f(g_{p-1} \oplus g_c)| \quad \text{式(2-12)}$$

若某种模式计算后所得到的  $U(G_p)$  小于或等于 2，则将其归为统一模式。在实际应用中，上式也可以表示为这个二进制模式串和它位移后的模式串按位相与运算的结果。更进一步地，我们可以为基本模式和统一模式创建一个查找表。并将除统一模式类别以外的其它模式归为另一种类，也就是混合模式类。

通过采取这样的改进，二进制模式的种类将大大减少，并且不会丢失任何图像信息。而且模式的数量由原来的  $2^P$  个减少到现在的  $\frac{P(P+1)}{2} + 1$  种，这里参

数  $P$  表示图像邻域内的采样点的数目。就拿  $3 \times 3$  邻域内有 8 个采样点的例子来说，二进制模式由最初的 256 种减少为 58 种，这样使得特征向量的维数减少了很多，并且可以降低高频噪声所带来的影响。Ojala 他们同时也认为，等价模式描述了图像的边缘等重要模式，如图 2.3 的几种模式所示，其中，黑色圆点来表示二进制位 1，白色圆点来表示二进制位 0。等价模式类别占总的模式类别中的绝大多

数，如果利用这些等价模式类别和混合模式类别的统计直方图，能够抽取更能代表图像本质属性的相关特征。

这里以基本的  $LBP^1$  为例，其统一模式有 58 种，其只占有 256 种基本模式的 23%，这就意味着这些统一模式出现的概率会很大。通过实验数据表明，对于一般的数字图像，统一模式占总的模式种类中比较大的比例。但是，同时我们也会发现，由于不同图像的差异，统一模式并不能在所有的图像中都占到很高的比重，在一些图像中，LBP 统一模式所占的百分比不足 60%，有些甚至比其更低。Ojala 等人在他们的关于人脸纹理的实验中，他们发现当采用  $LBP^1_8$  描述子时，所有的模式中统一模式所占得比例小于 90%，而在  $LBP^2_{16}$  模式中统一模式所占的比例大约 70% 左右。

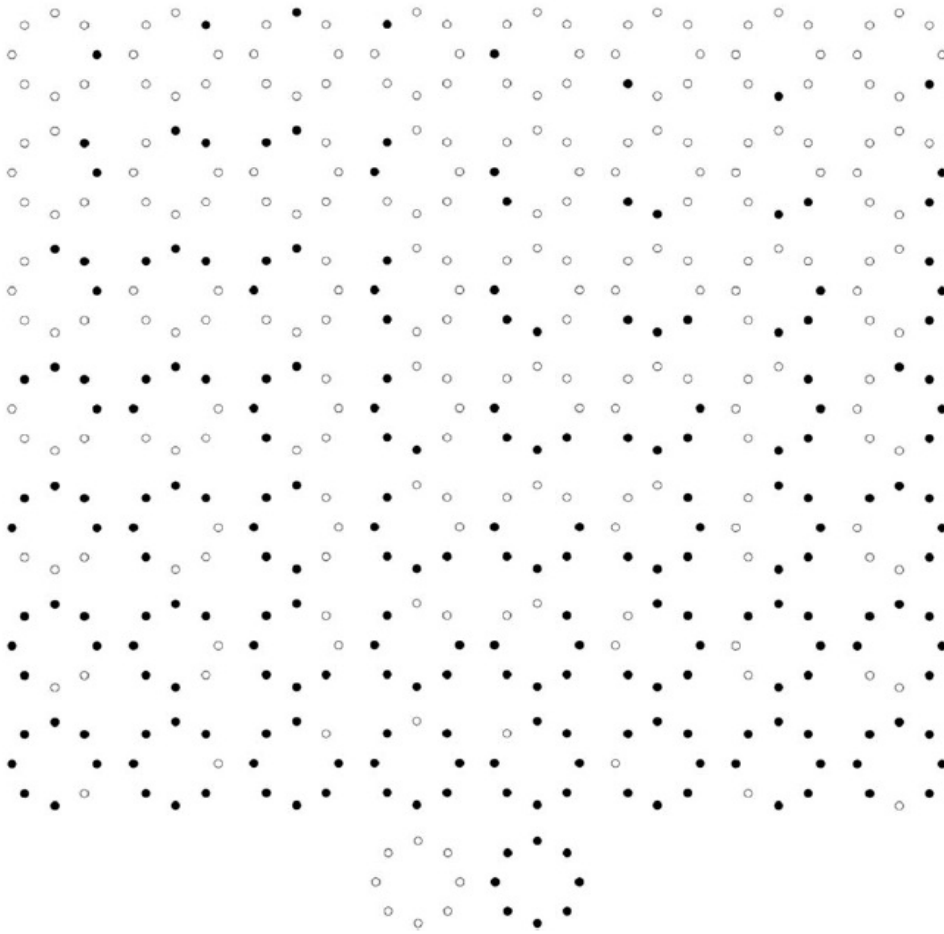


图 2.3  $P=8$  时的 LBP 统一模式

### 2.2.2 旋转不变的 LBP 描述子

从 LBP 的定义可以看出，LBP 描述子是亮度不变的，但它却不是旋转不变的。当对图像做旋转时，圆形邻域内的像素灰度值  $g_i$  在以  $g_c$  为中心， $R$  为半径



的圆周上运动。只要使  $f(g_i \square g_c)$  不全为 0 或者不全为 1，对图像的旋转就会产生不同的 LBP 值，我们可以从图 2.4 可以看出这一变化。因此，Maenpaa 等<sup>[32]</sup>人又将 LBP 描述子进行了一些扩展，他们提出了具有旋转不变性的 LBP 描述子，即通过不断旋转圆形邻域来得到一系列初始的 LBP 值，然后取其中最小值来作为该邻域的 LBP 值，用公式可以表示如下：

$$LBP_{P,R}^{ri} = \min\{ROR(LBP_{P,R}^{ri}, i) | i = 0, 1, K, P-1\} \quad \text{式(2-13)}$$

其中， $LBP_{P,R}^{ri}$  表示旋转不变的 LBP 描述子， $ROR(x, i)$  函数代表某个旋转函数，具体表示将  $x$  循环右移  $i$  ( $i < P$ ) 位。但对于邻域像素的标记顺序并没有太严格的约定，在这里，我们从中心像素的左上方的像素开始，顺时针标记。图 2.4 给出了求取旋转不变的 LBP 描述子过程的示意图，图中描述子下方的数字表示该描述子所对应的 LBP 值，图中所给出的 8 种 LBP 模式，经过旋转不变的处理后，最终得到的具有旋转不变性的 LBP 描述子的值为 15。也就是说，图中的 8 种 LBP 模式所对应的旋转不变的 LBP 模式都是 00001111。

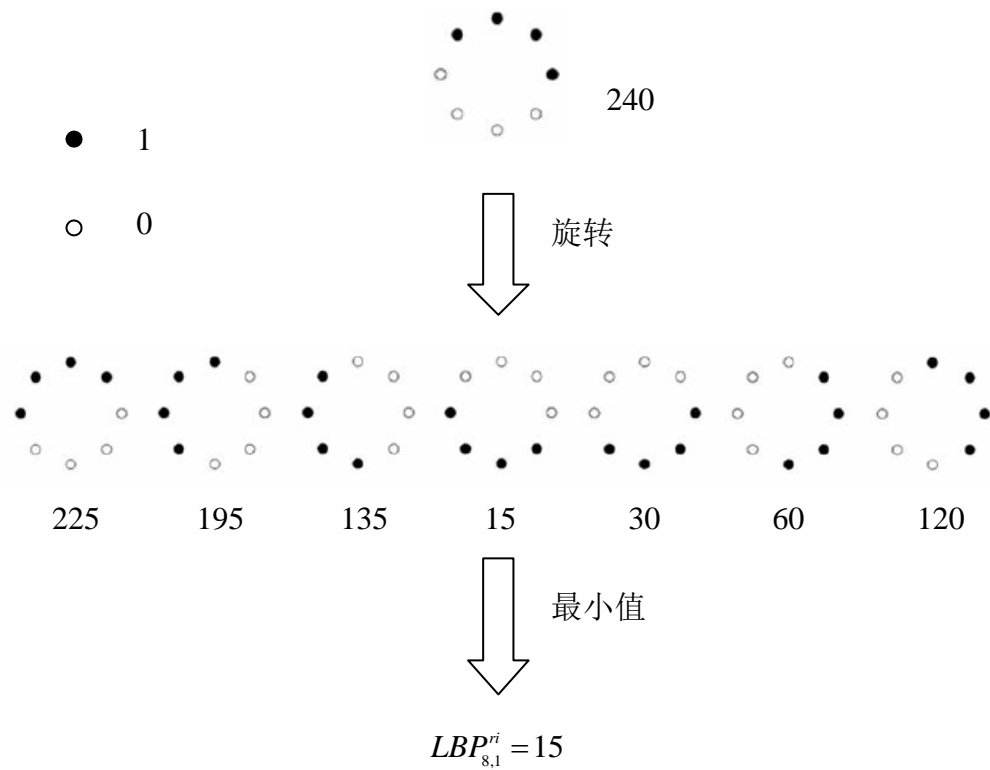


图 2.4 旋转不变 LBP 描述子示意图

通过加入旋转不变性的定义，LBP 描述子不仅对于图像旋转表现得更为鲁棒，并且 LBP 模式的种类也进一步减少，这使得纹理识别变得更加容易。但从另一方面来说，旋转不变的 LBP 描述子也存在最大缺陷，也就是其丢失了重要的方向信息，我们知道，对于某些场合来说，方向信息是相当重要的信息。然而，实际

中对于同种类纹理图像的分析,  $LBP^i$  描述子已经被证明是很有效的了。

此外,  $LBP^i$  描述子还可以和统一模式联合起来, 将等价模式类进行旋转从而得到具有旋转不变性的等价模式, 这将使可能的模式种类由原来的  $2^P$  个减少到  $P+1$  种。通常, 所有的非统一模式都被归为第  $P+1$  类, 有如下公式所示:

$$LBP_{p,k}^{riu2} = \begin{cases} \bigwedge_{p=0}^{P-1} f(g_p \square g_c), & \text{if } U(G_p) \leq 2 \\ P+1, & \text{otherwise} \end{cases} \quad \text{式(2-14)}$$

其中,  $U(G_p)$  表示从 0 到 1, 或者从 1 到 0 跳变的次数, 如公式(2-12)所示,  $LBP_{riu2}$  被称为具有旋转不变性的统一模式。经过统一模式和旋转不变的改进后, LBP 描述子除了具有了旋转不变性和亮度不变性的显著特点以外, 而且初始 LBP 的模式种类也得到较大幅度的减少。表 2.1 列出了几种不同的 LBP 描述子的模式维数。从表 2.1 可以看出, 经过统一模式和旋转不变模式的改进, 原始 LBP 模式的种类得到很大的降低。

表 2.1 几种 LBP 描述子的维数比较

LBP 类型	原始模式数	统一模式	旋转不变统一模式
$LBP_8^1$	256	58	9
$LBP_{16}^2$	65 536	542	17
$LBP_{24}^3$	16 777 216	554	25

### 2.2.3 LBP 的特点

由于纹理具有随机性、结构性的特点, 为了定量地描述纹理, 多年来人们建立了许多纹理分析算法来测量纹理的相关特性。概括起来, 这些方法可以分为两大类: 即统计分析法、结构分析法。统计分析方法是从图像的有关属性的统计分析出发, 主要是基于图像像素的灰度值的分布与相互关系, 来找出能反映这些关系的特征。其基本原理就是选择不同的统计量来对纹理图像的统计特征进行抽取。因此这类方法一般原理简单、较容易实现, 但是适用范围受到一些限制, 主要适合纹理图像中那些没有明显的规律性的结构图像, 尤其适合于具有随机性的、非均匀性的结构。而结构分析方法是从分析纹理图像的结构, 来从中获取相关结构特征。这类方法首先将纹理看成是由许多纹理基本单元按照一定的位置规则组成的, 然后提取纹理基元并计算纹理基元的位置规律, 当然也有直接去探求纹理构成的结构规律的。其目前主要用数学形态方法来处理纹理图像, 主要适用于规则性和周期性的纹理。

