

密级:_____



中国科学院大学
University of Chinese Academy of Sciences

硕士学位论文

基于变形迁移的表情克隆及其应用

作者姓名: 冀鼎皇

指导教师: 夏时洪 研究员

中国科学院计算技术研究所

学位类别: 工学硕士

学科专业: 计算机应用

研究所: 中科院计算技术研究所

2013年5月

Expression Cloning based on Deformation Transfer

By

Dinghuang Ji

A Dissertation/Thesis Submitted to

University of Chinese Academy of Sciences

In partial fulfillment of the requirement

For the degree of

Master of Engineering

Institute of Computing Technology, CAS

April 2013

声 明


我声明本论文是我本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，本论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

作者签名：  日期： 2013.4.25

论文版权使用授权书

本人授权中国科学院计算技术研究所可以保留并向国家有关部门或机构送交本论文的复印件和电子文档，允许本论文被查阅和借阅，可以将本论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编本论文。

（保密论文在解密后适用本授权书。）

作者签名：  导师签名： 日期：

摘要

高真实感人脸动画是 3D 电影和游戏领域的一项重要技术，现有的制作工艺大都依赖丰富的人脸表情模型库。而工业界中制作表情模型，往往需要有美术天赋的技术人员，长时间对模型局部的调整和整体的修复。这是一个非常费时费力的过程，因而一套好的表情模型库都代价不菲。除了电影和游戏，一些小成本的应用，如移动终端游戏、视频通话、远程交互等也在使用 3D 人脸动画技术来增加应用的趣味性和沉浸感。但是，由于成本限制，它们不能使用运动捕获设备来采集人脸运动数据，也没有丰富的表情模型库。为了解决这些问题，本论文设计了两套系统来简化人脸动画的生成过程：

1. 自动的表情克隆系统，该系统可以利用现有的表情模型库，将人脸模型的表情迁移到新模型上。利用这套系统可以有效地扩充表情模型库。
2. 实时的视频驱动人脸动画系统。该系统利用表情克隆系统产生的表情基，通过视频接口捕获用户表情变化，在虚拟人脸化身上产生对应的表情。利用这套系统可以非常方便的产生有较高真实感的人脸动画。

本文针对简化用户交互过程，提高系统的计算效率以及实用性等方面进行了研究，提出了以下研究成果。

(1) 提出了一种自动提取模型对应点的方法

表情克隆算法需要对现有模型和新模型进行配准，建立它们顶点和三角面片之间的对应关系。而模型配准一般都需要首先建立一些特征点的对应关系，比如找到模型之间眼角、嘴角等特征点的对应。现有的方法基本都是采用手工选点的方法。为了简化用户交互过程，本文提出一种自动选取模型对应点的方法。首先利用启发式方法选择一组模型对应特征点，然后利用径向基函数插值获得更多的对应点。使用该方法可以大大减少用户的手工交互操作。

(2) 提出了一种高效的模型配准优化方法

本文采用基于变形梯度的方法实现了表情克隆工作，该方法通过引入增广顶点来计算变形梯度。本文修改了变形梯度的矩阵表示，使得其计算过程不需要增广顶点。这一修正将建立模型配准过程中的优化变量减少到原来的三分之一，很大程度上提高了计算效率。虽然这种修正只是对变形梯度精确值的一种近似，实验表明，这种方法依然可以保持原算法的优化精度。

(3) 搭建了一套实时、低成本的视频驱动人脸动画合成系统

该系统使用普通的家用摄像头作为交互接口，用户只需要对摄像头做表情，便可以在 PC 上生成一个与自己有相似脸部表情和头部运动的虚拟人脸化身。该

系统实现了以下功能模块：实时人脸特征检测和姿态估计，可以较为准确的估计人脸表情变化和头部运动；运动参数化，可以将用户的脸部表情迁移到几何结构不同的虚拟人脸化身上面；人脸动画合成，使用表情基方法合成具有较高真实感的人脸动画。

关键词：人脸动画，表情库，表情克隆，模型配准，变形梯度，视频驱动

ABSTRACT

Realistic 3D facial animation generation is one of the key technologies in movie and game industry, current technology mainly depend on rich set of facial expression models which are called blend shapes. In the industry, building blend shapes often depend on talented artists' manually hand work. Iteratively adjust the local details and fix the whole model. Thus, a good set of blend shape often cost much expense. Other than movie and games, some low-budget applications, such as mobile device games, video chat, telepresence etc. also adopt 3D facial animation to make them more interesting and immersive. However, due to limit of budget, they can't use motion capture device to obtain expression data or hire artists to build rich blend shapes. In order to help solve these problems, we design two systems to simplify the process to generate facial animation.

1. Automatic expression cloning system, which make full use of existing blend shapes, and transfer expressions to new facial model. We can efficiently enrich blend shapes with this system.

2. Real time video driven facial animation system, which use blend shapes generated by expression cloning system, capture user's expression data, and generate similar expression with a 3D virtual avatar. Users can easily generate realistic facial animation with this system.

In order to simplify user interaction, improve the computation efficiency and practicality. This paper proposes the following contributions:

- (1) Propose an automatic correspondences selection method

Expression cloning algorithm require building correspondence of source model and target model, finding the correspondence between their vertices and faces, such as correspondences of eye corners and mouth corners. Existing methods mainly select correspondence manually. In order to simplify user's interaction, we propose an automatic method to select correspondence. First, select a set of correspondence by heuristic rules, then obtain more correspondence by RBF interpolation. This method can largely reduce user's manual work.

- (2) Propose an efficient model correspondence optimization method.

Our expression cloning system depend on a deformation gradient based method, which introduce augmented vertices for every triangular mesh to compute deformation

gradient. We propose a method to remove augmented vertices by modify matrix form of deformation gradient. This modification reduce the optimizing variables to one third of original method, which largely improve computation efficiency. Although this modification is only an approximation to exact value of deformation gradient, experiments show our method could still keep the optimization accuracy.

(3) Build a low cost, real time video driven facial animation generation system. The system use commodity camera as interaction interface, user make expressions to the camera, then he/she could get a 3D avatar making similar expressions. The proposed system has following functional modules: real time facial feature localization and gaze estimation, which capture expression change and head movement; motion parameterization, which transfer user's facial expression to virtual avatar; facial animation generation, which synthesize animation with blend shape method.

Key Words: Facial animation, blend shapes, expression cloning, model correspondence, deformation transfer

目录

图目录.....	VI
表目录.....	VII
1 绪论	1
1.1 研究背景.....	1
1.2 研究概述.....	5
1.2.1 表情基自动生成技术.....	6
1.2.2 运动数据获取技术.....	7
1.2.3 研究内容.....	8
1.2.4 主要贡献.....	9
1.3 论文组织.....	10
2 基于表情基的人脸动画合成方法研究综述	11
2.1 表情克隆方法.....	11
2.1.1 模型配准问题.....	11
2.1.2 运动迁移问题.....	16
2.1.3 其它表情克隆算法.....	18
2.2 视频驱动人脸动画研究现状.....	19
2.2.1 背景和领域知识.....	19
2.2.2 研究现状分析.....	20
2.2.3 相关工作分析.....	22
2.3 小结.....	23
3 自动表情克隆算法	25
3.1 变形梯度.....	25
3.2 变形梯度迁移.....	27
3.3 模型配准.....	30
3.3.1 初始对应特征选择.....	30
3.3.2 基于变形梯度的模型变形.....	31
3.3.3 确定配准关系.....	34
3.4 实验.....	35
3.4.1 程序实现细节.....	35
3.4.2 实验结果分析.....	35
3.5 小结.....	39
4 视频驱动人脸动画	41
4.1 总体框架介绍.....	41
4.2 技术细节.....	42
4.2.1 约束局部模型.....	42
4.2.2 表情运动参数化.....	43
4.2.3 人脸动画合成.....	46
4.3 系统效果.....	47
4.4 小结.....	48
5 总结和展望	49
5.1 全文工作总结.....	49

5.2 下一步研究方向.....	50
------------------	----

图目录

图 1.1 人脸动画在影视中的应用.....	2
图 1.2 Vicon 系统和演员	3
图 1.3 基于物理的肌肉建模效果.....	3
图 1.4 “诡异谷”效应.....	3
图 1.5 人脸动画在游戏中的应用.....	4
图 1.6 人脸动画在移动终端中的应用.....	4
图 1.7 人脸动画在远程交互中的应用.....	5
图 1.8 《指环王》、《阿凡达》、《本杰明奇事》等影片中的人物造型.....	6
图 1.9 人脸表情基示例.....	6
图 1.10 结构光设备示意.....	7
图 1.11[NRG 10]采用的图模型.....	8
图 2.1 表情克隆图示.....	11
图 2.2 模型配准和运动迁移.....	12
图 2.3 ICP 方法图示	13
图 2.4 模版匹配示例.....	14
图 2.5 适用于组合优化法的分段非刚性变形模型.....	15
图 2.6 适用于组合优化法的非刚性变形模型.....	15
图 2.7 表情克隆[JU 01]算法流程	16
图 2.8 三角坐标示意图.....	17
图 2.9 几种人脸运动捕获方法.....	21
图 2.10 文献[JJJ 03]的算法流程	21
图 3.1 自由形式变形(Free-Form deformation)	25

图 3.2 变形梯度示意图(图片取自文献[RJ 04]).....	26
图 3.3 变形梯度优化计算(图片取自文献[RJ 04]).....	28
图 3.4 文献[RJ 04]中对人脸模型的对应点选择方法.....	30
图 3.5 启发式方法选择的 14 组对应点.....	31
图 3.6 本文中利用启发式方法对人脸模型的对应点选择.....	31
图 3.7 三角面片配对计算.....	34
图 3.8 本文方法进行模型配准中间结果.....	36
图 3.9 表 7 所用各种模型示意图.....	37
图 3.10 表情克隆结果.....	38
图 3.11 应用本文算法到动物模型上的结果.....	39
图 4.1 视频驱动人脸动画数据流程图.....	41
图 4.2 CLM 方法流程.....	43
图 4.3 文献[JJJ 03] 表情运动参数化方法.....	44
图 4.4 稠密的表情基模型——对应于稀疏表情基(红点表示).....	45
图 4.5 RBF 函数不同形式对应的插值结果.....	46
图 4.6 视频驱动人脸动画实验结果.....	47

表目录

表 1 模型配准和注册问题概述.....	14
表 2 表情动画制作技术比较.....	19
表 3 变形梯度法进行模型配准应遵循的约束.....	32
表 4 本文方法与原文([RJ 04])方法的区别.....	33
表 5 算法主要功能数据结构和函数表.....	35

表 6 不同初始匹配点情况下的表面重构误差比较.....	36
表 7 不同迭代次数情况下的表面重构误差比较.....	36
表 8 对不同模型进行配准时算法运行时间比较.....	37

1 绪论

根据英国科学家的研究，四个月大的婴儿就可以识别人脸表情。作为人们日常生活中最常见的事物，人们对脸部表情的文字描绘和绘画展示更是历史悠久。近年来，随着数字艺术的发展，特别是 CG 技术的突破，人脸表情动画被广泛使用在 3D 动画电影和游戏中。1988 年 Pixar 的动画短片“Tin Toy”是第一部获得奥斯卡奖的动画电影，而这部电影的主角就是一个表情变化不定的婴儿。工业界的巨大成功也引起了学术界的广泛重视，自此，表情动画的研究也越来越深入。从早期的参数化人脸模型发展到复杂的基于物理的脸部肌肉模拟，人脸动画合成的真实程度不断得到提高，几乎达到了难以分辨真假的程度，学术研究成果也已被广泛地应用于工业界各个领域。例如运动捕获系统被用来获得人脸表情变化信息，Light Stage 等设备被用来捕获人脸表面光学属性以渲染更真实的人脸等。虽然人脸动画研究已经非常成熟，但由于人脸的表面变形复杂多变，需要高精度的运动捕获设备和复杂的建模算法才能进行准确模拟，高真实感的人脸动画往往要耗费大量人力物力。如何实现低成本、方便、高效的人脸动画生成，平衡效率与动画真实感，是一个具有挑战性的课题。

本文针对人脸动画生成中存在的问题，提出了一种新颖的人脸动画生成方案。该方案的核心算法是一种自动的表情克隆方法，该方法可以有效的重用现有的人脸表情库，极大的减少人脸建模工作量。利用该方法生成的表情基，本文又设计了一套视频驱动生成人脸动画的系统。该系统使用普通家用摄像头来捕获用户的表情变化，并通过虚拟的人脸化身生成较高真实感的人脸动画。本文将重点研究如何简化用户交互过程、提高系统的计算效率以及实用性等方面，实现效率和真实感的良好平衡。

1.1 研究背景

人脸动画的发展大致经历了一下几个阶段：人脸参数化模型，数据驱动人脸动画和基于物理仿真的人脸建模。早在上世纪 70 年代，Parke 等人提出了参数化 3D 人脸模型，作为人脸动画的奠基性工作，它可以通过少数几个参数生成一些特定的脸部表情。80 年代，Waters 等人又提出了基于质点弹簧系统的脸部肌肉模型，可以生成更丰富的面部表情。90 年代，Blatt 等人提出了人脸变形模型，通过建立人脸模型顶点之间的对应关系并对面部表情进行属性(胖瘦、方圆、肤色)建模，可以生成大量具有指定属性的人脸模型。进入新世纪，随着运动捕获系

统(如 Vicon, 见图 1.2)的兴起, 数据驱动人脸动画被广泛应用于影视制作中。运动捕获系统可以捕获很细微的面部肌肉运动, 从而可以生成更精细的表情动画。最新的人脸动画技术很多对面部肌肉进行材料力学建模, 并采用有限元方法进行求解。这种建模方法可以产生近乎真实的肌肉颤动效果(见图 1.3), 不过计算量相当大。人脸动画领域借用了机器人领域的一个概念“Uncanny Valley”(诡异谷效应见图 1.4), 也就是当人脸动画越来越接近真实情况时, 任何异常的表现都会令其真实感大大降低。

下面介绍一下人脸动画在工业界中的应用。

(1) 影视游戏

成功的动画电影离不开性格鲜明, 表情丰富的动画人物, 而人脸动画技术则是人物制作不可或缺的因素。动画电影领域的巨头 Weta Digital, Pixar, Disney, Digital Domain 等无一例外都采用了最先进的技术和设备来制作人脸动画。早期的人脸动画主要依赖手绘和雕塑师进行手工雕刻。后来, 在电影《指环王》系列中, 运动捕获系统被首次应用于实践中, 被用来生成咕噜的表情。电影《黑客帝国》则首次使用了皮肤渲染技术, 来制作一些非真实存在的人。近期的电影《阿凡达》以及《本杰明奇事》等更是将这些技术发挥到极致, 产生了人们意想不到的制作效果。



图 1.1 人脸动画在影视中的应用



图 2.2 Vicon 系统和演员



图 3.3 基于物理的肌肉建模效果

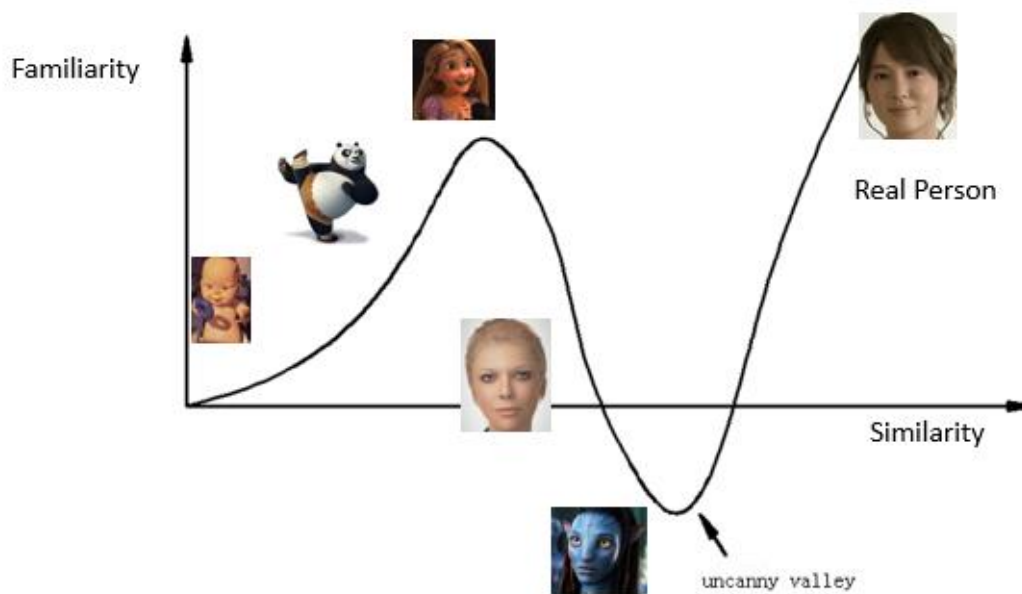


图 4.4 “诡异谷”效应

在大型 3D 游戏中，人脸动画也扮演着越来越重要的作用。许多游戏的宣传片中都包含非常逼真的人脸动画。由于游戏的实时性，在游戏中人脸动画不能得到最精细的展示，但是玩家也希望他们的游戏人物可以通过表情动画展现自己在游戏中的个性。于是，一些 RPG 游戏纷纷加入了人物表情动画，可以随着用户说话的内容改变表情。还有一些游戏甚至可以由用户自定义人物表情。



图 1.5 人脸动画在游戏中的应用

(2) 移动终端应用

随着手机、平板电脑等设备被越来越多的人使用，用户已经不再满足传统的电话、视频通话、游戏等应用。一些开发者将人脸表情动画加入到传统应用中，产生了意想不到的效果。《Talking Cat》等模仿类游戏使用了表情动画技术，增加了游戏的交互性和趣味性；通讯软件将表情动画加入短信发送的功能中，使得传统功能不再单调；还有将人脸动画用于视频通话的应用，这样做可以增加通话的乐趣，同时在匿名聊天中保护隐私。



图 1.6 人脸动画在移动终端中的应用

(3) 远程交互应用

当前，大多数远程交互情景如远程会议、远程医疗、远程手术等都只能通过视频通话方式进行。这种方式给人以距离感，难以达到最好的交互效果。北卡大

学的研究人员利用人脸动画技术开发了一套沉浸式远程交互系统。该系统由一些机器人组成，机器人的头部用特殊材料设计，远程用户的脸部表情被摄像头捕获后，通过人脸动画合成技术由机器人展示出来。这套系统可以使用户产生更好的沉浸式体验，具有良好的应用前景。

