

# 基于静态人脸图像的性别识别 方法研究

作者姓名\_\_\_\_\_张 巍\_\_\_\_\_

学校导师姓名、职称\_\_\_\_\_郑海红 副教授\_\_\_\_\_

企业导师姓名、职称\_\_\_\_\_刘瑞华 高工\_\_\_\_\_

申请学位类别\_\_\_\_\_工程硕士\_\_\_\_\_



学校代码 10701  
分 类 号 TP391

学 号 1303121914  
密 级 公开

# 西安电子科技大学

## 硕士学位论文

### 基于静态人脸图像的性别识别 方法研究

作者姓名： 张 巍

领 域： 计算机技术

学位类别： 工程硕士

学校导师姓名、职称： 郑海红副教授

企业导师姓名、职称： 刘瑞华 高工

学 院： 计算机学院

提交日期： 2016 年 3 月



# **Research of gender recognition based on static facial images**

A thesis submitted to  
XIDIAN UNIVERSITY  
in partial fulfillment of the requirements  
for the degree of Master  
in Computer Technology

By

Zhang Wei

Supervisor: Zheng HaiHong Associate Professor

Ruihua Liu Senior Engineer

March 2016



## 西安电子科技大学 学位论文独创性（或创新性）声明

秉承学校严谨的学风和优良的科学道德，本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果；也不包含为获得西安电子科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同事对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文若有不实之处，本人承担一切法律责任。

本人签名：\_\_\_\_\_ 日 期：\_\_\_\_\_

## 西安电子科技大学 关于论文使用授权的说明

本人完全了解西安电子科技大学有关保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权属于西安电子科技大学。学校有权保留送交论文的复印件，允许查阅、借阅论文；学校可以公布论文的全部或部分内容，允许采用影印、缩印或其它复制手段保存论文。同时本人保证，结合学位论文研究成果完成的论文、发明专利等成果，署名为西安电子科技大学。

保密的学位论文在\_\_\_\_年解密后适用本授权书。

本人签名：\_\_\_\_\_ 导师签名：\_\_\_\_\_

日 期：\_\_\_\_\_ 日 期：\_\_\_\_\_





## 摘要

随着全球化和网络的飞速发展,身份识别在日常生活中需求越来越大。在人类生物特征中,人脸易获取的特性使其在身份识别中应用广泛,而性别识别属于人脸识别中的一种,随着身份识别的发展也成为了研究的热点之一。

词袋(BoW)模型是一种广泛应用于文档分类领域的分类方法,最近几年一些研究人员将 BoW 的思想引入到图像分类领域,出现了视觉词袋、视觉词汇、视觉字典等,由于 BoW 模型简单并且行之有效的特点,本文对 BoW 模型进行了比较深入的研究。

性别识别是一个典型的二分类问题,本文首先介绍了影响性别识别的各种因素,然后给出了常用的人脸数据库以及对人脸图像的预处理方法。

本文通过对 BoW 模型、Adaboost 算法及朴素贝叶斯算法进行深入学习的基础上,提出了一种基于改良 BoW 模型并利用空间相似度(Spatially Constrained Similarity Measure ,SCSM)进行测量的 NBC(Naïve Bayes Classifier)分类器的方法进行性别识别。传统的稀疏 SIFT 特征在提取过程需要构建高斯尺度空间和高斯差分尺度空间,尺度空间的构造需要比较多的时间花费和空间花费,而 Dense SIFT 在特征提取过程中则不需要构建.Dense SIFT 对图像进行均匀采样,在实际的特征提取过程对不同的背景适应性也较强,提高了特征提取效率。

最后选择 CAS-PEAL 人脸数据库,在影响人脸识别的不同的变化条件下,如姿势、表情、光照和饰物,用基于稠密 SIFT 和稀疏 SIFT 的 BoW 特征提取模型,分别与朴素贝叶斯分类器和 Adaboost 分类器结合进行实验。实验结果表明基于 Dense SIFT 的 BoW 模型和 NBC 结合的性别分类效果比其他几种识别方法准确率平均提高 2%。

**关键词:** 性别识别, 朴素贝叶斯分类器, BoW, 视觉词汇, SIFT 特征



## ABSTRACT

With the rapid development of globalization and network, the identity recognition technique in everyday life is urgently demand. In the human biological characteristics, face features being easy get to make it widely used in identity recognition. Gender is one of the important face characteristics, and its recognition has become one of the hotspots.

BoW(bag of words) model is widely used in the field of document classification methods. In recent years, some researchers have tried to apply it to image classification, such as the-bag-of-words visual vocabulary and visual dictionary. Since BoW model is simple and effective features, we deeply study the BoW model in this paper.

Gender recognition is a typical binary classification problem. First, this paper introduces various factors that affects the gender recognition. Then the common face database and the pre-processing methods for facial images are given.

Based on the deeply study of BoW model, Adaboost algorithm and Naive Bayesian algorithm, this paper proposes a modified method of BoW model based on Dense SIFT and NBC using spatially-constrained similarity measure. Traditional sparse SIFT features need to build the Gaussian scale-space and difference of Gaussian scale space during the process of extraction, so constructing the scale-space require more time and space, while the extraction process of Dense SIFT feature is not required to build the space. Dense SIFT gets samples images uniformly. In the actual process of feature extraction, it also has strong adaptability to different backgrounds and improves the efficiency of feature extraction.

Finally by using the images on the CAS-PEAL face database, under different face changing conditions, such as posture, facial expressions, lighting and accessories, etc. the experiment is carried out with Naïve Bayes classifier and Adaboost classifier combined with BoW feature extraction model based on Dense SIFT and Sparse SIFT. The experiment results show that the model based on Dense SIFT's BoW combined with NBC improve the accuracy of 2% than several other recognition method.

**Keywords:** Gender recognition, Naïve Bayes Classifier, Bag of words, Visual words, Scale-invariant feature transform

## 插图索引

图 2.1 图像分类的简易流程图 .....	5
图 2.2 中值滤波示意图 .....	7
图 2.3 均值滤波前后对比 .....	8
图 2.4 直方图均衡化效果 .....	9
图 3.1 积分图 .....	12
图 3.2 $(x, y)$ 处的积分图 .....	14
图 3.3 区域 $D$ 像素和的计算 .....	15
图 3.4 bag-of-words 模型 .....	16
图 3.5 Gauss 金字塔 .....	18
图 3.6 高斯尺度空间和高斯差分尺度空间 .....	19
图 3.7 寻找空间极值点示意图 .....	19
图 3.8 图像金字塔 .....	20
图 3.9 SIFT 特征方向图 .....	23
图 3.10 K-Means 聚类 .....	24
图 4.1 SCSM 的流程图 .....	30
图 5.1 经典 BoW 模型示意图 .....	35
图 5.2 CAS-PEAL 人脸数据库中不同姿态示例图 .....	37
图 5.3 性别判定结果 .....	40



## 表格索引

表 3.1 初始的特征模板 .....	11
表 3.2 多尺度窗口下特征矩形的数目 .....	14
表 3.3 视觉字典 .....	16
表 5.1 实验数据样本的分布 .....	38
表 5.2 SIFT 和 Dense SIFT 在 NBC 和 Adaboost 下的性别识别率 .....	38
表 5.3 SIFT 和 Dense SIFT 在 NBC 和 Adaboost 下的运算时间 .....	38
表 5.4 识别率及测试时间 .....	39





## 符号对照表

符号	符号名称
$idf^2(k)$	视觉词汇在数据库中出现次数的倒数
$tf_Q(k)$	第 $k$ 个视觉词汇在 $Q$ 中出现的次数
$tf_D(k)$	第 $k$ 个视觉词汇在 $D$ 中出现的次数
$w(f_i)$	代表特征分量 $f_i$ 属于第 $k$ 个视觉单词
$S(x, c_i)$	空间尺度相似值
$h(x_i)$	弱分类器
$H(x_i)$	强分类器
$P(A B)$	事件 $A$ 在事件 $B$ 发生的条件下发生的概率
$P(A \cdot B)$	事件 $A$ 和事件 $B$ 同时发生的概率



## 缩略语对照表

缩略语	英文全称	中文对照
LoG	Laplacian of Gaussian	高斯拉普拉斯算子
MAP	Maximum A Posteriori	极大后验假设
SCSM	Spatially Constrained Similarity Measure	定义空间约束相似度测量
PAC	Probably Approximately Correct	概率近似正确
ML	Machine Learning	机器学习
NBC	Naïve Bayes Classifier	朴素贝叶斯分类器
RBF	Radial Basis Function	径向基函数
SVM	Support Vector Machine	支持向量机
LDA	Linear Discriminant Analysis	线性判别分析法



# 目录

摘要 .....	I
ABSTRACT .....	III
插图索引 .....	V
表格索引 .....	VII
符号对照表 .....	IX
缩略语对照表 .....	XI
<b>第一章 绪论</b> .....	1
1.1 研究背景和意义 .....	1
1.2 国内外基于人脸的性别识别研究现状 .....	1
1.2.1 基于人工神经网络进行的研究 .....	1
1.2.2 支持向量机(Support Vector Machine,SVM) .....	2
1.2.3 AdaBoost 分类算法 .....	2
1.2.4 LDA(Linear Discriminant Analysis)线性判别分析法 .....	3
1.3 论文组织结构 .....	3
<b>第二章 图像预处理和人脸数据库</b> .....	5
2.1 影响性别识别的因素 .....	5
2.2 常用人脸数据库 .....	6
2.3 图像预处理 .....	7
2.3.1 中值滤波 .....	7
2.3.2 均值滤波 .....	8
2.3.3 直方图均衡化 .....	8
2.3.4 尺度归一化 .....	9
2.4 本章小结 .....	10
<b>第三章 性别特征的提取方法</b> .....	11
3.1 基于整体特征的 AdaBoost 算法 .....	11
3.1.1 类 Haar(Haar-Like)特征 .....	11
3.1.2 类 Haar 矩形特征的表示和计算 .....	12
3.1.3 利用积分图法快速计算类 Haar 特征 .....	14
3.2 基于局部特征的 SIFT 特征提取算法 .....	15
3.2.1 Bag-of-words 模型 .....	15

3.2.2	Bag-of-words 模型在人脸图像上的应用 .....	16
3.2.3	SIFT 算法提取不变特征点 .....	17
3.2.4	K-Means 构造单词表 .....	23
3.3	本章小结 .....	24
第四章	性别识别的分类方法 .....	25
4.1	AdaBoost 分类器 .....	25
4.1.1	Boosting 方法 .....	25
4.1.2	AdaBoost 算法原理 .....	26
4.1.3	AdaBoost 算法训练过程 .....	27
4.1.4	AdaBoost 的优缺点 .....	27
4.2	朴素贝叶斯分类器(Naïve Bayes Classifier,NBC ) .....	28
4.2.1	概率论基础 .....	28
4.2.2	朴素贝叶斯分类原理及模型 .....	29
4.2.3	朴素贝叶斯分类器的设计 .....	30
第五章	基于改良 BoW 模型的性别识别算法 .....	35
5.1	经典的 BoW 模型 .....	35
5.2	基于改良 BoW 模型的性别识别算法 .....	36
5.2.1	改良的 BoW 特征提取模型 .....	36
5.2.2	改进的 NBC 分类器 .....	36
5.3	实验验证及结果分析 .....	37
第六章	总结和展望 .....	41
参考文献	.....	43
致谢	.....	47
作者简介	.....	49

## 第一章 绪论

### 1.1 研究背景和意义

我们身处的这个时代信息化高速发展，全球化的极速发展带来了大量的信息交流，在这其中非常重要且常见的一项技术就是身份识别，大量的人流移动需要身份检测和识别。随着信息化的飞速发展，传统的身份识别系统已经越来越不能满足现在的需求，基于计算机的生物特征识别技术随之就来到了我们身边。

概括来说生物特征识别技术包括：指纹、虹膜、人脸、声音、掌纹、手势等，在这些特征里，人脸是相对易于收集的特征。在自然正常状态下每一张人脸都是独一无二的，即便是同卵双胞胎也存在差异，也就是说人脸在身份识别中是具有相对唯一性的，每一张人脸都包含着丰富的信息，利用这些信息我们可以进行年龄、性别、身份、种族、表情等研究，而这些识别大多都成为了热门研究项目，其中人脸易于在识别对象未知的情况下对识别对象进行识别，具有友好性等，具有多种优势，基于脸部图像的性别识别的持续发展，也促进了身份识别的发展，在理论和实际等各方面的运用当中，基于人脸的性别识别都具有很重要的研究意义，不断有新的研究人员投入到这项工作中来。

随着上世纪计算机的诞生以及自诞生以来的飞速发展，利用计算机进行基于人脸的性别识别也越来越热门，并成为一项新的领域。利用计算机进行身份识别逐渐的出现在日常生活中，如门禁系统、未成年检测系统、特定人群检测系统、监控系统、支付系统、人口信息采集系统等，利用计算机进行的基于人脸的性别识别系统具有获取方便快捷、识别过程并发、识别结果直观、应用成本低性价比高等多种优势。

### 1.2 国内外基于人脸的性别识别研究现状

国外来说基于人脸的性别识别开始于 20 世纪八九十年代，刚开始发展并不快，进入 21 世纪后才发展比较快。从刚开始简单的性别分类逐步发展到一个研究项目，利用分类器来进行大量快速的分类成为了一种需求，进而促生了更多基于人脸的性别识别的各种方法的研究。

#### 1.2.1 基于人工神经网络进行的研究

人工神经网络自 1943 年提出后被广泛应用于人工智能中，简单讲是模仿人类大

脑中的神经网络进行工作，由许多神经元构成，每一个神经元具有单独逻辑计算和学习能力，并互相连接交织构成网状模型，根据神经元的功能不同，目前已经出现了 40 多种基于神经网络的模型。

