

基于三维模型的单目图像 序列人脸姿态跟踪

姓名: 梁国远

学号: 10180805

院系: 信息科学技术学院

专业: 信号与信息处理

研究方向: 计算机视觉与智能信息处理

导师: 查红彬 教授

二零零五年五月

3D MODEL BASED FACE POSE TRACKING FROM A MONOCULAR IMAGE SEQUENCE

BY

GUOYUAN LIANG

School of Electronic Engineering and Computer Science
Peking University

Directed by Prof. Hongbin Zha

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Signal and Information Processing
in the Graduate College of
Peking University, 2005

Beijing, China

©Copyright 2005 Guoyuan Liang. All rights reserved.

版权声明

任何收存和保管本论文各种版本的单位和个人，未经本论文作者同意，不得将本论文转借他人，亦不得随意复制、抄录、拍照或以任何方式传播。否则，引起有碍作者著作权之问题，将可能承担法律责任。

献给我的家人

摘要

人脸姿态跟踪(Face Pose Tracking)是智能人机交互和计算机视觉研究中的一项基本课题，也是近年来人们越来越感兴趣的一个研究方向。人脸姿态跟踪的主要目的是在一组包含人脸的图像序列中计算得到人脸在三维空间中的姿态参数。人脸姿态跟踪在表情识别、人脸识别、姿态理解、视频会议、智能监控、疲劳检测、虚拟现实、游戏和娱乐等方面都具有广阔的应用前景。

现有的人脸姿态跟踪方法大体上可以分为基于特征的方法和基于模型的方法。基于特征的方法假设在人脸姿态和人脸图像的某些特征之间存在唯一的对应关系，其目标是通过大量不同姿态的图像样本确定这种对应关系。这类方法需要大量的训练样本或进行插值操作，因此结果往往不够准确；此外还存在如何定义特征的问题，在定义的特征描述不充分的情况下难以获得满意的结果。基于模型的方法一般用某种三维结构或模型来近似人脸，基于二维/三维特征对应求解姿态参数。这类方法相对易于实现，精度较高，可是常常需要求解维数较高的非线性方程，因此难于获得闭合形式的解，在初始估计值不够准确的情况下，可能会出现陷入局部最小或出现不收敛的现象；此外，该类方法依赖于人脸特征检测的准确程度，往往容易受到光照变化和遮蔽的影响。

本文针对仅有一个摄像机获取的单目视频图像序列的情况，将人脸三维模型引入人脸姿态跟踪系统框架以提供深度信息。我们将姿态跟踪分为初始姿态估计和帧间姿态跟踪两个子过程。首先用仿射对应的方法得到初始帧的人脸姿态参数并作为姿态跟踪的起点；然后用三维模型的几何信息对线性灰度和深度约束方程加权得到更精确的帧间运动参数；为了消除光照变化和遮蔽的影响，在跟踪过程中逐帧自动进行特征点更新。对模特头像和真实人脸进行的实验结果表明该算法能在较大的变化范围内实现可靠的人脸姿态跟踪，尤其是对深度方向有明显变化的运动，该算法的作用表现得更加明显。与以前的工作相比，本文的贡献主要体现在以下几个方面：

- 提出了基于仿射对应原理得到精度较高的初始姿态估计值的方法。利用三维模型，首先建立了一个满足平面假设条件的虚拟人脸平面，因此不需要用摄像机获取一幅正面平行图像作为参考帧；然后旋转人脸三维模型以得到精确

度较高的特征点正面平行投影；接着三次旋转摄像机坐标轴并求出人脸全部的六个姿态参数；最后将基于仿射对应得到的参数作为初始估计值，通过一个优化迭代过程对姿态参数求精，从而使估计结果能够不依赖于平面假设条件。在计算过程中还用鲁棒方法进一步提高了估计的精度。

- 提出了用加权的线性深度灰度约束求解人脸姿态参数并进行姿态跟踪的方法。利用三维模型的几何信息，不仅可以得到帧与帧之间线性的深度和灰度约束方程，还能根据各个特征点所在区域的几何特性，计算出相应的权值并对约束方程加权。这种方法不仅避免了求解非线性问题，而且考虑了不同特征点对约束方程的不同影响，因此能进一步提高估计结果的准确性。
- 提出了一种自动的特征点动态更新方法。受人脸表情、光照的变化和遮蔽等因素影响，在连续跟踪的过程中，某些特征点可能出现较大的误差，显然人工确定误差的阈值是不合适的，我们提出了一种自动算法，能够逐帧判断特征点的可靠性，丢弃误差较大的特征点并及时补充新的特征点，从而能有效地消除遮蔽和光照变化引起的特征点跟踪误差积累。

为了验证该算法的有效性，我们分别在石膏像图像序列、模特头像图像序列和真实人脸图像序列上进行了各种实验，给出了相关的实验结果，就仍然存在的问题进行了分析，并且对未来的工作进行了展望。

关键字：人脸姿态跟踪，单目视频图像序列，人脸三维模型，仿射对应，深度灰度约束加权，动态特征更新。

Abstract

Face pose tracking, a basic research topic in the field of computer vision and intelligent human computer interaction, is becoming more and more attractive recently. The main objective of face pose tracking is estimating the 3D pose parameters from an image sequence with human faces. The technology of face pose tracking can be widely used in face recognition, expression recognition, gesture understanding, video conference, intelligent surveillance, fatigue detection, virtual reality, game and entertainments.

The existing approaches for face pose tracking can be classified into two categories: property-based and model-based. The property-based methods assume certain relationship between the 3D head pose and some properties of the 2D face images and a large number of training images are used to determine the relationship. These methods are simple but need numbers of training images or interpolation, so the results usually are not very accurate. Besides this, there is an extra problem: how to determine the property? No satisfied results can be achieved when the description for property is not adequate. The model-base methods usually approximate the face with some 3D models or structures, and calculate the pose parameters based on 2D/3D feature correspondences. These methods are easy to be implemented with higher accuracy, but it's difficult to find a closed form solution for a high dimensional nonlinear problem. When the initial guess is not accurate enough, it might be trapped into local minimum or not converge at all. Furthermore, these methods rely on the accuracy of the detected or tracked features, which are easy to be affected by illumination changes or occlusions.

This dissertation deals with the case with the monocular image sequences captured by a single camera by introducing the face 3D model into tracking system to provide the depth measurements. The whole process includes two steps. First, estimate the pose parameters for the initial frame by using the affine correspondences, and use it as the starting point for the continuous pose tracking process. Second, make use of the weighted linear brightness and depth constraints to calculate the interframe motion parameters which is more accurate. In order to diminish the errors caused by illumination changes and occlusions, we dynamically update the features frame by frame. Experiments performed on real face and model head image sequences show

our method can reliably estimate face poses in a wide range of head motion. The algorithm works well especially in the case with obvious depth variance. Compared with previous work, the main innovations in our work can be stated as follows:

- The initial face poses estimation is based on affine correspondences algorithm. By using 3D face model, we first construct a virtual face plane which satisfied the face plane assumption, so the fronto-parallel frame used as a reference frame is not needed. Then rotate the 3D face model to produce accurate fronto-parallel projections of features. After that, the full motion parameters are calculated by rotating the camera coordinate system three times. In the end, use the pose parameters as initial guess and apply a nonlinear optimization process to refine the rough results. That makes the estimation not depend on the condition of plane assumption. We apply robust statistics method to improve the accuracy of the calculation.
- We proposed an algorithm for calculate the depth gradient with respect to time. In order to solve the linear depth constraint equation, the depth gradient with respect to time should be known. With monocular image sequences, the frame-rate depth information is not available, so we rotate the 3D models many times to minimize the projection errors, therefore calculate the accurate depth gradient with respect to time.
- We present an automatic features updating technique. Because of the inevitable presence of face expression changes, illumination changes and occlusions while pose tracking, some features position may be inaccurate. Apparently, manually setting the threshold for the error is not applicable here. We proposed an automatic algorithm which can evaluate the reliability of the features frame by frame, discard the inaccurate features and generate some new features if necessary. This technique can effectively reduce the error accumulations caused by occlusions, non-rigid face motion and illumination changes.

In order to examine the effectiveness of our approach, we do some experiments on the bust images, model head image sequences and real head image sequences. The experimental results are given and the existing problems are analyzed, in the end, the future work is also discussed.

Keywords: Face pose tracking, monocular image sequences, 3D face model, affine correspondences, weighted brightness and depth constraints, dynamic feature updating.

目 录

摘要	I
Abstract.....	III
目 录	V
第一章 绪论.....	1
1.1. 人体运动的视频分析简介	1
1.2. 人脸空间姿态跟踪及其应用	2
1.3. 论文的主要工作和结构	5
第二章 人脸姿态跟踪方法概述.....	8
2.1. 概述	8
2.2. 基于特征的方法	9
2.2.1. 基于神经网络的方法	9
2.2.2. 基于特征空间的方法	10
2.2.3. 基于肤色发色特征的方法	12
2.3. 基于模型的方法	12
2.3.1. 基于特征点匹配求解空间姿态的方法	12
2.3.2. 基于几何结构的方法	13
2.3.3. 基于平面模型的方法	14
2.3.4. 基于圆柱模型的方法	15
2.3.5. 基于三维模型的方法	21
第三章 基于三维模型的人脸姿态跟踪框架.....	23
3.1. 概述	23
3.1.1. 引入人脸三维模型	23
3.1.2. 人脸姿态跟踪系统框架	24
3.2. 人脸三维建模	26
3.2.1. 人脸三维数据的获取	27
3.2.2. 人脸三维数据的处理	28
3.2.3. 人脸三维建模的结果	30
3.3. 透视投影模型	30
3.3.1 图像坐标系、摄像机坐标系与世界坐标系	31
3.3.2 透视投影成像	32
3.4. 摄像机标定	33
第四章 基于仿射对应的人脸初始姿态估计	36

4.1.	概述.....	36
4.2.	人脸初始姿态估计.....	37
4.3.1.	构造特征点正面投影.....	38
4.3.2.	计算仿射参数.....	42
4.3.3.	基于仿射对应的空间姿态参数粗略估计.....	43
4.3.4.	人脸姿态参数迭代求精.....	53
4.3.	鲁棒的姿态参数估计方法.....	54
4.4.	实验结果和分析.....	55
4.4.1.	人脸平面假设的验证实验.....	55
4.4.2.	模拟数据实验和分析.....	56
4.4.3.	石膏像图像与真实人脸图像实验.....	58
第五章	基于深度灰度约束加权的姿态跟踪.....	60
5.1.	概述.....	60
5.2	用 KLT 准则生成新的特征点.....	61
5.3	加权的深度和灰度约束方程.....	64
5.4.	基于快速傅立叶变换的二维特征跟踪.....	68
5.4.1.	Lucas-Kanade 类方法.....	68
5.4.2.	基于 FFT 的改进的 KLT 方法.....	69
5.5.	计算特征点的权值.....	70
5.6.	计算特征点的曲率权值.....	71
5.7.	帧间运动参数估计.....	73
5.8.	特征点动态更新.....	74
5.9.	实验结果和分析.....	76
5.9.1.	基于 FFT 的 KLT 特征点跟踪实验分析.....	77
5.9.2.	模特头像图像序列姿态跟踪实验分析.....	78
5.9.3.	真实人脸图像序列姿态跟踪实验分析.....	81
第六章	结束语.....	84
6.1.	总结.....	84
6.2.	展望.....	85
附录 I	Levenburg-Marquett 算法	86
图表目录	88
参考文献	90
已发表和录用的论文	97
参与的科研项目	97

致谢	98
----------	----

第一章 绪论

1.1. 人体运动的视频分析简介

人体运动的视频分析是智能人机交互和生物测定学的一个重要的研究领域，其主要的研究内容是通过一个或多个摄像机获取人体运动的图像序列，并在复杂的环境中对不同的人或者人体各部分，如脸、手、手指、脚等等，进行检测、跟踪并对其运动进行分析。

人们对复杂环境下人类运动和行为的分析和理解的兴趣由来已久，有许多具体应用与此相关，如人体各部分的运动分析、人脸检测[1,2,3]、人脸识别[4,5,6]、人脸跟踪[7]、手势分析[8,9,10]、步态识别[11,12]和表情识别[13,14,15]等，这些应用大都和智能人机交互、安全系统或者生物测定学研究有关，通常侧重于层次较低的人体检测、人体跟踪和人体局部的运动分析等等。另一类应用则将人看成一个整体，试图从较高的层次上对整个人或人群的行为进行分析和理解[16]，这类应用包括对环境中人的行为进行监控，从人的各种反应中发现并理解潜在的问题或者可能的因素，识别特定人或者一般人的某种行为模式，一般来说，这类应用往往建立在对低层次的人体运动检测和跟踪得到的信息进行综合分析的基础之上。除此之外，还有一类应用就是所谓的人类运动或行为建模，即建立某种模型对人体的各种运动和行为进行模拟。这类应用包括生成具有真实感的人脸图形和动画，医用数据(如 CT 或者 MR 数据)的可视化和虚拟运动分析与合成等等。

综上所述，人体运动视频分析有三个主要的研究领域：(1) 人体各部分的检测、跟踪和运动分析；(2) 作为一个整体的人的运动分析以及在更高层次上对人的行为进行分析和理解；(3) 对人体的运动和行为进行计算机建模，通过虚拟的模型来模拟现实。这三个研究领域是相辅相成的，只有对人体的运动和行为进行深刻的分析和理解，才能使计算机模型更加接近现实世界；反过来，通过逼真的模型才能模拟出现实世界中很难出现或是无法出现的情况。其中，人体的运动分析是实现虚拟世界和现实世界自由转换的中心环节。

对视频序列中的人体部位进行跟踪是对人体运动分析最基本的步骤，其主要

目的是在图像序列各帧中定位人体部位的位置和姿态，跟踪的结果可以用来对人体的运动和行为进行定量的数学分析。人体的运动，既可以看成是摄像机运动人静止，或是人运动摄像机静止，或是人和摄像机都在运动的结果。跟踪技术总体上分为二维跟踪和三维跟踪两大类。二维跟踪计算人体在每一帧图像上的二维位置；三维跟踪则求解出描述人体在三维空间中位置和姿态的三维运动参数。更深一层次的跟踪还包括对物体的变形进行跟踪。在跟踪的过程中，我们常常用到各种标记、相关性的度量、颜色或形状约束等。目前主要有两种跟踪运动人体的方法，一是基于运动的方法；一是基于模型的方法。基于运动的方法认为物体的运动随着时间变化存在某种一致性，通过鲁棒方法来求解。这种方法一般速度较快，不过无法保证跟踪的区域具有语义上的意义。基于模型的方法则将高层的语义知识赋予运动模型，往往计算量较大，往往还需要考虑比例变化、平移、旋转和变形等因素的影响。前一种方法需要提取某些区域的特征，而后一种方法则需要提供某些几何信息。

人脸是人体上最引人注目的部位，也是人与人之间交流思想，表达意愿的重要部位，因此人们对人脸的跟踪兴趣远远超过对其他部位的兴趣。不论是过去还是现在，不论是国内还是国外，对人脸跟踪的研究无疑是人体运动分析当中最重要和最有吸引力的内容。

1.2. 人脸空间姿态跟踪及其应用

人脸姿态估计（Face Pose Estimation）是指在摄像机获取的人脸图像序列中确定人脸在三维空间中姿态的技术和方法。人脸姿态估计作为智能人机交互和计算机视觉研究中的一项基本课题，近年来正不断引起人们的兴趣。人脸姿态估计技术在实践中，特别是在智能人机交互、基于模型的视频会议编码，虚拟现实、智能监控等方面都有广泛的应用前景。

智能人机交互是目前计算机和人工智能研究的热点。为了打破计算机与人（尤其是没有掌握高深的计算机专业知识的普通用户）之间的交流障碍，科学家们正在努力使计算机具备人类在视觉听觉等方面的某些功能。当前普遍采用的人机交互方式仍主要依赖于键盘和鼠标，其主要特点是以计算机为核心，让人来适应计算机，因此对于普通用户来说难于掌握，效率较低。因此人们迫切需要改变

目前使用计算机的方式，要求以人类习惯的、更自然的方式与计算机进行交流，从而实现以人为核心的应用模式，使计算机能够主动地适应人的要求，这正是智能人机交互研究要达成的目标。为实现该目标，不仅需要有硬件技术方面的发展，如计算能力的提高、显示技术的进步和各种智能接口设备的出现；而且还需要在语音分析与合成、人脸检测、人脸识别和验证、表情识别、人体运动分析、人体建模与动画等理论和算法研究方面的进步。最终目的就是使人们最终摆脱键盘和鼠标的束缚，使计算机更加智能化、人性化，从而能更好地为人们的生活、工作和学习服务。由于视觉信息具有直观、信息量大、易于保存等特点，并且硬件技术的快速发展使廉价的摄像设备（如web camera）的不断涌现，越来越多的人拥有了桌上视频系统，加上计算机视觉和图像处理技术的进步，因此如何使计算机视觉系统更好地理解人的行为并与人进行智能化的交互就成了亟待解决的问题。

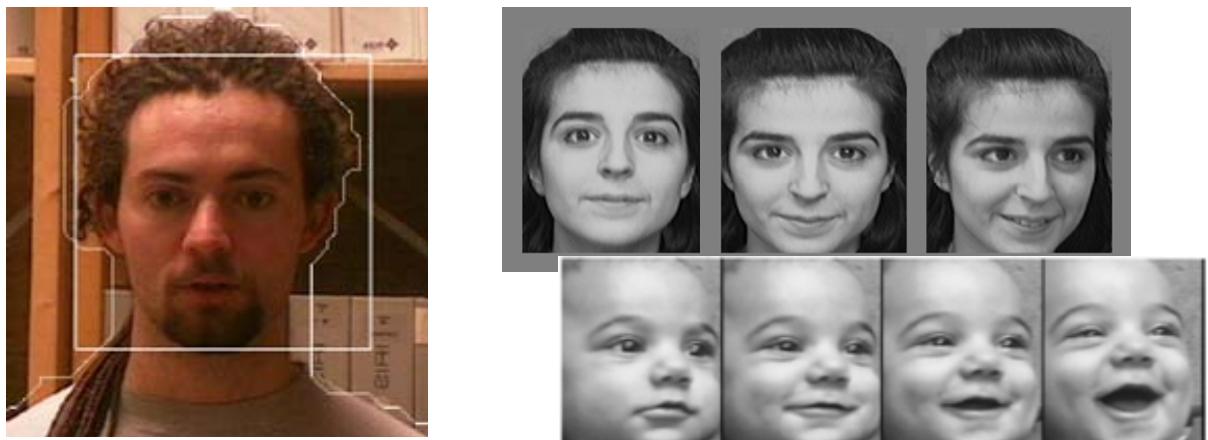


图 1-1 人脸识别和表情识别

众所周知，智能人机交互研究中的一个重要问题是要求准确地判断人在某一时刻注意力的焦点，从而使计算机更好地理解人的行为并作出相应的反应。人脸作为人体上最重要的部位之一，在人与人之间沟通与交流中扮演着重要的角色，其空间姿态对于表达情感、揭示心理状态和确定其注意力的焦点都具有非常重要的作用，我们常说的“垂头丧气”，“仰面大笑”等等就是对这种作用的生动描述。因此准确地判断人脸空间姿态对于理解人的行为具有非常重要的意义。

除此之外，许多智能人机交互应用，如人脸识别和表情识别（图 1-1）和视线跟踪（图 1-2），往往需要使摄像机和人脸之间保持某种特定的空间位置关系。以人脸识别为例，如果摄像机和人脸之间位置不合适，很可能造成人脸许多

部分出现遮蔽和变形，甚至完全不可见，从而影响人脸识别的可靠程度。

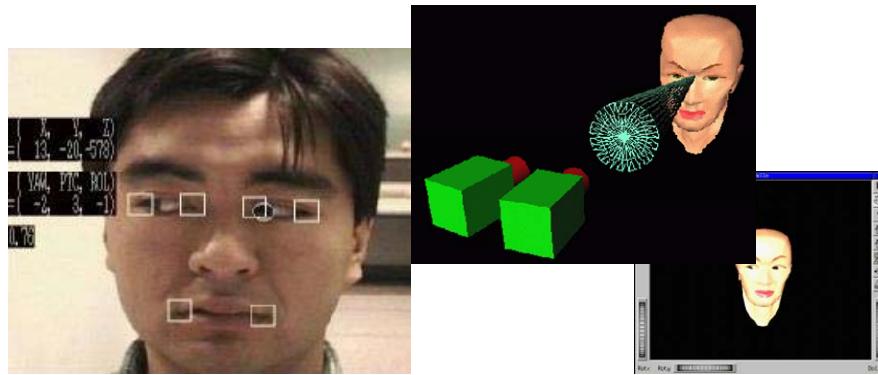


图 1-2 视线跟踪

此外，某些基于对称性的人脸识别方法也要求获得正面的人脸图像，这时就可以应用人脸姿态估计技术求得人脸的空间姿态参数，并调整摄像机的姿态参数使之和人脸之间保持一定的位置关系。对表情识别而言也是如此。对一些要求进行视线跟踪的应用，如果知道人脸的姿态，那么就能更容易地确定视线的方向。

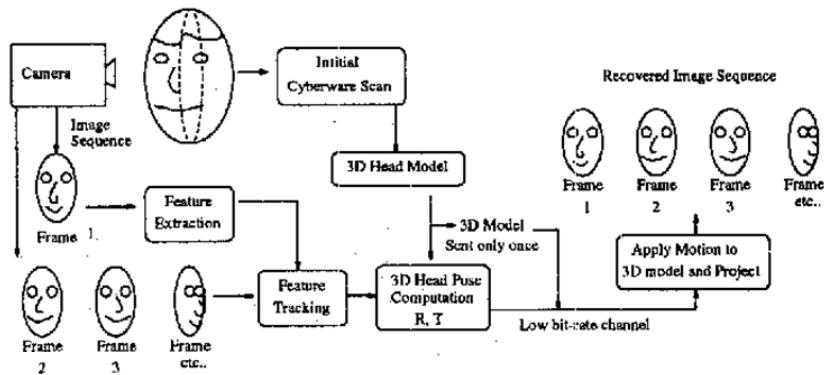


图 1-3 基于模型的视频会议系统

在基于人脸模型的视频会议系统中（图 1-3），与会者的人脸模型可以预先获得并传送到远端，然后在会议的进程中可以应用人脸姿态估计技术得到真实人脸的姿态参数，这样只需传送少量的姿态参数就可以在远端恢复出真实的人脸姿态，在网络带宽有限的情况下能够有效地减少传送大量图像数据给网络带来的负载。在安全监控系统（图 1-4）中，摄像机往往不能直接捕捉到被监控对象的正面。从而给监控工作带来困难。如果能获得人脸的三维空间姿态参数，就能够动态地调整摄像机的姿态使之始终处于对监控对象最佳的观测位置。此外，利用人脸空间姿态参数还能够使多个监视摄像机在时间和空间上协调合作，实现对被

监控对象的连续追踪。

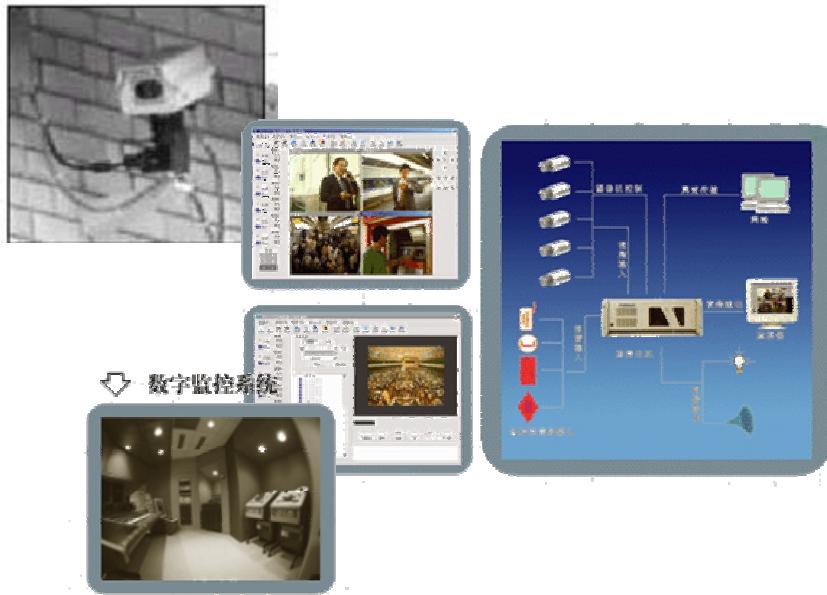


图 1-4 安全监控系统

娱乐和游戏是计算机技术最有吸引力的两个应用领域，尤其近年来网络游戏的盛行给新一轮网络经济带来了新的利润增长点。在三维游戏中，逼真的效果和简易的交互方式一直是玩家们梦寐以求的理想，目前基于游戏棒、游戏盒的交互方式远远不能满足玩家们的要求。通过计算人脸空间姿态参数，能够以图像对虚拟世界中的人物进行驱动，从而为游戏玩家们带来更真实的游戏体验。此外，在主题乐园的交互式的娱乐项目中，可以根据观众脸部的空间姿态有针对性地调整周围环境的音响和画面等效果，使观众获得身临其境的感受。

司机疲劳检测是人脸姿态跟踪另一个有价值的应用。司机驾驶的过程中可能出现疲劳，情况严重时可能导致事故。通过人脸姿态跟踪，我们能够发现司机可能存在疲劳的趋势并及时发出警报，从而能够避免事故的发生。

此外，还有一些基于心理学研究和应用也要求比较精确的人脸姿态估计。总而言之，人脸姿态估计技术的研究不仅有学术研究价值而且还有重要的应用价值。

1.3. 论文的主要工作和结构

现有的人脸姿态跟踪方法大体上可以分为基于特征的方法和基于模型的方法。基于特征的方法缺点在于需要大量不同姿态的图像样本而且结果不太准确。基于模型的方法相对易于实现，精度较高，可是基于三维/二维特征对应求姿态

参数，往往需要求解维数较高的非线性方程，需要较好的初始估计值，容易出现陷入局部最小和不收敛的现象，还容易受到光照变化和遮蔽的影响。这些问题构成了人脸姿态跟踪问题的难点。本文针对仅有一个摄像机获取的单目视频图像序列的情况，尝试将人脸三维模型引入人脸姿态跟踪系统框架，并解决了姿态跟踪中一些关键的技术问题。我们的应用系统框架包括初始姿态估计和帧间姿态跟踪两个子过程。首先用仿射变换的方法得到初始帧的人脸姿态参数并作为姿态跟踪的起点；然后用三维几何信息对线性灰度和深度约束方程加权得到更精确的帧间运动参数；为了消除光照变化和遮蔽的影响，在跟踪过程中逐帧自动进行特征点更新。对模特头像和真实人脸进行的实验结果表明该算法能在较大的变化范围内实现可靠的人脸姿态跟踪，尤其是对深度方向有明显变化的运动，该算法的作用表现得更加明显。

后续章节中，我们将陆续介绍已有的研究工作、系统框架的构成以及各个子过程的主要思想、算法实现并给出相关的实验结果分析。全文的结构安排如下：

第二章介绍了目前人脸姿态跟踪的两种主要方法，给出了相关的结果并对各种方法的优点和不足之处做了分析。

第三章介绍了我们提出的基于三维模型的人脸空间姿态跟踪框架，说明了框架内的各个组成部分及其之间的关系，并叙述了初始化工作中的人脸三维建模、摄像机标定方法和系统所用的透视投影模型。

第四章介绍了基于仿射对应原理的初始帧姿态估计的方法。重点分析了仿射对应的姿态参数估计的原理、通过非线性优化过程对姿态参数求精以及为了提高估计精度使用的鲁棒方法，最后给出了平面假设的验证实验结果、基于模拟数据的实验结果、基于石膏像图像和真实人脸图像的实验结果，并对实验结果做了比较详细的分析。

第五章介绍了初始估计完成之后利用帧间的深度和灰度约束加权进行姿态跟踪的方法。主要叙述了线性深度灰度约束的原理、权值的定义和计算方法、改进的基于FFT的KLT特征跟踪方法原理、基于三维模型的帧间的运动参数估计方法并提出了一种逐帧进行的特征点自动更新方法。最后给出了基于FFT的KLT特征跟踪方法的实验结果、基于模特头像和真实人脸图像序列的实验结果，对实验

结果进行了详细的分析，并给出了相关的结论。

第六章总结全文，就仍然存在的问题进行了分析，并且对未来研究工作进行了展望。

第二章 人脸姿态跟踪方法概述

2.1. 概述

目前国内外对于人脸姿态估计的课题已有大量的相关工作，每年都有的许多论文问世。本节将简要地介绍现有的一些方法及其特点，详细的讨论将在后面几章陆续进行。

现有的方法大体上可以分为两类：基于人脸特征的方法和基于模型的方法。

基于人脸特征的方法认为在人脸空间姿态和人脸图像的某些特征（如灰度，色彩，图像梯度等等）之间存在着某种特定的对应关系，因此通过大量的样本来建立这种关系。如 Chen[17]将人脸看成是由肤色和发色区域组成，然后根据大量的样本建立这些区域的特性（面积，质心，惯量）与相对应的空间姿态之间的对应关系。然后对输入图像提取这些区域的特性并进行匹配从而得到结果。类似的工作包括[18,19,20,21,22,23,24,25]。一般来说，这类方法需要大量的样本，结果往往不够准确。

基于模型方法的基本思想是利用某种几何模型来表示人脸的结构和形状，并通过提取某些特征，在模型和输入图像之间建立起对应关系，然后通过几何或其他的方法实现人脸空间姿态的估计。这里所使用的模型既可能是简单的几何形体，例如平面[26,27]、椭圆[28]、圆柱[29,30,31,32]，也可能是某种简单的几何结构[33,34,35,36]，也可能是通过激光扫描或其他方法获得的人脸三维模型[37,38,39,40,41,30]。由于简单的几何模型或结构不能准确地描述人脸的真实形状和特征，往往要求一些先验知识或者是使用正则化方法减小误差，因此，基于人脸三维模型的方法越来越受到重视。Lopez 等人[38]利用三维模型和图像上的多达 28 组特征点对实现人脸姿态的估计。该方法需要预先建立一个庞大的、包含各种姿态在内的模板库，计算开销很大，难以实时应用。Yang 等人[39]将三维模型引入一个实时的立体视觉系统，实现了精度较高的人脸姿态估计。该方法需要两台摄像机，并要求预先用半交互式的方法用立体视觉系统建立一个有真实感的人脸三维模型。Shimizu 等人[37]则定义了一个由统计方法得到的通用人脸模型，使用三维模型和图像上特定的曲线之间的对应关系并通过优化迭代过程得到人

脸姿态的估计。这种方法需要提取人脸图像上某些特定曲线并实现三维模型上的曲线与图像上的曲线之间的匹配。

近年来，随着深度数据获取设备（如激光扫描仪）的广泛应用，能够获得精度较高的人脸三维模型。由于三维模型包含丰富的几何信息，因此完全可能用较少的特征点实现人脸姿态的估计，得到精度更高、更鲁棒的结果。对于许多应用场合，如医院智能护理系统、视频会议系统、司机疲劳检测系统等，对象（如病人、与会人员、司机等）的三维模型都能够预先获得，而且还可以进一步考虑基于统计方法构造出通用的人脸三维模型并应用到不同个体。这些都使基于人脸三维模型的方法具有很好的应用前景。

此外还有一些基于立体视觉[42,43,44,45]的人脸姿态估计方法。其基本思想是用两台摄像机构成立体视觉系统[46]，利用一些约束条件（通常是极线约束条件）将特征跟踪过程中误差较大的特征点除去，在左右两幅视图上得到比较可靠的特征点对，然后利用这些点对通过线性或非线性的方法得到人脸的空间姿态参数。在此就不做详细介绍了。