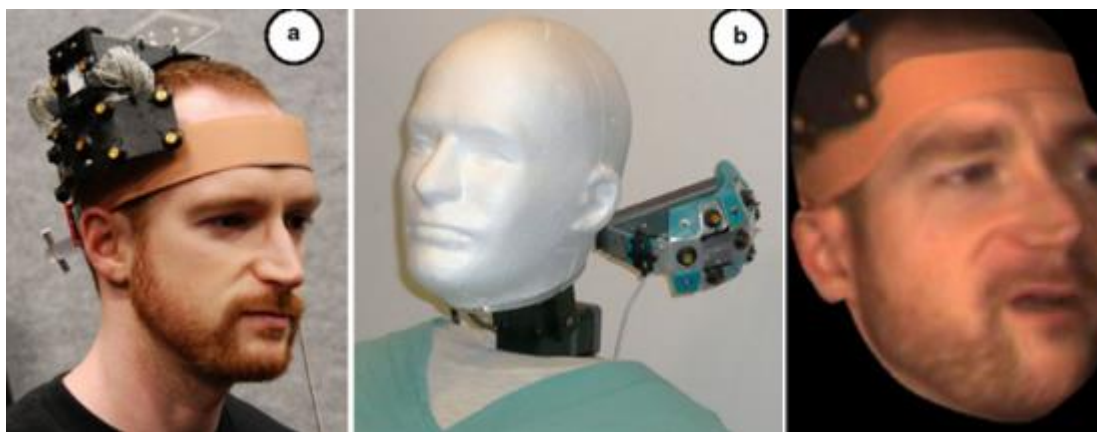


图 1.7 人脸动画在远程交互中的应用

1.2 研究概述

电影《指环王》、《阿凡达》和《本杰明奇事》中动画人物(图 1.8)的成功表明，丰富的表情是高真实感人脸动画生成的重要目标。在工业界中，一般普遍采用表情基的方法来生成 CG 人脸动画。所谓表情基也就是人脸的基本表情，如：喜、哀、怒、恐惧、惊讶、厌烦等(图 1.9)。这些表情基可以以线性[XTSYZ 2008]或者非线性[XSYZ 2011]等方式结合生成任意人脸表情。Ekman 博士等人在 70 年代系统地研究了决定人脸表情的因素，并基于人脸的生理结构对面部肌肉进行编号，提出了一套科学详尽的人脸表情编码理论 FACS[PW 78]。[CLS 2001]基于人脸肌肉运动单元设计了一套基本表情基。除此之外，[EC 02][JTDP 03][PKWS 03][DCFN 06]等文献通过分析运动捕获数据的空间分布，指出使用用户定制的表情基也可以获得良好的表情效果。为了产生更多丰富的人脸表面细节，[BV 99][ZSS 04][WLVP 09][HCTW 11]等文献利用扫描仪获取高精度人脸结构，通过对扫描数据进行配准获得高精度人脸表情基。[HJYC 13][CYSK 13]采用家用摄像头、Kinect 等设备，通过准确的人脸特征捕获，设计出高效、方便的生成人脸动画的系统。

为了获得尽可能真实的脸部表情运动，工业界广泛采用运动捕获系统来获得人脸表面的运动数据。这套系统首先需要在演员脸部按照一定规则贴一些反光点，然后演员按照预定动作对着多个同步相机做表演。运动捕获系统精度非常高，



图 1.8 《指环王》、《阿凡达》、《本杰明奇事》等影片中的人物造型

数据帧率也很大，捕获的数据具有很高的实用性。它的缺点也很明显：造价高昂，需要大量手工操作来贴点、调试设备等。最具代表性的运动捕获系统是 Vicon 光学系统，它被 Weta 等特效公司用于《指环王》、《阿凡达》、《猩球崛起》等影片的制作中。除了这种运动捕获系统，还有使用摄像头、结构光设备(见图 1.10)等进行面部表情捕获的方法。

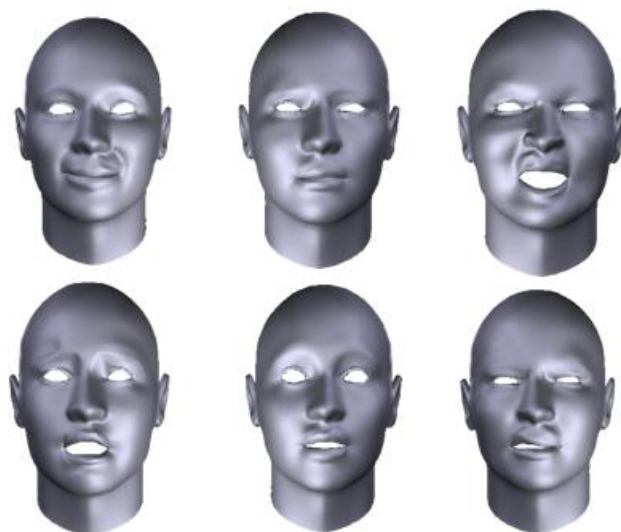


图 1.9 人脸表情基示例

本文主要针对表情基合成和运动数据获取两方面技术进行研究，力求获得一套低成本、方便、自动的人脸动画合成系统。下面对两方面技术进行现状分析。

1.2.1 表情基自动生成技术

在计算机中，各种人脸和其表情模型被数值化为顶点-面片模型。在本文中，所有模型都是三角面片模型。当然除此之外，几何模型还可以被表示成多边形、

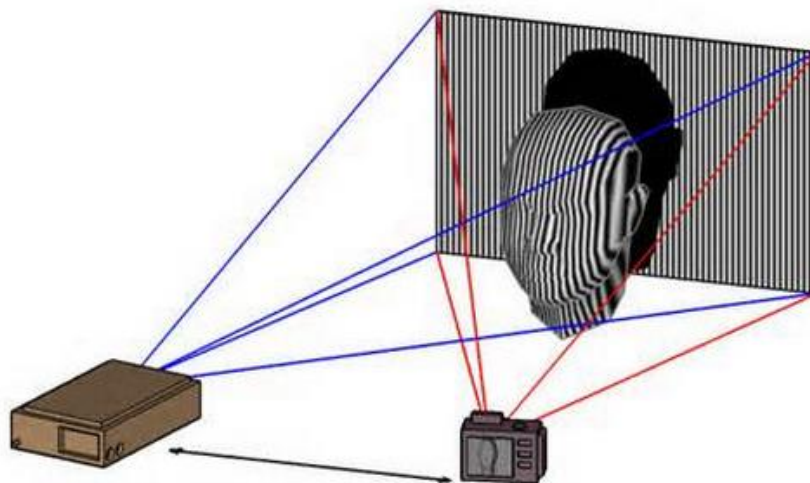


图 1.10 结构光设备示意

参数化表面(如 NURBS、B 样条)、细分表面、点云等。实际应用中,三角面片是被最广泛使用的形式,不仅因为其表示简单方便,还因为其它各种形式都可以很方便地转化为三角面片格式。因此,以顶点-三角面片模型开发的算法可以更容易地兼容多种数据格式。在 CG 人物制作中,其头像一般是由艺术家雕刻制作而成,再由 3D 扫描仪转化成数字格式(顶点-三角面片模型)。

为了保证模型表情的逼真,工业界使用的表情基往往有上千个。而这些表情基都需要美工或技术人员根据人物的头像模型去重新设计制作。即使对于熟练的技术人员,这也是一项相当繁琐和耗时的的工作。

学术界最早在 2001 年提出了表情克隆的方法,该方法可以将现有的模型表情迁移到另一个人脸模型的中性表情上,产生新模型的一套表情基。表情克隆可以有效地重用现有的表情模型库,因而引起了研究人员的广泛关注。[RJ 04]提出了一种更普适的方法,可以将具有相似拓扑结构的模型之间的变形进行相互迁移。该方法将模型的变形进行了准确的数学描述,首次定义了变形梯度这一模型来刻画产生变形的原因。[KJJX 06]通过引入体积和骨架约束,使得模型可以获得更大的变形。[KWYM 10]将变形梯度方法扩展为支持支持多个模型间进行变形迁移。[HTWP 10]采用美工制作的表情基作为约束,对变形梯度方法加以约束,可以获得满足用户约束条件的克隆表情基。

1.2.2 运动数据获取技术

运动数据按照数据来源来分有:视频数据、音频数据、深度数据等。音频数据主要用于语音驱动的人脸动画,[MB 99]建立了第一套成功的 voice puppetry

系统，采用 HMM 模型建立语音和唇形的关系。[ZSP 09][陈 03][范 10]等分别从不同角度进行了语音驱动人脸动画的研究。深度数据主要由结构光设备来获得，可以捕获人脸表面高精度的深度细节。[ZSS 04]运用结构光设备获取人脸表面精细的几何结构。[THL 09]使用高速结构光设备实时的生成人脸动画。

音频数据不能获得直观的脸部表情运动，深度数据获取需要特殊的设备。基于以上原因，视频驱动人脸动画得到了学术界的广泛关注。[EC 02]首次给出了视频驱动人脸动画的一个框架，使用光流法跟踪人脸表情运动，并通过手工的表情基选择建立视频中人脸表情与 3D 人脸表情基之间的关系。光流法检测表情运动对旋转和光照变化不敏感，但是光流法跟踪不鲁棒，容易丢失跟踪点，而且噪声较大。[JJJ 03]为了解决这个问题，建立了一套 Vicon 采集的运动数据库，通过数据库中无噪声的运动数据对视频采集的运动数据进行过滤，获得更平滑稳定的表情运动序列。[NRG 10]采用主动表观模型跟踪人脸特征点，并将人脸表情的变化抽象化为图模型，通过求解图模型中的最优路径来生成平滑的表情动画（见图 1.11）。[HJYC 13]结合视频运动数据和深度信息提出了一种基于适应性 PCA 模型的人脸动画合成框架，增强了人脸动画的表现力。[CYSK 13]设计了一种从 2D 视频运动估计 3D 人脸特征的方法，该方法可以处理比较夸张的表情以及脸部遮挡情况。

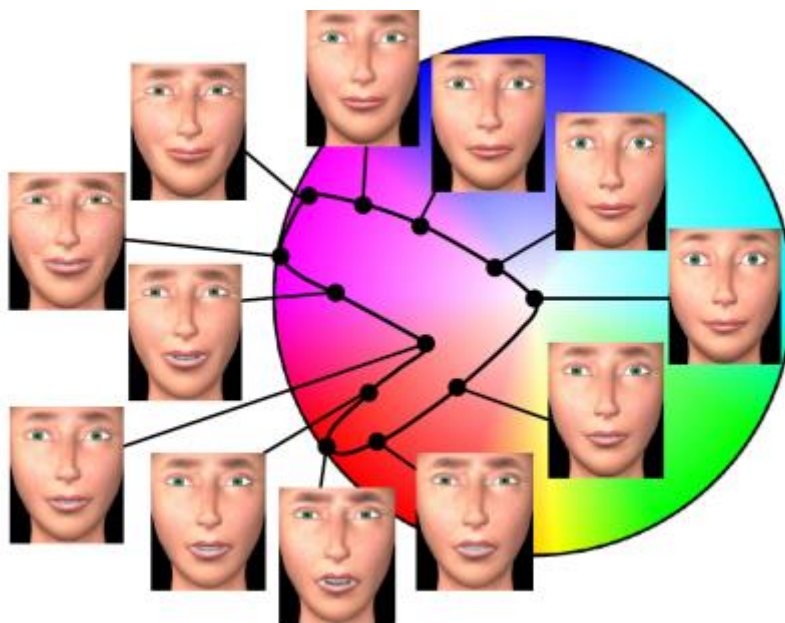


图 1.11 [NRG 10]采用的图模型

1.2.3 研究内容

本文的研究内容围绕两个子目标开展，分别是自动、高效的表情克隆算法和

实时、低成本的视频驱动人脸动画合成。具体目标是：

- (1) 对于表情克隆算法，表情克隆算法需要用户指定一些模型的对应点，通过自动的方法计算对应点可以减少用户交互操作，避免对应点指定错误的情况。表情克隆算法引入增广顶点来计算变形梯度，导致顶点数目变为原来的三倍，导致了计算效率的降低，我期望找到一种更好的变形梯度表示方式来减少计算量。
- (2) 对于视频驱动人脸动画，我希望开发一套新的技术框架来获取脸部表情数据，生成人脸动画。

1.2.4 主要贡献

本文研究在生成人脸动画过程中，如何简化用户交互过程，提高系统的计算效率以及实用性，提出了一种自动的表情克隆算法和视频驱动的人脸动画合成方法。在整个研究过程中有以下三方面研究成果：

(1) 提出了一种自动提取模型对应点的方法

表情克隆算法需要对源模型和目标模型进行对应关系配准，建立模型之间顶点的对应关系。而模型配准的初始阶段一般都需要用户给出一些基本特征点的对应关系，比如眼角、嘴角等特征点的对应。现有的方法基本都是采用手工选点的方法。为了简化用户交互过程，本文提出一种自动选取模型对应点的方法。

首先利用启发式方法选择一组模型对应特征点，比如利用 x, y, z 坐标的最大最小值选择头顶、鼻尖、耳朵尖，再根据头顶和鼻尖连线上的 z 坐标极值求其它几个特征点。用户在对源模型上手工选几个点后，就可以利用径向基函数插值获得目标模型上相应的对应点。使用该方法可以大大减少用户的手工交互操作。

(2) 提出了一种高效的模型配准优化方法

本文采用基于变形梯度的方法实现了表情克隆工作，由于空间三角形只有两条线性无关的边，不能获得变形梯度的准确值，该方法引入增广顶点来计算变形梯度。本文修改了变形梯度的矩阵表示，通过一个六自由度的矩阵近似原来八自由度的矩阵，从而使得其计算过程不需要增广顶点。这一修正将建立模型配准过程中的优化变量减少到原来的三分之一，很大程度上提高了优化计算的效率。虽然这种修正只是对变形梯度精确值的一种近似表示，实验表明，这种方法依然可以保持原算法的优化精度。

(3) 搭建了一套实时、低成本的视频驱动人脸动画合成系统

该系统使用普通的家用摄像头作为交互接口，用户只需要对摄像头做表情，便可以在 PC 上生成一个与自己有相似脸部表情和头部运动的虚拟人脸化身。该系统实现了以下功能模块：利用约束局部模型实现了实时人脸特征检测和姿态估

计,可以较为准确的估计人脸表情变化和头部运动;通过估计相机参数,将 2D 的表情运动数据变为 3D 的运动数据,同时建立用户脸部结构虚拟人脸化身模型结构的对应关系,将用户的脸部表情迁移到几何结构不同的虚拟人脸化身上面;人脸动画合成,使用带约束的表情基优化方法获得表情基合成参数,进而生成具有较高真实感的人脸动画。

1.3 论文组织

本文共分为五章,各章具体内容安排如下:

第一章:介绍本文的应用领域,研究背景,阐明研究目标、内容,并总结本文的主要贡献。

第二章:对表情克隆和视频驱动人脸动画相关研究进行综述。第一部分介绍了表情克隆需要解决的两个问题:模型配准和运动迁移。接着详细介绍了表情克隆早期的经典工作[JU 01]。并总结了其优点和局限性。文中还简要介绍了其它几种表情克隆方法的思想。第二部分介绍了视频驱动人脸动画方面的一些主要工作,通过功能模块实现的不同进行了分类,并阐述了现有方法的优点和不足。

第三章:论述本文的第一项研究内容——自动表情克隆算法。首先引入变形梯度的定义,再介绍原始的变形迁移算法的优化计算过程。接着介绍变形迁移算法计算模型配准的具体过程。最后阐述本文的创新点:自动对应点提取和修正的变形梯度优化算法。

第四章:论述本文的第二项研究内容——视频驱动人脸动画。将本系统分为三个功能模块:人脸特征跟踪和姿态估计,运动参数化,人脸动画生成。分别对三个模块进行详细的阐述。

第五章:总结了本文的研究内容,分析可能深入研究的问题和思路,给出下一步的研究方向。

2 基于表情基的人脸动画合成方法研究综述

2.1 表情克隆方法

表情克隆就是重用现有模型的表情基，并把这些表情迁移到新的模型上(图 2.1)。2001 年, Noh 等人提出这一表情克隆的概念, 并设计了算法成功对不同人脸模型之间实现了表情迁移。表情克隆的实现需要解决两个重要问题: 模型配准和变形迁移(图 2.2)。

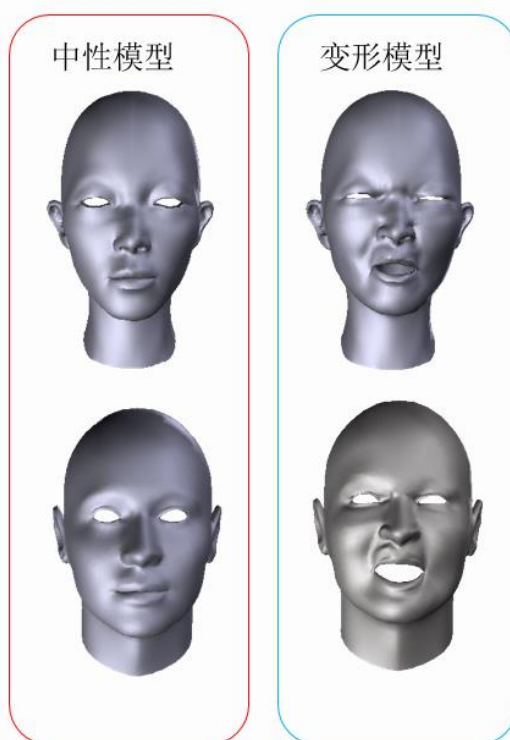


图 2.1 表情克隆图示

两个不同的人脸模型包含不同的顶点、面片数目, 几何拓扑结构不一样。模型配准要解决的就是源模型和目标模型上哪个(哪些)顶点、面片应该互相对应。这种对应关系可以是一对一、多对一、多对多。变形迁移解决的问题是, 在确定模型间配准关系后如何把变形(表情)从一个模型迁移到另一个模型上, 这中间涉及如何用数学的方法描述变形以及如何对模型进行变形。

2.1.1 模型配准问题

模型配准的目标是建立两个模型之间顶点一对一或者多对一的关系, 最终要将两个模型的拓扑结构变为一致。尽管人脸具有总体相似的几何结构, 但是使用三角面片表示的人脸模型其拓扑结构却会有较大差异:

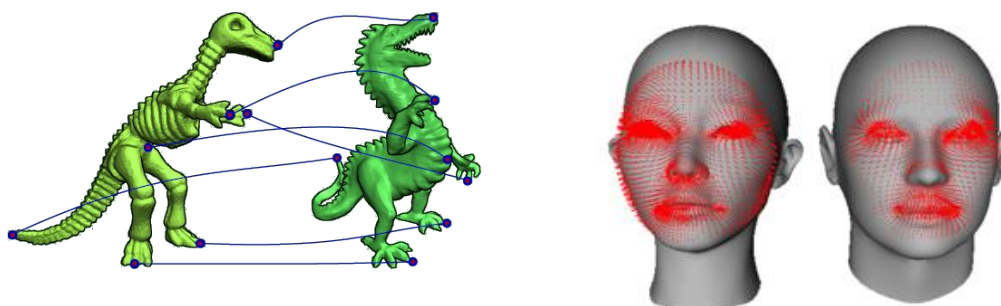


图 2.2 模型配准和运动迁移

- (1) 不同精细程度的模型对细节的刻画有区别。顶点数越多，模型越精细，对细节的表现力越强。这种情况下，顶点数少的模型一个面片就要对应好多个高精度模型的面片。
- (2) 即使是相同精细程度的模型，由于人脸表面结构的不同，所形成的几何结构也不同。如有些人嘴大，有些人嘴小，它们嘴部的面片同等精细程度下就要形状差异就很大。
- (3) 在同一个人脸模型上，面片的精细程度也有区别。脸部，下巴，后脑等处由于缺少细节，面片数较少。而眼角，嘴唇等处细节较多，精细程度较高。

好的模型匹配需要遵循两个基本原则：

相似性： 对应顶点或者面片需要看起来相似，比如两个模型的鼻尖、眼角、嘴角等特征位置应该作为匹配点。

同一性： 源模型某一顶点相邻区域的顶点对应目标模型的顶点也应该具有相邻关系，不能改变这种局部区域连接关系。

模型配准不等于模型注册，前者是找到两个模型之间顶点和面片的对应关系，后者是将两个模型进行空间上的对齐，不过两个问题是相辅相成的。模型配准解决好了就能找到精确的变形关系进行模型注册；模型注册如果解决好了就可以直接用最近点对来找配准关系了。从注册的角度来看，模型配准研究的两个模型之间可能存在以下变形关系：

- (1) 两个模型可以由刚性变换对齐；
- (2) 两个模型可以分区域进行刚性变换对齐；
- (3) 两个模型间是一种非线性的对齐关系。

其中，第一种情况主要应用于对同一个物体的不同几何表示进行对齐，比如使用不同精度的扫描仪得到的数据，或者是两组具有重叠区域的几何数据。第二种情况最常见的就是具有肢体的模型，如人、各种动物、家具等。其中人和动物的肢体活动是一种分段刚性的运动，只有关节处有非线性变形关系。这种情况下