最早利用人工神经网络进行性别分类的是国外的研究人员。如 Golomb 等，Golomb 的主要研究内容是在 1991 年的时候训练了两层神经元网络构成分类器，对  $30 \times 30$  的人脸图像进行分类，达到了在 90 幅人脸图像上 91.9% 的正确率。与 Golomb 同时进行的还有 Cottrel 等人，他们与 Golomb 不同的是在 160 张的人脸样本上先对样本进行了处理，即主成分分析(PCA)，从人脸样本中提取的特征作为输入，再训练 BP(Back Propagation)神经网络进行人脸分类。1994 年，Yen 等人用径向基函数(Radial Basis Function, 简称 RBF)和 BP 神经网络的方法，得出了在 1500 张  $50 \times 70$  的人脸数据集上正确率均可达到 90% 以上的结论，这两种神经网络的性能差不多。1995 年，Brunelli 等人设计了 HyperBF 神经网络，提取的人脸图像的几何特征用 HyperBF 神经网络分类得到 79% 的精准度。1996 年，Tamura 等人采用 3 种分辨率的样本分别进行分类，得出实验结果在 3 种分辨率最低的  $8 \times 8$  的人脸图像上获得了 93% 的正确率。1998 年，Gutta 等人把径向基神经网络和决策树结合起来，该分类系统在 FERET 人脸数据库的 3000 张人脸图像上进行性别分类，达到 96% 的精准率。2006 年，Fok Hing 等设计了基于分流抑制卷积神经网络(Shunting Inhibitory Convolutional Neural Networks)的人脸性别识别系统，在人脸检测和性别判别中都用该系统，但提取的特征不同，在 FERET 人脸数据库中达到了 97.2% 的正确率。

### 1.2.2 支持向量机(Support Vector Machine,SVM)

2002 年 Moghaddam 等人在 FERET 人脸库上的  $21 \times 12$  的人脸图像上引入基于 RBF 的 SVM 分类器，并用多种方法对同样的样本进行了实验，结果表明 SVM 达到了 96.6% 的正确率，优于线性分类器等其他的分类器。

### 1.2.3 AdaBoost 分类算法

AdaBoost 在 2001 年由 Viola 和 Jones 提出，他们并不是第一个提出的，但是在自适应的迭代算法中，这一次提出的算法非常经典，AdaBoost 算法通过把弱分类器级联的方法组成强分类器，提升了性能之后进行分类。2002 年 Shakhnarovich 等人把 AdaBoost 应用在性别识别上，完成了一个可以自动检测跟踪的性别识别系统，实验采用网络上的图像，在 4500 张图像上达到了 78% 的正确率，该系统的速度比基于 RBF



的 SVM 分类器快 1000 多倍，效果很好。2003 年，清华大学的武勃等采用基于查找表(Look Up Table, LUT)的弱分类器，实验发现该方法比用基于阈值的 AdaBoost 算法聚合快，收敛快，进而容易快速形成高效强分类器，在实验中采用 FERET 和网络中的图像，正确率达到 88%。AdaBoost 算法和目前的成熟算法比起来，速度相对快，具有较强的竞争力。

### 1.2.4 LDA(Linear Discriminant Analysis)线性判别分析法

线性判别分析法属于统计类方法，在模式识别领域里是广泛应用的，线性判别分析法是将高维的样本投影到最佳鉴别矢量空间，以达到降维和提取信息的作用，投影后的样本具有最大和最小的类内间距，那么样本在这个空间具有最佳的分离性。

2000 年，Lyons 等人用 Gabor 和 PCA 方法提取人脸特征，再用线性判别分析法进行性别分类，实验得到了 92%的正确率。2004 年，Jain 等用独立成分分析法提取人脸特征，用线性判别分析法在 FERET 人脸库的 500 张图像上得到 99.3%的正确率。2005 年，Buchala 等用 PCA 提取人脸特征，再用线性判别分析法在 FERET 人脸库得到 86%的正确率。

## 1.3 论文组织结构

本文主要对基于 BoW<sup>[1]</sup>(bag-of-words)模型的人脸性别分类方法进行研究，把广泛应用于文本检索中的 BoW 模型应用到基于静态人脸图像的性别识别问题中。并在经典的 BoW 模型基础上进行改进，使其在性别分类上有更好的效果。

第一章是本文的绪论，绪论中介绍了人脸性别的研究背景，研究意义，对国内外的人脸识别的研究状况做了概括叙述，并对本论文的组织结构进行概括的介绍。

第二章 对人脸性别识别系统进行简单的研究，介绍了影响性别分类的因素，主流的人脸数据库和图像预处理方法，并对预处理中的常见方法，如中值滤波，均值滤波，直方图均衡化和尺度归一化进行了详细的解释。

第三章 在第三章中重点介绍了类 Haar 基的知识和基于 BoW 模型的性别特征提取方法。首先对经典特征因子类 Haar 基的定义和计算方法进行叙述，后面介绍了 BoW 模型及 BoW 模型的原理，以及经典 BoW 模型中用到的 SIFT 特征子和 K-Means 聚类法。

第四章 重点介绍了性别识别的分类方法，有经典的 Adaboost 分类器，以及与 BoW 模型结合进行分类的朴素贝叶斯分类器(Naïve Bayes Classifier, NBC)。

第五章 进行实验设计，使用改进的基于 Dense SIFT 的 BoW<sup>[2]</sup>模型对人脸数据库 CAS-PEAL 上的 3000 张图像分为 6 组进行实验，实验采用 BoW+NBC, BoW+Adaboost,

基于 Dense SIFT 的 BoW,基于 Dense SIFT 的 BoW+Adaboost 四种分类器进行试验,通过仿真实验可以看出 Dense BoW<sup>[6]</sup>在性别识别中的识别时间和识别正确率都比其他方法有提升,平均达到 2%的准确度提升。

第六章 对本文进行总结,并对性别识别的进一步发展进行了讨论。

## 第二章 图像预处理和人脸数据库

基于人脸图像的性别识别过程是一个二分类问题。影响人脸检测的有光照、背景、噪声等，这对后面的工作都会造成不同程度的影响，所以要对人脸图像进行提前的处理，尽量减少各种因素带来的影响。在人脸图像预处理完之后，我们先判别图像中有没有人脸，没有人脸则检测下一张，如果存在人脸，则需要确定有几张人脸以及每一张人脸的位置。人脸存在大小不一，以及受到亮度的影响，图像就需要进行灰度归一化处理。这是后面工作的基础，把人脸数目和人脸位置找准才能顺利进行后面的实验。

如图 2.1 所示，是一个简单的图像分类过程，特征提取和分类器的设计是我们主要的讨论内容，在本章中我们先讨论特征提取之前的工作。

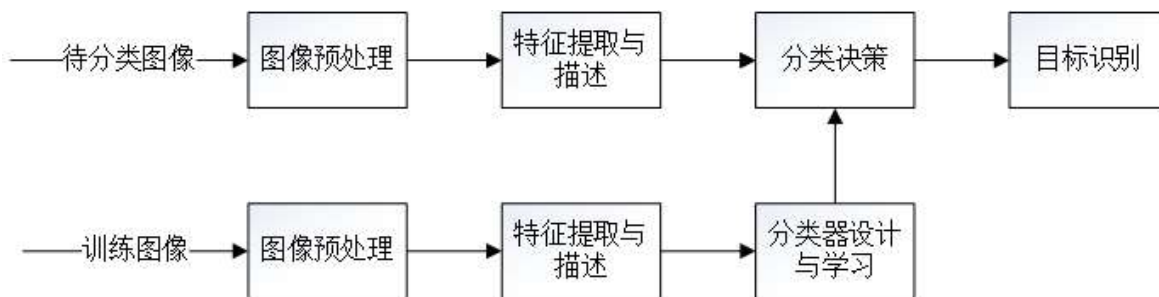


图 2.1 图像分类的简易流程图

### 2.1 影响性别识别的因素

性别识别是人类身份识别中的一种，在人类身份识别过程中有着众多因素影响身份识别。这些因素不仅仅影响性别的判断，其他种类的身份识别也受到不同程度的影响。

(1) 光照(illumination): 光照是每张图像都会存在的问题，图像光照是指物体被照明的朝向及光照的强度，这两种因素的变动对基于人脸灰度图像的分布会造成很大程度的影响，进而影响大多数基于灰度特征统计的辨认方法的效果，而对另一种基于几何特征的辨认方法来说，光照对基于几何特征的精确提取也造成不小的影响。

(2) 姿势(pose): 人的姿势的变动也很多，在数字图像处理中可以分为两类，第一类是二维平面上的姿态变动(平移、旋转)，第二类是三维空间上的姿态变动，对人

脸图像来说，姿势的变化使人脸图像上包含信息不同，可能使人脸的信息大量丢失，其中第二类对人脸辨别的效果影响更明显。

(3) 表情(expression): 人脸最重要的一个特征就是表情，表情代表着一个人此时的心情。在人脸识别中，相对于其余的目标辨别(object recognition)问题来说，表情特征的困难是指人脸的非刚性，包含了人脸多样的表情变换。

(4) 遮挡(occlusion / disguise): 识别过程中，还有一种特殊的难题是人脸在拍摄时会受到许多遮挡的影响，例如头发、胡须、刺青、眼镜、帽子、饰品等，减少了一部分人脸信息，增加了一些非人脸信息，不利于进行分类。

(5) 低分辨率(low-resolution): 指的是拍摄设备初始获取的图像像素比较少，分辨率比较低，包含的图像信息比较少，增加了识别的困难程度。

## 2.2 常用人脸数据库

### 1. FERET 人脸数据库

由 FERET 项目创建，数据库里包括 14051 张多姿态、光照的灰度人脸图像，在人脸图像辨识领域中是运用最普遍的一个人脸数据库，其中同一人的脸表情变化单一，基本是西方人。

### 2. PIE 人脸数据库

由卡耐基梅隆大学创建，包含 68 位志愿者的 41368 张多姿态、表情和光照的面部表情，姿态和光照变化在受控制情况下拍摄完成，应用也越来越广。

### 3. Yale 人脸数据库

由耶鲁大学计算机视觉与控制中心建立，包含 15 名志愿者在不同的光照、姿态、大小面部下的 165 张图片。

### 4. YaleB 人脸数据库

由 15 个人的 5850 张多姿态、多光照的图像组成，主要用于姿态和光照问题的建模和研究等。

### 5. ORL 人脸数据库

由 40 人的 400 张灰度图像组成，分辨率为  $92 \times 112$  像素，其中一些人是在不同时刻拍的，一些脸部表情和光照变化比较大，人脸姿态也有很大的变化。

### 6. CAS-PEAL 人脸数据库

由 1040 名中国人的 99450 幅头肩图像组成，涵盖了姿态、表情、光照和饰物这四种变化条件，一部分人的图像还有距离、背景和时间的跨度的变化，所有图像均在特定条件下拍摄。

## 7. AR 人脸数据库

由普渡大学机器人视觉研究中心的 Aleix 和 Benavente 两位教授建立, 包含了 70 个男性和 50 个女性共 120 多人的 4000 多幅正面人脸图像, 有光照、表情和遮挡物的变化。每个人间隔 2 周拍摄, 每次 13 张一共 26 张的图像。

## 2.3 图像预处理

在 2.1 中可以看到图像上有很多噪声, 进行图像预处理是进行身份识别过程的必要的一步。输入图像由于图像采集环境的不同, 比如光照明暗程度、设备性能的优劣、距离远近和焦距大小等, 往往存在有噪声, 对比度不够等缺点, 后面两种情况还使人脸在整幅图像中间的大小和位置不确定, 这还需要对人脸进行缩放平移和变换等。为了保证人脸图像中人脸大小, 位置以及人脸图像质量的一致性, 必须对图像进行预处理。下面简单介绍一些常见的图像预处理方法有人脸图像去噪, 增强, 摆正, 二值化, 灰度化和几何归一等。

### 2.3.1 中值滤波

中值滤波的理论在提出后被迅速应用在数字信号处理领域。中值滤波对脉冲噪声有较好的滤除作用, 中值滤波器是一种非线性的滤波器, 因为中值滤波属于非线性的信号处理方法。中值滤波的思想是把滤波模板中心像素点的灰度值设置为该点模板窗口内的所有像素点灰度值的中值, 使四周的像素值更靠近实际值, 进而打消掉孤立的噪声点。在滤除噪声的同时, 能够保护信号的边缘, 使之不被模糊。具体实现是用某一规格的二维滑动模板, 将滑动模板中的所有像素依照像素值的大小进行排列, 产生单调递增(或递减)的二维数据序列。



图 2.2 中值滤波示意图

简单步骤:

- (1) 将滤波模板在图像中漫游, 将模板中心与某个像素重合;
- (2) 读取模版中各个像素点的灰度值;

(3) 将 2 中的灰度值排序，从小到大；

(4) 取 3 中排序中的中间值，将模板的中心位置对应的像素值变为中间值，中间值在滤波模板中像素数是偶数的情况下，将排序后中间两个灰度值求和取算术平均值。如果是奇数的情况下，则正好取中间像素的灰度值。

对于常见的二维图像，滤波器的效果好坏与滤波模板的大小影响很大，要根据不同的图像和实验需求选择合适大小的滤波模板。

### 2.3.2 均值滤波

均值滤波与中值滤波不同，是线性滤波器，与中值滤波过程类似，但是在取值上是用邻域内所有像素的平均值作为滤波后的像素值，这样做的优点是可以有效地滤除图像中附加的噪声。但是很明显，由于取均值作为新的像素值，会把原来图像上的细节破坏掉，图像就会变得模糊。均值滤波主要包括算术均值滤波，几何均值滤波，谐波均值滤波和逆谐波均值滤波，其中几何均值滤波器所达到的平滑度可以与算术均值滤波器相比，在滤波过程中丢失更少的图像细节。更适合于解决脉冲噪声的问题的滤波器是逆谐波均值滤波器，它的缺点是必须要知道噪声是暗噪声还是亮噪声，以便于选择合适的滤波器阶数符号。



图 2.3 均值滤波前后对比

### 2.3.3 直方图均衡化

直方图均衡化的思想是把图像对应的灰度图像中像素数多的那部分灰度级加宽，压缩灰度图像中像素个数少的灰度级，这样就增大了像素取值的动态范围，提高了对比度和灰度色调的变化，使图像更加清晰。人脸图像由于受到光照不均匀、高光、偏

