

上海交通大学

SHANGHAI JIAO TONG UNIVERSITY

学士学位论文

THESIS OF BACHELOR



论文题目：基于弹力模型的人脸表情建模研究

学生姓名：张亿皓

学生学号：5080309498

专 业：信息工程

指导教师：林巍骁

学院(系)：电子信息与电气工程学院

基于弹力模型的人脸表情建模研究

摘要

计算机人脸表情建模和动画制作 (Computer Facial Modeling and Animation) 是计算机图形学领域中一个非常重要的课题。该课题自上世纪七八十年代兴起, 随着近些年来动画电影、电脑游戏等娱乐产业以及医学辅助分析、心理学等学科不断发展, 得到了十分快速的进步。人脸表情变换 (Facial Expression Morphing) 作为该领域的一部分, 又称为表情映射或表情合成。该方法能根据一个人已有的表情, 通过映射的方法, 合成出另一个人对应于该表情的结果, 使其拥有与第一个人相似的表情。在本文中, 我们基于传统带纹理的二维表情变换方法, 设计了一个新的模型来增强这种映射的逼真性。这个模型具体包括两个子部分, 即特征点弹性模型和肌肉分布模型。特征点弹性模型主要根据了人的脸部具有弹性的特性, 将表情的整体变换结果和局部变换结果通过弹力结合起来, 在几何上达到更逼真的效果; 肌肉分布模型运用了人脸肌肉的分布和运动特性, 对表情的纹理细节进行了强化。在本文的后半部分, 我们进行了一系列实验, 并且提出了基于用户学习 (User study) 的主观评价标准和基于特征人脸 (eigenface) 的客观评价标准, 从多个角度证明了新的模型能确实有效地增强表情变换的逼真性。

关键词: 表情变换, 人脸动画, 计算机图形学, 弹性, 脸部肌肉, 表情比例图像

RESEARCH ON FACIAL EXPRESSION MODELING BASED ON ELASTISITY MODEL

ABSTRACT

Computer Facial Modeling and Animation (CFMA) is a very important issue in the field of Computer Graphics. It first shows up in the seventies and eighties of the last century. In recent years, much progress has been made in this field due to the fast developments in the industries of animation movies and computer games, and in the subjects of medical assisting analysis and psychology. Facial expression morphing, also called expression mapping or synthesis, is a very important part in CFMA. With this method, one person's unknown expression can be synthesized by the same expression of another person, which really exists. In this paper, we design a new morphing model to make this synthesized result more vivid based on the traditional 2D expression transformation method including texture details mapping. The new model contains two parts. They are feature-points-elasticity model and muscle-distribution-based model. The feature-points-elasticity model combines global warping and local warping results with the elastic force between the feature points on human face. The muscle-distribution-based model strengthens the texture details of an expression according to the distribution and motion of muscles on human face. Later, we did a series of experiments to evaluate the performance of the new model. Meanwhile, we use the method of 'user study' as a subjective criteria and the 'eigenface' as a objective criteria to prove that the new model can indeed make the expression transformation more vivid and convincing.

Key words: expression transformation, facial animation, Computer Graphics, elasticity, facial muscle, Expression Ratio Image

目 录

第一章 绪论.....	1
1.1 计算机人脸建模和动画概况.....	1
1.1.1 人脸建模和动画的应用领域.....	1
1.1.2 人脸建模和动画的研究难点.....	1
1.1.3 人脸建模和动画领域的历史发展.....	2
1.2 人脸表情变换概况介绍.....	3
1.3 人脸表情变换技术分类及研究现状.....	3
1.3.1 2D 表情变换技术及研究现状.....	4
1.3.2 3D 表情变换技术及研究现状.....	5
1.4 本文的研究内容和章节安排.....	7
第二章 带纹理的二维表情变换算法流程介绍.....	8
2.1 带纹理的二维表情变换算法的目标与框架.....	8
2.1.1 带纹理的二维表情变换算法目标.....	8
2.1.2 带纹理的二维表情变换算法框架.....	8
2.2 脸部特征点的选取.....	9
2.2.1 手动标定特征点的方法.....	9
2.2.2 自动标定特征点的方法.....	10
2.2.3 手动与自动标点方式的比较.....	10
2.3 三角网络剖分算法.....	11
2.4 仿射变换.....	12
2.4.1 仿射变换概述.....	12
2.4.2 齐次坐标.....	13
2.4.3 基本变换的变换矩阵.....	14
2.4.4 仿射变换的变换矩阵.....	15
2.5 离散点的插值方法.....	16
2.5.1 最近点插值.....	16
2.5.2 双线性插值 (Bilinear Interpolation).....	16
2.6 在变换中加入纹理细节——表情比例图像 (ERI).....	17
2.7 本章小结.....	19
第三章 基于弹性和肌肉分布模型表情变换算法的实现.....	20
3.1 带纹理的二维表情变换算法实现.....	20
3.1.1 标定人脸的特征点.....	21

3.1.2 德劳内 (Delaunay) 三角剖分的实现	21
3.1.3 根据“全局”变换确定 B1 的特征点位置	22
3.1.4 不同插值方法下仿射变换的实现	22
3.1.5 利用表情比例图像移植纹理细节的实现。	23
3.2 “局部”表情变换	25
3.3 脸部弹性模型	26
3.4 基于肌肉分布的纹理增强模型	28
3.4.1 人脸的肌肉分布模型	28
3.4.2 “纹理强化圆”的选取	29
3.4.3 利用“高斯模型”强化纹理细节	30
3.4.4 高斯模型增强纹理的过程示意	30
3.5 本章小结	31
第四章 实验结果比较与评价	33
4.1 基于“弹性和肌肉分布模型”的实验结果及比较	33
4.1.1 基于人脸弹性模型的实验结果	33
4.1.2 基于肌肉分布的纹理增强模型实验结果	34
4.2 利用弹性和肌肉分布模型实现二维表情变换的其他示例	35
4.3 表情变换的评价标准	38
4.3.1 主观评价标准	38
4.3.2 客观评价标准	39
4.4 表情变换的应用实例	40
4.5 本章小结	41
第五章 总结与展望	42
5.1 论文总结	42
5.2 工作展望	42
参考文献	44
谢辞	45

第一章 绪论

1.1 计算机人脸建模和动画概况

在电子计算机问世以前，人们处理图像的方式仅仅停留在手工绘画或光学摄影的阶段，方法十分单一。计算机被发明以后，处理图像的方式得到了极大地丰富，数字图像处理（Digital Image Processing）这一学科也从那时开始蓬勃发展起来。起初，人们发现在处理某些数字图像的时候，需要用到图形和几何学的知识，因此图像的几何学处理方式慢慢出现了萌芽，并作为数字图像处理的一个辅助部分。渐渐地，随着图形学领域被挖得越来越深，它逐渐从数字图像处理中分离出来，形成了一门庞大而又复杂的独立学科——计算机图形学（Computer Graphics）。计算机人脸建模和动画（Computer Facial Modeling and Animation）正是其中非常重要的一个课题，并在近几年得到了飞速的发展。

计算机人脸建模和动画（以下简称“人脸建模和动画”）起源于上世纪七十年代，但真正取得大量成果和进步是在八十年代中后期。这个学科主要致力于建立人的头部或脸部模型，或者将其制作成动画，期望能够利用计算机虚拟技术尽可能逼真地还原现实的人脸姿态，以达到“虚拟现实”的效果。由于这项技术的特殊性，它还与其他一些领域有着密不可分的关系，例如医学和心理学。早在上世纪六十年代就有心理学家经研究提出^[1]，人类在表达自身情感时，文字内容仅仅起了7%的作用，利用话语表达起了39%的作用，而另外54%是靠人的表情和动作来表达的。可见，人脸的表情建模是多么重要！精准的建模甚至可以使人相信这是一个真实的世界，人们表达自己行为的方式也越来越多样化。因此，有越来越多的学者投身入这个行业，在近几年快速推动了这个行业的发展。

1.1.1 人脸建模和动画的应用领域

人脸建模和动画制作可以利用虚拟技术逼真地再现一个真实的世界，因此它的应用领域是相当广泛的。动画电影和计算机游戏（包括视频游戏）产业是这个领域最主要的推动力量。进入九十年代后，为了满足新的感官需要，人脸建模和动画制作技术被投入到了三维动画电影的生产中去，在早期产生了一系列优秀的作品，例如玩具总动员、怪物史莱克等。同时，游戏业的快速发展也起到了巨大的作用。现如今的游戏大多是3D游戏，其中涉及了大量人脸建模和动画技术，并且近两年微软公司新推出的Kinect体感摄像头更是将人脸建模技术推向了一个新的高度。同时，这种技术也能用于教育、通信领域，例如虚拟教师，视频会议等，可以用虚拟的人物代替真实的人进行交流。另外，人脸建模和动画也能很好地融入到医学研究领域中去，例如通过对病人头部模型和表情的分析，为其提供一种辅助治疗的手段，甚至可以将此方法用到法庭的辅助辨认当中去。总之，人脸建模和动画涉及到了生活的方方面面，在很大程度上改变着人们的生活方式。

1.1.2 人脸建模和动画的研究难点

虽然人脸建模和动画能够广泛地运用到各个领域中去，但是目前却还没有大量成熟的产品投入使用，最主要的原因就是由于人脸特殊的复杂性导致的。人的脸或头部很难用一般的几何形状通过数学方法描述出来，同时人脸上分布的大量肌肉组织更是带来了许多难以预测的微小变化。总结起来，人脸建模和动画的研究难点主要集中在以下三个方面：

(1) 不同人的脸型、五官特征、肤质和颜色是千差万别的，没有两个人的脸能够一模一样，这导致了很难用一个统一的模型将人脸表示出来。为了保证建模的精准，往往会包含大量的参数来表现这些特征，因此给计算带来了很大的障碍。

(2) 人的脸部并不是一个刚体，人脸皮肤下分布了大量的肌肉，由于这些肌肉不同的运动方式，可以使人脸在不同表情下呈现不同的几何形状，并且如皱纹等的纹理细节也会随之变化。

(3) 同一个人不同年龄段的脸部外观都是不一样的，脸的轮廓、五官的位置、形状，还有特别是皮肤的质地、纹理细节等特征，都会随着年龄的增长发生很大的变化。同一个人幼年期和老年期的外貌特点的差别是非常明显的。所以“年龄估计”这一课题在世界范围内都是一个难题。目前 UIUC 的 T. Wang 带领的小组在这方面取得了不错的成果。

1.1.3 人脸建模和动画领域的历史发展

对于脸部动作和表情的研究，最早可追溯到十七世纪，当时是从生物学的角度出发研究的。后来，随着计算机的发明诞生，人脸建模技术才有了较快的发展。在早期，有相当一部分科学家为脸建模和动画的发展奠定了坚实的基础，后来的新技术都是在这些基础上发展起来的。下面将简单介绍人脸建模和动画制作中一些经典的技术手段。

Parke^[2]无疑是人脸建模领域的先驱之一，他早在 1972 年就提出了一个非常强大的 3D 头部模型。Parke 提出的这项技术实际上是将人脸模型进行参数化表示，它能根据人脸主要的特征点信息，利用有限的参数表示出人的头部模型。这个模型是由一系列的 3D 点和这些 3D 点形成的网络所构成的。这些 3D 点由一组构型参数和一组表情参数控制，改变 3D 点的位置就可以改变人脸的模型。其中，构型参数可以控制人脸某些特征点的位置，例如眼睛、鼻子和嘴的相对位置。改变这些参数的值就可以形成一个新的基本头部模型；表情参数可以进一步控制脸部的动作，例如拉伸嘴唇或闭合眼睛。利用这种方法，可以得到同一个人甚至不同的人在各种表情下的脸部模型，进而产生一系列的脸部动画。这种做法的好处是将复杂的人脸模型参数化了，为进一步的计算带来了方便。

另一个具有里程碑意义的技术是由 Waters^[3]在八十年代提出的，这是一个基于人脸肌肉构造的新模型，依据了物理和生物方面的理论。与参数化的人脸模型不同的是，Waters 将脸部的运动分解为各部分不同肌肉的运动，通过这些肌肉运动的不同组合，可以变化出各种各样的脸部拓扑结构。目前，大多数基于物理的模型都参考了 Waters 提出的这一理论。

还有一大类方法是由 Paul Ekman 和 Wallace V. Friesen^[4]等人在 1978 年提出的面部行为编码系统 (Facial Action Coding System, FACS)。Ekman 等人经过长期的研究，发现可以根据不同肌肉的运动，将脸部运动分解成 44 种最基本的独立行为单元 (Action Unit, AU)，每一个 AU 都是由单独一块肌肉的运动形成的，是最基本的表情。这些 AU 就像是人脸模型空间的 44 个基底一样，几乎所有常见的表情都可以由这些 AU 按照不同比例混合而成，因此称为面部行为的编码系统。文章[5]就根据 FACS 系统，给出了一个 3D 人脸建模的实例。文章[5]中，作者进一步延伸提出了“融合形状 (Blend-shape)”方法，将 FACS 中的每个 AU 具体到单个用户上去，并通过给这些 AU 设置不同的权重，融合出该用户的任意一种表情模型。图 1-1 表示了这种形状融合的结果，其中左侧是具体到用户的 AU，右侧是根据各个 AU 的权重融合得到的人脸模型。

此外，还有其他许多非常经典且有效的方法，例如自由形式变形 (Free-form Deformation)，等等。以上谈到的这些方法，都是人脸建模和动画领域中十分基础的方法，是整个学科的奠基石，对今后人脸建模和动画制作的发展起到了十分巨大的推动作用。后人在总结前人方法的基础上，不断推陈出新，将这些经典方法扩展成为更加强大的建模手段，大大提高了这种虚拟技术的逼真性。

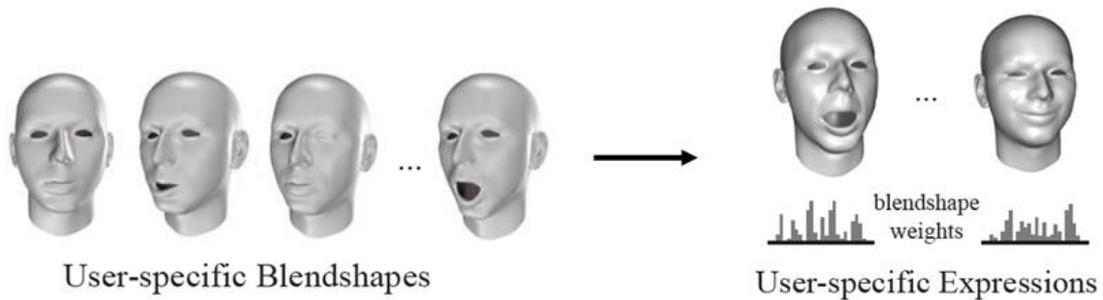


图 1-1 利用 FACS 的 AU 融合构建人脸模型的实例

1.2 人脸表情变换概况介绍

人脸表情变换（Facial expression transformation）是计算机人脸建模和动画的一个重要组成部分。人脸表情变换，又称为表情映射或表情合成，其基本的意图是将一个人的人脸变换为另一种形状，或将一种表情变换为另一种表情。表情变换的方法通常基于人脸的网格化或特征点的三角网络化来实现。

人脸表情变换技术被广泛运用在科学分析、图像合成中（例如 Photoshop），在动画电影、电脑游戏等娱乐产业中也占据了十分重要的地位，为虚拟现实的逼真化提供了强有力的技术支持。图 1-2 展示了一个十分有趣的人脸表情变换实例。



图 1-2 一个有趣的人脸表情变换实例

图 1-2 的左边一幅图像是前美国总统小布什，右边的一幅图像是电影明星施瓦辛格。我们期望能够通过图像变形（Image Morphing）手段，得到这两张人脸的“中间脸”。这张“中间脸”并不是左右两张图像的简单叠加，而是要满足轮廓及五官形状相同，没有重影的叠加结果。图 1-2 中间的一幅图像就是利用这种技术变换得到的“中间脸”结果。具体的方法是，首先通过插值，确定这张“中间脸”的几何形状，再分别将左右两张人脸以“中间脸”的形状为目标进行变形，之后将两个变形结果叠加起来，最终就得到了中间那个和小布什以及施瓦辛格“都有点像”的结果了。

1.3 人脸表情变换技术分类及研究现状

经过多年的发展，人脸表情变换的技术已经变得十分多样化了。这些技术所属的分派门类类别，各有不同。如果从表情变换是基于图像进行的还是基于三维模型进行的这一方面来看，主要可以分为 2D 表情变换技术和 3D 表情变换技术这两大领域。