很常见的一种应用是对齐一段连续时间内采集到的几何数据(点云)。人脸之间的注册具有典型的非线性对齐关系,因此最好的配准方法就是先把两个模型对齐到很接近,在用最近点对或其它方法来获得模型配准。表 1 给出了模型配准和注册所研究问题的形式表述。

下面介绍几种典型的计算模型配准的方法,ICP 方法、模版匹配法、组合优化法、变形配准法。

ICP 方法是 Iterative closed points 的简称,主要适用于可进行刚性变换对齐的两个点云模型。这种方法首先计算模型间顶点间的最近对应,然后通过最小化刚性变换后最近点距离的平方误差来获得刚性变换关系。算法思路如图 2.3 所示。

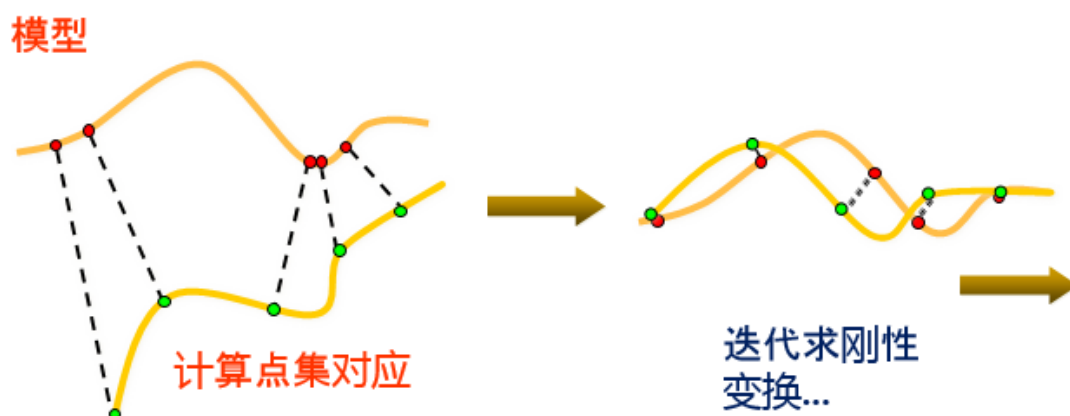


图 2.3 ICP 方法图示

ICP 方法容易陷入局部极小,而且只适用于噪点很少的情况。可以采用两种方法对其进行改进:

- (1) 手工或按照某种规则自动指定部分对应点。这些规则包括利用点集几何位置或者模型表面的曲率信息等。
- (2) 点集投票方法,选择符合最多点集匹配的变形关系。

模版匹配法的基本思想是用一个模版模型去匹配所有待配准的模型。举例来说就是如果要配准两个人脸模型的话,我们先找一个“一般”的低精度人脸模型,这个模型有眼睛、鼻子、嘴等结构,但是没有很多细节。然后再不断的拟合这个模型到高精度人脸模型上(如图 2.4 所示)。文献[EWP 28]采用这一方法建立具有相似拓扑结构的匹配关系,文献[BBZ 03]进一步通过补全扫描人体数据,建立起人体的一个参数化模型,可以通过少量参数生成高矮胖瘦等不同体形、具有相同拓扑结构的几何模型。文献[VT 98]利用这种方法将很多人脸模型进行配准,并进一步进行统计建模,获得一组人脸的参数化模型。这种方法的一个缺点是改变了原来模型的拓扑结构,而这种改变在某些情况下是有害的[EWP 28]。

输入	几何表示	点, 表面, 轮廓	
	数据类型	2D, 3D, 2D + time, 3D + time	
输出	关联表示	对应+变形	平移、旋转、缩放、仿射、非线性
		对应	映射、单射、多对多、概率对应
方法	计算变形	刚性(如 ICP)、非刚性变形 (如 thin plate spline)	
	计算对应关系	几何、测地距离 + 领域知识(如保距性等)	

表 1 模型配准和注册问题概述

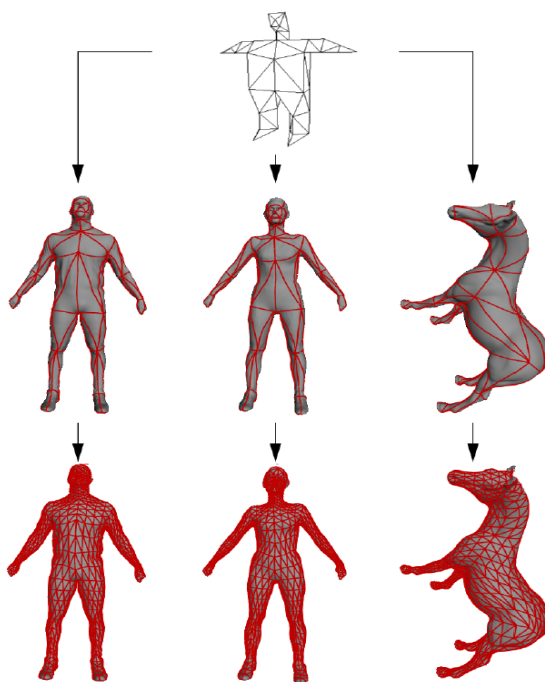


图 2.4 模版匹配示例

组合优化法主要被应用于计算分段非刚性变形(如图 2.5)的模型配准问题中。其优化目标是

$$\min \omega \sum SimDist(p_i, q_i) + (1 - \omega) \sum distort(p_i, p_j, q_i, q_j) \quad (1)$$

其中 p_i 和 q_i 是两个模型上的对应点。 p_j 和 q_j 是 p_i 和 q_i 的相邻顶点。优化目标的第一项是顶点的相似度， p_j 可以用欧式距离、形状上下文等几何特征来计算。第二项是同一性的体现，表示变形后同一模型上相邻顶点的距离不变。这一问题可以被看作两个集合间的点集匹配问题，能应用组合优化法求解[YY 11]。

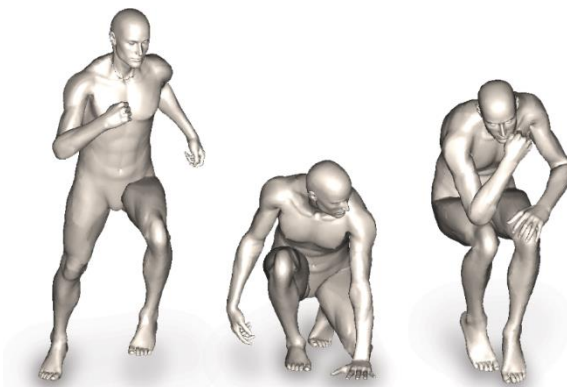


图 2.5 适用于组合优化法的分段非刚性变形模型

变形配准法。对于只存在非刚性对齐关系的模型(如图 2.6)，最好的配准思路就是先将两个模型相互变形到与对方几何形状一致，然后顶点对应关系就能直接根据最近欧式距离来计算了。

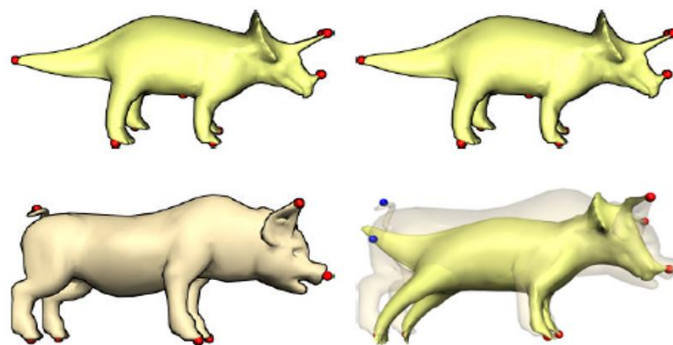


图 2.6 适用于组合优化法的非刚性变形模型

人脸表面属于高度非线性曲面，因此配准两个人脸模型最好使用变形配准法。下面介绍 2001 年 Noh 等人 在其表情克隆文章中提出的计算模型配准的方法(如图 2.7 所示)。具体步骤是：

算法 1：表情克隆模型配准

1. 选择模型上的一组对应点；
2. 建立 RBF 模型，进行初步对齐；
3. 通过变形模型和源模型间进行圆柱投影获得稠密顶点对应关系。

作者参考光学运动捕获时人脸表面的贴点方式，设计了 46 组对应点，并对源模型和目标模型进行手工标注。然后对 46 组对应点建立 RBF 网络，将目标模型

变形到源模型上去，建立两者的初步对齐。变形后的目标模型称之为变形模型。

RBF 网络是一种插值函数，其一般形式为：

$$\mathbf{F}(\mathbf{x}) = \sum_{i=1}^M \omega_i h_j(\|\mathbf{x} - \mathbf{x}_i\|), h_j(\mathbf{x}_i) = \sqrt{|\mathbf{x}_j - \mathbf{x}_i|^2} \quad (2)$$

(2)式表示一个有 M 个插值点的 RBF 函数，基函数 $h_j(x)$ 往往是个非线性函数。以上 RBF 函数对于一般的插值应用能够适合，这里由于插值点远远少于模型顶点数目，需要加入线性项修正，第五章中我还会讨论这个问题。

在得到初步对齐结果后，采用圆柱投影的方法，计算源模型顶点到变形模型上面片的交点，这些交点作为一组稠密的顶点配准。具体过程是：

1. 为变形模型计算一个包围圆柱；
2. 从变形模型每个顶点发出一条射线使其垂直于圆柱中线；
3. 这条射线(或者延长线)与源模型某一面片的交点就被看作变形模型与源模型的对应顶点。如果不存在交点则规定其对应点为源模型上欧式距离最近的顶点。

这种方法实质上属于模版匹配，其优点是方法直观、计算简单，缺点是改变了目标模型的拓扑结构(网格的布线结构)。

2.1.2 运动迁移问题

Noh 等人在表情克隆一文中采用了运动向量迁移的方法。思路就是把源模型上顶点的位移应用到目标模型上去。但是由于两个模型的拓扑结构不同，这些位移不能简单的应用。需要利用局部坐标对它们进行变换。步骤如下：

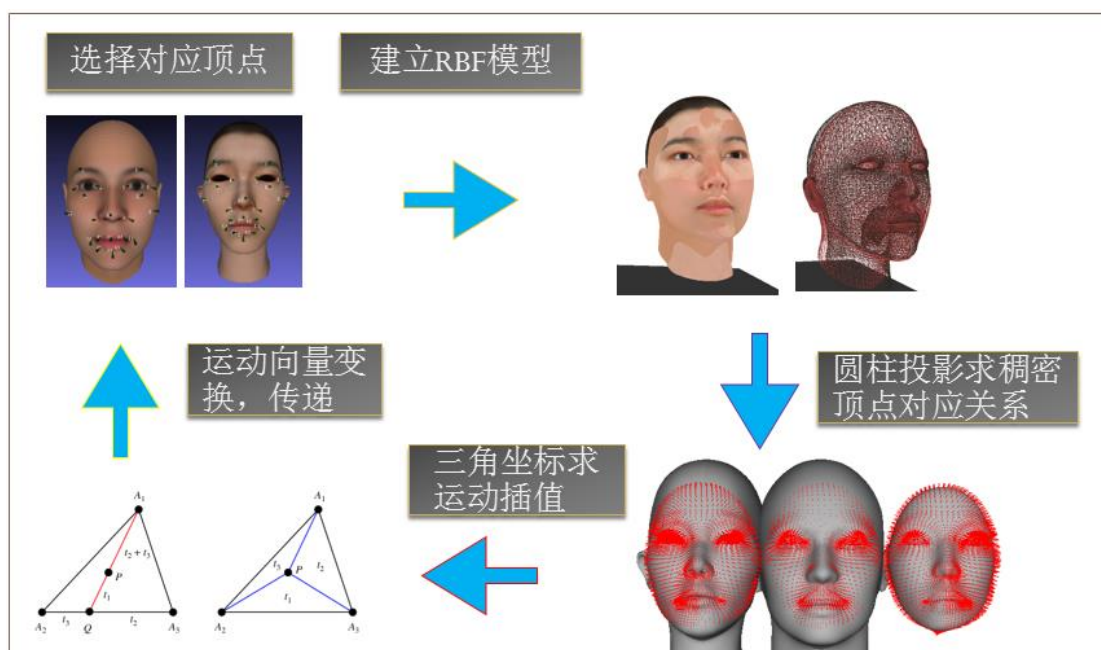


图 2.7 表情克隆 [JU 01] 算法流程

算法 2: 运动向量修正

1. 将变形模型在目标模型的对应顶点用源模型的顶点进行三角坐标表示;
2. 利用三角坐标获得变形模型每个顶点的运动向量;
3. 为变形模型和目标模型每个顶点构建局部坐标系, 通过世界坐标建立局部坐标系之间的对应关系。计算目标模型每个顶点的运动向量的旋转和缩放量;
4. 应用修正后的运动向量到目标模型每个顶点上获得新的模型。

下面详细介绍以上步骤:

三角坐标。如图 2.8 中所示, 三角形中的一个内点 P 可以表示为三个顶点 A_1, A_2, A_3 的线性组合, 组合系数可以由以下公式计算。 x_i, y_i 分别是 A_i 的坐标值。

$$\begin{aligned} t_1 &= \frac{(y_2 - y_3)(x - x_3) + (x_3 - x_2)(y - y_3)}{(y_2 - y_3)(x_1 - x_3) + (x_3 - x_2)(y_1 - y_3)} \\ t_2 &= \frac{(y_3 - y_1)(x - x_3) + (x_1 - x_3)(y - y_3)}{(y_2 - y_3)(x_1 - x_3) + (x_3 - x_2)(y_1 - y_3)} \\ t_3 &= 1 - t_1 - t_2 \end{aligned} \quad (3)$$

则变形模型的运动向量就可以用源模型的运动向量进行加权表示:

$$\mathbf{mv}_P = t_1 \mathbf{mv}_1 + t_2 \mathbf{mv}_2 + t_3 \mathbf{mv}_3, t_1 + t_2 + t_3 = 1 \quad (4)$$

局部坐标系建立。由于变形模型保持了目标模型的顶点对应关系和布线结构, 我们可以对它们的每个顶点建立一一对应的局部坐标系, 步骤如下:

1. 取与顶点 v_i 邻接的所有面片的法线的平均值作为 z 轴;
2. 取与顶点 v_i 相连的任意一条边 e_i , 将其投影到 z 轴的垂直平面, 取其投影作为 x 轴;
3. 通过求 z 轴和 x 轴的差积获得 y 轴。

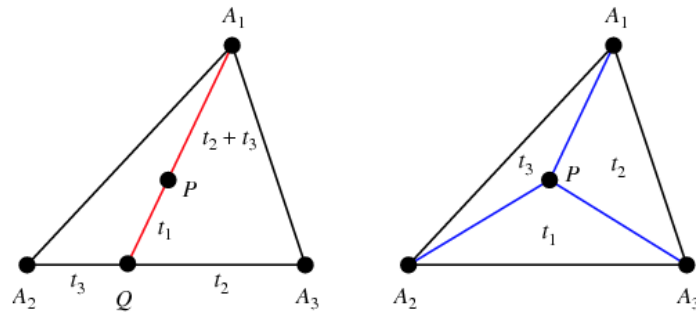


图 2.8 三角坐标示意图

对于每一对变形模型和目标模型上的对应顶点, 我们建立他们局部坐标系之间的变换关系。其中 x_w, y_w, z_w 是世界坐标系的 x, y, z 轴。 ${}^O_w R$ 是目标模型和世界坐标系的变换矩阵, ${}^D_w R$ 是变形模型和世界坐标系的变换矩阵。

$$\begin{aligned}
{}^O_W R &= \begin{bmatrix} x_w \cdot x_o & y_w \cdot x_o & z_w \cdot x_o \\ x_w \cdot y_o & y_w \cdot y_o & z_w \cdot y_o \\ x_w \cdot z_o & y_w \cdot z_o & z_w \cdot z_o \end{bmatrix} \\
{}^W_D R &= \begin{bmatrix} x_w \cdot x_d & y_w \cdot x_d & z_w \cdot x_d \\ x_w \cdot y_d & y_w \cdot y_d & z_w \cdot y_d \\ x_w \cdot z_d & y_w \cdot z_d & z_w \cdot z_d \end{bmatrix} \\
{}^O_D R &= {}^W_D R {}^O_W R
\end{aligned} \tag{5}$$

经过上述变换关系，我们可以得到 ${}^O_D R$ ，变形模型和目标模型顶点之间的坐标变换关系。

这一变换关系刻画的只是运动向量方向上的变化，我们还需要计算其大小的变化，Noh 等人采用的方法是计算每个顶点邻接面片的包围盒，通过变形模型和目标模型顶点之间包围盒的体积之比获得运动向量大小变化。

$$S_{x,y,z} = \frac{size_{x,y,z}(DeformedModelLocalBoundingBox)}{size_{x,y,z}(TargetModelLocalBoundingBox)} \tag{6}$$

最后，经过方向和尺度变换后的运动向量被应用到目标模型上产生新的表情模型。这里 M_i 是目标模型第 i 个顶点的运动向量， M'_i 是变形模型第 i 个顶点的运动向量。变形模型的运动向量由源模型运动向量的三角坐标加权得到。

$$\begin{aligned}
M_i &= S_{x,y,z} \cdot {}^O_D R \cdot M'_i = S_{x,y,z} \cdot {}^W_D R \cdot {}^O_W R \cdot M'_i \\
M' &= t_1 M'_1 + t_2 M'_2 + t_3 M'_3
\end{aligned} \tag{7}$$

2.1.3 其它表情克隆算法

文献[HTWP 10]采用了[RJ 04]提出的变形迁移方法进行表情克隆，对原文做了如下改进：使表情克隆产生的表情基比较像用户给出的表情基。作者通过修改[RJ 04]中优化过程，加入修正项来实现这个目的。这种方法为表情克隆算法提供了更高的灵活性，具有较高的实用价值。[KWYM 10]可以对具有多个分离组件的模型进行变形迁移，该方法自动计算各组件之间的空间位置关系，并在变形迁移后继续保持原位置关系的一致性。这种方法使表情克隆算法可以适用于更复杂的模型，

2.2 视频驱动人脸动画研究现状

2.2.1 背景和领域知识

虚拟人物表情动画被广泛地应用于电影、游戏、教育等领域。在科幻片《指环王》、《阿凡达》和《猩球崛起》中，利用动作捕获技术演员可以脱离沉重的面具、繁琐的化妆，更加轻松灵活地表演。在”Talking cat”等整蛊系列游戏中，表情动画被用来展现游戏与用户更生动的互动。在语言教育领域，微软 Bing 新推出的 Engkoo 系统利用语音驱动表情动画技术，将一些句子的口型更加真实地展现给用户，帮助用户学习英语。

上述表情动画的产生都依赖运动数据的获取，如 Vicon 系统捕获的运动数据、美工建立的人脸表情基、演员录制的语音数据等。因而这种表情动画生成方式被称为数据驱动人脸动画。其中，视频驱动人脸动画是数据驱动人脸动画中的一个重要分支。顾名思义，视频驱动的人脸动画就是借助摄像头，虚拟人物可以做出与演员(用户)相同的表情。表 2 是与其它的数据驱动人脸动画技术在数据获取和表情动画制作技术上的比较，可以看出它在实时性、精确性、演员(用户)舒适度和制作成本等因素上有比较高的综合优势。

	实时性	精确性	演员(用户)舒适度	制作成本
运动捕获设备 (如 Vicon)驱动	很高	很高	有较大束缚	很高
语音驱动	低	较低	无束缚	较低
视频驱动	高	高	无束缚	较低
手工调整表情基	低	低	无束缚	较低

表 2 表情动画制作技术比较

视频驱动人脸动画包含三个主要技术问题:

1. 如何实时准确地从视频中提取人脸运动信息;
2. 如何将视频中提取的运动信息转化为人脸动画的控制信号;
3. 如何利用人脸动画的控制信号生成真实感人脸表情动画。