光、侧光等导致的阴影、过明和过暗。可以看出直方图均衡化仅需要输入图像直方图就可以自动达到这种效果，是一个单输入的变换函数。

对于一张灰度图像，假设有  $N$  种灰度，第  $i$  种灰度出现的次数是  $n_i$ ， $n$  是所有像素数，则第  $i$  种灰度出现的概率是

$$p_x(i) = \frac{n_i}{n}, (i = 1, 2, 3 \dots N)$$

经过上式计算之后，所有像素概率之和为 1，也即归一化到 0, ..., 1。

下图是经过直方图均衡化的图像与原图的对比



图 2.4 直方图均衡化效果

直方图均衡化方法的优点是直方图均衡化应用在背景和前景都太亮或者太暗的图像上效果很好，可以将 X 光图像中的骨骼结构更好地显示出来，对曝光过度或者曝光不足照片而造成细节丢失的图像处理效果也不错。直方图均衡化的过程是可逆过程，方法简单明了，在有均衡化函数的情况下，那么就可以通过直方图均衡化恢复原始的直方图，也不需要太多时间和计算。

### 2.3.4 尺度归一化

人脸的尺度归一化是几何归一化，是不同于 2.3.3 中的直方图均衡化的另外一种归一化，包含了图像的平移、旋转、缩放和切割等。由于人脸图像在获取的过程中受到各种外界条件的影响，使人脸图像对比度、距离、光照、方向、姿势、背景噪声和尺寸等方面差异较大，目的是把不同条件下所得的图片统一变成我们所需要的标准格式。尺度归一化是进行人脸研究里必不可少的一个环节，常用的方法是根据人脸图像上人的瞳孔位置进行旋转，摆正之后用两个瞳孔间的距离来进行缩放以达到统一模板的目的。

## 2.4 本章小结

本章介绍了进行人脸性别研究的一些常见的理论和背景知识,影响性别识别的因素以及解决办法和目前比较常用的一些人脸数据库。人脸预处理是作为进行人脸识别的第一个门槛,是基础但是却不容忽视,掌握了这些,就知道性别识别研究需要努力的方向,并为下面性别特征的提取做好了准备。



## 第三章 性别特征的提取方法

在图像经过预处理后，我们需要对人脸进行特征提取，数学意义上的特征提取过程就是把测量空间到特征空间的一次映射变换，那么在映射变换中我们需要进行的是降维和保留主要分类特征。经过预处理的人脸图像特征维数比较高，不适宜进行运算，我们需要把维数降低到可计算的程度。而在降维的同时我们要保留重要的，也就是具有好的分类特性的特征。人脸性别识别是一个典型的二分类问题，而性别特征的提取可以分为两大类，基于整体特征和基于局部特征的，本章主要介绍一下经典的 AdaBoost 算法和基于 Bow 模型的局部特征提取法。


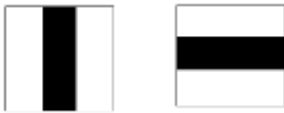

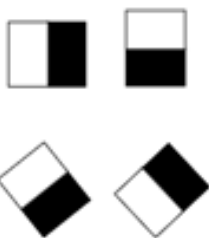
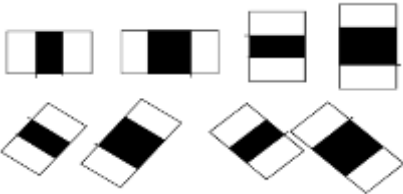
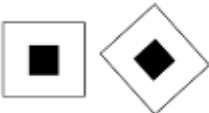
### 3.1 基于整体特征的 AdaBoost 算法

AdaBoost 算法利用基于 Haar 特征的弱分类器组成强分类器，再将强分类器进行级联，达到快速检测的效果，本节主要对类 Haar 特征进行介绍。

#### 3.1.1 类 Haar(Haar-Like)特征

Papageorgiou 首先在人脸脸上运用 Haar-like 特征，Viola 等设计了最初的特征模板。一共有 5 种基本的特征模板，可以分为 3 类，如表所示。每块模板由黑色矩形区域和白色矩形区域组成。

表 3.1 初始的特征模板

模板号	边缘特征	线性特征	特定方向特征
Viola 等人提出的特征模板	 (2,1)      (1,2)	 (1,3)      (3,1)	 (2,2)
扩展的特征模板			

Haar 特征值是求出白色、黑色矩形区域的像素和，两个值相减得出的差值，描述了黑白矩形区域之间的灰度差异，其中白色矩形区域和黑色矩形区域是全等的。

后来 Lienhart 扩展了特征模板，如表 3.1 所示。

### 3.1.2 类 Haar 矩形特征的表示和计算

(1) 要想计算类 Haar 矩形特征，需要先知道类 Haar 矩形怎么表示。

每一个矩形特征可以这样表示：

$$j = (x_j, y_j, R_j, S_j) \quad (3-1)$$

式中  $(x_j, y_j)$  为矩形顶点在检测窗口中的坐标， $R_j$  是指类 Haar 矩形滤波器的类型， $S_j$  是指类 Haar 举行滤波器的尺度。

(2) 矩形特征的计算

由图易见，对于  $n \times n$  的待检测矩形图像，如果矩形的左上角点  $a(x_1, y_1)$  和右下角点  $b(x_2, y_2)$  可以确定，那么这个矩形特征就可以确定。

该  $n \times n$  矩形需要满足  $(s, t)$  条件：

- 1) 矩形 x 方向的边长必须可以平均分为 s 段
- 2) 矩形 y 方向的边长必须可以平均分为 t 段

其中 s, t 均为正整数，我们称这样的矩形为条件矩形。

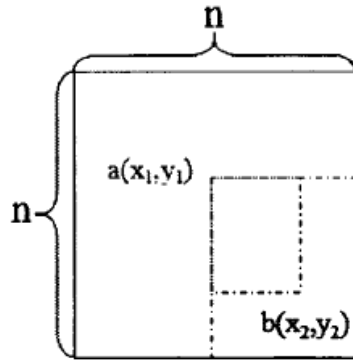


图 3.1 积分图

矩形的最小尺寸为  $s \times t$ 、 $t \times s$ ，最大尺寸为  $[n/s] \cdot s \times [n/t]$  或者  $[n/t] \cdot t \times [n/s]$ ，其中的  $[]$  为取整数运算符。

则对于任一满足  $(p, q)$  条件的矩形的求解步骤如下：

- (1) 确定矩形左上角顶点  $a(x_1, y_1)$  的位置，则  $x_1 \in \{1, 2, 3, \dots, n-s, n-s+1\}$ ， $y_1 \in \{1, 2, 3, \dots, n-t, n-t+1\}$ ；

(2) 确定了  $a(x_1, y_1)$  的位置后, 那么  $b(x_2, y_2)$  的位置只能在虚线区域内, 也就是  $a$  的右下方。则

$$x_2 \in \{x_1+s-1, x_1+2 \cdot s-1, \dots, x_1+(p-1) \cdot s-1, x_1+p \cdot s-1\} \quad (3-2)$$

$$y_2 \in \{y_1+t-1, y_1+2 \cdot t-1, y_1+3 \cdot t-1, \dots, y_1+(q-1) \cdot t-1, y_1+q \cdot t-1\} \quad (3-3)$$

其中,  $p = \left\lfloor \frac{n-x_1+1}{s} \right\rfloor$ ,  $q = \left\lfloor \frac{n-y_1+1}{t} \right\rfloor$ 。

由上述分析可知, 在  $n \times n$  的子窗口中, 满足  $(s, t)$  条件的矩形的总数目为

$$\Omega_{(s,t)}^n = \left( \left\lfloor \frac{n}{s} \right\rfloor + \left\lfloor \frac{n-1}{s} \right\rfloor + \left\lfloor \frac{n-2}{s} \right\rfloor + \dots + \left\lfloor \frac{s+1}{s} \right\rfloor + 1 \right) \cdot \left( \left\lfloor \frac{n}{t} \right\rfloor + \left\lfloor \frac{n-1}{t} \right\rfloor + \left\lfloor \frac{n-2}{t} \right\rfloor + \dots + \left\lfloor \frac{t+1}{t} \right\rfloor + 1 \right) \quad (3-4)$$

对于有 5 种特征模板的在  $n \times n$  的子窗口上特征总量的计算公式如下:

$$\Omega^n = \Omega_{(1,2)}^n + \Omega_{(2,1)}^n + \Omega_{(1,3)}^n + \Omega_{(3,1)}^n + \Omega_{(2,2)}^n \quad (3-5)$$

由旋转对称性可知  $(1,2)$  和  $(2,1)$  以及  $(1,3)$  和  $(3,1)$  具有旋转对称性, 上式可化为:

$$\Omega^n = 2\Omega_{(1,2)}^n + 2\Omega_{(1,3)}^n + \Omega_{(2,2)}^n \quad (3-6)$$

当  $n=16$  时,

$$\begin{aligned} \Omega_{(1,2)}^n &= \Omega_{(2,1)}^n \\ &= \left( \left\lfloor \frac{n}{s} \right\rfloor + \left\lfloor \frac{n-1}{s} \right\rfloor + \dots + \left\lfloor \frac{s+1}{s} \right\rfloor + 1 \right) \cdot \left( \left\lfloor \frac{n}{t} \right\rfloor + \left\lfloor \frac{n-1}{t} \right\rfloor + \left\lfloor \frac{n-2}{t} \right\rfloor + \dots + \left\lfloor \frac{t+1}{t} \right\rfloor + 1 \right) \\ &= \left( \left\lfloor \frac{16}{1} \right\rfloor + \left\lfloor \frac{16-1}{1} \right\rfloor + \dots + \left\lfloor \frac{1+1}{1} \right\rfloor + 1 \right) \cdot \left( \left\lfloor \frac{16}{2} \right\rfloor + \left\lfloor \frac{16-1}{2} \right\rfloor + \left\lfloor \frac{16-2}{2} \right\rfloor + \dots + \left\lfloor \frac{2+1}{2} \right\rfloor + 1 \right) \\ &= (16+15+\dots+1) \cdot (8+7+\dots+1) \\ &= 8704 \end{aligned}$$

$$\begin{aligned} \Omega_{(1,3)}^n &= \Omega_{(3,1)}^n \\ &= \left( \left\lfloor \frac{n}{s} \right\rfloor + \left\lfloor \frac{n-1}{s} \right\rfloor + \dots + \left\lfloor \frac{s+1}{s} \right\rfloor + 1 \right) \cdot \left( \left\lfloor \frac{n}{t} \right\rfloor + \left\lfloor \frac{n-1}{t} \right\rfloor + \left\lfloor \frac{n-2}{t} \right\rfloor + \dots + \left\lfloor \frac{t+1}{t} \right\rfloor + 1 \right) \\ &= \left( \left\lfloor \frac{16}{1} \right\rfloor + \left\lfloor \frac{16-1}{1} \right\rfloor + \dots + \left\lfloor \frac{1+1}{1} \right\rfloor + 1 \right) \cdot \left( \left\lfloor \frac{16}{3} \right\rfloor + \left\lfloor \frac{16-1}{3} \right\rfloor + \left\lfloor \frac{16-2}{3} \right\rfloor + \dots + \left\lfloor \frac{3+1}{3} \right\rfloor + 1 \right) \\ &= (16+15+\dots+1) \cdot (5+5+\dots+1+1) \\ &= 136 \cdot 40 \\ &= 5440 \end{aligned}$$

$$\begin{aligned} \Omega_{(2,2)}^n &= \left( \left\lfloor \frac{n}{s} \right\rfloor + \left\lfloor \frac{n-1}{s} \right\rfloor + \dots + \left\lfloor \frac{s+1}{s} \right\rfloor + 1 \right) \cdot \left( \left\lfloor \frac{n}{t} \right\rfloor + \left\lfloor \frac{n-1}{t} \right\rfloor + \left\lfloor \frac{n-2}{t} \right\rfloor + \dots + \left\lfloor \frac{t+1}{t} \right\rfloor + 1 \right) \\ &= \left( \left\lfloor \frac{n}{s} \right\rfloor + \left\lfloor \frac{n-1}{s} \right\rfloor + \dots + \left\lfloor \frac{s+1}{s} \right\rfloor + 1 \right)^2 \\ &= \left( \left\lfloor \frac{16}{2} \right\rfloor + \left\lfloor \frac{16-1}{2} \right\rfloor + \dots + \left\lfloor \frac{2+1}{2} \right\rfloor + 1 \right)^2 \\ &= 4096 \end{aligned}$$

$$\Omega^n = 2\Omega_{(1,2)}^n + 2\Omega_{(1,3)}^n + \Omega_{(2,2)}^n$$

$$\begin{aligned}
&=2 \times 8704+2 \times 5440+4096 \\
&=32384
\end{aligned}$$

同理可以得出不同大小的子窗口内特征矩形的总数目，如表

表 3.2 多尺度窗口下特征矩形的数目

窗口大小	16×16	20×20	24×24	30×30	36×36
特征矩形数目	32384	78460	162336	394725	816264

从表中容易可以看出，随着窗口的增大，特征矩形的数目也大幅度增加，当达到 36×36 时，特征矩形的数目已经超出像素数目本身很多了，这给计算特征值带来了很大的不变，在下节我们介绍积分图法来快速计算特征值。

### 3.1.3 利用积分图法快速计算类 Haar 特征

Viola 在 2001 年阐述了如何利用积分图在相同的时间内快速计算不同的类 Haar 特征值，极大地提高了检测速度。该思想如下：

设  $F(x)$  是  $f(x)$  的原函数， $f(x)$  在区间  $[a, b]$  上连续，则存在积分

$$\int_a^b f(x) dx = F(b) - F(a) \quad (3-7)$$

积分图的思想类似于这样一个简单的微积分。

对图像内一点  $(x, y)$ ，定义其积分  $ii(x, y)$  如下：

$$ii(x, y) = \sum_{x' \leq x} \sum_{y' \leq y} I(x', y') \quad (3-8)$$

3-8 式中  $I(x', y')$  是图像中坐标为  $(x', y')$  处的像素值，在灰度图像中就是  $(x', y')$  的点的灰度值，灰度值的值域是  $[0, 255]$ 。