1.3.1 2D 表情变换技术及研究现状

由于单幅图片没法呈现某个物体完整的三维信息,因此以单幅图片为基础建立人脸的三维模型进行处理是十分困难的。所以,基于图像进行人脸变换大多采用了 2D 表情变换技术。2D 表情变换技术的优点是不用建立复杂的三维人脸模型,而是仅在二维上处理人脸的几何变形。由于仅涉及一个平面的信息,标定的特征点以及要处理的参数远少于三维模型,因此此技术处理速度较快,复杂度不高。不过,其缺点是显而易见的:由于二维处理的方法缺少了深度等三维信息,因此人脸上的纹理细节的变化(例如皱纹,肤色的变化)往往得不到很好的体现,而且处理精度也不及三维模型来得高。运用 2D 变换技术可能导致一些表情变换结果显得比较呆滞,不自然。

对 2D 图形变换方法起到开创性推动作用的人是 T. Beier 和 S. Neely^[6]等人。他们在 1992 年首先提出了“基于特征的图像变换”这一技术。这种图像变形的技术在当时十分有用,它能准确快速地将一幅图像变形为规定好的另一种形状。Beier 等人提出的这一方法为之后二维图像更先进、更快速的变形技术做出了极大的贡献,目前很多变形方法都基于此。其基本的原理可以用图 1-3 所示的例子来说明。首先,这种方法需要对一张要变形的图片定义一个网格,就像图 1-3 (a) 中的字母“F”一样。网格的疏密程度可以根据实际的需要来定,网格划分地越密集,变形的精度就越高,相应的计算量也就更大。除了要在图像中定义网格之外,还要定义一些“特征线”,就像图 1-3 中带箭头的线段表示的那样。在 (a) (b) 两幅图中,对应的特征线称为一个“线对”。我们可以把这些线对想象成是整个网格的“手柄”,移动或转动这些“手柄”,网格的形状就会发生变化。再将这种变化作用到每一个像素之后,就实现了整张图像的变形。线对的数量也可以根据需求来确定。线对越多,可以做出的变化越丰富,复杂度也越高。图 1-3 定义了两个线对,并且显示了变动字母“F”上沿的线对后图像产生的变形效果。运用此种方法,很容易就可以实现图 1-2 展现的人脸融合结果了。

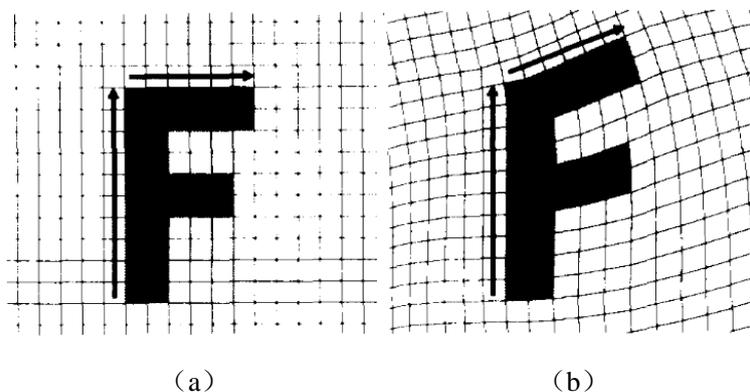


图 1-3 Beier 等人提出的图像变换算法实例^[6]

在 Beier 等人提出这种基于网格和特征线的变形方法后,很多人对此作了优化和推广。其中,最重要的一个推广是用三角网络代替网格,用三角形的各个顶点(称为特征点)代替特征线。这样,控制这些特征点的位置就可以任意地改变三角网络的几何结构,并通过对应三角形的仿射变换实现图像的变形(仿射变换的具体算法流程将在本文第二章中作详细介绍)。利用三角网络进行图像变形的好处是可以得到更加精准、并且更加随意的图形变换效果。

基于网格或三角网络的图像变换方式,很多人做了大量探索。Takeo Igarashi 等人提出了一种更加自由方便的变换方法,称为“尽可能刚性的形状操纵”(As-rigid-as-possible shape manipulation)^[7]。这种方法的基础仍然是基于三角网络的变形方式,但是操纵起来更加随心所欲,而且能最大程度地维持物体的本来面目,即尽可能维持物体的刚性,使其在变形过

程中不至于因为失真而变得扭曲。这种新的模型变形技术的主要理论依据是文章[8]提出的“尽可能刚性的形状内插”方法。根据这一理论,要想在变换某一物体的形状的过程中尽可能地维持其刚性,只要保证能够最大限度地维持三角网络中每个三角形的刚性就行了。图1-4就是文章[7]运用这种方法进行图像变形的一个实例。首先,需要进行变形的物体(即图中的“小人”)被分割成了一个三角网络。然后,在这个三角网络的各个顶点中,取出若干个点作为操纵变形的“手柄”,即图1-4第三幅图像中的三个空心圆点。通过改变这些“手柄”的位置就可以快速地改变“小人”的形状。当然,如果仅仅改变“手柄”的位置,但是其他三角顶点的位置维持不变的话,图像就一定会出现严重的失真。所以,作者利用刚性内插法对此作了优化。在这个过程中,三个“手柄”的位置是主动改变的,其他顶点的位置会跟着它们被动地改变,最后达到的效果是要尽可能地维持每一个三角形的刚性不变,或者说所有三角形的形状变化之和最小。这样,就出现了图1-4最后一幅图像的变形结果,在几何形状发生变化的基础上,基本维持了图像的本来面貌。

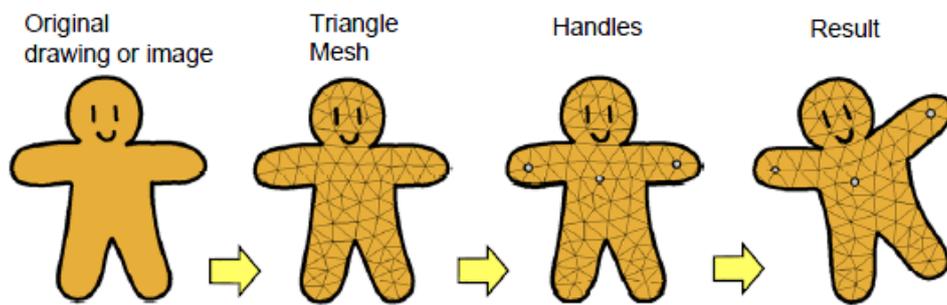


图 1-4 As-rigid-as-possible shape manipulation 变换实例^[7]

上面介绍的这些图像变形技术,主要都是根据人们的意愿对物体进行变形的,目标形状不需要一个明确的依据。除了这种方式外,2D 人脸表情变换还有另一种比较有趣的应用,称为表情移植。顾名思义,表情移植就是将一个人的某种表情移植到另一个人的人脸上去,主要还是通过特征点标定以及图像变形技术实现的。不过,仅仅运用传统的几何变形方法很难得到逼真的效果,因为这种方法虽然可以改变一个人脸部及各器官的几何形状,但是脸部的纹理细节,例如皱纹等因素,就很难得到体现了。这是 2D 表情变换的又一大劣势。不过,在这个领域,微软研究院的 Z. Liu 等人^{[9][10]}做的工作相当出色,他们提出了“表情比例图像(Expression Ratio Image)”这一概念。通过“表情比例图像”,人们在进行 2D 表情移植时,能够将人脸的纹理、光照信息一同移植到目标人脸上去,实现不亚于 3D 移植的效果。这种方法是本文主要工作的基础,详细算法将在第二章阐述。

1.3.2 3D 表情变换技术及研究现状

3D 表情变换技术是人脸变换领域的又一大技术门类。它通常需要建立人脸的三维模型,并结合几何变换、纹理贴图等渲染方式实现 3D 人脸的变换。相比于 2D 人脸变换技术,建立 3D 模型往往能够在几何上更加精确,而且通过纹理贴图的方式,2D 中容易丢失的纹理和光照细节也能得以保留。同时,2D 图像所不具有的深度信息,给 3D 模型提供了更高维度的处理信息。不过,3D 人脸建模的不足之处是它常常需要利用一些昂贵的硬件设备来采集 3D 信息,例如激光扫描仪,或光学、电子传感器等,并且三维模型拥有的大量参数处理起来也比二维图像复杂得多。然而,随着近年来硬件设备发展越来越迅速,这类问题逐渐在弱化。所以目前 3D 变换技术占据着主流的位置,应用十分广泛。

在介绍主要的 3D 表情变换技术之前,我们将首先介绍一种介于 2D 和 3D 之间的变换技术,称为“视角变换”(View Morphing)^[11]。之所以称它介于 2D 和 3D 之间,是因为这种技术仍旧只利用了普通的二维照相机,但是与传统方法不同的是它需要从多个视角(至少

两个视角) 拍摄同一个人的头部。实际上这就是一种变相的建立三维模型的方法, 很多学者都对此有所研究。在这种方法下, 同一个人脸部的多个视角的图像被采集到, 并通过插值的方式得到介于这些视角之间的人脸模型, 从而产生一系列动画。图 1-5 是文章[11]中演示的利用这种变换方法矫正画像《蒙娜丽莎》的一个例子。

在 3D 图像变换领域内, R. W. Sumner 等人同样利用了三角网络实现了三维物体之间动作或表情的移植。^[12]图 1-6 展示了他们利用三角网络变换将一只猫的姿态移植到一只狗的三维模型上去的过程(图 1-6 上方是猫的模型, 下方是狗的模型)。首先, 狗的三维模型被剖分成一个三角网络, 接着根据猫的模型, 利用变换矩阵对狗模型中的每一个三角形进行变换, 最后利用优化方法将得到的零散的三角形拼接成完整的模型。将这种方法作用于三维人脸, 就能实现 3D 表情变换。



图 1-5 利用“视角变换”技术矫正《蒙娜丽莎》示例^[11]

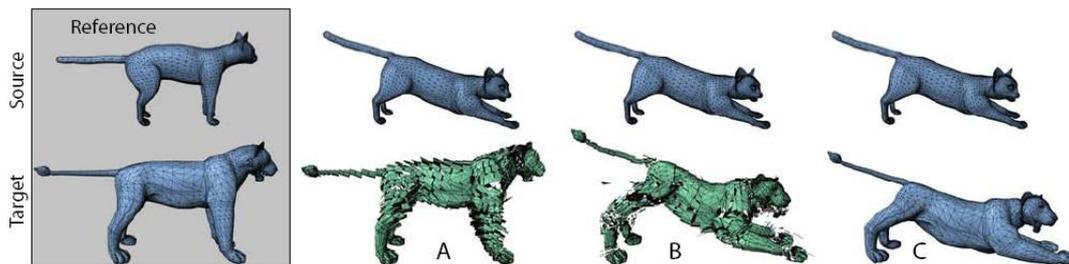


图 1-6 利用三角网络对三维模型的形态进行移植的示例^[12]

另一种实现 3D 表情变换的方法是文章[10]提出的, 这个工作是接着文章[9]的内容做的进一步深化, 也是本文主要要对比的方法之一。文章[10]提出的方法可以将 2D 变换和 3D 变换融合在一起。其主要的贡献是能将 2D 的纹理信息转化成 3D 模型, 这样 2D 图像变形过程中容易丢失的纹理和光照细节也能得以保留。图 1-7 表示了这种算法。在该算法中, 由于人皱眉产生的皱纹被建立一个三维深度模型, 建模的主要依据是皱纹部分像素的灰度值: 在皱纹区域, 颜色越亮, 代表三维模型的高度越高; 颜色越暗, 则表明三维模型的高度越低。采用这种方法来进行人脸建模和表情变换, 效果更加逼真。

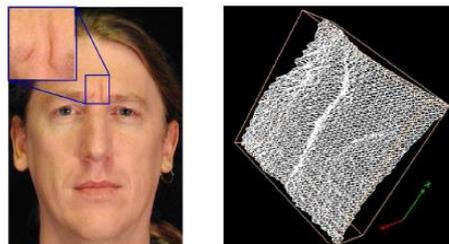


图 1-7 将二维纹理信息转化为三维深度模型^[10]

当然，上述提及的 3D 人脸表情变换算法只是冰山一角，大量学者都在做着此方面的研究，每年都会有各种各样不同的新算法、新模型出现，一一道尽并不现实。不过，有一点仍然需要强调，那就是近年来的 3D 建模技术的飞速发展离不开三维硬件采集设备的支持。值得一提的是，前两年由微软公司研制开发的 Kinect 体感摄像头，更是给 3D 建模领域带来了一次技术上的革命。Kinect 体感摄像头是世界上第一个能大规模投入商用的三维信息采集设备，它能在低价格的基础上保持高素质的三维信息采集效果。Kinect 摄像头不仅能够提取物体的深度图、区分前景和背景、提取人体模型，还能够自动识别并跟踪人体骨骼，功能十分强大，被广泛地运用于游戏娱乐行业和学术研究领域。上文提到的文章[5]就利用了 Kinect 摄像头提取人脸的深度图，并根据深度图建出人脸的三维模型。最后，文章[5]实现了一种十分有趣的人机交互方式：人们只要在摄像头前作出一些表情，就可以实时地控制计算机中的动画模型做出相同的表情，效果十分逼真。

1.4 本文的研究内容和章节安排

本文研究的重点是要实现 1.3.1 节中提及的二维人脸图像的表情移植算法，并且需要把表情中纹理细节的变化也一同包含到移植结果中去，使结果更加逼真。所以，文章[9]提出的“表情比例图像”是本文研究的基础。在此之上，本文还提出了一个新的模型——**二维表情变换的弹力和肌肉分布模型**，用来增强传统二维表情移植的逼真性，获得在主观上和客观上都更令人满意的结果。这个新模型的创新点和贡献主要包括：

(1) 提出了人脸的弹性拉力模型。传统的基于几何的人脸变换算法，往往只考虑了人脸特征点的全局变化，而忽略了针对不同用户的脸型和五官差异导致的局部变化。因此，本文引入了局部变换的概念，用于具体到每一个用户不同的脸型。最后，本文假设脸部相邻特征点之间存在拉力关系，并通过这种弹性关系将全局变换结果和局部变换结果结合起来，取得更令人信服的结果。

(2) 提出了基于脸部肌肉分布的纹理增强模型。由于滤波或其他特殊原因，表情的纹理细节部分（如皱纹）在移植的过程中可能会出现淡化或模糊的情况。因此，本文基于经典的脸部肌肉分布模型^[3]，通过观察肌肉运动对纹理细节产生的影响，提出了一套纹理细节的增强算法，用于抵抗移植过程中可能出现的纹理淡化或模糊现象。

本文通过大量的实验，并根据主观和客观两种评价标准，分析比较了各种不同表情移植算法和本文提出的新方法之间的优劣。

本文的后续章节，将做如下安排：

第二章，将首先详细阐述传统的带纹理细节的二维表情变换算法的一般流程。之后，会对算法中的各个步骤做一个详细的介绍，包括脸部特征点选取方法，三角剖分算法，仿射变换及插值算法，以及表情比例图像的详细介绍。这些内容，都将作为后续工作的基础。

第三章，将详细阐述二维表情变换的具体实现过程、可能遇到的问题以及解决的方法。之后，会着重阐述本文提出的新模型，即弹力和肌肉分布模型，包括它的原理以及实现的方案，并在最后给出初步的结果。

第四章，会通过一些实验，证实我们提出的新模型的有效性。并且，通过定义基于“用户学习”的主观标准和基于“特征人脸”的客观标准两种评价方式，将本文提出的新模型与其他二维表情变换算法进行比较。最后，通过一些实例，展示了这种表情变换技术在电影、视频制作等领域的实际意义和作用。

第五章，将对本文提出的表情变换模型做一个总结，并对未来的工作进行展望。

第二章 带纹理的二维表情变换算法流程介绍

在本章中我们将详细介绍传统的带纹理细节的二维表情变换算法的具体流程,以及流程中各步骤的详细算法和实现方式。具体地,这个算法又可以分为两个主要部分:(1)基于几何变形的二维表情变换算法;(2)基于“表情比例图像”^[9]的纹理改进算法。其中,基于几何变形的表情变换算法仅仅根据人脸特征点的位置信息,通过几何变形的方式对表情进行变换。这种方法忽略了表情的纹理细节,结果很难令人满意。而基于“表情比例图像”的变换算法则在几何变形的基础上,通过计算“模板表情”的表情比例图像,将纹理细节加到最终的结果上去,从而产生如皱纹、酒窝这样的光照、纹理特征,使变换结果更加真实可信。

本章内容安排如下:2.1节介绍本算法的总体框架和要实现的目标;2.2~2.5节介绍基于几何的表情变换方法,包括人脸特征点的选取、三角网络剖分算法、仿射变换、插值算法;2.6节将详细介绍文章[9]提出的“表情比例图像”(Expression Ratio Image, ERI),讲述如何将纹理细节加到变换结果上去,此算法是本文的一个重要基础。最后,将在2.7节对本章做个小结。