衡量这些技术的主要指标就是系统是否具有实时性，用户是否对该系统生成的人脸表情动画有认同感。

2.2.2 研究现状分析

视频驱动人脸动画涉及人脸运动信息提取、运动参数计算和传递、人脸动画生成等技术。下面就这三方面对目前的研究现状进行分析。

人脸运动信息提取

在[EC 02][JJJ 03] [NRG 10]等文章中, 作者采用光流等基于像素区域的仿射变形方法跟踪视频中人工选择的人脸表面特征点(如图 2.9a)。为了保证特征点跟踪的稳定性, [EC 02]在人脸上用颜色笔画了一些标记, [JJJ 03]设置了人工干预机制进行丢失特征点的重新跟踪, 并假设了干净的背景和高像素的视频质量。这种方法的好处是对人脸的旋转、平移运动不敏感, 缺点是跟踪点缺乏标记信息。

[TYJ 11]采用主动表观模型 AAM 来捕获人脸表面特征点(如图 2.9b), AAM 是一种对人脸特征点位置和像素值建模的统计模型。构建这个模型需要手工标记大量的人脸图像。其好处是跟踪的特征点具有明确的标记, 这样有助于获取精确的运动信息。在[TYJ 11]中, 由于 AAM 对光照较敏感, 作者加了边缘信息约束, 有效地降低了配戴眼镜等饰物对特征点检测的影响。

除了捕获稀疏的特征点运动, [THL 09]利用多目摄像头和投影仪设计了实时捕获人脸表面运动的结构光设备(如图 2.9c), [JJJ 03]发明了用 14 个标定好的相机捕获高精度(毛孔、皱纹) 脸特征的设备。这些设备可以用来获得很精细的模型和动画序列。

运动参数计算和传递

文献[EC 02]中采取了以下的方法:

1. 记录下跟踪人脸眼睛、眉毛、嘴巴的轮廓点位置, F 为跟踪的帧数;
2. 对 F 帧的运动向量采用聚类、求凸包或者求 PCA 主成分的方法求得一组表情基;
3. 对照第 2 步计算的 2D 表情基, 建立精细人脸模型表情基与其的一一对应关系;
4. 采用最小二乘对每一帧运动数据计算 2D 表情基的权值。

这种方法采用了表情基动画的思路, 通过表情基优化计算将 2D 的特征点序列与 3D 人脸模型建立对应关系。模型直观, 计算简单。但是, 这个方法的缺陷是: 对每个使用者都需要较长时间的训练过程, 手工交互多。由于其跟踪的轮廓点没有明确的语义信息, 并且受头部运动的影响, 优化计算得到的表情基权值不稳定。文献[JJJ 03]采用了同样的特征跟踪方法, 并定义了一组表情参数来表示运

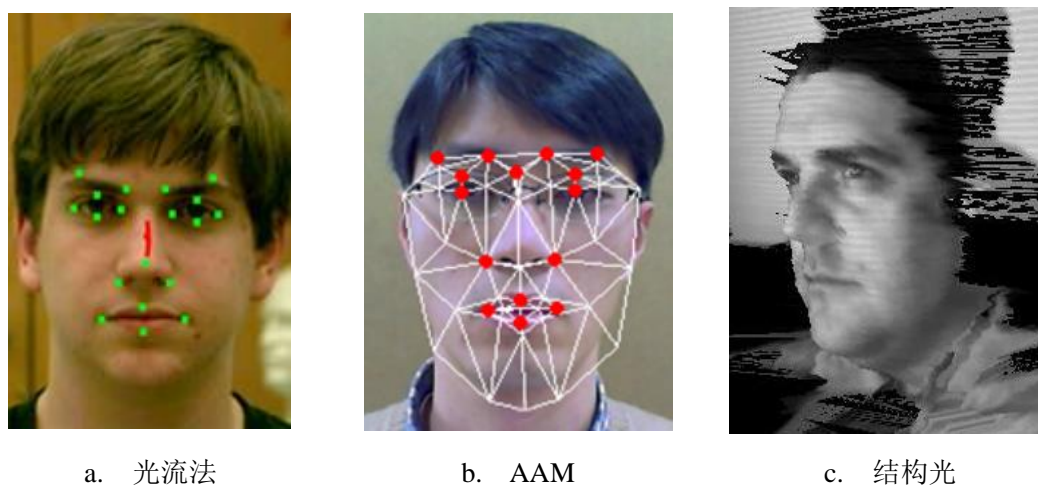


图 2.9 几种人脸运动捕获方法

动信息(例如上下嘴唇高度差,眉毛与眼睛距离等)。注意到运动参数不稳定的问题,作者借用了 Vicon 采集的运动数据库来对 2D 运动参数进行滤波。具体方法如下(如图 2.9 所示) :

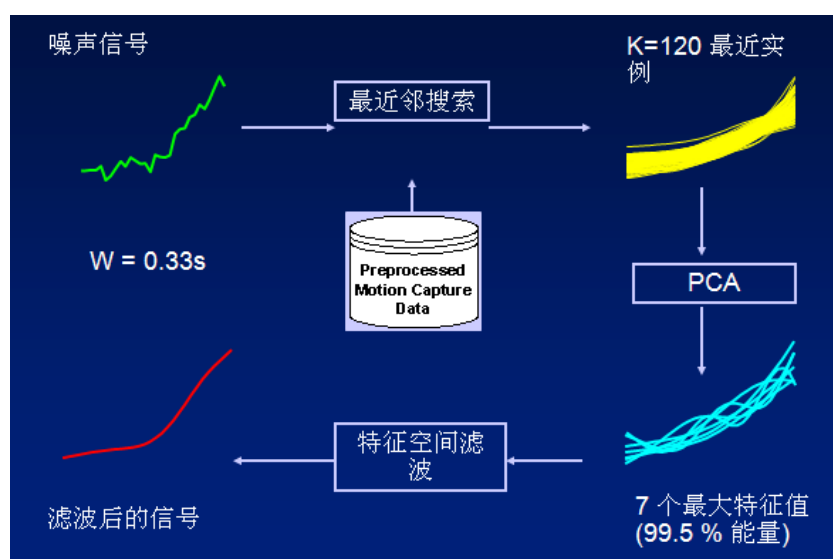


图 2.10 文献[JJJ 03]的算法流程

1. 对有噪声(不稳定)的运动参数序列在运动数据库中计算最相似的 K 个序列;
 2. 对 K 个序列建立 PCA 模型,保留 99%以上最大的特征值对应的 N 个特征向量;
 3. 将原运动参数序列投影到这 N 个特征向量上,重建原运动参数序列。
- 经过滤波后的运动参数序列可以通过计算其与 K 近邻距离来获得插值权重。文献[TYJ 11]定义了 11 组表情运动参数,这些参数的组合可以完整地表示

FACS 人脸表情编码系统的 41 组表情基。然后通过模型上取相同的表情运动参数。通过对视频捕获表情运动参数序列进行最小二乘优化,可以得到模型上的表情运动参数序列变化情况。

人脸动画生成

人脸动画生成主要包含两类技术:表情基和模型变形。前者通过表情基权值控制,具有直观的语义含义。后者通过直接对模型表面进行变形,达到表情变化的效果。

文献[EC 02]和[JJJ 03]采用了表情基技术,这方面的参考文献有很多,如 [LJP 10][HPS 06][XSYZ 2011]等。迄今为止,仍然是人脸动画领域最流行的制作方法。但是其主要缺点是需要人工制作很多表情基模型,而且这些模型的重用性往往不好。为了解决这些问题,[JU 01][RJ 04][HTWP 10]等提出了表情基克隆的一系列技术,提高了表情基的重用性。

模型变形技术主要是通过人脸模型表面若干稀疏顶点的运动,带动整个模型产生形变。Wei-Wen Feng 等人开发了利用少数特征点控制大量局部区域非线性形变的方法并成功应用到人脸模型上[WBYP 08]。该算法包括基于核 CCA 和泊松传播的方法基于实例进行快速的形变回归。[NRG 10]中提出一种线性融合形状变形的办法,通过手动提取并跟踪人脸表面特征点,将模型其它顶点的位移表示为这些特征点的位移加权和。[SZPG 05]提出 meshIK 的办法,通过驱动表情基上的少数特征点生成非线性表情形变,这种方法可以合成很自然的表面变形,从而生成质量更高的动画。文献[TYJ 11]利用测地距离 RBF 插值的方法进行局部变形,获得了较好的效果。

2.2.3 相关工作分析

[EC 02]首次给出了视频驱动人脸动画的完整框架,但是在此项工作需要每个用户进行表情基选择工作,跟踪的特征点位置噪声较大。

[JJJ 03]利用 Vicon 数据库对视频运动信号进行滤波,并通过优化求解表情基的权值,是一套成功的视频驱动人脸动画系统。但是其控制信号是表情参数在时间上的一个序列,有 0.3s 的延迟,难以做到实时。控制信号(15 维)维度较低,难以生成丰富的表情。

[NRG 10]采用主动形状模型跟踪人脸特征点,表情运动信号更准确、稳定。但是该系统需要对每个用户进行监督式表情数据流形学习,还要建立表情状态转移图。初始化工作比较费时。

2.3 小结

人脸动画合成因其在工业界的广泛应用一直受到学术界的关注。就本文所涉及的表情克隆和视频驱动人脸动画领域来看，尽管前人已经有很多研究成果，但还是存在以下问题需要解决：

(1) 自动表情克隆方法

现有的表情克隆算法大都需要用户进行较多的人工交互，通过自动的方法计算对应点可以减少大部分用户交互操作，避免对应点指定错误的情况。原始的表情克隆算法引入增广顶点来计算变形梯度，顶点数目变为原来的三倍，导致了计算效率的降低。因此，研究对原算法优化方法的改进、探索高计算效率的方法成为自动表情克隆方法需要解决的问题。

(2) 改进的视频驱动人脸动画

现有的基于视频的人脸特征跟踪算法大都只能获得 2D 运动信息，而我们需要驱动的人脸化身是 3D 的。因此，如何将视频跟踪获得的 2D 运动信号转化为驱动 3D 人脸表情运动的 3D 控制信号是我们需要解决的问题。另外，采用表情基方法生成人脸动画时，有可能出现表情基权值异常的情况，我们需要对这种情况进行特殊处理。

3 自动表情克隆算法

3.1 变形梯度

本文所说的变形是指对于拓扑结构相同的两个模型(顶点数和三角面片数一样,三角面片连接关系一致),它们的几何形状差异。在人脸动画领域,变形就是某一人脸模型对应的表情基。根据凸包理论,任意合理的人脸表情(凸包内点)总是可以由中性表情加上表情基(凸包边界点)的线性组合构成。

变形迁移所做的就是把一个模型的变形迁移到另一个不同拓扑结构的模型上去。传统的做法是美工按照源模型的变形方式,通过移动目标模型上的顶点来达到相同效果,比如采用 FFD 方法。这种方法把模型用很多空间网格点来近似,每个网格点都会对应模型上的一些顶点。移动一个顶点,空间网格会带动周围的顶点也会跟着进行运动。这种方法的优点是实现简单,计算方便。缺点是控制参数较多,而且不能严格按照用户的意愿进行变形。究其原因,我认为没有对变形进行细致的量化表述。因此,为了准确地进行变形迁移,我们首先要对变形进行合理的数学表述。为了增强其泛化能力,这种表述应该是独立于产生变形的算法的。

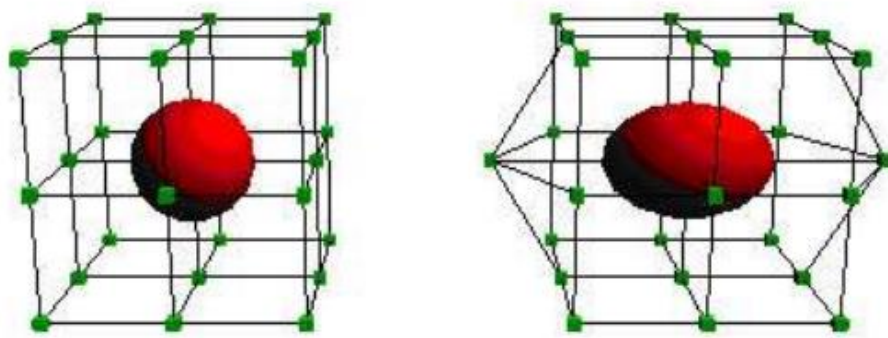


图 3.1 自由形式变形(Free-Form deformation)

一种可能的方法是采用模型和其变形模型顶点之间的位移表示变形。文献 [JU 01]就采用了这一方法来表示中性人脸模型和有表情的模型之间的变形,这些顶点间的位移被称作运动向量。虽然经过局部坐标变换,运动向量以一种更合理的方式被迁移到目标模型,但是这种方法经实验证明只适用于模型变形较小的情况。对于产生较大变形的情况,这种方法容易产生不自然的变形[RJ 04]。

文献[RJ 04]指出,更有意义的变形表述是,不仅知道变形的大小,还知道变形是如何产生的。要理解这句话需要看下面这幅图。图 3.2(a)中表示了一条直线

弯曲的变形, 其中灰色的线条是各顶点的位移。在图 3.2(b) 中我们直接把这些位移应用到一条更短的直线上, 这样产生的弯曲变形是不自然的, 因为它不符合原来模型的变形程度。如果我们换一个角度看, 变形是由这些直线上每条线段经过旋转某个角度 α , 再把长度变化 β 倍产生。那么我们应用这些旋转和尺度信息到新的模型上就会产生一个更自然的变形, 如图 3.2(c)。这种变形的变化就叫做变形梯度, 因为它可以看作是变形之间的增量改变。

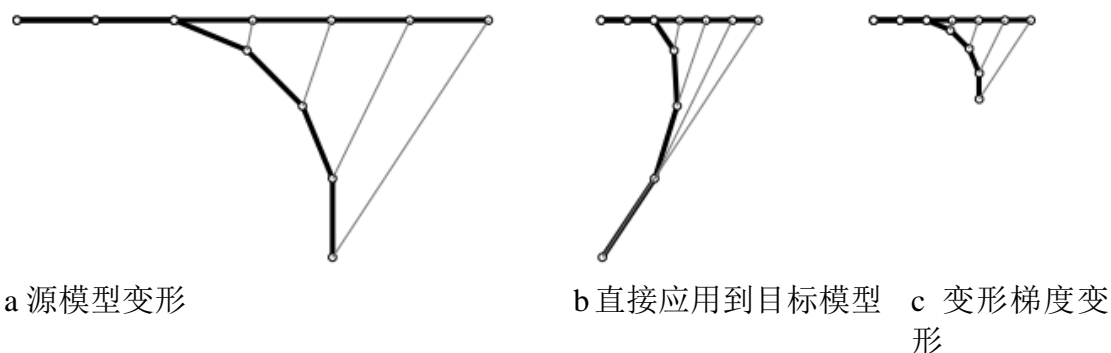


图 3.2 变形梯度示意图(图片取自文献[RJ 04])

为了从数学上表述变形梯度, 我们假设源模型上一个顶点 $p = [p_x, p_y, p_z]$ 经过变形后成为 $q = [q_x, q_y, q_z]$ 。定义变形操作子为 U , 则有下面的关系成立:

$$q = U(p) = \begin{bmatrix} U_1(p_1, p_2, p_3) \\ U_2(p_1, p_2, p_3) \\ U_3(p_1, p_2, p_3) \end{bmatrix} \quad (8)$$

当变形很小时, 其微分形式可以写成:

$$\frac{\partial U}{\partial p} = \begin{bmatrix} \frac{\partial U_1}{\partial p_1} & \frac{\partial U_1}{\partial p_2} & \frac{\partial U_1}{\partial p_3} \\ \frac{\partial U_2}{\partial p_1} & \frac{\partial U_2}{\partial p_2} & \frac{\partial U_2}{\partial p_3} \\ \frac{\partial U_3}{\partial p_1} & \frac{\partial U_3}{\partial p_2} & \frac{\partial U_3}{\partial p_3} \end{bmatrix} \quad (9)$$

这就是变形梯度的数学形式。它准确地刻画了变形的朝向、尺度等变化, 而且不受变形本身的影响, 因此是一种比较好的变形计算方式。

本文采用三角面片模型, 模型经过变形后, 每个三角形被变形为另外一个三角形, 因此变形梯度可以由三角形的三个顶点来计算得到。

$$\begin{aligned} V &= [v_2 - v_1, v_3 - v_1] \\ V' &= [v'_2 - v'_1, v'_3 - v'_1] \\ Q &= V' V^{-1} \end{aligned} \quad (10)$$

变形梯度 Q 由于把三个三维顶点映射另外三个三维顶点，所以它应该是个 3×3 的矩阵。然而，由于空间平面上的三个点决定的三条边只有两个自由度，因此上式不能完全确定变形梯度 Q 的九个变量。为了解决这一问题，[RJ 04]中为每一个三角面片引入另外一个顶点，组成空间四面体。我们称这个顶点叫做增广顶点。增广顶点的计算方法如下：

$$v_4 = v_1 + (v_2 - v_1 \times (v_3 - v_1)) / \sqrt{|(v_2 - v_1 \times (v_3 - v_1))|} \quad (11)$$

这样的话，使用四个顶点就能完全计算出变形梯度 Q 的九个变量了。正如前文所说，这九个变量包含空间四面体的旋转、缩放、扭曲等变换。

3.2 变形梯度迁移

目前我们还没有解决两个模型之间配准的问题，本文之所以这样组织是由于下一章会涉及这一节的内容，先介绍变形梯度迁移可以更好地理解下一章的内容。我们现在假设已经通过模型配准获得了源模型和目标模型之间顶点以及三角面片的对应关系。由于变形梯度的计算是对每一个面片而言的，我们只需要两个模型之间三角面片的对应关系：

$$M = (s_1, t_1), (s_2, t_2), \dots, (s_{|M|}, t_{|M|}) \quad (12)$$

上式中， s_i 和 t_i 表示源模型和目标模型第 i 组对应面片的索引，我们一共找到 $|M|$ 组对应三角面片。这里值得注意的是，上面的对应关系可能有一对一，一对多(比较少见)的关系。文中所用的迁移方法就是直接把源模型每个三角形的变形梯度应用到目标模型上。写成一个优化目标就是：

$$\min \sum_{j=1}^{|M|} \|S_{s_j} - T_{t_j}\|_F^2 \quad (13)$$

这里每个 S_{s_j} 和 T_{t_j} 表示源模型和目标模型上第 s_j 和 t_j 个三角形的变形梯度。

为了计算简单，我们可以直接把源模型的变形梯度带入变形梯度计算公式里求得目标模型上每个三角形顶点的新位置。然而，由于变形梯度里面没有顶点位置的信息，我们得到的新的模型具有很大的误差，很多面片都是破损的(如图 3.3(A)所示)。

既然没有顶点的位置信息，那么能不能把源模型顶点的位置赋值给目标模型呢，图 3.3(B) 展示了这一结果，从中我们能看到，虽然目标模型(狮子)具有变形的倾向，但是由于直接将源模型顶点的位移赋值给目标模型也会引入误差，所以这种方法还是行不通。

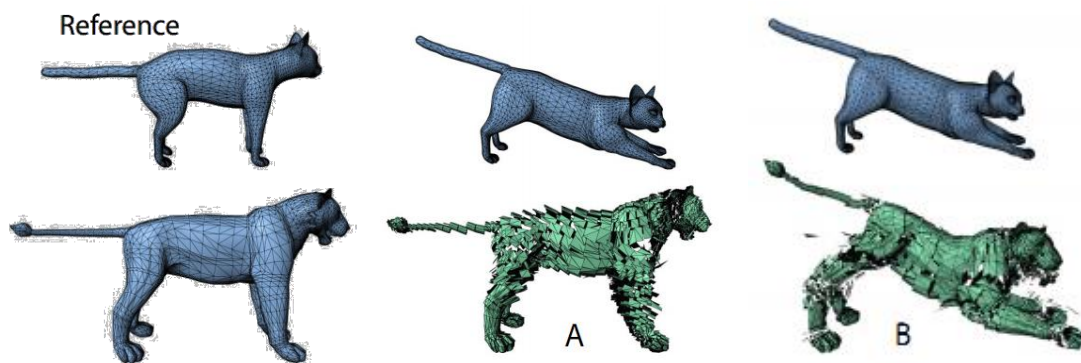


图 3.3 变形梯度优化计算(图片取自文献[RJ 04])