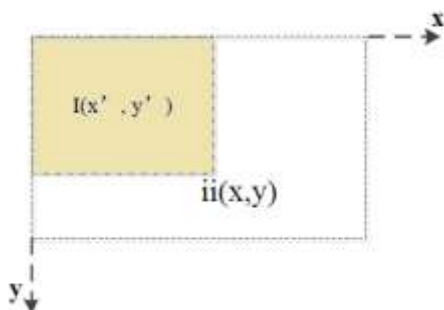


图 3.2  $(x, y)$  处的积分图

图 3.2 中坐标(x,y)的积分图是其左上角区域像素之和

ii(x,y)也可以通过迭代法进行求解：

$$ii(x,y) = ii(x-1,y) + s(x,y) \quad (3-9)$$

$$s(x,y) = s(x,y-1) + I(x,y) \quad (3-10)$$

$$s(x,y) = s(x-1,y) + s(x,y-1) - s(x-1,y-1) + f(x,y) \quad (3-11)$$

式中  $I(x,y)$  为图像在  $(x,y)$  处的像素值， $f(x,y)$  为原始图像灰度值  $s(x,y)$  为从  $(x,1)$  到  $(x,y)$  所有的像素和。

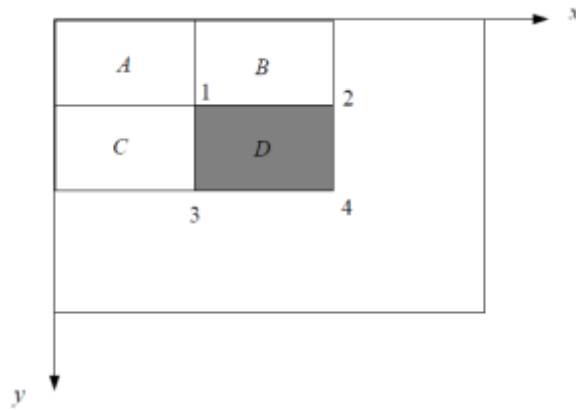


图 3.3 区域 D 像素和的计算

求区域 D 得像素和如图 3.3 所示，求区域 D 像素和可以通过下式：

$$I_D = \frac{[S(x_2, y_2) - S(x_2, y_1 - 1) + S(x_1 - 1, y_2) + S(x_1 - 1, y_1 - 1)]}{(x_2 - x_1 + 1)(y_2 - y_1 + 1)} \quad (3-12)$$

从 3-12 式中可以看出，特征值的计算要一个输入就能计算，即本身特征端点的积分图的值，与其他无关。任一矩形包含的像素点的灰度和都能在常数时间内完成，每次计算都只用索引四次积分图，进行四次简单的加减算法即可得到矩形特征值，也就是说计算时间是一定的，因此说积分图方法在相同的时间里提高了检测速度。

## 3.2 基于局部特征的 SIFT 特征提取算法

### 3.2.1 Bag-of-words 模型

Bag-of-words 模型是整理文档的一种方法,在信息检索领域已经运用广泛。对于每一个文档，都是由不同的词汇构成，如下面两个文档：

(1) Tom is a college student , and also Jerry.

(2) Jerry is an athlete too.

这两个文档可以用下面这个字典(Dictionary)来表示:

表 3.3 视觉字典

1	2	3	4	5	6	7	8	9	10	11
Tom	is	a	college	student	and	also	Jerry	an	athlete	too

该字典有 11 个不同的单词, 利用第一行的索引号, 则在两个文档中字典里的单词出现的次数可以用 11 维向量表示:

(1) {1,2,1,1,1,1,1,2,1,1,1}

(2) {0,1,0,0,0,0,0,2,1,1,1}

可以看出我们用一个 11 维向量表示出了字典中每一个单词在一个文档中出现的次数, 上表是 Bag-of-words 模型的应用, 两个文档检索的简单过程, 可以看出在检索过程中 Bag-of-words 模型忽略了文档的语义及顺序, 这是 Bag-of-words 模型在文档检索过程中的缺点, 但是这并不影响我们将 Bag-of-words 模型应用到人脸图像上来。

### 3.2.2 Bag-of-words 模型在人脸图像上的应用

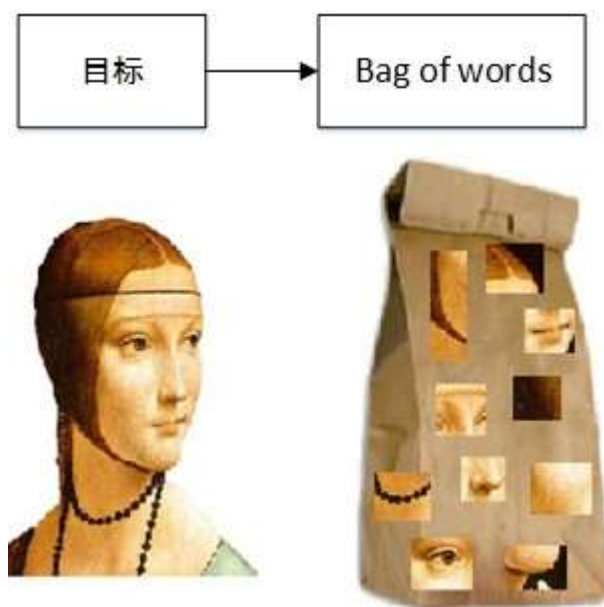


图 3.4 bag-of-words 模型

考虑一幅人脸，如 3.2.1 中所述，把人脸看做文档，那么我们也同样可以用视觉词汇来表示一幅人脸。人脸中的视觉词汇需要我们进行特征提取，可以从 3.2.1 中看到字典中的词汇具有独立，具有唯一性，类似的视觉词汇也就需要相对独立的，可以用来表示大部分人脸图像的特征。我们可以发现人与人的脸有不同，甚至会有很大的不同，但是每一个人都有鼻子、眼睛、嘴唇等一些细小的地方是观察不出来差别的，那么我们就可以把这些地方提取出来当做视觉词汇，进而构成视觉词典。起初 SIFT(Scale Invariant Feature Transform)算法用来进行图像匹配，提取的是图像中局部不变特征，运用广，由此我们用 SIFT 算法提取人脸图像中的不变特征点，这些不变特征点具有视觉词汇的特性，可以作为视觉单词构成单词表，每幅图像都可以用视觉词典来表示。

### 3.2.3 SIFT 算法提取不变特征点

1999 年，加拿大英属哥伦比亚大学的 David G.Lowe 教授提出了 SIFT(尺度不变特征转换)算法，并在 2004 年进一步完善了 SIFT 算法，可以进行图像匹配，该算法基于尺度空间，使用的特征是局部特征，在图像旋转、缩放和仿射变换后，提取的局部特征仍可以保持不变性。计算 SIFT 特征需要的是目标上局部外观的一部分关键点(Key Point)，这些关键点的提取与目标图像的大小和旋转没关系，具有良好的鲁棒性，并且 SIFT 特征将极值点转化为向量更加稳定，对有部分遮挡的图像检测率也相当高，运算速度也很快，按现在的计算机运行速度已经可以即时运算。

**SIFT 算法思想：**在图像的不同尺度下，SIFT 算法用图像的局部特征，在每一个尺度空间找到极值点，再提取出极值点的尺度、位置和旋转不变量。将这些极值点和不变量转化为向量来表示，即 SIFT 不变因子。

#### (1)构建尺度空间(Scale Space)

在尺度空间中寻找关键点(感兴趣点，角点等)，以及缩放和方向描述子得到特征。对一幅图像来说，可以有多种尺度，这些尺度是连续的。通过实验发现在一些尺度下会存在这样的极值点，这些点在其他尺度下是不能表示出来的，那么这些点就可以作为特征点，我们就需要用尺度空间来表示图像，这样才能更好的进行极值点的提取。

线性尺度空间的表示可以通过高斯卷积核来定义：

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \quad (3-13)$$

其中  $G(x,y,\sigma)$  是尺度可变高斯函数，\* 运算是在 x 轴方向与 y 轴方向做卷积， $I(x,y)$  表示输入图像。

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3-14)$$

其中 $(x,y)$ 为空间坐标， $\sigma$ 为尺度因子，图像的平滑程度决定了 $\sigma$ 的值， $\sigma$ 越大图像分辨率越低，反之越高。尺度空间越大就越能表示图像的概貌特征，反之代表的是图像的细节，能够准确的进行定位。

图像通过将具有可变核的高斯滤波函数  $G(x,y,\sigma)$ 和图像  $I(x,y)$ 做卷积得到尺度空间  $L(x,y,\sigma)$ ，可以用金字塔来形容尺度空间，如图所示：

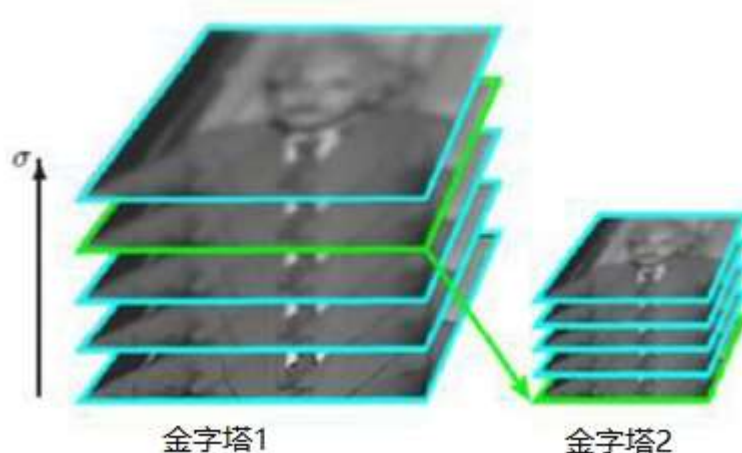


图 3.5 Gauss 金字塔

## (2)构建图像金字塔(Octave)

金字塔算法由 Burt 等人提出，最初应用在机器视觉和图像压缩方面。金字塔算法是基于高斯函数的，而图像金字塔是利用原始图像构建的一个图像序列，底层为待处理图像的高分辨率表示，向上的每一组的图像是由下一层降采样得到，一般上一层为下一层的低分辨率表示。

图像金字塔有抽样金字塔，均值金字塔，高斯金字塔(Laplacian of Gaussian)，高斯差分金字塔(DoG)四种。抽样金字塔构建速度快，但构建的金字塔图像混淆严重，像素的区域代表性不强。均值金字塔运算量比较小，有较好的视觉效果。高斯金字塔运算量偏大，具有良好的视觉效果。高斯金字塔中相邻的两层图像相减得到对应的高斯差分金字塔的一层，可以减少计算量，与高斯金字塔媲美。

## (3)构建高斯差分尺度空间



想要精确地在每一个尺度空间检找到合适的关键点,用尺度归一化的拉普拉斯函数的近似函数,即差分高斯函数  $D(x,y,\sigma)$ 来提取图像稳定的特征,DoG 算子如下:

$$\begin{aligned} D(x,y,\sigma) &= (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \\ &= L(x,y,k\sigma) - L(x,y,\sigma) \end{aligned} \quad (3-15)$$

其中  $k$  为常数,  $k=2^{1/s}$ , 代表两个相邻尺度之间的间隔,  $s$  为金字塔的每一层中含有的图像数目。DoG 算子直接使用高斯卷积核,比要使用两个方向的高斯二阶微分卷积核的 LoG 算子要省去了生成卷积核的计算量。DoG 算子可以直接利用高斯尺度空间的图像而不需要再生成新的尺度空间图像。DoG 是对 LoG 的近似简化,在保留了 LoG 的优点同时,一样具有良好的健壮性,尺度空间示意图如图所示:

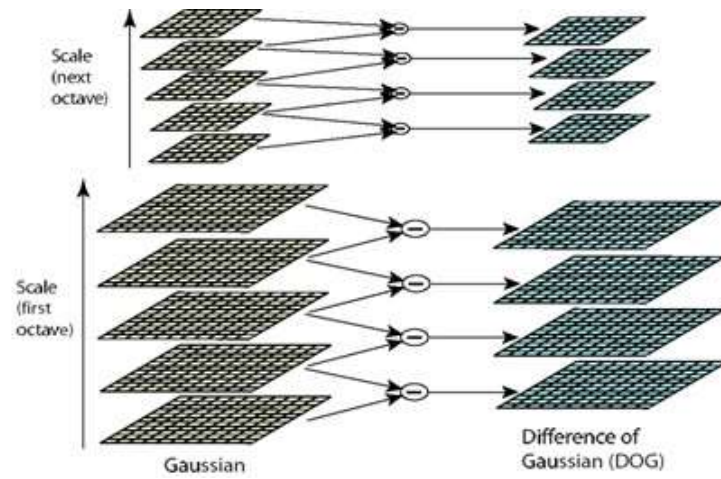


图 3.6 高斯尺度空间和高斯差分尺度空间

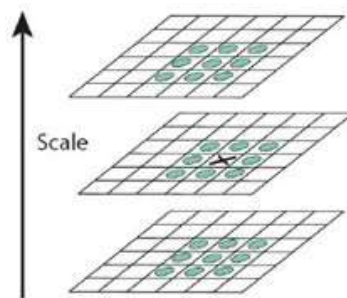


图 3.7 寻找空间极值点示意图

### (3)查找空间极值点

为寻找极值点,对图中每一个像素点来说,需要对周围 8 个和上下 18 个一共 26

个像素点进行比较，才能知道该点是不是极值点，假如这个点与 26 个周围像素点比较后是最大值或者最小值，则该点为图像在该尺度下的待定极值点。这样做的目的可以使所提取的特征在各尺度下都能保证其良好的独特性和稳定性。

