

# 大连理工大学本科毕业设计（论文）

## 视频序列中的人脸检测及跟踪定位实现 Human Face Detection and Tracking in Video Sequences

学 院（系）： 计算机科学与技术学院

专 业： 计算机科学与技术

学 生 姓 名： 郭金锋

学 号： 200881417

指 导 教 师： 徐刚

评 阅 教 师：

完 成 日 期： 2012 年 5 月 25 日

**大连理工大学**

Dalian University of Technology

## 摘 要

随着多媒体技术的发展，人脸图像在各种领域中具有越来越重要的作用，如人机交互系统、视频检测系统、人脸图像数据库管理系统、人脸检测与识别系统等。在众多人脸图像应用系统中，人脸检测与跟踪定位是重要的组成部分。例如在安全监控应用中，安保部门需要对监控录像中的内容进行实时的分析，包括将监控录像中的人脸标记出来并跟踪其运动。这种系统方便了安保部门进行分析，并提高了效率和准确率。在更高级的系统中，更包含人脸识别功能，例如，系统在识别出通缉的嫌疑犯或者恐怖分子后会自动报警。当然这一切功能的第一步就是实现人脸的检测和跟踪。简单地讲，人脸检测就是确定输入图像中在图像中是否存在人脸，并根据人脸的特征将人脸的大致位置用方框标记出来；人脸跟踪就是在输入的视频序列帧中，逐帧将人脸标记出来。当然人脸跟踪的前提应该是人脸检测。人脸检测和跟踪所面临的主要困难是在非理想条件下的检测和跟踪。本文主要研究了现在较流行的一些人脸检测与跟踪方法，并通过基于级联分类器的人脸检测算法进行人脸检测，并采用基于粒子滤波的人脸跟踪算法实现跟踪。

**关键词：**人脸检测；人脸跟踪；机器视觉；主成分分析；增强学习；粒子滤波

## Human Face Detection and Tracking in Video Sequences

### Abstract

With the development of multimedia technology, human face is becoming more and more important in many territories such as human-machine interaction, video detecting system, human face database managing system, human face detection and recognition system. Among most applications in human face territory, human face recognition and tracking play very important parts fundamentally. Take security monitoring applications for example, security department will have to analysis the video in real time, which includes marking and tracking human figure. With the help of such applications, people can easily do the analyzing work with higher efficiency and accuracy. In advanced systems, there will be human face recognition functions. Imagine that when certain terrorists or wanted suspects come into vision, system will automatically tell the security or police of the upcoming danger. The above functions will never be possible until human face detection and tracking is done in advance. Let's put this in a simple way, human face detection is to make sure if there is (are) human face(s) in the input figure, and, if there is (are), mark the area(s) where the system thinks human face lies. Human face tracking is to mark human face in continues video sequences. Apparently, tracking comes after detection. What is difficult in human face detection and tracking is to function in real scenes where condition is not always ideal. In this article, we mainly discuss some popular ways to detect and track human face in video sequences, and make a little system which is able to detect human face in input figure and track human face in video sequences.

**Key Words:** Human face detection, Human face tracking, Computer vision, PCA, AdaBoost, Particle Filter

## 目 录

摘    要 .....	I
Abstract .....	II
1 人脸检测与跟踪概述 .....	1
1.1 人脸检测的基本方法与过程 .....	2
1.2 人脸跟踪的基本方法与过程 .....	4
1.3 技术难点 .....	8
1.4 评价标准 .....	10
1.5 研究现状 .....	13
1.6 本文结构 .....	15
2 视频序列中人脸检测及跟踪定位理论与方法 .....	17
2.1 静态图像中的人脸检测 .....	17
2.2 基于特征的人脸检测方法 .....	18
2.3 基于图像的人脸检测 .....	25
2.4 视频序列中人脸的跟踪定位 .....	28
2.5 图像差分法 .....	28
2.6 光流法 .....	30
3 基于级联的类 Haar 特征人脸检测算法 .....	32
3.1 特征选取 .....	32
3.1.1 矩形特征 .....	32
3.1.2 积分图像 .....	34
3.2 分类器学习 .....	35
3.3 分类器级联 .....	36
3.4 人脸检测结果 .....	38
4 基于粒子滤波的人脸跟踪算法 .....	40
4.1 算法原理 .....	40
4.2 算法流程 .....	41
4.3 运动模型和观测模型 .....	42
4.4 空间更新 .....	43
4.5 跟踪结果 .....	43
结    论 .....	47
参 考 文 献 .....	48

致    谢 .....	50
--------------	----

## 1 人脸检测与跟踪概述

人脸检测 (face detection) 是指在任何人脸识别系统中从输入信息 (图像) 中确定所有存在的人脸的位置、大小、姿态的过程。人脸跟踪 (face tracking) 即在视频流中实时地、连续地估计人脸的位置和方向, 主要用于需要连续探测和检测识别人脸的场所。

在目前的人脸信息处理领域中, 已经将人脸检测和人脸跟踪这两种关键技术配合使用, 作为人脸信息处理的第一个环节, 可以说二者相辅相成既有区别又有联系。人脸检测的目的是在静止得图像中定位出人脸的区域, 而人脸跟踪是根据已定位的人脸在运动图像序列帧的后续帧中跟踪该人脸的运动。人脸检测是人脸跟踪的前提和基础, 而人脸跟踪是人脸检测的后续推广。

Yang 等<sup>[1]</sup>定义人脸检测是指给定任意一幅图像或者一组图像序列, 判定该图像或者图像序列中是否存在人脸, 若存在, 则确定其位置、大小、个数等信息和空间分布等。人脸检测实际上是一个两类人脸的鉴别问题, 即在一幅图像中只有“人脸”和“非人脸”之分。

人脸检测问题起源于人脸识别 (face recognition)。早期的人脸检测问题可以追溯到 20 世纪 70 年代, 经过几十年的发展已经日趋成熟。但是早期的人脸识别研究主要针对那些具有较强约束条件人脸图像 (如无背景的图像), 往往假设人脸位置已知或者很容易获得, 因此人脸检测问题并未受到重视。近几年随着国家公共安全、信息安全、电子商务等应用的发展, 人脸识别成为最有潜力的生物身份验证手段, 所有这些应用背景要求自动人脸识别系统能够对一般环境中的人连图像具有一定的适应能力, 由此所面临的一系列问题使得人脸检测开始作为一个独立的课题受到研究者的重视。人脸检测在旧内容的检索、数字视频处理、视觉监测等方面有着重要的应用价值。

人脸跟踪问题属于目标跟踪问题的范畴。目标跟踪问题是以动态图像处理技术为基础的, 动态图像处理技术是对物体移动时得到的连续图像进行分析和研究的技术。运动目标跟踪的目的是确定运动目标的运动轨迹。运动目标跟踪的实质即是通过传感器拍摄到的图像序列进行分析, 计算出目标在每一帧图像中的位置, 给出目标速度的估计。目标跟踪问题主要解决连续的图像帧之间的基于位置、速度、形状等有关特征的对应匹配问题, 其方法是对分割出的目标的运动轨迹和轮廓的变化进行持续的检测。

## 1.1 人脸检测的基本方法与过程

人脸检测所包含的内容十分广泛，从不同角度可以有很多分类方法，如下表 1.1 所示。

表 1.1 人脸检测方法分类

分类依据		类别	
图像类型	图像来源	静态图像	动态图像
	颜色信息	彩色	灰度
	镜头类型	头肩部图像	半身/全身图像
图像前景	人脸姿态	正面	侧面
	人脸数目	单人	未知
图像背景复杂程度		简单背景	复杂背景

根据图像背景的复杂程度可将人脸检测划分为两种类型：一类是在简单的人脸图像中检测人脸，这些图像包括这类证件上的照片，坐在摄像头前录制的包含人头和肩部的图像，以及用于进行人脸识别研究而专门拍摄的带有很大限制条件的人脸图像，这类人脸定位称为“简单背景人脸检测”；另一类则是在任意的图像中检测人脸，对图像没有任何限制条件，图像的背景、人脸个数、大小、出现的位置等都可以是任意的，称为“复杂背景下的人脸检测”。

根据 Yang 等<sup>[1]</sup>对人脸检测的定义，即根据图像的来源可将人脸检测划分为基于静态图像的检测方法和基于动态图像的检测方法。基于静态图像的检测方法是在静态图像中判断是否存在人脸，若存在，则定位人脸的位置；基于动态图像的检测方法是在视频图像序列中判断是否存在人脸，若存在，则动态的跟踪人脸。显然，后一种所采用的人脸检测方法比前一种要困难。

根据所采用的特征的颜色属性可以将人脸检测方法分为基于肤色特征的方法和基于灰度特征的方法两类。前者适用于构造快速的人脸检测和人脸跟踪算法，后者利用了人脸区别于其他物体的更为本质的特征，是人脸检测领域研究的重点。

根据人脸检测时采用的不用模型，可以将基于灰度特征的方法分为两大类：基于启发式（知识）模型的方法和基于统计模型的方法。基于启发式模型往往先要提取人脸的

特征，比如肤色、纹理、轮廓等，然后利用这些特征结合人们的先验知识进行检测。基于统计模型的方法往往将人脸灰度图像看作高维空间的一个点，然后根据其在特征空间中的分布规律，利用机器学习和概率统计的方法构造人脸检测器。

根据人脸检测时所处理的图像的不同可将人脸检测分为基于灰度图像的人脸检测和基于彩色图像的人脸检测两类。对于灰度图像，目前存在的方法主要分为3类：基于表象(appearance-based)的方法、基于特征(feature-based)的方法和基于模板(template-based)的方法。对于彩色图像，由于其能够提供更丰富的信息，因此可以采用有别于灰度图像的处理方法，利用彩色图像的自身特性进行人脸检测。

人脸图像所包含的模式特征十分丰富，包括：颜色特征（肤色、发色等）、变换域特征（特征脸、小波特征等）、模板特征（均值、方差、距离等）、轮廓特征（椭圆轮廓特征等）、结构特征（对称性、投影特征等）、镶嵌图特征（马赛克规则等）、启发式特性（头发、下颚等）和直方图特征（分布、距离等），如图 1.1<sup>[2]</sup>所示。这些特征中哪些是最有用的，如何利用这些特征，是人脸检测所要研究的关键问题。

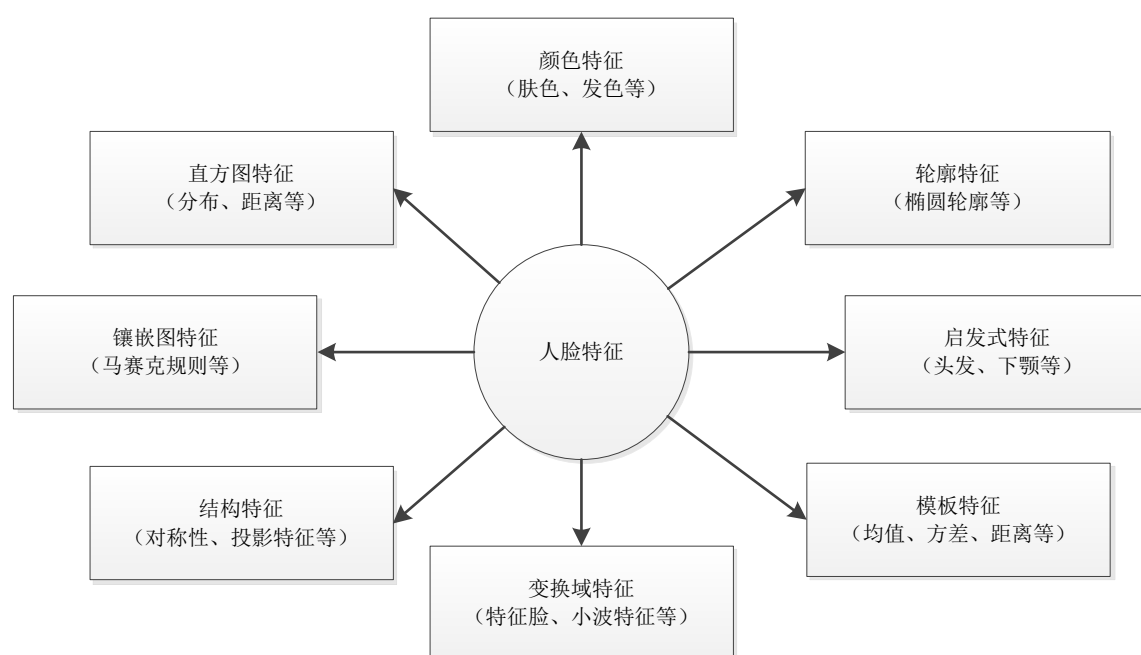


图 1.1 人脸模式的特征



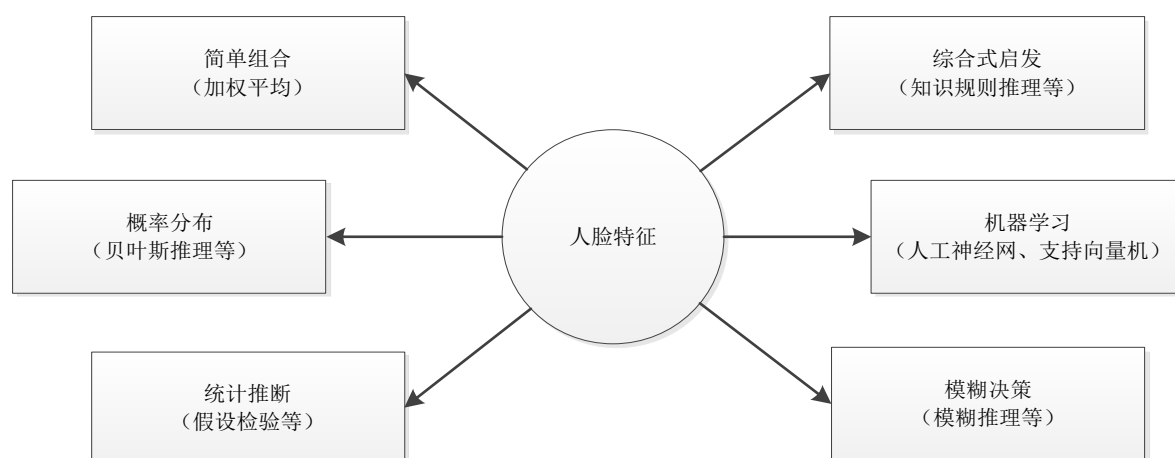


图 1.2 各种特征的综合

从上文可以看出，人脸模式具有复杂且细致的变化，因此，一般利用这些信息时需要采用多种特征综合的方法，包括：简单组合（加权平均）、启发式综合（知识规则推理等）、概率分布（贝叶斯推理等）、机器学习（人工神经网络、支持向量机）、统计推断（假设检验等）和模糊决策（模糊推理等）等，如图 1.2 所示<sup>[2]</sup>。

## 1.2 人脸跟踪的基本方法与过程

### 1.2.1 目标跟踪与人脸跟踪

目标跟踪问题是对物体移动时得到的连续图像进行分析和研究，以确定运动目标的运动轨迹。其具体方法是通过对传感器拍摄到的图像序列进行分析，计算出目标在每帧图像中的位置、速度、形状等信息，并对分割出的目标的运动轨迹和轮廓的变化进行持续的检测。

### 1.2.2 目标跟踪技术分类

目标跟踪技术的分类方法很多，可以通过跟踪环境（包括室内、室外）进行分类，也可以根据跟踪人数（单人、多人、人群）划分。这里从摄像机状态的角度对目标跟踪方法进行分人类，将目标（在此主要考虑人脸和其他身体对象）跟踪可以分为固定摄像机条件下的目标跟踪和非固定摄像机下的目标跟踪两类。

#### （1）固定摄像机条件下的目标跟踪

在固定摄像机条件下进行目标跟踪，目前有两种思路：

其一是将图像序列中的目标跟踪看成是每一帧图像中的人脸检测问题，图像序列中所包含的动态信息仅用来减小当前帧中目标检测的搜索空间，以提高检测速度。常用的方法有基于肤色的方法、基于运动的方法、基于差分的方法等。

其二是将目标跟踪看成是运动估计问题。这类方法更多地利用图像序列中的动态信息，对光照、遮挡有一定的适应性，但是这类方法往往需要第一帧中目标的准确位置，即“锁定”目标。这就需要人机交互方式或者使用人脸检测方法提供初始人脸位置。

## （2）非固定摄像机条件下的目标跟踪

非固定摄像机条件下的目标跟踪就是将摄像机安装在可控云台上，目标的检测与跟踪由同一个摄像头完成。据了解，目前还没有单独研究非固定摄像机条件下的目标跟踪问题。但是许多固定摄像机条件下人脸跟踪方法也可用于非固定摄像机的情况。

非固定摄像机方案从机器视觉的角度来看是十分复杂的，需要解决摄像机的校准、运动目标检测、目标定位、云台控制等问题。非固定摄像机对目标的跟踪必须是在线实时的，这显然与问题的复杂性相矛盾，因此实现非固定摄像机的运动目标跟踪是相当困难的。