因此，文献[RJ 04]提出把上面的优化目标看作一个整体来进行整体优化的方法。仔细观察变形梯度，可以发现其实它是可以写成模型变形后顶点向量的线性形式的：

$$V'_j = \begin{bmatrix} v'_2 - v'_1 \\ v'_3 - v'_1 \\ v'_4 - v'_1 \end{bmatrix} \quad (14)$$

$$V_j^{-1} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix}$$

$$Q_{3 \times 3} = V'_j V_j^{-1} = \begin{pmatrix} -(h_{11} + h_{12} + h_{13}) & h_{11} & h_{12} & h_{13} \\ -(h_{21} + h_{22} + h_{23}) & h_{21} & h_{22} & h_{23} \\ -(h_{31} + h_{32} + h_{33}) & h_{31} & h_{32} & h_{33} \end{pmatrix} * \begin{pmatrix} v'_1 \\ v'_2 \\ v'_3 \\ v'_4 \end{pmatrix}$$

其中， V'_j 是第j个三角形变形后三边的边长， V_j^{-1} 是第j个三角形变形前包含边长信息矩阵的逆。从上式我们可以看出，变形梯度和模型变形后顶点存在线性关系，整个优化目标可以转化成一个线性系统。因此我们可以将此问题中每一组对应面片转化为一个线性最小二乘问题，再把所有的变量和方程结合在一起求解。为了做到这一点，下面对这一过程进行形式化表述：

模型变形后的顶点 V'_1, V'_2, \dots, V'_n 被重组成一个 $3n \times 1$ 的列向量，这里的n是模型顶点数加面数(也就是增广顶点数目)：

$$x_1 = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad x_2 = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \quad x_3 = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad (15)$$

其中， x_1, x_2, x_3 是分别包含x-,y-,z-坐标的 $n \times 1$ 的列向量。整个大x向量顺序包含了x-,y-,z-分量。

类似于顶点变量，模型的所有变形梯度 T_1, T_2, \dots, T_m 也可以被重组为一个 $9m \times 1$

的列向量，这里的 m 是模型的面数。

$$f_1 = \begin{bmatrix} T_1(1, :)^T \\ T_2(1, :)^T \\ \vdots \\ T_m(1, :)^T \end{bmatrix} \quad f_2 = \begin{bmatrix} T_1(2, :)^T \\ T_2(2, :)^T \\ \vdots \\ T_m(2, :)^T \end{bmatrix} \quad f_3 = \begin{bmatrix} T_1(3, :)^T \\ T_2(3, :)^T \\ \vdots \\ T_m(3, :)^T \end{bmatrix} \quad f = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} \quad (16)$$

其中的 $T_i(j, :)$ 表示模型第 i 个面片变形梯度的第 j 行，是一个 1×3 的行向量。因此， f_1, f_2, f_3 分别包含了模型所有面片变形梯度的第 1,2,3 行数据，因此它们都是 $3m \times 1$ 的列向量。至此，整个优化目标可以被写成一个线性系统：

$$\begin{bmatrix} A_{3m \times n} & & \\ & A_{3m \times n} & \\ & & A_{3m \times n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} \quad (17)$$

其中每个矩阵 A 都是 $3m \times n$ 的矩阵，且每一行只有四个位置不为零。 A 矩阵是由每个面片的变形梯度矩阵构造的。

前文我们说过，变形梯度没有记录顶点的位置信息，所以在优化的过程中，我们必须提供一些相关信息。文献[RJ 04]中提到求解的过程中最少只需要提供一个顶点的位置即可，这些顶点的位置使用约束项的形式提供，例如：

$$\min \|f - Ax\|_F^2, \text{ s.t. } x_k = c \quad (18)$$

x_k 表示模型的第 k 个顶点(不能是增广顶点)，优化计算中我们可以选择脸上很少动的顶点(比如头顶，耳朵，鼻梁)作为约束顶点。我们通过修改方程的右边可以很容易地将约束项添加到优化目标里：

$$f_{\text{new}} = A(x_k - c) = f - Ac \quad (19)$$

因为矩阵 A 的行数大于列数($3m > n$, n 是一倍的面片数加顶点数)，因此这个线性系统是过定的，有唯一的最优解。

$$\begin{aligned} \min \|f - Ax\|_F^2 \\ A^T Ax = A^T f \\ x = (A^T A)^{-1} A^T f \end{aligned} \quad (20)$$

由于矩阵 A 每一行只有四个位置不是零，所以它是高度稀疏的矩阵。但是如果按照最小二乘方法来计算求 A 的伪逆，这样会改变矩阵的稀疏性。因此，从计算效率的角度来看更好的方法是采用对矩阵 $A^T A$ 做乔里斯基分解，使其分解为一

个上三角矩阵和一个下三角矩阵的乘积。而上(下)三角矩阵作为系数的线性系统求解是非常高效的。由于矩阵 A 只与中性表情模型有关，因此在做变形迁移时不需要重新计算 A 。

正如第二章中所介绍的，人脸属于非线性的表面，对人脸模型进行配准主流的方法是基于变形的配准方法。具体思想就是利用用户指定的一些对应顶点，将两个人脸模型逐步对齐到一起，然后再求配准关系。下面对这些步骤进行详细介绍。

3.3 模型配准

3.3.1 初始对应特征选择

文献[RJ 04]中采用了文献[JU 01]标记人脸特征点的方法，一共标记了 42 组对应点。

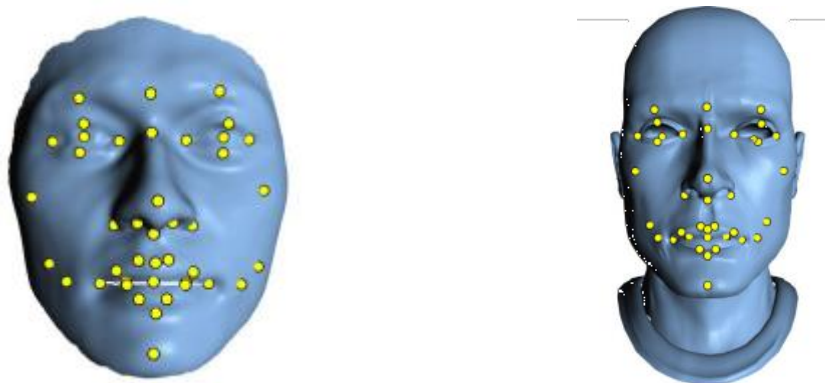


图 3.4 文献[RJ 04]中对人脸模型的对应点选择方法

但是，手动选择 42 组对应点不仅容易出错，也是一项比较费时的工序。因此，本文提出一种全自动方法来检测模型的 25 组对应点，并通过实验验证，这 25 组对应点已经可以很好地对模型进行配准。

首先，利用启发式的方法选择模型表面的 14 组对应顶点，它们分别是：

1. 鼻尖，模型 z 坐标最大的顶点；
2. 头顶，模型 y 坐标最大的顶点；
3. 左右耳尖，模型 x 坐标最小和最大的顶点；
4. 鼻梁，从鼻尖出发向 y 轴正方向上 z 坐标的第一个谷点；
5. 左右鼻衬，从鼻尖出发向 x 轴左右方向上 z 坐标的第一个谷点；
6. 上下唇尖，从鼻尖出发向 y 轴负方向上 z 坐标的两个峰点；
7. 鼻根，从鼻尖出发向 y 轴负方向上的第一个谷点；
8. 下巴，从鼻尖出发向 y 轴负方向上 z 坐标的最后一个峰点；
9. 左右近眼角，从鼻梁出发向 x 轴左右方向上 z 坐标的第一个谷点。

10. 后脑，模型 z 坐标最小的顶点；
使用启发式方法获得的特征点如图 3.5 所示：

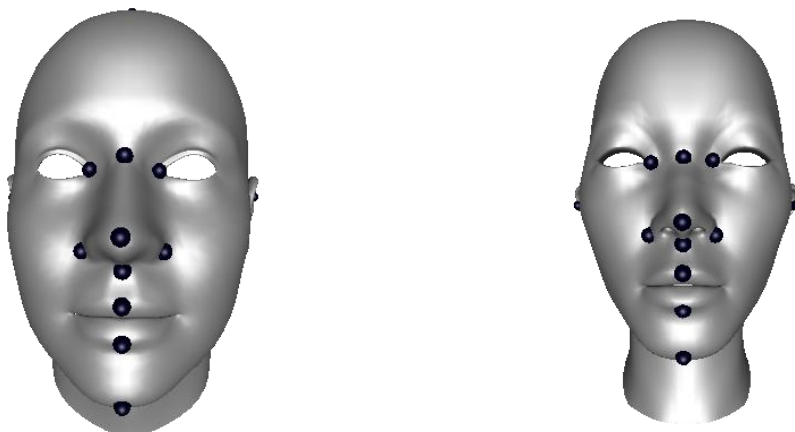


图 3.5 本文中利用启发式方法对人脸模型的对应点选择

但是仅用这 14 组特征点不足以得到很精确的模型配准。因此，我们采用下面的方法来获得更多的对应顶点：

1. 在一个模版模型上标记 25 个顶点，其中包含上述的 14 个顶点；
2. 训练模版与待标记模型 14 个顶点的 RBF 差值模型；
3. 将训练模版上的其它 11 个顶点带入 RBF 差值模型得到其插值位置；
4. 计算这些顶点与待标记模型最近的顶点，将这些顶点作为新的对应点。

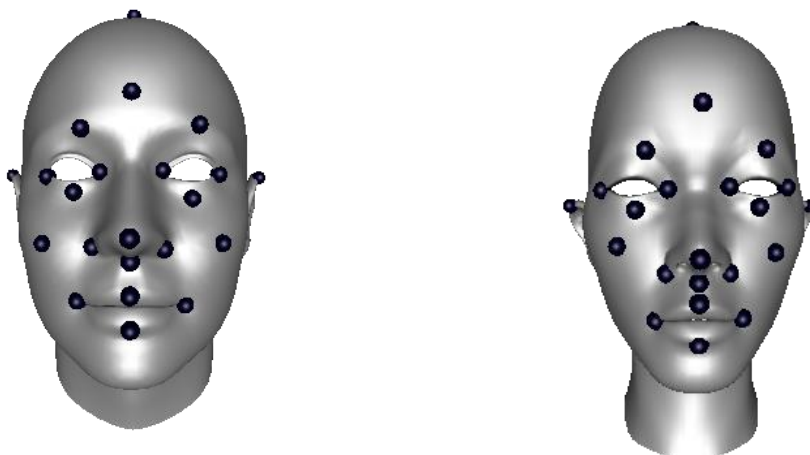


图 3.6 本文中利用启发式方法对人脸模型的对应点选择

3.3.2 基于变形梯度的模型变形

获得一组初始的对应点后，接下来要用上一章中介绍的变形梯度方法来进行模型的表面变形。这一方法对文献[JU 01]方法的优点是，它可以保持原有模型的拓扑结构，顶点连接关系不会随着模型形变而变化。

类似于文献[JU 01]提出的模型配准思路，本文也按照如下的步骤进行：

1. 利用预先指定的一些模型对应点进行两个模型的初步对齐；
2. 寻找两个模型间的最近点对，建立顶点之间的对应关系；
3. 利用第 2 步中的顶点对应关系，使两个模型进一步对齐；
4. 跳入第 2 步，直至收敛。

在模型变形过程中，Sumner 等人为了使两个不同拓扑结构的模型可以平滑地变形到彼此，设计了变形中应遵循的约束，如表 3 所示。

变形一致性	$E_S = \sum_{j=1}^{ \mathcal{J} } \left\ \mathbf{T}_j - \left(\sum_{k \in \text{adj}(j)} \frac{1}{ \text{adj}(j) } \mathbf{T}_k \right) \right\ _F^2$
变形合理性	$E_I = \sum_{j=1}^{ \mathcal{J} } \left\ \mathbf{T}_j - \mathbf{I} \right\ _F^2$
顶点相关性	$E_C = \sum_{i=1}^n \left\ \mathbf{v}_i - \mathbf{c}_i \right\ _2^2$

表 3 变形梯度法进行模型配准应遵循的约束

其中，变形一致性约束中， \mathbf{T}_j 表示第 j 个三角面片的变形梯度， $\text{adj}(j)$ 表示与第 j 个三角面片相邻的面片索引。最小化这一约束表示每个三角面片发生变形必须和周围的面片保持同步，不能发生一个面片变形过大的情况。

在变形合理性表示中， \mathbf{I} 表示单位矩阵，如果变形梯度为单位矩阵则表示不发生任何变形。最小化上述形式表示三角面片发生变形过程中宁可选择变形较小的变化。

在顶点相关性表示中，两个模型在经过对齐后，计算源模型顶点 \mathbf{v}_i 在目标模型上的最近点 \mathbf{c}_i ，而 \mathbf{v}_i 和 \mathbf{c}_i 这两个顶点的距离应该足够小。

综合以上三种约束，下面给出模型配准的优化方程：

$$\begin{aligned} \min_{\mathbf{v}_1 \cdots \mathbf{v}_n} \quad & \omega_S E_S + \omega_I E_I + \omega_C E_C & (21) \\ \text{subject to} \quad & \mathbf{v}_k = \mathbf{m}_k, k \in 1, \cdots, m \end{aligned}$$

其中，权重 ω_S ， ω_I ， ω_C 衡量了变形一致性、变形合理性和顶点相关性三种约束重要关系。在这三种约束之下，顶点 \mathbf{v}_k 应该被移动到 \mathbf{m}_k 。 \mathbf{v}_k 和 \mathbf{m}_k 由初始选定的对应点决定。同 2.2 节中的变形梯度优化问题一样，这些顶点的位置关系可以加入到优化方程中。

在求解这一优化问题的时候本文对[RJ 04]的方法进行了修改，[RJ 04]中使用了顶点和增广顶点作为优化变量，而实际上增广顶点的求解并没有实际意义，反

而使优化问题的求解效率更低。文献[MRR 06]指出，本问题的精确优化可以使用三角形的法线向量代替增广顶点形成的独立边。但是三角形的法线是三个顶点的非线性形式，为了得到模型顶点的线性优化形式，我们需要去掉变形梯度中的法线方向变量。下面提出利用矩阵的 QR 分解精简增广顶点的方法：

	变形	变形梯度	线性系统
原文方法	$V = \begin{bmatrix} v_2 - v_1 \\ v_3 - v_1 \\ v_4 - v_1 \end{bmatrix}^T$ $\tilde{V} = \begin{bmatrix} \tilde{v}_2 - \tilde{v}_1 \\ \tilde{v}_3 - \tilde{v}_1 \\ \tilde{v}_4 - \tilde{v}_1 \end{bmatrix}^T$	$H_{3 \times 3} = \tilde{V}V^{-1}$	$\begin{pmatrix} -(h_{11} + h_{12} + h_{13}) & h_{11} & h_{12} & h_{13} \\ -(h_{21} + h_{22} + h_{23}) & h_{21} & h_{22} & h_{23} \\ -(h_{31} + h_{32} + h_{33}) & h_{31} & h_{32} & h_{33} \end{pmatrix} * \begin{pmatrix} \tilde{v}_1 \\ \tilde{v}_2 \\ \tilde{v}_3 \\ \tilde{v}_4 \end{pmatrix}$
本文方法	$V = \begin{bmatrix} v_2 - v_1 \\ v_3 - v_1 \end{bmatrix}^T$ $\tilde{V} = \begin{bmatrix} \tilde{v}_2 - \tilde{v}_1 \\ \tilde{v}_3 - \tilde{v}_1 \end{bmatrix}^T$	$V = Q_{3 \times 2}R_{2 \times 2}$ $V^{-1} = R_{2 \times 2}^{-1}Q_{2 \times 3}^T$ $H_{3 \times 3} = \tilde{V}V^{-1}$	$\begin{pmatrix} -(a + d) & a & d \\ -(b + e) & b & e \\ -(c + f) & c & f \end{pmatrix} * \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{pmatrix}$

表 4 本文方法与原文([RJ 04])方法的区别

通过本文的方法，优化变量可以减少到原来的三分之一左右，大大提高了程序执行的速度。

在进行第一步优化时，因为我们还不能计算两个模型的最近顶点， ω_c 被设为 0。经过第一步对齐之后，初始的对应顶点被对齐到一起，整个模型也进行了变形。我们下一步要计算获得两个模型的最近顶点。具体的规则如下：

1. 计算变形模型上顶点 v_k 与目标模型上最相近的几个顶点
2. 比较变形模型上顶点 v_k 与目标模型上最近顶点法线的夹角，选择夹角小于 90 度的最近顶点。

通过这种方法找到的最近顶点可以避免一些错误的匹配点，比如上下嘴唇的顶点，它们的欧式距离很近，很容易进行误匹配。

在计算最近顶点的时候，为了减少计算量(目前的方法需要对目标模型上的每个顶点计算欧式距离)，我们也采用了[RJ 04]中提到的进行空间网格划分的方法。具体算法是：

算法 3: 模型空间划分算法

1. 计算目标模型的长方体包围盒；
2. 对包围盒进行分块，本文采用的分块方法是将顶点按照 x,y,z 轴方向上包围盒

边长的比例进行均匀分配, 例如: 7000 个顶点模型, x,y,z 轴方向上包围盒边长的比例为 5:3:2, 那么三个方向上分别应该划分 3500, 2100, 1400 个网格。每个网格存储包含在内部的顶点和其法线信息。由于模型是面片模型, 大多数网格里都没有顶点。我们按照每个网格在包围盒 x,y,z 轴方向上网格划分的索引存储非空的网格;

3. 在计算变形模型对目标模型的最近点时, 我们先计算变形模型某顶点应该落在包围盒的哪个网格, 然后按照索引获得其内部的顶点。如果该网格为空, 或者其内部的顶点都不满足法线朝向夹角的约束, 那么我们采用就近寻找的原则, 依据广度优先的顺序寻找 x,y,z 轴方向上的近邻网格。

找到两个模型的最近点之后, 我们逐步增加 ω_c 的值, 进行迭代优化。迭代若干步后, 可以得到变形模型和目标模型比较好的对齐。

3.3.3 确定配准关系

在得到比较精确的模型对齐关系后, 我们通过计算最近三角面片来获得两个模型之间面片的配准关系。

类似于最近顶点的计算, 我们用变形模型某一三角面片的中心坐标作为三角形的位置, 计算目标模型上与其最相近的几个备选面片。计算最近面片的时候, 为了避免得到错误的对应面片(比如上下嘴唇的面片之间的欧式距离很近, 容易得到错误对应), 我们需要规定源模型和目标模型的某个三角面片中心距离小于某个阈值, 并且其法线夹角 α 不能超过九十度(如图 3.7 所示)。