2.2. 基于特征的方法

基于特征的方法假定在人脸空间姿态和人脸图像的某些特征之间存在着某种特定的对应关系，这种关系往往用精确的数学关系式很难描述，或者是难以求解。为了确定这种关系，利用大量的已知姿态的人脸图像通过统计方法建立这种关系。图像的特征包括灰度，色彩，图像梯度，或者是图像灰度的某种变换（如图像在特征空间的投影），还可能包括图像某些特定区域的一些几何特性。

2.2.1. 基于神经网络的方法

一些基于特征的方法采用人工神经网络来构建三维空间姿态到二维图像的映射并以此来估计输入人脸图像的空间姿态。Hogg[19]将物体在空间的旋转看成是由绕 X、Y、Z 轴的旋转组成，对每种旋转都采集大量旋转角已知的二维图像作为原型图像，然后通过协作神经网络学习算法将这些旋转映射到 N 个两维的顺序参数空间(Order Parameter Space)，每个顺序参数空间对应于一个旋转角度域，并得到一条映射曲线。如图 2-1 所示。对于待求姿态的输入图像，将其用

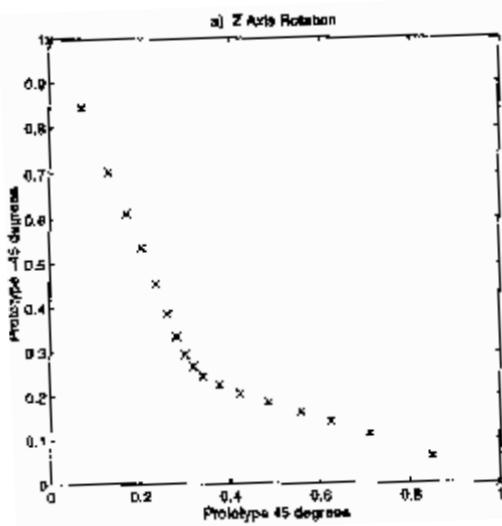


图 2-1 基于特征空间的方法

2.2.2. 基于特征空间的方法

另一种基于特征的方法是利用特征空间的性质，其基本思想是假设人脸空间姿态和人脸图像在特征空间的投影之间具有某种特定的对应关系，然后利用大量的图像样本得到这种关系。

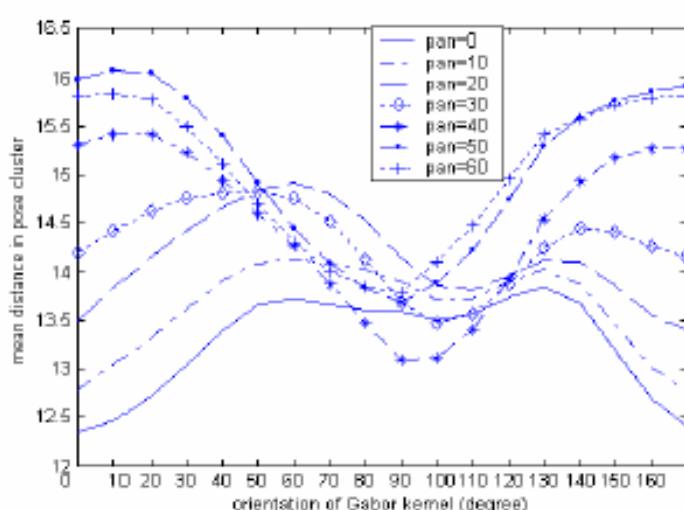


图 2-2 Gabor 滤波器的效果

照条件下得到的局部特征空间。在全局特征空间中，对每个人用三次样条插值算法得到对应于该人的一个连续的超表面，该曲面是空间姿态参数和光照参数的函数。同样在每个人的局部特征空间中，也得到这样的超表面。这样对于任一待求姿态的输入人脸图像，求解姿态参数的过程包括两个步骤，首先将其变换到全局特征空间中得到其对应的点 z ，求与该点距离最近的超曲面，从而确定输入图像

同样的方映射到这 N 个空间中去，选择能使该输入图像的映射点和映射曲线之间距离最短那个顺序参数空间所对应的角度域作为估计的结果值。Rae[21]等人也提出了一种基于神经网络的姿态估计方法，但该方法只能在实验室环境下工作，并且无法对训练图库中没有出现过的人进行姿态估计。

Murase[20]提出一种基于 PCA 的方法。首先获得 N 个人在 M 种不同的姿态和光照条件的组合下的图像，然后对这些图像用 KL 变换构造了两种特征空间，一个是所有 N 个人在 M 种不同的姿态和光照条件下得到的全局特征空间，另一种是对每个人在 M 种不同的姿态和光

属于哪个人。第二步将输入图像变换到该人的局部特征空间中，得到对应的点 z^p ，然后求出使 z^p 与超曲面距离最近的姿态参数和光照参数。显然该方法无法对没有在样本图库中出现的人进行姿态估计。

Wei[47]等人的方法则克服了这个困难，他们引入 Lee[48]提出的 Gabor 滤波器来消除不同的人脸形状的差异以及光照的影响从而得到精度比较高的姿态估计结果。首先对大量旋转角度已知的人脸图像用不同转角的 Gabor 滤波器滤波，得出结论：对于姿态接近正面的图像，水平 Gabor 滤波器能使其在 Gabor 特征空间的投影点的内聚性较好；对于其他姿态的图像用垂直 Gabor 滤波器的效果较好。如图 2—2 所示。

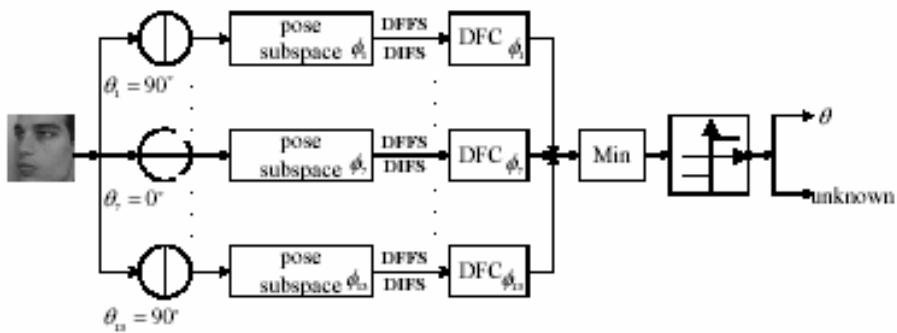


图 2—3 在 Gabor 特征空间中估计人脸姿态

假设每种已知的姿态在 Gabor 特征空间中的投影点满足多维高斯分布，引入 DBPM(Distributed-based pose model)模型描述 Gabor 特征空间中的各种姿态图像对应的投影点。这样任一输入人脸图像的姿态就可以由其在 Gabor 特征空间中的投影点与各个已知姿态的统计中心点之间的 Mahalanobis 距离决定。由于在高维的 Gabor 特征空间中直接求 Mahalanobis 距离非常困难，因此用 PCA 求得低维的特征子空间，定义两种距离 DIFS(distance in feature space) 和 DFFS(distance from feature space)，前者是低维的特征子空间上的 Mahalanobis 距离，后者则是欧氏距离。用由这两个量组成二维距离向量，就构成了一个描述人脸姿态的二维空间。这样求解任一输入人脸图像姿态的过程就包括两步（如图 2—3 所示），第一步是将该图像映射到各个已知姿态的二维子空间，第二步求使其映射点到各姿态的统计中心点的欧氏距离最短的那个姿态作为估计结果。这种方法对不在样本库中的人脸图像也能得到较好的估计结果，并且由于应用 Gabor 滤波和 DBPM 模型在一定程度上减少了人脸不同的表情和光照对估计结果的影响。

2.2.3. 基于肤色发色特征的方法

Chen[17]等人将人脸看成是肤色区域和头发区域的组合，计算出这些区域的几何特性（如面积、质心、转动惯量等等），并将人脸的运动看成是绕 X、Y、Z 轴的转动的组合。通过大量的统计样本得到不同年龄和性别的人，绕 X、Y、Z 轴的转角和这些几何特性的关系（如图 2-4 所示）。

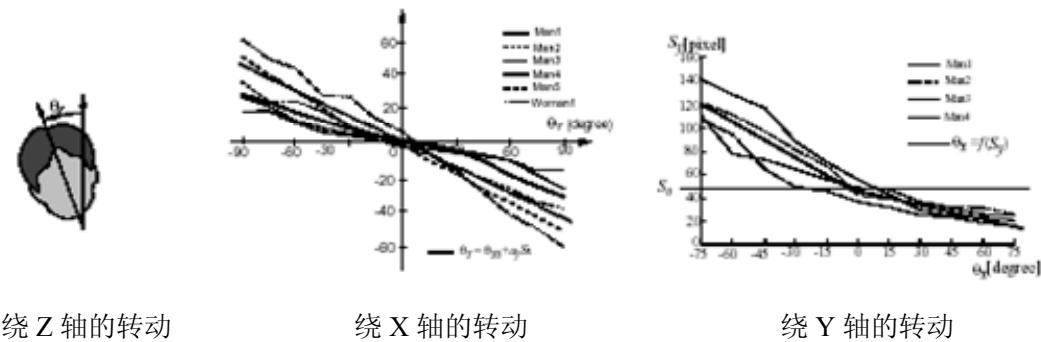


图 2-4 用肤色发色特征估计人脸姿态

Shioyama[49]也利用颜色信息来估算姿态，对肤色区域和头发区域提取是根据肤色分布模型和发色分布模型，对于人脸区域的每一个象素求出其为肤色点或发色点的相似度，并以此作为计算几何特性的加权值。

2.3. 基于模型的方法

一般情况下，由于单目图像提供的信息过少，往往难于准确地恢复人脸的空间姿态，因此有必要对人脸结构进行一些假设，提供一些附加的约束信息，从而使姿态估计的问题得到解决，这就是基于模型的方法的基本出发点。其主要思想是利用某种几何模型或结构来表示人脸的结构和形状，并对某些特征（如点、线等）在模型和输入图像之间建立起对应关系，然后通过几何或其他的方法实现人脸空间姿态参数的估计。

2.3.1. 基于特征点匹配求解空间姿态的方法

大多数基于模型的方法都是先用一些方法检测人脸上特定部位的特征点，并根据这些特征点构建相应的人脸模型，并利用特征点的二维/三维匹配从而确定人脸的空间姿态。最常用的特征点是眼角、嘴角、鼻尖等的易于探测并且特征比

较明显的部位。

究竟需要多少特征点对才能确定人脸的空间姿态呢？Fischler 和 Bolles[50]提出一种使用三个不共线的特征点的方法，他们证明了在全透视投影关系下三点对应问题(P3P)可能有多达四组解，Wolfe[51]后来给出这些解的具体形式。Huttenlocker 和 Ullman[52]则在弱透视投影的条件下证明了只用不共线的三个特征点就能确定人脸的空间姿态。Fischler 和 Bolles[50]还证明在全透视投影关系下四个特征点对应(P4P)不能够确定唯一的解，除非这四个点在一个平面上。Faugeras[53]则给出了一种直接求解空间姿态的线性方法，但该方法必须利用 6 个以上的不共面的特征点对，而且得到的旋转矩阵不能保证是单位正交的，即无法保证变换之后物体的刚性。尽管 Faugeras 后来给出了一个修正算法[53]，在一定程度上减小了误差，但仍然无法保证物体的刚性。Liu 和 Wong[55]认为 P4P 问题出现多解现象的可能性要远小于 P3P 问题，并且给出了一种在全透视投影条件下精确恢复空间姿态的方法。该方法是将连接摄像机焦点和特征点的向量的长度作为变量，并利用特征点之间的距离保持不变作为刚性约束条件，利用 Newton-Gauss 的迭代优化过程得到精确的解。这种方法必须要有一个合理的初始估计值，否则 Newton-Gauss 算法可能不收敛。

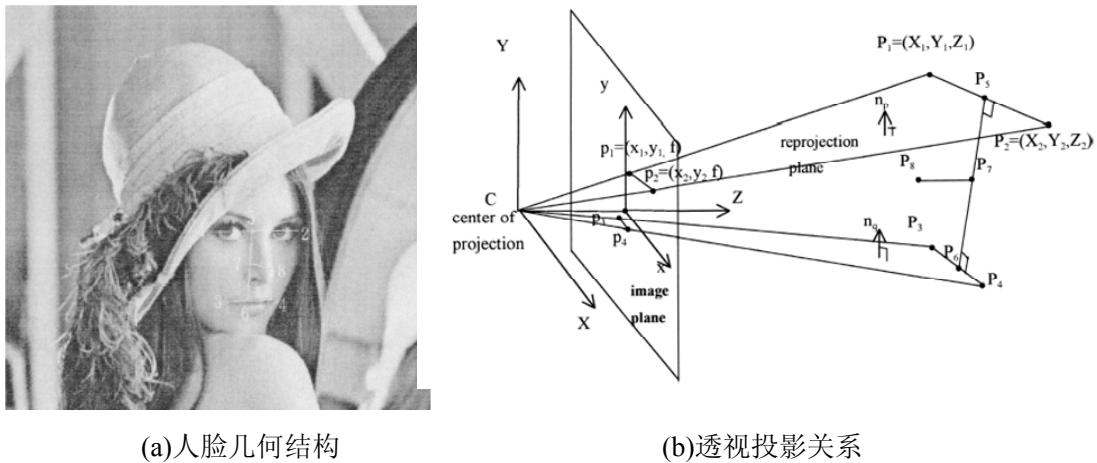


图 2-5 基于弱透视投影和简单几何结构的方法

基于上述基本原理，研究者们提出了各种人脸模型来求解人脸姿态。

2.3.2. 基于几何结构的方法

Ho [33] 使用图像中的四个特征点（两个外眼角和两个嘴角）构造一个简单的

人脸结构模型如图所示（如图 2—5(a)所示，1, 2 为外眼角，3, 4 为嘴角），并通过全透视投影关系（如图 2—5(b)所示）得到人脸空间姿态的解析解。Nikolaidis[35]则用眼睛和嘴的中点构造一个等腰三角形，并利用人脸对称性的原理求得空间姿态参数。

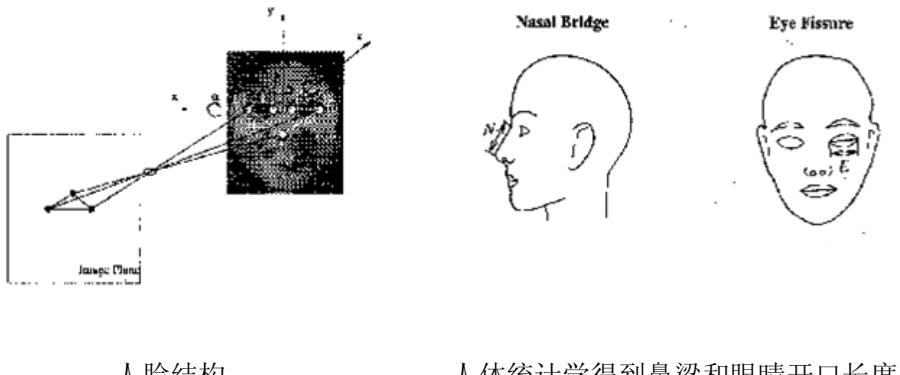


图 2—6 基于人体测量学和简单几何结构的方法

Horprasert[34]用四个共线的眼角和鼻尖共五个特征点来描述人脸结构（图 2—6），并分别求得人脸绕 X、Y、Z 轴的旋转角度。这种方法需要根据人体测量学得到的统计数据预先获得鼻梁和眼睛开口的长度。

2.3.3. 基于平面模型的方法

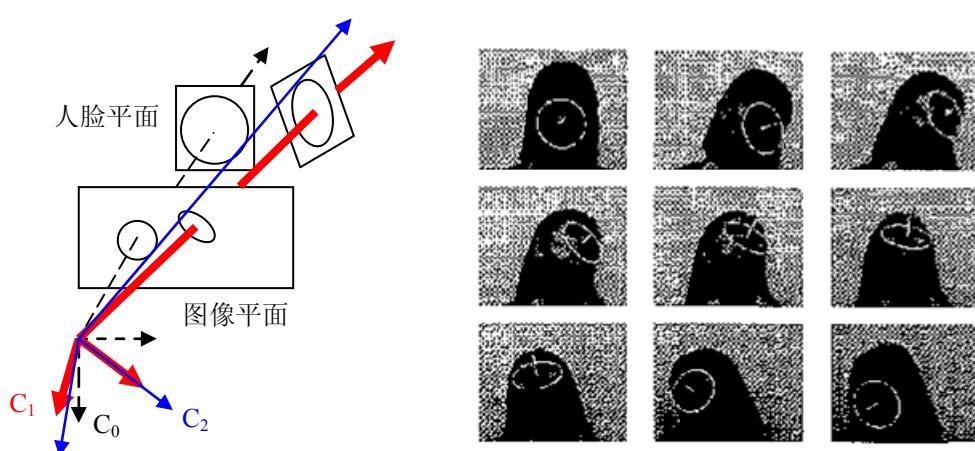


图 2—7 基于人脸平面假设和仿射对应的方法

Yao[26]近似地将人脸看成是平面，并进一步认为人脸平面在图象平面上的投影满足仿射变换的条件。通过正面平行帧和目标帧上眼睛、鼻尖、嘴等特征点位

置估算两帧之间的仿射参数，并利用圆—椭圆之间的仿射变换关系，让摄像机坐标系进行两次旋转使其 Z 轴与人脸平面的法线方向重合而得到人脸姿态。这种方法需要一幅人脸的正面平行帧作为参考帧，并且需要在摄像机离人脸较远的前提下平面假设才能成立，并需要预先通过摄像机标定过程得到焦距 f ，如图 2—7。

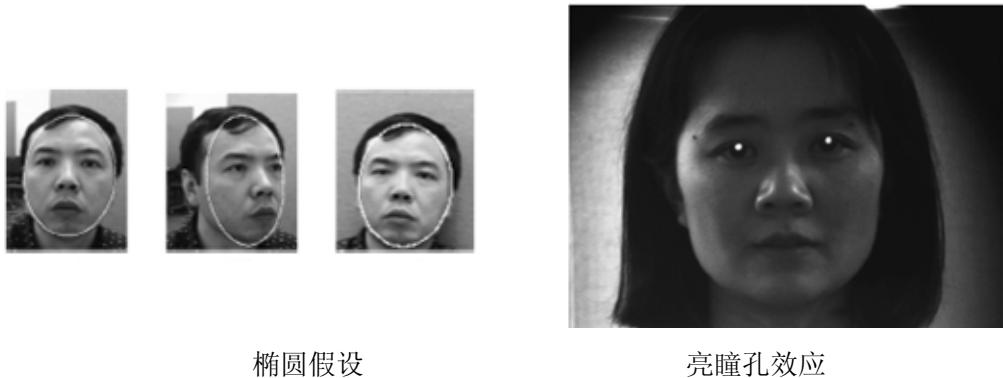


图 2—8 基于椭圆假设的方法

Ji[28]则将人脸看成是一个椭圆，这实质上是另外一种形式的平面假设（图 2—8 左）。在弱透视投影条件下进行几何分析得到人脸空间姿态。这种方法利用红外线照明产生的亮瞳孔效应（图 2—8 右）来检测瞳孔的位置。

Wang[56]将人脸看成是由多条平行线构成的，对平行直线的消失点位置进行几何分析从而得到人脸的姿态参数。该方法本质上是一种特殊的基于平面的方法，其对人脸的几何结构限制更加严格，在人脸接近正面平行状态的时候，由于消失点趋向无穷远处，因此对姿态参数的估计可能产生较大的误差。其他基于消失点的工作包括[57,58,59,60,61]。

2.3.4. 基于圆柱模型的方法

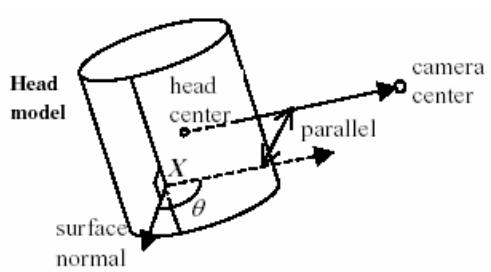


图 2—9 人脸圆柱模型

Xiao 等人[29]利用 Lucas-Kanade 方法[62,63]来求解人脸空间姿态。Lucas-Kanade 方法的本质是对三维点的空间运动引起的图像上的对应的投影点的位置变化做线性的近似，这样就可以通过相邻两帧图像模板区域象素灰度值的变化估计出

三维物体的空间运动参数。模板区域的各个象素对估计运动参数的贡献是不同的。这里存在两种情况，一是由于噪音、非刚性运动、遮蔽产生的错误的象素点，二是由于人脸表面几何形状的不一致造成的象素

点的灰度值非一致性。显然采用加权的方法可以有效地减小误差较大的象素的干扰，得到比较精确的空间运动参数。Xiao 对第一类象素采用 IRLS(iterative re-weighted least square)方法来得到相应的权值。在第二类象素中，离人脸表面的边缘比较近的象素有比较高的灰度值，由于 Lucas-Kanade 公式是一种线性假设，因而在这种情况下就会产生比较大的误差，所以应该有较小的权值。Xiao 采用圆柱模型来描述人脸（如图 2—9 所示），并以人脸表面三维点 X 的法线方向与摄像机中心和人脸中心连线之间的夹角 θ 作为权值的度量， θ 越小权值越高。

当给定实际人脸的尺度的时候，由于摄像机的焦距已通过校准过程获得，因此圆柱模型还可以用来求得人脸图像各个象素点对应的人脸三维点的坐标。并根据透视投影关系进一步得到运动参数。

这种方法的特点是算法简单，容易实现实时跟踪，但由于圆柱模型是人脸结构的粗糙近似，仅仅基于图像灰度信息得到的估计值显然不可能很精确，因此该方法在弱透视投影的情况下才能得到较好的结果，而对于中近景的应用则很难保证估计值的准确性。

2.3.5. 基于三维模型的方法

由于三维模型包含着丰富的几何信息，并且深度数据采集设备（如激光扫描仪）的性能不断提高，使获得精度较高的人脸三维模型成为可能。Lopez[38]等人提出的一种基于三维模型深度数据的方法（图 2—10），所用的三维模型是通过 Cyber Head Scanner 获得。先在人脸模型和图像上选取 28 个对应的特征点，找到所有不共线的三个特征点对的组合。用 Huttenlocker[52]方法，在弱透视投影的条件下得到空间姿态参数。

除了用特征点进行二维/三维的匹配的方法外，也有些研究者试图使用其它的人脸特征（如特征曲线）来进行匹配以求减少误差增加估计结果的准确性。

Shimizu 等人[37]构造了一个由统计方法得到的通用人脸模型（如图 2—11 所示），定义了两类边界：稳定边和变化边。稳定边在人脸空间姿态发生变化的



人脸的三维模型

选取的 28 对特征点

图 2-10 基于三维模型和特征对应的方法

时候不会变化（通常是眼睛、嘴唇、眉毛的轮廓）而变化边则会随姿态和摄像机参数的改变而改变（如人脸的轮廓）。并在模型上预先将这两类边提取出来。然后在输入的二维图像中提取这两类边界，并找到它们之间的对应关系。最后不断地减小曲线对之间的距离过程（即 ICC: Iterative Closest Curve）求得人脸姿态参数。由于曲线比点包含更多的几何信息（如曲率和长度）因而 ICC 的方法比通常的 ICP 方法更鲁棒。但是特征曲线的提取和匹配比特征点要复杂，因此这种方法的开销比较大，目前还没有看到针对连续的图像序列做实时姿态估计的结果。如图 2-11 所示。



图 2-11 提取出稳定边和变化边的人脸模型

另外一些基于三维数据的方法包括[64,65,66,67,68]。对人脸空间姿态估计技术这一研究课题，目前已有的方法大体上可以分为基于人脸特征的方法和基于模型的方法。基于人脸特征的方法需要对大量姿态参数已知的图像样本用统计方法得到估计结果，基于模型的方法则是通过定义某种几何模型来表示人脸的结构和形状，并在模型和输入图像之间建立起对应关系从而得到估计结果。由于基于模型，特别是三维模型的方法，既利用了人脸的几何信息又利用了输入人脸图像的信息，因此不需要大量的训练数据，就能得到比较精确的估计结果，因此基于三维模型的方法现在越来越受到重视，成为人脸姿态估计问题研究的重要方向。

第三章 基于三维模型的人脸姿态跟踪框架

3.1. 概述

在绪论中提到，人脸空间姿态跟踪的目标就是从摄像机获取的人脸图像序列中确定人脸在三维空间中姿态，具体说就是计算每一帧图像中人脸的运动参数。人脸的运动参数 $\Phi = (t_x \ t_y \ t_z \ \omega_x \ \omega_y \ \omega_z)^T$ ，包括三个旋转分量 $\Omega = (\omega_x, \omega_y, \omega_z)^T$ 和三个平移分量 $T = (t_x, t_y, t_z)^T$ 。

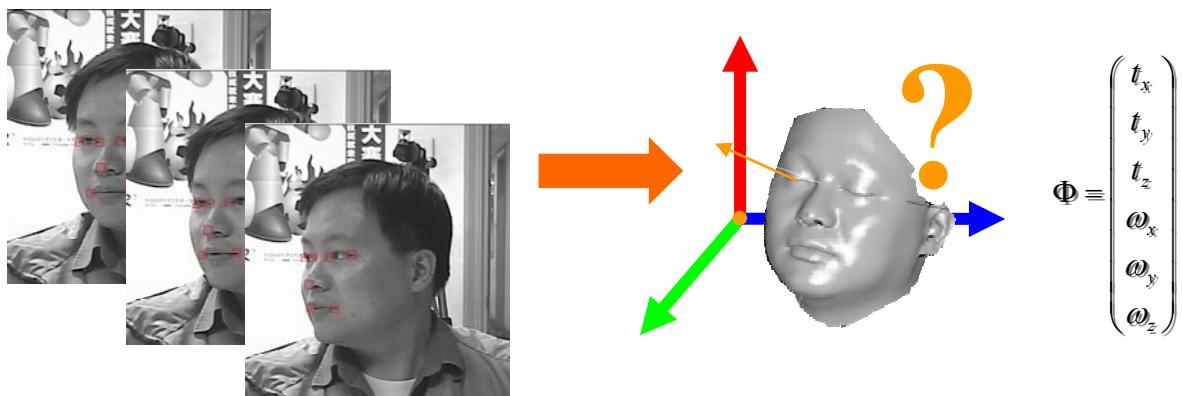


图 3-1 从视频图像序列中计算人脸的姿态参数

3.1.1. 引入人脸三维模型

上一章已经提到，基于三维模型的方法可以充分地利用人脸的几何信息从而不需要预先获得大量训练数据或者某些先验知识。近年来，随着深度数据采集设备（如激光扫描仪）的广泛应用，获得精度较高的人脸三维模型成为可能，因此可以利用三维模型丰富的几何信息，用较少的特征点实现人脸姿态的估计，得到精度更高、更鲁棒的结果。由于三维模型能够弥补单幅图像包含空间几何信息不足的困难[69]，因此可以和单目视觉系统配合来完成人脸姿态估计的任务，从而避免使用配置较复杂的双目或者多目视觉系统。从应用的角度来看，许多应用场景（如医院智能护理系统、视频会议系统、司机疲劳检测系统等），对象（如病人、与会人员、司机等）的三维模型都能够预先获得，并且还可以进一步考虑构造出通用的人脸三维模型并应用到不同个体。这些都使基于人脸三维模型的方法具有很好的应用前景，因此近年来基于三维模型的方法越来越受到重视。

目前三维人脸建模的方法，大致可以分为基于图像和基于几何这两类建模方法。基于图像的建模方法以单幅或多幅人脸图像，单目或多目视频序列作为输入数据，通过对图像数据的处理和分析，得到人脸三维模型，在处理的过程中也可能用到某种通用的人脸三维模型。相对于基于图像的建模方法，基于几何的建模方法一般通过深度数据采集设备(如三维激光扫描仪等)直接获得人脸的三维模型。

这两种建模方法各有优缺点。三维激光扫描设备得到的三维数据精度较高，有些扫描仪在扫描三维数据的同时还能够获取人脸图像并自动生产成纹理，这种功能对于许多应用系统的研究和开发很有意义；然而三维扫描设备比较昂贵、笨重不易装备、灵活性相对较小，尽管近来能够进行实时扫描的三维扫描设备已经开始出现，但仍处于研究阶段，大多数激光扫描设备还是难以获得实时的三维数据，这限制了其在实际中的应用。

基于图像的建模方法往往依赖于关于人脸的某种先验知识或图像之间的特征匹配和对应的准确程度，可能利用某种非线性优化框架得到结果，也有些基于图像的方法需要通用的人脸三维模型并定义某种变形规则或约束条件来拟合特定人脸，此外，得到的三维数据往往噪音较大，不够精确。然而基于图像的方法一般建模速度较快，某些基于立体视频序列的方法还能快速地获得帧速率的三维数据，此外，图像获取设备比三维扫描设备便宜得多，因此相当多的研究者还在致力于基于图像建模方法的研究。

3.1.2. 人脸姿态跟踪系统框架

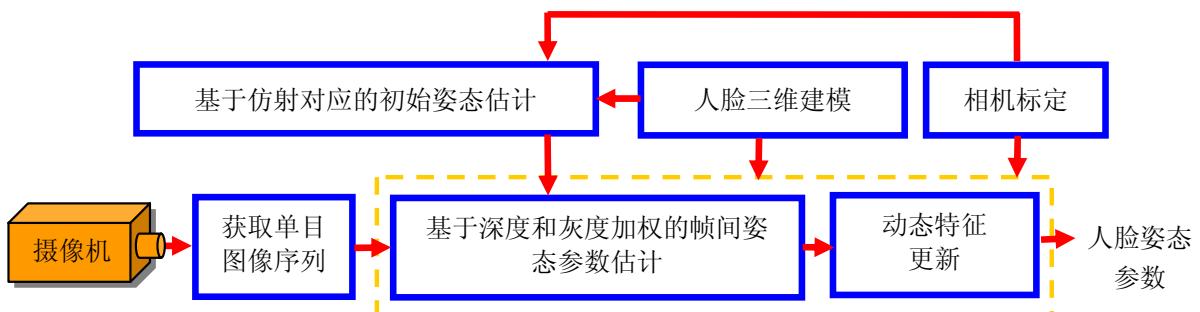


图 3-2 基于三维模型的人脸姿态跟踪系统框架

基于以上的认识，我们将研究的出发点设定为利用三维模型和单目人脸图像序列实现精确的人脸空间姿态估计并构建了一个实验系统。该系统结构如图 3-

2 所示。系统工作之前先进行一些前期的数据准备工作，包括人脸三维模型的获取、摄像机定标和初始特征点选取。

首先，我们采用一种激光扫描系统 VIVID 910 来获取人脸的三维数据。并对原始数据进行了必要的预处理。其次，本系统采用一台 MINTRON 变焦 CCD 摄像头获取人脸的视频图像。该摄像头的内参数已经通过一种自定标方法[43]获得。该方法控制摄像机进行两组移动（每组移动三次，正交移动），用线性的最小二乘法获得各个内参数。第三，特征点的初始选取。我们用手工的方式分别在三维模型和初始帧上选择一些特征比较明显的特征点，包括内外眼角和嘴角。当然可以考虑自动选择特征点的方式[70,71]。一般位于这些区域的特征点纹理信息丰富，适合进行跟踪[62,63]，我们主要进行姿态估计方法的研究，因此采用手工初始化并不会影响后面的过程。