2.1 带纹理的二维表情变换算法的目标与框架

二维表情变换算法起源较早,传统的算法大多基于网格或三角网络的变形,仅仅在几何形状上进行了改变。为了弥补这种方式的不足,Z. Liu提出了“表情比例图像”^[9]这一概念来将纹理细节一同加入到变换结果中去,实际结果较为出色。

2.1.1 带纹理的二维表情变换算法目标

该算法的主要目的,是要逼真地实现“表情移植”这一过程。表情移植的概念曾在1.3.1节中简要介绍过,是为了将某一个人的一种表情(如微笑)移植到另一个人脸上去,并且使结果尽量逼真。例如,如果要将人物A的微笑的表情移植到人物B的脸上去,可采取如下步骤:首先,给定人物A的中性表情 A_0 和微笑表情 A_1 ,以及人物B的中性表情 B_0 这三张图像。这里,中性表情是指一个人在严肃的状态下,脸部没有呈现任何有特殊含义的表情,这种表情就称为中性表情(即可理解为没有任何表情)。其次,分析和计算 A_0 变到 A_1 的过程中出现的变化,例如脸部特征点的位移和纹理的深浅变化,将这些变化以同样的方式加到表情 B_0 上去,最后合成出人物B微笑的表情 B_1 。由于人物A的表情相当于一个样板,B的表情根据这个样板发生变化,因此合成出来的B的表情应当与A是相似的。我们称 A_1 为模板表情,称 B_1 为移植表情或合成表情。

2.1.2 带纹理的二维表情变换算法框架

该算法的框架示意图如图2-1所示。

此流程框图分为上下两个部分。上半部分主要实现了基于几何变形的表情变换算法。通过该步骤,可以合成出我们期望得到的移植表情 B_1 的结果,但这个结果只是在几何形状上发生了变化,纹理部分却没能得到体现。下半部分主要利用了“表情比例图像”(ERI)将纹理加入到结果中去。以下将详细阐述此流程中的关键步骤和算法。

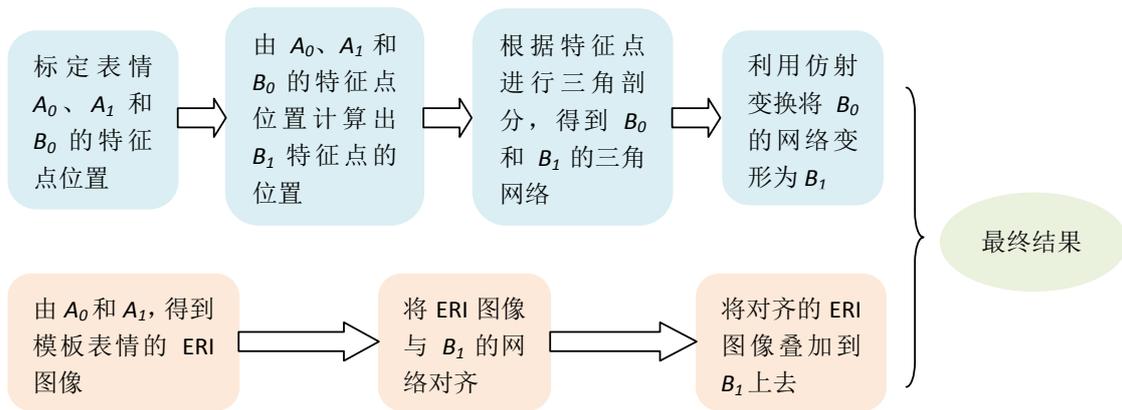


图 2-1 带纹理的二维表情变换算法框图

2.2 脸部特征点的选取

在本算法的第一步中, 我们需要标定人脸图像的特征点位置。这些特征点的定义并不是唯一的, 但是必须能够表示脸部某些关键位置的信息, 例如眼角、嘴角、眉毛、眼睛或嘴唇的轮廓, 以及鼻翼两侧等等。我们知道, 人脸部的构造是极其复杂的, 除了表层的皮肤以外, 皮下还有各种肌肉、骨骼等其他组织分布着, 对人体的头部进行精确的建模描述确实有困难。尽管如此, 宏观地来看, 人的脸部无非是由眼、耳、鼻、口、眉等五官组成, 掌握了这些关键部位的信息之后, 就能对人脸的几何结构有个大致的描述。提取脸部特征点的意义在于我们可以将复杂的脸部模型简化成一系列极其有限的参数, 通过调整这些参数来控制脸部的外观。这样就可以给计算机的数字化处理带来极大的便利, 处理速度也会提升。需要注意的是, 普通的基于几何的图像变换是不需要如此详细地标定这些特征点的, 而是只要任意地、均匀地选取一些点作为剖分三角网络的依据。但是人脸结构的特殊性要求我们进行特征点的提取, 以使得结果更加自然、更为精确。

标定人脸特征点的位置, 大致可以分为手动标定法和自动标定法。两种标定方式各有优缺点, 下面将逐一介绍。

2.2.1 手动标定特征点的方法

手动标定特征点的方法比较简单, 主要是通过人的直接观察, 标定出脸部图像具有特征性的关键位置, 如之前所说的眼角、嘴角等特殊位置。一般来说, 虽然是手动标定, 但是标定的点数不能过少, 要基本能够将人脸的几何信息完整地表示出来, 而且标定的点数越多, 变换结果越精确。通常, 手动标定的点需要包括脸部以及各五官的轮廓, 并在此基础上适当增加其他一些特征点。手动标点的方式并不唯一, 如图 2-2 就是一种可行的手动标点方式。

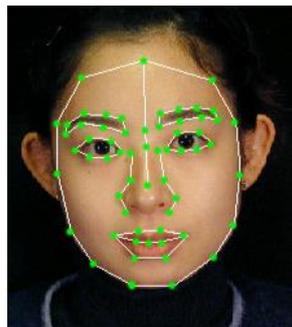


图 2-2 脸部特征点手动标定方法 (例子)

在图 2-2 中，特征点用绿色圆点表示，共 61 个点。这些点可以涵盖脸部基本的几何信息。白色的连线只是为了表示方便，在标点时并不是必要的。

2.2.2 自动标定特征点的方法

标定特征点的另一大类方法就是根据脸部识别进行自动标定。自动标定并不是凭空产生的，而是要建立在大量的训练数据上的。这些训练数据大部分同样需要手动标定，但是这些手动标定的过程只不过属于前期“准备”工作。计算机通过这些数据得到训练，进而对未标定的人脸进行识别，自动标出特征点。

有一种著名的算法可以实现这样的自动标点过程。这种方法称为“主动表观模型”（Active Appearance Model, AAM）^{[13][14]}。该算法是由 Cootes 等人在 1998 年提出的，基于先前的“主动形状模型”（Active Shape Model, ASM）算法，利用了人脸的几何信息和纹理、光照信息，通过著名的主成分分析（Principal Component Analysis, PCA）方法，建立出人脸的特征模型。该模型由于混合了人脸上的各种信息（不仅仅是几何特征），因此被称为人脸混合模型，是 AAM 模型的核心内容。利用这个模型，人们不光可以自动地对人脸的特征点进行标定，还能将其应用于人脸识别、表情识别、人脸建模等各种领域，为研究人脸提供了极其强大的技术支持。

AAM 模型的主要实现方法依然离不开训练数据的支持，是一种典型的机器学习（Machine Learning）的示例。AAM 将需要进行标定的人脸模型与已经存在的那些先验模型进行比较，匹配出最适合的模型并不断进行调整，最后将待标定的人脸特征点显示出来。AAM 算法的主要过程可分为建模和匹配两个部分：

（1）AAM 建模：AAM 算法所建立的模型被称为对象的主动表观模型。这个模型的基础是 Cootes 在之前提出的 ASM 形状模型。巧合的是，这种模型的进步与本文要讨论的表情变换算法是一致的：ASM 只考虑了人脸的几何特征，结果不够准确；而 AAM 在形状模型的基础上结合了重要的纹理信息，建立了更完整的人脸模型，提升了算法的精度。

（2）AAM 匹配：在 AAM 匹配计算中，首先，采用 PCA 方法来描述人脸特征点的动态变化。其次，根据 AAM 的训练模型和输入的测试模型的均方差来定义一个能量函数，用来评价 AAM 的匹配程度。在人脸定位匹配的过程中，根据模型的线性表达式并通过有效的匹配算法变化模型参数组，从而控制形状特征点的位置变化，生成当前新的 AAM 模型实例；再次，利用当前得到的能量函数值来更新模型的参数，如此反复迭代以实现能量函数的最小化，从而达到模型实例与输入图像相匹配的目的。最终得到的形状特征点位置就描述了当前人脸图像的特征信息。随着具体的 AAM 匹配算法的不断进步，从最初的基于梯度下降算法的匹配计算，发展到基于线性假定的 AAM 匹配计算，再到现在的基于 Lucas-Kanade 算法的反向组合，AAM 模型的精度和效率都得到了很大的提升。

2.2.3 手动与自动标点方式的比较

以上介绍的两种标定人脸特征点的方法，在实际的表情变换过程中均可以使用，各有其优缺点：

（1）手动标定特征点的一大优势就是其拥有很高的精确性，这是自动标点方式不能比的。手动标点依靠了人眼的观察，因此可以对特征点的选取进行比较精确的定位。在实际的操作过程中，由观察而确定的特征点位置的误差不会超过 2~3 个像素。但是，这种方法的缺点也显而易见，那就是极低的效率以及相当大的工作量。由于每一张图片都要手动标点，要把它用在动画电影制作或实时人机互动的应用中去的话，是很难胜任的。

（2）相比于手动标点，自动标点方式就没有那么精确了。由于人脸识别和模型匹配总有误差，因此限制了这种方式的精确性，特别是不同脸型的情况下脸部轮廓的标定是很粗糙

的。但是，这种方式快速、高效，并且能够做到实时检测，所以在需要实时性的人际交互应用中，这种方法使用较为广泛。

在具体的应用中，需要根据具体需求来确定使用哪种方式，甚至两者联合使用。

2.3 三角网络剖分算法

2.2 节已经介绍了特征点的标定方法，那么标定了人脸的特征点之后，该用什么方法才能对图像进行变形呢？这就需要用到三角网络剖分算法了。三角网络的剖分是大部分图像变形算法的基础，这种方法并不是通过某一变换函数同时对整张图像变形，而是把特征点当作顶点，首先将图像分解成一个个小三角形，然后对每一个小三角形分别作仿射变换，变形成为目标形状。图 2-3 可以很好地说明这一过程。例如，我们要将图 2-3 (a) 所示的图像变形成为 (b) (注意，这里的变形仅仅指几何形状的变化，图片像素的颜色不改变)，首先需要两张图像都进行三角剖分，三角形的顶点就是第一步中标定的特征点。接着，找到两幅图像对应的三角形，例如图 2-3 中用红色表示的部分，利用仿射变换将 (a) 中的三角形变形为 (b) 中的三角形。只要所有的三角形都进行这样的变换，最终就能实现对整张图像的变形。

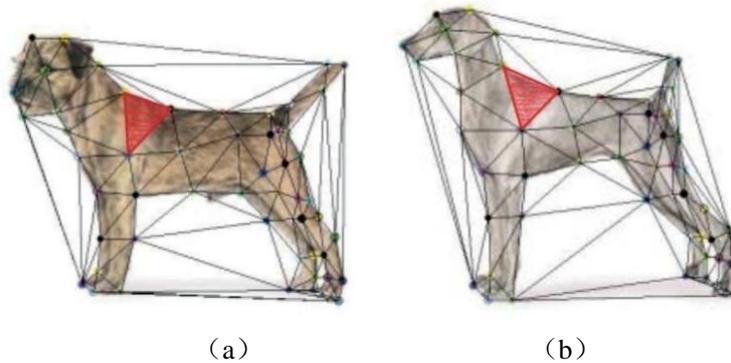


图 2-3 利用三角网络进行图像变形

将图像剖分成三角网络必须满足最基本的一个条件，即除了交于三角形顶点外，网络中任意两条边都不相交。如果两条边交于非顶点的位置，那么网络中就会凭空多出一个节点，这种情况是不允许的。例如对图 2-4 中的 7 个点进行三角剖分(特征点用黑色实心圆点表示)，(a) (b) (c) 这三种情况都是符合要求的，但 (d) 的剖分方法就是不符合要求的，因为其中两条边在非顶点的位置相交了(交点没有用黑色圆点表示)。

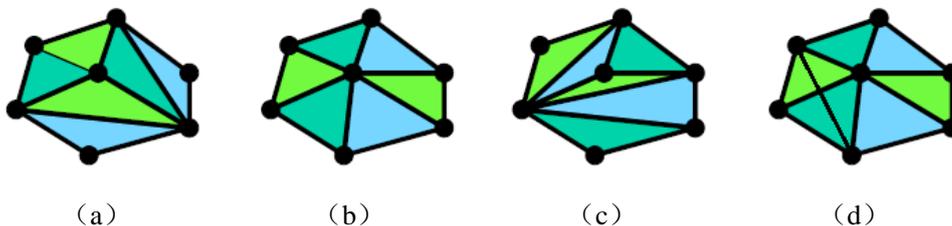


图 2-4 几种三角网络剖分方法

既然图 2-4 的前三种剖分方法都是合理的，那么究竟采用哪一种更好？直觉告诉我们，(b) 的效果要优于其他两种方法，因为它剖分地更为“均匀”。要达到这种最优化的剖分效果，可以采用“德劳内三角剖分”(Delaunay Triangulation) 算法^[15]。

Delaunay 三角剖分算法是一种“最接近规则化的”三角剖分方法，它不仅可以保证网络中任意两条边不相交，还能尽可能“平均”地生成这个网络，使三角形各个内角的角度值尽量接近，不会出现如图 2-4 (c) 那样的大角度和小角度的两极分化。Delaunay 三角剖分

需要满足下列两个特性：

(1) 空圆性：空圆性的定义是，在 Delaunay 三角网络中，任一三角形的外接圆内都不包含点集中的其他任何点。图 2-5 表示了一种不满足空圆性的剖分方法：如果取三角形 ABD 的外接圆，那么点 C 就会落在这个圆内；如果取三角形 BCD 的外接圆，那么点 A 就会落在这个圆内。无论是哪种情况，都证明了这种剖分方法不满足空圆性的特征。遇到这种情况，只要将对角线 BD 换成 AC，就必定能满足空圆性的要求，称为局部优化算法。图 2-6 即为改正后的三角剖分结果。

(2) 最小角最大化特性：“最小角最大化”是指，Delaunay 剖分算法产生的三角形的最小角是最大的，其他任何剖分方法形成的最小角都比 Delaunay 方法来的小。这个特性与空圆性有很大的联系，图 2-5 到图 2-6 的变化就反映了这个特性。很明显，在图 2-5 中，最小角是 $\angle ADB$ ，而在图 2-6 中，最小角是 $\angle ACB$ ，正确的剖分算法使最小角变大了。正因为这个特性，Delaunay 三角剖分算法能给人一种剖分得比较“均匀”的感觉。因此，这种算法是“最接近规则化”的剖分算法，均匀的剖分结果能给接下来的计算带来更高的精度。

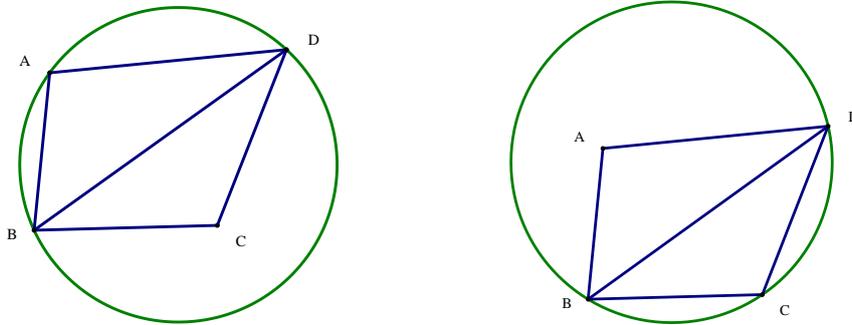


图 2-5 不满足空圆性的剖分方法

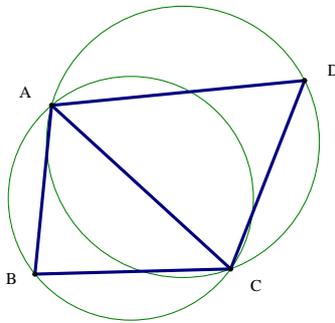


图 2-6 满足空圆性的剖分方法

Delaunay 三角剖分算法只是一种标准，实际实现的时候可以采用多种算法。经过优化，可以将该算法的时间复杂度控制在 $O(n \log n)$ ，算法可行性很高。