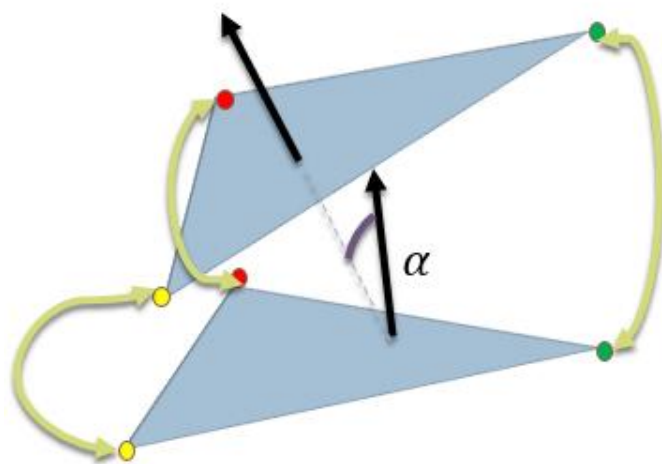


图 3.7 三角面片配对计算

实际计算中为了减少不必要的运算量, 我们也采用空间划分的方式将目标模型的所有面片分散到小网格中。才后采用类似于上一节的方法计算得到配对的三

角面片。在得到两个模型的面片对应关系后，我们就能利用变形梯度方法进行表情克隆了。

3.4 实验

3.4.1 程序实现细节

本文提出的算法使用 c/c++ 实现，显示功能使用 OpenGL 实现。在变形梯度方法的求解过程中，我使用了 UMFPACK 工具包。具体的数据和程序列表如下：

变量(功能)	使用的数据结构(函数)
A 矩阵(2.2 节)	cholmod_sparse 稀疏矩阵存储结构
顶点位置向量	cholmod_dense 稠密矩阵存储结构
稀疏矩阵转置	cholmod_transpose
稀疏矩阵相乘	cholmod_sdmult
求 $A \cdot A^T$	cholmod_aat
稀疏矩阵相乘行顺序重排	cholmod_sort
LU 分解	umfpack_di_symbolic umfpack_di_numeric
线性系统求解	umfpack_di_solve
乔里斯基分解	cholmod_analyze cholmod_factorize
线性系统求解	cholmod_solve

表 5 算法主要功能数据结构和函数表

其中进行矩阵分解时采用了两种方法，LU 分解和乔里斯基分解。实际应用中，发现它们具有近似的数值精度。

3.4.2 实验结果分析

在使用本文方法进行模型配准时， ω_C 的初始值为 0。 $\omega_S = 1$ ， $\omega_I = 0.001$ 。第二步之后 $\omega_C = 1:10:41$ ，这样经过五步迭代后就能得到一个比较精确的变形模型了。

为了比较本文与[RJ 04]工作效果上的差异，表 6 和表 7 给出使用两种方法的重构误差分析。由于变形模型和目标模型的顶点数目不同，我们选用第三章中计算最近点对的方法来计算重构误差。下面误差值是由以下公式得到。

$$error = \frac{1}{n} \sum \frac{dist(v_i, c_i)}{modelsize}$$

$dist(v_i, c_i)$ 是最近顶点之间的欧式距离， $modelsize$ 是模型的大小。

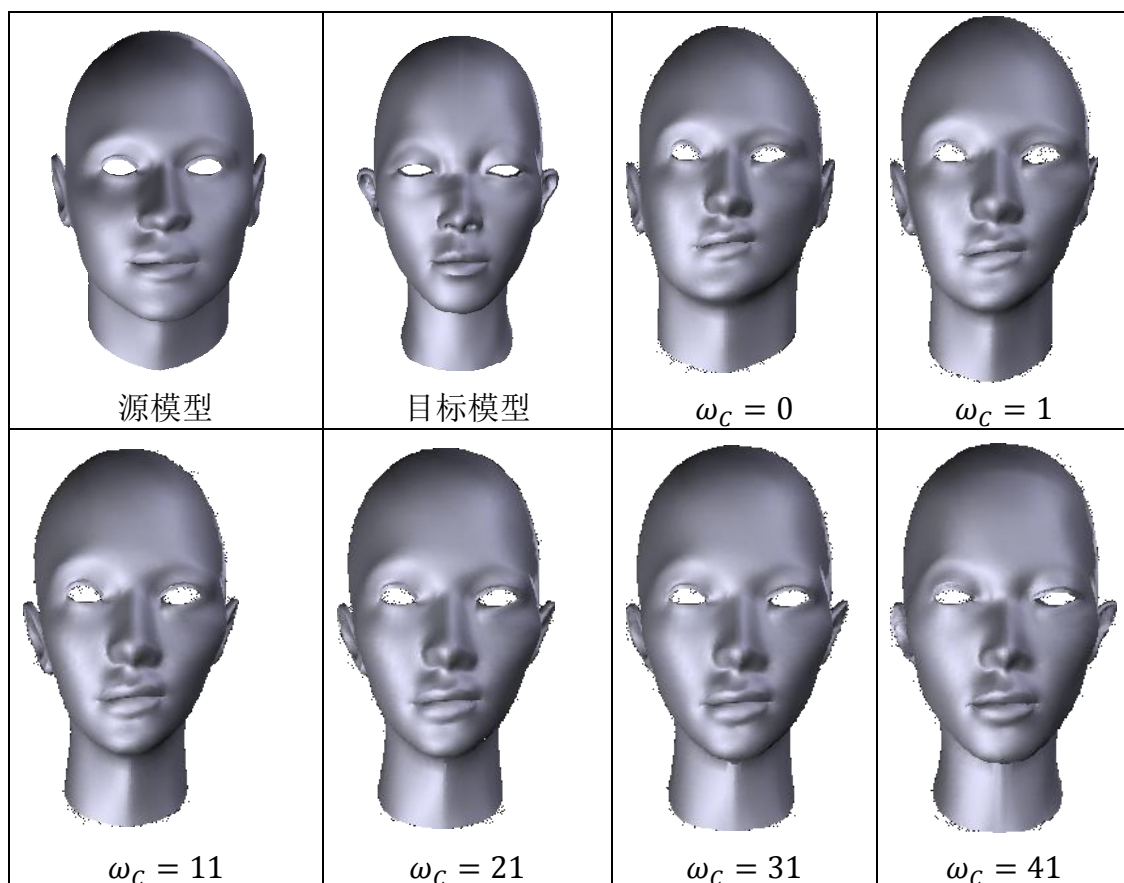


图 3.8 本文方法进行模型配准中间结果

我们选择两种不同的初始对应点标记方法，第二种情况手工标记 40 组对应点。可以看出本文的模型配准方法与[RJ 04]方法误差非常接近。

匹配点对数目	表面重构误差(相对于模型大小)改进方法	表面重构误差([RJ 04]方法)
25	2.3%	2.1%
40	1.6%	1.2%

表 6 不同初始匹配点情况下的表面重构误差比较

迭代范围	表面重构误差(改进方法)	表面重构误差([RJ 04]方法)
1:10:50	2.3%	2.1%
1:10:100	1.9%	1.6%

表 7 不同迭代次数情况下的表面重构误差比较

由于采用了本文方法，模型配准的时间被大大减少，表 8 采用不同顶点数的

模型进行模型配准，比较本文方法和[RJ 04]方法的时间效率。从表 8 中我们可以看出，本文方法比[RJ 04]方法提速了大约三倍左右。从实验数据来看，模型的面片数大约是顶点数的两倍，这与我们得到的结论相吻合。

源模型	目标模型	模型匹配时间 ([RJ 04]方法)	模型匹配时间 (改进方法)
人脸模型 1 7032 顶点	人脸模型 2 5832 顶点	25.3s	7.1s
人脸模型 1	狮子 275 顶点	17.6s	5.8s
人脸模型 1	鹿 2702 顶点	21.2s	6.3s

表 8 对不同模型进行配准时算法运行时间比较

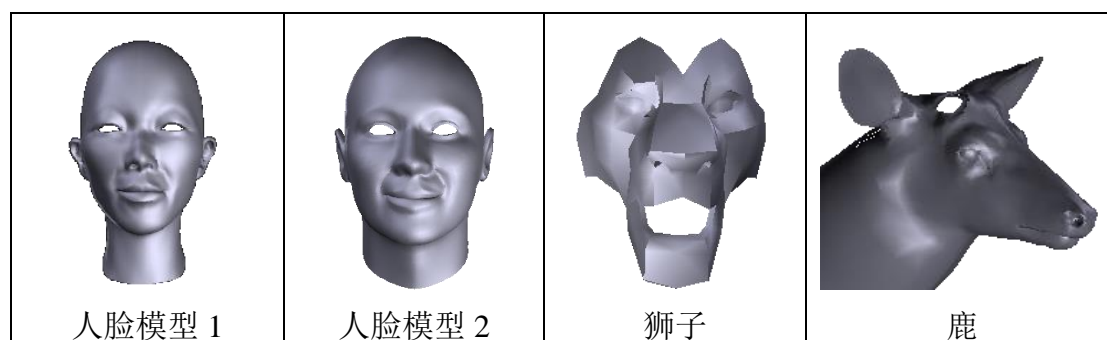
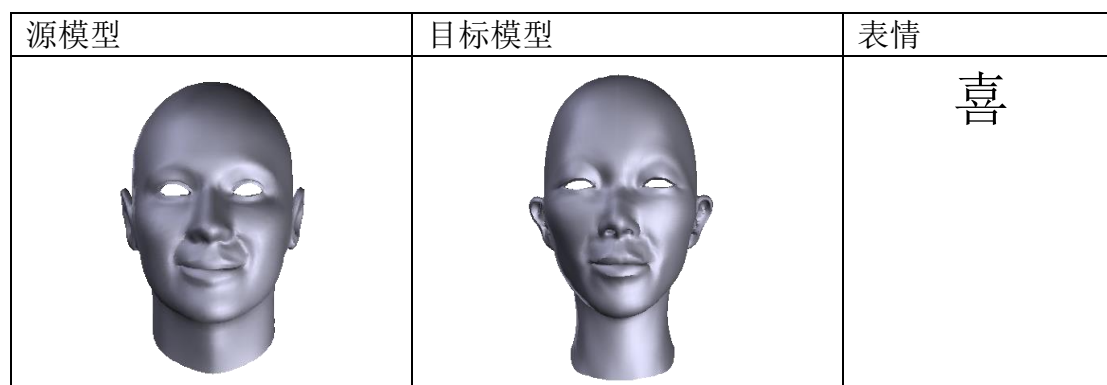


图 3.9 表 7 所用各种模型示意图

图 3.10 给出用本文方法进行人脸表情克隆的结果。



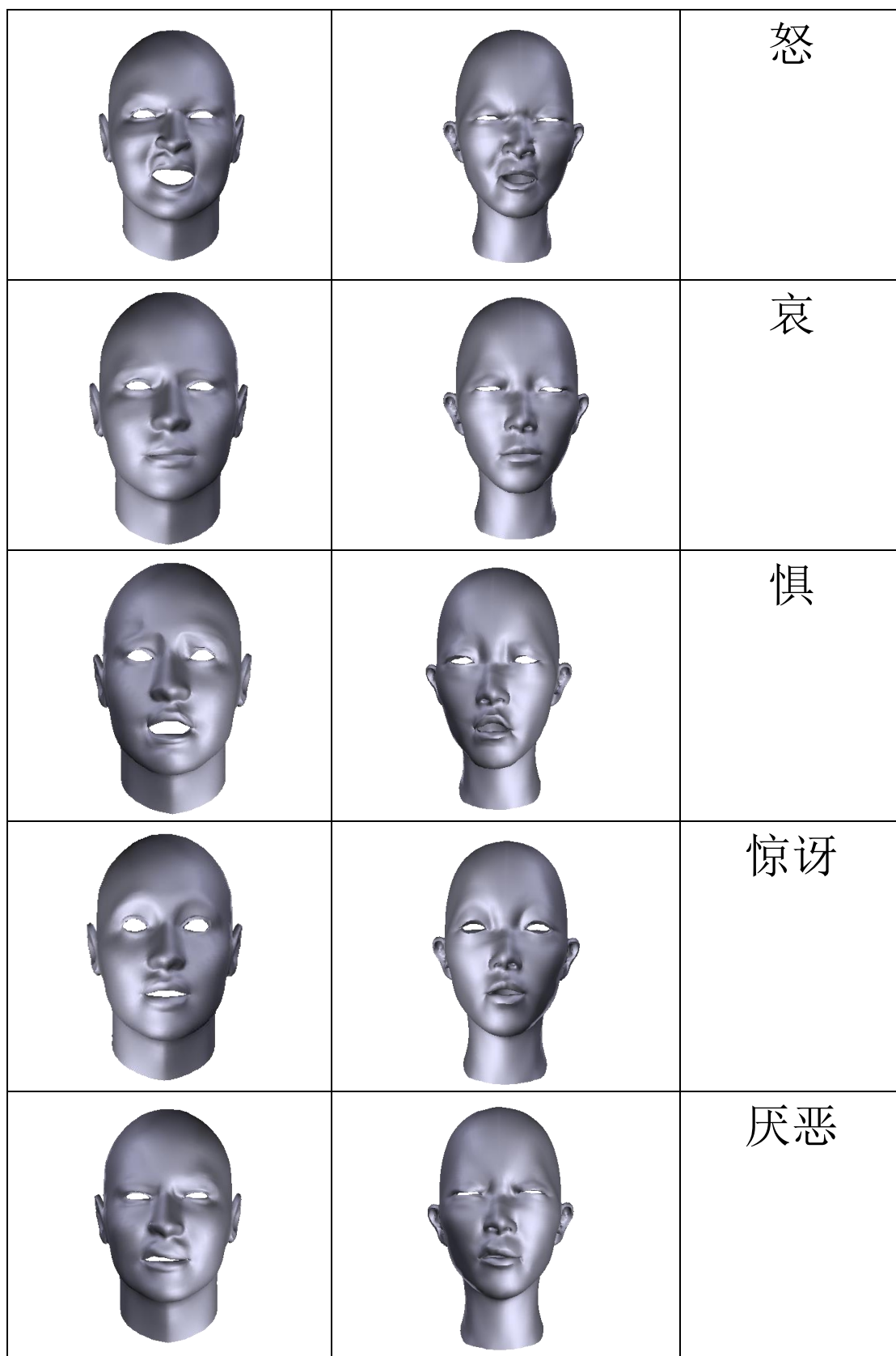


图 3.10 表情克隆结果

为了检验算法的普适性，我用狮子和鹿的模型进行了模型匹配和表情克隆实

验。由于动物脸部和人脸有很大的结构差异，我手动标记了 25 组对应点。结果如下图 3.11 所示。从图中可以看出，本文方法不能将动物脸部严格对齐到人脸，但是可以获得动物脸部的拟人化表示。表情迁移也只能获得比较近似的结果。一方面，这是因为人脸和动物脸部拓扑结构差异过大，另一方面，初始对应点不足也是模型对齐效果不佳的主要原因。

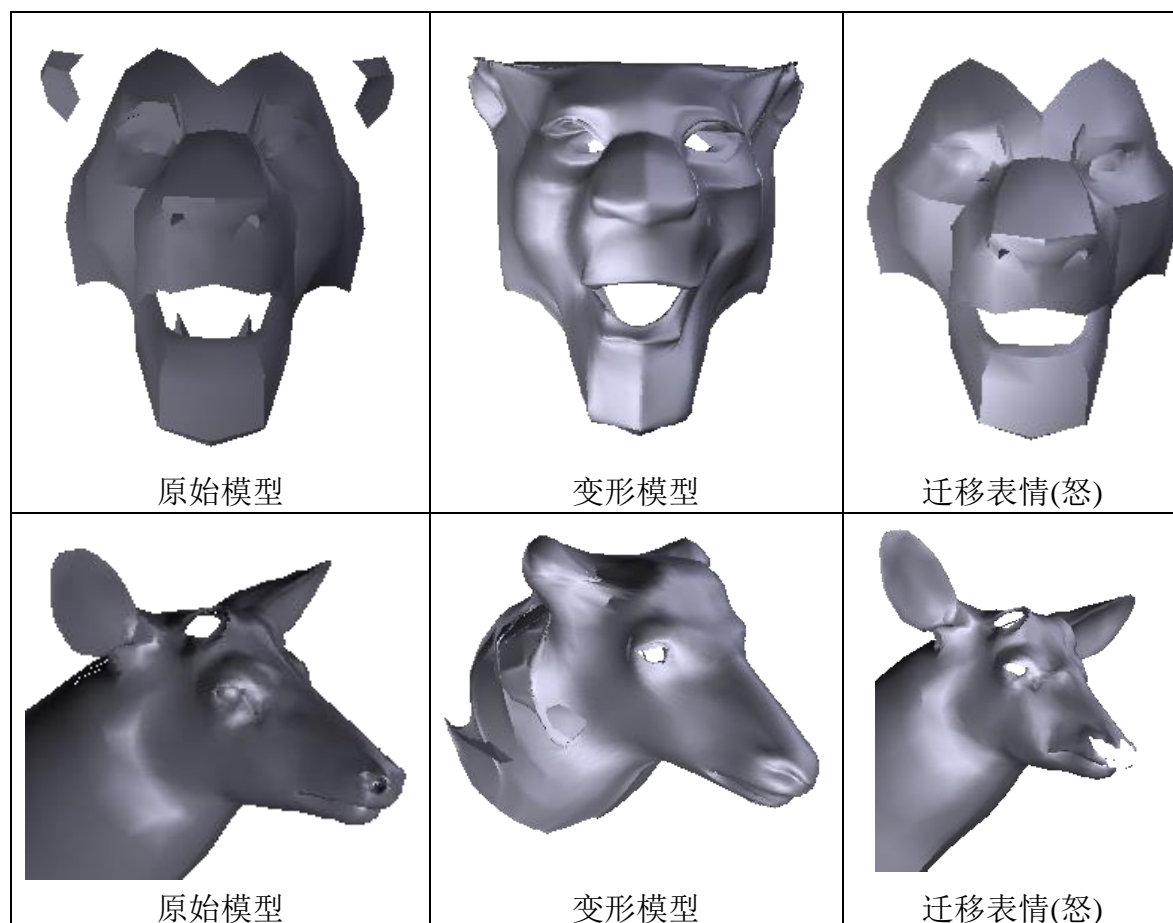


图 3.11 应用本文算法到动物模型上的结果

3.5 小结

本章首先介绍了几何形状之间变形的定义，接着给出了描述变形的两种数学表示方法：顶点位移差和形状之间的变形梯度。文中首先在假设模型配准关系已知的前提下，给出变形梯度迁移的理论表达形式。并通过改进优化目标，将其表示为一个变量为模型顶点位置的最小二乘问题。

第二部分主要介绍了利用变形梯度方法进行模型配准的过程，主要思路是利用初始设置的模型对应点逐步对齐全源模型和目标模型。本文对[RJ 04]方法进行了两个方面的改进：

- (1) 在进行初始对应点的选择上，本文提出了一种完全自动的选择方式。即

首先利用启发式思路选择十几组对应点，再利用 **RBF** 函数预测获得另外十几组对应点。

(2) 修改了原文中的优化形式，优化变量中去掉了增广顶点，减少了 $2/3$ 的计算量，大大提高了算法计算效率。

4 视频驱动人脸动画

本文第三章介绍的表情克隆工作可以很好地利用现有的人脸动画表情库，生成新模型的表情。这种方法可以应用于电影、游戏等的人脸动画制作当中。这一章中，我会给出表情克隆工作的一个具体应用:视频驱动人脸动画。