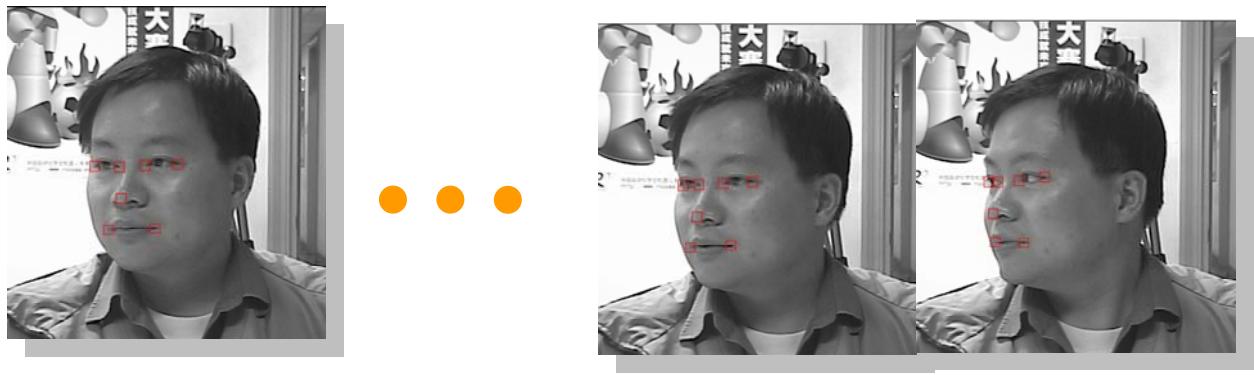


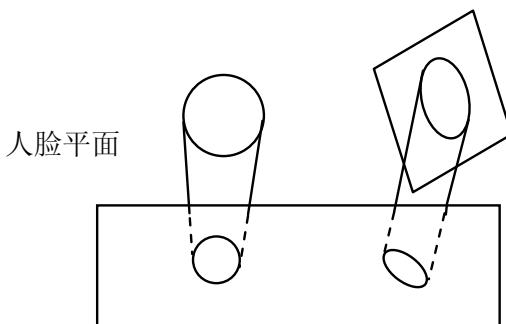
图 3-3 人脸姿态跟踪系统的两个组成部分

系统框架包括两个主要的部分，（一）用基于仿射对应的方法对单目图像序列的初始帧进行姿态估计；（二）以初始姿态估计为起点，利用帧间的约束关系进行姿态跟踪。如图 3-3 所示。

首先用一种基于仿射变换的方法[26]计算初始帧的人脸姿态并作为跟踪的起点。该方法的基本思想是选取人脸特定部位的特征点，如眼角、嘴角等，这些特征点近似位于一个平面，因此满足仿射变换的条件，然后计算出仿射变换参数并得到姿态参数的初始估计值，最后通过一个非线性的最优化过程得到姿态参数的最优解。

图 3-4 显示了该方法的基本原理。假设有一个圆位于正面平行的人脸平面

上, 它在图像平面上的投影也是圆, 当人脸姿态发生变化, 其投影就变成了椭圆。可以通过旋转图像平面和坐标轴使该投影再次变成圆从而得到姿态参数。



得到初始姿态的估计值以后, 我们利用帧与帧之间的线性灰度和深度约束关系求解帧间运动参数。上面已经提到, 基于二维/三维特征对应的方法从本质上来说是求解一个非线性最优化问题, 难于获得闭

合形式的解。如果假设帧之间运动是刚性并且幅度较小, 就可以将非线性问题简化为线性问题并且用线性最优化方法可靠地求解。此外, 在跟踪的过程中由于光照变化[27,72]或遮蔽的影响, 某些特征点可能变得不可靠, 因此在跟踪过程中逐帧地自动进行特征点更新, 这种操作能够有效地消除跟踪时间较长的情况下光照变化和运动过程中人脸表情的细微变化对姿态估计结果的影响。在以后的章节中我们将详细说明如何利用帧间的约束关系和人脸三维模型提供的深度信息实现连续的人脸姿态跟踪。

3.2. 人脸三维建模



图 3-5 人脸三维建模的过程

从人脸姿态跟踪系统的框架中, 我们可以看出, 算法中使用的三维人脸模型数据质量好坏将会直接影响到后面步骤计算结果的准确性, 因此获得精度较高的人脸三维模型是非常重要的。我们主要利用实验室购买的两种精度不同的扫描仪(手提式的精度较低的 FastScan 和高精度的 VIVID 910 获得人脸的三维数据)。

一般来说, 三维人脸建模主要包括几何建模和纹理建模两部分, 由于在我们

的研究课题中，关心更多的是人脸的几何信息，因此在我们的人脸三维建模并没有包括纹理建模的内容，主要考虑如何进行人脸的几何建模。

人脸几何建模过程包括人脸建模所需原始数据的获取及预处理。接下来，我们将简单介绍所使用的激光扫描设备；获取三维数据的方法；以及使用三维建模工具 Polyworks 对人脸几何数据的后续处理和建构人脸网格模型的过程。如图 3—5 所示。人脸几何模型的构建的完成，是下两章所要介绍的人脸姿态初始估计和人脸姿态跟踪的必要条件。

3.2.1. 人脸三维数据的获取



图 3—6 VIVID 910 激光扫描仪

我们所在的实验室有多种三维数据测量设备，包括接触型三维测量仪 MicroScribe，手持式三维扫描仪 FastScan，非接触式三维激光扫描仪 VIVID 700/910，中近景全周扫描仪 RIEGL LMS-Z210，三维激光远景扫描仪 Cyrax 2500 等。考虑扫描效率、精度和研究课题的要求，我们选择使用美能达(Minolta)公司生产的非接触式三维激光扫描仪 VIVID 910 来采集人脸几何数据。

图3—6是VIVID 910的基本构造，使用时放置在一个能够自由旋转的三角架上，对准人脸进行扫描。该扫描仪以光条方式通过一个圆柱透镜发射一条水平激光至被扫描物体。反射光线被内置CCD接受，之后通过三角化转换为距离信息。

通过使用一块Glavano镜，上述扫描过程通过水平光条在物体表面的垂直移动而重复，从而得到关于物体的三维数据。与以前的型号相比，该扫描仪能用硬件实现三维图像与彩色图像之间的纹理映射，这对于许多特定的应用是非常有意义的。

VIVID 910提供了三种类型的镜头：宽角度，中角度及远景镜头，可以根据被扫描物体的大小和扫描仪到物体的距离来选择，中角度镜头一般适合用于人脸的扫描。一些影响成像效果的参数，例如激光强度，颜色级别和白平衡等，都可以根据扫描环境的光照和被扫描对象的性质来具体设定。

获取三维数据的过程如下：首先以快扫模式试扫若干次以确定景深，之后再以精扫模式采集特定人脸数据。由于每次扫描仪扫描得到的图像都只包含人脸的某一部分，因此为了得到完整人脸的几何数据，需要从不同的方向扫描几次。

三维激光扫描仪在扫描人脸的时候存在一个问题，黑色的头发会吸收激光而导致其难以反射并被CCD接收，导致无法获得头发覆盖区域的三维数据。幸运的是，头发覆盖下的部分对我们研究的课题并不是很重要。其他一些应用，如真实感的人脸三维建模就需要采取一些特殊的措施来获得这些区域的三维数据。

3.2.2. 人脸三维数据的处理

由VIVID 910直接扫描得到的人脸三维图像只是原始的几何数据，为了应用到我们的系统中，需要对这些数据进行进一步的处理以得到一个质量较高的三维人脸几何模型。我们使用了一个三维几何建模软件包Polyworks来对原始数据进行处理，整个处理的流程包括配准、融合、编辑和压缩等几个步骤。

配准。如图3-7所示，实际扫描的数据可能是若干次不同的扫描得到的人脸不同部分的三维数据，这些数据彼此之间存在重叠的部分，因此数据处理的第一步就是将不同部分的数据统一变换到同一个坐标系下。Polyworks提供了交互式的工作方式，在重叠区域上选择对应点并通过迭代算法使坐标变换满足一定的精度要求。由于人脸在两次扫描中无法保持完全一致的表面，某些无法完全对准的部分可以通过手工剪裁的方式进行处理。

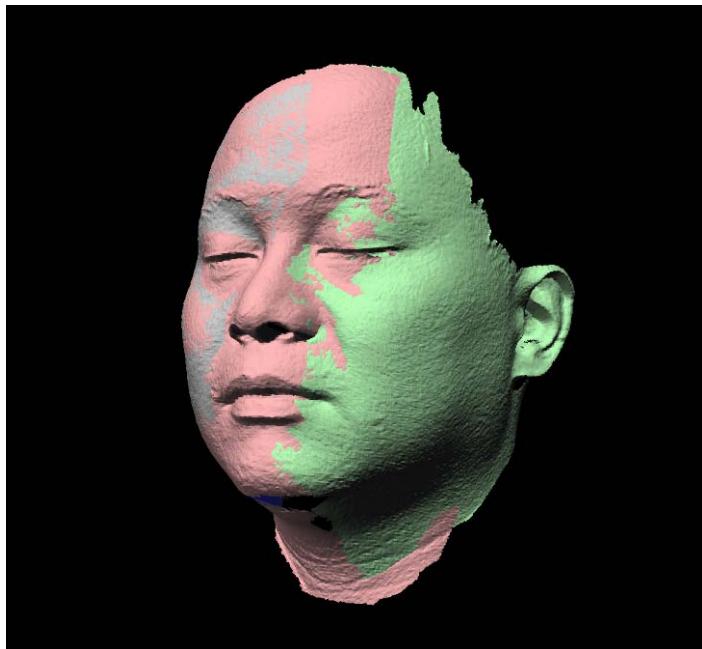


图 3-7 多块三维数据的配准

融合。完成配准操作后，使所有人脸三维数据都变换到同一坐标系下。然而在重叠区域内的三维数据仍然是由不同分块的三维数据分散构成的，为了得到单层的表面模型，需要消除重叠区域的多层数据，为此需要一个融合操作的过程将重叠区域内的多层数据合并为单层表示。最常见的方法就是设定某个阈值，并根据这个阈值进行融合操作。

编辑。完成融合操作以后，得到了一个基本为单层的比较完整的人脸三维模型。然而在扫描和处理的过程中，由于各种因素，三维数据可能存在一些缺陷：数据缺失，某些部分由于遮挡或者某些人体部位吸收激光而产生空洞；形状错误，某些几何形状复杂的部分无法被激光束有效地扫描，或是前面的配准融合操作不能使数据完全吻合，还可能是人为的因素而造成形状上的错误；拓扑错误，在建网的过程中，由于噪声数据和融合操作不彻底而造成一些网格拓扑结构的错误；此外还有噪声数据、不规则三角面片和孤岛数据的存在都会影响三维人脸模型的质量。为了解决这些问题，需要通过一些编辑操作来对数据进行处理。这些操作包括添加三角形、边交换、点合并、平滑、去噪、补洞、细分和曲面拟合等。

压缩。经过上述三个步骤的操作以后，我们可以得到一个结构完整缺陷较少的人脸三维模型，然而原始的三维模型分辨率很高，网格很致密，数据量动辄达到数十兆之多。考虑到我们研究课题的具体要求，适当进行压缩处理是必要的。

经过压缩之后的数据大小可以下降到几兆，网格数量也大大减少。

3.2.3. 人脸三维建模的结果

经过配准、融合、编辑和压缩操作之后，人脸的建模工作基本完成，最后再对三维数据做整体的平滑处理就得到了质量较好的人脸三维模型，如图 3-8 所示。

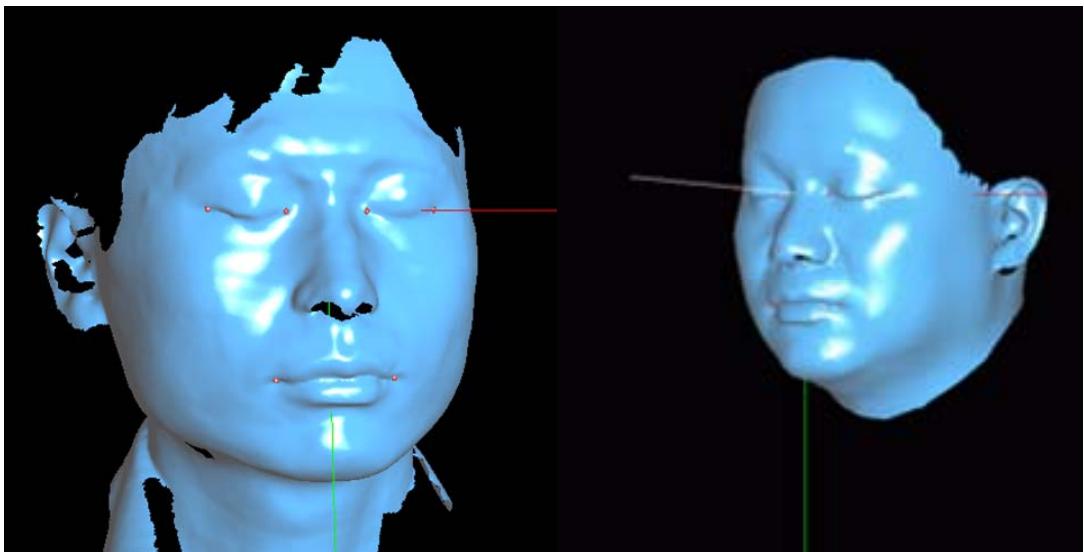


图 3-8 经过预处理之后精度较高的人脸三维数据

VIVID910 不仅能够获取三维数据，还能够实现自动的三维模型纹理映射。由于我们的系统并没有利用三维模型的纹理信息，因此我们所使用的三维模型是没有纹理的简单曲面。这种表面光滑精度较高的三维模型能够为我们的应用提供很好的深度测量值。

3.3. 透视投影模型

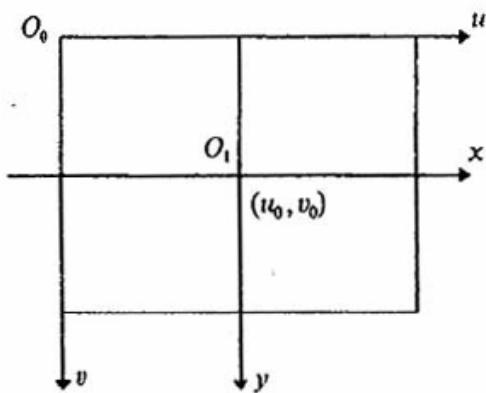


图 3-9 图像坐标系

当摄像机拍摄空间物体的时候，图像上某一点的位置与空间物体表面相应点的几何位置有关。这些位置的相互关系由摄像机的成像几何模型决定。成像几何模型是对实际的成像系统的近似，对于普通摄像机而言，最常用的成像模型就是线性模型，也叫针孔模型或透视投影模型。

3.3.1 图像坐标系、摄像机坐标系与世界坐标系

我们知道，摄像机采集的数字图像是以数组形式表示的。对灰度图像而言，每一个元素，即图像像素，其数值就是灰度值。对彩色图像而言，每个象素的亮度由红、绿、蓝三种颜色的亮度来表示。

图像坐标系是在图像上定义的一个直角坐标系，其中每个像素的坐标(u, v)分别是该像素在数组中的行数和列数，所以，(u, v)还是以像素为单位的图像坐标系的坐标。这里并没有用物理单位表示出该像素在图像中的位置，因此还需要再建立以物理单位（例如毫米）表示的图像坐标系。该坐标系以图像内某一点 O_1 为原点，x轴和y轴分别与u, v轴平行，如图3-9所示。在x, y坐标系中，原点 O_1 定义在摄像机光轴与图像平面的交点，该点一般位于图像中心处，但由于摄像机制作的原因，也会有些偏离，若 O_1 在u, v坐标系中的坐标为(u_0, v_0)，每一个像素在x轴与y轴方向上的物理尺度为 dx, dy ，则图像中任意一个像素在两个坐标系下的坐标有如下关系：

$$u = \frac{x}{dx} + u_0$$

$$v = \frac{y}{dy} + v_0$$

上式可以进一步用齐次坐标与矩阵形式表示为

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (3.1)$$

逆关系可写成

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} dx & 0 & -u_0 dx \\ 0 & dy & -v_0 dy \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \quad (3.2)$$

摄像机的透视投影成像几何关系可由图3-10表示。其中点O称为摄像机光心， X_c 轴和 Y_c 轴与图像的x轴与y轴平行， Z_c 轴为摄像机的光轴，它与图像平面垂直。光轴与图像平面的交点，即为图像坐标系的原点，由点O与 X_c ， Y_c ， Z_c 轴组成的直角坐标系称为摄像机坐标系， $O O_1$ 为摄像机焦距。

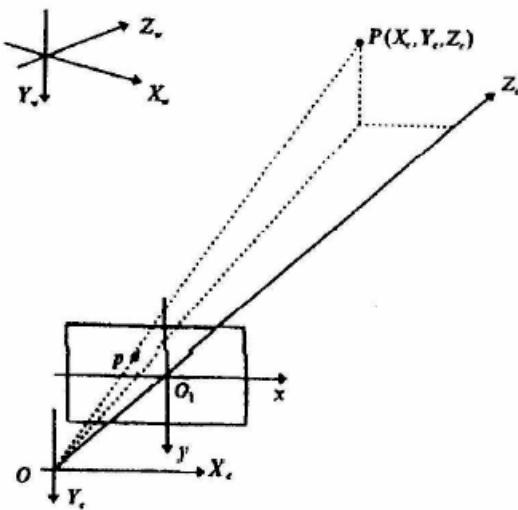


图 3-10 摄像机坐标系和世界坐标系

由于摄像机可安放在环境中的任何位置，我们在环境中选择一个基准坐标系来描述摄像机的位置，并用它描述环境中任何物体的位置，该坐标系称为世界坐标系。具体到我们的应用当中，世界坐标系其实是三维扫描仪本身所在的坐标系统。它由 X_w ， Y_w ， Z_w 轴组成。摄像机坐标系与世界坐标系之间的关系可以用旋转矩阵 R 与平移向量 t 来描述。因此，空间中某一点P在世界坐标系与摄像机坐标系下的齐次坐标如果分别是 $(X_w, Y_w, Z_w, 1)^T$ 和 $(X_c, Y_c, Z_c, 1)^T$ ，于是存在如下关系

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} = \begin{pmatrix} R & t \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} = M_1 \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \quad (3.3)$$

其中 R 为 3×3 正交单位矩阵； t 为三维平移向量； M_1 为 4×4 矩阵。

3.3.2 透视投影成像

空间任何一点P在图像上的成像位置可以用针孔模型近似表示，即任何点P在

图像上的投影位置 p , 为光心O与P点的连线OP与图像平面的交点。这种关系也称为透视投影或者中心射影。由比例关系有如下关系式:

$$\begin{aligned}x &= \frac{fX_c}{Z_c} \\y &= \frac{fY_c}{Z_c}\end{aligned}\tag{3.4}$$

其中, (x, y) 为 p 点的图像坐标, (X_c, Y_c, Z_c) 为空间点P在摄像机坐标系下的坐标, 我们用齐次坐标和矩阵表示上述透视投影关系:

$$Z_c \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix}\tag{3.5}$$

将式(3.2)与(3.3)代入上式, 我们得到以世界坐标系表示的P点坐标与其投影点 p 的坐标的关系:

$$\begin{aligned}Z_c \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} &= \begin{pmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \\&= \begin{pmatrix} \alpha_x & 0 & u_0 & 0 \\ 0 & \alpha_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} = \mathbf{M}_1 \mathbf{M}_2 \mathbf{X}_w = \mathbf{M} \mathbf{X}_w\end{aligned}\tag{3.6}$$

其中, $\alpha_x=f/dx$, $\alpha_y=f/dy$; \mathbf{M} 为 3×4 矩阵, 称为投影矩阵; \mathbf{M}_1 完全由 α_x , α_y , u_0 , v_0 决定, 由于 α_x , α_y , u_0 , v_0 只与摄像机内部结构有关, 我们称这些参数为摄像机内部参数; \mathbf{M}_2 完全由摄像机相对于世界坐标系的方位决定, 称为摄像机外部参数。

3.4. 摄像机标定

上面提到, 摑像机的参数由内参数和外参数组成。内参数是一些与摄像机本身有关的参数, 而外参数则是摄像机相对世界坐标系的运动参数。在人脸姿态估

计的过程中，我们往往预先需要知道相机的内参数，这些参数必须由实验和计算事先得到，实验和计算的过程就称为摄像机标定。一般的定标方法往往需要在摄像机前面放置一个已知形状和尺寸的物体，称为标定参照物。在很多应用场合，放置标定参照物是难以实现的，因此基于摄像机自身运动的标定方法，也称为自标定方法(self calibration)越来越受到重视。考虑到我们实验室拥有一台 Kawasaki 的机械臂，能使摄像机产生比较精确的运动，因此我们采用一种自标定方法对摄像机定标，该方法利用主动视觉和运动图像分析的思想和方法，控制摄像机进行三次正交的运动，得到摄像机的内参数[54]。首先来看摄像机在空间进行纯平移的情况，如图 3-11 所示。 I_1 和 I_2 分别是摄像机做平移运动之后得到的两幅图像， P 为空间中任意一点， p_1 和 p_2 为 P 点在两幅图像上的投影， O_1 和 O_2 为摄像机的投影中心。 p'_2 为按 p_2 在 I_2 上的坐标在 I_1 上画出的对应点，将 $p_1p'_2$ 联线称为 I_1 图上对应点的联线。由几何关系可以证明，如果摄像机做纯平移运动，那么空间所有点在 I_1 上对应点的都相交于一点 e 且 O_1e 是摄像机的运动方向。

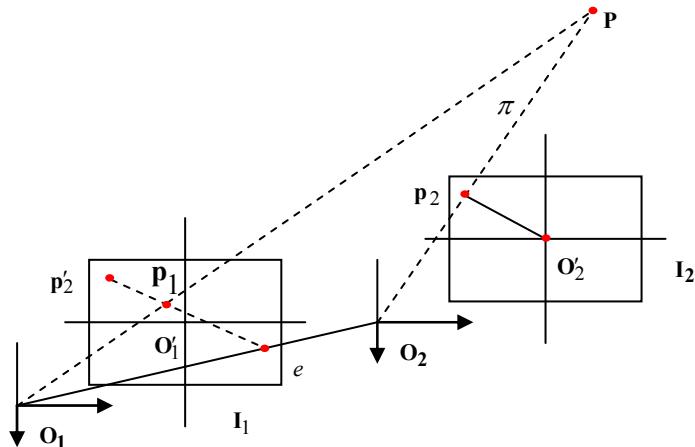


图 3-11 摄像机平移运动的几何关系

在定标的过程中，如果我们将摄像机固定在机械臂的末端执行器上并控制平台做三次纯平移运动，对于每一次平移，我们都能算出对应点联机的交点即 e_i 点，其表达式为 $e_i = ((u_i - u_0)dx, (v_i - v_0)dy, f)$ ，如果我们控制机械臂做三次移动方向正交的运动，即 $e_i^T e_j = 0 (i \neq j)$ ，那么将 e_i 的表达式代入正交条件关系式，即得到：

$$(u_1 - u_0)(u_2 - u_0)dx^2 + (v_1 - v_0)(v_2 - v_0)dy^2 + f^2 = 0, \quad (3.7)$$

$$(u_1 - u_0)(u_3 - u_0)dx^2 + (v_1 - v_0)(v_3 - v_0)dy^2 + f^2 = 0, \quad (3.8)$$

$$(u_2 - u_0)(u_3 - u_0)dx^2 + (v_2 - v_0)(v_3 - v_0)dy^2 + f^2 = 0, \quad (3.9)$$

定义两个中间变量 $t_1 = (\frac{dy}{dx})^2$ 和 $t_2 = (\frac{f}{dx})^2$, 则方程组转化为以 u_0, v_0, t_1, t_2 为变量的

非线性方程, 将(3.7)分别减去(3.8)和(3.9), 整理得到如下两个方程:

$$(u_2 - u_3)u_0 + (v_2 - v_3)v_0t_1 - (v_2 - v_3)v_1t_1 = u_1(u_2 - u_3), \quad (3.10)$$

$$(u_1 - u_3)u_0 + (v_1 - v_3)v_0t_1 - (v_1 - v_3)v_2t_1 = u_2(u_1 - u_3), \quad (3.11)$$

令 $t_3 = t_1v_0$ 则得到关于 u_0, t_1, t_3 的线性方程。但由于两个线性方程无法解出三个参数, 因此再控制机械臂进行另外三次正交运动, 同样得到两个关于这三个参数的方程, 这样四个方程就能解出 u_0, t_1, t_3 三个参数, 并由这三个参数得到摄像机的内参数如下:

$$v_0 = \frac{t_3}{t_1}, \quad \alpha_x = \frac{f}{dx} = \sqrt{t_2}, \quad \alpha_y = \frac{\sqrt{t_2}}{\sqrt{t_1}}, \quad (3.12)$$

第四章 基于仿射对应的人脸初始姿态估计

4.1. 概述

本章提出了一种基于仿射对应原理和人脸三维模型对单目视频图像序列中的初始帧进行空间姿态估计的方法。其主要思想是利用人脸的三维模型生成特征点正面平行投影，并估算输入帧和该正面平行投影之间的仿射变换参数，然后根据圆/椭圆之间的仿射对应关系得到描述人脸空间姿态的六个参数(三个旋转分量，三个平移分量)的粗略估计值，最后通过一个非线性的优化迭代算法得到精确的解。在非线性优化迭代求精确解的过程中，我们应用了鲁棒统计学的方法对二维/三维特征点对的匹配优劣程度进行估计并对优化目标函数各项加权从而提高了求解的精度。

从本质上说，基于二维/三维特征对应求解姿态参数的问题是一个非线性问题，一般难以获得闭合形式的解，往往需要通过非线性最优化的方法求解[73]，这类优化方法一般都需要较好的初始估计值，并通过多次循环迭代的过程得到最终解，如果初始估计值不准确的话还可能产生无法收敛或者陷入局部最小的情况，而且许多非线性最优化方法需要求解复杂的多阶导数构成的矩阵，计算量较大，因此不适合对图象序列的每一帧都做这样的操作。

在估计初始帧的姿态参数的时候，虽然有三维模型提供人脸几何信息，可是无法利用帧间的约束关系，因此一般情况下仍然需要利用非线性优化方法求解姿态参数。一些研究者为了避免这样的困难，就选择利用某些特殊的姿态状态，如人脸的正面平行状态，作为姿态跟踪的起点。这类方法往往用一些统计方法或者某种肤色区域分布模型得到正面人脸的检测子，并且通过一个初始姿态的检测步骤确定人脸序列中出现的正面平行状态的那一帧，再利用帧间的约束进行跟踪。这种方法的最大问题是依赖于人脸的初始状态，如果不出现正面平行状态的话，系统就无法开始跟踪；另外为了获得检测正面人脸，还需要获得大量的训练样本；除此之外，通过这种方法得到的正面平行状态只是一种近似，无法与借助人脸三维模型构造的正面平行投影相融合，因此可能会出现较大的误差。

考虑上述因素，为了避免检测正面人脸可能存在的问题，我们仍然基于非线

性最优化的方法来求解姿态参数，因为花费一些时间对初始帧进行姿态估计是可以接受的，对每一帧都进行这样的操作所带来的计算花费就无法让人接受。前面提到初始估计值对估计结果的重要影响，因此我们可以考虑用某种模型或结构来近似人脸并得到比较精确的初始估计值从而加快非线性优化算法的收敛过程。

我们选择平面作为人脸的近似，主要有两个原因，一是平面的几何性质比较简单，在透视投影的条件下与其投影之间的几何关系比较容易分析；二是对于大多数应用的情况，人脸距离摄像机的距离都远远大于摄像机的焦距，因此这种近似不会带来太大的误差。在得到初始姿态估计值以后，我们就可以不依赖平面的近似条件而用一般性的二维/三维特征对应的方法得到比较准确的姿态参数。

4.2. 人脸初始姿态估计

仿射对应算法是一种近似估计算法，这种方法有两个基本前提：一是平面假设，即当人脸离摄像机一定距离以上的时候，选择一些特定的特征点（如内外眼角，嘴角等）并近似地认为这些特征点位于一个平面上；二是刚性假设，即当人处于自然表情状态的时候，由于没有脸部肌肉的剧烈运动，所以可近似地认为人脸是一个刚性体，对于人脸初始姿态估计，我们可以认为初始帧上的人脸基本是一种中性表情的状态，与扫描得到人脸三维模型时候的状态是一致的。当然随着姿态跟踪的进行，人脸会出现一些微妙的变化，这时候完全刚性的条件就不成立了，但在摄像机采集图像足够快的时候，相邻帧之间仍然基本满足刚性约束条件，如果对每帧的特征进行可靠性判别，及时丢弃不可靠的特征点，这样基于刚体假设进行跟踪的时候就不会产生很大的误差。

Yao[26]在 2001 年提出了一种基于仿射变换的人脸姿态估计方法。该方法要求有一幅人脸的正面平行帧 (fronto-parallel frame) 作为参考帧。所谓正面平行帧就是指人脸正对摄像机，并且人脸平面与图像平面平行时得到的图像。这种方法存在两个较大的缺陷。第一，某些情况下很难得到人脸正面的图像，即使得到了也不能保证是正面平行的。这里可能存在两种情况：一是选取的特征点不一定在一个平面上，二是即使在一个平面上也不能保证这个平面与图像平面平行。第二，该方法最终只能求得人脸平面法线方向，无法恢复人脸的全部六个运动参数。原因是人脸平面法线方向并不能唯一地确定人脸平面的空间姿态。

为了解决这些问题，我们利用人脸三维模型提供的几何信息，生成了特征点正面平行投影，通过三次坐标轴的旋转恢复了全部六个姿态参数(三个旋转分量，三个平移分量)，作为优化迭代算法的初始估计值。相比较文献[26]的工作，我们在下面几个方面做了改进。第一，不再需要获取正面平行帧做参考帧，建立了一个能很好地通过各个特征点(内外眼角和嘴角)的平面，并将这个平面定义为人脸平面，这样即使人脸离摄像机较近也能满足仿射对应算法所要求的平面假设。第二，通过精确地旋转人脸三维模型使我们定义的人脸平面与图像平面平行，并在图像平面上进行透视投影，就可以得到精确度较高的特征点正面平行投影。显然，该投影是针对所选特征点及其所在平面而构造的，这显然比一般地用经验来判断某幅输入图像是否是正面平行图像更合理。第三，通过增加一次在人脸平面内的转动并进行相应的几何分析，可以得到人脸从正面平行状态运动到输入帧所处状态的全部六个运动参数(即人脸的姿态参数)。第四，利用三维模型提供的空间几何信息，通过一个优化迭代过程对从粗略估计的参数求精，这样做一方面避免了基于粗糙的平面假设得到的结果可能存在的误差，另一方面也充分了利用了平面假设得到的结果，将其作为一个非线性最优化过程的初始估计值，这样能够有效地提高了估计值的准确度并加快非线性最优化过程的收敛速度。

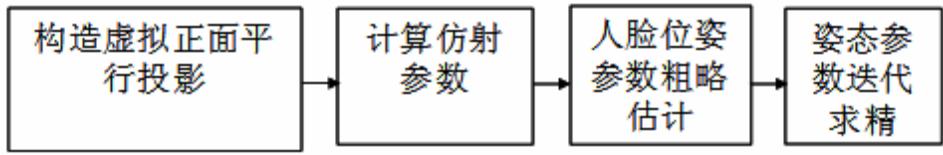


图 4-1 初始帧姿态参数估计流程图

图 4-1 是初始帧姿态参数估计的流程图。首先是利用三维人脸模型上的特征点拟合一个能很好通过各个特征点的平面作为人脸平面的定义；其次是根据圆-椭圆的仿射对应的条件，利用初始帧和虚拟正面平行投影之间的关系，计算出仿射变换参数；然后通过三次坐标轴的旋转并进行相关的几何分析，得到人脸的六个运动参数的粗略估计值；最后通过一个优化迭代的过程对粗略估计值求精。为了提高估计的准确性，在计算过程中还运用了鲁棒方法对不同特征的组合分别求解相应的姿态参数得到各个特征点的权值，并代入非线性优化方程中。