### 1.2.3 目标跟踪的常用方法

下面简要总结一下现有的几种主要的目标跟踪方法思想。

#### （1）基于运动的目标跟踪方法（moving-based tracking）

基于运动的目标跟踪方法是利用视频图像序列中，相邻帧之间目标的运动信息来对目标进行跟踪的一种方法。在该领域比较常用的方法包括光流法和差分法。差分法是比较简单的一种。他是通过当前图像和背景图像的差或者相邻两帧之间像素的差来进行目标是否运动的判断，当设置的阈值满足某种条件时就认为目标发生了运动。该方法适应于外界环境（如光照、背景）简单的不变的情况。光流法是在基于运动的目标跟踪方法中最有代表性的一种，也称之为基于梯度的方法，它主要根据图像序列中灰度的变化计算速度场，然后通过聚类分析等方法从速度场中估计目标的结构和运动规律。该方法可以直接从图像序列中提取出运动目标，并进行跟踪，计算也比较精确。但光流法的计算复杂、运算速度太慢，不足以满足实时的要求，当人脸具有较大幅度的旋转或大小变化时，往往会导致跟踪失败；另外受光线的影响比较大，在光线变化较大的视频序列中，误差比较严重。

#### （2）基于模型的目标跟踪方法（model-based tracking）

基于模板的跟踪方法是利用反映目标特征的点、线、区域将跟踪的目标拟合成一个几何模型，从而将运动目标的跟踪问题转化为识别目标问题的方法。对人体进行跟踪时，

通常有三种形式的模型，即线图模型、2D 模型和 3D 模型<sup>[3]</sup>。该方法由于能够充分利用目标的特殊性质，与只能利用目标一般特征的其他方法相比，跟踪的可靠性较强。这种优势在复杂环境下显得尤为突出，比如多物体，帧间的运动幅度比较大的情况。该方法也存在一定的局限性，因为在现实生活中要获得所有运动目标的精确几何模型是非常困难的，如果几何模型的精度过低，就会影响到运动分析的精度，限制了该方法的使用。同时计算量还是比较大的，因为它需要耗费大量的时间来计算复杂的模型，而且需要大量关于目标跟踪的先验知识，实时性较差。

### （3）基于特征的目标跟踪方法（feature-based tracking）

基于特征的目标跟踪方法主要利用特征位置的变化信息，包括特征点的选取和特征匹配两个方面。首先需要在图像序列中抽取各种角度、边界、轮廓灯对应的特征，该特征是对图像边缘进行特征提取，而不是直接与图像灰度相关，可以避免图像序列中光线变化的影响；然后在连续帧中利用一些约束条件例如平滑性约束来寻找特征点的对应关系，对那些特征进行匹配。其主要思想是抽取图像中足以表示目标信息的一组特征序列，与标准序列配准来确定目标机器在图像中的位置。这类方法主要取决于特征提取的好坏，根据需要选择合适的特征来进行跟踪，比较适用于目标纹理比较丰富，运动过程中纹理变化不大的情况。但是如果提取的特征序列过多，会降低效率；如果提取的特征过少，跟踪失效的可能性越大，影响跟踪结果。因此对特征序列的提取决定着该方法的好坏。

### （4）基于相关的目标跟踪方法（correlation-based tracking）

基于相关的跟踪方法的基本思想是在前帧寻找与上一帧中目标区域相关性最大的区域。具体做法就是将系统的基准图像在实时图像上以不同的偏移值位移，然后根据一定的相似性度量准则对每一个偏移值下重叠的基准图像和与基准图像同样大小的实时图像进行相关处理，根据相关判别准则和相关处理结果，判断目标在实时图像中的位置。这种方法在很多应用系统中被采用，一方面是因为它的简单性；另一方面是因为当图像中目标特征不明显时，其他跟踪方法都失效，而只有这种方法还能继续运作。不过，由于它利用图像间的灰度相关性作为区域相似性的判别依据，那么一旦连续帧间光强发生变化，或者目标运动的姿态发生改变，或者噪音较大，或者目标被遮盖物遮盖等对这种方法的影响都很大。该方法一般用于跟踪低速运动的目标，对于高速运动的目标，由于运算量大，它一般满足不了实时跟踪的要求。

### （5）基于变形模板的目标跟踪方法（deformable-based tracking）

变形模板是纹理或边缘可以按照一定限制条件变形的面板或曲线。变形模板有非参数变形模板和参数变形模板两种，参数变形模板又可以分为参数化变形模板和机遇原型的变形模板。在视觉跟踪过程中，常用的变形模板是由 Kass 在 1987 年提出的主动轮廓

模型，又称为 Snake 模型。该算法是一种非参数变形模板，非常适合可变形单目标的跟踪。对于多目标的跟踪更多的是采用基于水平集（Level-set）方法的主动轮廓模型。对于运动速度较大的对象跟踪，可以通过轮廓的运动投影作为下一帧中对象的粗轮廓，然后再使用 Snake 算法得到分割结果的方法。采用基于参数化变形模板的目标跟踪方法相对其他方法比较好，因为我们可以采用利用运动检测的方法减少模板匹配的搜索范围，然后进行相关性匹配，并在跟踪的过程中动态调整优化模板的思想，可以取得较好的效果。

#### 1.2.4 人脸跟踪的基本过程

人脸跟踪的快速算法主要是通过人脸跟踪的帧来实现，人脸跟踪的方法是将视频序列分为检测帧和跟踪帧两种。对于检测帧，利用检测算法确定图像中的人脸位置，作为本帧的结果并为下一帧提供参考；对于跟踪帧，综合利用肤色面部器官结构、前帧参考等信息，快速得到结果，从而实现实时跟踪。

##### （1）检测帧

检测真人脸椭圆区的确定方法如下：①对检测帧进行人脸检测，得到人脸的大致位置；②在被检测道德人脸区域，利用人脸的结构关系和人脸器官的水平投影特点确定嘴区和脸区；③在嘴区，应用数学形态学运算分割出嘴，并利用目标分析方法求取嘴目标的参数 MouthObject；④以嘴为基点，并利用人脸结构的对称性，对人脸结构进行准确定位，得到人脸椭圆区，同时输出人脸椭圆参数 FaceParameter；⑤将上述各参数保存，作为相邻后续跟踪帧的初始参数。

##### （2）跟踪帧

跟踪帧人脸椭圆区的确定方法描述如下：①用相邻前帧数出的嘴目标参数 MouthObject 和人脸参数 FaceParameter 确定当前帧的脸区和嘴区，当前帧嘴区的中心点与相邻前帧的嘴 2 中心点在空间上相对应，嘴区和脸区的尺寸与相邻前帧的相等。②对嘴区进行数学形态学运算，分割出最大的目标区。求取该目标区的表征参数，即周长、面积、长短径之比等，然后将其与相邻前帧的嘴相匹配。如果不匹配，则说明该目标不是嘴，即人脸跟踪失败，将前帧作为检测帧；否则 继续进行下一步。③根据脸区，用人脸结构准确定位方法求取当前帧的人脸椭圆区。计算该椭圆区与相邻前帧人脸椭圆区之间的均方误差，如果该均方误差大于某阈值，则该椭圆区为非人脸区，人脸跟踪失败，将当前帧作为检测帧；否则继续下一步。④输出当前帧的嘴目标参数 MouthObject 和人脸参数 FaceParameter，作为下一帧人脸跟踪的初始参数。

人脸跟踪的整体流程如图 1.3 所示。

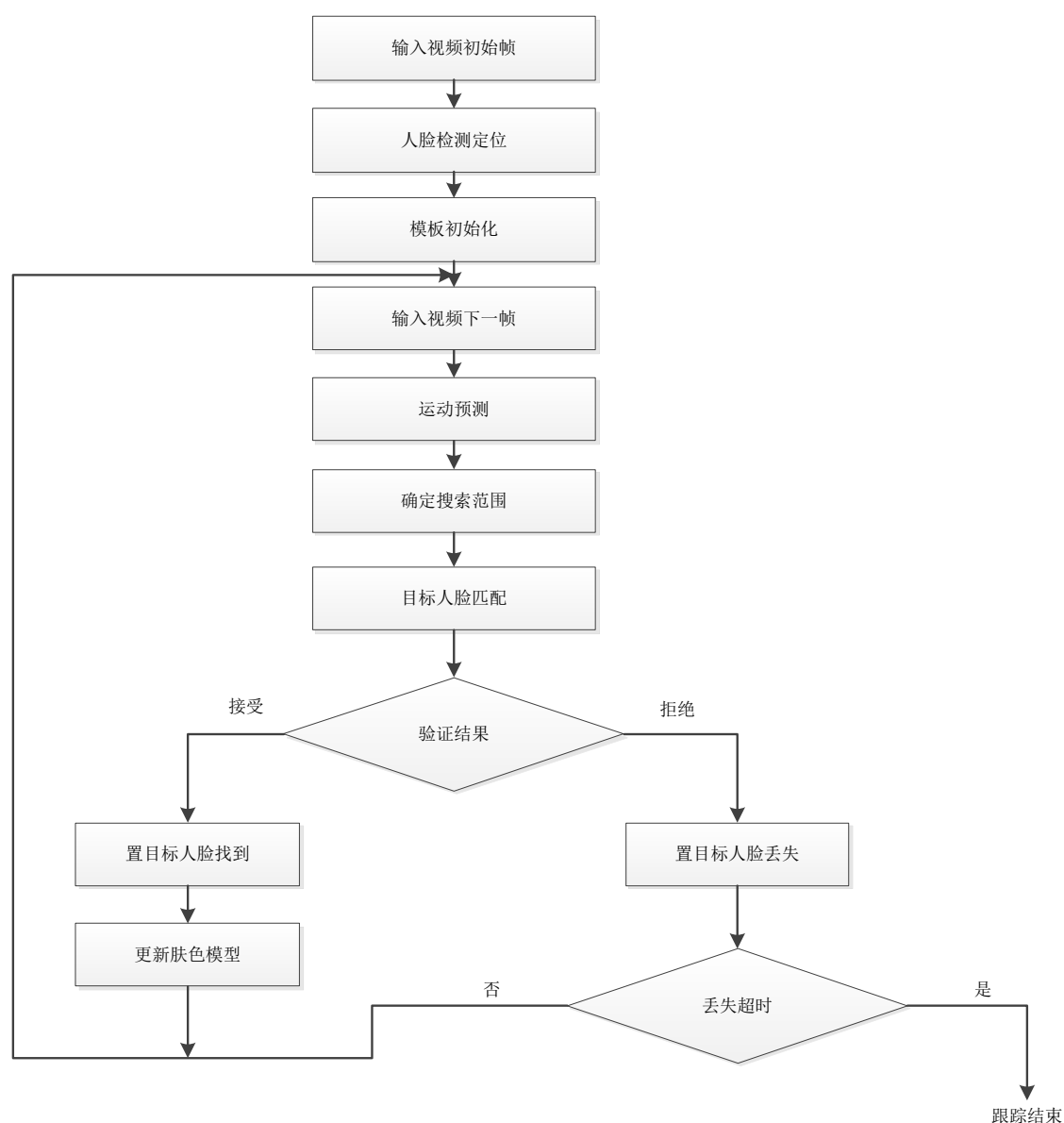


图 1.3 人脸跟踪的整体流程图

### 1.3 技术难点

现有的人脸检测识别系统在用户配合、采集条件比较理想的情况下可以取得令人满意的结果。但是，如果在用户不配合、采集条件不理想（如光照恶劣、有遮挡、图像分辨率低等）的情况下，现有系统的识别率将陡然下降。因此，将人脸识别技术应用到世纪中还存在着诸多挑战性的问题需要解决。

#### （1）复杂条件下的人脸检测和关键点定位问题

人脸检测和关键点定位是一个人脸识别系统的前端处理环节，此环节效果好坏直接影响着人脸识别系统的识别性能。而现有的多数特征定位算法的定位精度都随着光照、姿态等外界条件的变化而快速下降，因此研究复杂条件下人脸的检测和关键点定位算法是目前人脸识别研究中的当务之急。

### （2） 光照变化问题

光照问题是机器视觉存在已久的问题，在人脸识别中的表现尤为明显。尽管研究人员针对光照问题提出了一些解决方案，但是目前的光照处理方法远未达到实用的程度，还需要进一步的研究解决。

### （3） 姿态问题

与光照问题类似，姿态问题也是目前人脸识别研究中需要解决的一个技术难点。针对姿态的研究相对比较少，目前多数的人脸识别算法主要针对正面、准正面人脸图像，当发生俯仰或者左右侧面比较厉害的情况下，人脸识别算法的识别率也会急剧下降。因此，如何提高系统对姿态的鲁棒性也是人脸识别研究中一个极具挑战性的问题。

### （4） 表情问题

面部表情的识别作为生物特征识别的一种，是人机交互中不可缺少的一部分。但是计算机要自动的识别出人的表情却不容易，这是因为：首先，表情是一种很复杂的肌肉运动，每种表情是几十块面部肌肉共同运动的组合，很难用一个准确的数学模型来描述这些肌肉的运动；其次，面部表情的变化主要表现为面部特征点上的运动（如嘴角、眼角的运动），由于计算机视觉技术的限制，计算机不能准确的定位出这些特征点的位置，也就判别不出面部肌肉的运动状态；再次，表情的表现形式也因人而异同一个表情在不同的人面部上有不同的表现形式，而且同一个人的不同表情也没有明显的界限，所以计算机很难用统一的标准来精确的划分各种表情。因此，如何有效地识别带有表情的图像，从而极大地提高人脸识别的准确率，推动人脸识别的发展是一个非常重要的课题。

### （5） 遮挡问题

对于非配合情况下的人脸图像采集，遮挡问题是一个非常严重的问题。特别是在监控环境下，往往被监控的对象都会带着眼镜，帽子等饰物，使得被采集出来的人脸图像有可能不完整，从而影响了后面的特征提取与识别，甚至会导致人脸检测算法失效。如何有效地去除遮挡物的影响是一个非常重要的研究课题。

### （6） 年龄变化问题

随着年龄的变化，面部外观会发生变化，青少年尤其如此。对于不同的年龄段，人脸识别算法的识别率也不同。如何解决年龄变化对人脸识别算法的影响也是许多人脸识别系统必须要面对的一个问题。

### （7） 低质量照片问题

目前多数人脸识别系统在算法设计和模型训练方面往往都只针对图像质量很好的情况，但对监控、公安系统犯罪嫌疑人照片对比等应用而言，由于其来源不一，照片质量也很差（模糊、噪声、低分辨率等）。如何提高系统针对这种低质量照片的识别能力也是人脸识别亟待解决的问题之一。

### （8） 大规模人脸识别问题

随着人脸数据库规模的增长，人脸识别算法的性能将呈现下降趋势，如何维持或者提高大规模应用环境下的人脸识别算法的识别率是一个非常重要的问题。同时，大规模人脸数据库的建立、维护、编码和检索也是非常值得研究的问题。

### （9） 样本缺乏问题

基于统计学习的人脸识别算法是目前人脸识别领域中的主流算法，但是统计学系方法需要大量训练。由于人脸图像在高维度空间中的分布是一个不规则的流形分布，能得到的样本只是对人脸图像空间中的一个极小部分的采样，如何解决小样本下的统计学习问题有待于进一步的研究。

### （10） 海量数据学习问题

传统人脸识别方法在小规模数据中可以很容易的进行学习。但是对于海量数据，这些方法的计算复杂度太高，训练过程难以进行，甚至有可能导致系统崩溃。如何解决海量数据的学习问题是一个非常重要的问题。

### （11） 人脸信息采集设备带来的问题

摄像头是目前主流的人脸信息采集设备，他们的种类、型号、传感方式（CCD/CMOS）、采样和量化精度等各不相同，采集时刻的焦距、光圈、增益、白平衡等参数设置区别很大。这些变化因素使得获取人脸图像变化更大。如何解决这些问题也是值得关注的。

另外，需要注意的是：上述问题并不是单独作用的，很多都是组合型的，例如姿态和光照问题同时出现，会更进一步的增加问题的难度，而且目前出现的主要研究论文，往往都是试图鼓励的解决其中一个问题，而没有考虑到它们同时出现情况下的解决方案，这也是值得关注的一个问题。

## 1.4 评价标准

绝大多数人脸检测算法在设计过程中都需要有一定数量的训练图作为设定参数的依据或统计学系的样本，此外还需要一定数量的测试图库用于评测和改进性能。为了得到

一个公平的人脸检测评估办法，有必要使用一个标准的有代表性的测试集。在评价一个算法的时候，除了要求在相同的数据集上测试之外，还应考虑许多其他因素，如：

(1) 搜索范围。

脸部特征定位的搜索范围可以大致分为全图搜索和给定区域搜索两类，前者的难度显然高于后者。

(2) 定位结果评价标准。

研究者们对“定位准确”有着不同的评价标准，如 Rowley 等使用的标准为：与手工标记位置比较，误差在 4 个像素点以内，求得的尺寸误差不超过 20%。

(3) 应用范围。

不同的应用对算法性能要求的侧重点各不相同。如姿态识别对特征定位的准确率有较高要求，而实时应用则允许牺牲准确率以减小运算量；在用于确认人脸检测结果的时候，必须降低错误接受率，而自底向上的人脸检测算法则要求有较低的错误拒绝率等。