4.1 总体框架介绍

本章介绍一种实时、方便的视频交互方法来控制 3D 人脸化身生成表情。整体系统框架如图 4.1 所示。

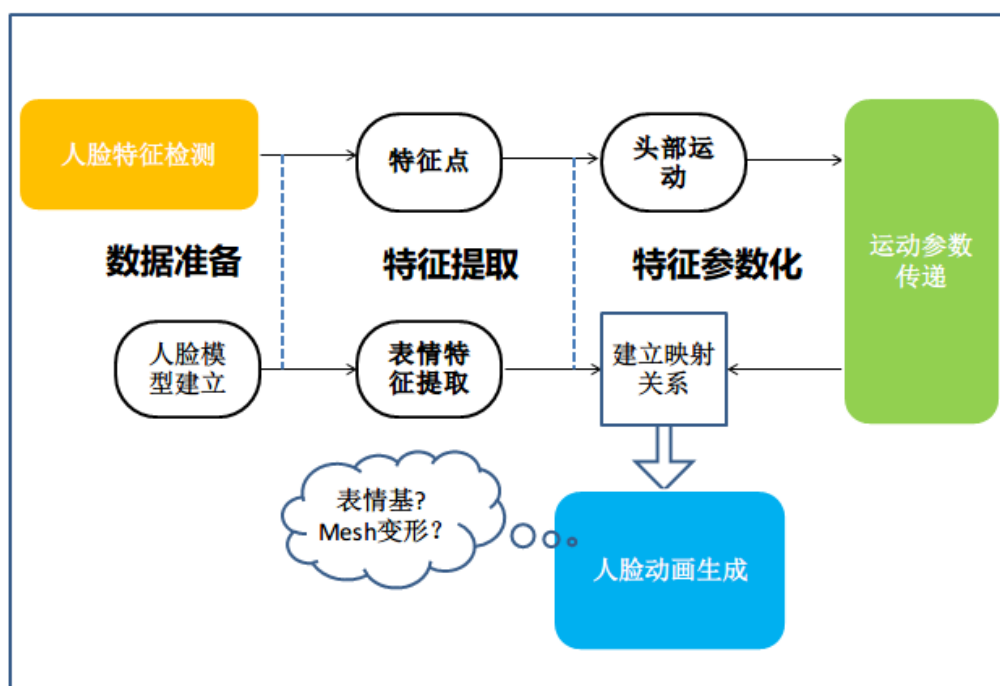


图 4.1 视频驱动人脸动画系统框架图

该框架包括以下处理流程：

数据准备。包括视频图像数据的获取和 3D 人脸化身模型准备。

特征提取。包括提取视频图像中人脸特征点以及 3D 人脸化身表情特征(对应的 3D 特征点)提取。

特征参数化。包括头部运动估计和表情运动迁移两部分。

人脸动画生成。通过表情基或者 mesh 变形等方法获得人脸动画。这部分可以直接使用表情克隆工作合成的表情基。

该系统的主要难点在于实时、鲁棒地跟踪人脸特征点，以及使用 2D 运动信号驱动生成 3D 人脸动画。为了进行实时鲁棒的人脸特征跟踪，本文采用了约束

局部模型。该模型通过对优化过程引入凸性约束，减少了优化变量，计算速度非常快；通过对脸部特征点几何位置以及周围领域像素进行统计建模，可以非常鲁棒地检测到面部特征点。为了获得 3D 运动信号，本文通过棋盘格法估计相机的内参矩阵，然后将 2D 运动信号反投影到 3D 的相机空间，并以此 3D 运动信号来驱动人脸动画生成。

本章 4.2 节详细介绍了本系统所用到的约束局部模型、运动参数化以及人脸动画生成的相关技术。

4.2 技术细节

4.2.1 约束局部模型

常见的人脸特征检测算法有光流法、主动形状模型(ASM)，主动表观模型(AAM)等。光流法检测跟踪到的特征点没有标记(如眼角、嘴角等)，而且光流法噪声较大。主动形状模型检测到的特征点具有标记，比较鲁棒，但该方法对光照、人脸姿态变化很敏感。主动表观模型通过建立特征点局部表观模型，可以更好的适应光照变化，但是它的求解过程中优化变量过多，检测速度很难有大幅度提高。[DT 06]引入了约束局部模型(CLM)，通过对表观模型加入凸性假设，将像素块表示为二次函数的系数，减少了优化变量数目，而且更符合人脸特征点的变化规律。图 4.2 是 CLM 方法的流程展示。

从图 4.2 中可以看出，约束局部模型由形状模型和表观模型构成。表观模型采用线性 SVM 来建立，对于每一个特征点，要对其附近的窗函数采样建立一个线性 SVM 分类器，该分类器的系数被用做滤波掩码。在检测人脸特征时，该掩码作用于窗函数采样上，获得一个响应图像 (Response Image) $R(x,y)$ 。假设 $[x_0, y_0]$ 为响应图像中像素值最大的点,我们对该区域可以拟合一个二次曲线。

$$r(x, y) = a(x - x_0)^2 + b(y - y_0)^2 + c \quad (22)$$

以使得

$$\min \epsilon = \sum_{x,y} [R(x, y) - r(x, y)]^2 \quad (23)$$

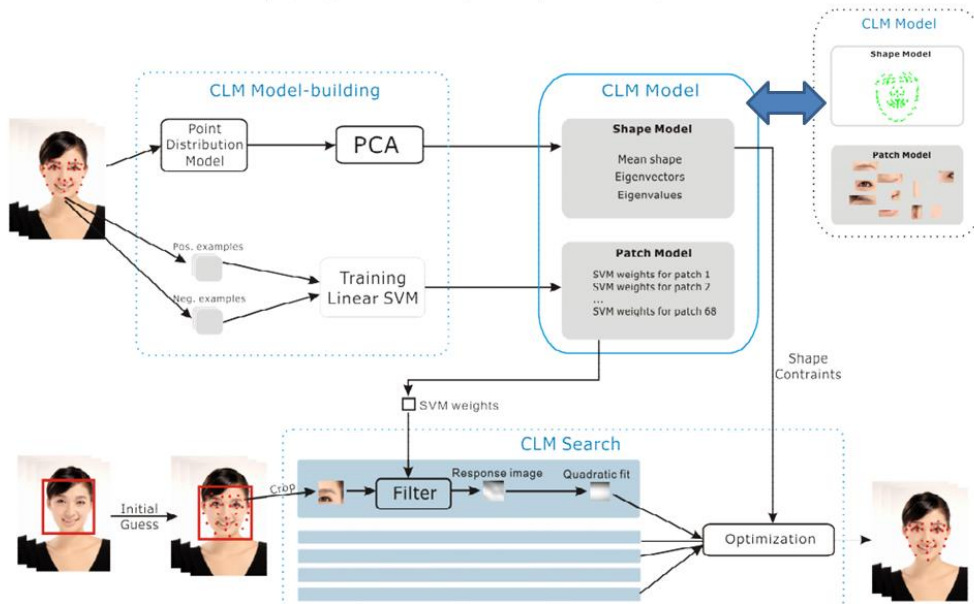


图 4.2 CLM 方法流程

通过引入 $\omega = [a, b, c]$ ，公式(23)可以写成

$$\epsilon = \|A\omega - R\|^2 = \omega^T A^T A \omega - 2R^T A \omega + R^T R \quad (24)$$

$\omega = [a, b, c]$ 中的 a, b 被加上大于 0 的条件以保证凸性。求解出 ω 后，即可将 (x, y) 作为变量，在原优化函数中进行求解

$$\min \sum_{i=1}^n r_i(x_i, y_i) - \beta \sum_{j=1}^k \frac{b_j^2}{\lambda_j} \quad (25)$$

公式(25)中第一项为每个特征点处的重构误差，第二项为特征点的结构约束。具体的算法如下：

算法 4: CLM 搜索过程

1. 检测人脸位置，猜测初始时特征点的位置；
2. 在特征点处采样，求取响应图像；
3. 依次优化公式(24)中的 $\omega = [a, b, c]$ 和公式(25)中的 (x_i, y_i)
4. 继续执行 2-3 步，直至两次优化的特征点位置差异足够小。

4.2.2 表情运动参数化

表情运动参数化需要将视频中提取的 2D 运动信号转换为驱动生成 3D 人脸表情的控制信号。文献[JJJ 03]和[TYJ 11]采用了提取表情控制参数的方法。他们提取了如特征点间距、特征点与某条线之间距离、特征点组成的线之间朝向等。这些控制参数对于 2D 和 3D 运动信号具有不变性，因而可以用来驱动 3D 表情生成。图 4.3 给出了 [JJJ 03] 中表情参数的一些实例。

对视频中提取的运动信息序列和数据库中检索到的 K 个最接近的运动信息序列, [JJJ 03]通过计算他们的欧式距离来估计一个插值权重。通过插值可以与 2D 特征点序列对应的 3D Vicon 数据点云, 该 3D 的 Vicon 数据点云接着被用来生成人脸动画。

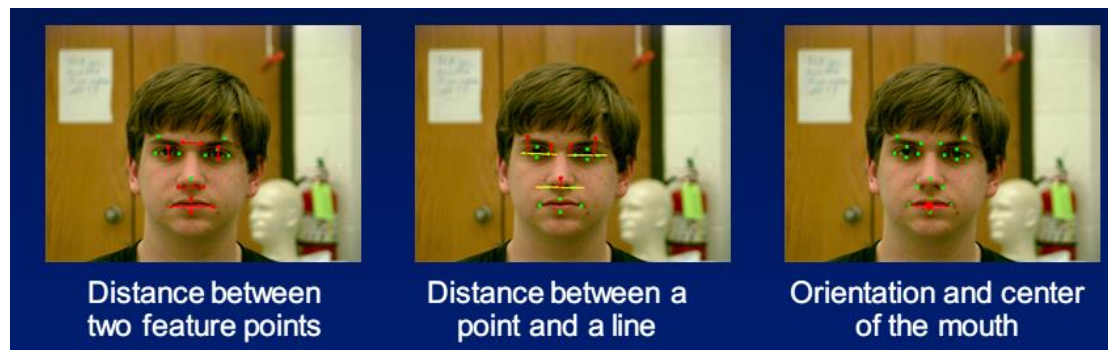


图 4.3 文献[JJJ 03] 表情运动参数化方法

但是, 这种表情参数化方法对于刻画局部表情效果欠佳, 而且不能反映单个特征点的位置变化。因此, 我们希望采用更直观的方法, 将 2D 的人脸特征位置直接转化为 3D 特征位置。具体的思路是:

1. 通过估计摄像头的参数恢复 2D 特征点的深度信息;
2. 计算深度恢复后 3D 特征点序列的旋转、平移等刚性运动信息;
3. 通过运动插值方法, 将不同拓扑结构模型上的运动参数进行迁移。
4. 使用稀疏表情基计算去除旋转、平移后人脸的表情运动权值;

首先, 我们可以用棋盘格法计算摄像头的内参矩阵。

$$\mathbf{A} = \begin{bmatrix} x_a & 0 & p_a \\ 0 & x_b & p_b \end{bmatrix} \quad (26)$$

其中 x_a 和 x_b 是相机在 x-y 方向上的焦距大小, p_a 和 p_b 是成像平面中点的坐标值。3D 的特征点位置 \mathbf{f}'_i 可以由 2D 特征点位置 \mathbf{f}_i 经过

$$\mathbf{f}'_i = \mathbf{f}_i * (\mathbf{A}\mathbf{A}^T)^{-1} * \mathbf{A} \quad (27)$$

得到。这样得到的 3D 坐标只是相机空间中的位置, 而不是真实世界坐标中的位置。我们采用该 3D 坐标来近似其真实 3D 位置。

人脸做表情时, 头部是在不断运动的。为了获得准确的表情参数, 我们需要去除头部运动的影响。一般情况下, 我们取开始采集时的第一帧作为中性帧, 头部旋转和平移等可以通过

$$[U, D, V] = Cov_{X,Y}, R = U * I * V' \quad (28)$$

算出。公式(28)中 X 表示中性帧顶点的位置, Y 表示其它帧顶点的位置。通过对

X 和 Y 的协方差矩阵进行 SVD 分解, 我们可以获得头部的旋转角度。通过乘以旋转矩阵的转置, 我们可以将 3D 特征点转到初始位置。去除头部朝向影响后, 我们用稀疏表情基对 3D 运动数据进行参数化。

稀疏的表情基指的是从表情基模型上获得的具有语义信息的 3D 特征点序列, 如图 4.4 所示。

由于演员(用户)脸部的拓扑结构和表情基模型的结构不一样, 我们首先要将稀疏的表情基插值到视频中恢复的 3D 特征点空间。步骤如下:

1. 取视频中恢复的 3D 特征点的某一帧 \mathbf{f}_1 作为演员运动的中性表情帧(可以通过程序交互的方式指定);
2. 建立稀疏的模型中性表情基 \mathbf{e}'_0 (输入)与视频中恢复的 3D 特征点的第一帧 \mathbf{f}_1 (输出)的 RBF 映射关系;
3. 使用 RBF 函数对其它稀疏的模型表情基 $(\mathbf{e}'_1, \dots, \mathbf{e}'_N)$ 插值得到所有稀疏运动表情基 $(\mathbf{e}_1, \dots, \mathbf{e}_N)$ 。

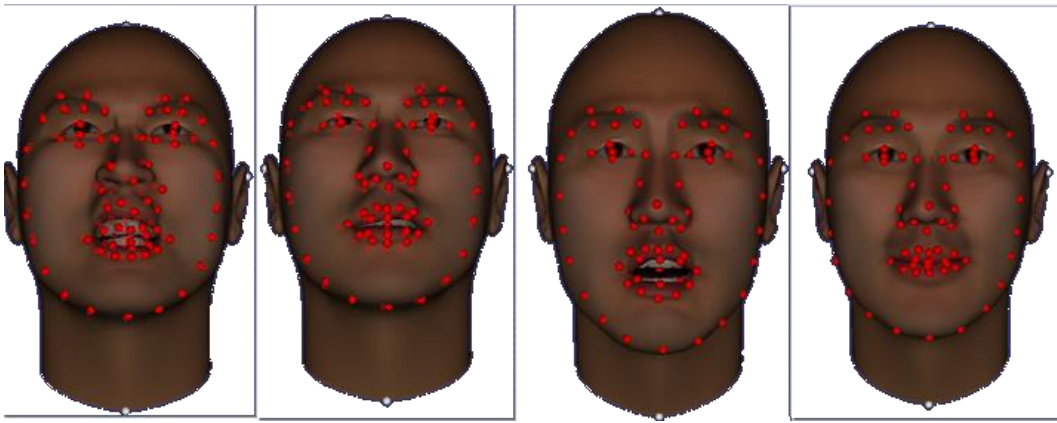
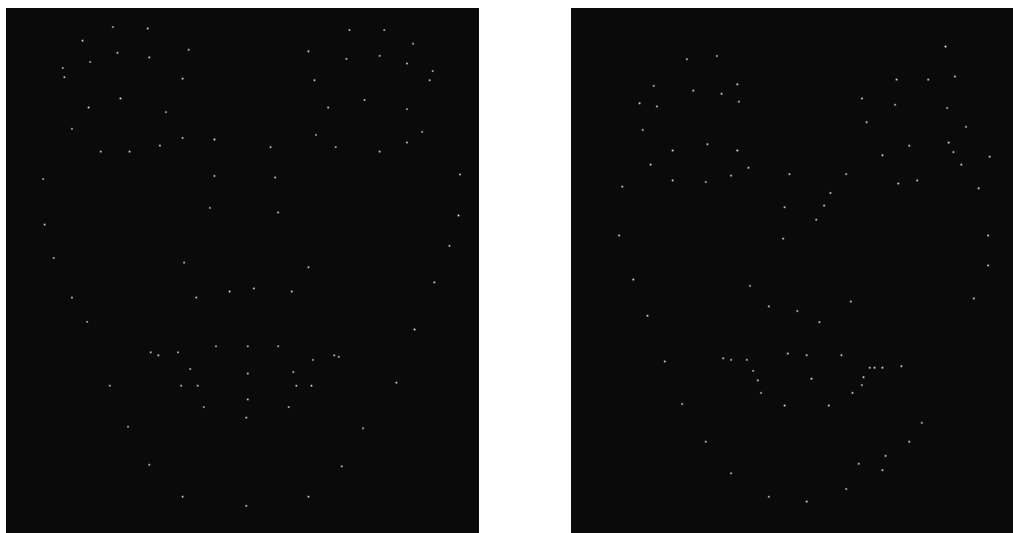


图 4.4 稠密的表情基模型——对应于稀疏表情基(红点表示)

第一章中已经介绍了 RBF 的形式(2), 但是实验证明对于人脸表情基该形式的 RBF 生成的稀疏运动表情基不能很好地体现稀疏的模型表情基的结构, 生成的人脸动画会严重变形(如图 4.5(b)所示)。因此, 我们为 RBF 函数加上一个线性项 $\omega_0 = [a_i, b_i, c_i]$ 。生成的结果如图 4.5(a)所示。

$$\mathbf{F}(\mathbf{x}) = \sum_{i=1}^M \omega_i h_j(\|\mathbf{x} - \mathbf{x}_i\|) + \omega_0 \mathbf{x} \quad (29)$$



a. 加了线性项的插值结果

b. 不加线性项的插值结果

图 4.5 RBF 函数不同形式对应的插值结果

如公式(30)所示，表情基权值的计算可以通过约束最小二乘的方法求得，即：

$$\min \sum_i^M \|\mathbf{F}_i - \sum_j^N w_j \mathbf{e}_j\|^2 \quad (30)$$

$$0 < w_i < 1, \text{ for } 0 < i < N$$

4.2.3 人脸动画合成

利用本文自动表情克隆方法获得的表情基模型，我们用表情基权值进行插值来合成人脸动画了。然而，直接使用表情基权值进行插值可能出现一些不自然的表情。文献[XTSYZ 2008]提出一种同时优化表情基和权值的方法。该方法基于的思想是：最小化运动数据重构误差的同时，保证优化表情基与初始表情基尽量接近。公式(31-33)给出了这一过程。

$$\min_w \sum_{j=1, \dots, M} \text{Error}_j + \sum_{i=1}^N \lambda_i \|\bar{\mathbf{e}}_i - \mathbf{e}_i\| \quad (31)$$

$$\text{Error}_j = \sum_{i=1}^N w_i^j \bar{\mathbf{e}}_i - (\text{trans}^{-1}(f_j) - f_1) \quad (32)$$

$$\min_{\bar{\mathbf{e}}_i} \sum_{i=1}^N \lambda_i \|\bar{\mathbf{e}}_i - \mathbf{e}_i\| \quad (33)$$

(31)式中的 trans^{-1} 表示去除绝对朝向的意思，(32)式中的 \mathbf{e}_i 为第 2 节计算得到的稀疏运动表情基 $(\mathbf{e}_1, \dots, \mathbf{e}_N)$ ， $\bar{\mathbf{e}}_i$ 为待优化的表情基。 λ_i 为手工给定的 balance 权值。

在(33)式的优化中， λ_i 取的值越大表示优化得到的 \bar{e}_i 越接近初始值 e_i 。 λ_i 取的值越小表示优化得到的 \bar{e}_i 对运动数据的重构误差越小。在实际应用中，这个参数可以调整来调节人脸动画的质量。这个优化问题采用迭代优化的策略来求解，即：

算法 5：表情基和权值同时优化过程

1. \bar{e}_i 初始值取为 e_i ， λ_i 取某个固定值。优化权值 w 。
2. 固定权值 w ，优化表情基 \bar{e}_i 。
3. 固定表情基 \bar{e}_i ，优化权值 w 。
4. 执行 2-3 步 20 次。

这里的表情基我们采用本文提出的自动表情克隆方法生成，用户只需要提供一个人脸模型，然后就可以使用摄像头来控制模型生成真实感人脸动画。

4.3 系统效果

本文中的视频驱动人脸动画系统使用 c/c++ 开发，运行的计算平台为 Intel®i7 2630QM@2.0GHz，内存 6GB，运行帧率为 15fps。图 4.6 展示了使用本文中的技术获得的视频驱动人脸动画序列。可以看出，本系统对于嘴部、眉毛等表情运动模拟的较好。由于本系统对眨眼动作跟踪效果不鲁棒，系统对眼部的动画生成质量还不理想。