事实上,任何纹理图像都包含了统计特征和规律特征这两种特征,因此,使用单一的统计分析方法或者结构分析方法都很难达到了人满意的分类效果<sup>[33]</sup>。然而 LBP 描述子可以认为是随机性方法和统计方法结合的代表。LBP 描述子并不是试图从一些单一的像素点来描述纹理,而是采用局部区域的模式来描述纹理。每个像素点都由一个与之最合适的局部邻域内的原始纹理形成的二进制数来标记。这样, LBP 模式也可以被认为是一种微观的结构。将结构方法和随机方法相结合正好始于这样的一个事实,也就是分配的微观结构可以被看作是一种统计布局规则。LBP 具有原始纹理和布局规则,因此, LBP 分布具有结构分析的特点;另一方面来说,这种分布又可以被看做是图像经过一种非线性滤波后的统计结果,这明显又具有统计分析法的一些特性。LBP 描述子为同时分析随机性纹理和周期性纹理开辟了一条有效的途径,因此, LBP 的方法能够广泛地应用到各类纹理图像的识别,并能够充分克服单一只采用统计方法或者结构方法的不足。

## 2.3 基于 LBP 的人脸纹理特征提取

LBP<sup>[30]</sup>是一种有效的纹理描述算子,它具有旋转不变性和亮度不变性的显著的特点。在近十几年的时间内, LBP 算子已经得到不断的发展和演化,并广泛地应用于纹理分类<sup>[31]</sup>、纹理分割<sup>[34]</sup>、人脸图像分析<sup>[27][29][35]</sup>等领域。Ahonen 等<sup>[29]</sup>将 LBP 描述子引入了人脸识别中,他们将人脸图像划分为几个互不重叠的区域,并利用 LBP 的局部特性和直方图的统计特性,在一定程度上实现了局部特征和全局特征的结合,该方法在 FERET 人脸库上取得了较好的实验结果。此后,基于 LBP 的人脸识别算法引起许多研究者的极大兴趣,并得到了广泛的关注。

在本节中讨论基于局部二值模式 (LBP) 和分块的 LBP (Block-LBP) 的人脸纹理特征提取。

### 2.3.1 基于 LBP 人脸纹理特征提取 局部二值纹理模式技术广泛应用于纹理分析,

近年来,该方法已经成功应用于人脸识别。从原始图像任取一点作为阈值,采用 LBP 算子计算后得该像素点的 LBP 值。将全部的像素点进行 LBP 运算后得到的仍然是一副图像,如图 2.5 (c)所示,可以看出经过 LBP 操作后,所改变的只是各个像素点的取值,我们将经过 LBP 运算后的图谱称为 LBP 图谱。在实际的 LBP 运用中,一般不将 LBP 图谱作为特征向量,然后用于分类识别,而是通过求取 LBP 图谱的统计直方图作为人脸美丽的纹理特征。

对于一幅人脸图像,为了能够更好的保持人脸的空间结构,通常提取人脸局

部特征的直方图信息来描述该人脸区域的特性。首先，需要对人脸进行分块，将每个人脸图像在空间上划分为多个互不重叠的矩形区域；然后，分别对每个区域的 LBP 编码进行直方图统计，并采用该直方图特征表示该区域的特征；最后，将人脸图像各个区域的统计直方图信息串联起来，组成直方图序列作为最终人脸纹理特征。图 2.6 给出了人脸图像的几种矩形区域划分的方法。

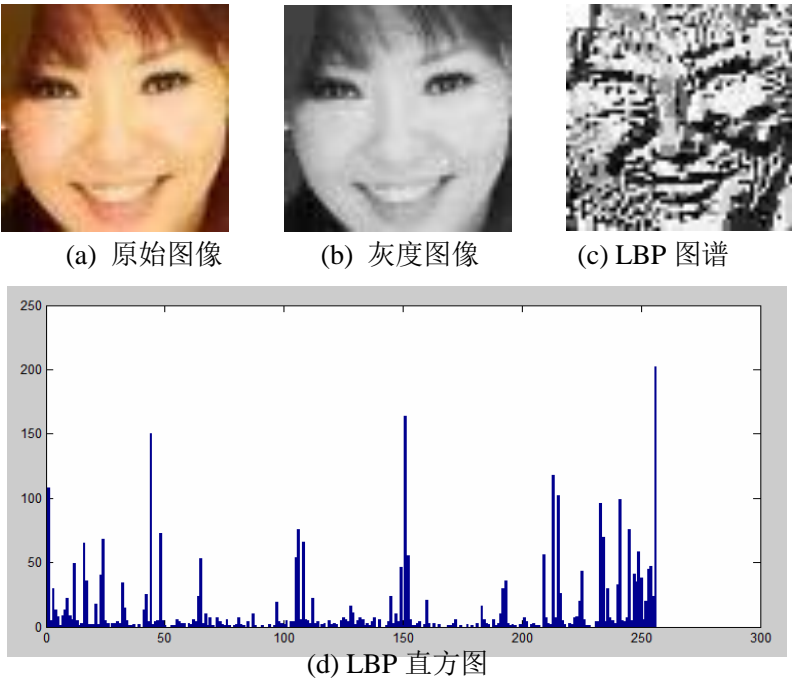


图 2.5 LBP 图谱和 LBP 直方图



图 2.6 人脸图像划分为 8·8，4·4，3·3 个子区域

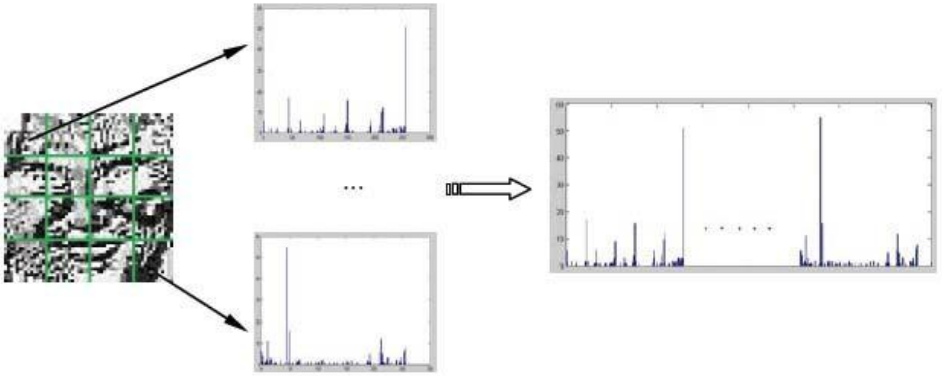


图 2.7 人脸图像各区域子直方图串联形成的整体直方图特征

在这里，对上述描述过程也可用表示如下的数学描述表示。

对于人脸图像  $I(x, y)$  的第  $k$  块子区域，在  $n$  个灰度级别上来统计直方图特征，其每个直方图分量  $h_{k,l}^j$  可定义为：

$$h_{k,l}^j = \sum_{(x,y) \in k} J\{LBP(I(x, y)) \square L_j\}, \quad j = 1, 2, \dots, n \quad \text{式(2-15)}$$

其中函数  $J\{\square\}$ ：

$$J\{B\} = \begin{cases} 0, & B \text{ is false} \\ 1, & B \text{ is true} \end{cases} \quad \text{式(2-16)}$$

这里， $L_j$  表示第  $j$  个灰度级别，一共选取了  $n$  个灰度级别，区域  $h_{k,l}$  的直方图统计特征可表示为：

$$h_{k,l} = (h_{k,l}^{L_1}, h_{k,l}^{L_2}, \dots, h_{k,l}^{L_n}) \quad \text{式(2-17)}$$

最后，将所有区域的局部直方图特征串连起来，所组成直方图序列来表示给定人脸图像  $I(x, y)$ ，即：

$$H_I = (h_{0,l}, h_{1,l}, \dots, h_{N \square, l}) \quad \text{式(2-18)}$$

其中， $N$  为人脸图像的分块区域数。图 2.7 给出了提取为人脸图像整体直方图特征的具体流程。

### 2.3.2 LBP 人脸纹理特征匹配

在基于 LBP 的人脸识别算法中，通常采用基于直方图的相似性度量，然后采用最近邻方法来分类。是用 LBP 提取图像直方图特征后，需要一种相似性度量的方法来计算两个直方图之间的相似度，在下面，将给出几种直方图的相似性度量最常用的方法。直方图的度量方法主要有这三种：直方图相交、Log 概率统计和 Chi 概率统计<sup>[36]</sup>。

对于任意两幅图像  $I_1$  和  $I_2$ ，用  $X$  和  $Y$  分别表示它们的直方图特征，则有如下的一些数学关系：

直方图相交：

$$D(X, Y) = \sum_{i=1}^L \min(X_i, Y_i) \quad \text{式(2-19)}$$

Log 概率统计：

$$L(X, Y) = \sum_{i=1}^L X_i \cdot \log Y_i \quad \text{式(2-20)}$$

Chi 概率统计:

$$\chi^2(X, Y) = \sum_{i=1}^L \frac{(X_i - Y_i)^2}{X_i + Y_i} \quad \text{式(2-21)}$$

其中,  $X_i$  和  $Y_i$  分别表示了两个 LBP 直方图的第  $i$  个 bin 的概率分布,  $L$  表示 LBP 直方图的总的 bin 数。先前的相关实验表明  $\chi^2$  概率统计性能优于直方图相交, 但是直方图相交计算量小。在我们的试验中, 我们选用直方图相交方法完成我们的工作。

## 2.4 小结

本章主要介绍了 LBP 描述子的原理及其发展。首先从纹理分析的方面对 LBP 描述子做了简要的概括; 其次, 重点介绍了 LBP 统一模式、旋转不变的 LBP 及其等价模式等; 接着对 LBP 的特点进行了概述。通过对 LBP 描述子的产生和发展进行深入研究, 我们得到了 LBP 的两个重要性质, 即旋转不变性和亮度不变性, 此外, LBP 同时具有随机分析和结构分析的特点, LBP 的这些特点对于提取图像的本质特征是十分有利的。最后, 分别采用基本的 LBP 和分块的 LBP 策略提取了人脸美丽图像的纹理特征。

## 第三章 基于 SIFT 和 SPM 的特征提取

尺度不变特征变换 (Scale Invariant Feature Transform, SIFT) 是 David Lowe 于 1999 年提出的局部特征描述子<sup>[37]</sup>, 并在 2004 年进行了进一步的发展和完善<sup>[38]</sup>。SIFT 特征独特性能好、信息量丰富, 并且对大多数图像变换有很强的不变性。在对包括 SIFT 算子在内的十几种局部描述子所做的不变性对比实验中, SIFT 描述子已被证实同类描述子中具有最强的键性。但从总体上归纳, SIFT 描述子具有以下特点:

(1) SIFT 特征是图像的局部特征, 对平移、旋转、尺度缩放、亮度变化等具有良好的不变性, 对视角变化、仿射变换也保持一定程度的稳定性。

(2) 独特性 (Distinctiveness) 好, 具有的丰富信息量, 适用于在海量特征数据库中进行快速、准确的匹配。

(3) 多量性, 即使少数的几个物体也可以产生大量 SIFT 特征向量。

(4) 速度相对较快, 经优化的 SIFT 匹配算法甚至可以达到实时的要求。

(5) 可扩展性, 可以很方便的与其他形式的特征向量进行联合。目前主要的人脸识别方法是基于传统的统计学习方法。然而由于大规模的人脸训练样本不容易得到, 并且传统的统计学习方法计算效率不太高, 因而基于单个样本的人脸识别策略开始引起大量关注。恰恰 SIFT 算法正好提供了这样一条崭新的道路。本文提出结合 SIFT 特征描述子和空间金字塔 (SPM) 策略对人脸图像美丽特征进行提取, 进而采用视觉词汇模型的方法完成人脸识别工作。

### 3.1 SIFT 特征提取

SIFT 算法首先在尺度空间进行特征检测来确定关键点 (Key Points) 的位置和关键点所处的尺度信息, 然后采用关键点邻域梯度的主方向作为该点的方向特征, 以实现描述子对尺度和方向的无关性。总的来说, 一幅图像的 SIFT 特征向量的生成算法总共包括以下 4 步:

(1)、检测尺度空间极值点, 初步确定关键点位置和所在尺度。

(2)、精确确定关键点的位置和尺度, 同时去掉低对比度的关键点、不稳定的边缘响应点, 以增强匹配稳定性、提高抗噪声能力。

(3)、为每个关键点指定方向参数, 使算子具备旋转不变性。

(4)、关键点描述子的生成, 即生成 SIFT 特征向量。下面分别对算法细节加以介绍。

图 3.1 给出了提取 SIFT 描述子的基本流程:

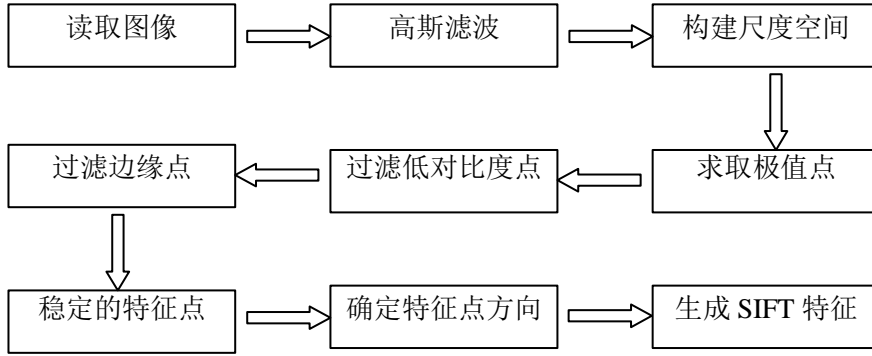


图 3.1 提取 SIFT 描述子的基本流程

### 3.1.1 构建尺度空间

尺度空间 (Scale Space) 的理论最早就出现在计算机视觉领域,当时提出的主要目的是模拟图像数据的多尺度特征。随后有些研究者又证明了尺度空间的滤波过程,从而证明了高斯核是实现尺度变换的唯一变换核。一幅图像的尺度空间就是由这幅图像在不同半径的高斯模糊下得到的结果来表示的。越接近底层的图像尺度层会越清晰,反而越接近顶层的图像尺度层就会越模糊。

一幅二维图像  $I$  的尺度空间可定义为将输入的图像和不同的高斯核做卷积运算,可表示如下:

$$L(x, y, \sigma) = G(x, y, \sigma) \square I(x, y) \quad \text{式(3-1)}$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad \text{式(3-2)}$$

其中,  $G(x, y, \sigma)$  是尺度可变的高斯核函数,  $(x, y)$  是空间的位置坐标,  $\sigma$  是尺度坐标。Lowe<sup>[38]</sup>在 SIFT 算法中采用了 DoG (Difference of Gaussian, 高斯差分) 尺度空间。DoG 是归一化的 LoG (Laplacian of Gaussian, 高斯拉普拉斯算子) 尺度空间的近似表示, 其是相邻尺度的图像相减而得到的高斯差分尺度空间, 其最大的优点是计算简单, 其数学表示为:

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) \square G(x, y, \sigma)) \square I(x, y) \\ &= L(x, y, k\sigma) \square L(x, y, \sigma) \end{aligned} \quad \text{式(3-3)}$$

其中,  $k$  是常数。



在实际的尺度不变特征点的提取中，SIFT 描述子将图像金字塔引入了尺度空间，图 3.2 给出了对图像金字塔的构建示意图。首先，采用不同尺度因子的高斯核函数对图像进行卷积来得到图像的不同尺度空间，将这一组图像作为金字塔图像的第一阶（octave）。接着对其中的 2 倍尺度图像，即相对于此阶的第一幅图像的 2 倍尺度，以 2 倍像素距离来进行下采样处理得到第二阶金字塔图像的第一幅图像，再对该图像使用不同尺度系数的高斯核函数进行卷积操作，来得到金字塔图像第二阶的一组图像。再以第二阶金字塔图像中的 2 倍尺度图像，其相对于此阶的第一幅图像的 2 倍尺度，以 2 倍像素的距离进行下采样来得到第三阶金字塔图像的第一幅图像，对该图像采用不同尺度系数的高斯核函数进行卷积，以获得金字塔图像第三阶的一组图像。这样依次类推，最终获得了需要的高斯金字塔图像。对每一阶相邻的高斯图像进行相减，就获得了高斯差分图像，也就是 DoG 图像。通常对 DoG 尺度空间中每个点与其相邻尺度、相邻位置的像素点逐个进行比较，所得到的局部极值位置即为特征点所处的位置及其对应的尺度。

总之，图像的高斯滤波保证了特征点不易受到噪声的影响，DoG 图像保证特征点不受亮度差的影响，高斯差分图像空间提取极值点也保证了尺度不变性。

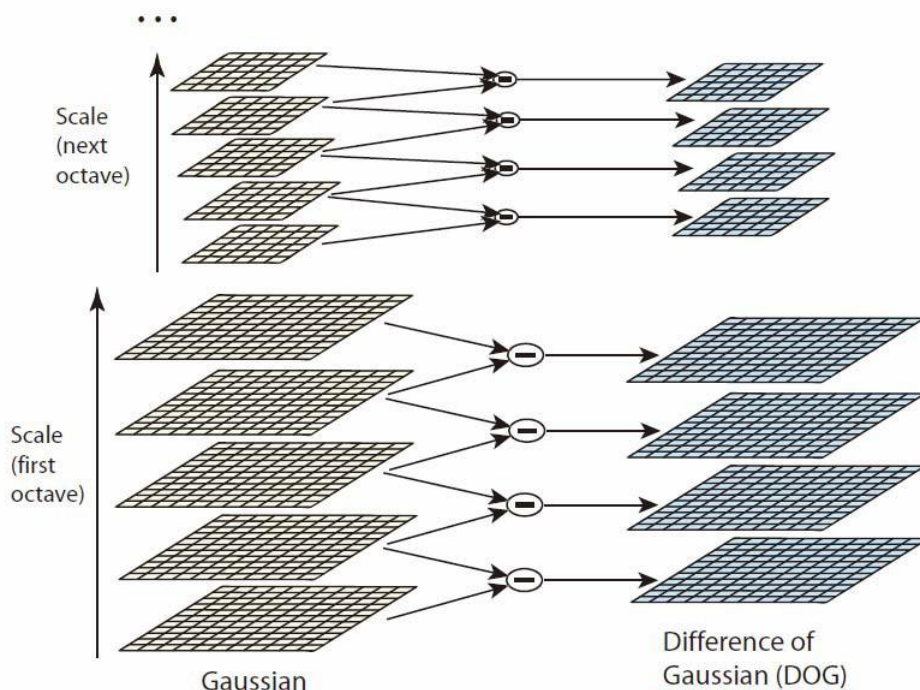


图 3.2 高斯图像与 DoG 尺度空间的构建

### 3.1.2 尺度空间极值点检测

对于图像上的特征点，计算其在每一尺度的下 DoG 的响应值，然后将这些值连接起来得到特征尺度轨迹曲线。特征尺度轨迹曲线的局部极值点即为该特征的

尺度。尺度曲线上完全有可能存在多个局部极值点，此时可认为该点有多个特征尺度。Lowe 在图像二维平面空间和 DoG 尺度空间中同时检测局部极值来作为特征点，这样使得特征具备良好的独特性和稳定性。

为了检测局部极值点，需要比较 DoG 尺度空间图像中的每个像素和它近邻的 26 个像素的值，这 26 个点是一个尺度空间图像中的 9 个像素点加上同尺度空间图像中的 8 个像素点，再加上下一个尺度空间图像的 9 个像素点。如图 3.3 所示，通过对每个像素点在其图像间及其 DoG 尺度空间的邻域中搜寻极值点，初步获得特征点的位置。如图 3.3 所示，在中间层用 X 来表示待检测点，同尺度的 8 个相邻点加上上下各相邻尺度对应的 18 个点总共 26 个点来构成的邻域(立方体) 中进行比较，来确保在尺度空间和二维图像空间都能够检测到极值点。这些极值点将作为粗糙的特征点参加下一轮的处理。

如果一个像素  $(x, y)$  是个局部极值点，则它必须是与其邻近的 26 个像素比较后得出的极值点。我们所需要的 SIFT 候选的关键点的集合，就是这样的所有的局部极值点的集合。

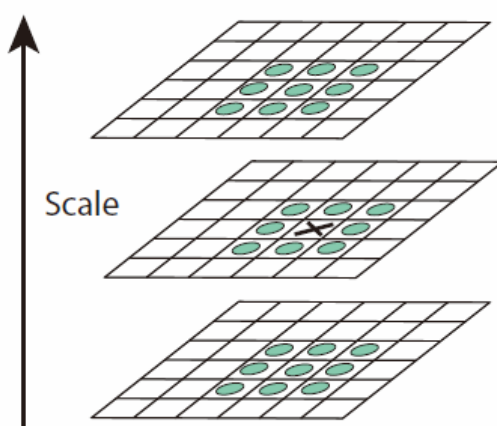


图 3.3 DoG 尺度空间局部极值检测

3.1.3 确定特征点的位置 通过极值检测方法得到的是候选关键点的集合。但是

为了得到更稳健的特征

点，还需要通过两步检测：(1)、这个关键点不能是低对比度的点，换句话说，关键点的亮度值必须要和其周围的点有明显的差异性；(2)、同时它也不能是边缘点，由于 DoG 算子可能会产生较强的边缘响应，由此而产生的边缘点是不稳定的，去除边缘点可以增强匹配的稳定性并能提高抗噪声能力。

由于 DoG 值对噪声和边缘较敏感，为了精确的确定特征点的位置和尺度，我们对上一步中检测到的极值点再进行三维二次函数拟和处理。使用 Taylor 极数将

尺度空间方程  $G(x, y, f)$  展开:

$$D(X) = D + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X \quad \text{式(3-4)}$$

其中,  $X = (x, y, f)^T$ ,  $\frac{\partial D}{\partial X} = \left[ \frac{\partial D}{\partial x}, \frac{\partial D}{\partial y}, \frac{\partial D}{\partial f} \right]$

$$\frac{\partial^2 D}{\partial X^2} = \begin{bmatrix} \frac{\partial^2 D}{\partial x^2} & \frac{\partial^2 D}{\partial xy} & \frac{\partial^2 D}{\partial xf} \\ \frac{\partial^2 D}{\partial yx} & \frac{\partial^2 D}{\partial y^2} & \frac{\partial^2 D}{\partial yf} \\ \frac{\partial^2 D}{\partial fx} & \frac{\partial^2 D}{\partial fy} & \frac{\partial^2 D}{\partial f^2} \end{bmatrix} \quad \text{式(3-5)}$$

对公式(3-4)关于  $X$  求一阶导数, 并令其等于零, 就可以得到函数  $D(X)$  的极值位置  $\hat{X}$ :

$$\hat{X} = - \frac{\frac{\partial D}{\partial X}}{\frac{\partial^2 D}{\partial X^2}} \quad \text{式(3-6)}$$

进一步为了去除对比度低的点, 只需把公式(3-6)代入公式(3-4)中, 得到:

$$D(\hat{X}) = D + \frac{1}{2} \frac{\partial D^T}{\partial X} X \quad \text{式(3-7)}$$

如果  $|D(\hat{X})| < 0.03$ , 就可以认为这个点是低对比度的点, 应该从特征点中剔除。

DoG 算子对边缘信息较敏感, 它可以提取到图像中的边缘、角点信息等, 但是也可能会带来一些不稳定的边缘点。高斯差分算子的极值在横跨边缘地方有较大的主曲率, 而在垂直边缘方向有较小的主曲率, 要寻找主曲率不满足条件的边缘点也就是不稳定的边缘点, 然后去除这些点来提高算法的稳定性。为了提高算法抗噪性能并同时增强匹配的稳定性, 通常采用海森矩阵 (Hessian) 来减少 DoG 函数的不稳定边缘点的方法。其具体流程如下所示:

边缘点的主曲率可以通过一个  $2 \times 2$  的 Hessian 矩阵  $H$  求出:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad \text{式(3-8)}$$

其中的导数  $D$  是由采样点的相邻之差来估计。并且  $D$  的主曲率和  $H$  矩阵的特征值成正比关系。由于 Hessian 矩阵  $H$  是  $2 \times 2$  的实对称矩阵, 故其必有 2 个实特征值。假定  $\lambda_{\max}$  为最大特征值,  $\lambda_{\min}$  为最小特征值, 则:

$$Tr(H) = D_{xx} + D_{yy} = \lambda_{\max} + \lambda_{\min} \quad \text{式(3-9)}$$

$$Det(H) = D_{xx} D_{yy} - (D_{xy})^2 = \langle \otimes \rangle \quad \text{式(3-10)}$$

设  $r = \frac{\langle \otimes \rangle}{\langle \otimes \rangle}$ , 则有:

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\langle \otimes \rangle)^2}{\langle \otimes \rangle} = \frac{(r \langle \otimes \rangle + (r+1)^2)}{r \langle \otimes \rangle^2} = \frac{(r+1)^2}{r} \quad \text{式(3-11)}$$

由于  $(r+1)^2/r$  的值在  $\langle \otimes \rangle$  时取最小值, 随着  $r$  的增大而增大。因此为了检测某点的曲率是否在一定的域值范围内, 只需要检测下式:

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r} \quad \text{式(3-12)}$$

如果特征点不满足上述不等式关系 (一般取  $r=10$ ), 就说明这个点是可能的边缘点, 就应该把它删除。

#### 3.1.4 为特征点分配方向值

在得到了比较鲁棒的特征点以后, 为了能够使这些特征点具有旋转不变的特性, 需要计算出每个特征点的主方向, 使后面的特征描述子都可以建立在这个主方向上, 从而消除因为旋转变换而带来的不一致性。

对于每个特征点, 首先计算出高斯滤波后的图像在这个特征点的小范围邻域内的所有像素点上的梯度幅值和梯度方向, 我们可以采用差分运算来近似梯度的求导运算, 如下面公式所示:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad \text{式(3-13)}$$

$$\theta(x, y) = \arctan\left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}\right) \quad \text{式(3-14)}$$

在实际的计算中, 需要在高斯空间内的特征点的邻域内进行采样, 来创建梯度方向的方向直方图。直方图每 10 度分为一个格子 (bin), 共 36 个格子。然后将邻域内的每个采样点按照梯度方向归入适当的 bin 中, 以梯度幅值  $m$  作为贡献的权重。最后选择直方图的主峰值作为特征点的主方向, 选取幅值达到主峰值 80% 以上的局部峰值作为辅助方向。这样会使得一个特征点可能被指定多个方向, 可以增强匹配的鲁棒性。如图 3.4 所示, 对特征点  $P$  的邻域中各像素的梯度值和方向进行计算, 得到八个梯度范围。其中, 幅值最大处也就是峰值所指的那个方向, 就代表了邻域梯度的主方向, 来作为特征点  $P$  的主梯度方向。

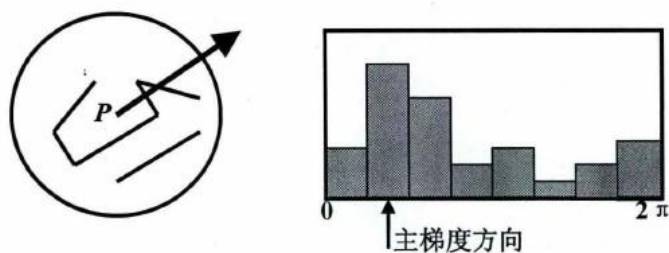


图 3.4 由梯度直方图确定的主梯度方向

综上所述，SIFT 算法提取到了四个重要参数：横坐标、纵坐标、尺度和主方向。主方向保证了特征点的旋转不变性，而尺度的提取方法也确保了特征点的尺度不变性，并且由该点的邻域可以确定一个兴趣区域（ROI）。

### 3.1.5 SIFT 特征描述子

接下来需要为检测到的每个特征点生成 SIFT 特征描述子，具体步骤如下：

(1) 首先以特征点为中心取  $16 \cdot 16$  的邻域作为采样窗口。

(2) 然后将采样窗口中的各像素点，根据坐标按高斯加权归入  $4 \cdot 4$  的区域网格，这里每个格子即为描述子的一维，窗口中的所有点对同一网格的贡献值之和来作为描述子在该维度的值。

(3) 将采样点与特征点的相对方向按照高斯加权，然后归入包含 8 个格子（bin）的方向直方图，这里每个格子同样作为描述子的一维，所有的点对同一个格子的贡献值之和来作为描述子在此维的值。这里需要注意创建方向直方图时使用的是采样点和特征点的相对方向，从而可以使 SIFT 特征具有旋转不变性。

(4) 由(2)和(3)，即获得  $128$  ( $4 \cdot 4 \cdot 8 = 128$ ) 维的 SIFT 描述子。这种由邻域内的采样点对方向直方图进行加权投票的思想不仅增强了算法的鲁棒性，同时也对特征点提取阶段的定位误差提供了较好的容错性。图 3.5 描述了对一个  $8 \cdot 8$  的网格窗口采样的情形，这时生成的描述子维数为  $2 \cdot 2 \cdot 8 = 32$ 。左图的中央为当前特征点的位置，蓝色的圈代表高斯加权的范围，每个小格代表窗口内的一个采样点，箭头方向表示该点的梯度方向，线段长度表示梯度幅值。

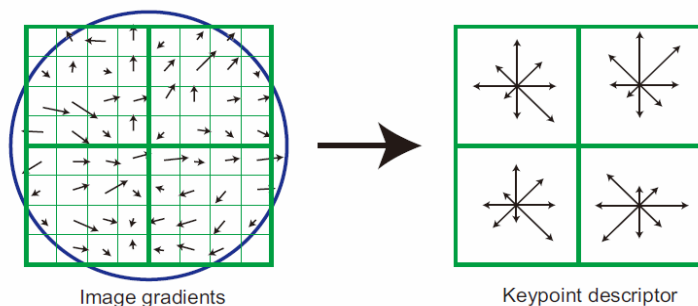


图 3.5 图像梯度和特征点描述子

此时, SIFT 特征向量已经去除了尺度变化、旋转等几何变形因素的影响。最后将特征向量的长度归一化, 以消除光照变化带来的影响。并且为了提高算法对非线性光照变化的鲁棒性, 需要将特征向量的每一维的大小限制在一个阈值  $Th$  (Lowe 的文章中推荐的值是  $Th = 0.2$ ), 若某一维的模大于这个值, 则令其等于  $Th$  并重新对向量进行归一化, 这样得到的才是最终的 SIFT 特征描述子。

图 3.6 分别给出了女性人脸图像和男性人脸图像的 SIFT 特征向量的示意图, 在这里对每个关键点采用  $4 \cdot 4 = 16$  总共 16 个种子点进行描述, 这样每个关键点的特征向量维数为 128 维, 长度表示 SIFT 特征向量的尺度, 箭头方向表示了 SIFT 特征向量的方向。

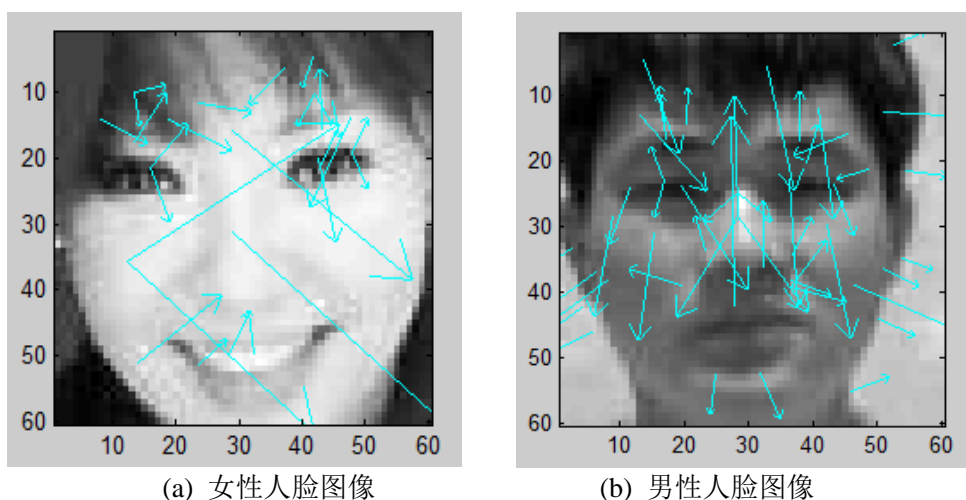


图 3.6 SIFT 特征向量的示意图

### 3.2 构建人脸图像的视觉词袋模型

词袋模型<sup>[39]</sup> (Bag of Words, BoW) 最早出现在文本分类的研究中, 并获得了很大的成功。在文本分类算法中, 首先可以通过统计每个已标注的正负例文本单词出现的词频形成向量, 然后使用这些正负例文本单词的词频向量对某些分类器进行训练获得一个文本分类器, 最后通过这个分类器对未标注类别信息的文本进行分类。文本分类器中的文本词频构建一般遵循以下的基本步骤: 首先, 通过语法分析算法把文本划分为一些单词的集合; 其次, 把某些具有相似意义的单词归为同一个单词, 例如把 “speak”、“speaker”、“spoke” 等归为同一个单词 “speak”; 再次, 把某些不具备类别信息的单词剔除, 比如单词 “a”、“an” 和 “the” 等; 最后获得文本对应的词频向量进行再文本分类操作。

BoW 模型用来表示图像模型, 我们就称之为视觉词袋模型<sup>[40]</sup> (Bag of Visual Words, BoV), 其需要将二维的图像信息映射成视觉关键词集合, 这样既保存了图像的局部特征又有效地压缩了对图像的描述。为了使用视觉词袋模型, 首先要

在学习阶段建立视觉关键词的集合，在训练图像集中，提取出所有局部特征（如 SIFT 特征）；然后对这些特征进行聚类，得到的结果是训练集中的普遍特征。我们称这些原型特征为“视觉词典”。

### 3.2.1 构建视觉词袋模型

BoV 模型的建立主要包括创建“视觉词典”，“视觉词”量化和投影的过程，最终获得“视觉词”的直方图表达。

(1) 提取人脸图像的局部特征，首先对数据库中的所有人脸图像进行 SIFT 特征点的密集采样。

(2) 学习“视觉词典”，对图像训练数据库中出现的所有 SIFT 特征点用 K-means 算法进行聚类，得到的聚类中心称为“视觉词”，把“视觉词”组成的集合叫做“视觉词典”。这样聚类以后特征相似的 SIFT 特征点会聚在一起。

(3) 量化图像特征和投影，将一幅图像的所有 SIFT 特征投影到“视觉词典”上得到“视觉词”出现的频率，图像可以由“视觉词”所组成的直方图表示。

本文在“视觉词典”构建过程中使用 K-means 聚类算法对图像显著区域的 SIFT 特征向量进行聚类，图像对应的“视觉词”个数目与代表图像各个聚类的聚类中心的数目一致，并且该图像的“视觉词”就是代表了聚类的显著区域特征向量。下面将介绍 K-means 聚类算法。

### 3.2.2 K-means 聚类算法

K-means 方法通过将  $N$  个观测值按照其与聚类中心的最近距离划分到  $K$  个聚类中心上<sup>[41][42]</sup>。其和高斯混合模型中的最大期望值算法类似，都是尝试通过不断地迭代求精找到数据的自然聚类。

给定一组观察值  $(x_1, x_2, \dots, x_N)$ ，其中每一个观察值都为  $d$  维实数值向量，K-means 聚类方法的目标是：将这  $N$  个观测值按照类内最小均方和的规则划分到  $K$  ( $K \ll N$ ) 个集合中，这里  $S = \{S_1, S_2, \dots, S_K\}$ ，优化的最小化目标函数：

$$S = \arg \min_S \sum_{i=1}^N \min_{x_j \in S_i} \|x_j - \mu_i\|^2 \quad \text{式(3-15)}$$

其中  $\mu_i$  为集合  $S_i$  的均值， $S = \{S_1, S_2, \dots, S_K\}$  便为我们寻找的  $K$  个聚类中心。

## 3.3 空间金字塔匹配

对于图像分类问题，BoV 模型<sup>[40]</sup>将图像表示成一个无序的局部特征集的特征



包方法,取得了令人关注性能<sup>[43][44]</sup>。然而,由于这些方法丢掉了所有的关于空间特征布局的信息,在描述性上具有一定的局限性。此外,BoV模型在字典学习中通常采用K-means等硬聚类方法;并从一开始就假设视觉词之间是独立的,特征点能以完全的概率投影到相应的概率中心上;而这显然是不符合实际的,如何将视觉词之间的相对位置关系融合到图像表达中,来弥补字典表达方式的不足一直是相关领域的科研人员的研究热点。

为了克服这些存在的问题,Lazebnik等<sup>[44]</sup>人提出了的空间金字塔匹配(Spatial Pyramid Matching, SPM)的策略,他们重新考虑了图像子区域的局部特征统计集合的全局不变的表达。通过引入基于核的识别方法,利用从Kurmyshev等<sup>[33]</sup>提出来的金字塔匹配策略改进相关的估计方法,并在计算全局范围内计算图像的相关性。这种方法通过将图像重复地划分成不断精细的子区域,并计算每个子区域中的局部特征的统计直方图。因而此算法产生的空间金字塔是无序的BoV模型的有效扩展,其表达方式简单并且计算不复杂。经过大量实验的比较可以发现,这种简单的操作可以明显的提高基本的特征包的表现性能,甚至超出了建立在复杂图形相关上的方法。此外,此算法简单、高效,在具有挑战性的数据上,诸如场景图像库、事件视频等,均表现出了出人意料的高的识别率。