(4) 其他因素。

近几年来出现的多种人脸检测的方法，只有一小部分在同一个数据集上做过测试。Yang 等中对人脸检测算法评估做了一个详细的综述。但是，即使是在相同的数据集合上不同的检测方法仍然无法进行性能优劣的评价。但是，多重因素的影响是的评估标准难以确定，也更加复杂化。第一，各种测试结果都依赖于各自不同的训练集合以及不同的调节参数，这样会直接影响到分类的性能；第二，对于实时系统而言，训练时间和执行时间同样重要，然而训练时间常常会被忽略；第三，不同的检测方法中扫描窗口的数量是不同的，因为要设计来应对不同的环境；第四，大部分算法的检测速率都没有明确的标准；第五，根据不同检测方法的不同目的会产生不同的评估标准。

下面是目前国际上公认的用于人脸自动识别系统性能评价的一些重要定义和指标：

(1) 检测 (detection)

指对人脸图像进行检测和定位的过程。

(2) 拒检 (detection rejection)

指不能正常检测到人脸，人脸不能定位以及人脸检测失败。

(3) 拒检率 (detection rejection rate, DRR)

指被拒检的人脸图像占统计综述的比例，用百分数表示。

(4) 对比 (matching)

指一人脸特征与另一人脸特征进行比较的过程。

(5) 匹配相似度 (similarity)



人脸特征对比的输出结果，代表参与对比的两个人脸特征的相似程度。用 0~1 的小数表示，该数字越大表示参与对比的人脸特征相似程度越大，该数字越小表示参与对比的人脸特征相似程度越小。

(6) 错误拒绝 (false rejection, FR)

指定某匹配相似度为判定阈值，在来自于同一个个体的人脸特征之间的比对，其结果（匹配相似度）小于设定阈值，即指授权人不能被接受的比率。

(7) 错误接受 (false acceptance, FA)

指定某匹配相似度为判定阈值，在来自于不同于一个个体的人脸特征之间进行比对，其结果（匹配相似度）大于设定阈值，即指非授权人判定为授权人的概率。

(8) 错误拒绝率 (false reject rate, FRR)

指发生 FA 的比对此数占总统计比对此数的比例，用百分数表示，也叫做拒真率。

(9) 错误接受率 (false accept rate, FAR)

指在发生 FA 的比对次数占总统计比对此数的比例，用百分数表示，也叫做错误通过率，或叫认假率。

(10) 相等错误率 (equal error rate, EER)

指在某给定匹配相似度下，FAR 与 FRR 相等时的错误率，即  $FAR=FRR$ 。

(11) 登录时间 (enrollment time)

从一幅人脸图像获取后，进行人脸检测、定位和特征提取所花费的时间，此时间是数学统计平均值，用毫秒 (ms) 表示。

(12) 对比时间 (matching time)

比较两个人脸特征所花费的时间，此时间是包含文件读写时间的数学统计均值，用毫秒表示。或者是将一个人脸特征与一定数量的人脸特征进行比对所花费的时间的总和，表示为毫秒/万人。

(13) 首选识别率 (first hit)

匹配相似度最大的人脸是正确的人的比率，即将识别结果按照匹配相似度从小到大排列，在第一位的人脸就是正确的被识别人的比率。

(14) 累计识别率 (first  $n$  hit)

正确的识别结果在前  $N$  个候选人中的比率，即将识别结果按照匹配相似度从小到大排列，在前  $N$  个结果中存在被识别人的比率。

## 1.5 研究现状

目前,针对人脸检测问题,国内外做了很多研究,国外主要是美国、欧洲国家、日本、新加坡、韩国等。著名的研究机构有美国 MIT 的 Media lab、AI lab, CMU 的 Human-Computer Interface Institute、Microsoft Research 等;国内的主要有清华大学、微软亚洲研究院、复旦大学、中科院计算技术研究所和中科院自动化研究所等。随着各方面研究人员对人脸检测研究的不断深入,每年在国际、国内的相关期刊和会议上发表的关于人脸检测的论文数量也在大幅度增长。国际著名的学术杂志,如 *IEEE Transaction on Pattern Recognition and Machine Intelligence*、*Computer Vision and Image Understanding*、*Pattern Recognition* 等,以及著名的学术会议如 IEEE ICFG (International Conference Automatic Face and Gesture Recognition)、ICIP (International Conference on Image Processing)、CVPR (Conference on Computer Vision and Pattern Recognition) 等。

人脸检测的研究大概分为三个阶段,表 1.5 概括地描述了人脸检测识别研究的发展简史与每个历史阶段代表性的研究工作及其技术特点。

第一阶段(1964 年~1990 年)。这一阶段人脸检测通常只是作为一个一般性的模式识别问题被研究,所采用的主要技术方案是基于人脸几何结构特征(geometric feature based)的方法。这几种体现在人们对剪影(profile)的研究上,人们对面部剪影曲线的结构特征提取与分析进行了大量的研究。人工神经网络也一度被研究人员用于人脸识别问题中。总体而言,这一阶段是人脸识别检测研究的初级阶段,重要级的成果不是很多,也没能得到实际应用。

第二阶段(1991 年~1997 年)。这一阶段是人脸检测识别的高潮期,硕果累累。不仅诞生了许多代表性的人脸检测识别算法,而且出现了若干商业化运作的人脸检测识别系统。总体而言,这一阶段的人脸检测识别技术发展非常迅速,所提出的算法在较理想的图像采集、对象配合、中小规模正面人脸数据库上达到了非常好的性能,也因此出现了若干知名的人脸检测识别商业公司。从技术方案上来看,2D 人脸图像线性子空间判别分析、统计表观模型、统计模式检测识别方法是这一阶段内的主流技术。

第三阶段(1998 年~现在)。上世纪 90 年代以来,对人脸检测识别方法的研究变得非常热门,吸引了大量的研究人员和基金支持。由于主流的人脸检测识别技术对光照、姿态等非理想采集条件或者用户不配合造成的变化等条件下的鲁棒性比较差。因此,光照、姿态问题逐渐成为研究热点。与此同时,人脸检测识别的商业系统进一步发展。总体而言,目前非理想丞相条件下(尤其是光照和姿态)、对象不配合、大规模人脸数据库上的人脸检测识别问题逐渐成为研究的热点问题。而非先行建模方法、统计学系理论、基于 Boosting 的学习技术、基于 3D 人脸建模与识别方法等逐渐成为备受重视的技术发

表 1.5 人脸检测识别研究历史发展

阶段	1964 年~1990 年	1991 年~1997 年	1998 年~现在
主要特征	作为一般检测识别问题研究, 基于特征的方法是主流	重点是在较理想的条件下、用户配合、中小规模数据库上的检测识别问题; 基于表观的子空间分析和统计方法是主流	重点是非理想条件下、用户不配合、大规模数据库的检测识别问题; 3D 和非线性是趋势
代表性的人脸检测识别技术与方法及其关键性事件和作品	已知的最早的人脸检测识别研究论文	特征脸 基于特征的方法与基于模板的方法对比	光照锥技术 SVM 用于人脸识别中
	首个半自动人脸检测识别系统	美国 DARPA 启动 FERET 项目 局部特征分析人脸识别方法发展成为 Visionics 公司 FaceIt 商业系统	3D 可变形模型 基于 AdaBoost 的人脸检测技术
	第一篇人脸检测识别方面的博士论文	基于双子空间的贝叶斯概率学习	流形学习 ISOMAP, LIE 两种非线性降维方法
	基于剪影分析的人脸识别	人脸识别的研究综述 Fisher Face	伯朗反射与线性空间分析 基于商图像的识别方法
	人脸的低维表示	弹性图匹配技术	人脸检测综述
		ASM/AAM 主动形状模型/主动外观模型	FRVT2000, 2003, 2006 测试
		基于特征的方法	基于模板的方法
		基于神经网络的识别方法	
		基于表观的 2D 人脸子空间分析与统计学习方法	
		非线性流形分析技术	
技术特点	基于 2D 图像模型的人脸检测识别	基于 3D 图像模型的人脸检测识别	

展趋势。

人脸跟踪在视频监控、人机交互视频会议等各种场合有着广泛的应用。近年来，伴随着视频会议、安全监控技术等的发展，迫切需要对环境中的人员进行跟踪，所以人脸跟踪技术越来越重要和越来越受到人们的重视，人脸跟踪算法的研究受到了研究者的广泛关注，提出了大量卓有成效的算法。但是，人脸跟踪算法是建立在人脸检测正确的结果基础上的，跟踪算法性能的提高，还有赖于检测算法的进一步成熟。

在人脸跟踪领域已经有了大量的相关研究。国外比较著名的研究机构有 MIT, CMU 等<sup>[4]</sup>，国外比较成功的跟踪方法有：Bala 利用视频的前后景分析，并结合彩色信息对整体人脸进行跟踪，同时，使用块匹配对双眼进行快速跟踪的方法；Menser 等则利用 PCA 进行快速定位人脸跟踪；Hager 等建立了整体人脸的低阶参数模型，补偿形变与光照，实时地跟踪人脸，并能容忍部分遮挡；Sobottka 等在 HSV 空间进行肤色分割，并利用脸部特征的对称性定位人脸。最近国外比较成功的算法有 Santana 等采用多模式级联的方式，实现快速人脸跟踪。在普通 PC 上，对 Forman、Miss American 等 11 个标准视频序列达到了 15~20fps 的跟踪速度。Lerdsudwichai 等使用一种基于肤色的均值转移跟踪（Mean Shift Tracking）算法，在普通 PC 上可以实现 12fps 的跟踪速度。

国内在人脸检测识别领域的研究起步较晚。近年来，国内很多高等院校和研究机构，如中科院计算技术研究所、中科院自动化研究所、清华大学、哈尔滨工业大学、浙江大学、南京理工大学、四川大学、吉林大学等，投入了人脸识别相关工作的研究，并取得了很大地进展，理论水平逐步与国际先进水平接轨。

人脸跟踪方面，国内的清华大学、中科院计算技术研究所、上海交通大学等都致力于人脸检测与跟踪相关的研究和应用开发。清华大学的艾海舟和智能技术与系统国家重点实验室做了关于实际环境的测试，采用了差分图像方法对于办公桌前的人像进行了定位跟踪，每次连续运行 5 分钟，共计 1123 帧  $128 \times 128$  的图像或 2160 帧的  $64 \times 64$  的图像，检测结果正确率大于 97%。

## 1.6 本文结构

本文主要讨论了人脸检测与人脸跟踪领域现有的较流行方法，从理论上分析了各种方法，并且在这些方法的基础之上，提出了一种人脸检测与人脸跟踪的方法。

第一章为绪论，提纲挈领的阐述了人脸检测与跟踪的基本方法、基本过程、技术难点、评价标准及研究现状等。第二章具体阐述了人脸检测和跟踪的理论和方法。第三章提出基于级联的类 Haar 特征人脸检测算法，在 MATLAB 中实现该了算法。第四章提出了基于粒子滤波的人脸跟踪算法，并在 MATLAB 中实现了该算法。结论部分对实验结

果进行了分析讨论，并对这种方法的优缺点进行了总结。最后，参考文献部分列出了文中相应内容所参考的主要文献。

## 2 视频序列中人脸检测及跟踪定位理论与方法

### 2.1 静态图像中的人脸检测

根据利用人脸知识的不同，现有的静态人脸检测方法可以分为基于特征（feature-based）的方法和基于图像（image-based）的方法两大类，如图 2.1 所示。前者是与 20 世纪 70 年代，它是先利用人脸的明显特征，如器官特征、几何特征、肤色、纹理等底层特征，然后通过对这些视觉特征之间的距离，角度或者区域的度量，将人脸图像视为一个高维向量，从而将人脸检测问题转化为高维空间中分布信号的检测问题，这种方法中视觉特征是关键，因此这类方法称为基于特征的检测。而后者利用现有的模式识别理论，把人脸检测问题视为一个广义的模式识别问题，这种方法不需要进行脸部特征的检测和分析，而是通过学习算法，利用隐含人脸的只是进行分类，通过训练过程将样本分为“人脸”和“非人脸”两类。这两类方法所采用的研究手段不同，所使用的应用场合也有所区别。

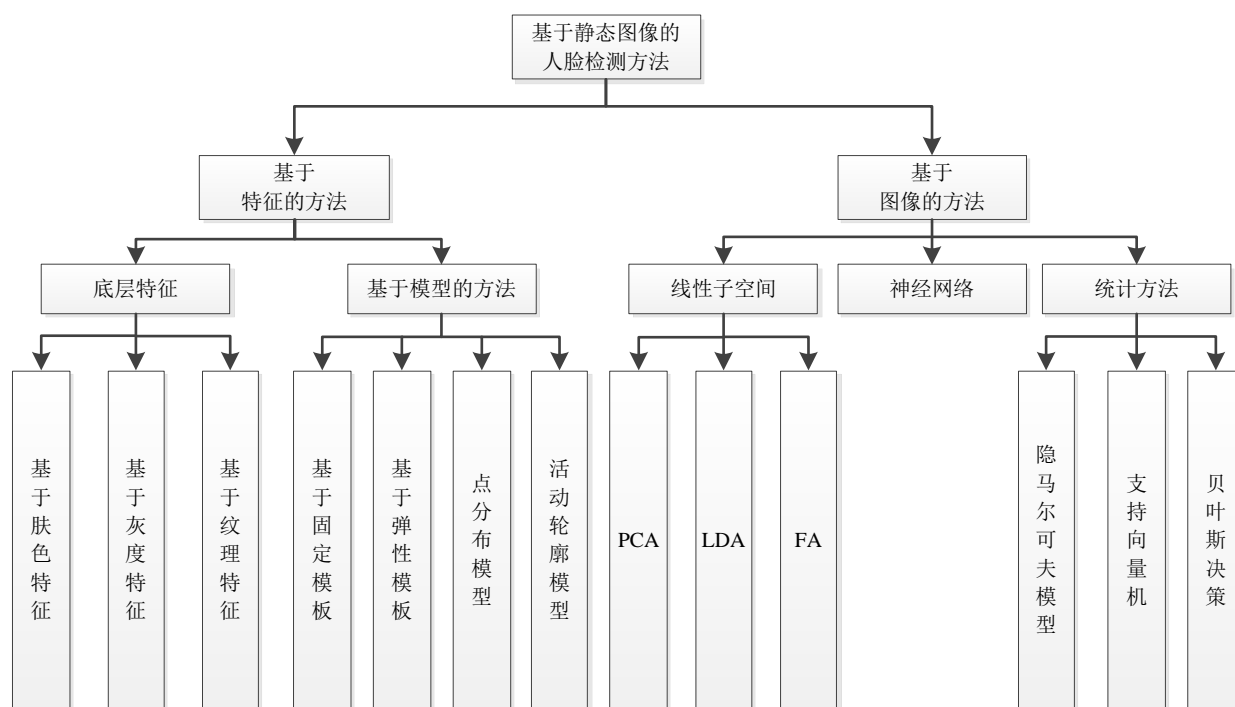


图 2.1 静态图像中人脸检测的方法分类

## 2.2 基于特征的人脸检测方法

根据使用特征类型的不同，基于特征的人脸检测方法主要分为基于底层特征的方法和基于模型的方法两种。

### (1) 基于底层特征的方法

这类方法的基本思想是利用图像的像素特征，如灰度特征（器官特征）、肤色特征、纹理特征等，对视觉信息进行分割。由于本身固有的特点，由底层处理得到的这些基本特征可能比较含糊，不够明确，而在特征分析中可以利用人脸的几何分布特性把这些视觉特征组织成人脸和面部特征的更具全局概念的表达。通过特征分析减少特征的含糊性后，关于人脸面部特征的位置确定问题可以得到解决。

#### ①灰度特征

灰度是人脸的亮度信息。灰度特征主要包括器官特征（人脸轮廓、对称性等）、人脸灰度分布特征（镶嵌图特征、直方图特征等）。

人脸的拓扑位置信息是最早被用来检测人脸的信息。虽然人脸在外观上变化很大，但都遵循一些普遍适用的规则，即五官分布的规则：都有两个眼睛，一个鼻子，一个嘴巴等，并且人脸具有对称性。检测图像中是否有人脸，即是否存在满足这些规则的图像块。这种方法首先是对人脸的器官可能分布的位置，再将这些位置点分别组合，用器官分布的几何关系准则对其进行筛选，从而找到可能存在的人脸。