2.4 仿射变换

对已知图像进行三角剖分后，接下来的问题就归结为如何将源三角形变换成目标三角形了，这个过程可以用仿射变换来实现。下面，将逐一介绍有关仿射变换的知识。

2.4.1 仿射变换概述

仿射变换 (Affine Transformation) 是计算图形领域中一种非常重要的处理几何图像的手段。几何图像经过仿射变换，可以实现任意的线性变形。仿射变换的形式多种多样，但总是

具有两个普遍的性质，即平直性和平行性。平直性是指一个图形经过仿射变换后，直线依旧是直线，弧线依旧是弧线，两者不会互相转化；平行性是指原图中平行的两条直线，经过仿射变换后依然能够维持平行的性质。与之相对的，另一种变换几何图形的方式称为透射变换。透射变换并不具有平行性，也就是说两条平行的直线经透射变换后可能不再平行。本文使用的方法是仿射变换。

利用仿射变换进行表情变换实际上体现了用线性逼近非线性的方法。我们知道，人脸的表情变化是一个非常复杂的非线性过程，但仿射变换只能实现线性范围内的变形。那么，三角剖分的作用就体现出来了。将一幅图像剖分成一个个小的三角形，虽然对每个小三角形都采用了线性变换算法，但是将这些小三角形组合起来后，最终实现的整体效果就很接近非线性。这就和用许多微小的直线段来逼近曲线的道理是一样的。这种“化整为零”的思想在很多领域内都用得到，应当得到重视。

仿射变换虽然形式多种多样，但都可以分解成一系列基本几何变换的组合。如图 2-7 所示，图中表示了仿射变换可以分解成的几种基本变换过程，从左到右依次是旋转（rotation）、错切（shear）、平移（translation）、缩放（scaling）。同时，图中不方便表示的反射（reflection）同样也是基本变换之一。这些子变换将在 2.4.3 节介绍。

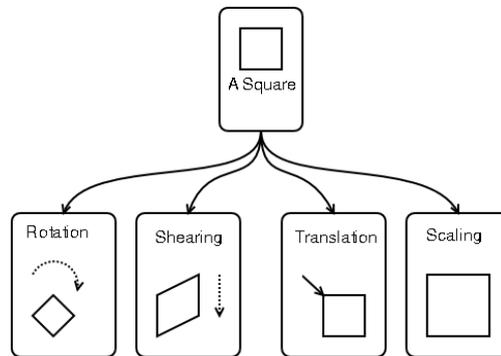


图 2-7 仿射变换的子过程

2.4.2 齐次坐标

简单的二维几何变换，例如平移、旋转和缩放，都可以用公式 (2-1) 中的普通矩阵形式来表示：

$$\mathbf{P}' = \mathbf{M}_1 \cdot \mathbf{P} + \mathbf{M}_2 \quad (2-1)$$

其中， \mathbf{P} 是变换前的位置坐标列向量， \mathbf{P}' 是变换后的位置坐标列向量，矩阵 \mathbf{M}_1 是一个包含乘法系数的 2×2 矩阵，主要可以实现缩放、旋转等功能， \mathbf{M}_2 是包含平移项的两元素列矩阵。为了利用这个公式产生先缩放、再旋转、后平移这样的变换顺序，必须一步步地计算变换的坐标。首先将坐标位置缩放，然后将缩放后的坐标旋转，最后将旋转后的坐标平移。这样的计算过程相当麻烦，在一些大型的设计和图形构造应用中都很难保证高效。因此，需要对公式 (2-1) 的各个部分进行重组，力争能够用某种单一的形式代替这种乘法和平移的复合形式。如果将 2×2 矩阵表达式扩充为 3×3 矩阵，就可以把二维几何变换的乘法和平移项组合成单一的矩阵表示。齐次坐标就是为了这个目的被提出的。

齐次坐标可以将二维平面中的双参数坐标扩充成为三参数坐标，从而满足单一矩阵表示的目的。齐次坐标的规则十分简单，只要将二维位置坐标 (x, y) 扩充成三维坐标 (x_h, y_h, h) ，其中 h 表示齐次参数，是一个非零值。齐次坐标存在如下等式：

$$x = \frac{x_h}{h}, \quad y = \frac{y_h}{h} \quad (2-2)$$

由于齐次参数 h 可以取任何非零值，因此对于每个二维坐标点 (x, y) ，可以有无数个等价的齐次坐标表达式。例如，二维坐标 $(2, 1)$ 可以表示为 $(2, 1, 1)$ ，或 $(4, 2, 2)$ ，或 $(6, 3, 3)$ ……这些表示方式都是等价的。为了计算的方便，可以统一将 h 设置为 1，这样任何一个二维坐标 (x, y) 表示成齐次坐标就是 $(x, y, 1)$ 。

利用齐次坐标表示位置，可以使我们用矩阵相乘的形式来表示所有的几何变换公式，这是图形系统中使用的标准方法。二维坐标位置用三元列向量表示，而二维变换操作用一个 3×3 矩阵表示，该矩阵称为变换矩阵。

2.4.3 基本变换的变换矩阵

前面提到，任何仿射变换都可以分解成几个基本变换的组合，而这些基本变换都可以用一个变换矩阵来表示。下面将分别介绍五种基本变换的变换矩阵。

(1) 二维平移矩阵

根据齐次坐标，二维平移过程可以用下面的矩阵乘法来表示。

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2-3)$$

或者，可以简写为

$$\mathbf{P}' = \mathbf{T}(t_x, t_y) \cdot \mathbf{P} \quad (2-4)$$

其中， $\mathbf{T}(t_x, t_y)$ 就是公式 (2-3) 中的 3×3 矩阵。

(2) 二维旋转矩阵

绕某一基准点 (x_r, y_r) 旋转角度 θ 的变换矩阵可表示如下。

$$\mathbf{R}(x_r, y_r, \theta) = \begin{bmatrix} \cos \theta & -\sin \theta & x_r(1 - \cos \theta) + y_r \sin \theta \\ \sin \theta & \cos \theta & y_r(1 - \cos \theta) - x_r \sin \theta \\ 0 & 0 & 1 \end{bmatrix} \quad (2-5)$$

特别地，绕原点的旋转变换矩阵可表示为

$$\mathbf{R}(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2-6)$$

(3) 二维缩放矩阵

关于某一基准点 (x_r, y_r) 进行缩放时，缩放变换矩阵可表示为

$$\mathbf{S}(x_f, y_f, s_x, s_y) = \begin{bmatrix} s_x & 0 & x_f(1 - s_x) \\ 0 & s_y & y_f(1 - s_y) \\ 0 & 0 & 1 \end{bmatrix} \quad (2-7)$$

其中， s_x 和 s_y 分别是 x 和 y 方向的缩放因子。缩放因子可以取任何正数值，当其大于 1 时将放大图像尺寸，小于 1 时将缩小图像尺寸。如果缩放因子 s_x 和 s_y 取相同的值，那么就能在缩放的过程中保持纵横比不变，这种变换称为一致缩放 (uniform scaling)。

(4) 反射变换

反射变换能够产生目标图像的镜像，反射的基准称为反射轴。式子 (2-8) 给出了反射轴分别是 $y=0$, $x=0$, $y=x$ 和 $y=-x$ 四种情况下的反射变换矩阵。

$$\begin{aligned} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} (y=0), & \quad \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} (x=0) \\ \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} (y=x), & \quad \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} (y=-x) \end{aligned} \quad (2-8)$$

对于平面内其他任意直线的反射，我们可以通过先平移，再旋转，再反射的变换组合来实现。对于这些过程的变换矩阵，已经在之前的部分介绍过了，不再赘述。

(5) 错切变换

错切是一种使对象形状发生变化的变换，经过错切的对象好像是由已经相互滑动的内部夹层组成。两种常用的错切变换是移动 x 坐标值的错切和移动 y 坐标值的错切。

相对于某一参考线 y_{ref} 的 x 方向的错切可用如下的矩阵生成。

$$\mathbf{SH}(sh_x, y_{ref}) = \begin{bmatrix} 1 & sh_x & -sh_x \cdot y_{ref} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2-9)$$

相对于某一参考线 x_{ref} 的 y 方向的错切可用如下的矩阵生成。

$$\mathbf{SH}(sh_y, x_{ref}) = \begin{bmatrix} 1 & 0 & 0 \\ sh_y & 1 & -sh_y \cdot x_{ref} \\ 0 & 0 & 1 \end{bmatrix} \quad (2-10)$$

其他形式的错切同样可以通过基本变形的组合来实现。

2.4.4 仿射变换的变换矩阵

平移、旋转、缩放、反射和错切都是仿射变换的特例，任何形式的仿射变换都可以表示成这五种变换的组合。例如，刚体变换 (Rigid Transformation) 同样作为仿射变换的一种特殊形式，必须维持物体的形状和大小不发生改变。因此，刚体变换可以表示为旋转和平移的组合。仿射变换维持了物体变换前后的平直性和平行性，在图形系统中作用十分广泛。

仿射变换的变换矩阵可表示为

$$\mathbf{T} = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \quad (2-11)$$

式 (2-11) 是仿射变换的一般形式。只要知道了这个变换矩阵，就可以实现将源三角形变换成为目标三角形的过程。如此一来，图像的整体变形就能得以实现了。

2.5 离散点的插值方法

2.4 节给出了仿射变换的一般形式。不过，这种变换是基于位置坐标是连续的情况的。实际上，我们知道图像是由一个个离散的像素点组成的，因此在对图像进行仿射变换时，可能出现变换后的坐标落在非整数点的位置，即发生了“落在像素点之间”的情况。这种情况下，我们就没办法确定变换后的整数点的灰度值究竟是多少。为了解决这个问题，就需要引入插值的概念。

首先，我们先要反向考虑仿射变换。如图 2-8 所示，只要求出仿射变换矩阵 \mathbf{T} 的逆矩阵 \mathbf{T}^{-1} ，就能通过目标图像中的整数点 (x', y') 反向求出在源图像中对应的坐标 (x, y) 。我们假设 (x, y) 是一个非整数点坐标，由于这个非整数点四周的几个像素的灰度值是已知的，我们就可以据此推算出 (x, y) 合理的灰度值，并将这个灰度值赋给 (x', y') 代表的像素。推算灰度值的过程称为插值。以下介绍两种基本的插值法。

2.5.1 最近点插值

最近点插值，顾名思义，就是将 (x, y) 的灰度值设定为距离其最近的一个像素的灰度值。这种方法的优点是简单明了，但是容易产生图形边缘锯齿化的负面效应。

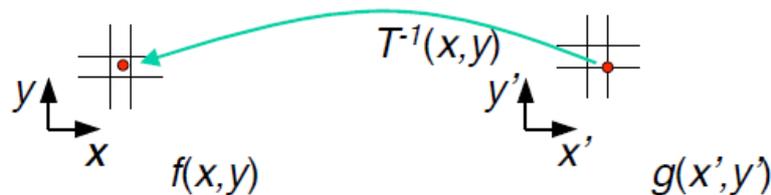


图 2-8 反向仿射变换

2.5.2 双线性插值 (Bilinear Interpolation)

双线性插值比最近点插值稍复杂一些，主要利用了非整数点 (x, y) 周围临近的 4 个整数点的灰度值作为推测的依据。双线性插值法的计算过程可用图 2-9 表示。

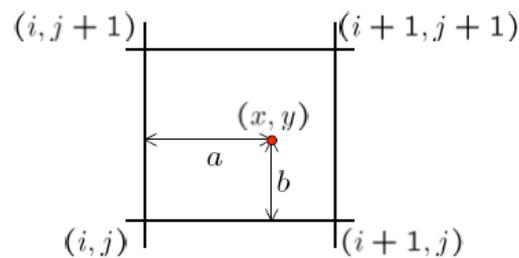


图 2-9 双线性变换图例

根据图 2-9，我们假设 $f(x, y)$ 代表点 (x, y) 处的灰度值，该点到左边整数点的垂直距离为 a ，到下方整数点的垂直距离为 b 。那么，可以用公式 (2-12) 的加权平均算法计算点 (x, y) 处的灰度值。

$$\begin{aligned}
 f(x, y) = & (1-a)(1-b) \cdot f(i, j) \\
 & + a(1-b) \cdot f(i+1, j) \\
 & + (1-a)b \cdot f(i, j+1) \\
 & + ab \cdot f(i+1, j+1)
 \end{aligned} \tag{2-12}$$

双线性插值利用了四个像素的灰度值进行推算,因此可以在一定程度上避免图形边缘锯齿化的负面效应,使变换结果变得更加平滑。不过,由于取了四个点的加权平均,变换图像会出现一定程度的模糊。

当然,除了以上两种插值方式外,还有其他诸如三立方插值这样的方法。一般的图像变换利用这两种方式已经足够了,所以其他方法不再介绍。

2.6 在变换中加入纹理细节——表情比例图像 (ERI)

2.2~2.5 节介绍的是基于几何变形的表情变换算法。这种算法提取了人脸特征点的位置,进而进行三角剖分,通过仿射变换的方法实现整个变形过程。从这个过程中我们可以看到,仅仅是人脸的几何形状发生了改变,而由于肌肉运动在皮肤表面产生的如皱纹、酒窝等的纹理细节都没有得到体现,因此结果不够真实。这是传统基于几何的表情变换算法的一大劣势。为了解决这个问题,Z. Liu 等人在文章[9]中引入了“表情比例图像”(ERI)的概念,将纹理细节加到了变换结果中去。图 2-10 显示了这种算法的优越性。

在图 2-10 中,(a)(b)(c)是已知的三张图像。(a)为人物 A 的中性表情 A_0 , (b)是其“思考”的表情 A_1 , (c)是人物 B 的中性表情 B_0 。我们需要将 (b) 中所示的模板表情映射到 (c) 的人脸上去,合成出第二个人“思考”的表情 B_1 。如图所示,(d)是仅仅采用几何变换的方法得到的表情变换结果。虽然这个结果相比于 (c) 而言确实是严肃了一些,特别是眼睛的形状发生了较明显的改变,但是依然很难让人相信这是一个人在“思考”。在加入了 ERI 图像后,就能得到图 (e) 中的结果。这时,我们看到眉毛之间出现了明显的由于“皱眉”而产生的纹理细节,脸部其他部分的光照特征也发生了细微的改变。这种结果能让人一眼就相信此人正在思考,结果十分逼真。

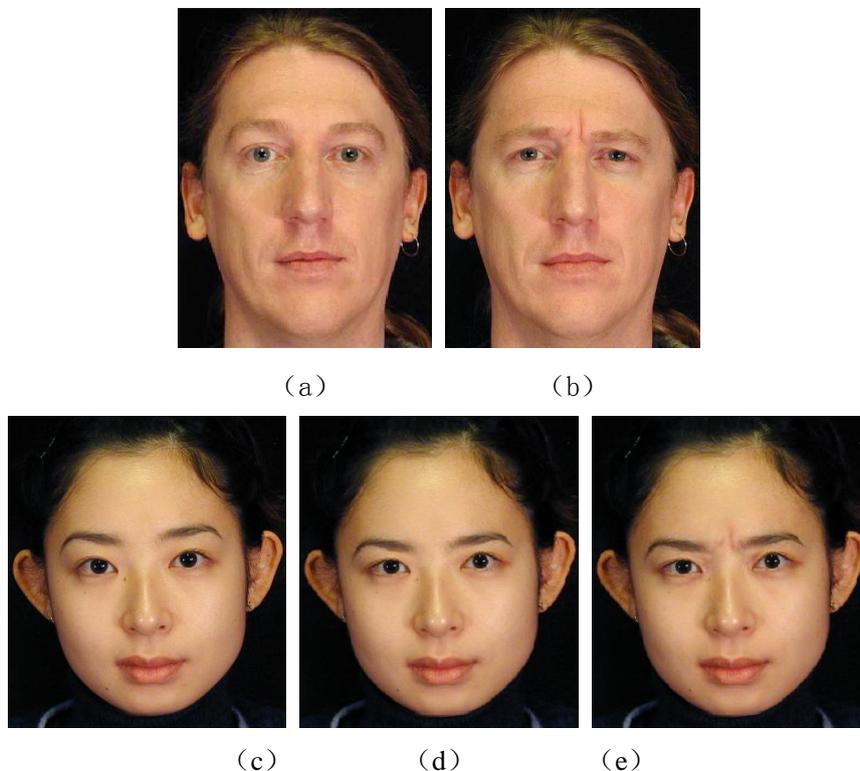


图 2-10 运用 ERI 图像进行表情变换实例

图 2-10 (d) 到 (e) 的提升就是由于加入了 ERI 图像。这里,我们假设 $P(u, v)$ 代表图像 P 中像素点 (u, v) 处的灰度值,并且为了方便说明,我们提前假设图像 A_0 、 A_1 和 B_0 已经对