### 3.3.1 金字塔匹配核算法

Grauman和Darrekk首次提出采用金字塔匹配核(Pyramid Match Kernel, PMK)算法来找两个向量集合间的近似对应关系<sup>[43]</sup>。简单的来说,金字塔匹配就是通过特征空间中放置一个越来越精细的网格序列并对不同分辨率级别下的匹配的数量之和进行加权。假定 $X$ 和 $Y$ 分别代表了 $d$ 维特征空间的两个向量的集合。在任意一个固定的分辨率上,如果两个点落在同一个方格中,就说这两个点是互相匹配的。更具体地,建立一个尺度为 $0,1,...,L$ 的网格序列,例如,在层次 $l$ 的网格的每个维数上有 $2^l$ 个单元格。定义 $H_x^l$ 和 $H_y^l$ 分别为这个分辨率下的 $X$ 和 $Y$ 的直方图。 $H_x^l(i)$ 和 $H_y^l(i)$ 表示从 $X$ 和 $Y$ 落在第 $i$ 个网格单元格内的点数。这样,在层 $l$ 上匹配的数目可用直方图交方法求得<sup>[43]</sup>:

$$I(H_x^l, H_y^l) = \sum_{i=1}^D \min(H_x^l(i), H_y^l(i)) \quad \text{式(3-16)}$$

为了方便表示,我们下面把 $I(H_x^l, H_y^l)$ 简写成 $I^l$ 。

由于在第 $l$ 层上的匹配的数量同样包括在更精细的第 $l+1$ 层的所有匹配的数量中,故,在 $l$ 层上的新的匹配的数量由 $I^l \square I^{l+1}$ 得出, ( $l=0,1,...,L \square 1$ )。第 $l$ 层的权重设为 $1/2^{L-l}$ ,权重值和这个层次的单元格的宽度成反比。由于在大一些的单元格中包含了较多的不相似的特征,需要对其匹配的数量进行惩罚。通过综合



考虑以上各因素，可以得到如下的金字塔匹配核的定义：

$$\begin{aligned}
 k^L(X, Y) &= I^L + \bigoplus_{l=0}^{L-1} \frac{1}{2^{L-l}} (I^l \square I^{l+1}) \\
 &= \frac{1}{2^L} I^0 + \bigoplus_{l=1}^L \frac{1}{2^{L-l+1}} I^l
 \end{aligned} \tag{3-17}$$

### 3.3.2 空间金字塔匹配策略

金字塔匹配核（PMK）是基于图像表达的无序的组合，它可使两个特征向量在高维空间的匹配变得更为精确，但是它会丢弃了空间信息。为此，Lazebnik 等人<sup>[44]</sup>提出了一种“正交”思想的方法，他们在二维图像空间中采用金字塔匹配，然后再在特征空间中用传统的 K-means 聚类方法。特别的，Lazebnik 将 SIFT 特征向量量化为  $M$  个离散的类型，也就是“视觉词典”的大小，并且假设只有同一类型的特征才能互相匹配。每一个通道  $m$  可以用两个二维特征向量集合  $X$  和  $Y$  表示，它们分别代表了类型  $m$  的特征在相应图像中的坐标。最终的核为各个子核的总和：

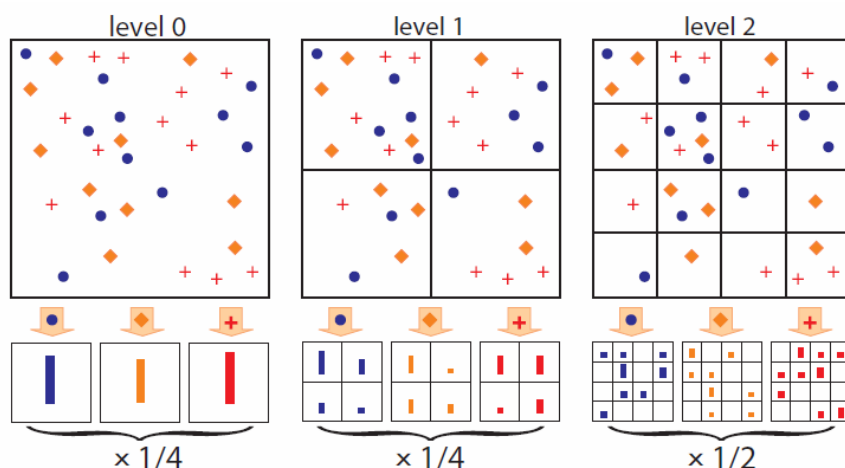
$$K^L(X, Y) = \bigoplus_{m=1}^M k^l(X_m, Y_m) \tag{3-18}$$

这种方法在保持“视觉词”的连贯性方面具有优势，实际上，当  $L=0$  时，该方法退化为视觉词袋（BoV）模型方法。图 3.7 给出了  $L=3$  层金字塔的图示。

由于所采用的金字塔匹配核是简单的直方图交集并求其加权和，而且通常对正数而言，有： $c \min(a, b) = \min(ca, cb)$ ，故可以通过将所有分辨率下的所有通道的近似加权直方图串联起来形成  $K^L$ ，进而形成为长向量形式的单一相交直方图。对具有  $L$  层和  $M$  个“视觉词”的图像，最终的向量维数为：

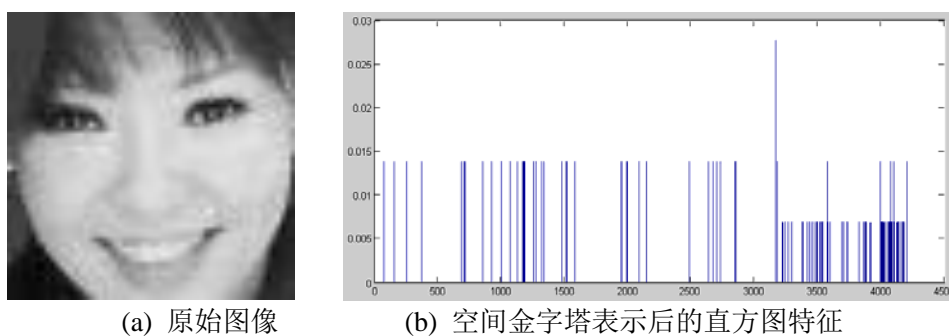
$$M \bigoplus_{l=0}^L 4^l = M \frac{1}{3} (4^{L+1} \square 1) \tag{3-19}$$

其中，计算核时的复杂度与特征维数成线性关系。在我们后面所进行的一系列实验中我们采用  $L=3$  和  $M=200$ ，来产生 4200 维的直方图交集。在计算中，由于直方图向量是非常稀疏的，故这些操作是非常高效的。

图 3.7 建立 3 ( $L=3$ ) 层金字塔的示例

图像有 3 种特征，分别用圆形、菱形和十字符号表示。在顶部，将图像在三个不同的分辨率级别上进行划分。然后，对于每个级别的分辨率和每一个通道计算落入到每一个空间仓库中的特征的数目，最后，根据公式 (3-17) 对每一个空间直方图进行加权<sup>[44]</sup>。

最后需要对特征向量进行归一化处理。为了获得最好的计算效率，采用图像中所有特征的权重之和来归一化所有的直方图向量，这样使得所有图像中的特征有相同的总数目。由于此方法使用的只是特征表示的密度，并不需要担心杂波等所带来的假特征，这个做法可以有效地处理图像尺寸的改变所带来的影响。例如图 3.8 所示，原始图像在 BoV 模型中经过空间金字塔表示后的直方图特征作为最终的特征向量。



(a) 原始图像 (b) 空间金字塔表示后的直方图特征  
图 3.8 原始图像在 BoV 模型中经过空间金字塔表示后的特征

### 3.4 小结

本章全面并详细的介绍 SIFT 描述子、视觉词袋模型 (BoV) 以及空间金字塔 匹配 (SPM) 策略，并在 BoV 模型下，提出了基于 SIFT 和 SPM 的人脸特征提取。

首先，描述了该算法的主要思想并给出了算法框架，列出了算子的主要特点。

其次，对 SIFT 算法的各个具体步骤，进行了详细分析和数学表示，包括尺度空间建立、特征点检测、特征点角度分配和描述、特征点匹配的每一步的具体实现过程，给出了设计流程图，并详细说明了如何实现。

再次，详细介绍了视觉词袋模型，主要包括“视觉词典”的创建、“视觉词”的量化和投影。

最后，提出了基于 SPM 的人脸特征提取方法，通过空间金字塔匹配策略，充分发掘了存在于人脸图像中的空间信息。



## 第四章 分类器设计

分类器设计是一般的分类系统的最后和最关键的一步，在图像处理和机器学习中起着重要作用。目前，有多种分类方法，主要包括神经网络方法、机器学习方法以及统计的方法等。在本节中主要介绍 K 近邻（K Nearest Neighbor, KNN）分类方法和支持向量机（Support Vector Machines, SVM）分类方法。

### 4.1 K 近邻算法

#### 4.1.1 KNN 原理

KNN 算法<sup>[45]</sup>最早是由 Cover 和 Hart 提出的，它是一种非参数的分类方法，现已被广泛应用于模式识别和人工智能的各个领域。其分类的基本思想是，给定一个待分类的样本  $x$ ，首先要找出与样本  $x$  最接近的或者最相似的  $K$  个已知类别标签的训练样本，然后再根据这  $K$  个训练样本的类别标签确定样本  $x$  的类别。算法基本步骤如下所示：

(1) 构建训练样本集合  $X$ 。

(2) 设定  $K$  的初值。 $K$  值的确定并没有一个统一的方法，一般方法是先利用经验知识确定一个初始值，然后根据实验结果不断调节，达到分类结果最优。

(3) 在训练样本集中选出与待测样本最近的  $K$  个样本。假定样本点  $x$  处于  $n$  维空间  $R^n$ ，样本之间的近邻一般是用欧氏距离来度量。假定第  $i$  个样本为  $x_i = (x_{i1}, \dots, x_{in}) \in R^n$ ，那么两个样本  $x_i$  和  $x_j$  之间的欧氏距离可以表示为：

$$d(x_i, x_j) = \sqrt{\sum_{m=1}^n (x_{im} - x_{jm})^2} \quad \text{式(4-1)}$$

(4) 给定一个待分类的样本  $x_b$ ，用  $x_1, \dots, x_K$  表示与  $x_b$  距离最近的  $K$  个样本，假设目标函数为： $f: R^n \rightarrow V$ ， $v$  表示第  $i$  个类别的标签，标签集合为  $V = \{v_1, \dots, v_s\}$ ，

$$\hat{f}_b(x_b) = \arg \max_{v \in V} \sum_{i=1}^K \text{TM}(v, f(x_i)) \quad \text{式(4-2)}$$

$$\text{TM}(p, q) = \begin{cases} 1, & \text{if } p = q \\ 0, & \text{otherwise} \end{cases} \quad \text{式(4-3)}$$

$\hat{f}_b(x_b)$  表示对  $f(x_b)$  的估计。

(5) 这里， $\hat{f}_b(x_b)$  便是待测样本  $x_b$  的类别信息。

#### 4.1.2 KNN 的优缺点

KNN 分类方法的主要优点有:

- (1) 算法简单直观, 容易实现。
- (2) 不需要产生额外的数据来描述规则, 它所采用的规则就是训练数据本身, 并不是要求数据的一致性问题, 因此可以存在一定程度的噪声。
- (3) 在做类别决策时, 其只与很少量的相邻样本有关。因此, 采用这种方法可以很好地避免样本数量的不平衡带来的问题。
- (4) 从分类过程来看, KNN 分类器最直接地利用了样本之间的关系, 从而减少了对类别特征选择不当对分类结果造成的一些不利影响, 这样可以最大程度地减少分类过程中的存在误差。

KNN 方法的不足之处主要包括:

- (1) 分类速度慢。最近邻分类器是基于实例学习的懒惰学习方法, 因为它是根据所给训练样本来构造的分类器, 首先需要将所有训练样本存储起来, 当需要进行分类时, 就实时地进行计算处理。需要计算待分样本与训练样本库中每一个样本的相似度, 然后才能求得与其最近的  $K$  个样本。对于高维样本或者样本库规模较大的情况, 其时间和空间复杂度较高。
- (2)  $K$  值的确定。KNN 算法必须预先指定  $K$  值,  $K$  值的选取比较费时间, 只能通过经验选取, 并且选择不当使分类精度不能得到保证。

### 4.2 支持向量机

支持向量机<sup>[46]</sup> (SVM) 是由 Vapnik 在 1995 年首次提出的, 是近几年来在统计学习理论中的 VC 维理论及其结构风险最小化原则的基础之上发展起来的一种高效的分类方法。它可以根据有限的样本信息在学习能力和模型的复杂度之间寻找最佳方案, 并且尽量提高学习模型的泛化能力, 目前它已广泛地应用于回归分析、模式识别、人工智能等领域。