另外，轮廓是人脸部的重要特征，边缘检测时提取人脸轮廓特征的首要步骤，通常可以采用 Sobel、Canny 等算法来提取图像边缘。由于噪声的存在，利用这些边缘检测算子检测到的图像边缘一般是不连续的，因此需要采用各种边缘连接算法把属于同一轮廓的各段边缘连接起来。多数情况下，人脸的外形轮廓近似于椭圆形，人脸检测可以借助椭圆形的检测来实现。Craw 等<sup>[5]</sup>首先在低分辨率图像中使用一个轮廓模板匹配出人脸的大致范围，然后再高分辨率图像中使用 Sobel 算子获得边缘的位置和方向，然后连接出完整的人脸轮廓。Govindaraju 等<sup>[6]</sup>把人脸抽象为 3 段轮廓线：头顶轮廓线（head contour）、左侧脸轮廓线（left contour）和右侧脸轮廓线（right contour）。对任意一幅图像，首先进行边缘检测，然后对细化后的边缘提取曲线特征，最后利用计算各曲线组合成人脸的评估函数来检测人脸。相对于其他灰度特征，轮廓特征对光照变化具有一定的鲁棒性，但是在强光照的变化会产生一些伪边缘。

除了拓扑信息和轮廓特征外，对称性也是识别人脸可利用的基本性质之一。人脸区域内的各个器官（如双眼、鼻子、嘴角等）是人脸的重要特征，这些器官的分布具有很强的点对称性。利用人脸结构的这种对称性可以有效地检测人脸。为了描述物体的点对称性，Reisfeld<sup>[7]</sup>定义了广义对称变换 GST（generalized symmetry transform）方法，然后

检测局部对称性强的点来进行人脸器官定位，然后考察人言中心点的强对称性和脸部特征的几何分布来检测人脸。此种方法对人脸偏转、表情变化、光照变化等条件不敏感，因而具有很好的鲁棒性，但计算量不大，由于只利用了个点的对称性，易产生大量的候选点，从而使定位的精确度受到影响。

从整体来看，人脸核心区域（眼睛、鼻子、嘴区域）具有独特的灰度分布特征，根据灰度分布特征可以建立若干规则，应用这些规则就可以将检测图像中是否存在人脸的问题转化为检测该图像中是否存在满足这些规则的图像块的问题。Yang 等首先提出了人脸镶嵌图（mosaic image，又称为马赛克图）特征。所谓镶嵌图，是指将原始图像的分辨率有规则地降低后得到的图像。他们首先把在镶嵌图中的人脸的五官区域划分成  $4 \times 4$  的马赛克子块，称为四分图，根据子块的特征初步确定人脸区域，然后进一步把人脸划分为更细的  $8 \times 8$  的八分图。人脸在四分图和八分图中具有一定的规则，依据子块的这些规则对图像从粗分辨率到高分辨率进行筛选，根据满足的这些规则的程度最终确定人脸区域，在此基础上还可以进一步找到人脸各个器官的位置。镶嵌图方法虽然检测率不是很高，但镶嵌图和分级搜索的思想在其他人脸检测方法中得到了广泛的应用。

## ②肤色特征

人脸的肤色是区别于非人脸检测的显著特征，也就是说，肤色特征具有相对的稳定性和大多数非人脸物体的颜色相区别，而且肤色不依赖于人脸面部的细节特征，对于旋转、表情、姿态以及图像的大小并不敏感，故通过肤色分割在一定程度上可以将人脸从大部分背景中区分开来。另外，运用彩色图像处理技术，人们可以很方便地根据肤色特征进行人脸检测。可以说，肤色特征在人脸检测中是最常用的一种特征。

肤色特征主要由肤色模型来描述。所谓肤色模型，指的是用一种代数的（解析的）或查表等形式来表达哪些像素的色彩属于肤色，或者表征出某一像素的色彩与肤色的相似程度。用肤色模型对肤色进行检测主要分为两个阶段：模型的建立于模型的运用。模型的建立主要是通过对大量肤色像素集进行统计分析，然后确定模型中的参数；而模型的应用，主要是通过已建立的肤色模型来判别所输入的像素或区域是否为肤色，或者给出其与肤色的相似程度。常用的肤色模型有高斯模型、混合高斯模型和直方图模型3种。使用任何形式的肤色模型与色度空间的选择有关，这是由于肤色在不同的色度空间中可能聚类的范围有所差异造成的。

高斯模型利用正态分布来拟合颜色的概率密度分布，简单高斯模型是假设肤色分布为单峰高斯分布的一种模型。它主要通过统计分析，预测高斯分布的参数，其中参数确定常用的方法有 EM 算法，Maximum-Likelihood 或通过统计直接求得彩色空间中每个分量的均值和方差。采用这种方法分两步走，首先选择方法确定模型的参数，其次就是利



用该模型来判别新的像素或区域是否为肤色。简单高斯模型考虑到了落入肤色模型范围内像素的点的分布情况，应用概率密度公式判断像素点属于肤色的概率。相对来说，他的肤色检测效率比较高，并且模型的参数也易于计算，但是速度较其他两种模型而言要慢，受肤色空间影响较大，只是对肤色分布的一种函数的拟合，并不是一种真正意义上的肤色分布的反映。

混合高斯模型采用多封的高斯分布来精确描述高斯分布。在这个模型中主要难点是对混合高斯模型的参数估计。其参数的估计常采用 EM 算法或 Elliptical K-均值算法。这种方法相对于高斯模型来说，肤色检测率要高得多，误检率也要小得多，但是模型的确定（模型的参数估计）较难，速度相对来说也比较慢。

利用直方图来描述肤色空间的分布。首先利用直方图对实验集上的肤色及非肤色像素分别进行统计，得到关于颜色在肤色和非肤色中的出现信息，然后在该直方图的基础上，运用贝叶斯决策规则或简单的选择阈值来建造肤色分类器。影响统计模型的主要因素有：训练集合大小、直方图大小。因为统计直方图与颜色空间无关，因而可以直接在 RGB 空间上运行，运算比较简单，并且由该模型建立的肤色分类器的检测速度较快。需要说明的是该模型的理论基础是建立在训练集足够大的基础上的，因此当训练集过小时，其推广能力会下降。

可以从以下两个方面考察色度空间：其一是在该色度空间中能否用给定的模型描述“肤色”区域；其二是色度空间中“肤色”与“非肤色”区域的重叠有多少。目前，在人脸检测中常用的色度空间主要有 RGB、rgb、HIS、HSV 模型、YUV、YCbCr、CIE L\*a\*b\* 等。

利用肤色特征进行人脸检测的一般过程是：首先，利用肤色模型进行肤色分割，缩小搜索范围；然后，对肤色区域进一步利用形状、尺度等集合特征，找出候选人脸区域；最后，在候选人脸区域使用眼睛、眉毛、鼻子等局部特征或者人脸的颜色和形状特征、对称信息等分析验证是否存在人脸。用这种方法进行人脸检测时，把肤色区域分割作为人脸检测的第一步，只要能形成相对完整的肤色区域，一般倾向下都可以大大减少计算量。但受不同光照条件、背景以及人的肤色差异等因素的影响，可能难以形成相对比较完整的肤色区域，这给后续的检测和验证。例如利用肤色特征进行人脸检测的过程中，眼睛区域的检测效果是最容易受光线影响的。尽管可以通过连通的方法来获得较为完整的肤色区域，在比较复杂的图像中，若环境颜色与肤色接近，也不能大幅度地缩小检测区域，肤色检测的优越性同样不能体现出来。

尽管肤色在不同种族，不同人之间，甚至同一个人在不同的光照灯条件下都有改变，但通过采集分析不同性别、不同年龄、不同肤色的人脸图像在颜色空间的分布情况，我

们发现他们的差异主要存在于亮度而不是色彩上。通过颜色的亮度归一化，不同人之间的肤色能得到明显的减弱。为了解决光照问题，Hsu 提出了通过抵消在图像中存在的色彩偏差，即将整个图像中的所有像素的亮度从高到低进行排列，统计 5% 的像素是否足够多（如大于 100），然后将他们的亮度作为“参考白”，即将它们色彩的 R、G、B 分量值都调整为最大值 255。针对不同光照进行补偿，然后在检测图像中的肤色区域，如图 2.2 所示，其中（a）为原始图像；（b）为光照补偿后的图像；（c）为光照补偿前进行的肤色分割的结果；（d）为光照补偿后进行的肤色分割结果。从图中可以看出，光照补偿后更加有利于肤色的分割。这样解决了彩色图像中偏光、背景复杂和多个人脸检测的问题，而且对图像中人脸的色彩、位置、尺度、旋转、姿态和表情等具有不敏感性。



图 2.2 用肤色模型提取肤色特征

目前多数的图像采集设备所使用的是 RGB 色彩空间，而这种色彩空间不利于肤色分割，因为肤色要受到亮度的影响，为了消除光照因素的影响， $YCbCr$  颜色模型具有将亮度分离的优点，聚类特性比较好。

实践证明，基于肤色模型的人脸检测方法应用广泛，对姿态不敏感，在单调背景下检测效率高，但是在光照和复杂背景下效果不理想，因此经常只能作为一种辅助的手段，借以缩小搜索空间，其要求是速度尽可能快，确保在人脸肤色不被过滤掉的基础上，尽可能多的过滤掉非人脸肤色。另外，现有的肤色过滤系统由于采用的是不同的颜色空间，而每个颜色空间又有其适用性，因此，不可避免的会产生很多虚警，如何降低虚警，是肤色过滤系统致力解决的主要问题。

### ③ 纹理特征

根据人脸具有特定的纹理特征这一性质，有人提出了基于纹理的人脸检测方法。主要以 Dai 等提出的基于空间灰度共生矩阵（space gray level dependence matrix, SGLD）的方法为代表，利用空间灰度共生矩阵纹理信息作为特征进行低分辨率的人脸检测，其中人脸的问力模型由一组不等式组成。若某区域中这组不等式都成立，则该区域为检测到的人脸区域。他们同时还将纹理模型与颜色信息相结合来提高人脸检测的精度。由于

纹理特征难以处理多视角问题，这类方法的研究目前尚不多见，也没有和其他方法相比拟的成功事例。

## （2） 基于模型的方法

基于模型的方法一般是：首先认为的建立一个标准感兴趣的目标模型，比较所有可能的待检测区域与目标模型的匹配程度，即计算输入图像与模板的相似度，然后确定一个相似度阈值，用以判断该输入图像中是否包含所需要的目标，从而检测出感兴趣的目标可能所在的位置区域，也就是判断检测区域为人脸。由于该方法是针对感兴趣的目标建立相应的模型的，因此其鲁棒性较好，但是目标模型的建立是相当困难的，同时该方法的计算代价比较大。基于模板的方法现也被广泛应用于人脸检测中，主要有 4 种形式的模型，即固定模板、弹性模板、点分布模型和活动轮廓模型。

### ①固定模板

固定模板匹配时一种经典的模式识别方法，其中心思想是将人脸的五官位置比例关系做成固定模板，用此模板在候选图像中逐点搜索匹配，计算输入图像的各区域与该模板在人脸边界、眼睛、鼻子、嘴巴等的相关程度，若它超过某个预先设定的阈值，则报告检测到人脸。

固定模板的方法实现起来比较简单，在早期的系统中采用的比较多。但是由于人脸特征的变化太大，所以一个固定的人脸模板无法对其充分描述，也不能有效地处理尺度、姿态和形状的变化。因此在实际应用中为了处理多角度和光照变化的模型，检测精度也不高。这种方法目前已不多见，有的系统将其作为粗检测或预处理的手段。另外，固定模板匹配还存在着对不同表情的人脸鲁棒性差的缺点，针对这一情况，人们提出了弹性模板匹配。

### ②弹性模板

弹性模板也称为可变性模板（deformable templates），它是由一个根据被测物体的形状而设定的参数变化的可调模板和与之相应的能量函数所构成，能量函数要根据图像的灰度信息、被测物体轮廓等先验知识来设计。其基本思想是根据待检测人脸特征的先验形状信息，设计一个用于若干参数描述的变参数可调的器官模型，并定义一个能量函数，通过调节模型参数是能量函数最小化，此时的模型参数即作为该器官的几何特征。

当使用弹性模板进行人脸检测时，首先使用参数化的曲线和表面，对人脸和人脸子特征的非刚性元素（如眼睛、鼻子和嘴唇）建立模板，然后采用形态滤波方法，找出原灰度图像的谷值区域、峰值区域和边缘区域的图像，并以此构造一个能量函数，接着，通过全局寻优方法使得模板动态的与图像交互，连续改变其形式和位置，令该能量函数最小化，是模板通过自我变形，逐渐地与人脸特征形状相匹配，最终找到一个最佳的可能拟合，此时的模板参数可以用来描述人脸特征，也就是检测到了一幅人脸。这种方法

的优点在于由于使用的弹性模板可调,能够检测不同的大小、具有不同旋转角度的物体。但是其缺点是检测前必须根据待检测物体的形状来设计弹性模板的轮廓,否则会影响到收敛的结果;当对图像进行全局搜索时,由于要动态的调整参数和计算能量函数,计算时间过大。

Yuille 等<sup>[8]</sup>提出了使用参数化的可变形模板的方法,提取人脸特征,检测眼睛和嘴巴。

总的来说,相比固定模板的检测方法,可变形模板对非刚性模式具有较好的适应性。但是,使用可变形模板匹配方法检测人脸,存在收敛缓慢、处理时间长,其参数选取依赖于经验,性能不够稳定。而且除了这些缺陷之外,基于可变形模板的人脸检测的成功与否,还在很大程度上依赖于模板初始位置的选择、定位。

### ③点分布模型 PDM

点分布模型(point distribution model, PDM)是 Cootes 等<sup>[9]</sup>于 1992 年提出的一种参数化的形状描述模型,它运用一组离散的控制点来描述对象的形状,并用 PCA 方法建立起各个控制点的运动模型,对控制点的基准位置和移动模式做了约束,从而保证模型的整体形变始终在可接受的范围内。之后,他还提出了主动形状模型(active shape model, ASM)的概念。他对形状和灰度表象建模,用建立的 ASM 在新的图像中定位易变的物体。根据样本统计特性建立起来的先验模型,反映了待分割轮廓的形状特征。通过对其参数的调整,既可以灵活地改变形状以适应目标形状不确定的特性,又将形状的变化限制在模型允许的范围内,从而保证模型改变时不会受到各种因素影响而出现不合理的形状。

考虑到人脸对象的特殊性,可以使用点分布模型 PDM 来刻画人脸对象的形状知识,将其用于静态图像中的人脸特征点定位。Lanitis 等<sup>[10]</sup>首先将这一方法应用于脸部特征定位及人脸识别,用于描述人脸的整体外观,包括所有的脸部特征,如眼睛、鼻子、嘴等。在使用 ASM 找出人脸的形状后,将人脸切割并归一到统一的框架,然后对这个与形状无关的人脸采用亮度模型来进行解释和识别。他们建立起一个有 152 个控制点组成的人脸模型,并对各个控制点附近的与形状无关的据不会读信息分别建立了 PCA 模型。搜索时,将基准模型放在目标图像附近,首先根据每个控制点当前位置附近的局部灰度信息计算出一个目标位置。与 Snakes 方法相比,此类方法在模型中引入了与目标特征有关的信息,从而降低了对噪声和局部变形的敏感度,但运算量仍然很大。

这类方法还有主动外观模型(active appearance model, AAM)<sup>[11]</sup>等,就人脸模型而言,ASM 建立了特征点局部的纹理模型和全局的形状模型,而 AAM 则建立了全局的纹理模型和形状模型。AAM 是 ASM 的进一步扩展是关于形状的一种统计模型,通过对 PDM 的主元分析,从训练数据集中得到关于形状变化的参量,将先验经验赋予 AAM。

对于一个新图像中的物体，AAM 能够通过迭代形变对其进行匹配，而该形变是受训练集合中所标记形状约束的。AAM 在人脸识别、计算机自动识别嘴唇形变等方面也都得到了较好的应用。

#### ④活动轮廓模型（Snake 模型）

1987 年由 Kass 等<sup>[12]</sup>提出了称为 Snakes 的活动轮廓模型（active contour models），它是一种边界检测和图像分割的方法，也可以用于脸部特征的定位。Snakes 模型是一种能量函数最小化的变形轮廓线，这些 Snakes 模型把图像中感兴趣物体轮廓看作一条连续、封闭的链条结构，为它设计了一类能量函数，以迭代方式求取能量最小值，从而获得最优轮廓。Snakes 模型锁定在图像特征附近，准确地将他们最小化。它是用来进行边界检测和图像分割的最常用的一种方法，也可以用于人脸部的定位。Snakes 模型是一条由  $n$  个控制点组成的连续闭合曲线，它用一个能量函数作为匹配度的评价函数，定义为

$$E_{snakes} = E_{internal} + E_{external} \quad (2.1)$$

式中  $E_{internal}$  和  $E_{external}$  分别为内部和外部能量函数，前者定义了模型的内部属性和内在运动趋势。定位前，首先将模型定位在目标对象预估计位置的周围，不断迭代使能量函数最小化，当内外能量达到平衡的时候即得到目标对象的边界。这种方法可以避免二值化方法时已发生的边缘断线问题，但缺点是运算量很大，容易受到初始位置的影响，对非凸体的检测效果不佳，而且由于模型本身不包含目标对象信息，因而当存在局部形变、噪声和遮挡等干扰时容易出错。针对这些问题，Wu 等<sup>[13]</sup>提出在搜索前，先根据一系列先验规则利用灰度和彩色信息找出脸部特征的大概位置，作为设定初始位置的基准。为解决非凸体的检测问题，他们根据各个脸部特征的形状特点，使用多个 Snakes 模型拟合出一个脸部特征，如在眼、眉、嘴的上下方和鼻的左、右、下 3 个方向各放置一个 Snakes 模型并使之各自收敛，直至相交。此外，他们还在能量函数中加入了边界、肤色相似性和红色相似性 3 种信息，以减少成像条件、表情及个体差异所带来的影响。在无遮挡、图像质量较好的情况下，他们的算法可以取得较好的效果，且对目标对象尺寸不敏感。