4.3.1. 构造特征点正面投影

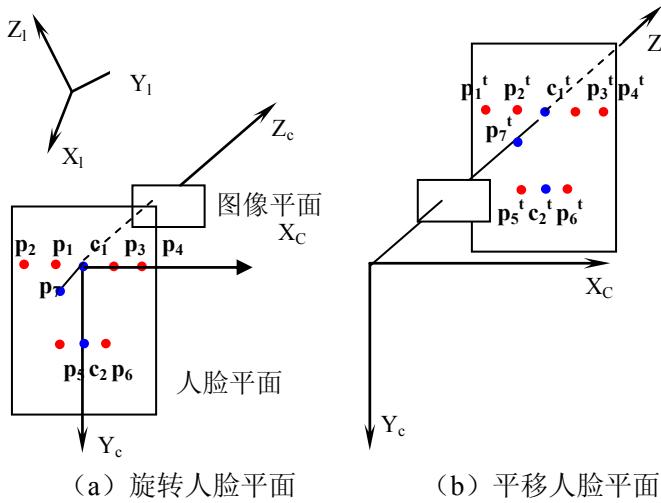


图 4-2 构造特征点正面平行投影

假设扫描得到的三维数据坐标系为本地坐标系 $\mathbf{X}_l\mathbf{Y}_l\mathbf{Z}_l$ 。不失一般性，我们可以把摄像机坐标系 $\mathbf{X}_c\mathbf{Y}_c\mathbf{Z}_c$ 作为世界坐标系，定义了 6 个特征点 $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \mathbf{m}_4, \mathbf{m}_5$ 和 \mathbf{m}_6 ，其中 \mathbf{m}_2 和 \mathbf{m}_3 为内眼角， \mathbf{m}_1 和 \mathbf{m}_4 为外眼角， \mathbf{m}_5 和 \mathbf{m}_6 为嘴角。显而易见，对大多数人而言，这 6 个特征点的位置基本在一个平面上。由这 6 个特征点可以用最小二乘法拟合一个平面，使各个特征点到这个平面的投影距离最短。

设人脸平面的方程由下式表示

$$f(x, y) = a_1x + a_2y + a_3 \quad (4.1)$$

最优化目标方程为

$$\min S = \sum_{i=1}^6 (f_i(x_i, y_i) - a_1x_i - a_2y_i - a_3)^2 \quad (4.2)$$

其中 $(x_i, y_i, f_i(x_i, y_i))$ 为特征点 \mathbf{p}_i 的坐标。用最小二乘法拟合平面得到平面参数 a_1, a_2, a_3 的值，我们计算出各个特征点在该平面上的投影点分别为 $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4, \mathbf{p}_5$ 和 \mathbf{p}_6 。设 \mathbf{c}_1 是 $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ 和 \mathbf{p}_4 的平均点， \mathbf{c}_2 是 \mathbf{p}_5 和 \mathbf{p}_6 的中点，如图 4-2(a)所示。

先将人脸模型调整成正面平行状态，然后对三维模型上的各个特征点在图像平面上做透视投影得到虚拟的正面平行投影。该操作可以通过以下三个步骤来完成：

(一) 旋转人脸模型

我们这样定义人脸的正面平行状态： \mathbf{c}_1 点与摄像机坐标系原点重合，人脸平面法线指向摄像机坐标系 \mathbf{Z} 轴负向。如 4-2(b) 所示。这样我们就可以通过构造一个旋转分量实现人脸平面在摄像机坐标系和本地坐标系之间的变换。

在 $\mathbf{c}_1\mathbf{c}_2 \times \mathbf{c}_1\mathbf{p}_3$ 上另外定义一个点 \mathbf{p}_7 ，使得 \mathbf{p}_7 和 \mathbf{c}_1 之间的距离为单位长度 1。坐标变换的结果使得 \mathbf{c}_1 位于摄像机坐标系的原点， \mathbf{p}_7 位于摄像机的 z 轴的负方向上。这样我们就可以得到 3 对不共线的三维空间点，它们在世界坐标系下的坐标和摄像机坐标系下的坐标都是已知的。因此可以求出相应的刚性变换系数（包括旋转矩阵 \mathbf{R} 和平移向量 \mathbf{t} ）。设这三个点在世界坐标系和摄像机坐标系下的坐标分别为 $\mathbf{c}_1, \mathbf{c}_2, \mathbf{p}_7$ 和 $\mathbf{c}'_1, \mathbf{c}'_2, \mathbf{p}'_7$ ，定义单位向量

$$\mathbf{n}_1 = \mathbf{p}_7 - \mathbf{c}_1, \quad \mathbf{n}_2 = \frac{\mathbf{c}_2 - \mathbf{c}_1}{\|\mathbf{c}_2 - \mathbf{c}_1\|}, \quad \mathbf{n}'_1 = \frac{\mathbf{p}'_7 - \mathbf{c}'_1}{\|\mathbf{p}'_7 - \mathbf{c}'_1\|}, \quad \mathbf{n}'_2 = \frac{\mathbf{c}'_2 - \mathbf{c}'_1}{\|\mathbf{c}'_2 - \mathbf{c}'_1\|}.$$

则 $(\mathbf{n}'_1 - \mathbf{n}_1)$ 与 $(\mathbf{n}'_2 - \mathbf{n}_2)$ 均与旋转轴 \mathbf{r} 垂直，于是

$$\mathbf{r} = \frac{(\mathbf{n}'_1 - \mathbf{n}_1) \times (\mathbf{n}'_2 - \mathbf{n}_2)}{\|(\mathbf{n}'_1 - \mathbf{n}_1) \times (\mathbf{n}'_2 - \mathbf{n}_2)\|} \quad (4.3)$$

由于根据上式求出的向量可能有两个方向，为了得到唯一确定的旋转轴和旋转角，我们不妨假设旋转角度不超过 π ，那么由 \mathbf{r} 与 \mathbf{n}_1 定义的平面法线和由 \mathbf{r} 与 \mathbf{n}_2 定义的平面法线之间的夹角也不超过 π 。定义向量

$$\mathbf{r}_1 = \frac{\mathbf{r} \times \mathbf{n}_1}{\|\mathbf{r} \times \mathbf{n}_1\|}, \quad \mathbf{r}_2 = \frac{\mathbf{r} \times \mathbf{n}'_1}{\|\mathbf{r} \times \mathbf{n}'_1\|} \quad (4.4)$$

那么唯一确定的旋转轴向量即为

$$\hat{\mathbf{r}} = \mathbf{r}_1 \times \mathbf{r}_2 \quad (4.5)$$

旋转角为 $\arccos(\mathbf{r}_1 \cdot \mathbf{r}_2)$ 。

根据 Rodrigues 公式可以确定唯一的旋转矩阵 \mathbf{R} 使人脸平面调整到正面平行的状态。 \mathbf{R} 为

$$\mathbf{R} = e^{[\mathbf{r}_a]_x} = \mathbf{I} + \frac{\sin\|\mathbf{r}_a\|}{\|\mathbf{r}_a\|} [\mathbf{r}_a]_x + \frac{1 - \cos\|\mathbf{r}_a\|}{\|\mathbf{r}_a\|^2} [\mathbf{r}_a]_x^2 \quad (4.6)$$

其中 $\mathbf{r}_a = \arccos(\mathbf{r}_1 \cdot \mathbf{r}_2) \cdot \hat{\mathbf{r}}$, $[\mathbf{r}_a]_x$ 为向量 \mathbf{r}_a 决定的反对称矩阵, 设

$$\mathbf{r}_a = (v_1, v_2, v_3)^T \quad (4.7)$$

,则 $[\mathbf{r}_a]_x$ 的具体形式如下

$$[\mathbf{r}_a]_x = \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix} \quad (4.8)$$

估计出旋转矩阵 \mathbf{R} 之后, 就可以用下式求解平移量

$$\mathbf{t} = \frac{1}{3}((\mathbf{c}'_1 - \mathbf{R}\mathbf{c}_1) + (\mathbf{c}'_2 - \mathbf{R}\mathbf{c}_2) + (\mathbf{p}'_7 - \mathbf{R}\mathbf{p}_7)) \quad (4.9)$$

(二) 沿摄像机坐标系 Z 轴正向平移人脸模型。

接下来沿着摄像机坐标系 Z 轴以任意正值 t_z 平移三维人脸模型, 如 4-2(b) 所示。后面的章节中将说明, t_z 的取值确实会影响三维人脸模型上的特征点的正面平行投影位置, 也会影响到所估算的仿射参数的线性分量, 然而却不会影响最终对人脸平面姿态的估计结果。实际上, 沿平移人脸平面只会引起其在图像平面上的投影成比例地变化, 因此平移大小的不同并不影响最终的人脸姿态参数估计值。所以不知道输入图像中人脸平面距摄像机焦点的距离真实值对于估计人脸平面姿态没有影响, 理论上取任意正值 l 即可。

在实际计算过程当中, 由于特征点对应和姿态参数的计算存在误差, 会导致不同的平移量 l 对估计结果有些影响。平移值较大的时候, 由于投影比例较大, 因此特征点的位置误差的影响就相对较小。因此我们可以用大约等于摄像机到人脸距离的值平移人脸平面, 然后进行一维搜索。我们使用进退法构造一维的搜索区间, 并用 0.618 法求得区间内能产生最小误差的那一组运动参数。这种搜索操作虽然会付出一些计算代价, 但一维搜索能够得到比假定为某个正值并进行平移时得到的结果更加准确。

(三) 透视投影

前面已经提到，如果采用线性摄摄像机模型（针孔模型），那么三维模型上的点 $\mathbf{X}=(x,y,z)$ 和图像平面上的点 $\mathbf{x}=(u,v)$ 之间的关系就可以用如下的透视投影公式表示：

$$\begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} = z_c \mathbf{P} \begin{pmatrix} \mathbf{X} \\ 1 \end{pmatrix} \quad (4.10)$$

\mathbf{P} 为透视投影矩阵， z_c 为点 \mathbf{X} 在摄像机坐标系下的 z 坐标。 \mathbf{P} 可以进一步分解如下：

$$\mathbf{P}=\mathbf{A}\mathbf{T} \quad (4.11)$$

其中 \mathbf{A} 为摄像机内参数矩阵。由于我们把摄像机坐标系看成世界坐标系，所以有 $z_c=z$, $\mathbf{T}=\mathbf{I}$ ，因此得到 $\mathbf{P}=\mathbf{A}$ 。由于内参数矩阵已经通过校准过程得到，因此由公式(4.10)可算出三维模型上的点在图像平面上的投影。这样即得到正面平行状态下的人脸模型特征点在图像平面上的投影。

4.3.2. 计算仿射参数

得到人脸虚拟正面平行投影之后，我们就可以利用当前帧和虚拟的正面平行投影之间的仿射对应关系，计算出相应的仿射参数。仿射对应的原理如图 3—4 所示。假设在人脸平面上有一个圆，当它位于正面平行状态的时候，在图像平面上的投影显然也是一个圆，当人脸平面的姿态变化到另外一个位置的时候，该圆在图像平面上的投影就变成了一个椭圆。圆和椭圆之间满足仿射变换的条件，即虚拟正面平行投影和当前帧之间相应的特征点之间存在仿射对应的关系，这样就可以用线性的最小二乘法计算出仿射参数。

用 $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_i, \dots, \mathbf{v}_N)$ 表示正面平行投影上的 N 个特征点，用 $(\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_i, \dots, \mathbf{v}'_N)$ 表示输入帧上相对应的特征点。任意的特征点对满足仿射对应公式

$$\mathbf{v}'_i = \mathbf{A}\mathbf{v}_i + \mathbf{b}, \quad i=1,2,\dots,N. \quad (4.12)$$

其中 \mathbf{A} 是仿射参数的线性部分, \mathbf{b} 是平移部分. $\mathbf{A}, \mathbf{b}, \mathbf{v}_i$ 和 \mathbf{v}'_i 分别定义为

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \quad \mathbf{v}_i = \begin{pmatrix} v_{ix} \\ v_{iy} \end{pmatrix}, \quad \mathbf{v}'_i = \begin{pmatrix} v'_{ix} \\ v'_{iy} \end{pmatrix}, \quad i=1,2,\dots,N. \quad (4.13)$$

那么求解由这两组 N 个特征点对定义的仿射变换参数即为求解下面的最优化目标函数

$$\min \|\mathbf{Km} - \mathbf{u}\| \quad (4.14)$$

其中 $\mathbf{m} = (A_{11} \ A_{12} \ b_1 \ A_{21} \ A_{22} \ b_2)^T$ 为仿射参数向量, \mathbf{K} 为正面平行投影上的特征点的坐标值构成的矩阵, \mathbf{u} 为输入帧上的特征点构成的向量,

$$\mathbf{K} = \begin{pmatrix} v_{1x} & v_{1y} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & v_{1x} & v_{1y} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ v_{Nx} & v_{Ny} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & v_{Nx} & v_{Ny} & 1 \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} v'_{1x} \\ v'_{1y} \\ \vdots \\ v'_{Nx} \\ v'_{Ny} \end{pmatrix} \quad (4.15)$$

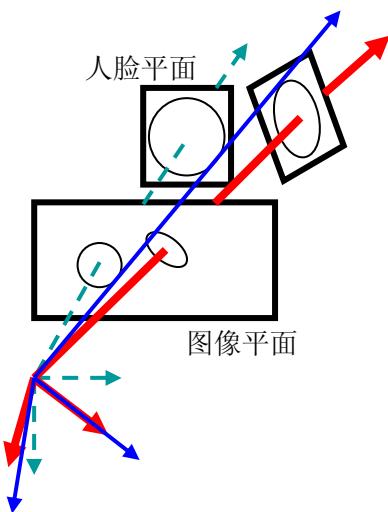
我们可以用最小二乘法求解这个线性最优化问题。得到的仿射变换参数向量可由下式计算

$$\tilde{\mathbf{m}} = (\mathbf{K}^T \mathbf{K})^{-1} \mathbf{K}^T \mathbf{u} \quad (4.16)$$

由式(4.15)可知, 仿射参数估计的准确程度取决于正面平行投影上的特征点位置和输入帧上的特征点位置的准确程度。如前所述, 我们的方法使用的特征点数 N 为 6, 其正面平行投影是通过对三维模型进行精确的旋转和平移并将模型上的三维特征点进行透视投影得到的, 因此比用经验判断的正面平行图像得到的特征点位置准确。第五章中我们将描述如何通过可靠的特征跟踪算法对每一帧都得到比较准确的特征点位置。

4.3.3. 基于仿射对应的空间姿态参数粗略估计

我们使用一种基于仿射变换的圆-椭圆对应的方法估算人脸平面的姿态参数。该方法的主要思路是: 设想在正面平行的人脸平面上存在一个假想圆, 其圆心正好位于摄像机光轴和人脸平面的交点, 显然其正面平行投影也是一个圆(由



于该圆的半径大小并不影响投影关系，因此可以假设其正面平行投影为单位圆，圆心在点 $(0, 0, f)$ 。当人脸平面的空间姿态发生变化后，它在图像平面上的投影就变成了一个椭圆。因此确定人脸平面的空间姿态参数的问题就转化成如何对摄像机坐标系进行合适的旋转使该圆的投影再次变成圆

图 4-3 旋转坐标轴三次得到人脸平面姿态参数 并且使该投影圆与人脸平面上圆的方向一致。我们可以对坐标系进行三次旋转来实现这一目标。首先构造一个椭圆锥，使它与图像平面的交线即为仿射变换之后得到的椭圆；然后对摄像机坐标系进行第一次旋转，由于椭圆锥和图像平面的交线为投影椭圆，第一次旋转的结果是使 Z 轴通过该椭圆的中心，且 X 轴和椭圆长轴重合；第二次旋转是绕 X 轴的旋转，使该椭圆锥与图像平面的交线变成圆，这样就得到了人脸平面的法线方向；第三次旋转是绕 Z 轴的旋转，目标是使旋转前后人脸平面上的圆的方向一致，为此需要定义一个标记点来标记圆的方位。第一和第二次旋转的过程如图 4-3 所示，其中 \mathbf{C}_0 为初始的摄像机坐标系(虚线表示)， \mathbf{C}_1 (粗实线表示)， \mathbf{C}_2 (细实线表示)分别为第一次旋转和第二次旋转之后的摄像机坐标系。

下面我们详细说明如何通过三次旋转得到人脸的姿态参数。

(一) 构造椭圆锥

首先假设在人脸的虚拟正面平行投影上存在一个假想圆，其中心位于点 (x_0, y_0) 上，其方程为

$$(x - x_0)^2 + (y - y_0)^2 = r^2 \quad (4.17)$$

其中 r 为圆的半径。实际上，我们将圆心设在摄像机的光轴和图像平面的交点 (u_0, v_0) 上，经过摄像机的标定，这个点的位置是已知的，即 $x_0 = 0, y_0 = 0$ 。此外，圆的半径 r 的大小并不重要，因为我们主要关心的是人脸的空间姿态，不同的假想圆半径并不会影响人脸姿态参数的估计值，为了简化计算，这

里可以取 $r=1$ 。因此上式的假想圆方程可以简化为

$$x^2 + y^2 = 1 \quad (4.18)$$

设 \mathbf{A} 和 \mathbf{b} 分别表示虚拟正面平行投影和初始帧之间的仿射参数的旋转分量和平移分量

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad (4.19)$$

因此位于虚拟正面平行投影和初始帧上的特征点对之间的位置关系由仿射变换方程决定

$$\mathbf{v}'_i = \mathbf{Av}_i + \mathbf{b}, \quad i=1,2,\dots,N. \quad (4.20)$$

其中

$$\mathbf{v}_i = \begin{pmatrix} v_{ix} \\ v_{iy} \end{pmatrix}, \quad \mathbf{v}'_i = \begin{pmatrix} v'_{ix} \\ v'_{iy} \end{pmatrix} \quad (4.21)$$

分别为虚拟正面平行投影和初始帧上的特征点坐标，N 为特征点数。将变换方程(4.20)代入式(4.18)即可得到经过仿射变换之后的椭圆方程

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \quad (4.22)$$

其中

$$A = A^2_{11} + A^2_{21}, \quad B = 2(A_{11}A_{12} + A_{21}A_{22}),$$

$$C = A^2_{12} + A^2_{22}, \quad D = 2(A_{11}b_1 + A_{21}b_2),$$

$$E = 2(A_{12}b_1 + A_{22}b_2), \quad F = b_1^2 + b_2^2 - 1 \quad (4.23)$$

为了确定人脸的姿态参数，我们需要构造一个椭圆锥，其顶点位于摄像机焦点，且其与图像平面的交线为仿射变换之后得到的投影椭圆。该椭圆锥方程如下

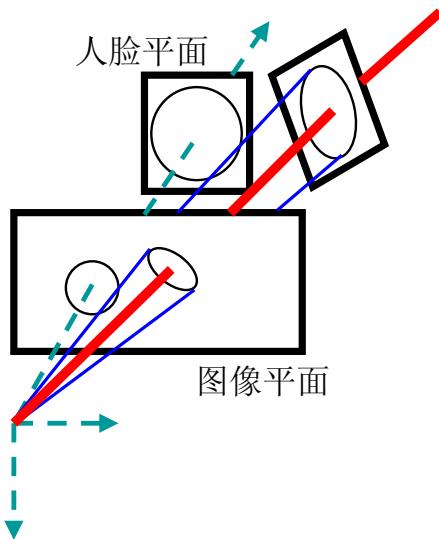


图 4-4 构造椭圆锥

$$Ax^2 + Bxy + Cy^2 + Dxz/f + Eyz/f + Fz^2/f^2 = 0 \quad (4.24)$$

其中 f 为摄像机的焦距, 如果取 $z=f$, 可得到式(4.22), 说明其与图像平面的交线即为式(4.22)确定的投影椭圆。该椭圆锥如图 4-4 所示, 由细实线所围成的曲面即为该椭圆锥。

(二) 第一次旋转摄像机坐标系

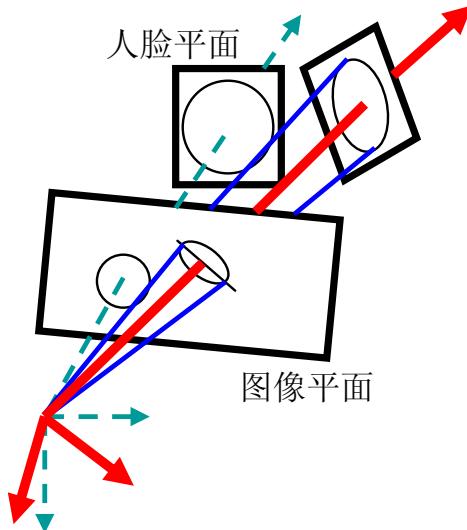


图 4-5 摄像机坐标系第一次旋转

为了使人脸平面与该椭圆锥的交线再次变成圆, 我们首先对摄像机坐标系进行旋转, 使 Z 轴穿过投影椭圆的中心, 且 X 轴与投影椭圆的长轴重合。这样做可以使第二次旋转操作变得很简单和直观, 只需绕投影椭圆的长轴进

行一次旋转就能达到目的。与之相比较，第一次旋转操作就不太直观，并非简单地绕 X 轴，Y 轴或 Z 轴进行旋转，而是绕空间某条直线进行旋转，其旋转由一个一般形式的单位正交矩阵描述。下面我们将说明如何求第一次旋转矩阵。

首先我们将投影椭圆方程(4.24)写成实对称二次型的形式

$$\mathbf{x}^T \mathbf{M} \mathbf{x} = 0 \quad (4.25)$$

其中 $\mathbf{M} = \begin{pmatrix} A & B/2 & D/2f \\ B/2 & C & E/2f \\ D/2f & E/2f & F/f^2 \end{pmatrix}$ 为实对称矩阵， $\mathbf{x}^T = (x, y, z)$ 为空间点的坐标向量。对于实对称二次型，存在对应于实正交矩阵 \mathbf{R}_1 的正交变换，把二次型变成标准型，我们这样定义 \mathbf{R}_1

$$\mathbf{R}_1 = (\mathbf{m}_2, \mathbf{m}_1, \mathbf{m}_3) \quad (4.26)$$

其中 $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$ 分别为实对称矩阵 \mathbf{M} 的三个特征向量，分别对应于特征值 λ_1, λ_2 和 λ_3 ，其中 $\lambda_1 > \lambda_2 > \lambda_3$ 。我们在构造实正交矩阵 \mathbf{R}_1 时，将最大的特征值 λ_1 对应的特征向量 \mathbf{m}_1 放在 \mathbf{m}_2 和 \mathbf{m}_3 之间，这样做的目的是为了使变换后摄像机坐标系的 X 轴和投影椭圆的长轴重合。

设存在如下的正交变换关系

$$\mathbf{R}_1 \mathbf{x}' = \mathbf{x} \quad (4.27)$$

其中 $\mathbf{x}' = (x', y', z')$ 为变换之后的空间点坐标，由于 \mathbf{R}_1 满足如下关系

$$\mathbf{R}_1^T \mathbf{M} \mathbf{R}_1 = \begin{pmatrix} \lambda_2 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \quad (4.28)$$

因此将该变换代入实对称的二次型式(4.25)，即得到标准化的二次型，即变换之后的椭圆锥方程

$$\begin{aligned}
 & (\mathbf{R}_1 \mathbf{x}')^T \mathbf{M} (\mathbf{R}_1 \mathbf{x}') \\
 &= \mathbf{x}^T \mathbf{R}_1^T \mathbf{M} \mathbf{R}_1 \mathbf{x}^T \\
 &= \lambda_2 x'^2 + \lambda_1 y'^2 + \lambda_3 z'^2 = 0
 \end{aligned} \tag{4.29}$$

当 $z' = f$ 的时候, 如果该椭圆锥与图像平面的交线为一个椭圆, 那么必有 $\lambda_3 < 0$, 且该椭圆的中心位于 $(0, 0, f)$, 所以上式可改写成

$$\frac{x'^2}{(-\lambda_2 / \lambda_3 f^2)} + \frac{y'^2}{(-\lambda_1 / \lambda_3 f^2)} = 1 \tag{4.30}$$

因为 $\lambda_1 > \lambda_2 > \lambda_3$, 所以 $-\lambda_2 / \lambda_3 f^2 > -\lambda_1 / \lambda_3 f^2$, 即该摄像机坐标系的 X 轴和投影椭圆的长轴重合, 这说明了式(4.26)中, 构造正交矩阵的时候, 我们为什么要选择将最大的特征值 λ_1 对应的特征向量 \mathbf{m}_1 放在 \mathbf{m}_2 和 \mathbf{m}_3 之间。

这样就得到了第一次旋转矩阵 \mathbf{R}_1 , 第一次旋转的过程如图 4-5 所示。细实线为椭圆锥曲面, 粗实线为第一次旋转之后的坐标系 \mathbf{C}_1 。注意这里的图像平面是刚性固定在摄像机坐标系的坐标轴上随着它旋转的, 因此旋转后的图像平面和虚拟的初始正面平行的人脸平面是不平行的。

(三) 第二次旋转摄像机坐标系

完成摄像机坐标系的第一次旋转之后, 摄像机坐标系的 Z 轴与椭圆锥的主轴重合, X 轴和投影椭圆的长轴重合, 接下来只要将 Y 轴和 Z 轴绕着 X 轴转动, 转过一定角度以后, 就能使椭圆锥和图像平面的交线成为圆, 如图 4-6 所示。 \mathbf{C}_1 (粗实线表示), \mathbf{C}_2 (细实线表示)分别为第一次旋转和第二次旋转之后的摄像机坐标系。 n_f 为人脸平面的法线, 设旋转角为 α , 则旋转矩阵 \mathbf{R}_2 为

$$\mathbf{R}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{pmatrix} \tag{4-31}$$

设第二次旋转后得到的空间点为 $\mathbf{x}'' = (x'', y'', z'')$, 那么旋转前后的空间点满足如下变换关系

$$\mathbf{x}'' = \mathbf{R}_2 \mathbf{x}' \tag{4-32}$$

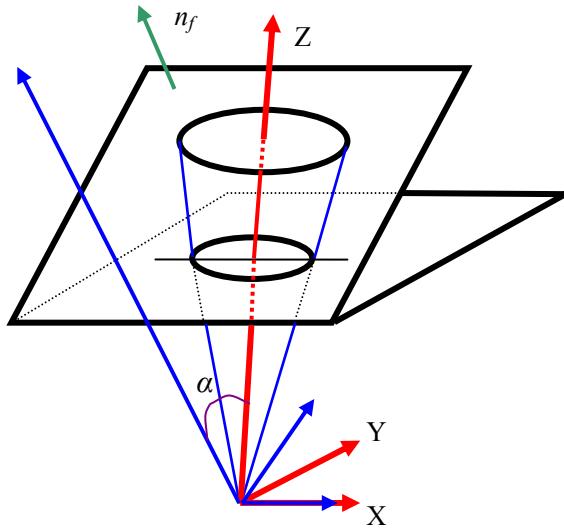


图 4-6 摄像机坐标系第二次旋转

将式(4-32)代入椭圆锥的方程(4.29)中就得到第二次旋转之后的椭圆锥方程

$$\begin{aligned} & \lambda_2 x''^2 + (\lambda_1 \cos^2 \alpha + \lambda_3 \sin^2 \alpha) y''^2 \\ & + (\lambda_1 \sin^2 \alpha + \lambda_3 \cos^2 \alpha) z''^2 \\ & + 2(\lambda_1 - \lambda_3) \sin \alpha \cos \alpha y'' z'' = 0 \end{aligned} \quad (4-33)$$

为了让椭圆锥与图像平面的交线为圆，显然要求 x^2 和 y^2 前的系数相等，因此有

$$\begin{aligned} \lambda_2 &= \lambda_1 \cos^2 \alpha + \lambda_3 \sin^2 \alpha \\ \Rightarrow \cos^2 \alpha &= \frac{(\lambda_2 - \lambda_3)}{(\lambda_1 - \lambda_3)} \end{aligned} \quad (4-34)$$

求解方程(4-34)可以得到四个 α 值，设 α' 为大于 0 度小于 90 度的解，则下面四个角度都为式(4-34)的解，

$$\alpha = \alpha', \quad \alpha = -\alpha', \quad \alpha = \pi + \alpha', \quad \alpha = \pi - \alpha'$$

如图 4-7 所示，四个带箭头的线段分别表示这四个解，其中两个解指向摄像机，两个解背离摄像机。这些解都能使椭圆锥与图像平面的交线为圆。由于人脸平面是可见的，这要求人脸平面的法线方向指向摄像机，所以应该把两个使人脸法线方向背离摄像机的解去掉。另外两个解暂时保留，待第三次旋转角求出后，根据得到的两组运动参数，分别求对各个特征点的投影误差之和，使误差较小的

那个解即为正确的解。

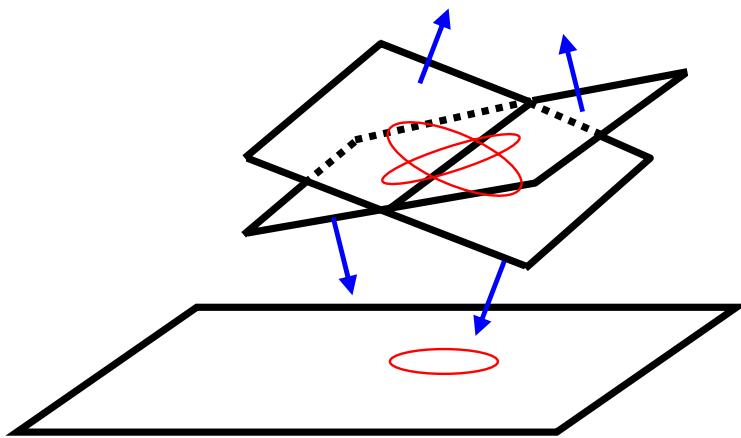


图 4-7 第二次旋转角的多解现象

(四) 第三次旋转摄像机坐标系

显然仅仅求出人脸平面的法线方向并不能唯一确定人脸在空间的姿态，原因是人脸还可能在平面内旋转，因此在两次旋转之后，必须增加第三次旋转—绕 Z 轴在人脸平面内的旋转。为了确定第三次的旋转角度，我们把前面提到的单位圆所在的平面定义为**单位圆平面**，并假设单位圆平面随着摄像机坐标系转动，且与坐标系之间的相对位置关系不变。显然在第一次旋转之前，单位圆平面与图像平面是重合的，在单位圆上定义一个标志点 \mathbf{m} ，其坐标为 $(0,1,f)$ ，令它为人脸平面上一个假想圆上的点 \mathbf{M} 在单位圆平面上的投影。如图 4-8 所示。

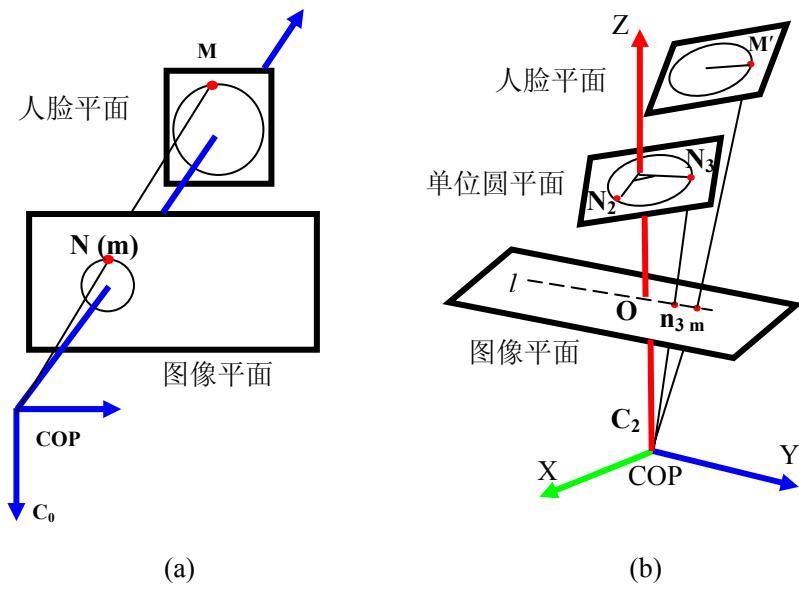


图 4-8 求摄像机坐标系的第三次旋转角

当人脸姿态发生变化使人脸平面上的点 \mathbf{M} 移动到 \mathbf{M}' 时, 根据前面提到的方法, 可以对摄像机坐标系进行两次旋转得到人脸平面的法线方向。设单位圆平面上的点 \mathbf{N} 经过两次旋转之后移到了 \mathbf{N}_2 , 如图 4-8(b)所示, 此时单位圆平面和人脸平面平行。显然还需要绕 \mathbf{C}_2 坐标系的 \mathbf{Z} 轴进行第三次旋转才能使人脸平面和单位圆平面的空间姿态完全一致。设点 \mathbf{N}_2 旋转了 β 角到 \mathbf{N}_3 , 并设 \mathbf{n}_3 和 \mathbf{m}' 分别为 \mathbf{N}_3 与 \mathbf{M}' 在图像平面上的投影, 显然为了使单位圆平面上单位圆的姿态与人脸平面上假想圆的姿态一致, 必有 \mathbf{n}_3 、 \mathbf{m}' 和 \mathbf{O} (\mathbf{Z} 轴与图像平面的交点)三点共线, 并且 \mathbf{n}_3 与 \mathbf{m}' 在 \mathbf{O} 的同一侧。