SVM 的基本思想是, 首先通过非线性变换将原始的输入空间映射到一个高维的特征空间, 然后在这个高维空间中求其最优分类超平面, 这种非线性变换是通过定义某种合适的核函数来实现的。此外, 发展的支持向量机还可以处理线性不可分的情况下的分类任务。

#### 4.2.1 线性可分的支持向量机

SVM 是从线性可分情况下的最优分类超平面发展而来的, 其基本的思想可以

用图 4.1 所示的两维的二分类情况来说明。图中，实心点和空心点分别代表两类样本， $H$  为把两类正确地分开的分类线， $H1$ 、 $H2$  分别为各类中距离分类线  $H$  最近的样本，并且平行于分类线  $H$  的直线，它们俩之间的距离叫做最大分类间隔 (margin)。这里所谓的最优分类线就是要求分类线不但能将两类样本正确分开，即训练错误率为 0，而且要使分类间隔 margin 最大。如果推广到高维空间，最优分类线就变成了最优分类面。

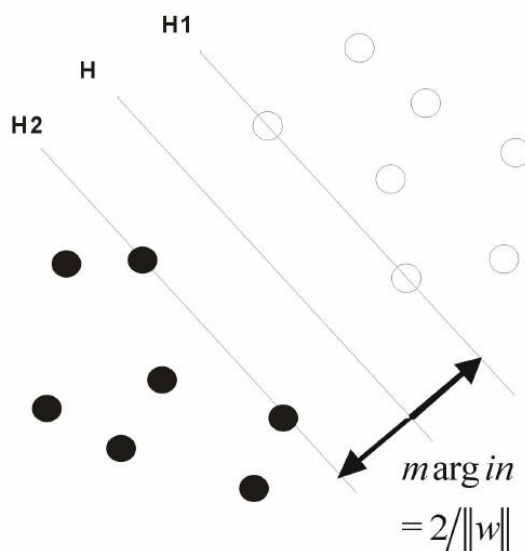


图 4.1 线性可分情况下的最优超平面

假定一组训练数据  $\{(x_i, y_i)\}, x_i \in R^N, x_i \in \{-1, +1\}, i=1, 2, \dots, n$ ，其中， $x_i$  表示第  $i$  个样本， $y_i$  表示样本  $x_i$  对应的类别。超平面  $H$  的方程为  $(w \cdot x) + b = 0$ 。如果超平面  $H$  能将两类样本无错误地分开，则使分类间隔最大的优化问题可用下面的约束优化问题来表示。其目标函数为：

$$\arg \min_{w, b} \frac{1}{2} \|w\|^2 \quad \text{式(4-4)}$$

约束条件为：

$$y_i [(w \cdot x_i) + b] \geq 1, i=1, 2, \dots, n \quad \text{式(4-5)}$$

其中，符号“ $\cdot$ ”表示向量点积操作。

由平面距离公式可知，分类间隔为  $\text{margin} = \frac{2}{\|w\|}$ ，要使它达到最大，则必须

使  $\|w\|^2$  在约束条件下取最小值，能使  $\frac{1}{2} \|w\|^2$  最小且满足公式(4-5)的分类平面就称为最优分类面。实际中，要在约束条件式(4-5)下求公式(4-4)的最小值，我们可以采用 Lagrange 函数：

$$L(w, b, \langle) = \frac{1}{2} \|w\|^2 + \sum_{i=1}^n \langle_i \{y_i [(w \cdot x_i) + b] - 1\} \quad \text{式(4-6)}$$

其中,  $\langle_i \in 0$  为各样本对应的 Lagrange 系数。求公式(4-6)的最小值, 根据 Wolfe 的对偶定义, 可以将上述问题转换为对偶问题。即在约束条件下:

$$\sum_{i=1}^n y_i \langle_i = 0 \quad \text{式(4-7)}$$

求下面目标函数的最大值的解  $\langle_i$ :

$$Q(\langle) = \sum_{i=1}^n \langle_i - \frac{1}{2} \sum_{i,j=1}^n \langle_i y_i \langle_j y_j (x_i \cdot x_j) \quad \text{式(4-8)}$$

其中,  $\langle_i$  为原始问题中与每个约束条件式(4-7)对应的 Lagrange 乘子。这实际上是一个不等式约束下的二次函数优化的问题, 故其存在唯一解。可以容易证明, 只有少部分的  $\langle_i$  不为零, 其所对应的样本特征向量  $x_i$  便是我们常说的支持向量 (Support Vector, SV)。

假设  $\langle_i^*$  是最优解, 则最优分类面的权系数向量为:

$$w^* = \sum_{x_i \in SV} \langle_i^* y_i x_i \quad \text{式(4-9)}$$

也就是最优分类面的权系数向量是训练样本中所有支持向量的线性组合。当我们得到支持向量和向量权系数  $w^*$  后, 对于分类器中的阈值  $b^*$ , 可以通过两类样本中任意一对支持向量并取其平均值得到。

$$b^* = \frac{1}{2} [w^* \cdot x^*(+1) + w^* \cdot x^*(-1)] \quad \text{式(4-10)}$$

其中,  $x^*(+1)$  和  $x^*(-1)$  分别表示两类样本中任意一个支持向量。

根据前面所得到的支持向量以及其相关的权系数等, 就可以得到上述问题的最优分类判别函数:

$$f(x) = \text{sgn} \left\{ (w^* \cdot x) + b^* \right\} = \text{sgn} \left\{ \sum_{x_i \in SV} \langle_i^* y_i (x_i \cdot x) + b^* \right\} \quad \text{式(4-11)}$$

#### 4.2.2 线性不可分的支持向量机

最优分类平面是在线性可分的前提下讨论的。然而, 对于线性不可分的情况, 就是某些训练样本不能满足公式(4-5)的约束条件, 为了能够解决这个问题, 可以在约束条件中增加一个松弛项  $\xi_i \in 0, (i=1, 2, \dots, n)$ , 此时, 软间隔分类超平面表示为在剔除那些错分样本后的最大分类间隔的超平面。



约束条件式(4-5)变为:

$$y_i[(w \cdot x_i) + b] \geq 1 + \xi_i, \xi_i \geq 0, i=1, 2, \dots, n \quad \text{式(4-12)}$$

目标函数式(4-4)改写为:

$$\arg \min_{w, b} \left( \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \right) \quad \text{式(4-13)}$$

其中的  $C$  为惩罚参数, 是某个给定的常数, 对错分样本惩罚程度起到控制的作用, 以达到在错分样本的比例和算法复杂程度之间的有效折衷。可以采用与求解最优分类面时相同的方法来求解公式(4-13)中的最优化问题, 同样是一个求二次函数的极值的问题, 其结果与线性可分情况下得到的公式(4-11)几乎完全相同, 只是条件关系式(4-7)变为:

$$\sum_{i=1}^n y_i \xi_i = 0, \quad 0 \leq \xi_i \leq C \quad \text{式(4-14)}$$

#### 4.2.3 非线性支持向量机

对于非线性情况, 支持向量机的基本思想是: 通过某种特定的非线性函数将训练样本由原始的输入空间映射到高维的特征空间, 也就是 Hilbert 空间; 然后在高维特征空间中构造出最优分类超平面, 使训练样本在高维空间中能够用线性判别函数来分类。一般说来, 对于非线性问题, 这种通过空间变换将非线性可分问题转化为高维线性可分时通常会引起“维数灾难”的问题。然后, 在非线性可分支持向量机中通过引入核方法巧妙地解决了此问题。

在处理非线性分类问题时, 支持向量机方法仅比线性分类情况多了一个非线性映射的环节。假定该非线性映射为:  $\Pi: \mathcal{H} \rightarrow H$ , 其中  $x$  为输入空间的输入向量,  $H$  为高维特征空间  $\mathcal{H}$  中的向量, 则公式(4-8)中的优化问题就可以转变为:

$$Q(\Pi) = \frac{1}{2} \left( \sum_{i=1}^n y_i \Pi(x_i) - \sum_{i,j=1}^n y_i y_j \Pi(x_i) \cdot \Pi(x_j) \right) \quad \text{式(4-15)}$$

在上面的对偶问题中, 只涉及到了训练样本之间的内积运算。根据泛函的有关理论, 只要一种核函数  $K(x_i, x_j)$  能够满足 Mercer 条件, 它就对应某一变换空间中的内积, 即  $K(x_i, x_j) = \Pi(x_i) \cdot \Pi(x_j)$ 。因此, 在最优分类面中通过选用适当的内积核函数  $K(x_i, x_j)$  就可以实现从低维空间向高维空间的映射, 从而实现了某一非线性变换经变换后的线性分类, 而计算复杂度并没有相应地增加。

此时的目标函数变可改写为:

$$Q(\gamma) = \sum_{i=1}^n \left[ \frac{1}{2} \sum_{j=1}^n \langle y_i y_j K(x_i, x_j) \rangle \right] \quad \text{式(4-16)}$$

对应的判别的分类函数变为:

$$f(x) = \text{sgn} \left\{ \sum_{x_i \in \text{SV}} y_i K(x, x_i) + b^* \right\} \quad \text{式(4-17)}$$

图 4.2 给出一个在输入空间中非线性可分的人造数据, 但通过高斯核函数后, 变得线性可分的样本。其中, 虚线上的特征点为支持向量 (SV)。

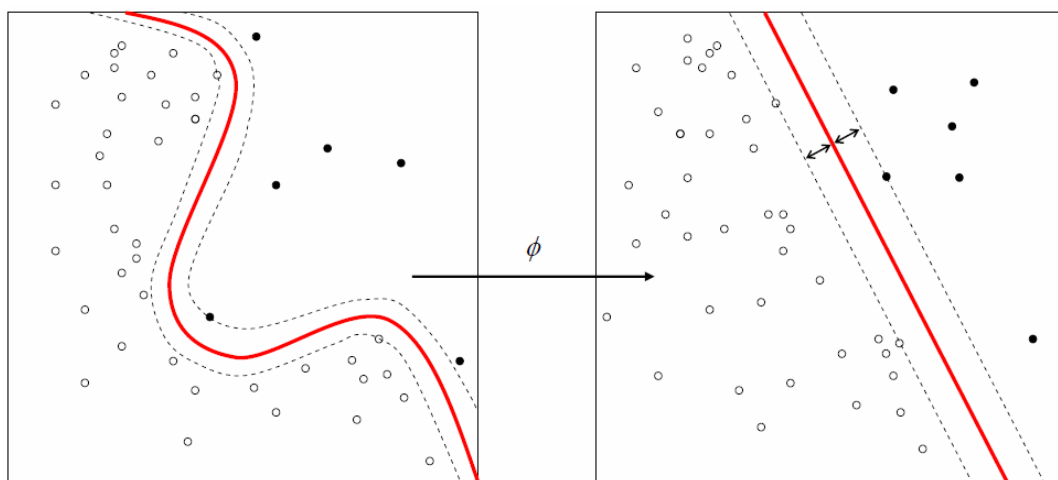


图 4.2 非线性可分支持向量机

由此可知, 对于是非线性可分问题, 利用核函数将样本集映射到某一高维空间, 使得样本集在高维空间中是线性可分的, 并将高维空间中的内积运算转化为原始空间中核函数的运算, 从而巧妙的避免了在高维空间中处理困难的问题。常见的核函数主要有下面几种:

(1)  $q$  阶多项式核函数:

$$K(x, x_i) = (x \cdot x_i + c)^q \quad \text{式(4-18)}$$

(2) 高斯径向基函数:

$$K(x, x_i) = \exp \left\{ -\frac{\|x - x_i\|^2}{2\sigma^2} \right\} \quad \text{式(4-19)}$$

(3) Sigmoid 核函数:

$$K(x, x_i) = \tanh \left( \gamma x \cdot x_i + c \right) \quad \text{式(4-20)}$$

(4) 直方图交核函数:

$$K(x, x_i) = \frac{1}{\|x - x_i\|} \quad \text{式(4-21)}$$

#### 4.2.4 支持向量机的优缺点

SVM 分类方法的主要优点有<sup>[47][48][49]</sup>:

(1) 解的唯一性和全局最优性: 支持向量机目标函数最终转化为了二次规划问题, 由凸二次规划的定理可知, 解是全局最优的并且唯一的。

(2) 泛化能力好: 由于 SVM 是基于结构风险最小化原则的, 从而保证了学习机器具有很好的泛化能力。

(3) 维数约减: 巧妙地解决了算法复杂度与输入向量维数密切相关的一些问题, 这大大克服了维数灾难问题。

(4) 小样本学习: 传统统计学习是基于样本趋于无穷大时来研究解的性质, 而支持向量机专门针对小样本情况, 它的最优解基于已有的样本信息, 而不是样本趋于无穷大时的统计的最优解。