值得注意的是，Snakes 模型并不是解决轮廓提取所有问题的安全性模型，在使用它之前，必须有一条初始轮廓。

总之，基于特征的人脸检测方法是利用先验知识考查人脸所包含的特征以及分布来判断待检测区域是否为人脸。其优点是快速，但其缺点是在复杂情况下不稳定，最大困难在于寻求相对较稳定的特征，因为图像的显式特征容易受到光照、噪声以及各种遮挡的影响，所以需要将人脸检测转化为不变特征量的寻找。一旦能找到符合要求的特征，

这类方法将会有突破性的成功和应用。但一般来讲，基于特征的方法不易检测图像中相对较小的人脸。

## 2.3 基于图像的人脸检测

由于人脸图像的复杂性，显式地描述人脸特征有一定困难，因此另一类方法——胡同游图像的方法越来越受到重视。与基于特征的检测不同的是，基于图像的方法无需检测各个面部特征。基于图像的人脸检测方法主要分为基于线性子空间的方法、基于人工神经网络的方法和基于统计的方法。

### (1) 基于线性子空间的方法

子空间分析的思想就是根据一定的性能目标来寻找一个线性或非线性的空间变换，把原始信号数据压缩到一个低维子空间，使数据在子空间中的分布更加紧凑，为更好地描述数据提供了手段，计算复杂度也大大降低。人脸图像是位于所有图像空间中的一个子空间，为了表示这个子空间，目前在人脸检测中常用的线性子空间方法主要包括主成分分析方法、线性判别分析（linear discriminant analysis, LDA）方法和因子分析（factor analysis, FA）方法等。

①主成分分析 PCA。主成分分析是模式识别中常用的经典方法，Pentland 等<sup>[14]</sup>首先将 PCA 引入到人脸检测。该方法根据图像的统计特征进行正交变换（K-L 变换）以消除原本各分量之间的相关性，去变换后所得到的最大的若干个特征向量来表示原来的图像保留了原图像差异的最大信息，这若干特征向量就称为主成分。换句话说，就是将表征人脸的高维向量映射到由若干个特征向量张成的子空间中，从而实现对人脸图像的最优解析和重构。在主成分分析 PCA 方法中，具有最大特征的若干特征向量被提取出来构成一个子空间，这些特征向量作为子空间的基，它们保留了元数据中主要的信息而被忽略了次要的部分。例如，从数据库中提取大量的人脸模式，经预处理后得到人脸的训练样本集，在这个人脸样本集中做主成分分析，得到一组特征向量，成为特征脸。设  $\Sigma = \{\xi_1, \xi_2, \dots, \xi_n\}$  为一组按特征值从大到小排序的特征脸，它构成了人脸特征空间的基， $f = f(x, y)$  是一幅人脸图像。被测区域通过预处理后，可以表示为特征脸的线性表示。即

$$\tilde{f} = x_1 \cdot \xi_1 + x_2 \cdot \xi_2 + \dots + x_n \cdot \xi_n \quad (2.2)$$

在进行人脸检测的过程中，被检测的图像区域根据上公式得到人脸的子空间表示  $\tilde{f}$ ，由此可以计算与原图像的信噪比，即

$$\beta = 10 \log \frac{f}{f - \tilde{f}} \quad (2.3)$$

根据其是否大于某一阈值，可以判断被检测部分是否包含人脸。

另一方面，在人脸识别中利用的是主元子空间（特征脸），而人脸检测中作为分类器利用的是次元子空间（特征脸空间的补空间），用待检测区域在次元子空间上的投影能量，也即待检测区域到特征子空间的距离作为检测统计量，距离越小，表明越像人脸。Pentland 等还发现人脸在特征脸空间的投影聚焦比较紧密，因此利用若干张特征脸前将人脸向量投影到主元子空间  $F$  和其正交的补空间  $\bar{F}$  中，相应的距离度量分别称为 DIFS（distance in feature space）和 DFFS（distance from feature space）。对于人脸检测问题，由于没有考虑“非人脸”样本的分布，需要同时使用 DIFS 和 DFFS 才能取得较好的效果。

②线性判别分析 LDA。后来发展的 LDA 方法<sup>[15]</sup>不同于主成分分析，它能够以使样本的可分性最好为目标，寻找一族最优的线性变换使每类的类内离散度最小，并且使各类间的离散度达到最大，因此从理论上来说比较适用于模式识别问题。经典的线性判别分析中使用的是 Fisher 准则函数，所以线性判别分析又称为 Fisher 线性判别分析（fisher linear discriminant analysis, FLDA）。LDA 方法同样通过某一非线性变换把人脸图像区域属性空间的属性向量映射到高维的特征空间，计算被测图像区域与该特征空间的距离是否小于某阈值来判断其是否包含人脸。但由于它利用了类别归属信息，选择与类内散布正交的矢量作为特征脸空间，因此能压制图像之间的鱼识别信息无关的差异，强调了不同人脸之间的差异，同时弱化了同一人脸由于光照、视角和表情变化而引起的变化，获得了比特征脸更好的识别效果。从而在一定程度上克服了 PCA 方法的某些缺陷，对光照及人脸表情等具有一定的鲁棒性。

③因子分析 FA。因子分析方法<sup>[16]</sup>是有心理学发展而来的一种线性子空间方法，类似于 PCA 方法。与之不同的是 FA 虽然也是从原变量中提取信息，以使多元问题降维，但它是从一定的模型出发，找出几个反应原有变量的公共因子，并力求使它们有较为合理的专业解释。也就是说，PCA 方法要使用到每个项目的所有变异，而 FA 方法可以克服 PCA 方法对噪声的敏感性。

## （2）基于人工神经网络的方法

人工神经网络（artificial neural network, ANN）<sup>[17]</sup>是一种有效的模式识别方法，它利用神经网络的学习能力，对“人脸”样本集和合“非人脸”样本集合进行学习，以产生分类器，从而达到人脸检测的目的。该方法是把模式的统计特性隐含在 ANN 的结构和参数中，对于人脸这类复杂的、难以显示描述的模式，基于 ANN 的方法具有独特的优势。

用神经网络进行人脸检测的好处是很容易训练一个用于检测人脸模式的系统。但是缺点是网络结构需要大范围的调整（层数、结点数、学习速率等）才能获得期望的性能。

## （3）基于统计的方法

除线性子空间方法和神经网络方法外，还有一些其他的统计方法用于人脸检测，基于统计的人脸检测方法不是针对人脸的某一特征，而是从整个人脸的角度出发，利用统计学原理，将人脸图像是为一个高维空间中分布信号的检测问题，从成千上万张人脸图像中提取出人脸共有的一些规律，利用这些规律来进行人脸检测。属于这类方法的有隐马尔可夫模型 HMM、支持向量机 SVM、贝叶斯决策和基于 AdaBoost 的方法等。

表 2.3 基于统计方法的检测结果统计

方法	样本规模	Sung 测试集上的结果 (图像: 23, 人脸: 149)	Rowley 测试集上的结果 (图像: 130, 人脸: 507)
基于事例学习 (Sung 等)	人脸样本 4150 个 非人脸样本 6189 个	79.9% (5 个错误报警)	N/A
基于线性子空间 (Yang 等)	人脸样本 16810 个 非人脸样本 8422 个	FA: 89.4% (3 个错误报警) FLD: 91.5% (1 个错误报警)	FA: 92.3% (82 个错误报警) FLD: 93.6% (74 个错误报警)
基于人工神经网络 (Rowley 等)	人脸样本 15750 个 非人脸样本 8000 个	84.5% (8 个错误报警)	86.2% (23 个错误报警)
基于概率模型 (Schneiderman 等)	人脸样本 118920 个 非人脸样本 1552 个	89.0% (3 个错误报警)	90.5% (33 个错误报警)
基于支持向量机 (Osuna 等)	N/A	74.2% (20 个错误报警)	N/A

梁路宏等<sup>[2]</sup>总结了基于统计方法的一些检测结果的比较。基于统计模型的方法大多适用于复杂背景图像中的人脸检测。目前基于统计模型的方法主要针对正面端正人脸的检测。评估不同方法的性能需要共同的测试集，目前比较公认的测试集是 Sung 等的 MIT 测试集和 Rowley 等的 CMU 测试集（含盖了 Sung 等的测试集）。如表 2.3 所示列出了几个具有代表性的算法在该测试集上的检测结果统计，由于基于统计的方法对训练样本具有较强的依赖性，因此表中同时给出了各种方法使用的样本规模。



基于统计模板的方法是目前比较流行的方法，这种方法大多是由于复杂背景图像中的实时人脸检测，是解决复杂的人脸检测问题的有效途径。首先，它不依赖于人脸的先验知识和参数模型，可以避免不精确或不完整的知识造成的错误；其次，它采用实例学习的方法获取模型的参数，统计意义上更为可靠；最后，通过增加学习的实例可以扩展检测模式的范围、提高检测系统的鲁棒性。但是由于其前期训练时间较长，故有待改进。

## 2.4 视频序列中人脸的跟踪定位

在视频监控系统或远程电视会议的实际应用中，从动态的视频序列图像中检测人脸是一项重要工作。利用运动信息检测人脸的难度在于，在一般的图像序列中，往往会有很多相互独立的运动着的物体，在这种情况下，通过运动信息不但不能达到减小搜索空间的目的，而且由于大量不正确的候选区域的引入，实际上会错误的引导系统，降低检测的可靠性。然而对大多数物体来说，在视频中的运动具有连续性，验证运动的连续性对我们检测人脸有着积极地作用。一般来说，基于序列图像的人脸检测方法有图像差分法和光流法等。

在序列图像中，通常人相对于背景来说总是运动的，因此利用运动信息就可简单有效地将人脸从背景中分割出来。图像序列的相邻帧间变化是运动目标检测和识别的基础，通过限定图像采集的背景，利用图像差分法就可以有效地检测出人脸<sup>[18]</sup>。

视频中的人脸检测具有与静态图像人脸检测不同的特点：

(1) 视频其实是由数量很多的序列静态图像组成，一段几十分钟的视频多达数十万帧图像，因此检测速度称为不得不考虑的目标。

(2) 视频提供了更多的信息，比如颜色、运动信息等，怎样合理利用这些信息。

(3) 视频中还设计对象跟踪的概念，怎样去跟踪检测到的物体

## 2.5 图像差分法

当系统输入的人脸图像为动态图像序列时，则可以利用人脸或人脸的器官相对于背景的运动来检测人脸，比如利用眨眼或说话的方法来实现人脸与背景的分离。因此，在运动目标的检测中，图像差分 (image difference) 方法是最简单的检测运动人脸的方法。由于图像序列的计算远比静态图像的计算复杂、耗时，所以随着计算机的高速发展和视频监控等应用的需要，在近几年基于动态图像序列的人脸检测方法才逐渐成为一个热点的研究话题。

图像差分运算定义为：

$$f_d(x, y, t_1, t_2) = f(x, y, t_1) - f(x, y, t_2) \quad (2.4)$$

式中,  $(x, y)$  是图像中像素的坐标,  $f(x, y, t)$  是  $(x, y)$  在  $t$  时刻的强度。由于式中只涉及对应像素强度的相减运算, 因此算法简单, 运算量小。

对图像的差分结果进行阈值处理, 可以得到一个新的输出:

$$f_{out}(x, y, t_1, t_2) = \begin{cases} 1, & f_d(x, y, t_1, t_2) \geq T \\ 0, & f_d(x, y, t_1, t_2) < T \end{cases} \quad (2.5)$$

这里,  $T$  作为一个适当的阈值。阈值处理后输出的图像是一个二值图像。图像中为 0 的像素对应在前后两时刻没有发生变化的地方, 图像中为 1 的像素对应两时刻发生变化的地方。

此时的二值图像给出了图像中的运动信息, 通过对二值图像的分析, 如将二值图像进行水平和垂直投影, 便可以得到图像中目标活动区域。图像差分是检测相邻两帧图像序列之间变化的最简单方法, 一般有两种情况:

#### (1) 当前图像与固定背景之间的差分 (减背景法)

减背景法是常用的运动目标检测方法, 其基本思想是将当前每一帧图像与事先存储或者实时得到的背景图像相减, 若像素差值大于某一个阈值, 则判定此像素是出现在运动目标上的, 且相减操作后得到的结果直接给出了目标的位置、大小、形状等信息。对于这种将同一背景不同时刻两幅图像进行比较, 即用序列中的每一帧与一个固定的静止的参考帧 (不存在任何运动物体) 进行图像差运算。然而自然景物环境永远不会很静止 (例如, 风吹动树枝和树叶、太阳位置改变导致阴影的变化), 因此该方法抑制噪声能力较差。

使用减背景法进行运动目标检测通常会遇到如下一些问题:

① 背景的获取。背景图像的获取最简单的方法是在场景没有运动目标的情况下进行, 但实际应用中无法满足这种要求, 如对高速公路和城市交通的监控。

② 背景的扰动。背景中可以含有轻微扰动的对象, 如树枝、树叶的摇晃, 扰动部分不应被看作是前景运动的目标。

③ 外界光照条件的变化。一天中不同时间段光线、天气等的变化对检测结果的影响。

④ 背景中固定对象的移动。背景里的固定对象可能移动, 如场景中的一辆车开走、一把椅子的挪开, 对象移动后的区域在一定时间内可能被误认为是运动目标, 但不应该永远被看作是前景运动的目标。

⑤ 背景的更新。背景中固定对象的移动和外界光照条件的变化会使背景图像发生变化, 需要及时对背景模型进行更新。

⑥ 阴影的影响。通常前景目标的阴影也被检测为运动目标的一部分，这样将会影响对运动目标的进一步处理和分析。

所以说，这种目标检测方法的优点是计算简单、易于实时，位置准确，只需要获取当前的一幅图像，但它要求背景绝对静止或基本无变化（噪声较小），不适用于设想头运动或者背景灰度变化很大的情况，因而适用场合有限，另外其不足之处还在于受环境光线变化的影响较大，在受非空环境下需要加入背景图像更新机制。

## （2） 当前连续两幅图像（时间间隔 $\Delta t$ ）之间的差分（相邻帧差分法）

运动目标检测的目的是在目标图像提取出运动区域，为了有效地提取运动区域，一般情况下，需要实时更新背景，即把当前的背景图像数据和目标图像数据融合，更新下一帧背景图像的数据，以得到自适应背景，从而获取差分图像。然后二值化该灰度差分图像来提取运动信息。

相邻真差分法非常适合于动态变化的环境，但相邻帧差分法不能够完整地分割运动对象，不利于进一步的对象分析和识别。这种运动目标检测方法的优点是它只对运动物体敏感，实际上只检测相对运动的物体，同时因两幅图像的时间间隔较短，差分图像受光线变化影响较小，检测有效而稳定；缺点是检测出的物体的位置不精确，其外接矩形在运动方向上被拉伸，这实际上是由相对运动与物体的位置并非完全一致引起的。相对运动受物体本身的运动速度及相邻图像之间的时间间隔影响。因此这种方法常常会引起两种不良后果：一种是两帧之间物体重叠部分没有检测出来，即只检测出物体的一部分；另一种是检测出物体在两帧中的信息，即检测出的目标比真实的物体大很多。

利用差分方法进行运动目标检测，在外界环境较好时，如运动速度不快、无阴影等干扰，运动目标检测效果很好；而对于存在外界干扰情况，检测运动的区域将大于目标的区域，或者可能将阴影检测为运动目标。总的来说，图像差分法极易受噪声污染，光照和蛇降级位置的改变以及摄像机的电子噪声，都会产生很多错误数据，因此差分图像适用于对图像进行全据粗略处理以便解释的注意力引向场景中出现“活动”的区域。而且当目标手遮挡或背景光照变化以及有多个运动目标时，这种方法就会失效。这时可以考虑用光流法（optical flow）或基于光流场的不连续性等方法。