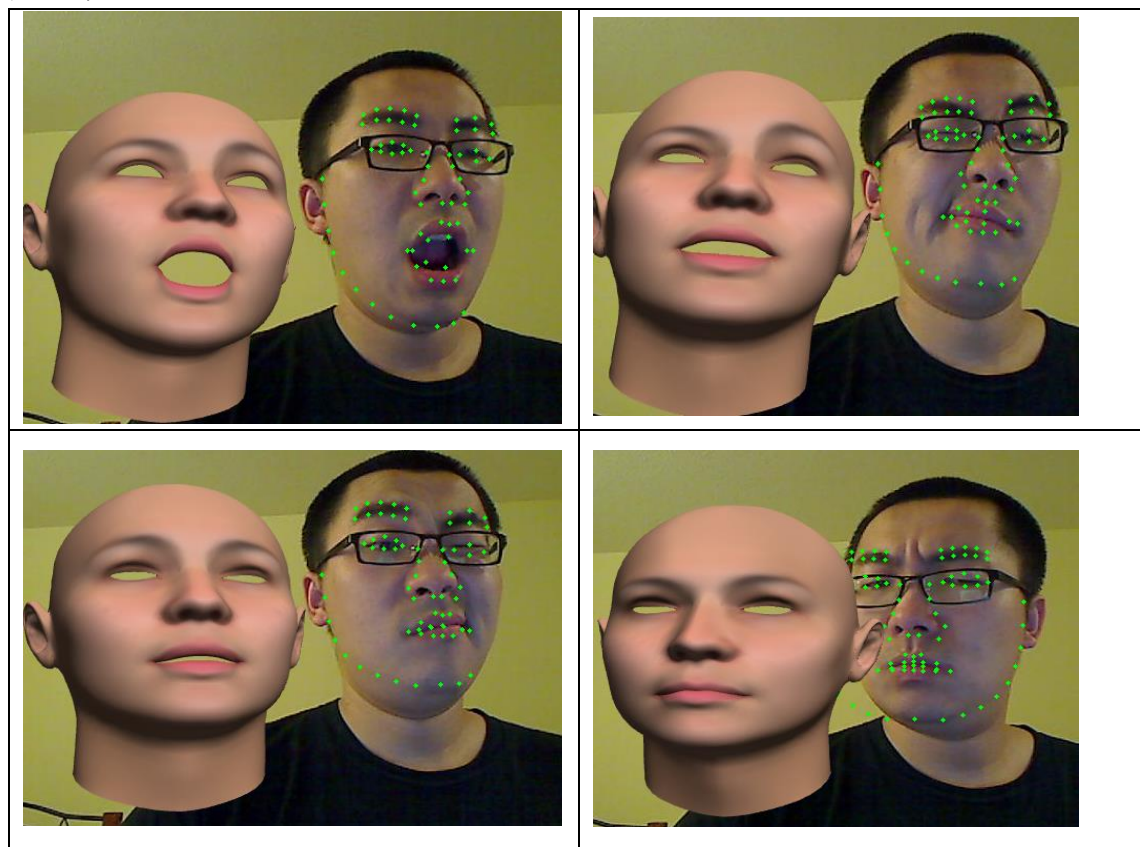


图 4.6 视频驱动人脸动画实验结果

4.4 小结

本章设计了一套基于 PC 平台的实时视频驱动人脸动画系统，包括人脸特征跟踪模块、运动参数传递模块和人脸动画生成模块。

本文提出的主要贡献有：

- (1) 提出一种 2D 特征到 3D 特征的计算方法，通过估计相机参数，还原出图像平面 2D 特征在 3D 空间中的位置。
- (2) 采用同时优化表情基和权值的实时优化算法，将运动数据的重构误差进一步减少，以获得高质量人脸动画。

5 总结和展望

5.1 全文工作总结

高真实感人脸动画被广泛应用于 3D 电影和游戏制作中，工业界中常用的技术大都依赖丰富的人脸表情模型库。而表情模型的制作需要美工长时间的调整，代价不菲。除了 3D 电影和游戏，移动终端游戏、视频通话、远程交互领域等也开始应用 3D 人脸动画技术来增加应用的趣味性和沉浸感。由于成本限制，它们难以使用运动捕获设备采集人脸运动数据，也难以根据自己的需要聘请美工制作丰富的表情模型库。为了解决这些问题，本论文设计了两套方法来简化人脸动画的生成过程：自动表情克隆方法和它的应用——视频驱动人脸动画。通过细致的研究，取得了三项研究成果：

(1) 提出了一种自动提取模型对应点的方法

表情克隆算法需要对源模型和目标模型进行对应关系配准，建立模型之间顶点的对应关系。而模型配准的初始阶段一般都需要用户给出一些基本特征点的对应关系，比如眼角、嘴角等特征点的对应。现有的方法基本都是采用手工选点的方法。为了简化用户交互过程，本文提出一种自动选取模型对应点的方法。

首先利用启发式方法选择一组模型对应特征点，用户再对源模型上手工选几个点后，就可以利用径向基函数插值获得目标模型上相应的对应点。使用该方法可以大大减少用户的手工交互操作。

(2) 提出了一种高效的模型配准优化方法

本文采用基于变形梯度的方法实现了表情克隆工作，由于空间三角形只有两条线性无关的边，不能获得变形梯度的准确值，该方法引入增广顶点来计算变形梯度。本文修改了变形梯度的矩阵表示，通过一个六自由度的矩阵近似原来八自由度的矩阵，从而使得其计算过程不需要增广顶点。这一修正将建立模型配准过程中的优化变量减少到原来的三分之一，很大程度上提高了优化计算的效率。实验表明，这种方法依然可以保持原算法的优化精度。

(3) 搭建了一套实时、低成本的视频驱动人脸动画合成系统

该系统使用普通的家用摄像头作为交互接口，用户只需要对摄像头做表情，便可以在 PC 上生成一个与自己有相似脸部表情和头部运动的虚拟人脸化身。该系统实现了以下功能模块：利用约束局部模型实现了实时人脸特征检测和姿态估计，可以较为准确的估计人脸表情变化和头部运动；通过估计相机参数，将 2D 的

表情运动数据变为 3D 的运动数据，同时建立用户脸部结构虚拟人脸化身模型结构的对应关系，将用户的脸部表情迁移到几何结构不同的虚拟人脸化身上面；人脸动画合成，使用带约束的表情基优化方法获得表情基合成参数，进而生成具有较高真实感的人脸动画。

5.2 下一步研究方向

本文提出的表情克隆并不是完美的解决方案，对于模型拓扑差异过大的情况不能解决的很好。我认为未来值得研究的方向有：

1. 继续探索对于模型拓扑差异过大的情况的解决方法；
2. 研究基于物理的模型变形迁移，这样可以大大丰富模型表情库。比如在进行表情克隆时，加入对顶点的力约束，可以获得面部肌肉颤动的效果。
3. 研究多媒介驱动生成人脸动画的框架，如同时采用摄像头、红外相机、深度相机、声音传感器等设备，这样的话可以获得更多相关信息，或许可以为研究人脸动画合成提供新的思路。

参考文献

- [BBAW 07] BICKEL B., BOTSCH M., ANGST R., MATUSIK W., OTADUY M., PFISTER H., GROSS M.: Multi-scale capture of facial geometry and motion. In ACM SIGGRAPH 2007 papers (2007), p. 33. 3
- [BBZ 03] B. Allen, B. Curless, and Zoran Popović. The space of human body shapes: reconstruction and parameterization from range scans. ACM SIGGRAPH 2003 Papers. Pages 587–594
- [BV 99] BLANZ, V., AND VETTER, T. 1999. A morphable model for the synthesis of 3d faces. In Proceedings of SIGGRAPH 99, ACM Press/Addison-Wesley Publishing Co., SIGGRAPH 99, 187 - 194.
- [CLS 01] CHOE, B., LEE, H., AND SEOK KO, H. 2001. Performance-driven muscle-based facial animation. The Journal of Visualization and Computer Animation 12, 67 - 79.
- [CYSK 13] C. Cao, Y. Weng, S. Lin, K. Zhou, "3D Shape Regression for Real-time Facial Animation", ACM Trans. on Graphics (SIGGRAPH), 2013.
- [DCFN 06] DENG, Z., CHIANG, P.-Y., FOX, P., AND NEUMANN, U. 2006. Animating blendshape faces by cross-mapping motion capture data. In Proceedings of I3D 2006, ACM, New York, NY, USA, 43 - 48.
- [DT 06] D Cristinacce, T. Cootes. Feature Detection and Tracking with Constrained Local Models. BMVC 2006
- [EC 02] E. Chuang and C. Bregler. Performance driven facial animation using blendshape interpolation. Computer Science Technical Report, Stanford University, 2002. CS-TR-2002-02.
- [EWP 28] E. Praun, W. Sweldens and P. Schröde. Consistent Mesh Parameterizations. Proceedings of the 28th annual conference on Computer graphics and interactive techniques. Pages 179–184
- [HJYC 13] H. Li, J. Yu, Y. Ye, C. Bregler. REALTIME FACIAL ANIMATION WITH ON-THE-FLY CORRECTIVES. ACM Transactions on Graphics, Proceedings of the 40th ACM SIGGRAPH Conference and Exhibition 2013, 07/2013 - SIGGRAPH 2013
- [HPS 06] HAVALDAR P. Sony pictures imageworks. In SIGGRAPH '06:

- ACM SIGGRAPH 2006 Courses (New York, NY, USA, 2006), ACM, p. 5.
- [HTWP 10] Hao Li, Thibaut Weise, Mark Pauly. Example-Based Facial Rigging. Proceedings of SIGGRAPH 2010.
- [HCTW 11] HUANG, H., CHAI, J., TONG, X., AND WU, H. 2011. Leveraging motion capture and 3d scanning for high-fidelity facial performance acquisition. ACM Trans. Graph. 30 (August), 74:1 - 74:10.
- [JU 01] J. Y. Noh and U. Neumann. Expression cloning. Proc. of ACM SIGGRAPH' 01, pages 277 - 288, 2001.
- [JTDP 03] JOSHI, P., TIEN, W. C., DESBRUN, M., AND PIGHIN, F. 2003. Learning controls for blend shape based realistic facial animation. In Proceedings of the SCA 2003, Eurographics Association, 187 - 192.
- [JJJ 03] Jin-xiang Chai, Jing Xiao and Jessica Hodgins. Vision-based Control of 3D Facial Animation. SIGGRAPH Symposium on Computer Animation 2003
- [LJP 10] Lewis, J.P. Direct Manipulation Blendshapes. Computer Graphics and Applications, IEEE. July-Aug. 2010. Volume: 30 , Issue: 4 pp: 42 - 5.
- [KWYM 10] K. Zhou, W. Xu, Y. Tong, M. Desbrun. Deformation Transfer to Multi-Component Objects. Computer Graphics Forum (Eurographics 2010).
- [KJXX 06] K. Zhou, J. Huang, J. Snyder, X. Liu, H. Bao, B. Guo, H. Shum. Large Mesh Deformation Using the Volumetric Graph Laplacian. ACM Transactions on Graphics (SIGGRAPH 2005), 496-503.
- [MB 99] M. Brand. Voice pupperty. Proc. of ACM SIGGRAPH' 99, pages 21 - 28, 1999.
- [MRR 06] Mario Botsch, Robert Sumner, Mark Pauly, Markus Gross. Deformation transfer for detail-preserving surface editing. Vision, Modeling & Visualization 2006, pp. 357-364
- [NRG 10] Nicolas Stoiber, Renaud Seguier, Gaspard Breton. Facial Animation Retargeting and Control based on a Human Appearance Space. Computer Animation and Virtual Worlds, vol. 21, pp. 39-54, 2010.
- [PKWS 03] PYUN, H., KIM, Y., CHAE, W., KANG, H. W., AND SHIN, S. Y. 2003. An example-based approach for facial expression cloning. In Proceedings of SCA 2003, Eurographics Association, 167 - 176.

- [PW 78] P. Ekman and W.V. Friesen. Facial Action Coding System. Consulting Psychologists Press, 1978.
- [RJ 04] R. W. Sumner and J. Popovi'c. Deformation transfer for triangle meshes. ACM Trans. Graph., 23(3):399 - 405, 2004.
- [SZPG 05] SUMNER R. W., ZWICKERM., GOTSMAN C., POPOVI' C J.: Mesh-based inverse kinematics. ACM Transactions on Graphics 24, 3 (2005), 488 - 495. (SIGGRAPH 2005 Conference Proceedings).
- [THL 09] Thibaut Weise, Hao Li, Luc Van Gool, Mark Pauly. FACE/OFF: LIVE FACIAL PUPPETRY. Proceedings of the Eighth ACM SIGGRAPH / Eurographics Symposium on Computer Animation 2009, 08/2009
- [TYJ 11] Taehyun Rhee, Youngkyoo Hwang, James Dokyoon Kim, Changyeong Kim. Real-time Facial Animation from Live Video Tracking. Eurographics/ ACM SIGGRAPH Symposium on Computer Animation (2011).
- [VT 98] V. Blanz and T. Vetter. A Morphable Model For The Synthesis Of 3D Faces. Proceedings of the 26th annual conference on Computer graphics and interactive techniques. Pages 187-194
- [WBY 08] Wei-Wen Feng, Byung-Uck Kim, Yizhou Yu. Real-Time Data Driven Deformation Using Kernel Canonical Correlation Analysis. ACM Transactions on Graphics (Special Issue for SIGGRAPH 2008).
- [WLVP 09] WEISE, T., LI, H., VAN GOOL, L., AND PAULY, M. 2009. Face/off: live facial puppetry. In Proceedings of SCA 2009, ACM, 7 - 16.
- [XTSYZ 2008] X. Liu, T. Mao, S. Xia, Y. Yong, and Z. Wang. Facial animation by optimized blendshapes from motion capture data. Computer Animation and Virtual Worlds 2008. Volume 19 Issue 3-4, September 2008 pp 235-245
- [XSYZ 2011] X. Liu, S. Xia, Y. Fan and Z. Wang. Exploring non-linear relationship of blendshape facial animation. Computer Graphics Forum, doi: 10.1111/j.1467-8659.2011.01852.x, 2011
- [YY 11] Y. Sahillioglu and Y. Yemez. Coarse-to-Fine Combinatorial Matching for Dense Isometric Shape Correspondence. Computer Graphics Forum (Proc. SGP), Vol. 30, No. 5, pp. 1461-1470, 2011
- [ZSS 04] ZHANG, L., SNAVELY, N., CURLESS, B., AND SEITZ, S. M. 2004. Spacetime faces: high resolution capture for

- modeling and animation. ACM Trans. Graph. 23 (August), 548 - 558.
- [ZSP 09] G. Zoric, K. Smid, I. S. Pandzic. Towards facial gestures generation by speech signal analysis using HUGE architecture. Multimodal Signals: Cognitive and Algorithmic Issues Lecture Notes in Computer Science, 2009, Volume 5398/2009, 112-120.
- [陈 03] 陈益强, 高文, 王兆其, 姜大龙, 基于机器学习的语音驱动人脸动画方法, 软件学报, 2003, 14(2): 215-221
- [范 10] 范懿文, 柳学成, 夏时洪. 人脸表情动画与语音的典型相关性分析, 第十六届计算机辅助设计与图形学学术会议, 277-285, 2010.

致 谢

合抱之木，生于毫末；九层之台，起于垒土；千里之行，始于足下。看着眼前这篇洋洋洒洒的数万字论文，研究生三年来的千种滋味，万般感受逐渐涌上心头。

在这里，我首先要感谢我硕士论文的指导教师—夏时洪研究员。早在2008年参加所里本科毕设答辩时，夏老师对我的提问和指导便使我人生中第一次体会到了学术研究的严谨认真精神。后来，经历了研一的迷茫和追寻，我在研二时加入了夏老师领导的研究组。开始进行人脸动画方面的项目和研究。夏老师对我学术的指导，主要体现在每周的例会上。在例会上，夏老师从我项目的细节到研究的问题、意义都给出了深入细致的指导，逐步拨开了我心中对研究存在的疑云，夏老师对我的鼓励也提高了我的信心。在我对自己研究方向踌躇不定时，夏老师支持我前往浙大和电子科大参与学术研讨活动，去上海参加学术会议，希望我能多多了解学术界的动向，开拓自己的眼界。并主动帮助我联系工业界的技术、资金支持，帮助我参与到移动增强现实等新的研究领域。后来由于我个人的原因，决定放弃读博、出国求学后，他仍然给与了鼓励和支持。我对夏老师的宽广胸怀非常钦佩。

我同样要感谢带领我进入实验室的指导老师—王兆其研究员。王老师对研究方向的大局把握和严谨认真的学术态度使我获益匪浅。王老师每次与我面谈时都会对我提出鼓励，并在我做的不好的地方给出深刻的建议。时至今日，每当我回想到他坚毅的眼神和亲切的笑容，心中便充满了动力。令我最感动的是王老师对我选择出国求学方面的支持，老师的恩情我会铭记在心。

感谢朱登明老师在我刚进所时给与的鼓励帮助，以及在戏曲动画项目上对我的指导。

感谢毛天露老师在上海世博会群体仿真项目上给我的指导和帮助，以及在项目报告申请和实验室学术报告上给我的建议。

感谢蒋浩老师对我项目上的指导，以及在我入党时给与的帮助。

感谢邱显杰老师、魏毅老师、李淳芄老师、刘卫玲老师、黄河老师对我在工作和生活上提供的帮助。

感谢李锦丹老师、刘玲老师、周世佳老师、宋守礼老师、李琳老师、冯刚老师、李丹老师、张平老师，在我顺利完成硕士阶段所给予的帮助。

感谢同窗三年共同拼搏的实验室兄弟姐妹们。王文中师兄作为我进实验室以来的大师兄与我在所里相邻而坐，给了我非常大的影响。王师兄常告诉我，做任何事一定要力求尽善尽美。他还经常在机器学习，机器视觉等理论知识方面给我细致的讲解。师兄对流浪猫的照顾更显示了他的博大爱心。感谢李挺、束博、李杨、刘米兰、周峰、宋海林、张桂连等师兄师姐给与我的帮助，我会永远记住1319的一年半生活。

感谢柳学成、张桂娟、洪义、宗丹、吴二元、马万里、赵建军、吕磊、石敏、刘华俊、陈明伟、叶青、黄武、任海滨等师兄师姐等的照顾和帮助。

感谢同届同学范懿文、费心宇、时进进、周果、刘宁三年来对我的帮助和照顾。

感谢王华同学对我平常的照顾帮助，她勤奋的工作精神和认真的研究态度值得我学习。感谢苏乐同学对我的帮助，他谦逊的品格和低调踏实的作风令人钦佩。

致 谢

感谢王涵同学对我在完成硕士论文中的帮助，她给我提出的意见让我完善了本文的写作格式。感谢王承博、路一峰等同学对我硕士论文完成提供的帮助。

感谢吕晓磊、沈亮、王从艺、王智勇、冯笑冰、邓一文等同学的帮助，Siggraph这种大山靠你们征服了。

感谢小师妹杨悦为实验室带来了快乐，希望你的未来之路一帆风顺。

最重要的是，我要感谢我的父母。父母的抚养之恩比山高，比海深。父母对我做人品格的教育令我敢于在波涛汹涌的社会大潮中乘风破浪、奋勇前进。谢谢父母全力支持我读研，让我没有后顾之忧。我一定会更加努力的工作，让父母退休之后的生活更上一层楼。

最后感谢我成长过程中伤害过我的人，你们让我看清自己，愈挫愈勇。

冀鼎皇

2013年5月

作者简介及在学期间发表的学术论文与研究成果

作者简介：（一般应包含教育经历和工作经历）

1986年3月29日出生于山西省晋中市榆次区。

2005年9月——2009年7月，在中国科技大学计算机学院（系）获得学士学位。

2009年9月——2013年5月，在中国科学院计算技术研究所攻读硕士学位。

已发表（或正式接受）的学术论文：（书写格式同参考文献）

申请或已获得的专利：（无专利时此项不必列出）

“一种人脸运动跟踪方法和系统以及一种增强现实方法”，夏时洪，冀鼎皇，魏毅。专利号：ZL 2011 1 0335178.8.

参加的研究项目及获奖情况：

1. 72xuan 虚拟家庭装修项目；
2. 上海世博会群体仿真项目；
3. 戏曲动画项目；
4. 基于移动终端的3D人脸化身。
5. 高通公司移动增强现实大赛三等奖