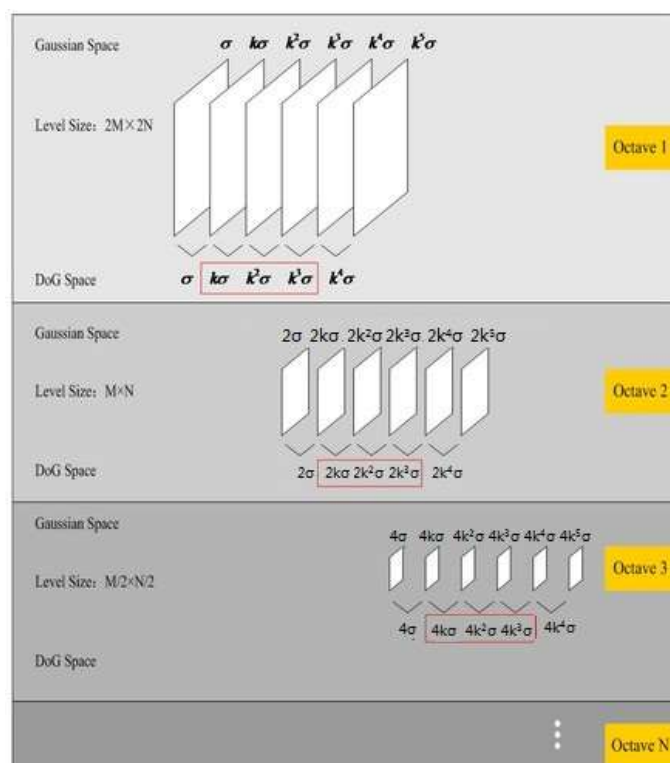


图 3.8 图像金字塔

图像金字塔需要有连续性才可以进行计算，为此在每一组的最上方用高斯模多产生 3 张图像，组成新的高斯金字塔。在新的金字塔中，每一组就包含  $s+3$  张图像，对应的高斯差分尺度金字塔包含  $s+2$  张图像。图像金字塔的连续性解释如下：假设每个塔层有 3 张图，即  $s=3$ ，则  $k = 2^{1/s} = 2^{1/3}$ ，那么高斯尺度空间和高斯差分尺度空间分别有 3 个和 2 个分量，在高斯差分尺度空间中，第一个八度(octave)两项分别是  $\sigma, k\sigma$ ；第二个八度的两项分别是  $2\sigma, 2k\sigma$ ；极值存在于 3 者或 3 者以上，两个是不能求出极值的，只有在高斯空间中增加一些高斯模糊项，才可以构成一列如  $\sigma, k\sigma, k^2\sigma, k^3\sigma, k^4\sigma \dots$  类似的序列，极值是上下都有的情况下才存在的，那么我们就可以用高斯差分尺度空间中的中间三项  $k\sigma, k^2\sigma, k^3\sigma$  求出极值，类似在下一八度中有中间的三项是  $2k\sigma, 2k^2\sigma, 2k^3\sigma$ ，第一项是  $2k\sigma = 2^{4/3}$ ，就能和前一个八度的最后项  $k^3\sigma = 2^{3/3}$  组合起来满足连续性尺度空间，这就是为什么每次要在 Gaussian space 中增加 3 项，增加后每个图像金字塔一共有  $S+3$  层图像，对应的高斯差分尺度金字塔有  $S+2$  张图像。

每一个 Octave 中的每一个极值点的坐标(Key point candidate)和尺度都需要保留：  
 $(x_1, y_1, \sigma_1), (x_2, y_2, \sigma_1), (x_3, y_3, \sigma_1), \dots, (x_n, y_n, \sigma_1), (x_{n+1}, y_{n+1}, \sigma_2), \dots, (x_N, y_N, \sigma_s)$

可以看到  $\sigma_1$  尺度下有  $n$  个极值点.....，一共  $N$  个极值点。对于每个像素点来说，与其比较的是周围的 26 个像素点，难免形成局部极值，而计算出的特征点的在空间分布会很集中，聚集在一个小的范围内，很容易出现聚类。当聚类发生时，聚类形成的特征点所包含的信息或许是图中一个或几个目标特征，代表的是局部的特征。但是我们需要的是可以变现图像的整体特征的特征点，当出现聚类时这些特征就不具有极值点的特性。这就需要我们对上面的这些极值点进行筛选。

#### (4)精确定位特征点

在上面计算得到的待定极值点中，需要找出不满足要求的极值点，进行筛选。有低对比度的极值点和不稳定的边缘响应点，需要运用拟合三维二次函数的方法，准确确定出关键点的位置和达到亚像素精度的尺度，使这些极值点的匹配可以更稳定，并能提高抗噪声能力。

删除边缘响应点，边缘处的极值点对于噪声极为敏感，高斯差分函数有很强的边缘响应，极易受噪声影响，为了确保特征匹配的稳定性和抗噪性，还需删除不稳定的边缘响应和噪声。不稳定的 DoG 算子的极值点，边缘切向方向的主曲率较大，而在边缘垂直方向的主曲率较小。为了得到稳定的特征点，只删除高斯差分尺度空间中响应值低的点还不行。由于 DoG 对图像中的 3 边缘很敏感，反应较强，只要有特征点出现在图像的边缘上，那么这些点就不能用来作为最终的极值点，是需要删除的不稳定的极值点。首先位于图像边缘上的点不容易定位，容易出现歧义；其次这样的点会受到噪声的干扰，使稳定性变差。一个平滑的 DoG 响应峰值具有在横跨边缘的地方大主曲率，在垂直边缘的地方有较小主曲率的特征。通过  $2 \times 2$  的 Hessian 矩阵  $H$  可以得到主曲率：

$$H(x,y) = \begin{bmatrix} D_{xx}(x,y) & D_{xy}(x,y) \\ D_{xy}(x,y) & D_{yy}(x,y) \end{bmatrix} \quad (3-16)$$

上式中， $D_{xy} = \frac{\partial}{\partial x \partial y}$  的值可以由采样点相邻差估计得到。其中  $D$  的主曲率与  $H$  的特征值成正比。设  $\alpha, \beta$  分别是最大和最小的特征值，通过  $H$  矩阵计算它们的和， $H$  矩阵的行列式计算它们的乘积：

$$\text{Tr}(H) = D_{xx} + D_{yy} = \alpha + \beta \quad (3-17)$$

$$\text{Det}(H) = D_{xx} D_{yy} - D_{xy}^2 = \alpha \cdot \beta \quad (3-18)$$

令  $\alpha = \gamma\beta$ ，则：

$$\frac{\text{Tr}(H)^2}{\text{Det}(H)} = \frac{(\alpha + \beta)^2}{\alpha \beta} = \frac{(\gamma\beta + \beta)^2}{\alpha \beta} = \frac{(\gamma + 1)^2}{\gamma} \quad (3-19)$$

上式的结果只与两个特征值的比值相关，与具体特征值无关。可以看出  $(\gamma + 1)^2/\gamma$  的值取最小是在两个特征值相等的时候， $(\gamma + 1)^2/\gamma$  随着  $r$  的变化而变化， $r$  增大，则  $(\gamma + 1)^2/\gamma$  增大，反之减小。所以，想要判断主曲率的比值是否小于阈值  $\gamma$ ，只需判断下式即可：

$$\frac{\text{Tr}(\mathbf{H})^2}{\text{Det}(\mathbf{H})} < \frac{(\gamma + 1)^2}{\gamma} \quad (3-20)$$

在 Lowe 的论文里， $\gamma$  取值是 10，那么主曲率比值比 10 大的特征点就会被删除，其余将保留。

#### 关键点方向分配

经过上面精确处理后余下的特征点称为关键点。有了关键点后，为了实现特征点描述子的旋转不变性，需要给每个特征点附加一个方向，而且这个方向的计只依赖于局部图像信息，利用这样的方向来表示特征点描述子，那么这个特征点描述子便是旋转无关的。

只对关键点处像素进行方向估计还不够，为了计算关键点处的主方向，借助关键点邻域像素的梯度方向分布特征为每个关键点定义方向参数。采用邻域像素梯度直方图，使邻域像素通过加权考虑进来，就可以更加精确地估计主方向。

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \quad (3-21)$$

$$\theta(x,y) = \arctan ((L(x+1,y) - L(x-1,y)) / (L(x,y+1) - L(x,y-1))) \quad (3-22)$$

3-21,3-22 式代表坐标  $(x,y)$  的点梯度模值和方向的计算公式， $L$  用的尺度是每个关键点自己所在的位置的尺度。

用直方图统计的是以关键点为圆心的邻域窗口内的各个像素的梯度方向。范围为  $0 \sim 360^\circ$ ，每个直方柱代表 10 度，一共有 36 个直方柱，邻域点与关键点的距离越近，其对梯度方向权值的贡献就越大，直方图中的最大值表示该关键点地方的邻域梯度的主方向，计算出的这个方向就是该关键点的方向。

目前，最终可以用的图像关键点都被检测到了，每个关键点一共包含三种数据：位置、所处尺度和方向。三种数据一起决定了一个 SIFT 特征区域。

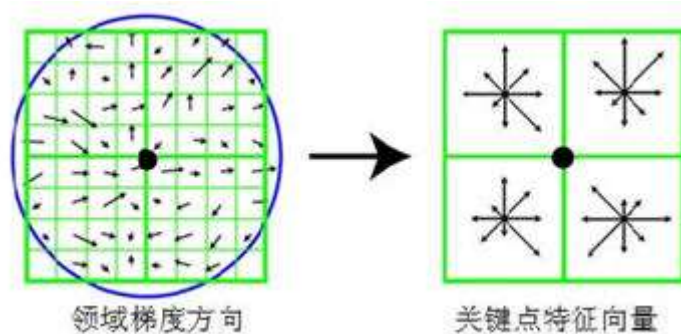


图 3.9 SIFT 特征方向图

### (5)生成描述子

对每一个 SIFT 特征区域，求该区域的描述子，即图像的特征向量。采用向量形式来确切描述图像自身的特征，可以使使其具有高区分性，且对光照和视角的变化尽可能保持不变。

如图所示，要创建一个关键点描述子，先要计算关键点位置附近的每个图像样本点的梯度大小和方向。Lowe 在实验中使用  $4 \times 4$  方格的描述子，每格有 8 个分量(方向)，每个分量的幅值是通过一个  $4 \times 4$  子区域内各点梯度方向构成的直方图加权计算得到的。

每个  $4 \times 4$  的图像小块在  $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ$  和  $360^\circ$  这 8 个方向的梯度方向直方图，这种方法运用的是把邻域方向性信息结合的思想，对算法抗噪声的能力进行提升，也使得对定位有误差的特征进行匹配时具有较好的容错性。每一个关键点就能生成 128 个数据，组成了 128 维度的 SIFT 特征向量,即特征描述子。此时，得到的 SIFT 特征描述子，已经解决了平移、尺度变化、旋转等几何变形因素带来的问题，对光照的影响和视角的改变也进行了相应的处理。

### 3.2.4 K-Means 构造单词表

我们从 3.2.3 中得到了  $N$  个 128 维的特征描述子，我们需要把这些特征描述子进行聚类构成视觉词汇字典。

聚类，简单地说就是把相似的东西分到一组。在聚类的时候，我们并不关心每一类是什么，我们需要实现的目标是把相似的东西聚到一起，因此，一个聚类算法通常只需要知道如何计算相似度就可以开始工作了。在机器学习中聚类属于无监督学习的一种。

聚类的算法思想：

假设给定数据集  $(x_1, x_2, \dots, x_n)$ ，已经给定正整数  $k(k \leq n)$ ，每个  $x_i$  是  $d$  维特征向量

(1) 从给定数据集 $(x_1, x_2, \dots, x_n)$ ，随机选取一个特征向量作为起始的一个聚类中心 $c_1$ 。

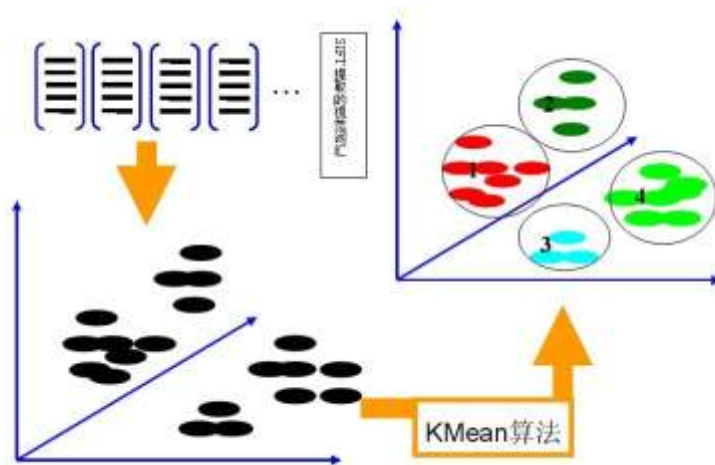
(2) 用下式来计算下一个数据点 $c_i$ ，被选为聚类中心的概率， $c_i \in (x_1, x_2, \dots, x_n)$

$$P = \frac{D(x)^2}{\sum_x D(x)^2} \quad (3-23)$$

式中  $D(x)$  表示数据点  $x$  到已选最近聚类中心的最短距离。

(3) 重复(2)直到找到  $k$  个符合条件的中心为止。

**K-Means** 算法基于样本间相似性度量，把相似度高的进行聚类，**K-Means** 中参数  $K$  代表总簇数，需要把一共  $N$  个对象划分成  $K$  个簇，这  $K$  个簇具有簇内相似度高，簇间相似度低的特性。依据词汇向量距离的大小，**K-Means** 算法就可以将 **SIFT** 提取的视觉词汇向量中词义相近的词汇进行合并，组成单词表



基础词汇(Visual Words)。如图 3.10 所示

图 3.10 K-Means 聚类

### 3.3 本章小结

在本章特征提取中，我们对经典的类 **Haar** 特征的求法进行了理论验证，在本文中采用 2-矩阵特征，利用积分图法快速对矩形特征进行求解。之后介绍了在文档中应用广泛的 **Bow** 模型，并将 **Bow** 模型应用到人脸特征提取上，利用 **SIFT** 提取不变因子作为视觉词汇，用 **K-Means** 进行聚类。

可以看出视觉字典在 **Bow** 模型尤为关键，视觉字典的构造法还存在很多问题，本文将在第五章提出改进的 **Bow** 视觉字典的构造方法。

## 第四章 性别识别的分类方法

### 4.1 AdaBoost 分类器

1984 年, Valiant 将计算机复杂理论和决策理论结合, 首先提出了新的学习模型——可能性近似正确(Probably Approximately Correct,PAC)学习模型。Valiant 在 PAC 学习模型中提出了强学习和弱学习这两种重要概念, 大大推动了机器学习理论的发展。而 PAC 也逐渐发展成为集成学习和统计学习的理论基础。PAC 的发展带来了通过弱学习来训练强学习的机器学习的新方法, 以及后来的 Boosting 算法也是将弱学习提升为强学习。弱学习训练过程简单, 人们发现通过迭代弱学习的训练, 可以产生强学习, 这使得机器学习也更加简单, 为机器学习的训练方法开辟了新的方向。