齐了。对齐的意思是指两张图像特征点的位置以及三角网络的形状完全相同，对齐过程可以通过几何变形来实现。那么，带纹理细节的表情移植结果 B_I 就可以用如下公式计算得到

$$B_I(u, v) = B_0(u, v) \cdot \frac{A_1(u, v)}{A_0(u, v)} \quad (2-13)$$

其中，分式 $\frac{A_1(u, v)}{A_0(u, v)}$ 就是 ERI 图像的雏形，代表了两幅图像之间的纹理差异。

当然，公式 (2-13) 是在三幅图像都对齐了的假设下得到的结论。为了真正推出 ERI 图像的计算式，首先要经过对齐这一步骤。在这里，我们需要将图像 A_0 进行变形，使它的网络能够与 A_I 对齐，定义 A_0 变形后的结果为 A_g 。ERI 完整的定义如下

$$ERI = \frac{A_1(u, v)}{A_g(u, v)} \quad (2-14)$$

由图 2-10 (a)、(b) 两幅图像得到的表情比例图像如图 2-11 所示



图 2-11 表情比例图像 ERI 示例

最后，要完整地实现二维表情变换算法，只要根据如下步骤进行就可以了：

- (1) 将 A_0 的网络与 A_I 对齐，计算得到表情比例图像 ERI
- (2) 通过 A_0 和 A_I 特征点之间的位移关系，计算得到 B_I 的特征点位置
- (3) 将 B_0 的网络与 B_I 对齐，得到变形后的结果 B_g ，即为图 2-10 (d) 表示的结果
- (4) 将 ERI 图像与 B_I 对齐，得到变形结果 ERI'
- (5) 计算最终结果： $B_I(u, v) = B_g(u, v) \cdot ERI'(u, v)$ 。结果如图 2-10 (e) 所示。

不过，采用表情比例图像进行表情的纹理移植是有一定限制的，这种限制往往表现在图像的整体光照强度上。根据公式 (2-13) 表示的那样，利用 ERI 图像进行表情移植的核心其实是求得同一人脸在不同表情下的灰度之比，用这种灰度的比值来反应纹理的强弱和分布信息。ERI 图像中小于 1 的像素表示纹理会变暗，大于 1 的像素表示纹理会变亮。将这种亮度的变化叠加到另一个人的脸部就能“复制”出纹理细节的变化效果，这看似是没有问题的。但是，如果模板人脸和目标人脸图像的平均光照强度不同，例如前者总体偏暗而后者总体偏亮，那么将 ERI 图像叠加到目标人脸上去后，由于目标人脸亮度的基数很大，因此相乘以后，这种纹理色彩的变化会被过分放大，容易导致不必要的噪声。

另外，如果目标人脸是一个虚拟人物，例如阿凡达，它皮肤颜色的基调是蓝色，和我们正常人类的差别很大。这种情况下用正常人类的 ERI 图像进行表情移植时，也会出现色彩上的失真问题。所以将 ERI 图像用在虚拟的动画电影应用中，还是有其局限性的。我们在具体的实现过程中也会遇到各种各样影响结果的问题，解决这些问题的方法将在下一章具体实现时详细讲述。

2.7 本章小结

本章主要讲述了传统的带纹理细节的二维表情变换算法的流程,并介绍了流程中各个关键步骤的详细理论依据。带纹理的二维表情移植算法主要分为基于几何变形的表情变换算法和纹理的移植方法,前者在 2.2~2.5 小节中得到了详细的阐述,包括如何通过手动或自动的方式标定人脸的特征点、如何用德劳内三角网络剖分算法得到“最优化”的三角网络、如何利用仿射变换进行图像的整体变形,以及在离散的情况下如何用插值算法确定非整数点的灰度值;后者在 2.6 这一小节中做了详细的理论介绍,阐述了纹理移植的基本思想,即如何通过表情比例图像将模板人脸的纹理细节移植到目标人脸上去,使结果更加自然和逼真。不过,不管是基于几何变形的表情变换算法,还是基于 ERI 图像的纹理移植算法,都只是理论依据,在实际实现的过程中会存在各种各样的问题。第三章将通过表情变换的具体实现过程来解释和解决这些问题,并且通过提出基于弹力和肌肉分布的新模型,使表情变换达到更出色的效果

第三章 基于弹性和肌肉分布模型表情变换算法的实现

在第二章中，我们详细介绍了传统带纹理的二维表情变换算法的一般流程，这部分主要涵盖了基本算法的理论依据。但是，在实际实现的过程当中，往往会遇到各种各样复杂的问题，仅仅依靠理论知识是没有办法完全解决的。因此，本章将会在基本理论的基础上，通过一个实例着重讲述带纹理细节的二维表情变换算法在实际应用中的实现过程，解释其中可能遇到的问题，并提供可行的解决方法。当然，解决的办法并不唯一，而本章提供的解决思路基本能够满足要求，使得变换结果尽可能逼真、自然。

同时，在观察和分析传统表情变换算法的不足之处的基础上，本章提出了一种基于人脸弹力和肌肉分布的新模型，来优化表情变换的结果，这同时也是本文最重要的创新点和贡献。新的模型主要包括两个子模块，其一是基于人脸弹性的拉力/推力模型，另一个则是基于脸部肌肉分布的纹理增强模型。基于人脸弹性的拉力模型主要解决了传统的“全局”（Global）几何变换过程中可能出现的五官比例不协调的问题。在这个模型中，我们首先根据目标人脸的结构对脸部各个部分进行局部调整，使其更加符合模板人脸的五官特征，形成一个“局部”（Local）几何变换结果。接着，根据人脸部具有弹性的特征，通过相邻特征点之间的拉力（或推力）将“全局”和“局部”两个变换结果结合起来，形成一种“受力平衡”的状态，以此优化几何变换的结果；在基于肌肉分布的纹理增强模型中，我们根据三维建模领域经典的脸部肌肉模型^{[3][16]}，分析了肌肉运动对纹理细节可能产生的影响，并提出了表情比例图像的高斯模型（Gaussian Model）对纹理细节进行强化。

本章将做如下安排：3.1 节将会通过一个实例将第二章介绍的算法具体实现出来。3.2 节会简要介绍局部变换的原理并给出一个示例。3.3 节将会介绍基于弹性的拉力模型的概念和实现方法。3.4 节会介绍基于肌肉分布的纹理增强模型的理论 and 实现。最后，会在 3.5 节对本章做个小结。

3.1 带纹理的二维表情变换算法实现

本节将通过一个实例，如图 3-1 所示，介绍带纹理的二维表情变换算法的具体实现过程，解释可能遇到的问题，并给出解决的方案。

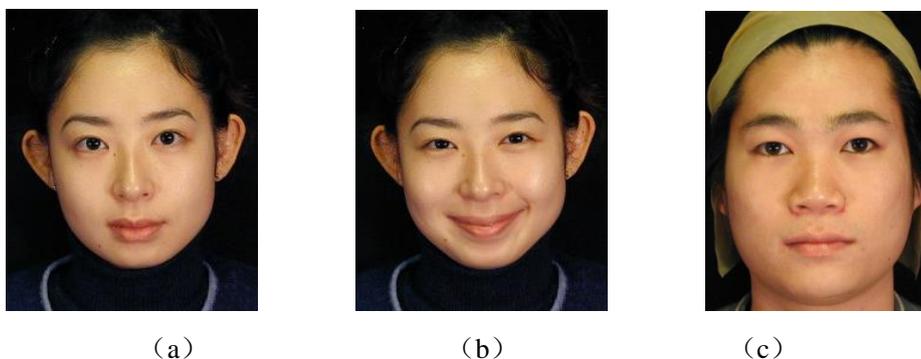


图 3-1 二维表情变换的已知图像

在图 3-1 中，(a) 为中性表情 A_0 ，(b) 为模板表情“微笑” A_1 ，(c) 为目标人脸的中性表情 B_0 。本节要介绍的表情变换的实例，就是将 (b) 所示的表情移植到 (c) 中的人脸上

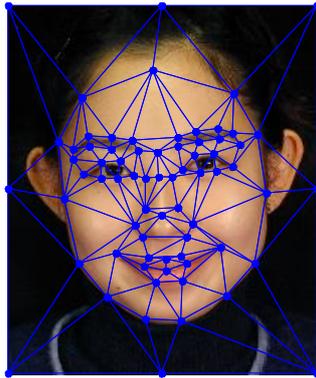


图 3-3 德劳内三角剖分算法示例

3.1.3 根据“全局”变换确定 B_1 的特征点位置

在进行仿射变换前,我们首先需要通过 A_0 、 A_1 和 B_0 的特征点位置确定 B_1 的特征点位置,然后才能将图像 B_0 变形为 B_1 的形状 (B_0 通过几何变换与 B_1 对齐的结果为 B_g , 曾在 2.6 节提及)。根据传统的方法,我们可以采用非常简单的“全局”变换来确定 B_1 的特征点位置。其基本思想是,先计算得到 A_1 特征点和 A_0 特征点之间位置的相对变化,然后将这种位置的相对变化作用在 B_0 的特征点上,移动后得到 B_1 的特征点位置。具体说来,可以细分为脸部轮廓特征点的移动和脸部内侧特征点的移动两个步骤:

(1) 脸部轮廓特征点的移动:在这个步骤中,我们计算 A_0 和 A_1 脸部轮廓的特征点相对于整张图片边缘的相对移动,将这种相对移动作用到 B_0 的轮廓特征点上。

(2) 脸部内侧特征点的移动:在这个步骤中,我们计算 A_0 和 A_1 脸部内侧的特征点相对于脸部轮廓的相对移动,将这种相对移动作用到 B_0 的内侧特征点上。

得到了 B_1 的特征点位置后,我们就能进行下一步的仿射变换了,将 B_0 的网络与 B_1 对齐。

3.1.4 不同插值方法下仿射变换的实现

仿射变换和插值方法的原理已经分别在 2.4 节和 2.5 节中详细介绍过了,这里我们就不再赘述了。以图 3-1 (c) 为例,我们试图分别用最近点插值和双线性插值方法对此图像进行仿射变换,使它的网络与 B_1 的对齐,得到结果 B_g 。这里的变形过程为“全局”变形,结果 B_g 可以初步得到 (c) 微笑的样子,虽然这种结果只是几何上发生了变化。图 3-4 (a)、(b) 分别显示了用最近点插值和双线性插值法得到的初步变形结果。

图 3-4 中两幅图片最大的差异在于嘴。正如在 2.5 节所说的那样,双线性插值法得到的嘴部轮廓比较平滑,而用最近点插值得到的嘴的轮廓就显得棱角分明了。虽然在大部分图像变换的应用中为了得到更平滑的效果,双线性插值往往用得比最近点插值更加广泛。但是,本文的实验对结果的要求是比较特殊的,我们恰恰需要一个棱角分明的结果,而图 3-4 (b) 的嘴的周围部分出现了一定程度的模糊,似乎使得整个嘴的形状“变厚”了,这是我们不希望看到的。所以,在本文的后续试验中,我们选择了最近点插值作为仿射变换的一部分,而非双线性插值。

从图 3-4 中,我们还可以发现,无论采取哪种插值方法,仅仅进行几何变形都是远远不够的。图 3-4 所示的两个结果均显得非常呆滞,没有立体感。虽然嘴角上扬了,但是要让人相信这个人物在微笑,恐怕是很难的。

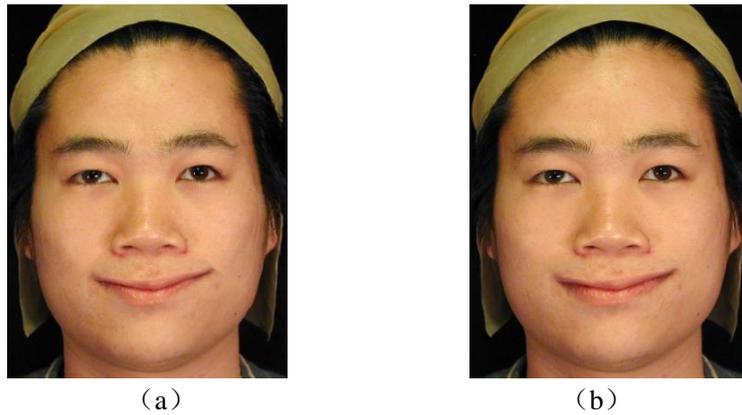


图 3-4 基于几何变形的表情变换示例：(a) 最近点插值 (b) 双线性插值

3.1.5 利用表情比例图像移植纹理细节的实现。

根据公式 (2-14)，为了得到模板表情的 ERI 图像，我们首先需要将 A_0 和 A_I 对齐，并得到 A_0 对齐后的结果 A_g 。这一过程同样可以用上面几个小节介绍的方法来实现。将 A_0 和 A_I 对齐后，我们得到如图 3-5 所示的结果。其中 (a) 为中性表情 A_0 ，(b) 为 A_0 对齐后的结果 A_g ，(c) 为模板表情 A_I 。

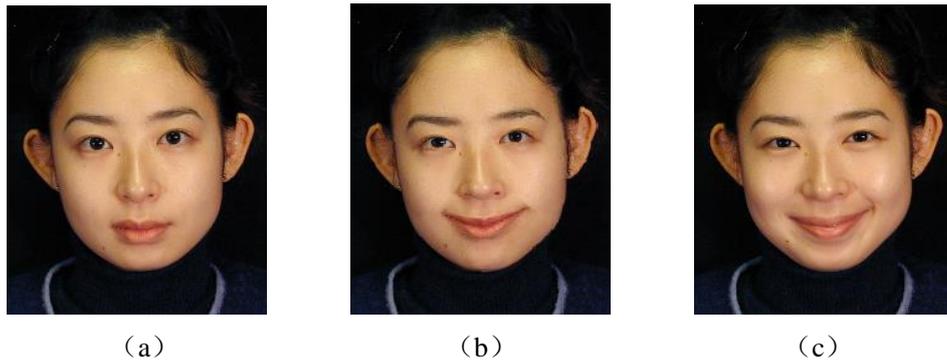


图 3-5 中性表情与模板表情的对齐过程

我们可以明显看到，对齐后的结果 (b) 和 (c) 的差距还是很大的。这种差距不仅仅表现在纹理细节的缺失上，就连只从几何变形的角度看，效果也相当不自然。这一方面是由于手动标定的特征点数量有限，标定更多的特征点可以达到更精确的效果；另一方面是由于仅仅基于几何进行表情变换确实是不够的，需要加入 ERI 图像让结果更加真实。

接着，将图 3-5 (c) 和 (b) 进行对应像素点相除运算，就能得到模板表情的 ERI 图像，如图 3-6 所示。



图 3-6 模板表情的 ERI 图像

图 3-6 所示的模板表情的 ERI 图像并不能直接进行使用, 因为图片中很明显存在大量噪声, 与图 2-11 表示的理论情况相去甚远。原因可解释如下: 例如, 在背景的深色部分, 某个像素点在一幅图像中的灰度值是 2, 在另一幅图像中的灰度值是 3。两者相除, 得到的 ERI 值为 1.5。虽然两者灰度值的绝对之差仅为 1, 肉眼几乎无法分辨这种差异, 但是相除之后差异就被放大了, 达到了 1.5。如果再把这个 ERI 值乘到另一个灰度为 100 的像素上, 产生的绝对误差就会达到 50! 如果我们把上述例子中的原始灰度值之差 1 看作是噪声的话, 那么由此可以得出一个结论: ERI 图像的噪声更容易在低灰度值的区域出现。

为了优化 ERI 图像, 去除图片中的大部分噪声, 我们采取了 (1) 截取和 (2) 滤波两种方式:

(1) 截取: 其实, 我们真正关心的是人脸内侧的区域, 而人脸外侧的背景、服装等部分, 在表情移植中是没有任何作用的, 巧合的是大部分噪声都在这些区域中。所以, 我们可以干脆将 ERI 图像中属于人脸轮廓外部的区域截掉, 仅保留人脸内侧的部分。这样, 就避免了大部分噪声。如图 3-7 (a) 所示。

(2) 滤波: 虽然进行了截取, 但是人脸内侧多少还是存在一些噪声的。去除这部分噪声就要用到滤波的方法了。不过, 直接对整个脸部进行滤波是不可行的, 因为这样好不容易得到的纹理细节也会被当成噪声一同滤掉。因此, 我们需要对图像的不同部分采用不同程度的滤波强度, 例如对纹理部分进行弱滤波, 减少对纹理细节的影响, 而对非纹理部分进行强滤波, 尽可能去除噪声。