由于 \mathbf{m}' 和 \mathbf{O} 的坐标可以经由仿射变换得到, 而且 COP 的坐标为 $(0,0,0)$, 因此由 \mathbf{O} , COP 以及 \mathbf{m}' 三个点确定的平面 \mathbf{P} 的方程是已知的。那么单位圆上的点到平面 \mathbf{P} 的距离必为 β 的函数, 设为 $d(\beta)$ 。

求解方程

$$d(\beta)=0 \quad (4.35)$$

可求得旋转角 β 的两个解, 由于 \mathbf{n}_3 与 \mathbf{m}' 在 \mathbf{O} 的同一侧, 所以取与 \mathbf{m}' 之间距离较短的那个解。这样就得到第三次的旋转矩阵 \mathbf{R}_3 , \mathbf{R}_3 的表达式如下

$$\mathbf{R}_3 = \begin{pmatrix} \cos \beta & -\sin \beta & 0 \\ \sin \beta & \cos \beta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (4.36)$$

这样通过三次旋转, 我们就得到了使人脸从正面平行姿态变换到待求帧人脸姿态的旋转矩阵 \mathbf{R}

$$\mathbf{R} = \mathbf{R}_3 \mathbf{R}_2 \mathbf{R}_1 \quad (4.37)$$

由于摄像机的内参数已知, 根据透视投影方程(3.6), 可以使用线性的最小二乘算法得到相应的平移分量 \mathbf{t} 的粗略估计值。这样我们就得到了全部的六个运动参数。

下面说明构造特征点正面平行投影的时候, t_z 的取值不影响对人脸姿态的估计。

设 \hat{t}_z 为输入图像中人脸平面距离摄像机焦点的实际距离, \hat{t}_z 是未知的, 设 $\hat{\mathbf{t}}_z = k\mathbf{t}_z$, k 为比例因子。对于三维模型上的某个特征点 \mathbf{P}_i , 如果人脸平面距离摄像机焦点的距离为 t_z 时其在正面平行投影上的点为 \mathbf{p}_i , 相应的仿射参数为 \mathbf{A} 和 \mathbf{b} , 那么由透视投影的几何关系, 当距离为 \hat{t}_z 时其在特征点正面平行投影的点就为 $\frac{1}{k}\mathbf{p}_i$, 根据公式(4.16)可得到相应的仿射参数为 $k\mathbf{A}$ 和 \mathbf{b} 。

前面已经得到摄像机坐标系的第一次旋转矩阵 \mathbf{R}_1 为

$$\mathbf{R}_1 = (\mathbf{m}_2, \mathbf{m}_1, \mathbf{m}_3)$$

第二次的旋转角 α 为

$$\cos^2 \alpha = \frac{\lambda_1 - \lambda_3}{\lambda_2 - \lambda_3}$$

其中 $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$ 为矩阵 \mathbf{M} 的特征向量, $\lambda_1, \lambda_2, \lambda_3$ 为矩阵 \mathbf{M} 的特征值, \mathbf{M} 为

$$\mathbf{M} = \begin{pmatrix} A & B/2 & D/2f \\ B/2 & C & E/2f \\ D/2f & E/2f & F/f^2 \end{pmatrix}$$

其中 A, B, C, D, E, F 均为仿射参数的函数, f 为摄像机的焦距。

由前所述, 在实际的 z 轴平移 \hat{t}_z 下的仿射参数为 $k\mathbf{A}$ 和 \mathbf{b} , 将其代入可以发现其对应的上述矩阵为 $k^2\mathbf{M}$, 该矩阵的特征方程为

$$|k^2\mathbf{M} - \lambda\mathbf{E}| = 0 \quad (4.37)$$

而原来矩阵 \mathbf{M} 的特征方程为

$$|\mathbf{M} - \lambda\mathbf{E}| = 0 \quad (4.38)$$

显然由式(4.37) 和(4.38)得到 $k^2\mathbf{M}$ 的特征值是 \mathbf{M} 的 k^2 倍, 特征向量相等。由公式(4.26)和(4.34)可知, \mathbf{R}_1 和 α 都不会改变, 也就是说, \mathbf{t}_z 的大小并不影响对人脸平面法线方向的估计。直观地说, 在透视投影模型下, 随着人脸平面和图像

平面距离的不同，得到的投影之间是比例放缩的关系，因此并不影响最终的姿态参数。

4.3.4. 人脸姿态参数迭代求精

由于平面假设只是对人脸结构的粗糙近似，虽然我们已经构建了一个能很好地通过各个特征点的平面并得到了比较准确的虚拟正面平行投影，可是在初始帧上的二维特征点所对应的三维特征点并不位于一个平面上，所以二维特征点的位置不可避免地存在误差，因此通过以上步骤算出投影矩阵仍然是比较粗糙的估计值。为了摆脱平面假设的束缚，求出比较精确的姿态参数，我们把粗略估计值作为一个非线性最优化迭代求精过程的初始估计值。

定义优化目标函数如下：

$$\min D = \sum_{i=1}^N \omega_i [(x'_i(\mathbf{X}, \mathbf{P}) - x_i)^2 + (y'_i(\mathbf{X}, \mathbf{P}) - y_i)^2] \quad (4.39)$$

其中 $x'_i(X, P)$ 与 $y'_i(X, P)$ 是空间点 \mathbf{X} 在输入帧上投影点的坐标值， \mathbf{P} 为投影矩阵。 x_i 和 y_i 为在待输入帧上跟踪得到的对应特征点。目标函数的物理意义就是两者之间的欧氏距离的平方，反映了三维特征点的二维投影误差。 ω_i 是权值，表示该二维/三维特征匹配的可靠程度。如何确定 ω_i 值将在下一节讨论。

我们使用非线性的最小二乘优化算法 LM(Levenburg-Marquett)[74]方法对该目标函数进行迭代优化。迭代的过程就是不断地更新投影矩阵，使目标函数的值最小，也就是使三维特征点的投影不断地向输入帧的特征点靠近，即一种 ICP 迭代算法。

LM 方法是一种求解非线性最小二乘问题的最优化算法，其基本思想是将非线性问题转化为一系列线性最小二乘问题来求解，在初始估计值偏离真实值不远的条件下能得到比较好的结果，附录 I 详细介绍了该方法的思想和求解步骤。

由于摄像机的内参数已知，可以直接估计投影矩阵 \mathbf{P} 中的六个姿态参数（三个旋转分量，三个平移分量）。然而由于旋转矩阵中的各个元素之间存在着约束关系，直接对这些元素做非线性最优化会产生比较大误差，而且直接将投影矩阵分解并得到独立的旋转分量也很困难[53]。由于采用仿射对应技术可以得到旋转

矩阵和平移向量的初始估计值，选择三个坐标已知的三维点，对其做旋转操作，就能得到变换之后的三维点的坐标，然后用 4.3.1 节中的方法能很容易能计算出一个三维向量 \mathbf{q} ，它的方向是旋转轴的方向，它的模等于旋转角的大小。因为 \mathbf{q} 的三个分量是相互独立的，因此对 \mathbf{q} 的分量做最优化能够简化最优化求解的过程，提高姿态参数估计的准确程度。

4.3. 鲁棒的姿态参数估计方法

由于特征点的初始位置不可避免地存在误差，因此一些二维/三维特征点对应并不准确，必须在最优化的目标函数中加以考虑。如果直接将线性或非线性的最小二乘法应用到这些匹配上，由于错误的匹配，得到的运动参数值可能误差极大，完全没用处，最小二乘法不是鲁棒的估计方法。

我们采用鲁棒的最小中值法对目标函数的各项加权如方程(4.39)所示。由于不共线的三组特征点对就可以决定一组姿态参数，因此对于所有三组特征点对的组合，可以得到其鲁棒标准方差为

$$\sigma = \Phi \cdot \min_{i=1 \dots N} \text{median} \sqrt{(x_i(\mathbf{X}, \mathbf{P}) - x_i)^2 + (y_i(\mathbf{X}, \mathbf{P}) - y_i)^2} \quad (4.40)$$

其中 $\Phi = 1.4826$ ，是校正系数。由此得到对每组特征点对其权值 ω_i 可由下式计算

$$\omega_i = \begin{cases} 1 & d_i \leq c \cdot \sigma \\ 0 & d_i > c \cdot \sigma \end{cases} \quad (4.41)$$

其中

$$d_i = \sqrt{(x_i(\mathbf{X}, \mathbf{P}) - x_i)^2 + (y_i(\mathbf{X}, \mathbf{P}) - y_i)^2}$$

是初始帧上三维特征点的投影和对应的二维特征点之间的欧氏距离，反映了三维特征点的二维投影误差。 $c \in [2.5, 3.5]$ 是可选的比例系数。在系统中通过多次实验发现选择 2.8 可以取得比较好的效果。

由于错误的匹配已经用最小中值法检出，因此那些匹配较好的特征对对应运动参数的估计起的作用更大，相比用简单的最小二乘法，能得到更准确的运动参数。

4.4. 实验结果和分析

本章对人脸初始姿态参数估计问题进行了比较全面的研究，并提出了基于三维模型和仿射对应原理的人脸初始姿态估计算法，为了检验该算法的有效性，我们搭建了相关的实验平台进行了一系列实验。首先利用多个实际扫描的人脸三维模型对人脸平面的假设条件做了验证实验；然后用我们的初始姿态参数估计算法分别对模拟数据、石膏像图像和真实人脸图像进行了实验。

4.4.1. 人脸平面假设的验证实验

在基于仿射对应的人脸的初始姿态参数粗略估计的算法中，我们利用了人脸平面的假设，即通过选择某些特定位置的特征点进而可以用平面来近似人脸，这样就能满足平面投影的仿射对应的条件。由于实际的人脸形状千差万别，对应具有不同形状特征的人脸是否仍然满足这个条件，需要进行实验的验证。

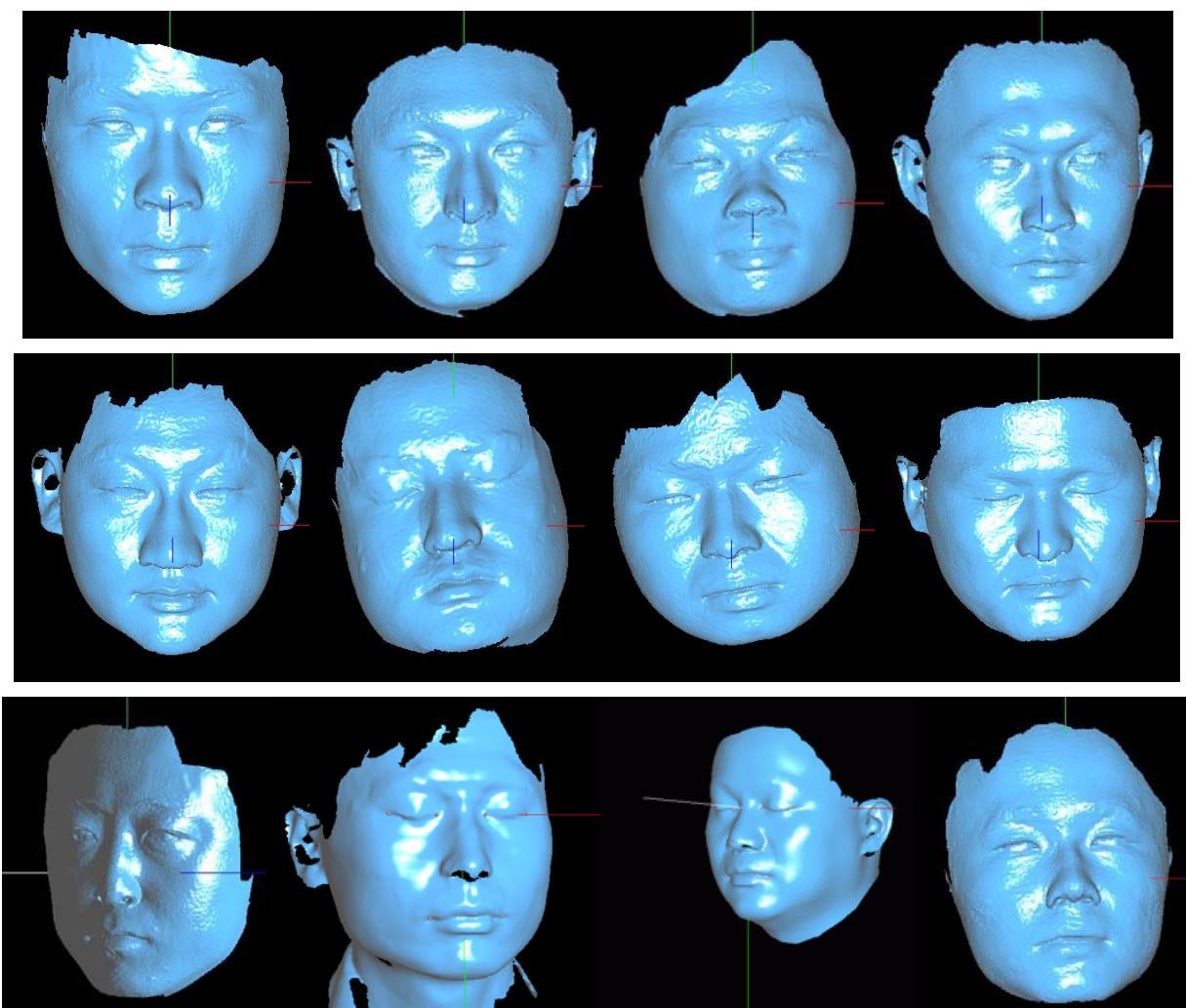


图 4-9 人脸平面假设验证实验所用的三维模型

我们实验室拥有各种人脸的三维模型数据，其中大多数采集自实验室的老师和同学。我们选择了 12 个不同形状的人脸模型，如图 4—9 所示，分别选择内外眼角、嘴角作为特征点，用本章前面提到的方法拟合了人脸平面，并且计算得到与拟合的人脸平面之间的最大距离和人脸尺度的比值，结果如图 4—10 所示，对于大多数的人脸模型而言，其最大的拟合误差都在 3% 以下，可见在人脸表情变化不很剧烈的情况下，基于平面的假设是合理的。

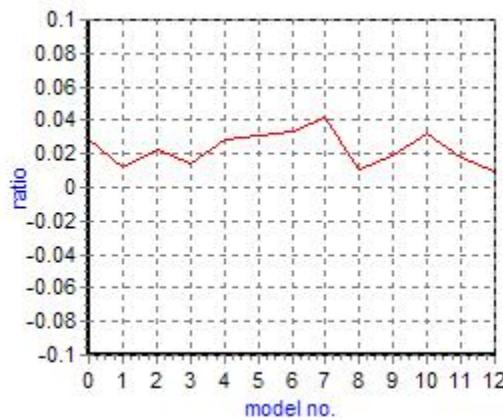


图 4—10 人脸平面假设验证实验结果

4.4.2. 模拟数据实验和分析

为了检验基于仿射对应的人脸姿态初始估计算法的有效性，我们利用图形学的方法生成了一些模拟数据并进行实验。我们首先假定存在一个虚拟的摄像机，其内参数和人脸距离摄像机的距离均已知；接着在三维空间中精确地定义人脸上六个特征点(分别表示内外眼角、嘴角)，保证它们位于同一平面上且该平面与图

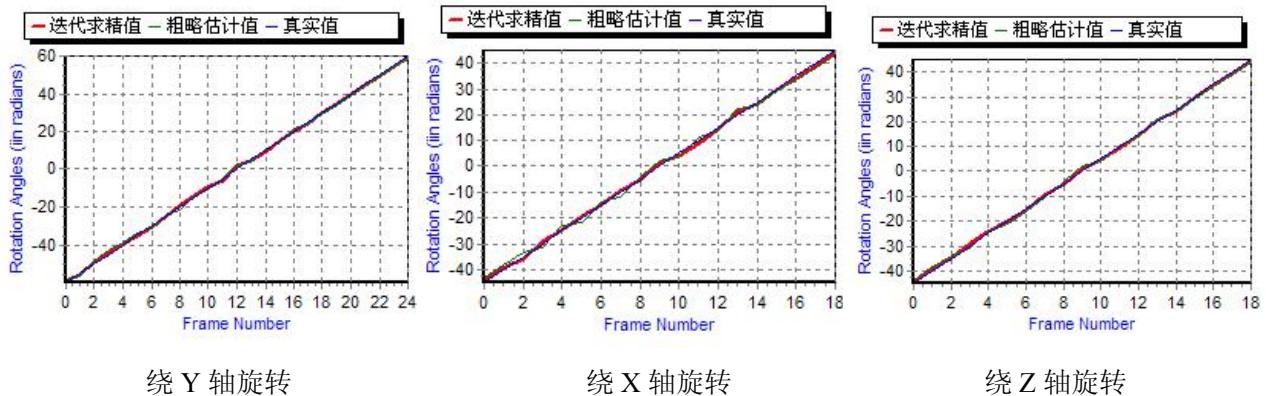


图 4—11 模拟数据实验结果

像平面平行；然后分别让它们绕 Y 轴旋转，旋转角从 -60° 到 60° ，绕 X 和 Z

轴旋转，旋转角从 -45° 到 45° ，旋转间隔角度为 5° 。由于各特征点在图像平面上的投影位置可以根据透视投影模型精确地算出，然后利用这些数据用我们的初始帧姿态参数估计方法进行计算，得到的实验结果如图 4-11 所示，其中点划线为真实转角，细线为用仿射对应方法得到的粗略估计值，粗线为迭代求精之后的估计值。从图中可以看出粗略估计值、迭代求精后的值与真实值非常接近。

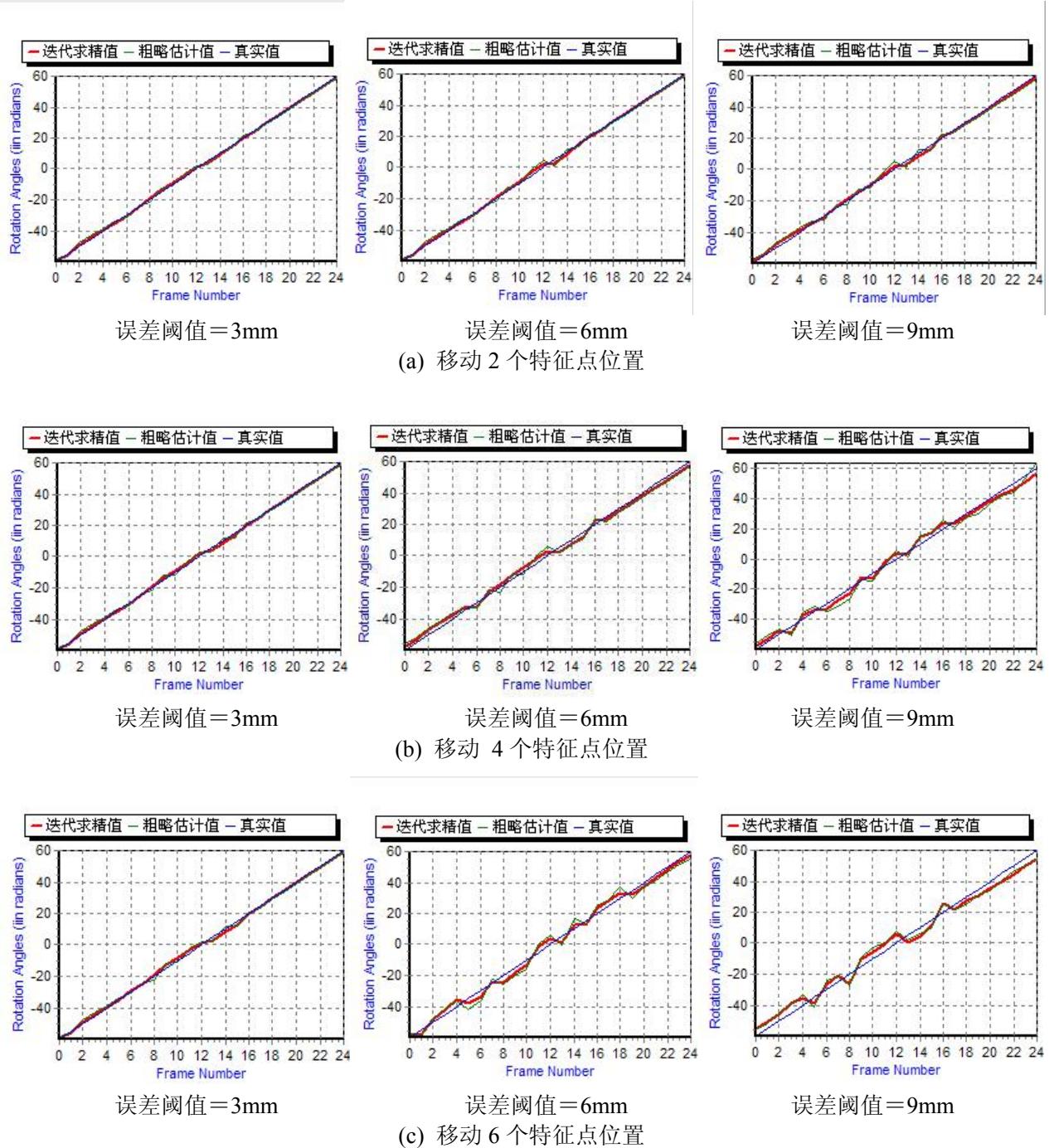


图 4-12 特征点位置出现误差时的实验结果(均为绕 Y 轴旋转的情况)

为了观察特征点位置存在误差时对估计结果的影响，以绕 Y 轴旋转（转角从

—60° 到 60°) 为例, 我们移动部分特征点的位置, 使它们不满足平面假设的条件并用我们的方法计算姿态参数。为了进行定量的估计, 我们将出现位置误差的特征点数目分别设置为 2, 4 和 6 个; 位置误差的阈值分别为 3, 6 和 9mm, 随机选取出现误差的特征点, 其相应的误差大小也在误差阈值内随机选取, 这样得到如图 4—12 所示的实验结果。图中(a)–(c)每一行对应于错误的特征点数目分别为 2, 4 和 6 时的估计结果, 每一列分别对应于误差的阈值为 3, 6 和 9mm 的情况。每幅图中的点划线为真实转角, 粗线为用仿射对应方法得到的粗略估计值, 细线为迭代求精之后的估计值。由实验结果可以看出: (一) 接近正面平行状态的时候估计值的误差较大, 这主要是由于在接近正面平行状态的时候, 如果特征点位置出现误差, 对仿射参数会产生较大的影响, 因此造成初始估计值不够准确; (二) 出现位置误差的特征点数目多少对粗略估计值有较大的影响, 出现位置误差的特征点数目较少时, 估计结果几乎没有什么改变, 主要原因是我们在鲁棒方法已将不太准确的特征点检出; (三) 特征点位置的误差阈值对估计结果有较明显的影响, 但对于一般情况, 人脸能较好地满足平面假设的条件, 因此不会出现很大的误差。

4.4.3. 石膏像图像与真实人脸图像实验

为了检验我们的初始姿态参数估计方法对真实图像的有效性, 我们分别用石膏像图像和真实人脸图像进行了两组实验。

首先是石膏像图像。我们将石膏像放在旋转角度能精确读出的转台上, 选择一些特定转角旋转得到图像, 并用基于仿射对应的方法进行计算, 如图 4—13 所示, 如果 α 为人脸平面法线方向在水平方向转角, β 为人脸平面法线方向与水平面的夹角, 那么可以发现估计结果和真实值非常接近。这表明我们的方法具有较高的精度。

其次是真实人脸图像, 我们从两组记录不同人的人脸图像序列中选择一些图像, 求出其中人脸的旋转角度 α 和 β 的值, 并用人脸三维模型模拟估计的结果。如图 4—14 所示, 可以看出估计值的模拟结果和真实图像在视觉上感觉非常接近。这表明基于仿射变换的方法能得到较好的结果。

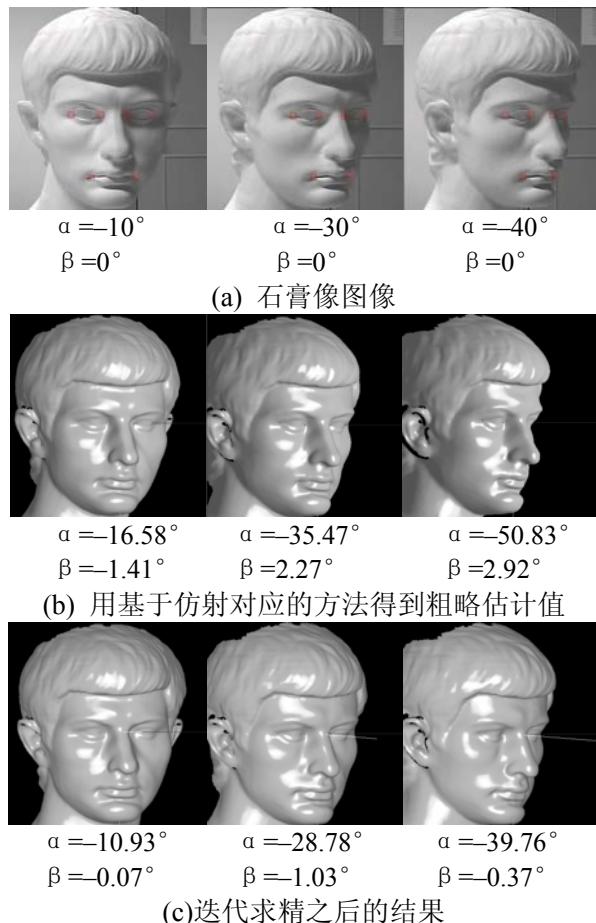
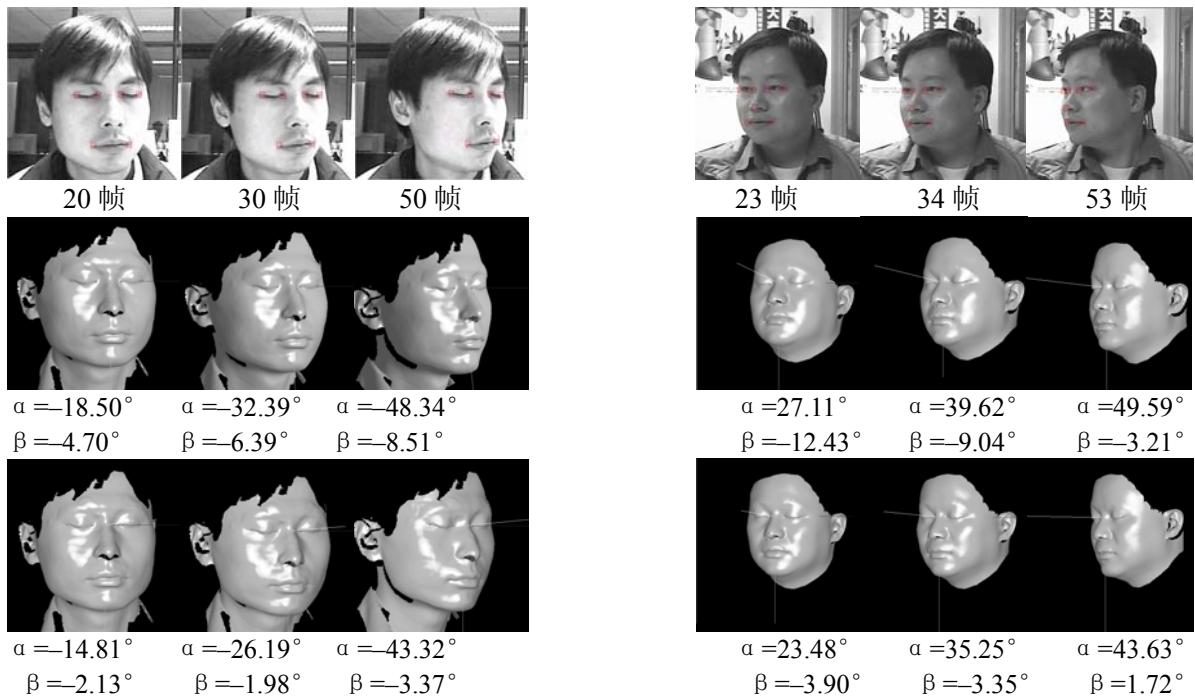


图 4-13 用基于仿射对应的方法计算石膏像图像的姿态参数

图 4-14 用基于仿射对应的方法计算真实人脸图像的姿态参数
(从上到下: 原始图像, 粗略姿态估计值, 迭代求精结果。)

第五章 基于深度灰度约束加权的姿态跟踪

5.1. 概述

本章将在前一章所叙述的图像序列初始帧姿态参数估计的基础上，讨论如何利用帧与帧之间的约束关系对连续图像序列中的各帧进行姿态跟踪。

对图像序列的初始帧进行姿态估计之后，我们得到了比较准确的人脸初始姿态参数，可以以初始姿态作为起点，立刻开始姿态跟踪的过程。从理论上来讲，可以对图象序列的每一帧都重复第四章提到的算法，从而得到每一帧的人脸姿态参数，然而由于基于非线性最优化过程和鲁棒方法的求解过程需要反复迭代几十次甚至几百次才能得到精度较高的结果，计算代价较大。

如果图像序列相邻两帧之间的变化较小，且物体的运动满足一定的条件，就有可能利用帧与帧之间的几何或灰度的关系，得到线性的约束方程，这样就能够用类似线性最小二乘法这样的线性优化方法得到结果。由于求解线性方程比非线性方程计算量小得多，因此利用线性约束关系求解姿态参数对目标是跟踪实时运动的应用就具有很重要的意义。

因此为了增强实际应用的可能性，我们考虑利用帧间的线性的灰度约束关系来求解姿态参数。由于人脸三维模型能提供稳定的深度度量值，因此如果将灰度约束的思想推广到深度，就能得到类似的深度线性约束方程。联合使用两种不同的约束，并对位于人脸不同部位的特征点加权，我们就把基于二维/三维特征对应的非线性问题转化成线性问题，从而能够用比较简单的线性方法求解，从而大大降低了计算的代价；另一方面由于深度度约束能提供比灰度约束更稳定的约束关系，因此能够比单独使用灰度约束得到更好的结果。

姿态跟踪过程的另外一个重要的问题是进行可靠的二维特征跟踪。为了获得精度较高的姿态参数估计，我们必须考虑两个不利因素的影响，一是人脸区域相当的部分缺乏纹理信息，二是输入图像中人脸的姿态变化范围较大，运动的过程中可能出现部分人脸遮蔽的现象，同时人脸表面也可能出现一些局部的变化，这些都使二维特征点的匹配和跟踪的可靠性降低，此外还可能出现特征点跨越连续表面边界的情况，这些因素都会对姿态参数的估计产生不利的影响，因此