## 2.6 光流法

光流反映了帧间运动的重要信息，在表情的动态分析中有较多的应用。基于光流方法的缺点在于，光照不均和脸部非刚性运动等因素将会影响特征提取的结果。

光流法是运动图像分析的重要方法，光流是图像中亮度模式的显示速度分布，是由于目标与观察者之间的相对运动引起的。在理想条件下，光流对应于运动场，光流场的分析广泛应用于运动场，光流场的分析广泛应用于估计观察者的自主运动、运动目标的检测与分割、图像稳定和深度重建等，此类方法的“瓶颈”在于光流得可靠计算。

当物体运动时，在图像上对应物体的亮度模式也在运动，从而称光流是图像亮度模式的视在运动。在光流场中不同的物体会有不同的速度，大面积背景的运动会在图像上产生较为均匀的速度矢量区域，这为具有不同速度的其他运动物体的检测带来了方便。

光流法检测采用了目标随时间变化的光流特性，Barron 使用光流法检测前方的车辆，使用光流法分析，同样可以检测后方的车辆，由于被检测车辆与背景的主运动方向完全相反，Bigun 等在这种情况下获得了较好的检测效果。利用光流法可以通过计算位移向量光流场来初始化目标的轮廓，从而使基于轮廓的跟踪算法有效地检测和跟踪目标。Bigun 等利用特征光流进行运动目标的分割和跟踪，存在计算出的特征光流精确度不够而导致跟踪不稳定，并且算法对硬件系统有较高的依赖性。

值得注意的是，有些光流计算结果只在物体边缘与背景有较明显的速度差，而在比较均匀的物体内部则差别不大。因而，简单的速度分割后，还需要一定的后续处理工作才能完整的检测出运动的物体。

光流法的优点是能够检测独立运动的对象，不需要预先知道场景的任何信息，并且可用于摄像机运动的情况，对目标在帧间的运动的限制较少，可以处理大的帧间位移。但多数光流法很耗时、计算复杂，且抗噪声性差，除非有特殊的硬件支持，并且很难实现实时的检测。也就是说，光流法的时间开销很大，其实时性和实用性较差。

### 3 基于级联的类 Haar 特征人脸检测算法

在人脸检测的算法中，最经典且最有效的当属 Viola 和 Jones 提出的基于级联分类器的人脸检测算法<sup>[19,20]</sup>。此方法不仅速度快，而且效率极高，因此本文的人脸检测部分采用了该算法。

该算法框架主要由三部分组成。第一是特征提取，该算法没有直接使用图像的强度信息，而是采用 Haar 特征。为了快速计算任意大小的图像的 Haar 特征，引入了积分图像的概念来表示图像。由于积分图像的计算只需几个像素点的值，因此任意大小或位置的 Haar 特征的计算也极其快捷。第二是利用 Adaboost 分类器，从一个很大的特征库中，挑选出小部分更具判决性的特征，简单而有效。对于任一图像子窗口，Haar 特征的数目非常大，甚至多于像素的个数。为了确保分类速率，学习的过程必须排除大量的无用特征，只专注于那些决定性的特征。每次学习都选择一个新的弱分类器，也可看作是选择了一个新的特征。第三是级联多个强分类器，只计算脸部可能出现的区域，忽视背景区域，再一次提高了检测的速度。下面将对这三部分分别详细介绍。

#### 3.1 特征选取

##### 3.1.1 矩形特征

选取特征而不是直接利用像素值来表示图像有很多优点，如特征可以更好的表示图像的纹理信息，梯度信息，颜色信息等，会在一个更高维的空间内更好的描述图像的本质。而对于该算法，利用特征会使得效率更高。

Viola 等人使用了简单的 Haar 特征<sup>[21]</sup>，其中矩形区域必须是水平或垂直相邻的，如图 3.1 所示。

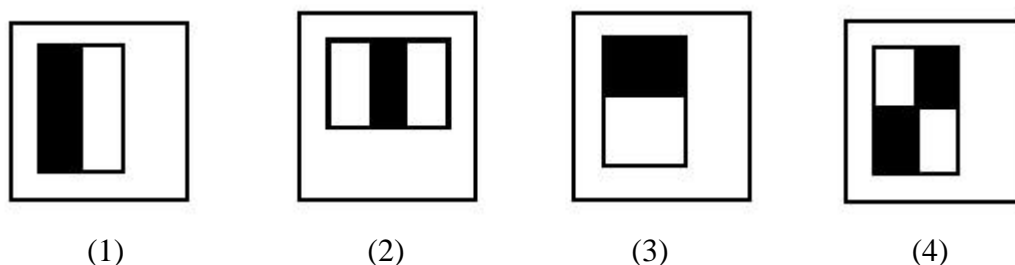


图 3.1 Viola 使用的 4 种矩形特征

- (1) 两个矩形，特征值为黑色区域像素值之和与白色区域像素值之和的差值。

- (2) 三个矩形，特征值为外面两矩形内像素值之和减去中间矩形内的像素值和。
- (3) 四个矩形，特征值为两个对角矩形内的像素值之和与另两个的差值。

后来 Lienhart 等<sup>[22]</sup>人对此进行了改进，又提出了另外一些类 Haar 特征，如图 3.2 所示。

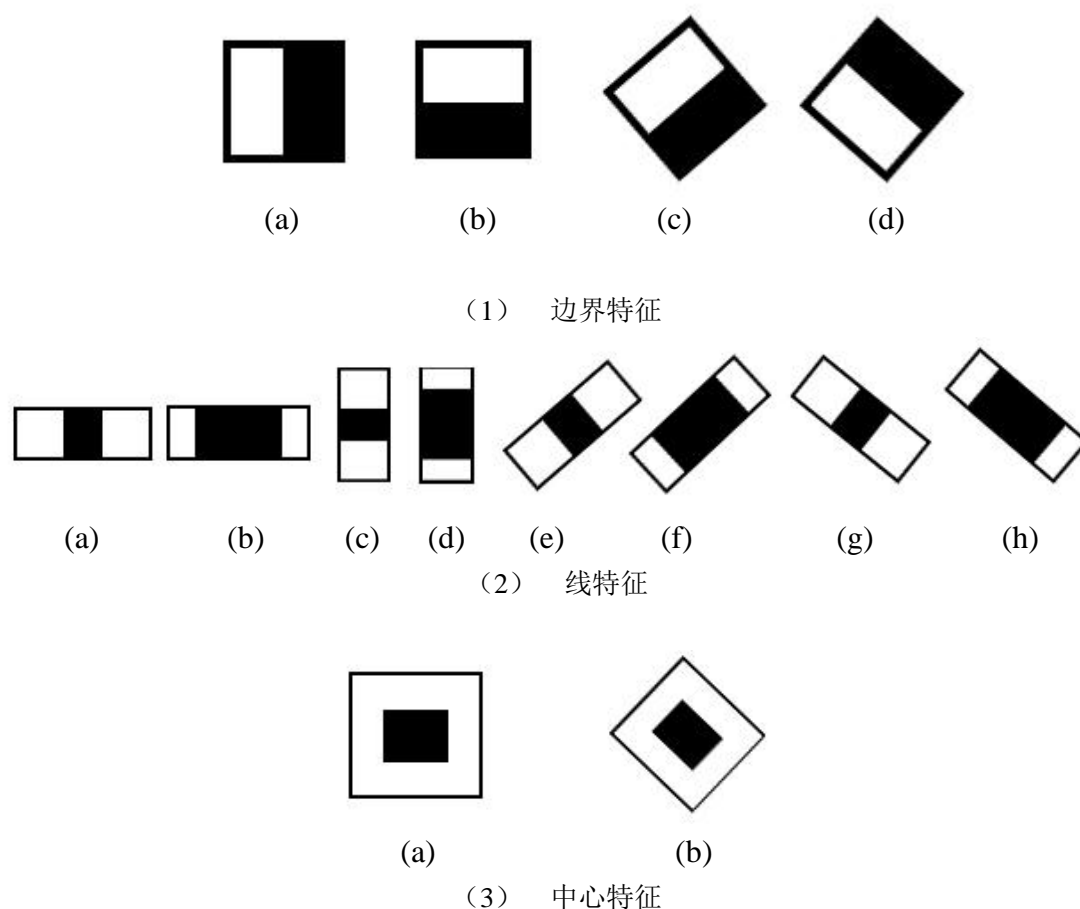


图 3.2 其他类 Haar 特征

从外观上看，这些特征只是一堆堆带条纹的矩形，但实际上意义很大。将上面的任意一个矩形放到人脸区域上，然后，将白色区域的像素和减去黑色区域的像素和，得到的值称之为人脸的特征值。如果把这个矩形放到一个非人脸区域上，那么计算出来的特征值应该和人脸特征值是不一样的，而且越不一样越好，所以这些方块的目的就是把人脸特征量化，以区分人脸和非人脸。

特征矩形特征的算法十分简单，但是这样简单的特征背后隐含着一个巨大的难点，会给我们的分类器训练过程带来很大的困难，那就是矩形特征的数量十分的庞大。下面

简单讨论下该数量，如果最小的矩形特征宽为  $w$ ，高为  $h$ ，图像的宽度为  $W$ ，高为  $H$ 。  
定义：

$$X = [W/w], Y = [H/h] \quad (3.1)$$

并且规定矩形特征在不同尺度下保持一定的宽高比，那么可以由如下的公式计算出矩形特征的个数：

$$X \cdot Y \cdot (W + 1 - w \frac{X+1}{2}) \cdot (H + 1 - h \frac{Y+1}{2}) \quad (3.2)$$

如果图像的大小为  $24 \times 24$ ，最小矩形特征的宽为 2，高为 1，保持这样的宽高比的矩形特征就有 43200 个。其他的特征计算可以由同样的方法得到，特征的种类越多，累积的图像特征越多。数目庞大的特征必须要找到一种简单的计算方式才能在实际中使用，否则算法就会因计算量过大而失去生命力。下面将简单介绍快速计算的方法——积分图像。

### 3.1.2 积分图像

如前面所述，由于矩形特征的数量非常之多，如果直接计算无疑是既浪费时间又浪费空间的，甚至是超出运算量的，导致溢出。将图像表示成积分图像的形式，这个概念的引入成功的解决了这个难题。

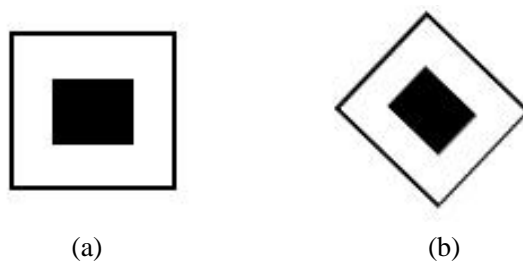


图 3.3 积分图像示意图

如图 3.3 所示，像素  $(x, y)$  点的积分图像定义为该点左上方矩形区域内所有像素点的和，即：

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (3.3)$$

这里  $ii(x, y)$  是值积分图像， $i(x, y)$  是原图像。使用如下的迭代公式，再遍历一遍原图像后可以得到原图像的积分图像。

$$s(x, y) = s(x, y - 1) + i(x, y) \quad (3.4)$$

$$ii(x, y) = ii(x - 1, y) + s(x, y) \quad (3.5)$$

这里 $s(x, y)$ 是像素的行累加和，而且为了方便计算，将图像边缘的累加和定义为 $s(x, -1)=0$ ，边缘处积分图像的值 $ii(-1, y)=0$ 。

利用积分图像的概念就可以快速的计算前面的矩形特征，如图 3.4 所示。矩形 D 内的像素和可以由四个参考矩形计算得到。位置 1 的积分图像是矩形 A 的像素和，位置 2 的是 A+B，位置 3 的是 A+C，位置 4 的是 A+B+C+D。所以 D 的像素和可以有  $4+1-(2+3)$  得到。因此二矩形特征可以通过 6 个参考矩形求得，三矩形特征可以通过 8 个参考矩形求得，四矩形通过 9 个求得。利用积分图像，就可以通过简单的加减运算得到图像的矩形特征，省去了传统算法中重复计算的麻烦。

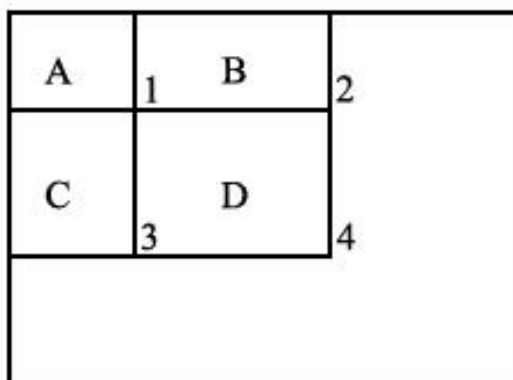


图 3.4 利用积分图像快速计算矩形内像素和

## 3.2 分类器学习

给定一个包含正负样本的训练特征集，很多机器学习算法都可以用来学习一个分类器来区分正负图像，本文采用的是 Adaboost 算法<sup>[23]</sup>。

Adaboost 算法，即自适应 Boosting 算法。它是利用不断增强学习的思想，每次训练一个弱分类器，这个弱分类器的表现只需要比随机猜想效果好，当增加弱分类器到强分类器中时，强分类器的表现都会变好，而对所选的弱分类器不做过多要求，只需要比随机猜想好。强分类器具体表现为对样本的判断能力一直在增强，错误率在下降，这些逐渐产生的弱分类器通过权重联合，每次训练都会调整训练样本权重，调整权重的原则是依据“更关注被错分的样本”，若当前强分类器对训练样本其中一些错分，那么就将这些被错分的样本的权重加大，将那些正确分类的样本权重减少，这样在寻找下一个弱分类器时会更加重视被分错的样本，这样从整体上才能达到不断加强学习的效果。



弱分类器是要通过对训练样本特征的学习，找到最好的分类阈值，使得错误分类的样本数最少。一个弱分类器 $h_j(x)$ ，由特征点 $f_j$ ，阈值 $\theta_j$ 和表示不等式方向的奇偶项 $p_j$ 组成：

$$h_j(x) = \begin{cases} 1, p_j f(x) < p_j \theta_j \\ 0, \text{其他} \end{cases} \quad (3.6)$$

这里  $x$  是输入的训练样本特征。

Adaboost 算法的流程如下：

给定训练样本集： $(x_1, y_1), \dots, (x_m, y_m); y_i \in \{-1, +1\}$

初始化样本权重： $D_1(i) = 1/m$

For  $t=1, \dots, T$

(1) 根据最小训练样本错误率找到一个弱分类器：

$$h_t = \arg \min_{h_j \in H} \varepsilon_j = \sum_{i=1}^m D_t(i) [y_i \neq h_j(x_i)]$$

(2) 判断此 $\varepsilon_t$ 是否大于  $1/2$ ，若大于  $1/2$  停止迭代，因为若大于  $1/2$ ，这个弱分类器比随即猜想的效果都要差，就没有任何意义。

(3) 计算分类器权重：

$$\alpha_t = \frac{1}{2} \log \left( \frac{1 - \varepsilon_t}{\varepsilon_t} \right)$$

(4) 更新样本权重：

$$D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t} \quad (Z_t \text{ 为归一化系数})$$

(5) 最后输出强分类器为：

$$H(x) = \text{sign} \left( \sum_{t=1}^T \alpha_t h_t(x) \right)$$

End

利用 Adaboost 分类器可以对前面所述的众多矩形特征进行筛选，挑出最具判决性的特征，极大的提高了运算速度。

### 3.3 分类器级联

Adaboost 算法可以有效的从众多弱分类器中挑选出最具有分类意义的组合形成强分类器，并且实验证明这个最终的强分类器具有很好的检测率。但是基于矩形特征的分类算法数量庞大，需要较多的特征才能实现分类，而且分类器中，弱分类的数量过多反而会使得分类器泛化的能力变弱，影响实验效果，所以就引入了层叠分类器的概念。层叠分类器是由多个 Adaboost 强分类器组成，每一层分类器就是一个小型的 Adaboost 强

分类器，小型的分类器的要求是在第一层的时候分类的要求较低，随着层数的增加，分类器的分类条件限制越强，即分类越逼近准确。前几层的较简单的分类器可以排除大部分的非目标窗口，从而后面较复杂的分类器只要针对那些更难区分的子窗口进行运算，就可以达到较低的误检率。由于难于区分的窗口需要用很多的特征才能进行分类，采用层叠的方式可以一层一层逐步进行更细化的分类，使得复杂的分类器只注重细节分类，忽略了粗分类的特征，这样使得检测速度相对的提高了。