1990 年 Schapire 提出了来源于 PAC 的 Boosting 算法, Schapire 证明可以将多个 PAC 分类器集成可以具有 PAC 强分类器的能力。1995 年, Schapire 和 Freund 在 Boosting 算法的基础上首先提出了 AdaBoost(Adaptive Boosting,自适应集成学习)算法, 证明了 AdaBoost 算法的收敛性。2001 年, Viola 与 Jones 首次将 AdaBoost 算法应用到人脸检测领域, AdaBoost 发展至今仍然是人脸检测的主流方法, 在音视频和文本信息检索等领域也有很多应用。

在本文的第三章中已经对类 Haar 特征进行了介绍, 本节主要对 AdaBoost 算法进行一些讲解, AdaBoost 在二值分类和模式识别中都有较强的表现, 正确率相比其他传统方法有显著提升。

#### 4.1.1 Boosting 方法

机器学习(Machine Learning)是人工智能的核心, 这是一门多领域的交叉学科。简单来说就是希望计算机能够模拟人的学习能力, 可以根据已有的环境和事物中发现规律并自动获取识别能力, 进而自动通过在学习获得知识技能, 不断改善性能, 最终实现自我完善的能力。弱学习方法是指比随机猜测准确率高的虚席方法, 相对的, 可以显著提高分类正确率的称做强学习。Kearns 提出了弱学习算法转化为强学习算法的思想, 在文章中证明了在样本集足够的条件下, 可以将弱学习算法通过集成学习得到强学习方法。

在 Boosting 算法问世前, 有两种和 Boosting 算法思想相似的算法:

一种是 bootstrapping 方法, 其简单步骤:

- 重复地从样本集  $D$  中采集  $n$  个样本;



- 对每次采样得到的子样本集进行统计学习，得到假设 $H_i$ ；
- 经过重复前两步，将若干个假设进行组合，形成最终的假设 $H_{final}$ ；
- 用 $H_{final}$ 进行分类。

另一种是 bagging 方法，其简单步骤：

- 从整体样本集中，抽样  $n < N$  个样本，训练生成分类器；
- 分类器进行投票得出结果。

上面的两种类似 boosting 的方法都仅仅将弱分类器进行简单的结合，分类器组合的效果并没有发挥出来。1990 年 Boosting 问世，Schapire 还因此获得了 2003 年的哥德摩尔奖(Godel Price)。

Boosting 算法步骤：

- (1) 从整体样本集合  $D$  中不放回地随机抽取  $n_1 < n$  个样本，组成  $D_1$ ，用抽取的  $n_1$  个样本训练弱分类器  $C_1$ ；
- (2) 从  $D$  中不放回地抽取  $n_2 < n$  个样本，并将  $C_1$  分错的一半样本加入组成集合  $D_2$ ，训练弱分类器  $C_2$ ；
- (3) 从  $D$  中再次抽取样本，组成  $D_3$ ，训练弱分类器  $C_3$ ；
- (4) 用  $C_1 C_2 C_3$  投票得到最终结果。

可以从训练过程中看出，Boosting 方法中样本集互不相同，前一个分类器决定了后一个分类器一部分的样本，分类正确留再原样本集，错误则需要将错误的一半样本分入下一个分类器的样本集中，这样以来，后面的分类器则不需再对已经分类争取的样本分类，专注于上一个分错的样本，通过不断迭代提升性能。Boosting 的不足之处是需要预先知道弱分类器的分类下限，而后面的弱分类器关注于识别困难的样本容易造成过拟合，使得分类器性能不稳定，达不到预期效果。

#### 4.1.2 AdaBoost 算法原理

从上面的 Boosting 算法介绍中可以看出 Boosting 算法存在的两种问题，首先，需要调整训练集的组成，使弱分类器可以在特定的训练集上进行训练。AdaBoost 算法可以在弱分类器训练过程中自适应的调整分类器的错误率。在训练过程中，AdaBoost 算法会选择分类精准度最高的分类器作为这次迭代的弱分类器，选择的弱分类器的分类精准度肯定超过 50%。这是 Boosting 算法所不具备的特点，AdaBoost 算法无需刻意对训练集进行筛选，即不需要知道样本集的下限。AdaBoost 算法中的弱分类器 都是经过 AdaBoost 方法训练得到的，不需要预先知道弱分类器的分类下限才能分类。



其次，如何将上面训练得到的弱分类器进行级联，并构成强分类器。而在 AdaBoost 中完美地解决了这个问题。分类器的级联策略指通过级联一系列强分类器达到高效、快速进行人脸检测的目的。级联强分类器使得 AdaBoost 算法更有效，精度也更高。而每个强分类器都可以对样本集进行分类，随着迭代的进行，后面的分类器获得的样本会越来越少，这样，后面的分类器的工作量就会很少。分类器级联策略的关键思想是通过调整参数值使样本集中的正样本可以通过每一级分类器，而负样本大部分都会被分类器拒绝，位于前面的分类器获得的矩形特征较少、运算速度快，排除了大部分的负样本，那么经过后面分类器的样本数很少，自然使得后面的分类器达到很低的错误报警率。

### 4.1.3 AdaBoost 算法训练过程

AdaBoost 的核心算法思想是对同一个训练样本集合训练出不一样的弱分类器，再将这些弱分类器进行级联，构成强分类器。训练过程简述：

- ◆ 初始给定样本集：\$(x\_1, y\_1) \dots (x\_m, y\_m)\$，其中 \$x\_i \in X\$，\$y\_i \in Y = \{-1, +1\}\$，\$y\$ 的值代表正负样本；
- ◆ 初始化权重 \$D\_1(i) = \frac{1}{2l}, \frac{1}{2n}\$，其中 \$l\$ 是正样本数量，\$n\$ 是负样本数量；
- ◆ 进入循环：For \$t = 1 \dots T\$，\$T\$ 是迭代次数；
- ◆ 根据分布 \$D\_t\$，在训练集上训练弱分类器 \$h\_t(x\_i)\$；
- ◆ 用分类器 \$h\_t(x\_i)\$ 对初始分布进行分类；
- ◆ 选择本次迭代出的最佳弱分类器 \$h\_t(x\_i)\$：\$h\_t(x\_i)\$ 满足 \$\epsilon\_t = \min \sum D\_t(i) |h\_t(x\_i) - y\_i|\$；
- ◆ 计算本次迭代出的最佳弱分类器 \$h\_t(x\_i)\$ 的权重：\$\alpha\_t = \frac{1}{2} \ln \left( \frac{1 - \epsilon\_t}{\epsilon\_t} \right)\$
- ◆ 更新训练样本权重 \$D\_{t+1}(i) = \frac{D\_t(i) \exp(-\alpha\_t y\_i h\_t(x\_i))}{Z\_t} = \frac{D\_t(i)}{Z\_t} \times \begin{cases} e^{-\alpha\_t}, & \text{if } y\_i = h\_t(x\_i) \\ e^{\alpha\_t}, & \text{if } y\_i \neq h\_t(x\_i) \end{cases}\$  
其中 \$Z\_t\$ 是归一化因子，满足 \$\sum\_{i=1}^n D\_{t+1}(i) = 1\$；
- ◆ 如果 \$t < T\$，则转到(3)，继续迭代。
- ◆ \$T\$ 次循环迭代之后，产生了 \$T\$ 个弱分类器，级联组成最终的强分类器，用函数表示：

$$H(x) = \text{sign}(\sum_{t=1}^T \alpha_t h_t(x_i)) \quad (4-1)$$

- ◆ 其中 \$\text{sign}\$ 为符号函数，\$\alpha\_t\$ 在最终的强分类器中体现了话语权，\$\alpha\_t\$ 值越大，那么 \$h\_t(x\_i)\$ 对分类就具有决定性作用，反之作用越小。

### 4.1.4 AdaBoost 的优缺点

优点

- AdaBoost 分类器是一种分类精度很高的分类器；
- 弱分类器构造非常简单，当运用简单分类器分类时，分类出来的结果是能够理解的；
- 可以使用各种方法构建弱分类器，AdaBoost 算法是基于一种框架的思想算法；
- 不需要进行特征筛选，不要担心过拟合问题。

缺点

- 训练时间长，从前面介绍类 Haar 基可以看到 AdaBoost 算法在训练弱分类器时采用的是穷尽搜索法，这样就会得到大量的矩形特征，花费巨大的时间和空间成本；
- 样本权值分布扭曲，在 AdaBoost 算法的训练过程中，弱分类器是通过改变权值来训练的，若分类器对样本集分类正确，则会减少该样本集的权值使其在下次训练中不受到很大关注，反之，如果分类错误就会增大权重。然而当训练集中包含一些困难样本时，AdaBoost 算法也会增大其权重，就会导致过学习现象的发生，也就是样本权值分布扭曲。

## 4.2 朴素贝叶斯分类器(Naïve Bayes Classifier,NBC)

贝叶斯方法是机器学习的一种重要方法，朴素贝叶斯分类方法具有简单、高效等优点，朴素贝叶斯假定分类数据的各项属性是独立的，这符合人脸的特性，而朴素贝叶斯有着贝叶斯定理这样坚实的理论基础。

### 4.2.1 概率论基础

#### (1) 条件概率和乘法定理

在样本空间  $S$  中， $A, B$  是两个随机事件，在事件  $B$  发生的情况下事件  $A$  发生的概率称为事件  $A$  在给定事件  $B$  的情况下的条件概率，即后验概率。记作  $P(A|B)$ ； $P(A)$  为无条件概率，即先验概率。

条件概率公式：

$$P(A|B) = \frac{P(A \cdot B)}{P(B)} \quad (4-2)$$

乘法定理公式：

$$P(A \cdot B) = P(A|B)P(B) \quad (4-3)$$

对于  $n$  个事件  $A_1, A_2, A_3, \dots, A_n (n \geq 2)$  类似有：

$$P(A_1, A_2, \dots, A_n) = P(A_n | A_1, A_2, \dots, A_{n-1}) P(A_{n-1} | A_1, A_2, \dots, A_{n-2}) \dots P(A_1) \quad (4-4)$$

## (2) 全概率公式和贝叶斯定理

在样本空间  $S$  中,  $A_1, A_2, \dots, A_m$  为样本空间  $S$  的一个划分, 且  $P(A_i) > 0 (i=1, 2, 3, \dots, n)$ , 则对任意事件  $B$ , 有

$$P(B) = \sum_{i=1}^n P(A_i) P(B|A_i) \quad (4-5)$$

$S$  上式成为全概率公式。

由全概率公式及条件概率公式能够得到:

$$P(A_j|B) = \frac{P(A_j B)}{P(B)} = \frac{P(A_j) P(B|A_j)}{P(B)} \quad (4-6)$$

## (3) 事件独立性

设  $A, B$  是两个独立事件, 当  $A, B$  的发生互不影响时, 即  $P(A|B)=P(A)$ , 此时有:

$$P(A \cap B) = P(A|B)P(B) = P(A)P(B) \quad (4-7)$$

则称  $A, B$  为相互独立事件。

类似的, 对于  $n$  个事件  $A_1, A_2, A_3, \dots, A_n (n \geq 2)$ , 如果有

$$P(A_1 \cap A_2 \cap A_3 \dots A_n) = P(A_1)P(A_2)P(A_3) \dots P(A_n) \quad (4-8)$$

则称为  $A_1, A_2, A_3, \dots, A_n$  是相互独立事件。

## (4) 极大后验假设与极大似然假设

朴素贝叶斯定理提供了一种检验假设  $h$  概率的办法, 基于  $h$  的先验概率 (prior probability), 给定假设下观察到的不同数据概率以及观察到的数据本身。  $P(h)$  表示没有训练数据前假设  $h$  的先验概率, 表示关于  $h$  是正确假设的概率的背景知识。  $P(D)$  表示  $D$  的先验概率, 即在没有确定某一假设成立时  $D$  的概率。由贝叶斯公式求得计算后验概率  $P(h|D)$  的公式:

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)} \quad (4-9)$$

极大后验假设 (Maximum A Posteriori, MAP): 分类器设定类别集合  $C$  为一个待选的假设集合, 并在其中寻找给定的训练样本  $x$  能够达到可能性最大的某个假设  $C_i \in C$ , 而这种假设或类别就被称作为极大后验假设, 记作  $C_{map}$ 。

在一些情况下, 可假设式中的每个假设都有相同的先验概率, 此时只要对  $P(D|h)$  找到极大可能假设。则  $P(D|h)$  被称做极大似然假设, 记为  $h_{ML}$ 。

## 4.2.2 朴素贝叶斯分类原理及模型

朴素贝叶斯分类的思想: 对于给出的待分类项, 假设所有的属性是相互独立互不影响的。在没有更多的可用信息时, 该项属于哪一类的概率最大, 则朴素贝叶斯算法

就把该项分为此类。朴素贝叶斯分类思想简单，大概流程如下：

(1) 设待分类项  $x = \{a_1, a_2, a_3, \dots, a_m\}$ ,  $x$  是  $m$  维的特征向量  $a_i$  互相独立互不影响。

(2) 有类别集合  $C = \{y_1, y_2, \dots, y_n\}$ 。

(3) 计算  $P(y_1|x)$ ,  $P(y_2|x)$ ,  $\dots, P(y_n|x)$  的值。

(4) 如果  $P(y_k|x) = \max\{P(y_1|x), P(y_2|x), \dots, P(y_n|x)\}$ ,

则  $x \in y_k$ 。

在上述过程中，我们在第三章中已经得到了 128 维的特征向量，性别识别的分类是两种，即  $C = \{0, 1\}$ , 0 代表男性，1 代表女性。那么问题的关键就是求解第三步，下面一节我们将对人脸图像上的朴素贝叶斯分类器的建立进行详细说明，并求得第三步中的概率。

### 4.2.3 朴素贝叶斯分类器的设计

朴素贝叶斯分类器假定一个属性的值与其他属性的值对给定类的影响是独立的，这个限制条件比较苛刻，我们的人脸特征具有互相独立的特性，在 BoW 的条件下满足朴素贝叶斯的基本条件。