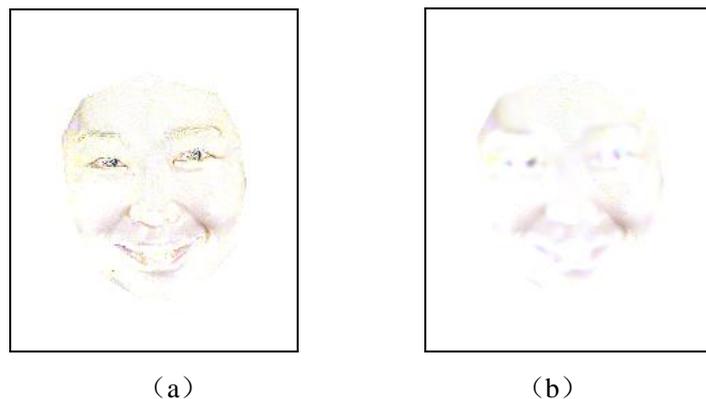


图 3-7 ERI 图像优化结果: (a) 截取 (b) 滤波

如何根据图片自动区分某个部分到底是不是纹理部分? 这可以通过计算图 3-5 (b) 和 (c) 中对应像素点的邻域的相关性来解决。如果在这个邻域内, (b) 和 (c) 的相关性越大, 则表明它们越相似, 是纹理细节的可能性就越小; 反之, 如果相关性越小, 则表明越不相似, 是纹理细节的可能性就越大。计算某个像素点 (u, v) 的相关性 $C(u, v)$ 的方法如下:

给定一个像素点位置, 我们在图 3-5 (b) 和 (c) 中分别取一个邻域窗口 (例如 11×11 大小的邻域)。我们定义图 (b) 在这个邻域内的点为 X_i , 图 (c) 的点为 Y_i 。计算相关性的公式如下

$$C(u, v) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \cdot \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (3-1)$$

$C(u, v)$ 总是在 0 和 1 之间取值。

我们再定义一个滤波权重 $W(u, v)$, 作为滤波强度的依据。

$$W(u, v) = 1 - C(u, v) \quad (3-2)$$

$W(u, v)$ 越小，滤波强度越大，反之越小。

图 3-8 为模板表情的滤波权重 $W(u, v)$ 的直方图。通过观察直方图，我们可以取阈值 w_0 ，对滤波权重大于这个阈值的 ERI 像素，进行弱滤波；小于这个阈值的 ERI 像素，进行强滤波。最后得到的滤波结果如图 3-7 (b) 所示。与 (a) 相比，噪声得到进一步滤除。

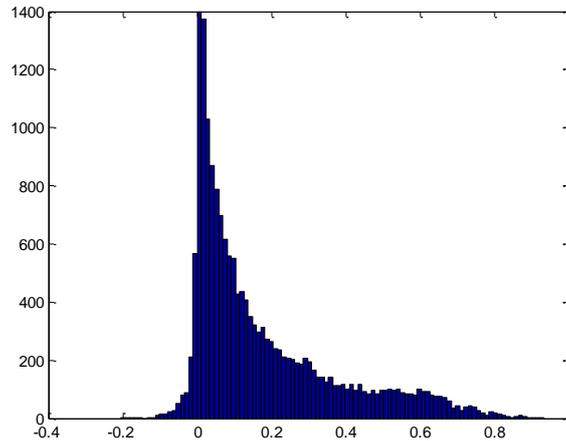


图 3-8 滤波权重直方图

ERI 图像经过上述步骤的优化后，就可以投入下一个步骤的使用了。

最后，只需要将图 3-7 (b) 所示的 ERI 图像与图 3-4 (a) 的所示的结果对齐并相乘，就能得到最终带纹理细节的“全局”表情变换结果了，如图 3-9 所示。



图 3-9 带纹理的“全局”表情变换结果

这个结果比图 3-4 (a) 有了很大的提升，纹理和光照的细节得到了体现。例如，在嘴角处，出现了几何变形无法得到的深色皱纹，在脸颊处也有明显的突起感。同时，在嘴唇中央和额头上，出现了几处细微的反光。通过叠加上 ERI 图像，整个人脸一下子变得生动了起来，并且富有立体感，而不再像图 3-4 中的结果那样呆滞。不过需要注意的是，由于滤波这一过程，纹理细节多多少少被削弱了。进一步加强纹理细节的方法将在 3.4 节中介绍。

3.2 “局部”表情变换

“全局”(Global) 变换结果是根据模板人脸的中性表情 A_0 和某种特定表情 A_1 的特征点间的相对位移来决定变换后目标人脸 B_1 的特征点位置的。这种方法的优点是运算起来方便

快捷，但有可能出现目标人脸 B_I 某些五官比例不协调的问题。例如，在图 3-4 中，可以发现经全局变换后目标人脸的嘴似乎被拉得“太宽”了。原因其实很简单：在图 3-1 中，由于目标人脸 (c) 的脸型本来就比模板人脸 (a) 来得大，而我们又是根据“相对”位移来确定 B_I 特征点的位置的，所以嘴角的绝对位移就显得过大了。因此，在全局变换的基础上，我们又提出了针对目标人脸各个器官进行修正的“局部”(Local) 变换法。S. Basu 曾在文章[17]中详细讨论了如何通过三维建模对人的嘴唇的动作进行跟踪的方法，他对人的嘴唇进行了很精细地建模，通过这个模型详细分析了嘴部的构造与运动的特征。在本文中，我们参考了这种方法，对变换结果 B_I (图 3-4) 的眼睛、眉毛和嘴唇等部分进行了几何修正，例如通过长宽比、弧度的对应关系，使结果尽量与模板人脸的五官接近。最终，我们可以得到如图 3-10 所示的“局部”几何变换结果。



图 3-10 局部几何变换结果

与图 3-4 所示的结果相比，目标人脸的嘴确实变窄了很多。但是由于模板人脸的嘴本身很小，用在目标人脸上似乎又觉得“太窄”了。为了平衡“全局”和“局部”这两种变换效果，我们将在 3.3 节提出一个新的模型，称为“人脸的弹力模型”，来达到更好的效果。

接下来的 3.3 和 3.4 这两小节，将会重点介绍本文提出的一种新算法，即二维表情变换的弹力和肌肉分布模型，分别用来在几何上和纹理细节上对结果进行优化提升。这种新算法的框图如图 3-11 所示。

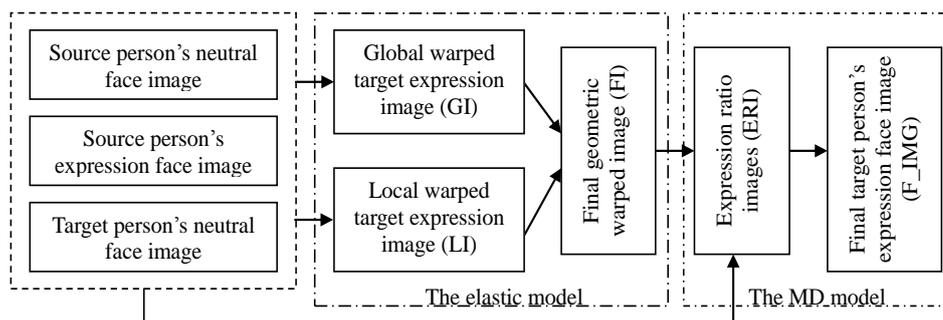


图 3-11 二维表情变换的弹力和肌肉分布模型算法框图

在图 3-11 中，第二个虚线框部分代表了脸部的弹性模型，第三个虚线框部分代表了基于肌肉分布的纹理增强模型，下面具体介绍。

3.3 脸部弹性模型

我们知道，人脸是一个十分复杂的结构，它并不是一个刚体，而是可以在皮下分布的各种肌肉的控制下，产生一系列复杂的运动，由此带动脸部几何形状和纹理细节的共同变化。

所以，我们可以认为人脸是一个充满弹性的结构，弹性的来源是各部分肌肉的运动，表现形式是皮肤的变形和褶皱。早在 1981 年，Platt 等人就发表了论文^[18]说明可以将人脸看成一个弹性网络结构，并且构造了一个“质点-弹簧”模型，用来模拟这个弹性网络。在“质点-弹簧”模型中，Platt 认为人脸的运动是由肌肉伸缩产生的力在弹性网络中的传递引起的。因此，对脸部弹性的分析是表情变换和建模中非常重要的一个环节。

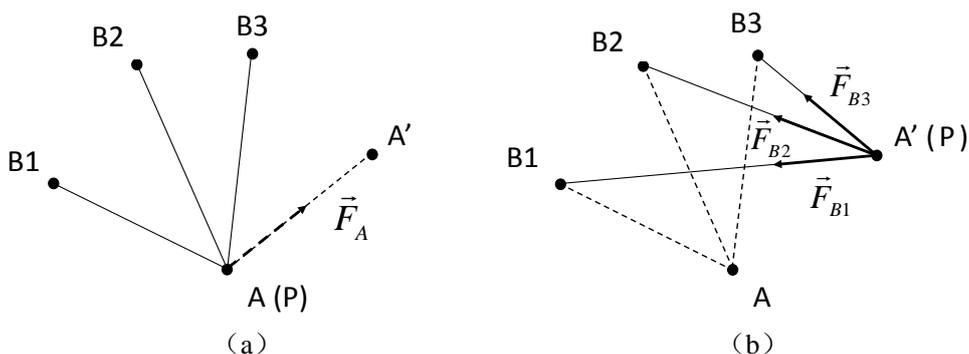
在本文提出的这个新模型中，同样利用了人脸具有弹性这一性质。利用这个弹性模型，我们可以将 3.1 和 3.2 节介绍的“全局”和“局部”几何变换结合起来。与 Platt 的工作略有差别的是，我们不打算通过对脸部肌肉的建模来确定弹性网络中质点的位置，而是简单地将标定的特征点看做是一个个质点。三角网络中相邻的特征点之间可能会存在拉力或者推力（统称弹力）来维持某些特征点的位置，使其达到受力平衡状态，并且假设不相邻的特征点之间不存在这样的弹力。对比“全局”和“局部”表情变换结果，由于我们只调整了眼睛、眉毛、嘴唇的特征点位置，因此有一部分特征点位置发生了变化，而另一部分点的位置固定不变。无论在全局还是局部变换中，可动点的受力都不平衡。最终的目标是要调整可动点的位置，使其达到受力平衡状态。

我们提出的弹性模型可由图 3-12 来描述。在图 3-12 中，点 P 是一个位置待调整的 movable 特征点。A 是点 P 在全局变换结果中的位置，A' 是点 P 在局部变换结果中的位置。B1、B2 和 B3 是三个固定点，同时是 P 的相邻点。我们假设在 P 与其相邻点 B1、B2 和 B3 之前存在拉力（或推力）关系，同时在 A 和 A' 之间也存在由 A 向 A' 的拉力。相邻点 B1、B2 和 B3 试图将 P 拉向全局变换位置 A，而 A' 试图将 P 拉向其局部变换位置 A'。所以，当点 P 落在位置 A 时（图 3-12 (a)），相邻点 B1、B2、B3 和 P 之间不存在任何弹力关系，但是 A 和 A' 之间的拉力会把 P 往 A' 的方向拉；相反，当点 P 落在位置 A' 时（图 3-12 (b)），A' 就不会对 P 产生拉力作用（可以把弹簧原长想象成 0），但是相邻点 B1、B2、B3 就会将 P 拉回全局位置 A。如此一来，只要点 P 受到的来自各个方向的力达到平衡，那么最终的位置就能被确定下来，如图 3-12 (c) 所示。点 P 在受力平衡状态下的位置可由如下的胡克定理计算得到。

$$\sum_i \vec{F}_{Bi} + \vec{F}_{A'} = 0$$

$$\sum_i \bar{k}_{Bi} \Delta l_{Bi} + \bar{k}_{A'} \Delta l_{A'} = 0 \quad (3-3)$$

其中， \vec{F}_{Bi} 和 $\vec{F}_{A'}$ 分别是来自相邻点 Bi 和局部变换位置 A' 的弹力。 Δl_x 表示了 P 移动时和点 X 之间距离的变化量，即可理解为弹簧伸长量。 \bar{k}_x 表示了每一根弹簧的弹性系数，并且这些弹性系数可以根据不同的需要进行调整，以决定平衡结果是更接近于全局变换结果还是局部变换结果。



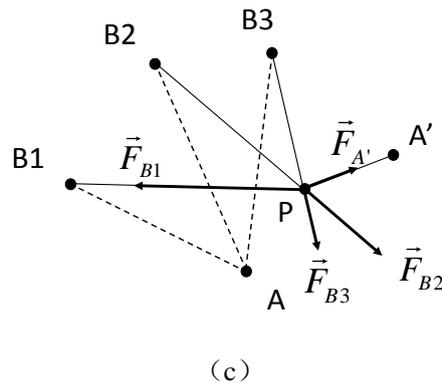


图 3-12 弹性模型示意图

经过弹性模型的平衡，表情变换在几何上的结果能够得到进一步提升，使这个结果与具体用户的脸部特征更加贴近、更加协调。

3.4 基于肌肉分布的纹理增强模型

利用 ERI 图像，可以将传统基于几何变形的表情变换算法所没有考虑到的纹理、光照细节加入到移植结果中去，使结果更加生动、真实可信。不过，本章曾提到，由于需要对 ERI 图像进行滤波去除其中的噪声，所以纹理细节在一定程度上还是被削弱了。削弱的程度取决于滤波的强度和滤波权重 $W(u, v)$ 的选择。所以说，只利用 ERI 图像进行纹理移植，与真实的模板人脸上的纹理强度依然存在不小的差距。为了进一步提升结果中纹理细节的深刻程度，我们提出了基于肌肉分布的纹理增强模型，来弥补 ERI 算法的不足。

3.4.1 人脸的肌肉分布模型

本文曾在 1.1.3 节中提到过，Waters 在八十年代提出了一个具有里程碑意义的人脸肌肉模型，用来解决三维人脸动画的建模问题。Waters 之所以提出这个概念，是因为他在长期的研究中发现了人脸部肌肉对于分析人脸表情和运动起到了十分关键的作用。确实，人的脸部运动离不开肌肉的控制，而掌握了人脸的肌肉模型就几乎掌握了人脸建模方法的一大半。Waters 建立肌肉模型的一大依据就是医学领域中关于人脸肌肉的解剖学概念，并给出了如图 3-13 所示的真实的人脸肌肉分布图。

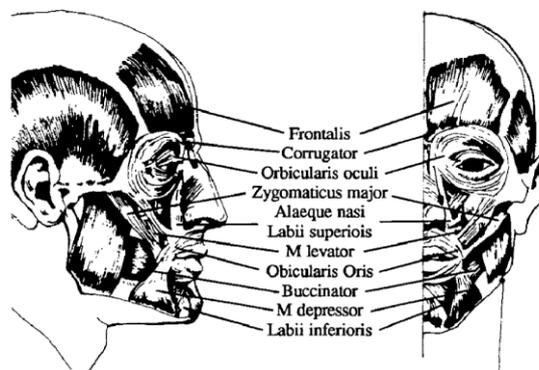


图 3-13 依据解剖学的人脸肌肉分布图

在 Waters 通过人脸肌肉的真实分布建立起一种新的人脸动画建模方法后，后人又在此基础上做了很多努力，包括不断修改这个模型，不断推广以适应更广泛的要求。文章[16]就提出了一种更先进的基于几何的肌肉模型建模方法。在文章[16]，作者结合了人脸的几何特征，对人脸主要的肌肉模块进行三维建模，并且通过改变这些肌肉的形状进而控制整个人脸