层叠分类器的示意图如图 3.5 所示：

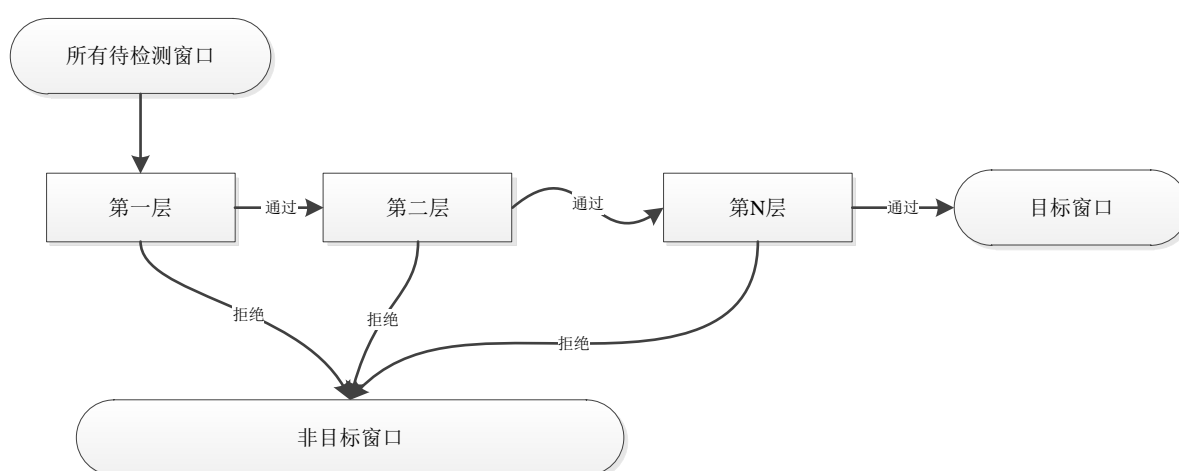


图 3.5 层叠分类器

层叠分类器的检测过程类似于决策树<sup>[24]</sup>，第一层分类器把判决出的目标结果送入第二层分类器继续判决，接着第二层分类器再把判决出目标结果送入第三层，依此类推。任何一层分类器判决出的非目标结果被直接排除，不再进行判决。所以当前层的强分类器面对的都是相对于上一层来讲更难分类的任务，因为能通过前面所有层的样本相对于普通样本来讲是更加难以区分的。

只有通过层叠分类器所有层的子窗口才被认为是目标。简单的非目标窗口在前面几层被就滤掉，越到后面的几层待检测的窗口越少，这使得后面的分类器可以专注于复杂窗口的判别，减小了复杂分类器的负担，提高了检测精度与速度。

### 3.4 人脸检测结果

我从网上下载了一些包含人脸的图片进行检测，实验证明该算法能检测出绝大部分的人脸，具有很强的鲁棒性，并且速度极快。此外，该算法不仅能识别单个人脸，也能检测出一幅图像中的多个人脸。部分单帧检测结果如下图 3.6、3.7 所示。

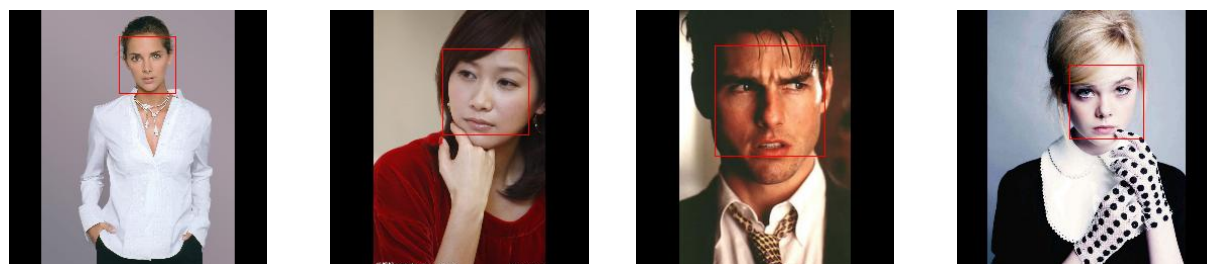


图 3.6 彩色图像人脸检测结果

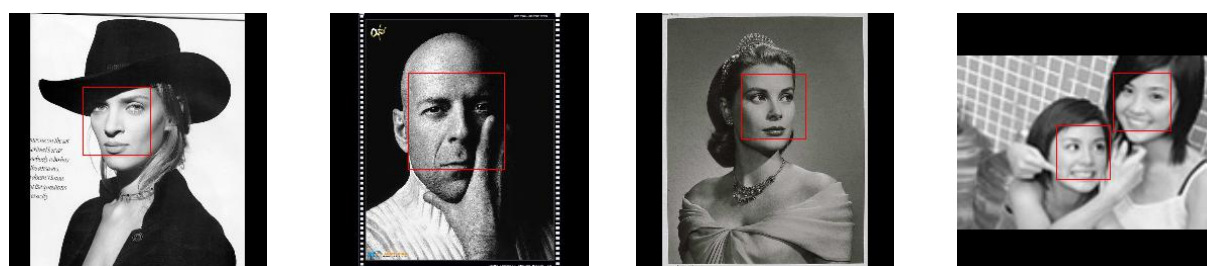


图 3.7 灰度图像人脸检测结果

为与经典算法进行比较，本文采用 MIT+CMU 人脸数据库，使用两种方法对其进行比较检测。表 3.1 给出了本方法与经典方法之一的基于神经网络人脸检测算法的检测率比较。从图 3.8 中可以看出，基于级联的类 Haar 特征人脸检测算法在分类器层数较高的情况下能够实现比基于神经网络人脸检测算法更高的检测率。

表 3.1 检测结果比较

方法	分类器层数							
	10	31	50	65	78	95	167	422
基于级联的类 Haar 特征人脸检测算法	<b>76.10%</b>	<b>88.40%</b>	<b>91.40%</b>	<b>92.00%</b>	<b>92.10%</b>	<b>92.90%</b>	<b>93.90%</b>	<b>94.10%</b>
基于神经网络人脸检测算法	83.20%	86.00%	86.80%	87.20%	88.70%	89.20%	90.10%	89.90%

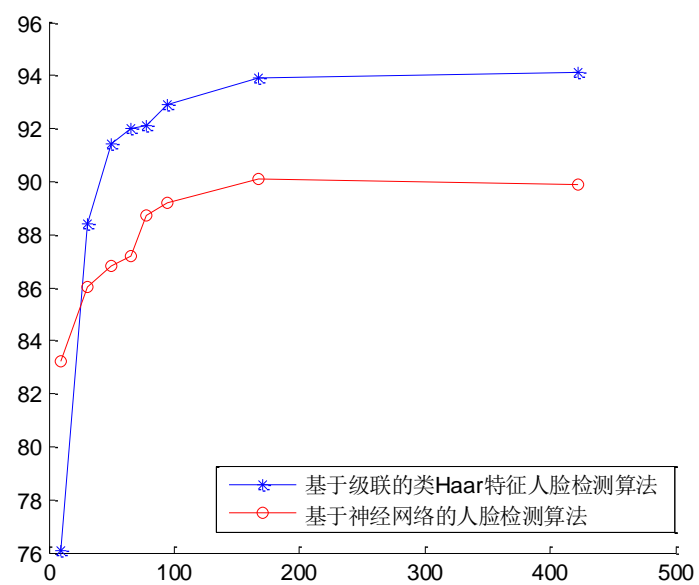


图 3.8 检测结果比较

## 4 基于粒子滤波的人脸跟踪算法

### 4.1 算法原理

采用粒子滤波算法进行跟踪一般需要两步，即计算特征脸和人脸跟踪。首先是计算特征脸，设人脸图像  $I(x,y)$  为二维  $N \times N$  的灰度图像，用  $N^2$  维向量  $\Gamma$  表示。人脸图像训练集为  $\{\Gamma_i | i = 1, \dots, M\}$ ，其中  $M$  为训练集中图像总数。这  $M$  幅图像的平均值向量为：

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i \quad (4.1)$$

每个人脸  $\Gamma_i$  与平均人脸  $\Psi$  的差值向量  $\Phi_i$  为：

$$\Phi_i = \Gamma_i - \Psi; i = 1, \dots, M. \quad (4.2)$$

训练图像的协方差矩阵可表示为：

$$C = A^T A \quad (4.3)$$

其中  $A = [\Phi_1, \dots, \Phi_M]$ 。

特征脸由协方差矩阵  $C$  的正交特征向量组成。对于  $N \times N$  人脸图像，协方差矩阵  $C$  的大小为  $N^2 \times N^2$ ，对它求解特征值和特征向量是很困难的，一种取而代之的方法是解  $M \times M$  个较小的矩阵。首先计算  $M \times M$  矩阵  $L$  的特征向量  $v_l (l=1, \dots, M)$ ：

$$L = A^T A \quad (4.4)$$

矩阵  $C$  的特征向量  $u_l (l=1, \dots, M)$  由差值图像  $\Phi_i (i=1, \dots, M)$  与  $v_l (l=1, \dots, M)$  线性组合得到：

$$U = [u_1, \dots, u_M] = [\Phi_1, \dots, \Phi_M][v_1, \dots, v_M] = AV \quad (4.5)$$

实际上， $m (m < M)$  个特征脸足够用于人脸识别或跟踪。因此，仅取  $L$  的前  $m$  个最大特征值的特征向量计算特征脸。 $m$  由门限值  $\theta_\lambda$  确定：

$$J = \min_r \left\{ r \mid \sum_{i=1}^r \lambda_i / \sum_{j=1}^M \lambda_j > \theta_\lambda \right\} \quad (4.6)$$

取得的这  $m$  个特征向量就形成了特征脸张成的子空间。

其次是基于特征脸的人脸跟踪。基于特征脸的人脸跟踪过程由两个阶段组成：训练阶段和识别阶段。在训练阶段，每个已知人脸  $\Gamma_k$  映射到由特征脸张成的子空间上，得到  $m$  维向量  $\Omega_k$ ：

$$\Omega_k = U^T (\Gamma_k - \Psi) \quad k = 1, \dots, L \quad (4.7)$$

其中  $N_c$  为已有跟踪结果个数。

在识别阶段，首先把待识别的所有候选图像  $\Gamma$  映射到特征脸空间，得到向量  $\Omega$ ：

$$\Omega = U^T (\Gamma - \Psi) \quad (4.8)$$

然后计算所有候选图像的重构误差，取重构误差最小的图像块作为当前的跟踪结果。

## 4.2 算法流程

第三章的算法检测到人脸之后，用矩形框标记出来，用像素的灰度值作为图像特征。由于前几帧目标变化一般不大，可以利用简单的欧式距离直接比较，获得 2~10 帧的跟踪结果。然后利用这 10 帧的结果训练得到 PCA 的特征向量基，并用此空间基进行下面的跟踪。为适应环境的改变，每隔 5 帧重新训练得到新的特征基，并舍弃一些较旧的跟踪结果。

算法流程图如下图所示：

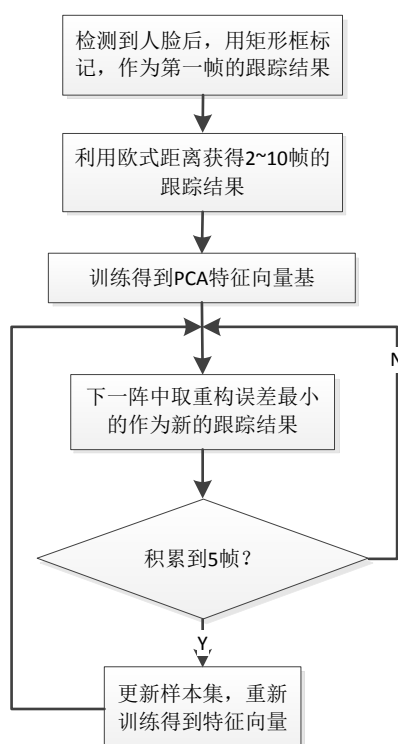


图 4.1 人脸跟踪算法流程图

### 4.3 运动模型和观测模型

本文算法的整体框架是基于贝叶斯推断模型，即试图用所有已知信息来构造系统状态变量的后验概率密度。状态变量  $X_t$  描述了在时间  $t$  目标的仿射运动参数，即目标位置。给定一系列观测图像集  $\tilde{I}_t = \{I_1, \dots, I_t\}$ ，我们的目标是估计隐藏状态变量  $X_t$  的值。由贝叶斯定理可得：

$$p(X_t | \tilde{I}_t) \propto p(I_t | X_t) \int p(X_t | X_{t-1}) p(X_{t-1} | \tilde{I}_{t-1}) dX_{t-1} \quad (4.9)$$

因此跟踪过程可分为两部分：运动模型  $p(X_t | X_{t-1})$ ，用系统状态转移预测状态的后验概率密度，也即两个状态之间的转换；观测模型  $p(I_t | X_t)$ ，用最近的观测值进行修正，即已知  $X_t$  估计的  $I_t$ ，判断其与目标的相似度。

运动模型，简单的说就是如何从上一帧的目标位置得到本帧的候选。本文采用的是基于粒子滤波的采样技术，即在上一帧的目标中心周围按高斯分布向四周撒播粒子，进行采样。但是在得到所需的后验概率密度时没有采用所有粒子的加权的方法，而是对每个粒子单独考虑，分别计算其为真实目标的概率。此外，为避免粒子滤波中粒子退化的问题和重采样的技术，本文采用每帧重新撒播粒子的方法，使计算更方便。

两连续帧之间的目标运动可以用仿射变换来估计。本文我们使用了 6 参数的仿射变换来建模  $X_{t-1}$  到  $X_t$  的状态转移，其中  $X_t$  表示被跟踪的目标在  $t$  帧的状态。令  $X_t = (x_t, y_t, \theta_t, s_t, \alpha_t, \varphi_t)$ ，其中  $x_t, y_t, \theta_t, s_t, \alpha_t, \varphi_t$  分别表示在时间  $t$  时的  $x, y$  方向的偏移量，旋转角度，高宽比和斜切方向。 $X_t$  中的每个参数都被认为是  $X_{t-1}$  周围的高斯分布，并且各自都是独立的，即

$$p(X_t | X_{t-1}) = N(X_t; X_{t-1}, \Psi) \quad (4.9)$$

其中  $\Psi$  是一个对角协方差矩阵，它的每个元素对应着各个仿射参数的协方差，即  $\sigma_x^2, \sigma_y^2, \sigma_\theta^2, \sigma_s^2, \sigma_\alpha^2, \sigma_\varphi^2$ 。

观测模型即判断哪一个候选最接近真实目标，作为本帧的跟踪结果。本文的方法为取重构误差最小。已经得到了 PCA 的特征向量基，把每个候选都投影到该向量基上，然后重构，与原数据比较，差异最小的就认为是当前的目标位置。

## 4.4 空间更新

随着跟踪的进行，目标姿态和环境都在不断的变化。在复杂背景中，存在着大量噪声、背景光照变化、地貌起伏、遮挡等情况，导致实时拍摄的目标在不断地变化。只有对模板进行合理的更新，才能在一定程度上克服这些变化对跟踪效果的影响。

在匹配跟踪算法中，模板是目标图像形态的一个记忆，是维持连续跟踪的纽带。跟踪过程中，目标图像的变化是必然的，模板更新也是必须的，是图像匹配跟踪的关键。模板更新方法通常有三种。第一种方法是固定模板方法，以目标第一次出现时的图像形态（即第一帧检测到的目标）作为模板，并且在以后的跟踪过程中保持不变，这种方法记住的是目标的初始形态，不能适应环境照度和目标形态的变化，在长时间跟踪过程中不稳定。第二种方法是逐帧更新模板方法，以当前最佳匹配子区图像作为模板，而抛弃目标以往的所有信息，这样会因为遮挡的发生或是匹配误差的累积，造成跟踪的漂移甚至失败。第三种方法则是前面两种方法的互补，既保留第一帧的信息，也加入最新帧的变化。

固定模板方法有其自身的优势，它计算量小，同时不会产生模板漂移问题，但它不能适应目标长期缓慢变化，造成跟踪的不稳定；而逐帧更新模板可以跟上目标缓慢的变化，但在复杂背景下，若跟踪过程中某一帧图像质量较差，或者前一帧跟踪质量不佳，都会影响后续帧的跟踪，造成跟踪误差积累，从而导致跟踪错误甚至跟踪失败。所以，我们应尽量扬长避短，更智能地进行模板更新。

本文采取的更新方法为：始终保留第一帧的图像特征，因为第一帧是最准的，包含的信息量最精确。同时在跟踪的过程中保留目标位置，当累积到 5 帧后，舍去前面的跟踪结果。利用最新的几帧结果和第一帧信息重新训练学习 PCA 的特征向量基。这样做既保证了目标不会偏移，因为第一帧结果一直参与学习，又能适应环境的变化，因为同时引入了最新帧的信息。

## 4.5 跟踪结果

为此，特地住在所住的寝室找同学拍了一段视频，视频中包含了光照不理想、人脸尺度变化、遮挡物等变化条件。另外也对网上公开的数据库进行了跟踪测试。分别如图 4.2、4.3、4.4、4.5 所示。从跟踪结果来看，该算法能够适应多种不同的条件，具有很好的鲁棒性，跟踪速度也很快。