给定一张特定的图像  $Q$ ，那么这张图像可以用一个矩形框  $B = \{x_c, y_c, w, h, \theta\}$  如图所示， $D$  是我们数据库中一张图像。

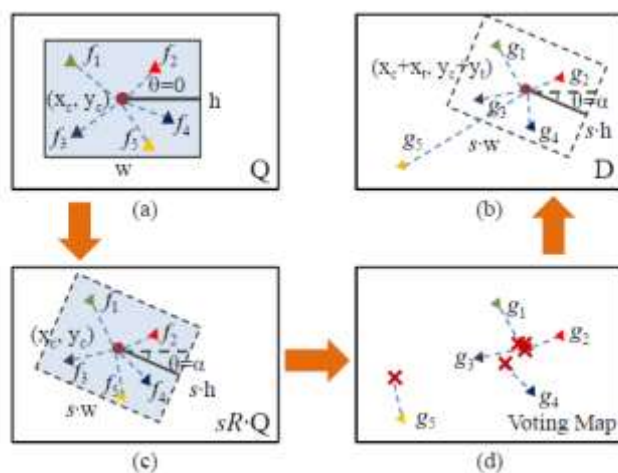


图 4.1 SCSM 的流程图

$(x_c, y_c)$  是矩形框中心的坐标， $w, h$  分别是矩形框的宽和高， $\theta$  可以由图中看出是矩

形框相对于水平方向的夹角。为了将这张图  $Q$  与其他图像进行比较，我们需要对其进行缩放、旋转和平移。设  $s$  表示缩放因子， $t=(x_t, y_t)$  是平移量， $T$  是旋转矩阵。 $B'$  是进行缩放、旋转和平移之后矩形框的位置，则

$$B' = T(B) = \{x_c + y_t, y_c + y_t, s \cdot w, s \cdot h, \theta = \alpha\} \quad (4-10)$$

这样我们就可以对测试图像( $Q$ )和数据库( $D$ )的图像进行变换，则我们要求得两幅图像中特征的相似度。定义空间约束相似度测量(Spatially Constrained Similarity Measure, SCSM):

$$S(x, c_i) = \sum_{k=1}^N \sum_{\substack{(f_i, g_i) \\ f_i \in Q, g_i \in D \\ w(f_i) = w(g_i) = k \\ \|T(L(f_i)) - L(g_i)\| < \varepsilon}} \frac{idf^2(k)}{tf_Q(k) \cdot tf_D(k)} \quad (4-11)$$

- $S(x, c_i)$  即相似度;
- $k$  代表视觉词汇字典中第  $k$  个视觉单词;
- $N$  是视觉字典的总量, 即 **K-Means** 聚类后生成的种数和;
- $w(f_i) = w(g_i) = k$  代表特征分量  $f_i$  和  $g_i$  均属于第  $k$  个视觉单词;
- $L(f)$  是二维空间中  $f$  的位置,  $T(L(f))$  是  $Q$  在经过变换后的坐标;
- $\|T(L(f_i)) - T(L(g_i))\| < \varepsilon$  代表经过  $T$  变换后  $Q$  和  $D$  中的特征应该很接近;
- $idf^2(k)$  是视觉词汇在数据库中出现次数的倒数, 即第  $k$  个视觉词汇的频率;
- $tf_Q(k)$  是第  $k$  个视觉词汇在  $Q$  中出现的次数;
- $tf_D(k)$  是第  $k$  个视觉词汇在  $D$  中出现的次数;

至此我们得出输入的一张图像  $Q$  与训练集中的图像  $D$  中的相似度  $s_i(x), y \in \{0, 1\}$ , 其中 0 代表男性作为正样本, 1 代表女性作为负样本。对于每一个样本和一个小的常量  $\epsilon$ , 存在一个阈值  $t_i$  使得

$$P(y = 1 | s_i(x) > t_i) \geq 1 - \epsilon \quad (4-12)$$

$$P(y = 0 | s_i(x) > t_i) \leq \epsilon \quad (4-13)$$

式中  $t_i$  是一个确定的阈值,  $\epsilon$  是一个非常小的正小数。

那么在  $C_i$  的周围就存在一个超球体, 把满足条件  $s_i(x) > t_i$  的特征都包括进来, 那么有很大概率可以断定  $x$  属于  $y=1$  这一类, 若不满足条件  $s_i(x) > t_i$ , 即  $s_i(x) \leq t_i$ , 我们不能断定  $x$  属于哪一类, 也就是说通过该特征我们不能对目标进行分类, 该特征贡献为 0, 即

$$P(y=1|s_i(x) \leq t_i) = P(y=0|s_i(x) \leq t_i) \quad (4-14)$$

在测试阶段，我们假设有  $m$  个正样本样例，那么相似性  $s_i(x)$  可以这样定义：

$$(s_1, \dots, s_m) = \frac{P(s_1, \dots, s_m | y=1)}{P(s_1, \dots, s_m | y=0)} \quad (4-15)$$

假设  $S_i$  是互相独立的，我们在这里取对数，可以得到朴素贝叶斯对数相似比：

$$\begin{aligned} \log L(s_1, \dots, s_m) &= \sum_{i=1}^m \log \frac{P(s_i | y=1)}{P(s_i | y=0)} \\ &\propto \sum_{i=1}^m \log \frac{P(y=1|s_i)}{P(y=0|s_i)} \end{aligned} \quad (4-16)$$

假设有  $n$  个样本满足条件  $s_i(x) > t_i$ ，那么剩下的  $m-n$  个样本就满足  $s_i(x) \leq t_i$ ，有  $\log \frac{P(y=1|s_i)}{P(y=0|s_i)} = 0$ ，据此有

$$\log L(s_1, \dots, s_m) = \sum_{i:s_i(x) > t_i} \log \frac{P(y=1|s_i)}{P(y=0|s_i)} \geq n \log \frac{1-\epsilon}{\epsilon} \quad (4-17)$$

实际上如果有更多的样本近似于测试样本，那么这个对数相似比会更高，那么  $x$  就更有可能判断是男性，所以我们用这样一个对数相似比来进行分类。

那么下面我们就可以进行分类器的设计。为了计算朴素贝叶斯对数相似比，我们需要一个更加详细的关于  $\epsilon$  的形式，如下定义：

$$P(y=1|s_i(x) > t_i) = 1 - \epsilon e^{-f(s_i(x))} \quad (4-18)$$

$$P(y=0|s_i(x) > t_i) = \epsilon e^{-f(s_i(x))} \quad (4-19)$$

式中  $f(s_i(x))$  是关于  $s_i(x)$  的单调递增函数，在实际应用中，我们选择  $f(s_i(x)) = s_i(x) - t_i$ 。则有

$$\begin{aligned} \sum_{i:s_i(x) > t_i} \log \frac{P(y=1|s_i)}{P(y=0|s_i)} &= \sum_{i:s_i(x) > t_i} \log \frac{1 - \epsilon e^{-f(s_i(x))}}{\epsilon e^{-f(s_i(x))}} \\ &= \sum_{i:s_i(x) > t_i} \log \left( \frac{1}{\epsilon} e^{s_i(x) - t_i} - 1 \right) \end{aligned} \quad (4-20)$$

因为  $\epsilon$  非常小，当  $s_i(x) - t_i > 0$  时，上式可以大约得到下式

$$\begin{aligned} \sum_{i:s_i(x) > t_i} \log \frac{P(y=1|s_i)}{P(y=0|s_i)} &= \sum_{i:s_i(x) > t_i} \log \frac{1}{\epsilon} e^{s_i(x) - t_i} - 1 \\ &\approx \sum_{i:s_i(x) > t_i} \log \frac{1}{\epsilon} e^{s_i(x) - t_i} \\ &= -n \log \epsilon + \sum_{i:s_i(x) > t_i} (s_i(x) - t_i) \end{aligned} \quad (4-21)$$

第一个量是常数，第二个量可以由前面聚合得到。

对于每一个正样本 $C_i$ 和它对应的分类器，阈值 $t_i$ 需要确定下来。为了可以有批评的对阈值进行学习，除了正样本集，我们也训练了负样本集 $N'$ 。

对于给定的负样本集和一个特定的样本 $c_i$ ，我们需要求得一个阈值 $t'_i$ ，使得  $P(s_i(x) > t_i | x \in N')$ 取最小值。如果

$$t_i \geq \max_{j \in N} s_i(x_j) \quad (4-22)$$

则  $P(s_i(x) > t_i | x \in N') = 0$

当满足上式约束后，我们可以在保证分类准确的情况下扩大 $C_i$ 的超球体，这样可以尽可能从正样本集中包括到。

$$\bar{t}_i = \arg \max_{t_i} P(s_i(x) > t_i | x \in P) \quad (4-23)$$

$$\text{s. t. } t_i \geq \max_{j \in N} s_i(x_j) \quad (4-24)$$

显而易见，当 $t_i$ 更小时，上式中的目标函数久更大。因此我们选择最终的阈值 $t_i$ 为：

$$t_i = \max_{j \in N} s_i(x_j) \quad (4-25)$$

也就是说，阈值 $t_i$ 是取样本 $c_i$ 与负样本训练集的相似值中最大值。





## 第五章 基于改良 BoW 模型的性别识别算法

本章首先介绍经典的 BoW 模型，然后介绍基于 Dense SIFT 的改良 BoW 模型，最后介绍性别识别算法。

### 5.1 经典的 BoW 模型

传统的 Bow 模型如图 5.1 所示，简单来说，对训练集和测试集分别进行识别、特征提取，用训练集的特征构造视觉词典，用视觉词典来训练 NBC 分类器，再将测试集的视觉词汇用视觉词典就行特征相似度判定，最后进行性别分类。

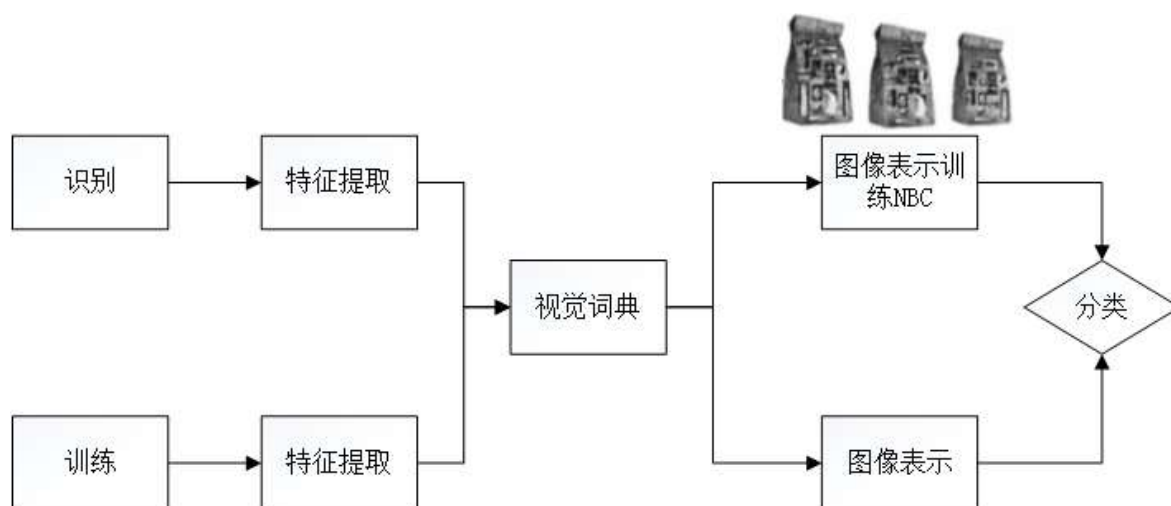


图 5.1 经典 BoW 模型示意图

从 3.2 节可知，传统的 BoW 模型中存在如下问题：（1）特征提取过程中用稀疏 SIFT 特征，需要构建高斯尺度空间和高斯差分尺度空间，在构建尺度空间过程中需要大量的空间和计算花费；（2）在分类器构造过程中，传统的 NBC 分类器中的度量在 BoW 模型中并不适用，需要找到合适的度量值进行特征相似度的计算。

为解决上述问题，本章提出了改进的特征提取方法和用来衡量特征相似度的新的空间相似度测量的方法。

## 5.2 基于改良 BoW 模型的性别识别算法

### 5.2.1 改良的 BoW 特征提取模型

在 Bow 模型中我们可以看到,需要大量计算的是 SIFT 特征子的提取,为了提高 Bow 模型的检索速度,在这里我们用 Dense SIFT 代替稀疏 SIFT,作为新的特征子。稀疏 SIFT 特征和 Dense SIFT 特征都是 128 维向量,但 Dense SIFT 与稀疏 SIFT 不同的地方是 Dense SIFT 只用了提取稀疏 SIFT 特征的一步,即当都找到了所有特征点的具体位置后, Dense SIFT 算法直接生成 128 维的特征向量。稀疏 SIFT 特征需要构建高斯尺度空间,并利用高斯模糊构造连续的图像金字塔,花费比较高。Dense SIFT 则不需要进行高斯尺度空间的建立, Dense SIFT 特征法将图像均匀分割成  $n \times n$  个均匀的小图像区域,每一个均匀分块的小区域都作为兴趣点进行特征提取。

Dense SIFT 算法提取图像中所有设定好的特征点的特征值,在每个特征点的周围设置一个  $16 \times 16$  的窗口,生成  $4 \times 4 \times 8$  的特征向量,生成一个 128 维的特征向量,与 SIFT 是相同的。

### 5.2.2 改进的 NBC 分类器

在第四章中介绍了经典的 NBC 分类器,在本文中我们采用空间相似度这一定义来计算性别特征之间的相似度,进而进行分类。空间相似度定义如下:

空间约束相似度测量(Spatially Constrained Similarity Measure,SCSM):

$$S(x, c_i) = \sum_{k=1}^N \sum_{\substack{(f_i, g_i) \\ f_i \in Q, g_i \in D \\ w(f_i) = w(g_i) = k \\ \|T(L(f_i)) - T(L(g_i))\| < \varepsilon}} \frac{idf^2(k)}{tf_Q(k) \cdot tf_D(k)} \quad (5-1)$$