SVM 分类方法的不足包括<sup>[47][48][49]</sup>:

(1) 需要求解二次规划问题: 核矩阵的规模是样本数的平方, 当样本数目比较大时, 存储核矩阵需要较大的存储空间, 从而用 SVM 进行训练数据造成了一定的困难, 不过这个问题随着计算机硬件的发展会有所改进。

(2) 核函数及其对应核参数的选择: 核函数模型及参数选择问题, 缺乏理论上的指导, 实际中大多靠经验的方式确定核函数及其模型参数。目前, 模型选择仍是一个公开的难题。

### 4.3 小结

本章首先介绍了 K 近邻算法的相关理论依据, 同时详细地分析了 K 近邻算法的优缺点和常用的一些改进方法。其次, 重点介绍了支持向量机分类算法, 支持向量机是建立在统计学习理论的基础之上, 可以寻找具有较好区分能力的支持向量, 由此构造的分类模型可以类与类之间的分类间隔达到最大化, 具有优秀的学习能力和泛化能力。同时由于采用了二次规划函数来寻优, 因而可以获得全局最优解。由于采用了核函数完成映射, 巧妙地解决了“维数灾难”的问题, 使得支持向量机的复杂性与样本维数无关, 因此非常适合于处理非线性分类问题。



## 第五章 实验结果与分析

本节通过一系列实验来验证提出的人脸美丽分类方法，主要包括人脸美丽数据库的创建，基于 LBP 纹理特征的人脸美丽分类、基于 SIFT 和 SPM 的人脸美丽分类、实验结果分析以及对本节的总结。

### 5.1 人脸美丽数据库

实验中采用的所有人脸图像均来自互联网的社交网站（如 Facebook）。为了验证我们提出的人脸美丽分类想法，我们创建了男性人脸美丽图像库和女性美丽图像库，图像大多数为正面，但存在着较大的姿势、年龄、种族等变化。男性人脸美丽图像库共有 5400 幅图像，图像大小为 60·60；女性人脸美丽图像库共有 5000 幅图像，图像大小也为 60·60。图 5.1 和 5.2 分别给出了 10 幅男性人脸图像样本和女性人脸图像样本。



图 5.1 男性人脸美丽图像

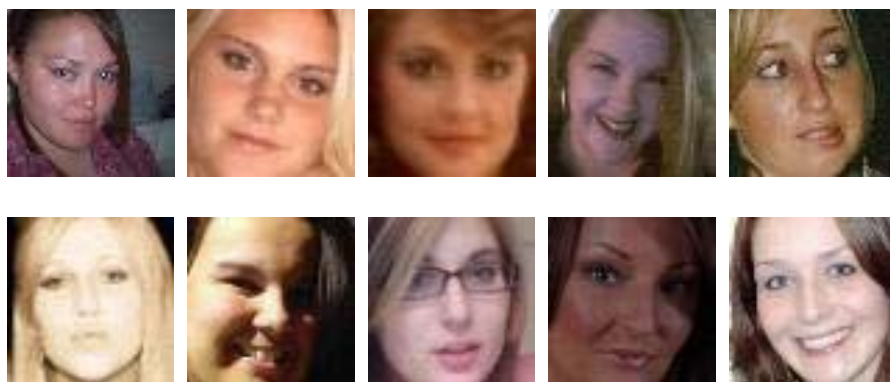


图 5.2 女性人脸美丽图像

更重要的是，为了对人脸图像的美丽程度进行学习和分类，我们就需要对人

脸图像的美丽程度进行人工打分,裁判队员由 5 名男性和 5 名女性总共 10 人组成,这 10 人分别为每幅图像的美丽程度进行打分,所给分数范围是 0 到 100 之间,表示由最不美丽到最美丽的过渡。在打分前,裁判队员首先浏览所有的图像,以便能够对图像库中图像的美丽吸引力有一个全面的掌握,然后每张随机出现的图像打分。最后,这 10 人的平均评分作为该人脸图像的美丽程度的分数。此外,我们又将分数归划分为 5 个等级,即 1、2、3、4、5,其具体含义如表 5.1 所示:

表 5.1 人脸图像美丽的等级及含义

评分等级	5	4	3	2	1
表示含义	优秀	好	中等	一般	不好

同时,图 5.3 给出了男性人脸美丽库和女性人脸美丽库中不同美丽等级的图像所占的比重,我们可看出,大多数人脸图像美丽处于美丽等级 4 中。

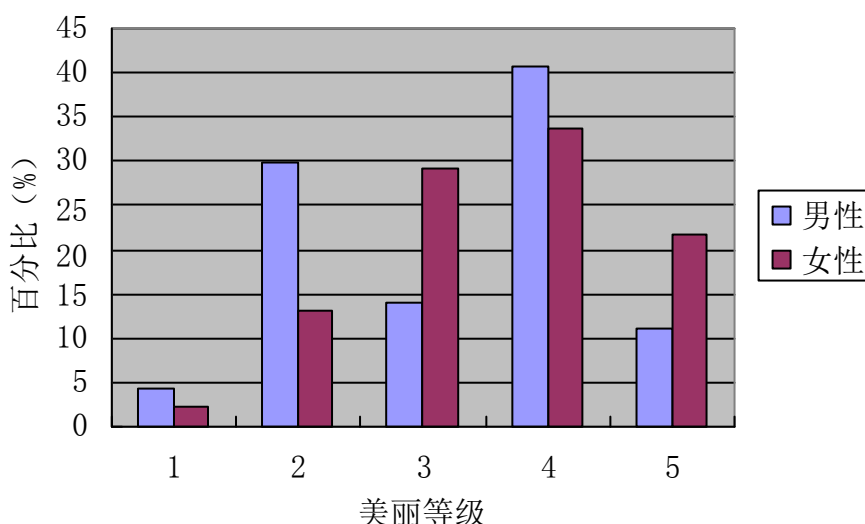


图 5.3 人脸库中各种美丽等级的图片所占的比例

## 5.2 基于 LBP 的人脸美丽分类

实验中在提取人脸的 LBP 纹理特征时,我们采用最基本的 LBP 描述子形式,即  $LBP^{8,1}$  这种类型的 LBP 特征。为了利用人脸图像的空间信息,对 LBP 图谱进行  $4 \cdot 4$  的分块 (Block),最后将各块的直方图串联起来得到总的直方图特征,为了方便,我们记做 Block-LBP。为了对 LBP 和 Block-LBP 纹理特征分类,我们采用了 K 近邻分类器和支持向量机分类器。

首先,为了验证 LBP 和 Block-LBP 的人脸美丽描述能力,我们做了如下实验,

我们随机选取 50% 的图像作为训练集，剩下的一半图像组成测试集，如此重复 5 次实验，然后求出这 5 次试验的平均分类结果作为最终分类结果。此外，在参数设置上，对于 KNN 分类器，我们选用  $K = 61$ ，而对于 SVM 分类器，我们采用 LibSVM<sup>[50]</sup> 工具包实现分类任务，并选用直方图交核，惩罚项  $C = 50$ 。实验结果如表 5.2 和表 5.3 所示。

表 5.2 LBP 和 Block-LBP 在男性人脸美丽库中分类性能比较 (%)

分类器	LBP	Block-LBP
KNN	37.06	37.69
SVM	30.90	33.95

表 5.3 LBP 和 Block-LBP 在女性人脸美丽库中分类性能比较 (%)

分类器	LBP	Block-LBP
KNN	32.08	32.69
SVM	26.47	28.75

从表 5.2 和表 5.3，我们可以观察到：(1)、Block-LBP 性能明显优于 LBP 特征，增幅在 0.6% ~ 3% 之间，然而这种优势在 KNN 分类器表现不是很明显；(2)、KNN 分类器在人脸美丽分类中识别率高于 SVM，最高达到 7 个百分点，显示出了 KNN 更适合了人脸美丽分类；(3)、SVM 能更好的挖掘图像特征的空间信息，引入空间信息后增幅达 3%；(4)、计算机对女性美丽程度的认知能力明显低于对男性美丽的认知，差距在 5 个百分点左右。

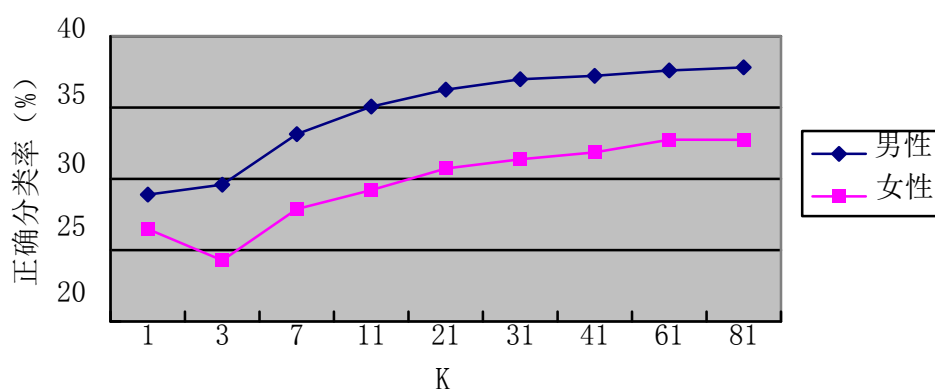


图 5.4 Block-LBP 在 KNN 分类器不同  $K$  下的分类结果

为了研究参数  $K$  对 KNN 分类器性能的影响，图 5.4 给出了 Block-LBP 特征在 KNN 分类器的不同的  $K$  下的分类结果。从图中可以看出，当  $K$  大于 61 以后，

人脸美丽分类率基本保持稳定了, 故在实验中我们选择  $K = 61$ 。

### 5.3 基于 SIFT 和 SPM 的人脸美丽分类

实验中在提取基于 SIFT 和 SPM 的特征时, 我们采用如下的参数: 在提取 SIFT 特征时, 在  $16 \cdot 16$  的图像块上提取一个 SIFT 描述子; 在视觉词袋模型中学习“视觉词典”时, 我们选用了  $M = 200$  个“视觉词”; 在空间金字塔匹配中, 采用了 3 层金字塔表示, 即  $L = 3$ , 这样经空间金字塔表示的图像特征的维数是 4200。同样为了方便表示, 对于没有经过空间金字塔表示的特征我们记为 SIFT, 而经过空间金字塔表示的特征记为 SPM-SIFT。然后, 我们采用了 KNN 分类器和 SVM 分类器完成人脸美丽的分类。

在实验设置上, 与上一节类似, 即随机选取 50% 的图像作为训练集, 剩下的图像组成测试集, 如此重复 5 次实验, 然后求出这 5 次试验的平均分类结果作为最终分类结果。还有, 在参数设置上, 对于 KNN 分类器, 我们选用  $K = 61$ , 而对于 SVM 分类器, 我们采用 LibSVM 工具包<sup>[50]</sup>实现分类任务, 并选用直方图交核, 惩罚项  $C = 50$ 。实验结果如表 5.4 和表 5.5 所示。

表 5.4 SIFT 和 SPM-SIFT 在男性人脸美丽库上分类性能比较 (%)

分类器	SIFT	SPM-SIFT
KNN	34.90	38.07
SVM	26.43	28.55

表 5.5 SIFT 和 SPM-SIFT 在女性人脸美丽库上分类性能比较 (%)

分类器	SIFT	SPM-SIFT
KNN	31.54	32.08
SVM	24.29	25.08

同样地, 从表 5.4 和表 5.5, 我们可以得到: SPM-SIFT 性能明显优于 SIFT 特征; KNN 分类器在人脸美丽分类中识别率高于 SVM; 计算机对女性美丽程度的认知能力明显低于对男性美丽的认知。

### 5.4 实验结果分析

从表 5.2、表 5.3、表 5.4、表 5.5 以及图 5.4 中的实验结果, 我们看出对人脸



图像美丽分类是一个挑战性的工作，因为人脸图像存在着光照、姿势、年龄、性别和种族的差异，这也使得我们的研究非常有意义。更重要的是，通过实验，我们有下面几个重要结论：

- (1) **Block-LBP** 性能优于 **LBP** 特征，**SPM-SIFT** 性能优于 **SIFT** 特征。
- (2) **KNN** 分类器在人脸美丽分类中识别率高于 **SVM**；
- (3) **SVM** 能更好地对图像特征的空间信息进行分类；
- (4) 计算机对女性美丽程度的认知能力明显低于对男性美丽的认知，这也与我们人类的认知基本相符合。

## 5.5 小结

本节通过一系列实验来验证提出的人脸美丽分类方法，首先介绍了人脸美丽数据库的创建，接着交代了实验设计中的参数设置，然后进行了基于 **LBP** 纹理特征的人脸美丽分类、基于 **SIFT** 和 **SPM** 的人脸美丽分类，在分类中，选用了 **KNN** 和 **SVM** 分类器，并对实验结果进行了分析和总结。