模型的外观。如图 3-14 就是文章[16]通过计算机仿真建立的一部分人脸肌肉模型的例子。

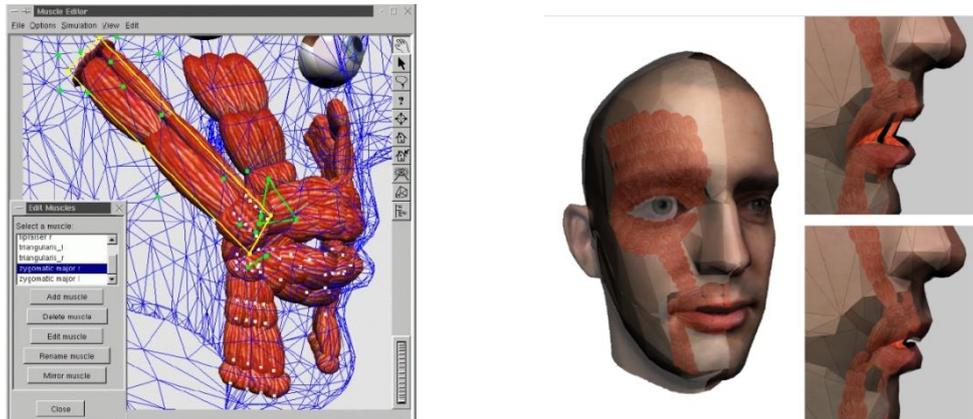


图 3-14 文章[16]建立的三维肌肉模型示例

通过研究这些模型，我们同样可以将它们运用到增强纹理细节的算法中去。在本文提出的基于肌肉分布的纹理增强模型中，我们主要通过观察那些容易产生对应表情纹理细节的肌肉运动，来有针对性地加强这些区域的纹理效果。严格地说，我们的这种方法应当称为“伪肌肉”模型更为恰当，因为我们并不严格分析每一种表情下肌肉的运动规律，而是只对那些容易通过肌肉运动而产生纹理细节的部分进行重点加强。因此，这种方法不仅可以建立统一的方法对所有表情的纹理细节进行强化，而且还灵活多变，能针对不同的人作出具体的调整，将本方法的优势发挥到最大。

3.4.2 “纹理强化圆”的选取

为了确定该对人脸哪些部分进行纹理强化，我们提出了“纹理强化圆”的概念。“纹理强化圆”的意思是指，我们根据“伪肌肉”模型在人脸上选定一些圆形区域，在这些圆形区域内，我们希望纹理特征能够得到大幅强化，而在圆形区域外，我们又希望图像的色彩不受太大影响，尽量维持原貌。在圆形区域的边界处，我们希望这种过渡尽量平滑，避免出现颜色的跳变。针对不同的表情，我们还可以单独对某些区域进行强化。在图 3-15 中，我们举了两个例子，来说明在不同的表情下可能选择的“纹理强化圆”区域。

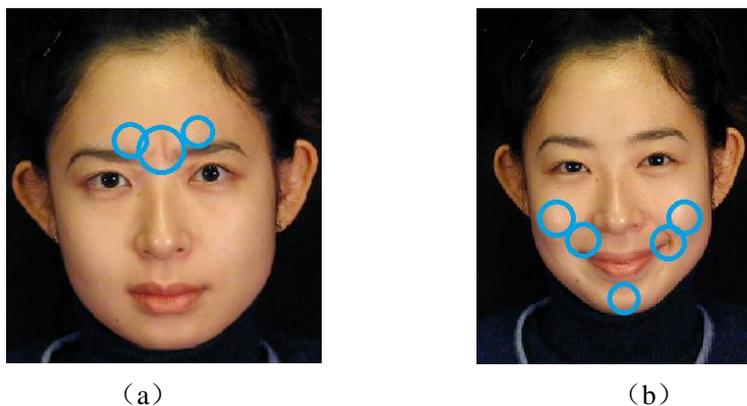


图 3-15 两种表情下“纹理强化圆”的简单选取示例

在图 3-15 中，为了显示清楚，并没有把所有的“纹理强化圆”画出来，而是仅选择了一部分比较关键的圆来说明问题。在图 3-15 (a) 的思考表情中，由于“皱眉”这一动作，纹理大部分出现在眉毛之间和眉毛上沿偏中间的部分。在图 (b) 所示的微笑表情中，纹理出现的区域更广，在脸颊、嘴角和下巴处都可能出现。对于其他表情，可以通过相似的办法

选取这些“纹理强化圆”区域。

3.4.3 利用“高斯模型”强化纹理细节

选定了“纹理强化圆”后，就需要对圆内的纹理效果进行增强了。增强纹理的过程依然用到了表情比例图像。为了满足“纹理强化圆”的三个定义，即（1）圆内纹理得到增强；（2）圆外图像基本不变；（3）圆边界过渡平滑，我们需要找出一种增强函数，它具有类似于低通滤波器的特性，并且圆的边界正好落在该低通滤波器的 3dB 点处（在通信领域，对于那些不是标准门函数的低通滤波器，常常采用 3dB 点作为该滤波器的截止频率），这样就可以满足“纹理强化圆”的三个要求了。经过筛选，高斯滤波器就满足这个条件。因此，我们提出了一种基于高斯函数的“高斯增强模型”来实现这一过程。

在介绍“高斯增强模型”前，我们首先需要定义一个新的 ERI 图像，称为平移 ERI 图像。原始 ERI 图像的中心值为 1。ERI 值小于 1 时，对应像素点的颜色会变暗，大于 1 时，颜色会变亮。只有当 ERI 值等于 1 时，对应像素点的颜色才能维持不变。这会给之后的计算带来一点麻烦。为了避免中心值为 1 的麻烦，我们需要将原始的 ERI 值减小一个单位，得到平移的 ERI 图像。这样，中心值就成为了 0：当平移 ERI 值为正时，像素点颜色变亮；为负时，颜色变暗；等于 0 时，颜色不变。平移 ERI 图像 $s_ERI(u, v)$ 的计算式如公式 (3-4) 所示。

$$s_ERI(u, v) = ERI(u, v) - 1 \quad (3-4)$$

对于某一个“纹理强化圆”，我们假设 (u_0, v_0) 为其圆心坐标，圆的半径用 $r(u_0, v_0)$ 表示。这样，用于增强纹理的高斯模型 $G(u, v)$ 就能作如下定义：

$$G(u, v) = 1 + h(u_0, v_0) \cdot e^{-\frac{(u-u_0)^2 + (v-v_0)^2}{2\sigma(u_0, v_0)^2}} \quad (3-5)$$

其中， $h(u_0, v_0)$ 为针对该圆形区域的纹理增强因子。 $h(u_0, v_0)$ 取得越大，纹理增强地越明显。 $\sigma(u_0, v_0)$ 是高斯函数的方差，可以控制该函数的影响范围。通过调整 $\sigma(u_0, v_0)$ 的大小，可以使高斯函数的 3dB 点正好对应于圆的边界。 $\sigma(u_0, v_0)$ 可由公式 (3-6) 确定。

$$\sigma(u_0, v_0) = \frac{1}{\sqrt{\ln 2}} r(u_0, v_0) \quad (3-6)$$

接着，增强后的平移 ERI 图像 $s_ERI'(u, v)$ 可由公式 (3-7) 得到。

$$s_ERI'(u, v) = G(u, v) \cdot s_ERI(u, v) \quad (3-7)$$

因为高斯模型的值总是大于 1 的，所以可以使平移 ERI 图像中正值更正、负值更负。换句话说，可以使图像中亮的更亮、暗的更暗。纹理细节因此能够得到强化。

最后，只要将增强后的平移 ERI 图像平移回去，就能得到增强后的 ERI 图像了。

3.4.4 高斯模型增强纹理的过程示意

本小节将会通过一个例子清晰地说明利用高斯模型增强纹理的过程。

由于 ERI 图像是二维的，直接表示不方便，因此，在这个例子中，我们定义了“ERI 曲线”这一概念。“ERI 曲线”是 ERI 图像的一维表示，也就是通过一根曲线来类比 ERI 图像的二维平面，由此简化问题的说明过程。同样，高斯模型也被简化为单坐标表示。

图 3-16 表示了一个高斯模型的例子。图 3-17 依次表明了利用该高斯模型增强纹理细节的过程。

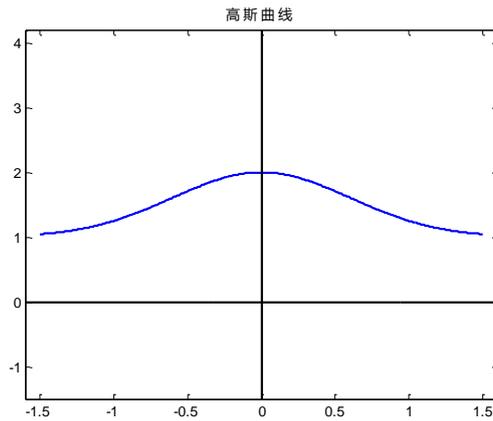


图 3-16 高斯曲线示例

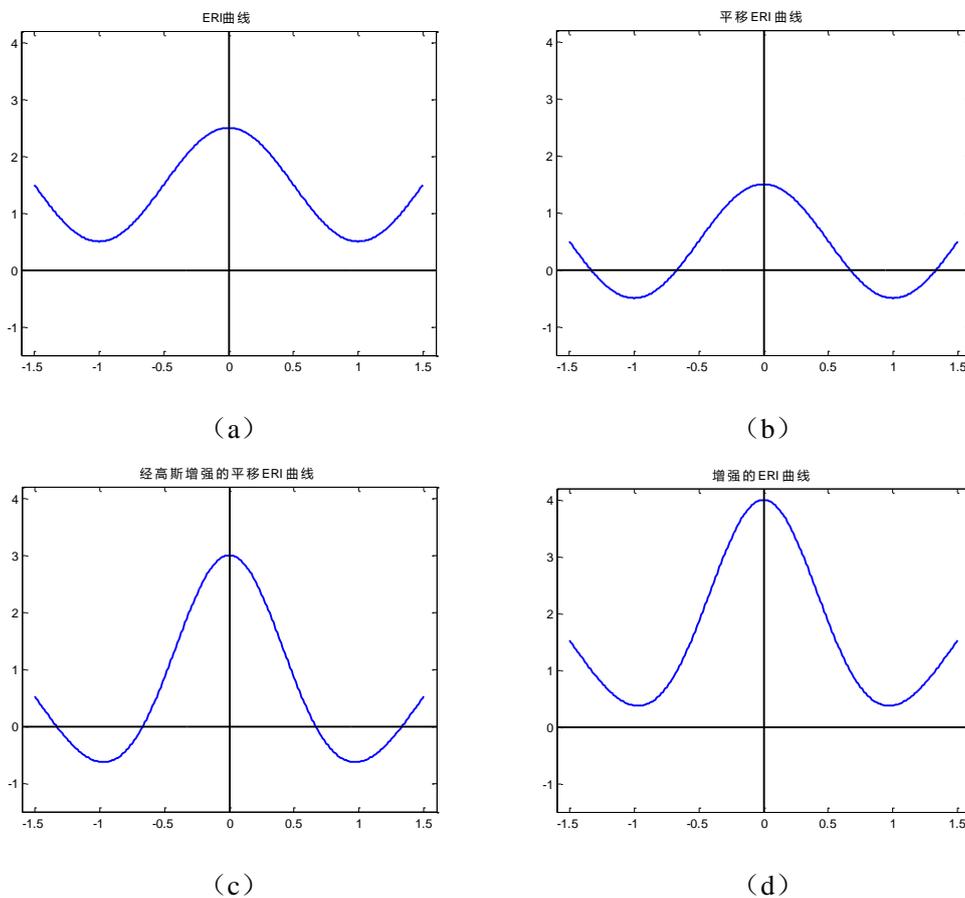


图 3-17 增强纹理细节过程示意

图 3-17 (a) 为原始的 ERI 图像，而 (d) 为增强后的最终结果。从结果中可以明显看到，ERI 图像在中心部分得到了最大程度的强化，并且这种强化随着离开中心点越来越远而逐渐趋缓。强化的总体效果是使 ERI 曲线凸起的部分更加凸起、凹陷的部分更加凹陷。因此，从这个例子中可以看出，高斯模型确实能够起到增强 ERI 图像的作用，从而增强表情变换中的纹理和光照细节。

3.5 本章小结

在本章的 3.1 节中，我们首先根据第二章介绍的理论知识，将传统带纹理细节的二维表

情变换算法具体实现了出来。在实现的过程中，我们遇到了各种各样的问题，并面临很多种情况的选择。通过实践、分析和比较，我们提出了一系列解决方案，有效地解决了这些问题，让结果尽量符合理论预期。之后，我们提出了一种新的模型，通过几何变换和纹理细节两方面对传统的表情移植方法进行了算法上的优化。在 3.3 节中，我们主要介绍了人脸的弹性模型，通过这个模型将全局和局部几何变换的结果结合了起来，实现了几何上的优化；在 3.4 节中，我们介绍了基于肌肉分布的纹理增强模型。通过选择“纹理强化圆”的位置，利用高斯模型对 ERI 图像进行了强化。在下一章节中，我们将会通过一系列实验比较这些算法的效果，并提出评价标准评判算法的优劣程度。

第四章 实验结果比较与评价

在本章中，我们首先会根据第三章提出的新模型，分别给出（1）利用人脸弹性模型得到的表情变换结果以及（2）利用肌肉分布模型得到的结果，并将这两个新结果与传统的方法做比较。其次，我们会将新提出的模型与文章[9][10]中的主要算法做比较，通过主观的观察和描述对不同的实验结果进行对比，作出初步优劣判断。另外，我们还会从主观和客观两个方面分别提出评估表情变换算法的标准，通过这些标准证明我们提出的新模型的优越性。

4.1 基于“弹性和肌肉分布模型”的实验结果及比较

基于人脸弹性和肌肉分布模型的表情变换算法包括两个方面，即对几何变形的优化和对纹理细节的提升。前者利用特征点之间的弹力将“全局”和“局部”变换结果结合起来，后者利用“高斯模型”对 ERI 图像进行增强。实验结果将在本节分别给出。

4.1.1 基于人脸弹性模型的实验结果

同样以图 3-1 所示的表情为例，我们首先采用弹性模型对图 3-1 (c) 中的中性表情进行几何变换，得到平衡了“全局”和“局部”几何变换后的结果，显示在图 4-1 中。同时，为了方便比较，我们把图 3-1 (c) 通过“全局”变换和“局部”变换的结果一同显示在图 4-1 中。其中，图 4-1 (a) 为全局变换的结果，(b) 为局部变换的结果，(c) 为用弹性模型平衡后的结果。

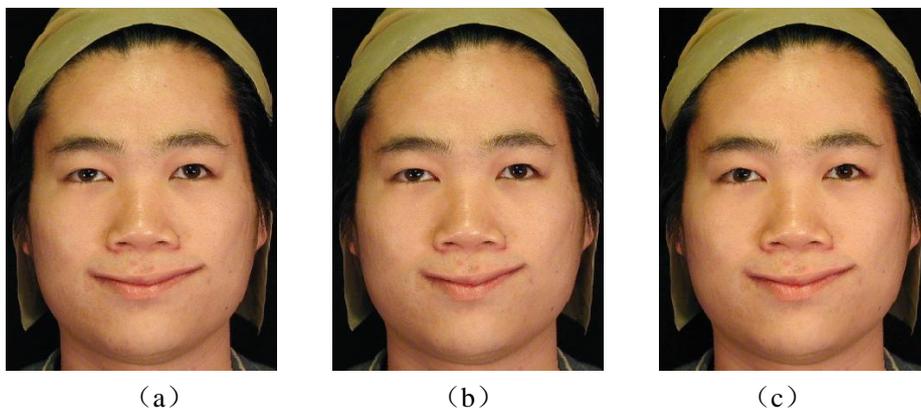


图 4-1 基于弹性模型的表情变换示例一

我们可以看到，(a) 中的嘴巴由于全局变换变得过宽了，相反的是，(b) 中的嘴巴由于局部变换变得过窄了。所以，在图 (c) 中，利用弹性模型平衡后的结果显得比较“适中”，既没有过宽也没有过窄，从整张人脸的大小来看显得较为协调。经过仔细观察，(c) 所示的结果比 (a)、(b) 都要来得自然可信一些。这证明了弹性模型确实能够解决全局变换和局部变换中存在的问题，而且用此模型仅仅通过几何变换就能取得不错的效果。

图 4-2 显示了用弹性模型对几何变换的结果进行平衡的又一个例子，同样是要将图 3-1 (b) 所示的模板表情进行移植。在图 4-2 中，(a) 为目标人脸的中性表情，(b) 为全局几何变换的结果，(c) 为局部几何变换的结果，而 (d) 为用弹性模型平衡后的最终结果。

从结果中可以清晰地看到，平衡后的结果 (d) 展示了一种最佳的效果，并且这个例子

比上一个例子更能说明问题。在图 4-2 (b) 中, 由于目标人脸在拍摄时脸部有一些倾斜, 但是在全局变换的结果中嘴巴仍然是水平的, 因此造成了“歪嘴”的现象; 而在图 4-2 (c) 的局部变换结果中, 问题同样出现在嘴巴过小过窄的现象上, 这是由于模板人脸的嘴巴本身比较小的缘故。同时, 目标人脸的眼睛本身是比较大的, 但是通过局部变换后, 眼睛也显得过于小了。无论是图 (b) 还是 (c), 各五官的形状和布局都显得很失调, 给人不自然的感觉, 更别说它们的逼真性了。不过, 这些问题在图 (d) 中都得到了很好地解决: 嘴的大小变得适中了, 也没有像 (b) 那样“歪”。同时, 眼睛的形状也得到了平衡, 在保持原来大眼睛的特点下, 又与真实微笑时应呈现出来的样子更接近了。总体而言, 利用弹性模型对几何变形进行平衡后, 移植的效果得到了明显的提高, 与原来的结果相比更加自然、更加可信。