- $S(x, c_i)$  即相似度;
- $k$  代表视觉词汇字典中第  $k$  个视觉单词;
- $N$  是视觉字典的总量,即 K-Means 聚类后生成的种数和;
- $w(f_i) = w(g_i) = k$  代表特征分量  $f_i$  和  $g_i$  均属于第  $k$  个视觉单词;
- $L(f)$  是二维空间中  $f$  的位置,  $T(L(f))$  是  $Q$  在经过变换后的坐标;
- $\|T(L(f_i)) - T(L(g_i))\| < \varepsilon$  代表经过  $T$  变换后  $Q$  和  $D$  中的特征应该很接近;
- $idf^2(k)$  是视觉词汇在数据库中出现次数的倒数,即第  $k$  个视觉词汇的频率;
- $tf_Q(k)$  是第  $k$  个视觉词汇在  $Q$  中出现的次数;

- $tf_D(k)$  是第  $k$  个视觉词汇在  $D$  中出现的次数;

在实验中进行训练的 NBC 采用空间相似度进行测量, 用公式 5-1 进行计算。

### 5.3 实验验证及结果分析

为了验证在本文提出的改良 BoW 模型比经典 BoW 模型在识别上的性能更优越, 本文采用 CAS-PEAL 人脸数据库的部分图像, 这些图像包含不同的变化条件, 实验运行在双核, 主频 2.50GHZ, 内存 4G 的计算机上, 开发平台为 Visual C++。

CAS-PEAL 人脸数据库包含了 1040 个人的 99594 张人脸图片, 其中有 595 名男性和 445 名女性。在实验中分别采用了不同姿势、光照、表情等图像, 其中姿势有 21 种, 15 种光照变化, 3 种眼镜和 3 种帽子的装饰变化, 以及 5 种表情变化。针对不同的情况把实验图像进行分类, 其中数据集是相同的, 包含了男性图像 300 张和女性图像 200 张。

如图 5.2 所示, CAS-PEAL 数据库中人脸有各种不同的形态, 实验中我们将不同的因素取 500 张作为训练集, 测试集选取非训练集的图像进行测试。其中, 表情、姿态、装饰、光照和无变化各 500 张图像, 都不相同。训练集采用相同的男女混合测试集, 随机挑选的人脸图像, 包含各种姿态。如表 5.1 所示。



图 5.2 CAS-PEAL 人脸数据库中不同姿态示例图

我们对选取的训练集用稀疏 SIFT 和 Dense SIFT 分别提取特征, 构造 BoW 模型, 再训练 NBC 和 Adaboost 分类器, 进行测试。利用识别率衡量算法的性能, 不同算法得到的性别识别率如表 5.2 所示, 从表中可以看出: 除了光照变化条件下识别率有明显的下降, 其他变化条件下, 都达到了 90 以上的识别率, 这是 SIFT 特征的特征, 对光照变化敏感, 而在其他的条件下, 识别率都是比其他的分类方法高的, 可见基于改进 BoW 模型和改进 NBC 分类器的性别识别算法最优。

表 5.1 实验数据样本的分布

数据集	训练集	测试集
表情变化	500	500
姿态变化	500	500
装饰变化	500	500
光照变化	500	500
无变化	500	500

表 5.2 SIFT 和 Dense SIFT 在 NBC 和 Adaboost 下的性别识别率

测试人脸集	分类器	特征提取方法	
		SIFT	Dense SIFT
表情变化	Adaboost	90.24%	92.75%
表情变化	NBC	92.55%	93.44%
姿态变化	Adaboost	90.81%	93.13%
姿态变化	NBC	91.64%	93.78%
装饰变化	Adaboost	92.78%	92.16%
装饰变化	NBC	93.45%	92.88%
光照变化	Adaboost	82.66%	91.67%
光照变化	NBC	87.83%	90.53%
无变化	Adaboost	91.32%	93.31%
无变化	NBC	92.71%	94.45%

在进行训练的同时，我们对性别分类的运行时间也进行了统计，如表 5.3 所示，其中的运行时间是整个系统运行 15 次得出的平均时间。从表中可以看出：由于特征提取过程的改良，使得本文提出的模型在运行时间上有了缩短，相比其他的分类方法，都取得了更快的识别速度。

表 5.3 SIFT 和 Dense SIFT 在 NBC 和 Adaboost 下的运算时间

测试人脸集	分类器	特征提取方法	
		SIFT	Dense SIFT
表情变化	Adaboost	1.63s	0.78s
表情变化	NBC	1.17s	0.98s
姿态变化	Adaboost	2.82s	1.59s
姿态变化	NBC	2.29s	1.24s
装饰变化	Adaboost	3.68s	2.65s
装饰变化	NBC	2.81s	1.98s
光照变化	Adaboost	2.91s	1.87s
光照变化	NBC	2.57s	1.64s
无变化	Adaboost	0.95s	0.73s
无变化	NBC	0.78s	0.63s

对每一组实验对象均在电脑上分类多次求平均，如表 5.4 所示，从表中可以看出在经过系统运行了 15 次后，Dense SIFT+NBC 在实验中取得了最好的识别率和最短的平均测试时间。BoW 作为新加入到人脸性别识别领域的新模型取得了不错的效果，与 NBC 的结合上达到了比较好的分类结果，不管是分类准确率还是运行时间与经典的 Adaboost 算法都可以进行抗衡。

表 5.4 识别率及测试时间

方法	识别率	测试时间(s)
SIFT+NBC	90.15%	1.67
Dense SIFT+NBC	92.73%	1.25
SIFT+Adaboost	89.36%	2.19
Dense SIFT+Adaboost	90.06%	1.88

在实验分析中，可以发现，头发作为一种在人脸图像中不可缺少的特征也对性别识别有一定的判断力。我们在后面对头发也进行了特征提取并进行了实验，结果证明头发对性别识别的准确率有所提升，可以将头发这项特征也加入到视觉词典中去。

下图是在 C++ 下运行得到的示例图：



图 5.3 性别判定结果

## 第六章 总结和展望

基于静态人脸图像的性别研究从上世纪 80 年代就开始了，经过这么久的的发展，形成了比较成熟的算法体系。本文主要研究了针对性别识别比较新的一种模型——BoW 模型。

人脸图像的特征提取方法有基于局部和总体的，本文着重分析了基于局部的 SIFT 算法和基于总体特征的 Adaboost 算法。这些算法都是经典算法，到目前为止已经有很多版本，而 BoW 模型是最近几年从文档检索中引进图像识别的，研究并没有太多，本文尝试将经典算法与 BoW 模型结合，并用改进的 Dense SIFT 特征进行提升，希望通过对特征值提速来实现分类器提速。而在上面的实验和实验结果中也看得到，取得了不错的效果。在实验过程中我们发现还有更多的因素可以影响性别识别的准确率，后面的工作可以朝着更多的特征提取方向来做。而 BoW 作为新加入人脸识别领域的一种方法，还有很多可以提升的地方，在本文中实验的图像来自中国人人脸的数据库，希望以后研究 BoW 模型的研究人员可以对 BoW 模型再做深入的研究，使其应用到更广泛的场景中，使其可以更好地为人脸识别服务。





## 参考文献

- [1] Belhumeur P N, Jacobs D W, Kriegman D J, et al. Localizing parts of faces using a consensus of exemplars.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(12):545 - 552.
- [2] Brubaker S C, Wu J, Sun J, et al. On the Design of Cascades of Boosted Ensembles for Face Detection[J]. International Journal of Computer Vision, 2008, 77(13):65-86.
- [3] Triggs B. Efficient object detection using cascades of nearest convex model classifiers[C]// Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2012:3138-3145.
- [4] Dai S, Yang M, Wu Y, et al. Detector Ensemble[C]// 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2007:1-8.
- [5] Felzenszwalb P F, Girshick R B, David M A, et al. Object Detection with Discriminatively Trained Part-Based Models[J]. Pattern Analysis & Machine Intelligence IEEE Transactions on, 2010, 32(9):1627-1645.
- [6] Heisele B, Serre T, Poggio T. A Component-based Framework for Face Detection and Identification[J]. International Journal of Computer Vision, 2007, 74(2):167-181.
- [7] Huang C, Ai H, Li Y, et al. High-Performance Rotation Invariant Multiview Face Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007, 29(4):671-86.
- [8] Jain V, Learned-Miller E. Online Domain Adaptation of a Pre-Trained Cascade of Classifiers[C]// IEEE Conference on Computer Vision & Pattern Recognition. 2011:577-584.
- [9] Kalal Z, Matas J, Mikolajczyk K. Weighted Sampling for Large-Scale Boosting.[J]. Bmvc, 2008.
- [10] Kostinger M, Wohlhart P, Roth P M, et al. Annotated Facial Landmarks in the Wild: A large-scale, real-world database for facial landmark localization[C]// Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on. IEEE, 2011:2144-2151.
- [11] Lampert C H. Detecting objects in large image collections and videos by efficient subimage retrieval.[C]// Proceedings / IEEE International Conference on Computer

- Vision. IEEE International Conference on Computer Vision. 2009:987-994.
- [12] Leibe B, Leonardis A, Schiele B. Combined Object Categorization and Segmentation With An Implicit Shape Model[J]. Eccv Workshop on Statistical Learning in Computer Vision, 2004:17--32.
- [13] Li J, Wang T, Zhang Y. Face detection using SURF cascade[C]// Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on. IEEE, 2011:2183-2190.
- [14] Lin Z, Brandt J. A Local Bag-of-Features Model for Large-Scale Object Retrieval[M]// Computer Vision – ECCV 2010. Springer Berlin Heidelberg, 2010:294-308.
- [15] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.
- [16] Mikolajczyk K, Schmid C, Zisserman A. Human Detection Based on a Probabilistic Assembly of Robust Part Detectors[M]// Computer Vision - ECCV 2004. Springer Berlin Heidelberg, 2004:69-82.
- [17] Muja M. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration.[C]// In Visapp International Conference on Computer Vision Theory & Applications. 2009:331--340.
- [18] Wu Y. Object retrieval and localization with spatially-constrained similarity measure and k-NN re-ranking[C]// Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2012:3013-3020.
- [19] Shen X, Lin Z, Brandt J, et al. Mobile product image search by automatic query object extraction[C]// Proceedings of the 12th European conference on Computer Vision - Volume Part IV. Springer-Verlag, 2012:114-127.
- [20] Sivic J, Zisserman A. Video Google: A Text Retrieval Approach to Object Matching in Videos[C]// null. IEEE Computer Society, 2003:1470.
- [21] Subburaman V B, Marcel S. Fast Bounding Box Estimation based Face Detection[J]. Idiap, 2010.
- [22] Viola B P, Jones M. Rapid object detection using a bossed cascade of simple features[C]// in Computer Vision and Pattern Recognition 2001, CVPR 2001. 2010.
- [23] Wang X, Han T X, Yan S. An HOG-LBP human detector with partial occlusion

- handling[C]// Proceedings / IEEE International Conference on Computer Vision. IEEE International Conference on Computer Vision. 2009:32-39.
- [24] Wu Z, Ke Q, Sun J, et al. Scalable Face Image Retrieval with Identity-Based Quantization and Multi-Reference Re-ranking.[J]. IEEE Transactions on Software Engineering, 2011, 33(10):1991-2001.
- [25] Zhang C, Zhang Z. A Survey of Recent Advances in Face Detection[J]. MSR-TR-2010-66, 2010:1-17.
- [26] Zhu X, Ramanan D. Face detection, pose estimation, and landmark localization in the wild[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2012:2879-2886.
- [27] 黄珍. AdaBoost 算法在人脸检测中的应用研究[D]. 兰州理工大学, 2013.
- [28] 李晓翠. 基于 BoW 模型的人脸识别方法研究[D]. 哈尔滨工程大学, 2014.
- [29] 张晓明. 基于 SIFT 特征的人脸表情识别研究[D]. 南京邮电大学, 2014.
- [30] 黄倩. 基于 BOW 和视觉注意模型的图像分类及其应用研究[D]. 中国科学技术大学, 2015.
- [31] 董彩超. 基于 Haar-like 特征的表情识别和时效性分析[D]. 天津大学, 2012.
- [32] 李长风. 基于 AdaBoost 算法的人脸检测研究[D]. 兰州理工大学, 2014.
- [33] 彭兴媛. 朴素贝叶斯分类改进算法的研究[D]. 重庆大学, 2012.
- [34] 王国才. 朴素贝叶斯分类器的研究与应用[D]. 重庆交通大学, 2010.
- [35] 阿曼. 朴素贝叶斯分类算法的研究与应用[D]. 大连理工大学, 2014.



## 致谢

本论文是在郑海红导师的悉心指导下完成的，从论文的选题到论文的撰写，无不渗透着导师的心血，首先诚挚地感谢郑老师！感谢郑老师在研究生的学习生活中对我的关心和指导，郑老师严谨认真的科研态度一直激励着我努力学习，是我永远的榜样！

感谢所有在学校帮助过我的老师和同学，是你们让我在这里感受到了温暖，让我在研究生的生活中体验到了家的感觉。两年多的时间里，每当我有问题的时候，郑老师总是耐心指导我，帮助我，带着我学会了更多的知识。郑老师在科研项目中一丝不苟的态度，用实际行动教会我们努力的方向。实验室的师兄师姐也给予了我很大的帮助，从刚开始不懂人脸识别领域到后来的熟知，离不开各位师兄师姐耐心的帮助。在这里也要感谢众多先行的研究人员，两年多的时间里我看了很多识别方面的文章和教材，没有这些前辈的努力，我不可能这么快地进入到人脸识别中来，在此表示由衷的感谢！

值此论文完稿之际，谨对郑海红导师的辛勤培育以及谆谆教诲表示最衷心的感谢！



## 作者简介

### 1. 基本情况

张巍，男，河南内乡人，1991年1月出生，西安电子科技大学计算机学院计算机技术专业2013级硕士研究生。

### 2. 教育背景

2007.08～2011.07 郑州大学，本科，专业：信息与计算科学

2013.08～2016.3 西安电子科技大学，硕士研究生，专业：计算机技术

### 3. 攻读硕士学位期间的研究成果

#### 3.1 软件著作权

[1] 郑海红,张辉,张巍. 目标标注软件: 中国,2014SR124150. 2014.08.20.

[2] 郑海红,张巍,张辉. 逆半调图像处理软件: 中国,2014R159551. 2014.10.24