必须采取一些策略保证特征跟踪的可靠性。

我们知道，基于二维/三维特征对应求解姿态参数方法对特征位置的准确程度要求较高，如果特征跟踪的结果不够准确，尽管我们可以采用各种鲁棒方法将某些误差较大的特征点检出，但特征跟踪的质量仍然在很大程度上决定了姿态参数估计的准确性。传统的 KLT(Kanade-Lucas-Tomasi)[62,63]方法是在图像序列相邻两帧间的变化比较小的条件下，由两幅图像对应特征区域之间各像素点的灰度变化来求解二维特征运动参数的一种方法，其本质上是利用一阶的泰勒展开对相邻帧特征点区域的灰度变化做线性的近似。在传统的 KLT 方法的基础上，我们提出了一种基于快速傅立叶变换的，改进的 KLT 算法，通过快速傅立叶变换，对频域中的低频部分进行最优化计算，从而减少了传统 KLT 算法收敛所需的迭代次数，同时也提高了二维特征跟踪的准确程度。

由于在跟踪过程中可能存在光照变化、遮蔽等不利因素的影响，造成某些特征点变得越来越不可靠，我们提出一种自动的动态特征更新的方法，该方法无需逐帧指定误差阈值，能根据误差的大小自动地得到可靠的特征点集合。

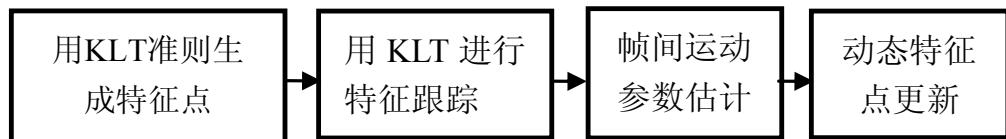


图 5-1 基于深度和灰度约束加权的人脸姿态跟踪

基于深度和灰度约束加权的人脸姿态跟踪过程，由图 5-1 所示，主要由以下几个步骤组成。首先，在完成初始姿态参数估计后，利用 KLT 准则在初始帧上生成一定数量的特征点；其次，用改进的 KLT 方法对部分二维特征进行跟踪；再次，利用帧间的线性深度和灰度约束加权求解帧间运动参数；最后用动态特征更新技术自动判断特征的可靠性，并及时丢弃误差较大的特征，同时生成新的特征补充到特征集合中。

5.2 用 KLT 准则生成新的特征点

KLT 方法[62,63]是由 Kanade 等人提出的在灰度图像序列中选取特征点并对二维特征进行跟踪的参数化的方法。后来 Lucas 和 Tomasi 等人将它推广为一类

在图像相邻帧之间的特征区域灰度相似的条件下，由两幅图像之间的灰度变化求解二维特征运动参数的方法。

这里首先介绍 KLT 特征选取方法。当摄像机拍摄人脸运动的图像时，受光照、人脸表面的形状和皮肤的光线反射特性等因素的影响，图像中人脸区域的灰度会产生复杂的变化。如果摄像机捕获图像的速度足够快，那么对于相邻帧而言，由于各种影响灰度变化因素的相似性，在局部区域内的灰度变化是极其相似的，因此我们可以认为，相邻两帧的局部区域之间存在沿 X 和 Y 方向的位移，这就是所谓的二维特征平移运动模型。这意味着 t 时刻图像上的某个特征点 $\mathbf{x} = (x, y)$ 在 $t+1$ 时刻运动到了 $\mathbf{x}' = (x - d_x, y - d_y)$ ，其中 $\mathbf{d} = (d_x, d_y)$ 为二维特征的平移运动参数向量，该特征点的灰度值在运动前后是近似相等的，即

$$J(\mathbf{x}) = I(\mathbf{x} - \mathbf{d}) + n(\mathbf{x}) \quad (5.1)$$

其中 $J(\mathbf{x}) = I(\mathbf{x}, t+1)$ 为 $t+1$ 时刻的特征点 \mathbf{x} 的灰度值， $I(\mathbf{x} - \mathbf{d}) = I(\mathbf{x} - \mathbf{d}, t)$ 为 t 时刻的该特征点的灰度值， $n(\mathbf{x})$ 为相应的噪声。

显然我们要选择合适的运动参数向量 \mathbf{d} 使在特征点 \mathbf{x} 周围的某个特征窗口 W 内如下的二重积分得到的残差最小

$$\varepsilon = \int_w (I(\mathbf{x} - \mathbf{d}) - J(\mathbf{x}))^2 \omega d\mathbf{x} \quad (5.2)$$

式中的 ω 为对特征区域内不同像素点的加权方程。如果相邻两帧之间的运动比较小的情况下，可以将 $I(\mathbf{x} - \mathbf{d})$ 在 \mathbf{x} 点进行一阶泰勒展开

$$I(\mathbf{x} - \mathbf{d}) = I(\mathbf{x}) - \mathbf{g} \cdot \mathbf{d} \quad (5.3)$$

\mathbf{g} 是梯度向量，于是我们可以将式(5.2)重新写成如下的形式

$$\varepsilon = \int_w (I(\mathbf{x}) - \mathbf{g} \cdot \mathbf{d} - J(\mathbf{x}))^2 \omega d\mathbf{x} = \int_w (h - \mathbf{g} \cdot \mathbf{d})^2 \omega d\mathbf{x} \quad (5.4)$$

其中 $h = I(\mathbf{x}) - J(\mathbf{x})$ 。可以看出，残差是平移向量 \mathbf{d} 的二次方程，这个最优化问题可以得到闭合形式的解。为了使残差最小，对式(5.4)等号两边，求其对 \mathbf{d} 的一阶导数，得到

$$\int_w (h - \mathbf{g} \cdot \mathbf{d}) \mathbf{g} \omega d\mathbf{A} = 0$$

由于 $(\mathbf{g} \cdot \mathbf{d})\mathbf{g} = (\mathbf{g}\mathbf{g}^T)\mathbf{d}$, 而且在特征窗口区域中假设 \mathbf{d} 为常量, 因此得到

$$\left(\int_w (\mathbf{g}\mathbf{g}^T) \mathbf{d} \omega d\mathbf{A} \right) \mathbf{d} = \int_w h \mathbf{g} \omega d\mathbf{A}$$

上式可以写成

$$\mathbf{G}\mathbf{d} = \mathbf{e} \quad (5.5)$$

其中

$$\mathbf{G} = \int_w (\mathbf{g}\mathbf{g}^T) \omega d\mathbf{A}, \quad \mathbf{e} = \int_w h \mathbf{g} \omega d\mathbf{A}$$

式(5.5)是特征跟踪计算中的基本计算步骤, 对于特征窗口内的所有象素都可以计算出其沿 x 和 y 方向的梯度, 因此可以得到实对称的交叉梯度矩阵 \mathbf{G} , 同时对于特征窗口内的所有象素都能计算出两帧之间灰度差并得到向量 \mathbf{e} 。这样就能计算出运动参数 \mathbf{d} 的值。得到 \mathbf{d} 以后, 移动特征窗口, 再重复以上过程, 直到 \mathbf{d} 小于某个阈值, 这表明相邻两帧的特征窗口已经匹配成功, 将前面重复过程中每一轮得到的 \mathbf{d} 都加起来就得到最终的平移运动参数。

从以上的讨论可以知道, 对于 KLT 的特征跟踪方法, 图像中不同区域跟踪的效果是不一样的, 因此研究者们提出了各种选取特征的方法, 如检测角点, 或是图像高频部分对应的点。这些方法都是假定特征已经预先用一种独立于特征跟踪的方法获得, 然后再用特征跟踪方法进行跟踪, 这显然不够合理, 我们应该针对特征跟踪方法的特点选取特征点和特征窗口。换句话说, 所谓好的特征窗口就是那些能被可靠跟踪的特征窗口。

回到 KLT 的特征跟踪的基本方程(5.5), 要可靠地求解平移运动参数 \mathbf{d} , 就必须使 2×2 的参数矩阵 \mathbf{G} 中的各个参数大于系统噪音且具有较好的条件数, 也就是说 \mathbf{G} 的特征值要尽可能大, 使方程(5.5)能够可靠地求解。

因此, 我们可以定义一个阈值 λ , 使 \mathbf{G} 的两个特征值都大于这个阈值

$$\min(\lambda_1, \lambda_2) > \lambda \quad (5.6)$$

我们可以用这个准则来选取特征点， λ 越大表明 \mathbf{G} 的两个特征值都大，这表明特征窗口存在两个梯度变化较大的方向，即该区域的交叉纹理信息非常丰富，这些区域的点能够用 KLT 方法可靠地进行跟踪。如果 λ_1 和 λ_2 一个大一个小，则意味着该区域沿某个方向的梯度较大，这时候就容易出现开孔问题，这样的区域不适合选作特征窗口，如果被选为特征的话，往往会出现沿梯度方向滑移的现象。如果 λ_1 和 λ_2 都较小，表明该区域几乎没有纹理信息，用作特征是不合适的。

当初始帧姿态参数估计完成后，我们就用上面描述的 KLT 准则在二维图像上选取一定数量的特征点并将其反投影到三维模型上，这样就得到了一定数量准确对应的特征点对，在以后的深度灰度约束加权求解人脸帧间运动参数的时候，我们就有较多的特征点可以使用，这样避免了特征点数量较少的时候，个别特征点对应不准确可能产生的姿态参数估计误差。图 5—2 显示了用 KLT 准则生成三维/二维特征点。

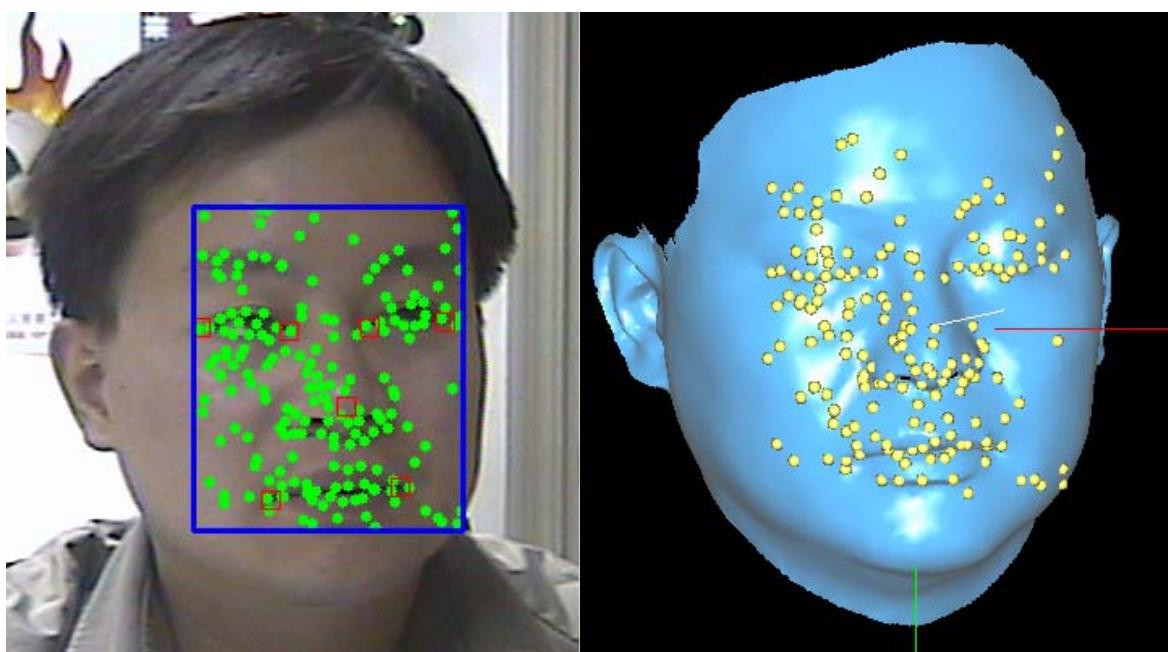


图 5—2 用 KLT 准则生成二维和三维特征对

5.3 加权的深度和灰度约束方程

基于透视投影模型和微小的刚性帧间运动假设，可以导出线性的灰度和深度约束方程。在推导过程中，我们用 $\mathbf{x} = (X, Y, Z)^T$ 表示三维特征点，其在三维空间中的运动速度为 $\mathbf{v} = (V_x, V_y, V_z)^T$ ，二维特征点为 $\mathbf{x} = (x, y)^T$ ，其在图像平面上的运动速度为 $v = (v_x, v_y)^T$ ， f 为摄像机的焦距。

(一) 灰度约束方程

Bergen[75]是最早提出灰度约束的研究者之一，以后许多研究者[76]陆续在一定的条件下给出灰度约束的具体应用。灰度约束方程是由相邻帧之间特征点位置的灰度变化不大的假设推导而来的。假设在 t 时刻图像上的特征点 \mathbf{x} 的灰度为 $I(\mathbf{x}, t)$ ， $t+1$ 时刻 \mathbf{x} 运动到 \mathbf{x}' ，我们近似的认为相邻两帧特征点之间的灰度变化很小，因此得到

$$I(x, y, t) = I(x + v_x, y + v_y, t + 1) \quad (5.7)$$

对右式做一阶泰勒展开得到

$$I(x, y, t) = I(x, y, t) + I_x v_x + I_y v_y + I_t \quad (5.8)$$

其中 I_x 、 I_y 和 I_t 分别是灰度对 x 、 y 和 t 的方向导数，用矩阵的形式表示上面的式子

$$-I_t = \begin{pmatrix} I_x & I_y \end{pmatrix} \begin{pmatrix} v_x \\ v_y \end{pmatrix} \quad (5.9)$$

基于透视投影模型，二维和三维特征点坐标之间的关系可以由 $x = \frac{fX}{Z}$ 和 $y = \frac{fY}{Z}$ 表示。分别对这两个式子求对时间 t 的导数，即可得到二维运动和三维运动之间的关系

$$v_x = \frac{dx}{dt} = \frac{f}{Z} V_x - \frac{x}{Z} V_z, \quad v_y = \frac{dy}{dt} = \frac{f}{Z} V_y - \frac{y}{Z} V_z \quad (5.10)$$

进一步写成矩阵的形式

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} f & 0 & -x \\ 0 & f & -y \end{pmatrix} \begin{pmatrix} V_x \\ V_y \\ V_z \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} f & 0 & -x \\ 0 & f & -y \end{pmatrix} \mathbf{V} \quad (5.11)$$

如果假设帧间的运动是刚性而且微小的，那么可以导出特征点的三维速度和其帧

间运动参数之间的关系。用 $\mathbf{T}=(t_x, t_y, t_z)^T$ 表示运动参数的平移分量， $\boldsymbol{\Omega}=(\omega_x, \omega_y, \omega_z)^T$ 表示旋转分量，其方向为旋转轴的指向，其模 $|\boldsymbol{\Omega}|$ 为旋转角度的大小。对于微小的转动有

$$\mathbf{V} = \mathbf{T} + \boldsymbol{\Omega} \times \mathbf{X} \quad (5.12)$$

由于两个向量的叉积等于一个向量和一个反对称矩阵的乘积，因此上式可以写成

$$\mathbf{V} = \mathbf{T} + \boldsymbol{\Omega} \times \mathbf{X} = \mathbf{T} - \hat{\mathbf{X}} \boldsymbol{\Omega} = (\mathbf{I} - \hat{\mathbf{X}}) \begin{pmatrix} \mathbf{T} \\ \boldsymbol{\Omega} \end{pmatrix} \quad (5.13)$$

其中 $\hat{\mathbf{X}} = \begin{pmatrix} 0 & -Z & Y \\ Z & 0 & -X \\ -Y & X & 0 \end{pmatrix}$ ，令 $\mathbf{Q} = (\mathbf{I} - \hat{\mathbf{X}})$ 且 $\boldsymbol{\Phi} = (\mathbf{T}^T \quad \boldsymbol{\Omega}^T)^T$ ，上式可以写成

更简单的形式

$$\mathbf{V} = \mathbf{Q} \boldsymbol{\Phi} \quad (5.14)$$

将 (5.11)中的 \mathbf{v} 用(5.14)的右式代入，并将(5.11)的右式代入(5.9)，则可得到线性的灰度约束方程

$$-I_t = \frac{1}{Z} \begin{pmatrix} f I_x & f I_y & -(x I_x + y I_y) \end{pmatrix} \mathbf{Q} \boldsymbol{\Phi} \quad (5.15)$$

对每对二维/三维特征点，其二维/三维坐标和摄像机焦距均已知，设特征点数目为 N ，可以将所有特征点的线性灰度约束方程组成矩阵的形式，运动参数的估计问题就等价于求解下面的最优化问题

$$\min \| \mathbf{W}(\mathbf{b} - \mathbf{H} \boldsymbol{\Omega}) \|^2 \quad (5.16)$$

其中 \mathbf{b} 是一个 N 维的向量，其元素的值为每个特征点的 $-I_t$ ， \mathbf{H} 为一个 $N \times 6$ 矩阵，每一行都是一个六维向量，由 $\frac{1}{Z} \begin{pmatrix} f I_x & f I_y & -(x I_x + y I_y) \end{pmatrix} \mathbf{Q}$ 给出。

由于位于人脸不同区域的特征点对运动参数的估计起的作用有所不同，有些点的影响应该小些，因此在式(5.16)中引入一个 $N \times N$ 维的对角权值矩阵 \mathbf{W} 来描述不同

特征点的影响，其定义为

$$\mathbf{W} = \begin{pmatrix} \sqrt{\omega_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\omega_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\omega_N} \end{pmatrix} \quad (5.17)$$

其对角线上的每一个元素为各特征点的权值，如果特征点数 N 大于 6，则可以用加权的线性最小二乘法求解上面的最优化问题。下一节将说明如何确定权值。

(二) 深度约束方程

公式(5.15)体现了相邻帧之间各特征点的灰度一致性关系。如果每个特征点的三维信息是已知的，将灰度一致性的思想扩展为深度一致性，就可以得到相邻帧间的深度约束。假设三维特征点 \mathbf{x} 运动到了新的位置 \mathbf{x}' ，其二维投影点也从 \mathbf{x} 运动到 \mathbf{x}' ，再次利用微小的刚性帧间运动假设，三维特征点的深度变换可以由类似于式(5.7)的形式表示

$$Z(x, y, t) + V_z = Z(x + v_x, y + v_y, t + 1) \quad (5.18)$$

用推导灰度约束方程过程中使用过的相同方法，可以得出矩阵形式的深度一致性约束方程

$$-Z_t = (Z_x \quad Z_y) \begin{pmatrix} v_x \\ v_y \end{pmatrix} - V_z \quad (5.19)$$

其中 Z_x 、 Z_y 和 Z_t 分别是深度对 x 、 y 和 t 的方向导数。再次用透视投影模型将二维速度和三维速度相关联，就可以得到和(5.15)类似的线性的深度约束方程

$$-Z_t = \frac{1}{Z} \begin{pmatrix} f Z_x & f Z_y & -(Z + x Z_x + y Z_y) \end{pmatrix} \mathbf{Q} \Phi \quad (5.20)$$

从上面的讨论可以看出，灰度约束方程是建立在帧之间特征点灰度一致性的假设之上，一般而言，灰度很不稳定，容易受到光照、遮蔽和其他一些微妙因素的影响。深度约束利用三维模型的深度信息，其变化比较稳定，能提供比较可靠的约束，因此将深度和灰度约束方程联立，并用加权的线性最小二乘法求解这个最优

化问题可以得出比单独应用灰度约束得到更好的结果。

上述结果是基于刚性约束的条件下推导出来的，基于刚性约束的运动估计方法还有[77,78,79]，一些研究者将深度约束应用到立体视觉系统中[80]或利用基于模型的灰度约束来估计物体的运动[81]，此外还有一些研究人员[82,83,84,85,86,87,88]试图对非刚性的运动进行估计。

5.4. 基于快速傅立叶变换的二维特征跟踪

系统在初始化时已经获得特征点的初始位置，下面对于输入的视频图像序列进行二维特征跟踪从而得到每帧的特征点位置。

5.4.1. Lucas-Kanade 类方法

假设观察一幅图像 $\mathbf{I} = (\mathbf{u}, t)$ ，其中 $\mathbf{u} = (u, v)$ 为图像中的某个象素，在 $t + 1$ 时刻， \mathbf{u} 运动到了 $\mathbf{u}' = \mathbf{F}(\mathbf{u}, \mathbf{d})$ 位置，其中 \mathbf{d} 为物体的运动参数向量， $\mathbf{F}(\mathbf{u}, \mathbf{d})$ 为参数化的运动模型(可以是仿射模型，或者是简单的二维平移运动模型 $\mathbf{d} = (d_x, d_y)$)，正如 Kanade 最初提出该方法的情形)，如果假设两帧之间变化很小，且照明条件不变，那么有

$$\mathbf{I}(\mathbf{F}(\mathbf{u}, \mathbf{d}), t + 1) = \mathbf{I}(\mathbf{u}, t), \quad (5.21)$$

则运动参数 \mathbf{d} 满足以下的最优化条件，

$$\min \varepsilon(\mathbf{d}) = \sum_{\mathbf{u} \in \Omega} (\mathbf{I}(\mathbf{F}(\mathbf{u}, \mathbf{d}), t + 1) - \mathbf{I}(\mathbf{u}, t))^2 \quad (5.22)$$

其中 Ω 是 t 时刻的模板窗口。为了便于求解运动参数 \mathbf{d} ，对 $\mathbf{I}(\mathbf{F}(\mathbf{u}, \mathbf{d}), t + 1)$ 做一阶的泰勒展开

$$\mathbf{I}(\mathbf{F}(\mathbf{u}, \mathbf{d}), t + 1) = \mathbf{I}(\mathbf{u}, t) - \mathbf{I}_u \cdot \mathbf{d} \quad (5.23)$$

将(5.23)式带入(5.22)，将该非线性问题简化为线性问题，即可求得运动参数 \mathbf{d}

$$\mathbf{d} = -(\sum_{\Omega} (\mathbf{I}_u \mathbf{F}_d)^T (\mathbf{I}_u \mathbf{F}_d))^{-1} \sum_{\Omega} (\mathbf{I}_t (\mathbf{I}_u \mathbf{F}_d)) \quad (5.24)$$

其中 \mathbf{I}_t 和 \mathbf{I}_u 分别为模板窗口内的象素在时域和空域上的梯度， \mathbf{F}_d 为运动模型 \mathbf{F} 对运动参数向量 \mathbf{d} 的偏导数。

5.4.2. 基于 FFT 的改进的 KLT 方法

由上述分析可知，KLT 方法本质上是对模板窗口区域内图像的运动做一个线性近似，因此在灰度曲面曲率变化比较大的象素周围会产生较大的误差，需要一个多次迭代的过程才能估算出准确的运动参数。为了减少迭代计算的开销，一般的解决方法是对模板窗口区域的每个象素按照一定的规则加权。这里我们提出了一种基于傅立叶变换的改进的 KLT 方法，它无需对每个象素加权，并能有效地减少迭代次数，提高迭代的精度。

假设运动模型为二维平移，那么在模板窗口内 KLT 的误差方程为

$$\varepsilon = \int_W [i_t(x, y) - \frac{\partial i_t(x, y)}{\partial x} d_x - \frac{\partial i_t(x, y)}{\partial y} d_y - i_{t+1}(x, y)]^2 dx dy \quad (5.25)$$

其中 $i_t(x, y)$ 和 $i_{t+1}(x, y)$ 分别为 t 和 $t+1$ 时刻在模板窗口 W 内象素的灰度值， d_x 和 d_y 为人脸在水平和垂直方向的位移。对该误差方程做傅立叶变换，根据帕西瓦尔定理，空域能量与频域能量守恒，则有

$$\begin{aligned} \varepsilon &= \int_W [i_t(x, y) - \frac{\partial i_t(x, y)}{\partial x} d_x - \frac{\partial i_t(x, y)}{\partial y} d_y - i_{t+1}(x, y)]^2 dx dy \\ &= \int_W \|I_t(\omega_x, \omega_y) - I_{t,x}(\omega_x, \omega_y) d_x - I_{t,y}(\omega_x, \omega_y) d_y - I_{t+1}(\omega_x, \omega_y)\|^2 d\omega_x d\omega_y \end{aligned} \quad (5.26)$$

其中 $I_t(\omega_x, \omega_y), I_{t+1}(\omega_x, \omega_y), I_{t,x}(\omega_x, \omega_y), I_{t,y}(\omega_x, \omega_y)$ 分别为对 $i_t(x, y), i_{t+1}(x, y), \frac{\partial i_t(x, y)}{\partial x}, \frac{\partial i_t(x, y)}{\partial y}$ 做傅立叶变换之后的结果。

由(5.26)式可以看出，对频域中误差方程的模做最优化等效于对空域中误差方程做最优化。运动参数的计算误差主要来源于空域中灰度曲面曲率变化比较大的区域，这些区域往往对应于图像的高频部分，因此如果我们在频域中只对误差方程的低频部分做优化就可以明显提高迭代的精度，这样用较少的迭代次数就能收敛，从而提高了跟踪的效率。可是傅立叶变换本身也会增加计算开销，因此要

权衡选择大小合适的傅立叶变换区域。通过反复试验发现用 8×8 的傅立叶变换区域时能够得到比较好的结果。

5.5. 计算特征点的权值

上面已经提到，图像噪音、光照变化以及遮蔽都会引起图像灰度分布的不一致，影响特征点的可靠性，而且位于人脸不同位置的特征点可能受到影响大小也是不同的。这种影响应该用权值的形式加以体现并反映到人脸运动参数的估计中。

首先考虑那些位于人脸侧面的特征点，由于它们非常容易被遮蔽或是在二维图像上跨越不同表面的边界，在这些情况下，对这些特征点的二维投影用光流法进行跟踪容易产生较大的误差，因此应当赋予较小的权值。

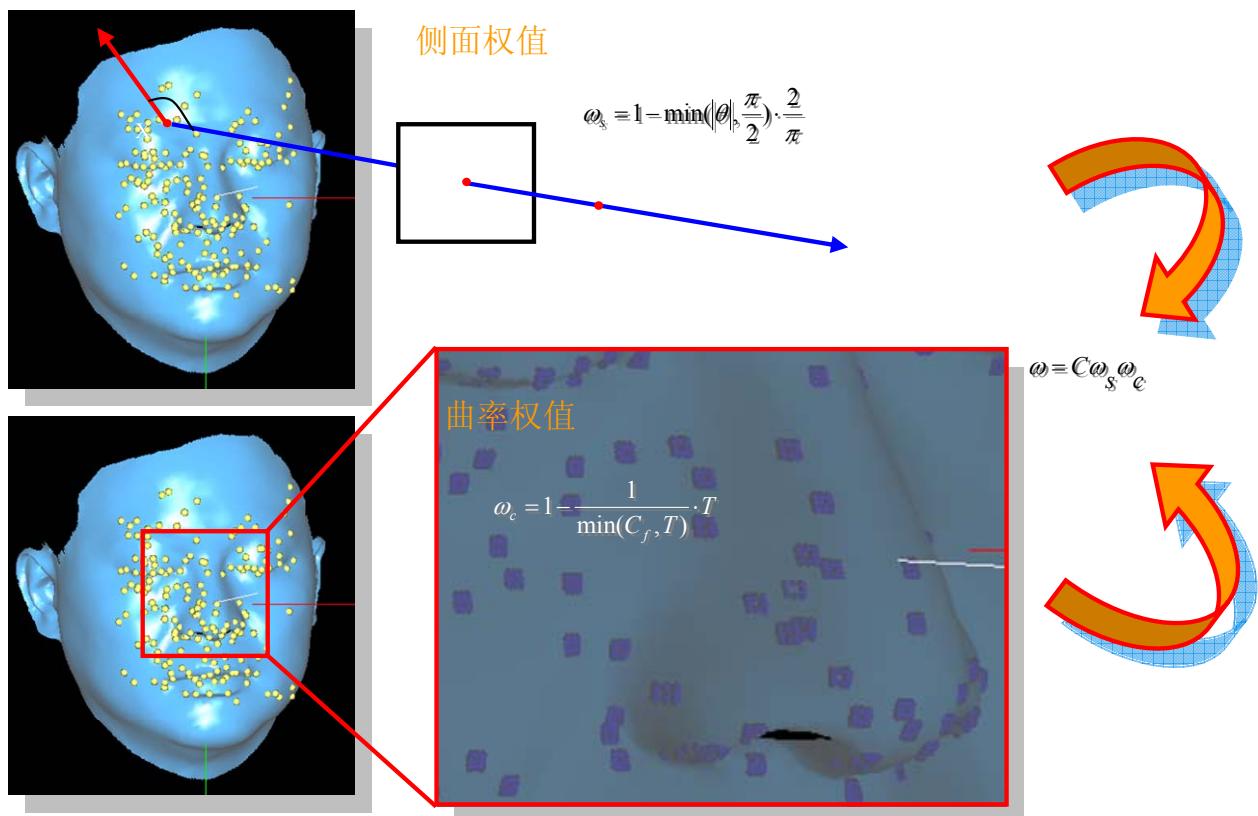


图 5-3 计算特征点权值

图 5-3 上半部分显示了确定侧面权值的方法。三维特征点 \mathbf{x} 其二维投影是 \mathbf{x} ， O 是摄像机的焦点， θ 是人脸表面在 \mathbf{x} 点的法线方向和 $\overrightarrow{\mathbf{X}O}$ 之间的夹角。定义侧面权值的表达式为

$$\omega_s = 1 - \min(|\theta|, \frac{\pi}{2}) \cdot \frac{2}{\pi} \quad (5.27)$$

当 $\theta \geq \frac{\pi}{2}$ 时, ω_s 等于 0, 表示 \mathbf{X} 不可见。当 $\theta < \frac{\pi}{2}$ 时, 权值随着 θ 线性地变化,

θ 越小权值越大。

在推导深度和灰度约束方程的过程中, 我们对 $t+1$ 时刻的深度和灰度值做了一阶泰勒展开。对深度和灰度值在时域和空域上呈现线性变化的情况, 一阶泰勒展开产生的误差较小。当相邻帧之间的时间间隔较小的时候, 时域上能够保持较好的线性变化。在空域上变化则取决于特征点所在的区域的曲率大小, 较小曲率的区域表明其变化越接近于线性, 因此应该对人脸姿态的估计起到更大的作用。因此可以用下面的式子来计算曲率权值

$$\omega_c = 1 - \frac{1}{\min(C_f, T)} \cdot T \quad (5.28)$$

其中 C_f 是人脸曲面在特征点位置的高斯曲率, T 是预定义的阈值。当 $C_f \geq T$ 时, $\omega_c = 0$, 表示该点处曲面的曲率太大不能满足线性变化的条件, 这时用一阶泰勒展开作为近似会产生较大的误差。当 $C_f < T$ 时, C_f 越小 ω_c 越大, 表示曲面越接近平面姿态参数估计起的作用越大。图 5-3 也显示了位于人脸表面上不同曲率区域的特征点。

综合考虑以上两种因素, 可以得到总的权值计算公式

$$\omega = C \omega_s \omega_c \quad (5.29)$$

其中 C 是比例常量。

5.6. 计算特征点的曲率权值

由于已经获得人脸的三维模型, 因此我们能够计算出人脸曲面上三维特征点位置的曲率。

为了得到精度较高的曲率值, 我们先在三维特征点周围的区域选取一些采样

点，由于在深度约束方程中还要计算深度对X轴和Y轴的方向导数，因此我们沿X轴和Y轴方向选取较多采样点，以使计算结果尽可能精确。由于直接对离散的采样点是无法进行曲率计算的，必须将离散的点连续化，即形成曲面。

由离散的点产生曲面的计算方法有很多，有各种插值算法和拟合算法。在这里我们选择拟合算法用曲面逼近数据点，那么在噪声点附近，由于噪声点周围的非噪点的均衡作用，就会减少拟合曲面和真实曲面的误差，从而减少由噪声点产生的误差。

我们用二次曲面来作为特征点周围区域的人脸曲面的近似，并计算该点的高斯曲率。设该二次曲面方程为

$$Z = f(x, y) = a_0 + a_1x + a_2y + a_3x^2 + a_4y^2 \quad (5.30)$$

其中 $\{a_i\}_{i=0}^4$ 为待定的系数。我们利用在人脸三维模型上选取的采样点，用最小二乘法得到这些系数的值。

我们之所以选择二次函数来对采样点进行曲面拟合原因主要是考虑二次函数性质比较简单，这将简化后面的曲率计算，有利于大数据量的计算。其次是曲面拟合是在局部范围内的进行的，二次函数不会产生很大的误差。