## 第六章 结束语

### 6.1 总结

美是人类所广泛感知的一个概念,人脸美是人类社会活动中最常感知的对象,人脸美能给人带来心灵的愉悦。人类对自身之美的研究已有几千年的历史了,人类对美的不断追求,促进了人的审美意识不断提高,同时也促进了社会的发展。认知心理学近年来大量研究实验发现人们对于什么脸是美的存在着高度的一致性,这种高度的一致与文化、种族、年龄、性别无关。

本文基于图像处理和机器学习方法提出了人脸美丽分类的研究。利用图像处理和机器学习方法来进行人脸美丽研究的报道比较少,但最近几年来已经开始受到不少研究者的重视。本文针对人脸美丽这一研究问题,并给出相应的客观的量化描述。主要工作包括以下四部分:

(1) 创建人脸美丽图像库。从一些流行的社交网站上收集大量的男性和女性人脸图像,建立较大规模的男性人脸美丽和女性人脸美丽数据库,其中男性人脸美丽库有 5400 幅,女性人脸美丽库中有 5000 幅图像,以便后续实验的进行,对图像进行灰度化处理,以消除光照的影响。

(2) 基于 LBP 的纹理特征提取。利用局部二值纹理模式(LBP)和分块的 LBP 描述子来提取人脸的纹理作为美丽特征,来表示人脸美丽的吸引力。

(3) 基于 SIFT 和 SPM 的特征提取。首先提取一幅图像的关键点作为 SIFT 特征,然后在视觉词袋模型下对每个 SIFT 特征是特征点进行向量量化,然后统计其词频,最后,我们采用空间金字塔匹配策略来对图像进行精细的划分,这样便引入了 SIFT 描述子的空间信息。

(4) 分类器的设计。对于一副给定的人脸图像,为了给出较客观的等级评价,我们使用 K 近邻(KNN)分类器和支持向量机(SVM)对提取的人脸美丽特征进行分类,然后对实验进行分析。

实验结果表明,提出的人脸美丽分类方法比较客观的评价人脸图像,验证了我们提出方法的可行性。并得出下面的几个结论:

- (1) Block-LBP 性能优于 LBP 特征,SPM-SIFT 性能优于 SIFT 特征。
- (2) KNN 分类器在人脸美丽分类中识别率高于 SVM;
- (3) SVM 能更好地对图像特征的空间信息进行分类;

(4) 计算机对女性美丽程度的认知能力明显低于对男性美丽的认知,这也与我们人类的认知基本相符合。

## 6.2 工作展望

在图像处理和机器学习领域，人脸图像美丽分类是一个新兴起的研究问题。本文的研究是对采用图像处理和机器学习方法对人脸美丽分类研究的一次有意义的尝试，虽然本文所作的一些研究，获得了一些结果，但仍然有许多问题需要进一步深入探讨。本文需要进一步研究的工作有：

- (1) 寻求更好的人脸美丽特征提取方法，如 LBP 的最新发展，Gabor 纹理，梯度直方图等特征。
- (2) 在人脸美丽特征提取过程中，加入颜色信息。
- (3) 在视觉词袋模型的，采用稀疏编码（Sparse Coding）方法来创建“视觉词典”和量化“视觉词”。

## 致谢

首先我要向我的导师韩保君副教授致以衷心的感谢！他为指导本文的工作倾注了大量心血，在我整个研究生期间的学习和论文工作中都得到了他的指点，给了我很多指导性的见解和富有启发性的建议。他对课题的整体把握为我工作的完成提供了最重要的帮助，并在我的研究生学习期间为我提供了许多方便和支持。在研究过程中他还教给了我很多科研的方法、思路，让我养成了良好的科研习惯。

此外，我还要感谢所有在我研究课题中给予过关心、指导和帮助的老师、同学和同事。

最后，感谢我的家人和朋友，正是他们自始至终的关心、理解、支持和鼓励，才使我坚持不懈地完成了学业。



## 参考文献

- [1] Aharon I., Etcoff N., Ariely D., *et al.* Beautiful Faces Have Variable Reward Value: FMRI and Behavioral Evidence [J]. *Neuron*, 2001, 32(3): 537-551
- [2] O'Doherty J., Winston J., Critchley H., *et al.* Beauty in a Smile: the Role of Medial Orbitofrontal Cortex in Facial Attractiveness [J]. *Neuropsychologia*, 2003, 41(1): 147-155
- [3] Rhodes G. The Evolutionary Psychology of Facial Beauty [J]. *Annual Review of Psychology*, 2006, 57(8): 199-266
- [4] Thornhill R., Gangestad S.W. Facial Attractiveness [J]. *Trends in Cognitive Sciences*, 2006, 10(1): 9-13
- [5] Langlois J.H., Kalakanis L, Rubenstein A.J., *et al.* Maxims or Myths Of Beauty? A Meta-analytic and Theoretical Review[J]. *Psychological Bulletin*, 2000, 126(3): 390-423
- [6] Koscinski K. Current Status and Future Directions of Research on Facial Attractiveness [J]. *Anthropological Review*, 2009, 72 (1): 45-65
- [7] Aarabi P., Hughes D. The Automatic Measurement of Facial Beauty [C]. *IEEE International Conference on Systems, Man and Cybernetics*, The Hague, The Netherlands, IEEE Press, 2001: 2644-2647
- [8] 毛慧云. 人脸美丽吸引力的特征分析与机器学习[D]. 华南理工大学博士学位论文, 2011
- [9] 孙凯. 计算人脸医学美学研究[D]. 哈尔滨工业大学硕士学位论文, 2009
- [10] 赵萍. 人脸美化技术的研究与实现[D]. 大连理工大学硕士学位论文, 2010 [11]
- Etcoff N. *Survival of the Prettiest: The Science of Beauty* [M], New York, Doubleday, 1999
- [12] Langlois J.H., Roggman L.A. Attractive faces are only average [J]. *Psychological Science*. 1990, 1(2): 115-12
- [13] Thomas R.A, Michael R.C. Averaged Faces Are Attractive, but Very Attractive Faces Are Not Average [J]. *Psychological Science*, 1991, 2(2): 123-125
- [14] Perrett D.I., Burt D.M. Penton-Voak I.S., *et al.* Symmetry and Human Facial Attractiveness [J]. *Evolution and Human Behavior*, 1999, 20(5): 295-307
- [15] Rhodes G, Proffitt F., Grady G.M., *et al.* Facial Symmetry and the Perception of Beauty [J]. *Psychonomic Bulletin & Review*, 1998, 5(4): 659-669

- 
- [16] Swaddle J.P., Cuthill I.C. Asymmetry and human facial attractiveness: Symmetry may not always be beautiful [J]. *Proceedings of the Royal Society, London*, 1995, 261(1360): 111-116
  - [17] Farkas L.G., Cheung G. Facial asymmetry in healthy north american caucasians: An anthropometrical study[J]. *Angle Orthodontist*, 1981, 51(1): 70-77
  - [18] Peck S., Peck L., Kataja M. Skeletal asymmetry in esthetically pleasing faces[J]. *Angle Orthodontist*, 1991, 61 (1): 43-48
  - [19] Zaidel D. W., Cohen J. A. The Face, Beauty, and Symmetry: Perceiving Asymmetry in Beautiful Faces [J]. *International Journal of Neuroscience*, 2005, 115(8): 1165-1173
  - [20] Eysenck S., Dror G., Ruppel E. Facial attractiveness: beauty and the machine [J]. *Neural Computation*, 2006, 18(1): 119-142
  - [21] Kagan A., Dror G., Leyvand T. A humanlike predictor of facial attractiveness [J], *Processing of Neural Information Processing Systems*, 2006, 649-656
  - [22] Gunes H., Piccardi M., Jan T. Automated Classification of Female Facial Beauty by Image Analysis and Supervised Learning [C]. *Proceedings of International Society for Optical Engineering*, 2004: 968-972
  - [23] Gunes H., Piccardi M., Jan T. Comparative beauty classification for pre-surgery planning systems [C]. *IEEE International Conference on Man and Cybernetics*, the Hague, The Netherlands, 2004: 2168-2174
  - [24] Chang F., Chou C.H. A Bi-prototype Theory of Facial Attractiveness. *Neural Computation* [J], 2009, 21(3): 890-910
  - [25] Davis B.C., Lazebnik S. Analysis of Human Attractiveness Using Manifold Kernel Regression [C]. *IEEE International Conference on Image Processing*, San Diego, California, USA, 2008: 109-112
  - [26] Chen F., Zhang D. A Benchmark for Geometric Facial Beauty Study [J]. *Lecture Notes in Computer Science*, 2010, 6165(1): 21-32
  - [27] T. Ahonen, A. Hadid, M. Pietikainen. Face recognition with local binary patterns [C]. *Proceedings of European Conference on Computer Vision*, Prague, Czech Republic, 2004: 469-481
  - [28] X. Feng, M. Pietikainen, A. Hadid. Facial expression recognition with local binary patterns and linear programming [J]. *Pattern Recognition and Image Analysis*, 2005, 15(2): 546-548
  - [29] T. Ahonen, A. Hadid, M. Pietikainen. Face description with local binary patterns: application to face recognition [J]. *IEEE Transactions on Pattern Analysis and*



- Machine Intelligence, 2006, 28(12): 2037-2041
- [30] T. Ojala, M. Pietikainen, D. Haxwood. A comparative study of texture measures with classification based on feature distributions [J]. Pattern Recognition, 1996, 29(1): 51-59
- [31] T. Ojala, M. Pietikainen, T. Maenpaa. Multiresolution gray scale and rotation invariant texture classification with local binary patterns [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971-987.
- [32] T. Maenpaa, M. Pietikainen. Texture analysis with local binary patterns. Handbook of Pattern Recognition and Computer Vision [M]. World Scientific, Singapore, 2001: 711-736
- [33] Sanchez-Y. R, E. Kurmyshev, F. Cuevas. A framework for texture classification using the coordinated clusters representation [J]. Pattern Recognition Letters, 2003, 24: 21-31.
- [34] L. Ma, L. Zhu. Integration of the optimal Gabor filter design and local binary patterns for texture segmentation [C]. Proc. of the 2007 IEEE International Conference on IntegrationTechnology, Shenzhen, 2007: 408-413
- [35] L Jin, Q.S Liu, H.Q. Lu. Face detection using improved LBP under bayesian framework[C]. Proc. of Third International Conference on mage and Graphics. Hong Kong, 2004: 306-309
- [36] 张文超. 局部 Gabor 二值模式人脸表示与识别方法研究[D]. 哈尔滨工业大学博士学位论文, 2007
- [37] Lowe DG. Object recognition from local scale-invariant features[C]. Proceedings of the International Conference on Computer Vision, 1999, 1150-1157
- [38] Lowe DG. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110
- [39] David D. Lewis. Naive (Bayes) at Forty: The Independence Assumption in Information Retrieval [C]. Proc. of 10th European Conference on Machine Learning, Chemnitz, 1998, 4-15
- [40] J. Yang, Y. Jiang, A. Hauptmann, *et al.*. Evaluating bag-of-visual-words representations in scene classification [C]. Proceedings of the international workshop on Workshop on multimedia information retrieval, 2007
- [41] Hartigan, J.. Clustering algorithms[M]. New York, USA, 1975.
- [42] MacKay, D.. An Example Inference Task: Clustering [M]. Information Theory, Inference and Learning Algorithms, 284-292
- [43] Grauman, K.and T. Darrell. The pyramid match kernel: Discriminative classifi-

cation with sets of image features, 2005

- [44] Lazebnik S., Schmid C., Ponce J.. Beyond bags of features:spatial pyramid matching for recognizing natural scene categories [C]. Proceedings of the Conference on Computer Vision and Pattern Recognition, Los Alamitos, California, 2006: 2169-2178
- [45] 孙岩, 吕世聘, 王秀坤等. 基于结构学习的 KNN 分类算法[J]. 计算机科学, 2007, 34(12): 184-187
- [46] Vapnik V. The nature of statistical learning theory [M]. New York: Springer Verlag, 1995
- [47] 武伟. 基于 SVM 的图形分类研究[D]. 西北师范大学硕士学位论文, 2009
- [48] 赵莹. 半监督支持向量机学习算法[D]. 哈尔滨工程大学博士学位论文, 2010
- [49] 李忠伟. 支持向量机学习算法研究[D]. 哈尔滨工程大学博士学位论文, 2006
- [50] C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1-27:27, 2011