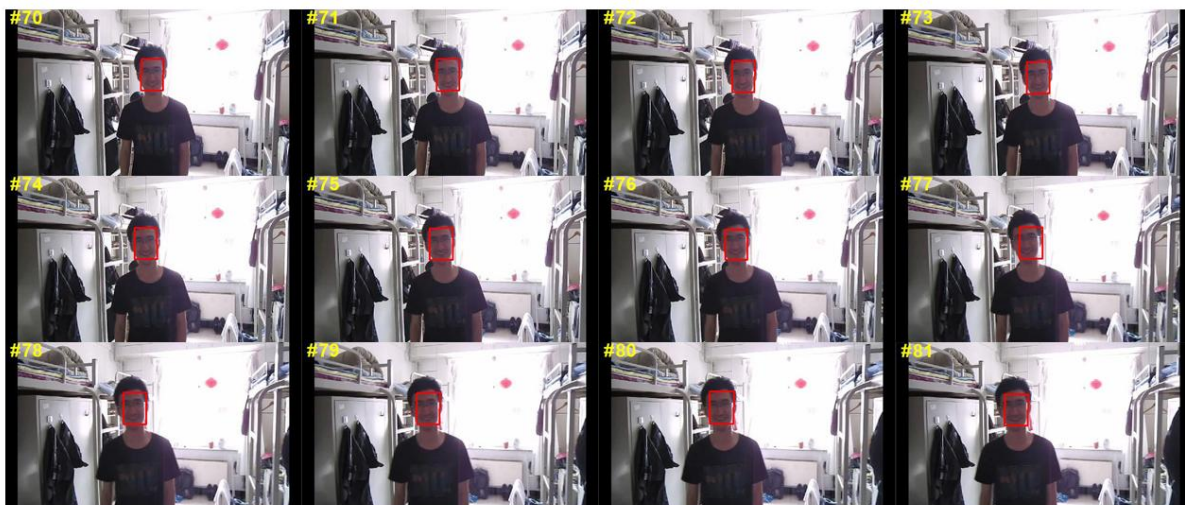


图 4.2 视频序列帧中人脸跟踪的结果（一）

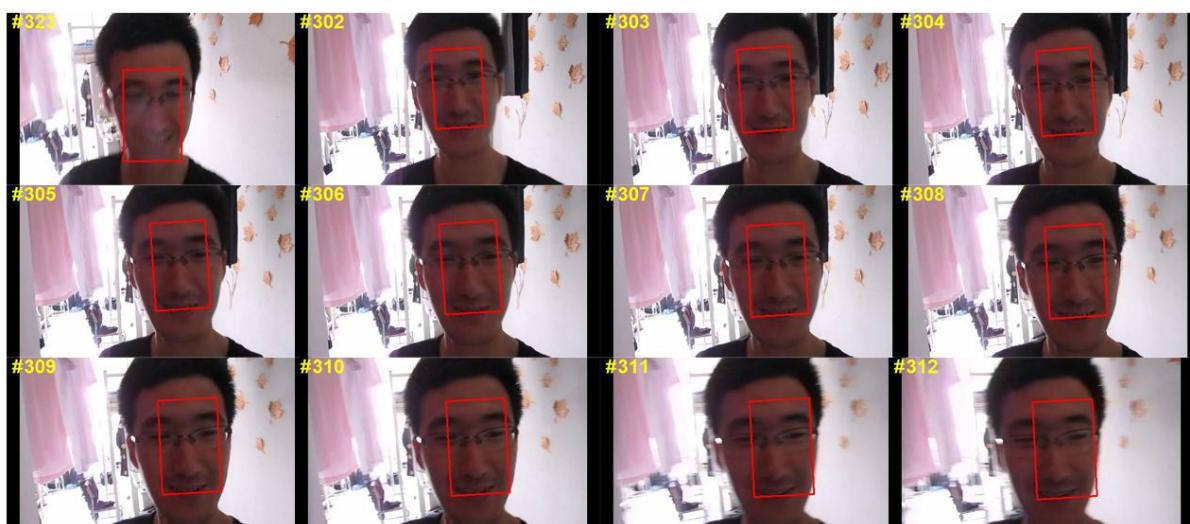


图 4.3 视频序列帧中人脸跟踪的结果（二）



图 4.4 视频序列帧中人脸跟踪的结果（三）

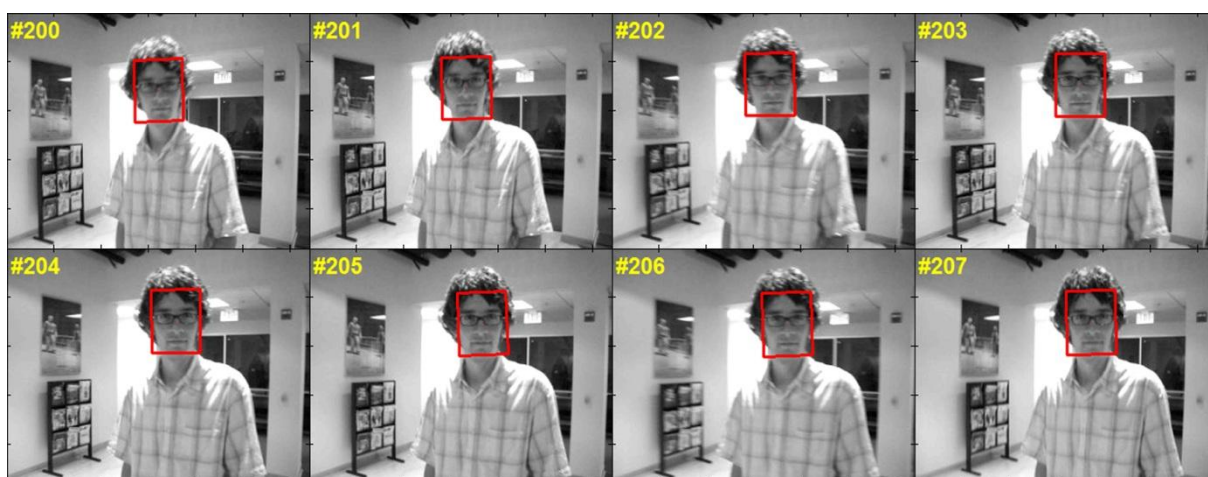


图 4.5 视频序列帧中人脸跟踪的结果（四）

为和其他经典算法比较，分别用基于粒子滤波的跟踪算法、均值偏移跟踪算法、基于分类器的跟踪算法对这两段视频进行了跟踪测试，对三种算法的中心偏差进行了计算。比较结果如图 4.6、4.7 所示。从图中可以看出，采用基于粒子滤波实现跟踪的算法，在两段视频中都实现了较好的跟踪，而另外两种算法都出现了不同程度的目标丢失情况。跟踪速度上本算法并无明显优势，但是在跟踪效果上却明显好于这几种经典的方法。

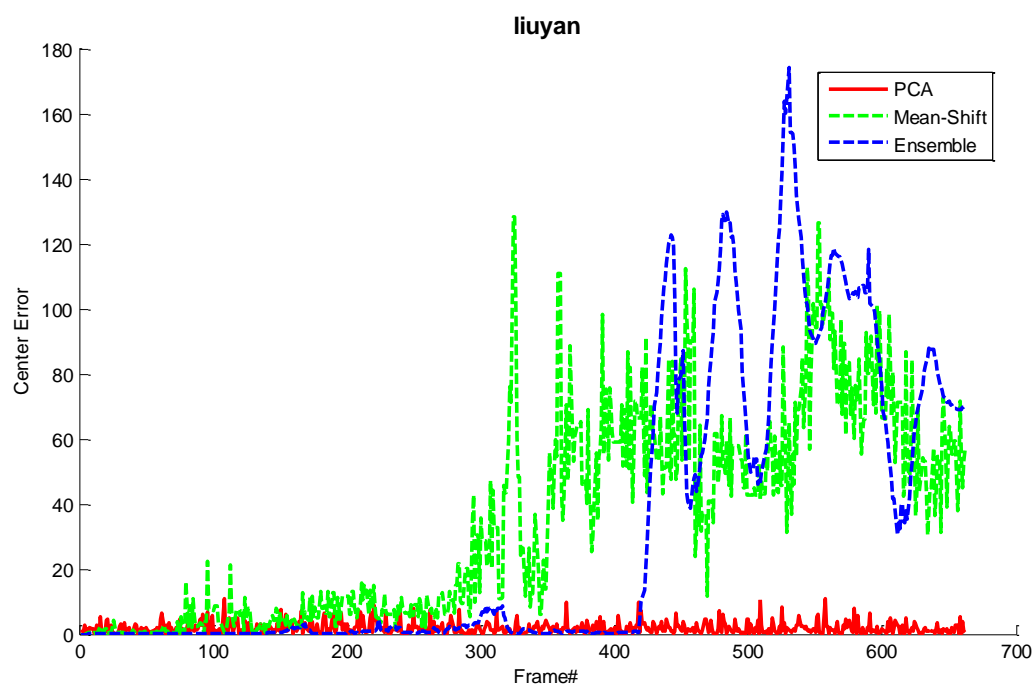


图 4.6 三种算法在视频 1 中的跟踪结果比较

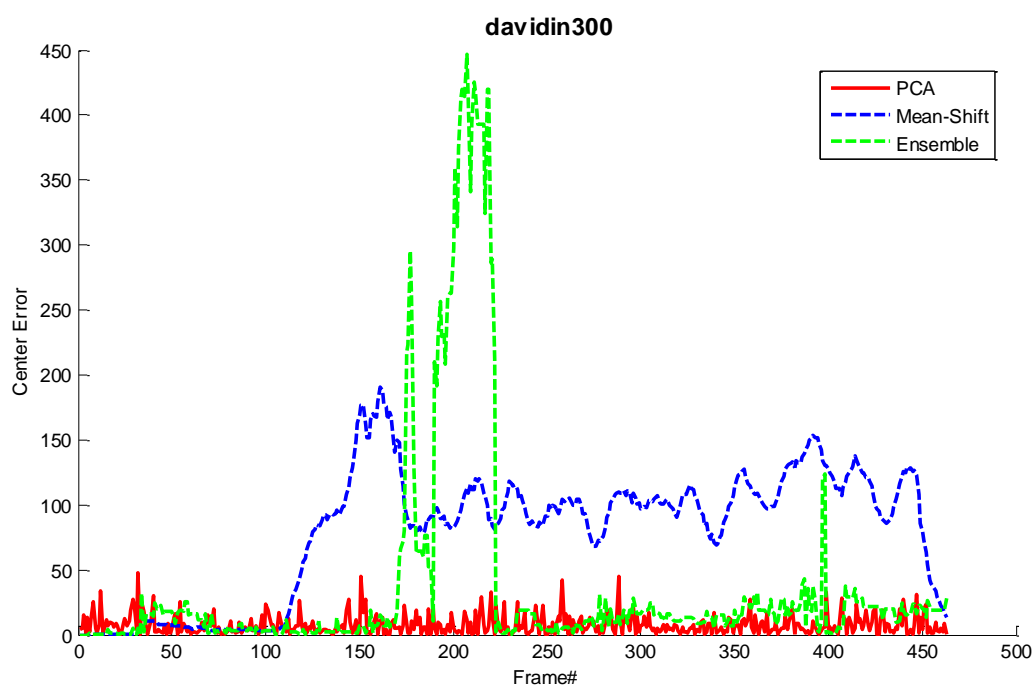


图 4.7 三种算法在视频 2 中的跟踪结果比较

## 结 论

人脸检测与跟踪技术由于其在生物识别领域的特殊地位，其应用范围越来越广泛，引起的关注也越来越多。人脸检测技术和目标跟踪技术作为技术的核心，发展程度也日趋成熟。能够准确检测出人脸并跟踪，不仅是当今各种人脸识别系统的前提，其自身更是很具有广阔的应用前景。

本文从最原始的方法入手，纵览了人脸检测与跟踪领域的研究历史、技术难点，讨论了时下流行的各类方法，讨论了各类方法的核心内容、优缺点等，并提出了基于级联的类 Haar 特征人脸检测算法与基于粒子滤波的人脸跟踪算法，录制了相关视频进行了检测跟踪实验并获得了成功。

本文的重点是基于级联的类 Haar 特征人脸检测算法和基于粒子滤波的人脸跟踪算法。在一段视频中，将第一帧用基于级联的类 Haar 特征人脸检测算法将人脸检测出来，在接下来的序列帧中，用基于粒子滤波的人脸跟踪算法实现第一帧中所检测出来的人脸的跟踪定位。基于级联的类 Haar 特征人脸检测算法在人脸检测中具有较强的鲁棒性，检测效率和准确率较高；基于粒子滤波的人脸跟踪算法在跟踪中实现的效果较好，不易出现失败或偏移等现象。这两种算法的结合，符合相关设计要求，较好的实现了视频序列中人脸检测和跟踪定位。

然而在一些存在景深效果的照片中，人脸的检测效果较差，主要是由于照片的景深部分本身比较模糊，提取特征比较困难所致。这类照片中的跟踪也就很难实现。如何调整第一帧的检测效果以提高效率和准确率，是解决这类问题的方法之一。

## 参 考 文 献

- [1] Yang M H, Ahuja N, Kriegman D. A survey on face detection methods[EB/OL], 1999, <http://vision.aiuiuc.edu/mhyang/papers/survey.ps.gz>.
- [2] 梁路宏, 艾海舟, 徐光佑, 等. 人脸检测研究综述[J]. 计算机学报, 2002,25(5):p449~458.
- [3] 贺长伟. 关于运动人体跟踪算法的研究[D]. 济南:山东大学, 2007.
- [4] 刘青山. 人脸跟踪与识别的研究[D]. 中国科学研究生院, 2003.
- [5] Craw I, Ellis H, Lishman J R. Automatic extraction of face features[J]. Pattern Recognition Letters, 1987,5(2):p183~187.
- [6] Govindaraju V. Locating human faces in photographs[J]. International Journal of Computer Vision, 1996,19(2):p129~146.
- [7] Reisfeld D, Yeshurun Y. Robust detection of facial features by generalized symmetry[C]. Proceedings of the International Conference on Pattern Recognition, Hague, Netherlands, 1992:p117~120.
- [8] Yuille A L, Hallinan P W, Cohen D S. Feature extraction from faces using deformable templates[J]. International Journal of Computer Vision, 1992,8(2):p99~111.
- [9] Cootes T F, Taylor C J. Active Shape Models - 'Smart Snakes'[C], Proc. British Machine Vision Conference, 1992:p266~275.
- [10] Lanitis A, Taylor C J. A unified approach to coding and interpreting face images[C], Proceeding of the IEEE Computer Society Conference on Computer Vision, 1995:p368~373.
- [11] Cootes T F, Edwards G J, Taylor C J. Active appearance models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001,23(6):p681~785.
- [12] Kass M, Witkin A, Terzopoulos D. Snakes: active contour models[C], Proceeding of the IEEE Computer Society Conference on Computer Vision, 1987,1:p59~268.
- [13] Wu H, Yokoyama T, Pramadiahanto D, et al. Face and facial feature extraction from color image[A], Proc. Automatic Face and Gesture Recognition[c], 1996:p345~350.
- [14] Turk M, Pentland A. Face recognition using eigenfaces[C]. Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Lahaina, Mawaii, 1991:p586~591.
- [15] Martinez A, Kak A. PCA versus LDA[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(2): p228~233.
- [16] Yang M H, Ahuja N, Kriegman D. Face detection using mixtures of linear subspace[C]. Proceedings of the International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 2000:p70~76.
- [17] Rowley H A, Baluja S, Kanade T. Neural network-based face detection[J]. IEEE Trans Pattern Analysis and Machine Intelligence, 1998, 20(1): p23~38.
- [18] 艾海舟, 王栓, 何克忠. 基于差分图像的人脸检测[J]. 中国图像图形学报, 1998, 3(12):p987~992.
- [19] Viola P, Jones M. Rapid Object Detection using a Boosted Cascade of Simple Features[C]. In IEEE CVPR, 2001.
- [20] Viola P, Jones M. Robust Real-Time Face Detection[J]. In IJCV, 2004(2), p137~154.

- [21] Papageorgiou C, Oren M, and Poggio T. A general framework for object detection[C]. In ICCV,1998.
- [22] Lienhart R, Maydt J. An Extended Set of Haar-like Features for Rapid Object Detection[C]. In IEEE ICIP, 2002,p900~903.
- [23] Freund Y, Robert E S. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Journal of Computer and System Sciences. 1997, 55(1):p119~139.
- [24] Quinlan J. Induction of decision trees[J]. Machine Learning, 1986, 1:p81~106.



## 致 谢

本论文是在徐刚老师的精心指导和关心下完成的，在相关理论知识的学习、毕业设计开题、方案选择和设计、论文撰写和修改等过程中，得到了徐老师富有启发性的建议和精心的指导。徐老师待人真诚、思维敏捷、严谨求实，为我树立了很好的榜样。

非常感谢信通学院的李甫和孙明浩同学，感谢他们在我毕业设计整个过程中对我的无私帮助，感谢他们在我遇到困难时总是能为我一一解答。由于所学内容并不相关，这两位同学为我在理论学习阶段提供了很大的便利，给予了很大的帮助。

非常感谢计算机学院的刘岩同学，感谢他配合我拍摄了相关的跟踪视频。

最后感谢辛勤养育我和教育我的父母，以及我人生路上的路灯我的姐姐，感谢家人对我的关心和支持，我和我的一切都要归功于我的家人。