根据微分几何的理论[89]，我们可以得到曲面的第一基本形式

$$I = Edu^2 + 2Fdudv + Gdv^2$$

和第二基本形式的表达式

$$II = Ldu^2 + 2Mdudv + Ndv^2$$

其中其中 L 、 M 、 N 、 E 、 F 和 G 是对应的参数。第一基本形式表示的是曲面上两个无限邻近的点的距离大小，而第二基本形式表示的是曲面上两个无限邻近点中，一点到另外一点的切面的距离的大小，它表明了曲面和切面的离差的特征，也反映了曲面在空间中的弯曲程度。

定义法曲率为

$$k_n = \frac{II}{I}$$

由法曲率的表达式可知, 法曲率的值决定了在某三维特征点沿某一固定方向的弯曲程度。法曲率存在两个极值, 每一极值点称为曲面在该三维特征点的主曲率, 其对应的方向为主方向。用两个主曲率的乘积即可得到该三维点的高斯曲率的表达式

$$C_f = k_1 k_2$$

根据求解主曲率的公式, 可以得到人脸曲面上的某个三维点 $(x, y, f(x, y))$ 处的高斯曲率 C_f 可以用下式计算

$$C_f = \frac{LN - M^2}{EG - F^2} \quad (5.31)$$

其中 L 、 M 、 N 、 E 、 F 和 G 分别为曲面第一基本形式和第二基本形式的参数, 可以根据拟合的二次曲面方程的各个系数计算出来, 略去推导过程, 得到计算各参数的式子如下

$$\begin{aligned} L &= \frac{2a_3}{\sqrt{1 + (a_1 + 2a_3x)^2 + (a_2 + 2a_4y)^2}}, \\ M &= 0, \\ N &= \frac{2a_4}{\sqrt{1 + (a_1 + 2a_3x)^2 + (a_2 + 2a_4y)^2}} \\ E &= 1 + (a_1 + 2a_3x)^2, \\ F &= (a_1 + 2a_3x) \cdot (a_2 + 2a_4y), \\ G &= 1 + (a_2 + 2a_4y)^2 \end{aligned} \quad (5.32)$$

将(5.32)中的 C_f 代入(5.31), 即可得到每个特征点的高斯曲率值。

5.7. 帧间运动参数估计

在式(5.20)中, 由于人脸的三维模型已知, 所以深度对 x 和 y 的方向导数能够很容易计算出来, 问题是如何计算深度对时间 t 的方向导数。有的研究者[76]利用立体视觉系统获得帧速率的人脸深度信息, 并计算出图像上每个象素点的深度

对时间 t 的方向导数。由于我们使用的单目系统无法获得帧速率的深度信息，因此采用一种不同的方法求解姿态参数，主要步骤如下：

- 1) 初始的姿态参数估计之后，设用 KLT 方法所生成的所有特征点的集合为 $P = \{\mathbf{p}_i\}_{i=1}^N$ 。从 P 中随机地选取 n 个点 ($n < N$) 构成一个特征点组合 $G_j = \{\mathbf{p}_{gi}\}_{i=1}^n, n < N$ 且 $\mathbf{p}_{gi} \in P$ ，反复 M 次选出 M 个特征点组合，并得到集合 $G = \{G_j\}_{j=1}^M$ 。
- 2) 对每个特征点组合 G_j 根据灰度约束方程(5.15)得到一组人脸运动参数 Φ_j ，所有 M 组参数构成集合 $\Phi = \{\Phi_j\}_{j=1}^M$ 。
- 3) 对每组运动参数 $\Phi_j \in \Phi$ ，计算出所有三维特征点的二维投影与二维特征点之间的误差中值，然后选取 M' 个能产生最小误差中值的运动参数 Φ_j ，组成新的集合 $\Phi' = \{\Phi_j\}_{j=1}^{M'}, M' < M$ 。
- 4) 对 Φ' 中的每组运动参数计算出 P 中所有特征点在 $t+1$ 时的深度值，然后用第 8 节中描述的方法对所有特征点得到能产生最小误差的深度均值。
- 5) 用深度均值计算深度对时间 t 的方向导数，然后联立方程(5.15)和(5.20)并用加权的线性最小二乘法得到运动参数的最终估计值。

完成上述步骤后，根据得到的帧间运动参数值，将三维模型旋转和平移到新的位置。对序列中的每一帧都重复以上的过程就实现了连续的人脸姿态跟踪。

5.8. 特征点动态更新

在特征跟踪的过程中，由于受各种因素的影响，某些特征点位置可能会出现较大的误差。尽管已经对可能出现较大误差的特征点赋予较小的权值，但由于导致误差产生的因素的不确定性，更好的方法是逐帧地判断每个特征点的可靠性并及时丢弃那些误差较大的点，并在必要的时候生成新的特征点。由于在跟踪的过程中不可能手工确定误差的阈值，因此采用一种自动的特征更新算法。该算法如图 5—4 所示。算法的主要步骤如下：

- 1) 对图像序列中的每一帧(除了第一帧)，完成运动参数的估计之后，计算出二维

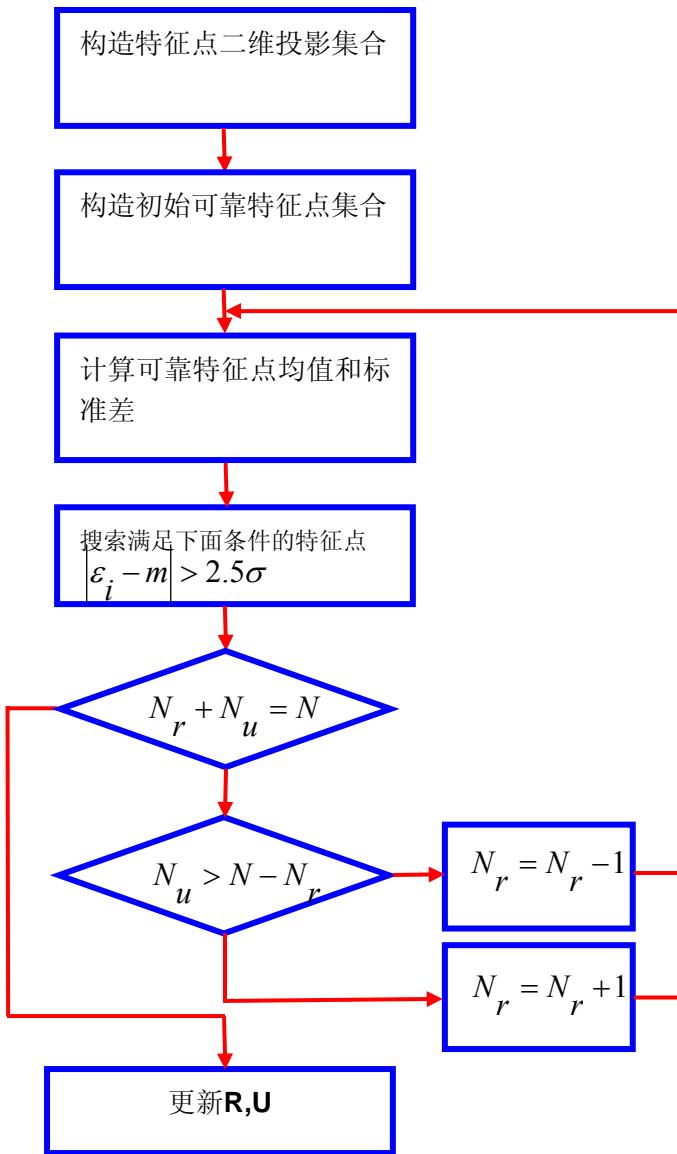


图 5-4 动态特征点更新算法

特征点和三维特征点的二维投影之间的误差，并得到按升序排列的误差集合

$$E = \{\varepsilon_i\}_{i=1}^N.$$

- 2) 开始进行跟踪时，假设可靠的特征点数与所有特征点数的比值为某个特定的值 r (例如 0.9)。构造可靠特征点集合 R ，包含 $N_r = N \times r$ 个投影误差最小的特征点；同样可以构造不可靠特征点集合 U ，包含剩下的 $N - N_r$ 个投影误差较大的点。
- 3) 计算 R 中特征点的误差均值 m 和标准差 σ 。
- 4) 在 E 中搜索满足条件 $|\varepsilon_i - m| > 2.5\sigma$ 的元素，假设其个数为 N_u 。
 - I) 如果 $N_r + N_u = N$ ，那么转到步骤 5，当前的 R 即为可靠的特征点集合。
 - II) 如果 $N_u > N - N_r$ ，那么令 $N_r = N_r - 1$ ，将 R 中的最后一个元素加入 U 的起始位置上并转到步骤 3。

III) 如果 $N_u < N - N_r$, 那么令 $N_r = N_r + 1$, 将 U 的第一个元素加到 R 的末尾, 并转到步骤 3。

5) 确定了可靠的特征点集合之后, 丢弃 U 中不可靠的特征点, 并根据 KLT 准则 [62,63]生成相同数目的新特征点补充到 R 中去。

对于图像序列中的每一帧都进行上述的特征点更新操作, 并利用更新后的特征点估算帧间运动参数。特征点动态更新算法能够有效地消除误差积累, 提高长时段的人脸姿态跟踪过程的可靠性和准确性。

5.9. 实验结果和分析

本章对人脸帧间姿态参数估计问题进行了比较深入的研究, 并提出了基于三维模型和帧间深度和灰度约束加权的人脸姿态跟踪算法以及动态特征点更新算法。为了检验这些算法的有效性, 我们搭建了相关的实验平台, 针对这些算法设计了一系列实验。

(一) 基于 FFT 的改进的 KLT 特征点跟踪方法的实验分析。传统的 KLT 方法基于微小的帧间运动假设, 利用相邻帧之间的灰度变化关系, 通过多次迭代的过程求解二维运动参数。我们提出的基于 FFT 的改进的 KLT 特征点跟踪方法则对 FFT 变换之后频域中的低频部分进行最优化操作。为了比较这两种方法的性能, 我们对几组实验图像序列分别用传统的 KLT 方法和改进的 KLT 进行特征跟踪, 比较了跟踪的准确程度、迭代收敛的次数。

(二) 模特头像图像序列实验。由于真实人脸的姿态参数难以准确测量, 因此我们将纹理的模特头像放在旋转角度能准确测出的转台上, 按照一定的转角间隔, 拍摄多幅图像组成真实值已知的图像序列, 用包括我们的方法在内的四种不同的方法进行姿态跟踪。我们分别获得了绕 X, Y, Z 轴的进行旋转的图像序列得到相应的估计结果并进行了分析。

(三) 真实人脸图像序列实验。最后为了直观地检验我们的算法在真实人脸图像序列上的有效性。我们采集了几组真实人脸运动的图像序列, 包括绕 X, Y, Z 轴转动和自由运动的情况。由于真实人脸数据无法准确测量, 我们将估计结果分别以图表和三维模拟图的形式表示, 从视觉效果上证明了我们的方法能够得到

比较精确的估计结果。

5.9.1. 基于 FFT 的 KLT 特征点跟踪实验分析

为了提高二维特征的跟踪精度，我们提出了一种基于 FFT 的改进的 KLT 方法。为了检验这种方法的有效性，我们选择了一组人脸帧间运动幅度较大的图像序列(共约 80 帧,图 5—5 为其中选出的一些帧)，对位于人脸纹理信息较丰富区域的六个特征点分别用 SSD(sum-of-squared-difference)方法[63]，传统的 KLT 方法和我们提出的改进的 KLT 方法进行跟踪，并对实验结果做了比较。以第 30 帧为例，从表一可以看出改进的 KLT 方法得到的特征点位置比较准确，误差均小于 5 个象素，而 SSD 方法则最高达 18 个象素，传统的 KLT 方法则介于其中。可见使用改进的 KLT 方法能有效地提高估计的准确程度。



图 5—5 帧间运动幅度较大的图像序列

表 5—1 第 30 帧特征点位置跟踪误差
(表格中的值为跟踪误差的绝对值, 单位为象素)

	SSD 方法	传统的 KLT 方法	基于傅立叶变换的 KLT 方法
右眼外眼角	18	8	4
右眼内眼角	5	5	3
左眼内眼角	7	3	1
左眼外眼角	11	7	4
右嘴角	11	5	5
左嘴角	8	4	2

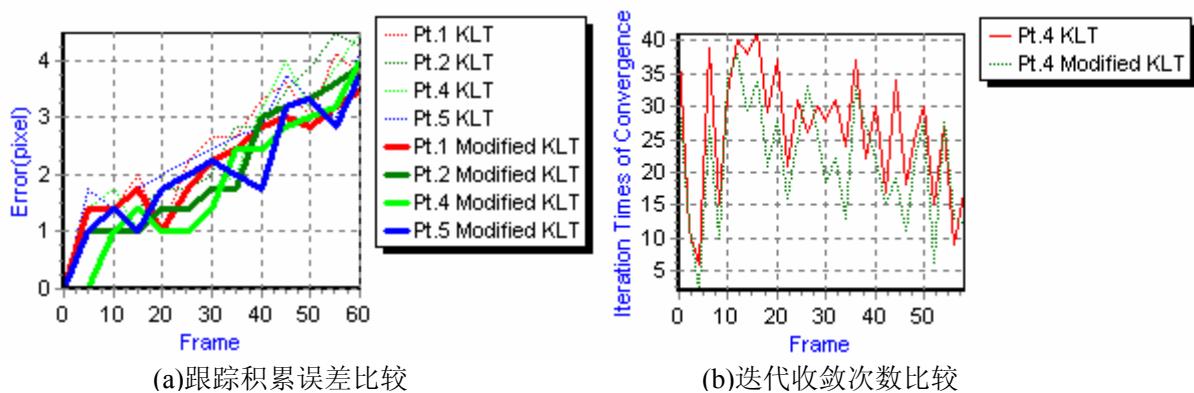


图 5—6 KLT 和改进的 KLT 方法对二维特征跟踪的比较

图 5—6 显示了对另一组帧间运动幅相对较小的人脸图像序列的二维特征进行跟踪的结果。(a)中是两个内眼角和嘴角进行跟踪的积累误差随帧数变化的曲

线。细线是用 KLT 方法得到的结果，粗线是用改进的 KLT 方法得到的结果。虽然对于帧间运动较大的情况，改进的 KLT 方法也不能消除误差积累的现象，可是由图可见，积累误差的值要小于相应用传统 KLT 方法得到的结果。(b)是 KLT 迭代收敛次数随帧数变化的曲线。实线为用传统 KLT 得到的结果，虚线为用改进 KLT 得到的结果，可以看出，对大多数帧，改进的 KLT 方法能比传统的 KLT 方法更快收敛。

5.9.2. 模特头像图像序列姿态跟踪实验分析

由于真实人脸的姿态参数难以实时测量，因此为了检验我们提出的人脸姿态跟踪算法的精确程度，我们对模特头像的图象序列进行了实验。为了能准确读出头像的旋转角度，我们将模特头像放在一个能进行 360 度旋转的转台上，转台的转动可以由一个线控系统准确地进行控制，其旋转角度能够通过仪器直接读出。通过调节转台和模特头像之间的相对位置，可以分别模拟头像绕 X、Y 和 Z 轴的旋转。

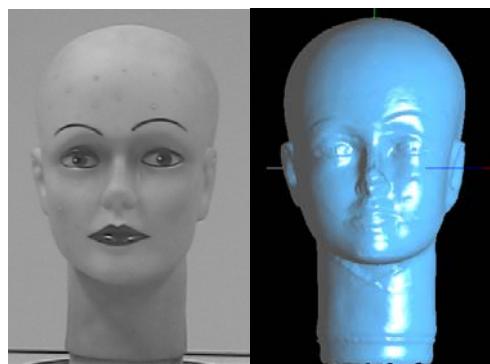


图 5-7 模特头像图象及其三维数据

实验共用了三组图像序列，所有序列中的每帧均是单独拍摄的，这样做的目的是保证帧与帧之间的运动参数都是通过仪器读出的，是已知的。第一组序列中模特头像开始是正面朝向摄像机，然后绕着 Y 轴先后做 -70° 、 $+70^\circ$ 、 -70° 和 $+70^\circ$ 旋转 (每帧 1 度，共 280 帧) 回到初始位置。第二组和第三组序列起始状态也是正面朝向摄像机，然后分别绕 X 和 Z 轴旋转，转角从 -45° 到 $+45^\circ$ (每帧 1 度，共 180 帧)。

图 5-7 显示了实验中使用的模特头像的图像和三维数据。图 5-8 显示了三组模特头像图象序列的极值位置。第一行三幅图分别为第一组图像序列(绕 Y 轴

旋转)的两个极值位置以及初始位置。第二行是第二组图像序列(绕 X 轴旋转)的两个极值位置以及初始位置。第三行是第三组图像序列(绕 Z 轴旋转)的两个极值位置以及初始位置。

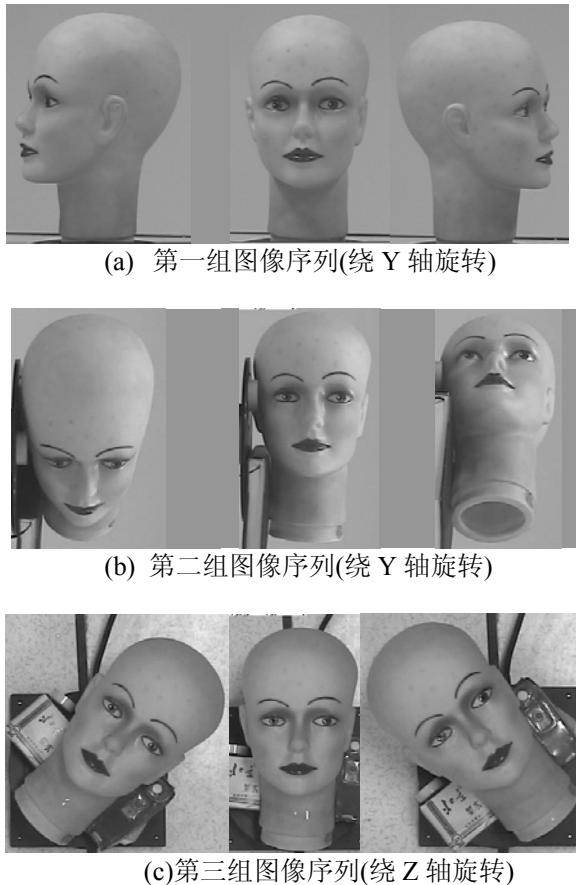


图 5-8 三组模特头像图象序列

用 BC 表示灰度约束, DC 表示深度约束, DFU 表示动态特征点更新。为了进行比较, 我们对每一组图像序列的运动参数都用以下四种方法分别求解: 1) 只用 BC 的方法; 2) 联合使用 BC 和 DC 的方法; 3) 加权的 BC 和 DC 方法; 4) 我们的方法: 加权的 BC、DC 和 DFU 的方法。

图 5-9, 5-10, 5-11 分别显示了第一到第三组图像序列中模特头像的转角和各帧之间的关系。图中转角的真值用实线表示, 虚线则表示用各种方法得到的估计值, 四张图是依次是使用上述四种方法得到的结果。

对三组序列来说, 只用灰度约束的方法得到的结果很不准确, 原因是灰度约束只利用了图像的灰度信息, 非常容易受到噪声、光照变化以及遮蔽的影响。联合使用灰度和深度约束的方法能够使结果得到明显的改善, 除了绕 Z 轴的旋转以

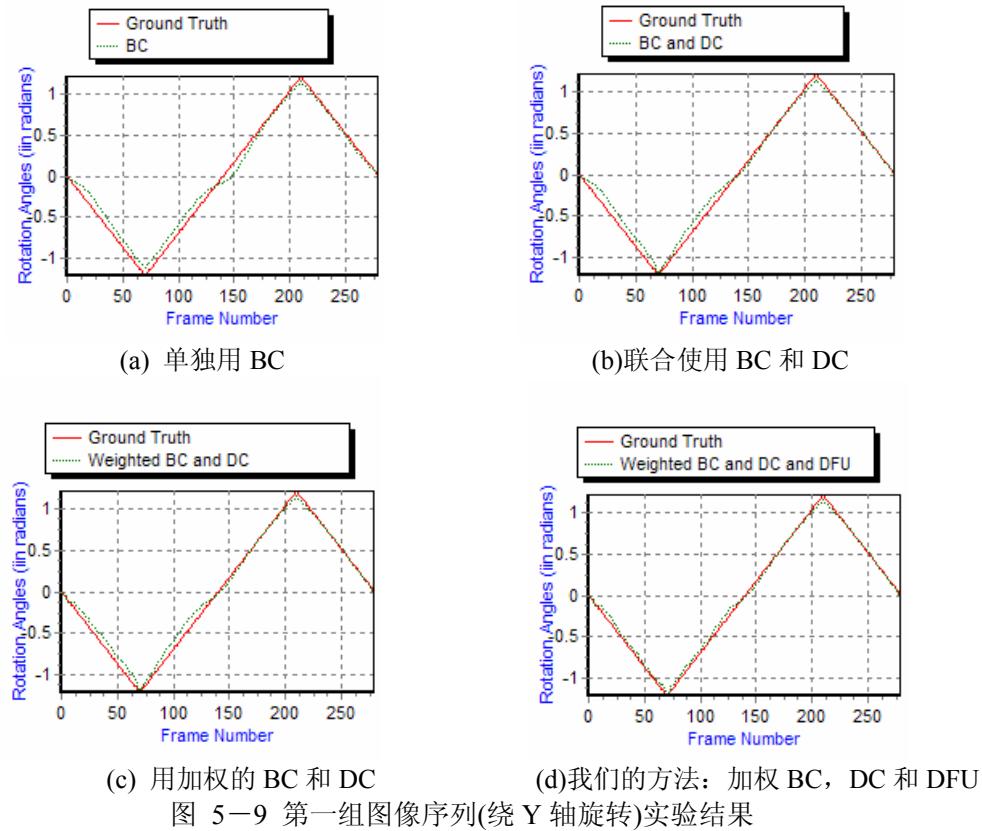
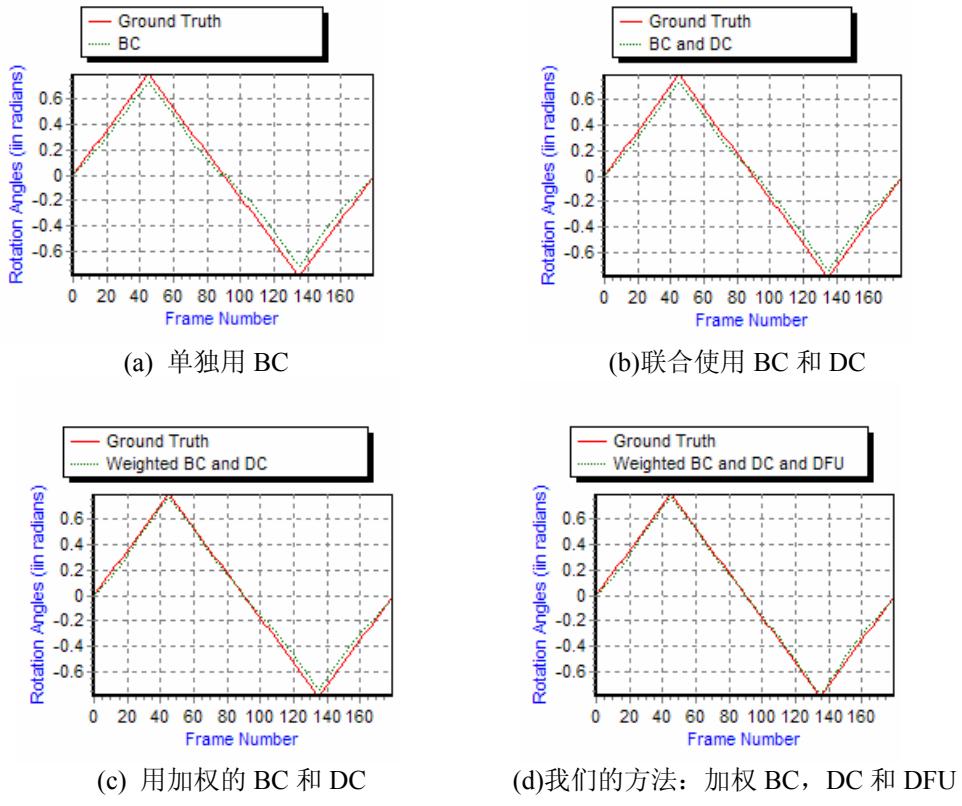


图 5—9 第一组图像序列(绕 Y 轴旋转)实验结果

图 5—10 模特头像的转角真实值(实线)和估计值(虚线)的比较
第二组图像序列(绕 X 轴旋转)实验结果

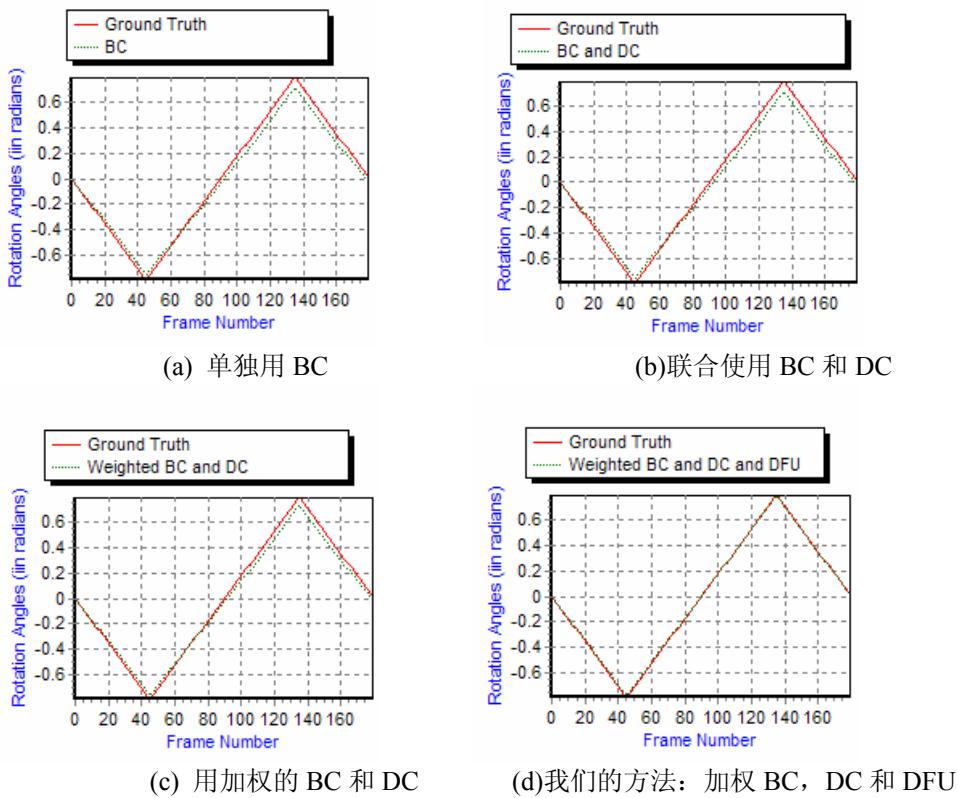


图 5-11 模特头像的转角真实值(实线)和估计值(虚线)的比较
第三组图像序列(绕 Z 轴旋转)实验结果

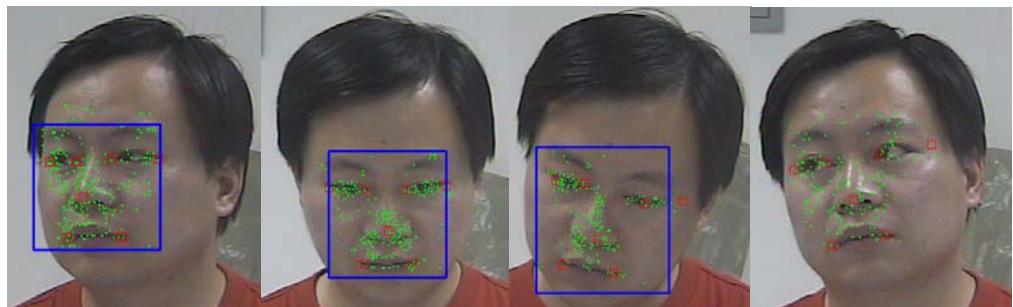
外，因为这种运动没有明显的深度方向的变化。基于加权的灰度和深度约束的方法能有效地减少误差，尤其是当人脸运动到两个极值点的时候，此时最容易出现遮蔽和不可靠的特征点。使用动态特征更新的最显著的效果是能够消除较长时间跟踪产生的积累误差。在我们的三组序列的实验中，最大的误差和整个旋转角的比值约为 3%。

5.9.3. 真实人脸图像序列姿态跟踪实验分析

为了直观地检验我们的算法在真实人脸图像序列上的有效性。我们采集了几组真实人脸运动的图像序列，包括绕 X, Y, Z 轴转动和自由运动的情况。由于真实人脸数据无法准确测量，我们将估计结果分别以图表和三维模拟图的形式表示，从视觉效果上证明了我们的方法能够得到比较精确的估计结果。

我们用摄像机分别记录人脸绕 X 轴、Y 轴和自由运动的三组图像序列(每组 190 帧)。图 5-12(a)显示了从这三组序列中分别选取的四帧图像,还有分别使用

联合灰度深度约束方法(b)和使用我们的方法(c)得到的结果。可以很直观地看出



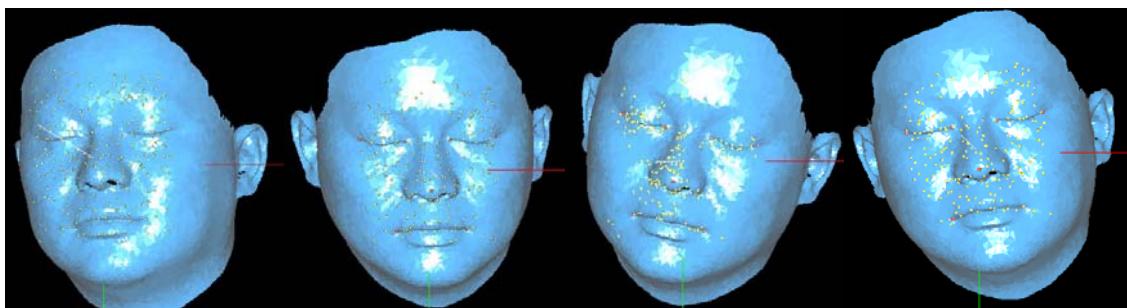
序列 1 第 90 帧

序列 2 第 145 帧

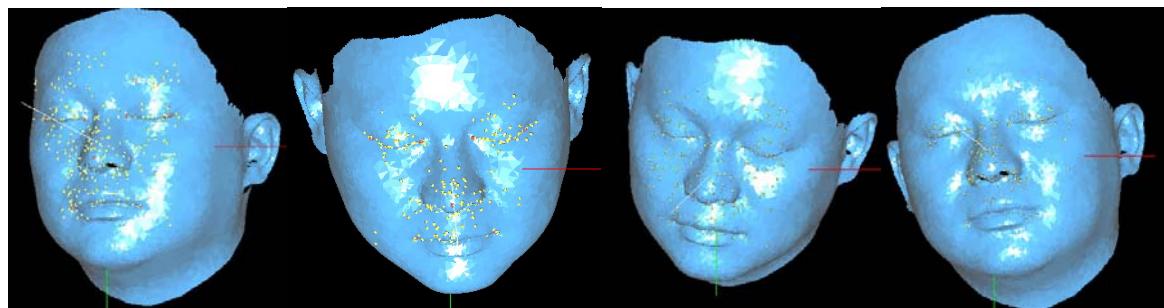
序列 3 第 187 帧

序列 3 第 122 帧

(a) 分别从三组序列中选择的几帧图像

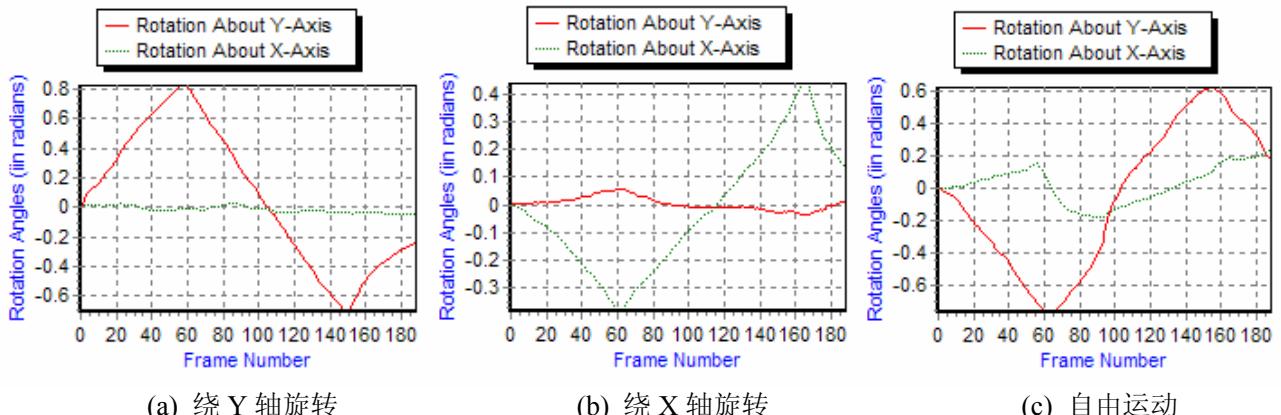


(b) 用 BC 和 DC 得到的估计结果



(c) 用我们的方法得到的估计结果

图 5-12 对三组真实图像序列的实验结果



(a) 绕 Y 轴旋转

(b) 绕 X 轴旋转

(c) 自由运动

图 5-13 转角真实值(实线)和估计值(虚线)的比较
三组真实图像序列实验结果

我们的方法得到估计结果比较准确，说明加权和特征动态更新算法具有比较明显的减小误差的作用。图 5—13 是对这三组图像序列的估计结果，其中实线是绕 Y 轴转角，虚线是绕 X 轴转角。由于该方法通过多次旋转模型并用统计的方法得到深度对时间的导数，因此该算法的运算速度较慢，在 PIV 2.4G 的微机上处理速度约为 0.45 秒/帧。

第六章 结束语

6.1. 总结

本文对单目视频图像序列中人脸姿态跟踪的问题进行了比较深入的研究，提出了一种用人脸三维模型提供深度度量，进而求解人脸姿态参数的方法。

该方法将姿态跟踪过程分为两个步骤，一是基于仿射对应原理的初始帧姿态估计；二是基于帧间线性的灰度深度约束加权的人脸姿态跟踪。在系统工作之前先进行初始化的工作，包括人脸的三维建模，摄像机标定和初始特征点对的选取。

基于二维/三维特征点对应计算姿态参数实际上是求解一个非线性问题，一般来说，难于获得闭合形式的解，往往需要给出一个初始估计值，并通过一个迭代优化的过程来求解。如果初始估计值不够准确，则可能陷入局部最小或不收敛。我们通过选取人脸特定部位的特征点，构造能很好通过这些点的平面作为人脸的近似；然后基于仿射对应的原理，将摄像机坐标系旋转三次得到人脸姿态参数的粗略估计值；最后将粗略估计值代入一个非线性优化框架中求解姿态参数。在求解的过程中，我们用鲁棒方法检测出不准确的特征匹配，以权值的形式体现在非线性优化目标函数中。

由于引入了人脸三维模型，我们的方法不象一般的仿射对应方法那样需要获取正面平行的参考帧；构造的虚拟正面平行投影能更好地满足仿射对应的条件；三次旋转摄像机坐标系不仅得到人脸平面的法线还能得到全部的运动参数；将基于仿射对应得到的结果，作为运动参数的初始估计值，能够摆脱平面假设的限制，得到更精确的结果。

初始帧姿态估计结束以后，我们用 KLT 准则在人脸上生成更多的特征点并反投影到三维模型上得到相应的特征点对，接下来就进入人脸姿态跟踪的第二个步骤，帧间运动参数估计。我们首先基于帧间微小的刚性运动的假设，推导出线性的深度和灰度约束方程；然后根据三维模型提供的几何信息对位于人脸不同部位的特征点求出相应的权值并对深度灰度约束加权；为了得到深度对时间的方向导数，我们将三维模型旋转多次，使投影误差最小，从而得到深度对时间的方向导数；联立深度灰度约束方程，用加权的最小二乘法求解帧间运动参数；最后为了

解决光照变化、遮蔽和人脸表面变化的影响，我们提出一种算法，能逐帧自动判断特征的可靠性，及时丢弃误差较大的特征点，并补充新的特征点用于跟踪过程。这种算法在一定程度上能降低跟踪过程中人脸形状的微小变化带来的不利影响。在二维特征跟踪的过程中，我们对特征区域做 FFT 变换，并对低频成分做最优化操作，这样就能减少由于特征区域灰度的非线性分布对二维运动参数造成的不利影响，提高二维跟踪的精度。

我们的方法存在两个主要的缺点。一是依赖于特定的人脸三维模型，。虽然在许多特定的应用场合，如医院智能护理系统、视频会议系统、司机疲劳检测系统等，系统处理对象（如病人、与会人员、司机等）是特定的，其三维模型能够预先获得，因此基于特定的人脸模型的算法具有一定的实用价值。但是依赖于特定模型在很大程度上降低了该方法的实际性。二是计算代价较高。由于无法获得帧速率的深度信息，我们不得不多次旋转三维模型以得到满意的深度测量值，此外非线性的优化迭代过程和 FFT 也造成了一定的计算开销。

6.2. 展望

由于人脸三维模型包含了丰富的几何信息，将其引入姿态跟踪系统，与图像信息相结合，就可能得到更精确、更鲁棒的人脸姿态估计结果。随着三维扫描技术的不断发展，实时扫描技术正慢慢走向成熟。将来我们可以用实时三维扫描为许多应用提供快捷而精确的三维信息。

针对我们系统的问题，将来可以从以下几个方面做进一步的研究：一是针对基于特定人三维模型的弊端，可以考虑基于统计方法构造通用的人脸三维模型，并通过某种变换拟合到特定人脸上从而得到与特定人相关的几何信息；二是利用实时扫描技术得到帧速率的深度测量，从而可以直接求解线性约束方程，这样就能避免反复旋转模型带来的开销。三是研究如何处理当人脸运动速度较快或表面形状发生较大变化的时候对姿态跟踪的不利影响。

附录 I Levenburg-Marquet 算法

Levenburg-Marquet 是一种求解非线性最小二乘问题的最优化算法在给定的初始值不偏离真实值太远的条件下可以通过一个迭代过程得到局部最优解。

假设有非线性最小二乘问题

$$\min F(x) = \sum_{i=1}^m f_i^2(x) \quad (\text{I-1})$$

其中 $f_i(x)$ 是 x 的非线性函数。因此需要将非线性问题转化为一系列线性最小二乘问题来求解。其基本思路是：设 $x^{(k)}$ 是解的第 k 次近似，在 $x^{(k)}$ 将 $f_i(x)$ 线性化，然后运用线性最小二乘求解最优值的公式得到极小点 $x^{(k+1)}$ 作为非线性最小二乘问题解的第 $k+1$ 次近似，再从 $x^{(k+1)}$ 出发重复以上过程得到非线性最小二乘的局部最优解。在 $x^{(k)}$ 对 $f_i(x)$ 做一阶泰勒展开得到

$$\varphi_i(x) = f_i(x^{(k)}) + \nabla f_i(x^{(k)})^T (x - x^{(k)})$$

令

$$\phi(x) = \sum_{i=1}^m \varphi_i^2(x) \quad (\text{I-2})$$

则可以用 $\phi(x)$ 的极小点来近似 $F(x)$ 的极小点。

对线性极小问题 $\min \phi(x) = \sum_{i=1}^m \varphi_i^2(x)$, 令

$$A_k = \begin{bmatrix} \nabla f_1(x^{(k)})^T \\ \vdots \\ \nabla f_m(x^{(k)})^T \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1(x^{(k)})}{\partial x_1} & \frac{\partial f_1(x^{(k)})}{\partial x_2} & \dots & \frac{\partial f_1(x^{(k)})}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_m(x^{(k)})}{\partial x_1} & \frac{\partial f_m(x^{(k)})}{\partial x_2} & \dots & \frac{\partial f_m(x^{(k)})}{\partial x_n} \end{bmatrix},$$

$$b = \begin{bmatrix} \nabla f_i(x^{(k)})^T x^{(k)} - f_i(x^{(k)})^T \\ \vdots \\ \nabla f_m(x^{(k)})^T x^{(k)} - f_m(x^{(k)})^T \end{bmatrix} = A_k - f^{(k)}, \text{ 其中 } f^{(k)} = \begin{bmatrix} f_1(x^{(k)}) \\ f_2(x^{(k)}) \\ \vdots \\ f_m(x^{(k)}) \end{bmatrix}$$

则(I-2)可以写成

$$\phi(x) = (A_k x - b)^T (A_k x - b) \quad (I-3)$$

则 $\phi(x)$ 的平稳点满足 $A_k^T A_k (x - x^{(k)}) = -A_k^T f^{(k)}$ ，这时的迭代方向为

$$d^{(k)} = -(A_k^T A_k)^{-1} A_k^T f^{(k)} \quad (I-4)$$

在实际计算中，有时 $A_k^T A_k$ 会出现奇异或接近奇异的情况，这时求解 $(A_k^T A_k)^{-1}$ 会很困难甚至无法进行，因此用 Marquet 方法对 $A_k^T A_k$ 进行修正，基本思想就是将一个正定对角阵加到 $A_k^T A_k$ 使原矩阵变成条件数比较好的对称正定矩阵，得到修正之后的的迭代方向为

$$d^{(k)} = -(A_k^T A_k + \alpha_k I)^{-1} A_k^T f^{(k)} \quad (I-5)$$

这样就得到 $\phi(x)$ 的极小点 $x^{(k+1)} = x^{(k)} + d^{(k)}$ 作为下一轮迭代的初始值，控制迭代误差 $\|x^{(k+1)} - x^{(k)}\| < \varepsilon_0$ 则停止迭代得到该非线性最小二乘问题的局部最优解。

图表目录

图 1-1 人脸识别和表情识别.....	3
图 1-2 视线跟踪.....	4
图 1-3 基于模型的视频会议系统.....	4
图 1-4 安全监控系统.....	5
图 2-1 基于特征空间的方法.....	10
图 2-2 Gabor 滤波器的效果.....	10
图 2-3 在 Gabor 特征空间中估计人脸姿态.....	11
图 2-4 用肤色发色特征估计人脸姿态.....	12
图 2-5 基于弱透视投影和简单几何结构的方法.....	13
图 2-6 基于人体测量学和简单几何结构的方法.....	14
图 2-7 基于人脸平面假设和仿射对应的方法.....	14
图 2-8 基于椭圆假设的方法.....	15
图 2-9 人脸圆柱模型.....	15
图 2-10 基于三维模型和特征对应的方法.....	22
图 2-11 提取出稳定边和变化边的人脸模型.....	22
图 3-1 从视频图像序列中计算人脸的姿态参数.....	23
图 3-2 基于三维模型的人脸姿态跟踪系统框架.....	24
图 3-3 人脸姿态跟踪系统的两个组成部分.....	25
图 3-4 基于仿射变换的人脸姿态估计方法.....	26
图 3-5 人脸三维建模的过程.....	26
图 3-6 VIVID 910 激光扫描仪	27
图 3-7 多块三维数据的配准.....	29
图 3-8 经过预处理之后精度较高的人脸三维数据.....	30
图 3-9 图像坐标系.....	30
图 3-10 摄像机坐标系和世界坐标系.....	32
图 3-11 摄像机平移运动的几何关系.....	34
图 4-1 初始帧姿态参数估计流程图.....	38
图 4-2 构造特征点正面平行投影.....	39
图 4-3 旋转坐标轴三次得到人脸平面姿态参数.....	44
图 4-4 构造椭圆锥.....	46
图 4-5 摄像机坐标系第一次旋转.....	46
图 4-6 摄像机坐标系第二次旋转.....	49

图 4-7 第二次旋转角的多解现象.....	50
图 4-8 求摄像机坐标系的第三次旋转角.....	50
图 4-9 人脸平面假设验证实验所用的三维模型.....	55
图 4-10 人脸平面假设验证实验结果.....	56
图 4-11 模拟数据实验结果.....	56
图 4-12 特征点位置出现误差时的实验结果(均为绕 Y 轴旋转的情况)	57
图 4-13 用基于仿射对应的方法计算石膏像图像的姿态参数.....	59
图 4-14 用基于仿射对应的方法计算真实人脸图像的姿态参数.....	59
图 5-1 基于深度和灰度约束加权的人脸姿态跟踪.....	61
图 5-2 用 KLT 准则生成二维和三维特征对.....	64
图 5-3 计算特征点权值.....	70
图 5-4 动态特征点更新算法.....	75
图 5-5 帧间运动幅度较大的图像序列.....	77
表 5-1 第 30 帧特征点位置跟踪误差.....	77
图 5-6 KLT 和改进的 KLT 方法对二维特征跟踪的比较.....	77
图 5-7 模特头像图象及其三维数据.....	78
图 5-8 三组模特头像图象序列.....	79
图 5-9 第一组图像序列(绕 Y 轴旋转)实验结果.....	80
图 5-10 模特头像的转角真实值(实线)和估计值(虚线)的比较	80
图 5-11 模特头像的转角真实值(实线)和估计值(虚线)的比较.....	81
图 5-12 对三组真实图像序列的实验结果.....	82
图 5-13 转角真实值(实线)和估计值(虚线)的比较	82

参考文献

- [1] M. H. Yang, D. J. Kriegman, N. Ahuja, Detecting faces in images: a survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34-58, 2002.
- [2] H.A. Rowley, S. Baluja, T. Kanade, Neural network-based face detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):23-38, 1998.
- [3] H.A. Rowley, S. Baluja, T. Kanade, Rotation invariant neural network-based face detection, In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 963 – 963, 1998.
- [4] R. Chellappa, C. L. Wilson, S. Sirohey, Human and machine recognition of faces: a survey, In: *Proceedings of the IEEE* 83(5): 705-740, 1995
- [5] T. Fromherz, P. Stucki, M. Bichsel, A survey of face recognition, *MML TechnicalReport*, No 97.01, Dept. of Computer Science, University of Zurich, Zurich, 1997.
- [6] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using Laplacianfaces, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3): 328 – 340, 2005.
- [7] T. Darrell, G. Gordon, M. Harville, J. Woodfill, Integrated person tracking using stereo, color, and pattern detection, In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 601 – 608, 1998.
- [8] V. I. Pavlovic, R. Sharma, T. S. Huang, Visual interpretation of hand gestures for humancomputer interaction: a review, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7): 677-695, 1997.
- [9] B.W. Miners, O.A. Basir, M.S. Kamel, Understanding hand gestures using approximate graph matching, *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 35(2): 239 – 248, 2005.
- [10] P.R.G. Harding, T.J. Ellis, Recognizing hand gesture using Fourier descriptors, In: *Proceedings of the 17th International Conference on Pattern Recognition*, 3:286-289, 2004.
- [11] M. S. Nixon and J. N. Carter, D. Cunado, P. S. Huang, S. V. Stevenage, Automatic gait recognition, *IEEE Colloquium on Motion Analysis and Tracking* (Ref. No. 1999/103), 3/1 - 3/6, 1999.
- [12] P.S. Huang, Automatic gait recognition via statistical approaches for extended template features, *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 31(5): 818 – 824, 2001.
- [13] M. Black, Y. Yacoob, Recognizing facial expressions in image sequences using local parameterized models of image motion, *International Journal of Computer Vision*, 25(1):23-48, 1997.

- [14] L. Ma, K. Khorasani, Facial expression recognition using constructive feedforward neural networks, *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 34(3):1588 – 1595, 2004.
- [15] X. Zhou, X. Huang, B. Xu, Y. Wang, Real-time facial expression recognition based on boosted embedded Hidden Markov Model, In: Proceedings of Third International Conference on Image and Graphics, 290 – 293, 2004.
- [16] H. Hienz, K. Grobel, G. Offner, Real-time hand-arm movement analysis using a single video camera, In: Proceedings of 2nd International Conference on Automatic Face and Gesture Recognition, 323-327, 1996.
- [17] Q. Chen, H. Wu, T. Fukumoto, M. Yachida, 3D head pose estimation without feature tracking, In: Proceedings of Third IEEE International Conference on Automatic Face and Gesture Recognition, 88-93, 1998.
- [18] T. Darrel, B. Moghaddam, A.P. Pentland, Active face tracking and pose estimation in an interactive room, In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 67-72, 1996.
- [19] T. Hogg, D. Rees, H. Talhami, Three-dimensional pose from two-dimensional images: a novel approach using synergetic networks, In: Proceedings of IEEE International Conference on Neural Networks, 2:1140-1144, 1995.
- [20] H. Murase, S.K. Nayar, Visual learning and recognition of 3D objects from Appearance, *International Journal of ComputerVision*, 14(1): 5-24, 1995.
- [21] R. Rae, H.J. Ritter, Recognition of human head orientation based on artificial neural networks, *IEEE Transactions on Neural Networks*. 9:257-265, 1998.
- [22] Z. Liu, Z. Zhang, Robust head motion computation by taking advantage of physical properties, In: Proceedings of Workshop on Human Motion, 73-77, 2000.
- [23] D.G. Lowe, Integrated treatment of matching and measurement errors for robust model-based motion tracking, In: Proceedings of Third International Conference on Computer Vision, 436-440, 1990.
- [24] D.B. Genney, Visual tracking of known three-dimensional object, *International Journal of Computer Science*, 7(3):243-270, 1992.
- [25] T.S. Jebara, A. Pentland, Parametrized structure from motion for 3D adaptive feedback tracking of faces, In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 144 – 150, 1997.
- [26] P. Yao, G. Evans, A. Calway, Using affine correspondence to estimate 3-D facial pose, In: Proceedings of IEEE Int. Conf. on Image Processing, 3: 919-922, 2001.
- [27] G.D. Hager, P.N. Belhumeur, Efficient region tracking with parametric models of geometry and illumination, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(10):1025-1039, 1998.
- [28] Q.Ji, 3D face pose estimation and tracking from a monocular camera, *Image and Vision Computing*, 20:499-511, 2002.
- [29] J. Xiao, T. Kanade, J.F. Cohn Robust full-motion recovery of head by dynamic templates and re-registration techniques, In Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, 593-600, 2002.
- [30] M.La Cascia, S. Sclaroff, V. Athitsos, Fast, reliable head tracking under varying

- illumination: an approach based on registration of texture-mapped 3D models, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(4):322- 336, 2000.
- [31] M.La Cascia, S. Sclaroff, Fast, reliable head tracking under varying illumination, In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1:23-25, 1999.
- [32] M. La Cascia, J. Isidoro, S. Sclaroff, Head tracking via robust registration in texture map images, In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 508 – 514, 1998.
- [33] S.Y. Ho, H. Huang, An analytic solution for the pose determination of human faces from a monocular image, *Pattern Recognition Letters*, 19: 1045-1054, 1998.
- [34] A.T. Horprasert, Y. Yacoob, L.S.Davis, Computing 3D head orientation from a monocular image sequence, In: Proceedings of SPIE-The International Society for Optical Engineering 25th AIPR workshop: Emerging Applications of Computer Vision, 2962:244-252, 1996.
- [35] A. Nikolaidis, I. Pitas, Facial feature extraction and determination of pose, <http://www.citeseer.nj.nmec.com/cs>, pp.1-6.
- [36] A. Gee, R. Cipolla, Estimating gaze from a single view of a face, In: Proceedings of the 12th IAPR International Conference on Pattern Recognition, Conference A: Computer Vision & Image Processing., 1:758 – 760, 1994.
- [37] I. Shimizu, Z. Zhang, S. Akamatsu, K. Deguchi, Head pose determination from one image using a generic model, In: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, 100-105, 1998.
- [38] R. Lopez, T.S.Huang, 3D Head pose computation from 2D images: template versus features, In: Proceedings of IEEE International Conference on Image Processing, 2:599 -602, 1995.
- [39] R. Yang, Z. Zhang, Model-based head pose tracking with stereo vision, In: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, 255-260, 2002.
- [40] G. Liang, H. Zha, H. Liu, Affine correspondence based head pose estimation for a sequence of images by using a 3D model, In: Proceedings of IEEE 6th International Conference on Automatic Face and Gesture Recognition, 632-637, 2004.
- [41] M. Berger, T. Auer, G. Bachler, S. Scherer, A. Pinz, 3D model based pose determination in real-time: strategies, convergence, accuracy, In: Proceedings of 15th International Conference on Pattern Recognition, 4:567-570, 2000.
- [42] M. Xu, T. Akatsuka, Detecting head pose from stereo image sequence for active face recognition, In: Proceedings of Third IEEE International Conference on Automatic Face and Gesture Recognition, 82-87, 1998
- [43] Q. Delamarre, O. Faugeras, Finding pose of hand in video images: a stereo-based approach, In: Proceedings of Third IEEE International Conference on Automatic Face and Gesture Recognition, 585-590, 1998.
- [44] R. Newman, Y. Matsumoto, S. Rougeaux, A. Zelinsky, Real-time stereo tracking

- for head pose and gaze estimation, In: Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 122-128, 2000.
- [45] E. Sung, Stereo head/face tracking and pose estimation, In Proceedings of 7th International Conference on Control, Automation, Robotics and Vision, 3:1609-1614, 2002.
- [46] J. Woodfill, B. Von Herzen, Real-time stereo vision on the PARTS reconfigurable computer, In: Proceedings of The 5th Annual IEEE Symposium on FPGAs for Custom Computing Machines, 201-210, 1997.
- [47] Y. Wei, L. Fradet, T. Tan, Head pose estimation using Gabor eigenspace modeling, In: Proceedings of International Conference on Image Processing, 1:22-25, 2002.
- [48] T. Lee, Image representation: using 2D gabor wavelet, IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(10): 959-971, 1996.
- [49] T. Shiyoyama, Q. Chen, H. Wu, T. Shimada, 3D head pose estimation using color information, In: Proceedings of IEEE Conference on Multimedia Computing and System, 1:697-702, 1999.
- [50] M.A. Fishler, R. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Comm.ACM, 24(6):381-395, 1981.
- [51] W. Wolfe, D. Mathis, C. Sklair, M. Magee, The perspective view of three points, IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(1):66-73, 1991.
- [52] D.P. Huttenlocker, S. Ullman, Recognizing solid objects by alignment, presented at the DARPA Image Understanding Workshop, 1114-1122, 1988.
- [53] O. Faugeras, Three Dimensional Computer Vision. Cambridge: MIT Press, 1993.
- [54] S.D. Ma, Z.Y. Zhang. Computer Vision: Computation Theory and Algorithm Basis. Beijing: Science Press, 1998 (in Chinese)
(马颂德,张正友, 计算机视觉——理论和算法, 科学出版社, 1998)
- [55] M.L. Liu, K.H. Wong, Pose estimation using four corresponding points, Pattern Recognition Letters, 20:69-74, 1999.
- [56] J. Wang, E. Sung, Pose determination of human faces by using vanishing points, Pattern Recognition, 34:2427-2445, 2001.
- [57] J.B. Huang, Z. Chen, J.Y. Lin A study on the dual vanishing point property, Pattern Recognition, 12:2029-2039, 1999.
- [58] F. Schaffalitzky, A. Zisserman, Planar grouping for automatic detection of vanishing lines and points, Image and Vision Computing, 9: 647-658, 2000.
- [59] P. Parodi, G. Piccioli, 3D shape reconstruction by using vanishing points, IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(2):211.217, 1996.
- [60] B. Brillault, O. Mahony, New method for vanishing points from detection, CVGUP: Image Understanding 54(2):289-300, 1991.

- [61] M.J. Magee, J.K. Aggarwal, Determining vanishing points from perspective images, *Computer Vision Graphics Image Process.* 26:256-267, 1984.
- [62] B.D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, In: *Proceedings of International Joint Conference Artificial Intelligence*, 674-679, 1981.
- [63] J. Shi, C. Tomasi, Good features to track, In: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 593-600, 1994.
- [64] V. Lepetit, J. Pilet, P. Fua, Point matching as a classification problem for fast and robust object pose estimation, In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:244-250, 2004.
- [65] J. Yao, W. Cham, Efficient model-based linear head motion recovery from "movies", In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:414-421, 2004.
- [66] S.Y. Park, M. Subbarao, Pose estimation and integration for complete 3D model reconstruction, In: *Proceedings of Sixth IEEE Workshop on Applications of Computer Vision*, 143-147, 2002.
- [67] D.A. Simon, M. Hebert, T. Kanade, Real-time 3-D pose estimation using a high-speed range sensor, In: *Proceedings of IEEE International Conference on Robotics and Automation*, 3:2235-2241, 1994.
- [68] N. Burtnyk, M. Greenspan, Signature search method for 3-D pose refinement with range data, In: *Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 312-319, 1994.
- [69] R. Zabih, J. Woodfill, M. Withgott, A real-time system for automatically annotating unstructured image sequences, In: *Proceedings of International Conference on Systems, Man and Cybernetics*, 1993. 'Systems Engineering in the Service of Humans', Conference, 2:345 – 350, 1993.
- [70] J. Deng, F. Lai, Region-based template deformation and masking for eye-feature extraction and description, *Pattern Recognition*, 30(30): 403-419, 1997.
- [71] T. Cootes, G. Edwards, C. Taylor, Active appearance models, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681-685, 2001.
- [72] G.D. Hager, P.N. Belhumeur, Real-time tracking of image regions with changes in geometry and illumination, In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 403-410, 1996.
- [73] T.Y. Tian, C. Tomasi, D.J. Heeger, Comparison of approaches to egomotion computation, In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 315-320, 1996.
- [74] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical recipes in C: the art of scientific computing*, Cambridge: Cambridge University Press, 1998.
- [75] J. Bergen, P. Anandan, K. Hanna, R. Higorani, Hierarchical model-based motion

- estimation, In: Proceedings of European Conference on Computer Vision, 237-252, 1992.
- [76] M. Harville, A. Rahimi, T. Darrell, G. Gordon, J. Woodfill, 3D pose tracking with linear depth and brightness constraints, In: Proceedings of IEEE 7th International Conference on Computer Vision, 1: 206-213, 1999.
- [77] S. Basu, I. Essa, A. Pentland, Motion regularization for model-based head tracking, In: Proceedings of International Conference on Pattern Recognition, 3:611-616, 1996.
- [78] C. Bregler, J. Malik, Tracking people with twists and exponential maps, In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 8-15, 1998.
- [79] B.K.P. Horn, E.J. Weldon, Direct methods for recovering motion, International Journal of Computer Vision, 2:51-76, 1988.
- [80] J.Y. Shieh, H. Zhuang, R. Sudhakar, Motion estimation from a sequence of stereo images: a direct method, IEEE Transactions on Systems, Man and Cybernetics, 24(7):1044 – 1053, 1994.
- [81] G.P. Stein, A. Shashua, Model-based brightness constraints: on direct estimation of structure and motion, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(9):992–1015, 2000.
- [82] M. Black, Y. Yacoob, Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion, In: Proceedings of International Conference on Computer Vision, 374-381, 1995.
- [83] A. Yilmaz, K. Shafique, M. Shah, Estimation of rigid and non-rigid facial motion using anatomical face model, In: Proceedings of 16th International Conference on Pattern Recognition, 1:377-380, 2002.
- [84] A. Del Bue, F. Smeraldi, L. Agapito. Non-rigid structure from motion using non-parametric tracking and non-linear optimization, In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04), 1:8-8, 2004.
- [85] A. Del Bue, L. Agapito. Non-rigid 3D shape recovery using stereo factorization, In: Proceedings of Asian Conference of Computer Vision (ACCV2004), 1:25-30, 2004.
- [86] L. Torresani, D. Yang, E. Alexander, C. Bregler, Tracking and modeling non-rigid objects with rank constraints, In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1:I493-I500, 2001.
- [87] C. Bregler, A. Hertzmann, and H Biermann, Recovering non-rigid 3d shape from image streams, In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 690–696, 2000.
- [88] M. Brand, Bhotika R, Flexible flow for 3d nonrigid tracking and shape recovery, In: Proceedings of IEEE Conference on Computer Vision and Pattern

Recognition, 315–22, 2001.

- [89] X.M. Mei, J.Z. Huang. Differential Geometry. Beijing: People's Education Press, 1981.(in Chinese)
(梅向明, 黄敬之. 微分几何. 北京: 人民教育出版社, 1981 年)
- [90] D. DeCarlo, D. Metaxas, The integration of optical flow and deformable models with applications to human face shape and motion estimation, In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 231-238, 1996.

已发表和录用的论文

-
- [1] G. Liang, H. Zha, H. Liu, Affine correspondence based head pose estimation for a sequence of images by using a 3D model. In: Proceedings of IEEE 6th International Conference on Automatic Face and Gesture Recognition (FG 2004), 632-637, 2004.
(EI 检索)
- [2] G. Liang, H. Zha, H. Liu, 3D model based head pose tracking by using weighted depth and brightness constraints, In: Proceedings of Third International Conference on Image and Graphics (ICIG 2004), 481-484, 2004.
(EI 检索)
- [3] 梁国远, 查红彬, 刘宏. 基于三维模型和仿射对应原理的人脸姿态估计方法. 计算机学报, 28(5):792-800, 2005.
(EI 检索)
- [4] 梁国远, 查红彬, 刘宏. 基于三维模型和深度灰度约束加权的人脸姿态跟踪. 计算机辅助设计和图形学学报. (已录用)
(EI 检索)

参与的科研项目

-
- [1]用于患者行为实时跟踪的护理机器人主动视觉研究, 国家自然科学基金项目, 2002-2004, 项目号: 60175025。

致谢

在论文完成之际，首先对我的导师查红彬教授表示深深的感谢。四年来，是查老师将我一步一步引入了计算机视觉的研究领域，使我从一个门外汉成长为具有独立科研能力的研究人员。查老师严谨的治学态度、渊博的学识和高尚的品格给我留下了深刻的印象，他的谆谆教诲对我今后的工作、学习和生活将产生深远的影响。

感谢刘宏副教授一直以来对我的支持和鼓励，他不仅在科研工作上对我悉心指导，而且在生活上也给予了无微不至的关心。每次和他的讨论，无论是学术还是非学术的内容，都令我获益匪浅。

感谢实验室的崔立农老师和孙立老师在我使用实验设备时提供的热情帮助。

感谢各位师兄弟张益贞、武宇文、冯所前、吴奇、王君秋、董宁、余泽、邓学智和已经毕业的师兄弟于行洲、王鹏、皮文凯、林飞等同学对我在生活和学习中的帮助，和你们在一起的日子充满了生活的乐趣。

感谢我的大哥梁国飞，在我多年的求学生涯中，一直是他承担起照顾父母和家庭的重任，使我得以安心学业。

感谢女友对我真挚热情的鼓励和毫无保留的支持，她陪伴我度过了一生中最美妙的时光。

最后要感谢我的父母，感谢他们二十多年来对我的养育和照顾，感谢他们对我始终如一的支持。

北京大学学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明： 所呈交的学位论文， 是本人在导师的指导下， 独立进行研究工作所取得的成果。除文中已经注明引用的内容外， 本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体， 均已在文中以明确方式标明。本声明的法律结果由本人承担。

论文作者签名： 日期： 年 月 日

学位论文使用授权说明

本人完全了解北京大学关于收集、保存、使用学位论文的规定，即：
按照学校要求提交学位论文的印刷本和电子版本；
学校有权保存学位论文的印刷本和电子版，并提供目录检索与阅览服务；
学校可以采用影印、缩印、数字化或其它复制手段保存论文；
在不以赢利为目的的前提下，学校可以公布论文的部分或全部内容。

(保密论文在解密后遵守此规定)

论文作者签名： 导师签名：
日期： 年 月 日