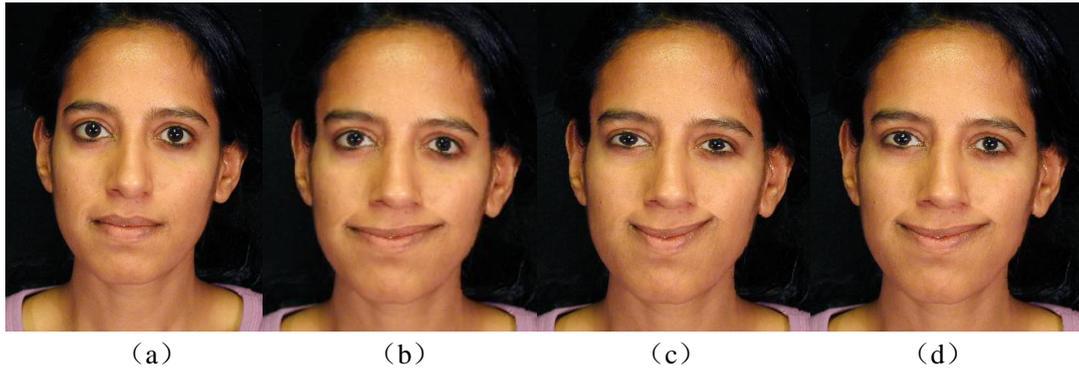


图 4-2 基于弹性模型的表情变换示例二

通过以上两个例子, 我们充分证明了人脸弹性模型对于平衡几何变换结果所起的积极作用。接下来, 将给出通过肌肉分布模型增强纹理细节的实验结果。

4.1.2 基于肌肉分布的纹理增强模型实验结果

利用 3.4 节介绍的增强 ERI 图像的方法, 我们将实验结果显示在图 4-3 中, 并与传统的用 ERI 图像移植纹理的结果进行了比较。其中, (a) 为“全局变换+普通 ERI”的结果, (b) 为“弹性平衡+普通 ERI”的结果, (c) 为结合了本文两个新模型的“弹性平衡+增强 ERI”的结果。

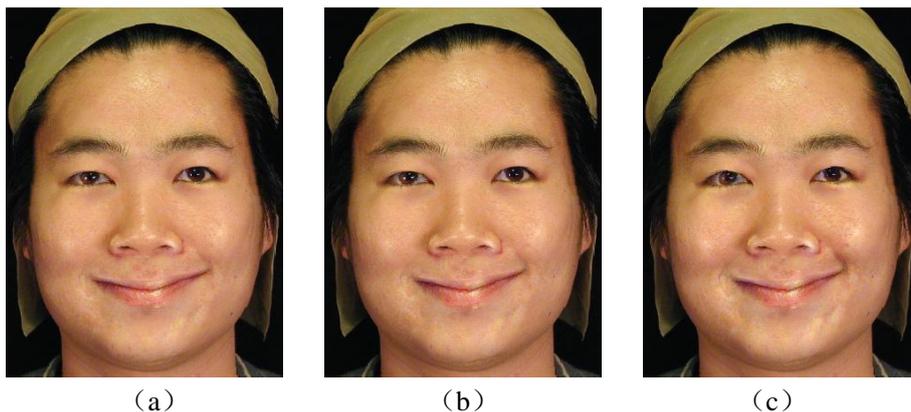


图 4-3 基于肌肉模型的表情变换示例一

同样, 在 (b) 与 (a) 的比较中, 由于 (b) 利用了弹性模型进行平衡, 所以在纹理细节相同的情况下, (b) 的结果显得稍好一些。不过, (a) 和 (b) 表情中的纹理细节都不够明显, 特别是右边嘴角处的这一皱纹, 作为最显著的特征, 与模板人脸比起来还是“稍欠火候”。在图 (c) 中, 这些皱纹被明显地加深了, 我们可以清晰地看到图片右侧 (人物左侧) 的嘴角处出现了一条深深的皱纹, 而且脸颊上皮肤的凸起感也更加强烈了, 整个结果显得更为生动有趣。

同样，基于图 4-2 (a) 的中性人脸，我们给出了第二个利用肌肉模型增强纹理细节的例子，如图 4-4 所示。同样地，(a) 表示了“全局变换+普通 ERI”的结果，(b) 表示“弹性平衡+普通 ERI”的结果，(c) 表示“弹性平衡+增强 ERI”的结果。

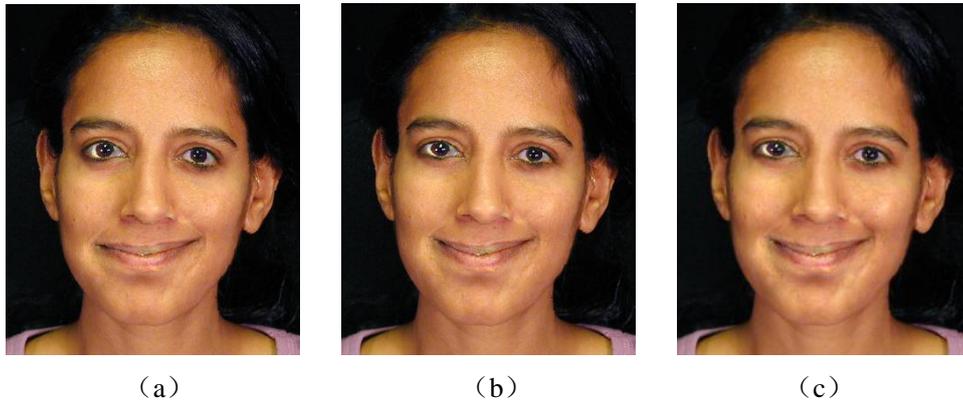


图 4-4 基于肌肉模型的表情变换示例二

这个例子同样说明了肌肉分布模型对纹理细节的增强效果。与上一个例子相似的是，比较结果 (b) 和 (c)，可以发现嘴角处的皱纹确实变深了，脸颊部分也更富有立体感。除此之外，下巴处的纹理细节也有略微的增强。种种变化，都给表情的真实性和生动性增添了一份色彩。

以上两个例子均说明了基于肌肉分布的纹理增强模型的有效性。至此，我们通过实验结果证实了本文提出的新模型确实能在几何上和纹理上给二维表情变换带来更好的效果，使这种表情移植的算法更适合投入到实际的应用中去，包括人机交互、远程会议、虚拟教学和电视电影制作等。这个新模型是本文的主要创新点和贡献，是本文最重要的内容之一。

4.2 利用弹性和肌肉分布模型实现二维表情变换的其他示例

本节中，我们将会给出更多的用本文提出的新模型进行人脸表情变换的示例，并且将这些结果与文章[9]和[10]中的主要结果进行对比，得出各个算法在表情移植中的优劣对比。

“表情比例图像 ERI”这一概念就是在文章[9]中提出的。在本文中，“全局变换+普通 ERI”的变换结果就对应了文章[9]中的算法，例如图 4-3 和 4-4 的图片 (a)；文章[10]的核心方法是对人脸进行三维建模，通过对三维模型的操作，实现人脸表情的变换过程。对于二维图像，文章[10]利用了人脸某部分灰度值的深浅信息将纹理细节也建成三维模型。对文章[10]提出的方法更具体的讨论可参见绪论 1.3.2，本节中不再赘述。

图 4-5、4-6、4-7、4-8 为利用弹性和肌肉分布模型实现二维表情变换的其他四个例子，将分别进行讨论。





图 4-5 表情变换结果一

图 4-5 (a) 为模板表情，我们需要将这个表情进行移植。(b) 为要移植的目标人脸的中性表情，(c) 为用文章[9]提出的用“全局变换+ERI”方法得到的结果，(d) 为用本文提出的弹性和肌肉分布模型进行变换得到的结果。从 (c) 和 (d) 两个结果中可以看出，用我们的方法得到的表情变换结果与模板表情更像。这两个结果最大的区别在于嘴的形状以及眼睛的形状，还有眉毛之间和嘴角边的纹理细节的差异。在图 (d) 中，嘴巴相比于图 (c) 更有往上翘的感觉，眼睛也更像“眯起来”的样子。自习观察的话，还能发现嘴角和两眉间的纹理显得更加清晰和深刻。在这个例子中，我们的方法取得了比文章[9]的方法更好的效果。

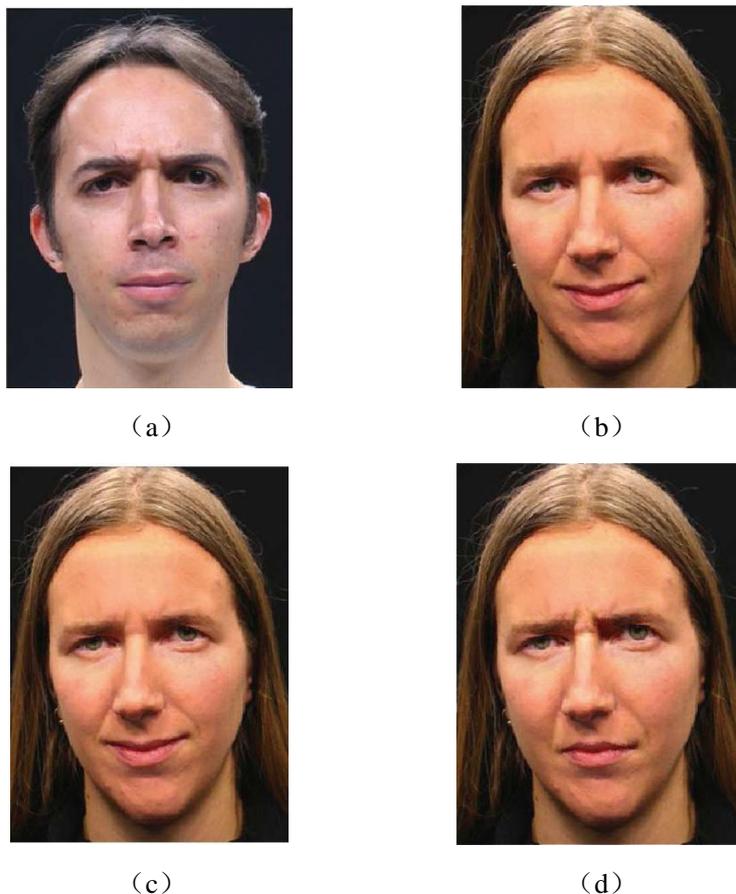


图 4-6 表情变换结果二

在图 4-6 中，(a) 为模板表情，(b) 为用文章[9]的方法得到的结果，(c) 是用文章[10]的方法得到的结果，而 (d) 是用本文提出的方法得到的结果。图 (b) 和 (c) 的效果几乎

没有差别，但是图（d）中的区别是非常明显的。首先，嘴型的区别表现地最为强烈。由于目标人物在中性表情中嘴角本就有一些上翘，所以文章[9]和[10]中的算法把这个问题也带到了结果中去。但是经过本文提出的方法的修正后，图（d）中人物的嘴型明显就回归到了与模板表情的嘴非常相似的地步，而且两眉之间肌肉隆起这一纹理细节也得到了一定程度的强化。总体来说，图（d）展示的结果与（a）的表情是十分相似的，并且做到了生动、自然，能让人一眼就相信这是个真实的表情而非合成的结果。所以，图 4-6 这一例子非常能说明问题。

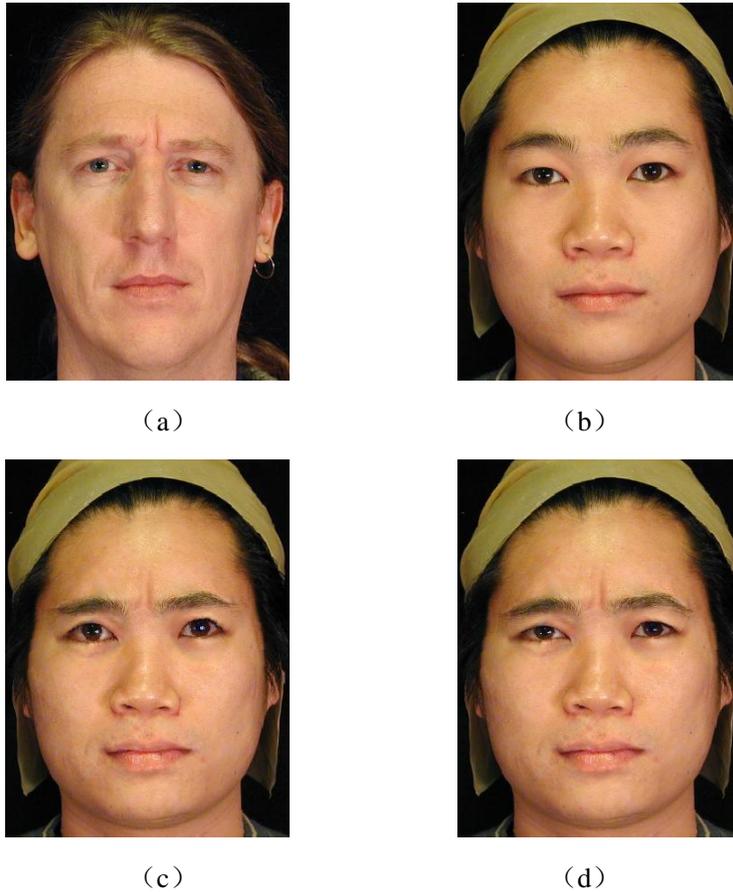


图 4-7 表情变换结果三

图 4-7 中，（a）为模板表情，（b）为目标人脸的中性表情，（c）是用文章[9]的方法得到的结果，而（d）是用本文提出的方法得到的结果。这个例子中的主要差别是眼睛的形状和眉毛间的皱纹差异。与（c）相比，（d）中眼睛的轮廓相对较小一些，并且眉毛之间的两条纹理也更深刻一些。



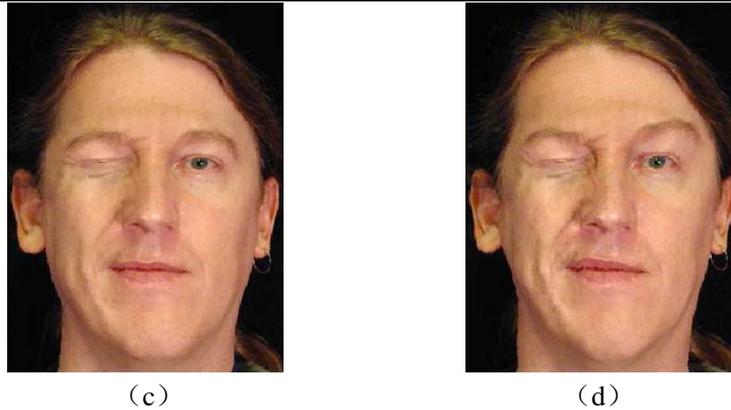


图 4-8 表情变换结果四

图 4-8 中，(a) 为模板表情，(b) 为目标人脸的中性表情，(c) 是用文章[10]的方法得到的结果，(d) 是用本文提出的方法得到的结果。对比图 (c) 和图 (d)，我们发现 (d) 中左侧（人物右侧）嘴角上扬的幅度更大，而且人物左边的眉毛上翘的程度也更大。这两个几何上的修正结果，都使图 (d) 中的表情与模板表情 (a) 更加接近。同时，人物右眼内侧的纹理也有明显的加深，使结果更加逼真。

4.3 表情变换的评价标准

在 4.2 节中，我们给出了一系列表情变换的结果，并且这些结果使用了不同的算法。为了比较不同算法的优劣程度，并且将这些比较结果用数字明确地表示出来，我们提出了基于“用户学习”的主观评价标准和基于“特征人脸”的客观评价标准两种方法来对表情变换的结果作出明确评价。

4.3.1 主观评价标准

主观评价标准是通过不同人的主观感受对结果作出评判的方法。本文涉及的人脸表情变换和建模是一个特殊的领域，因为这个领域中产生的结果大多是要直接“给人看”的，所以用户的主观感受成了评价算法优劣的一大依据。因此，我们提出了基于“用户学习”（User Study）的主观评价标准，来对实验结果作出一定意义上相对客观的评价。

基于“用户学习”的主观评判方法是，将这些结果分别给一定数量的用户看，用户根据自己的观察给这些结果打分，取平均分数作为最终的结果，得分越高就代表对应的算法效果越好。为了避免结果有失偏颇，用来给结果打分的用户必须达到一定数量。因此，即使每个用户的评价是主观的，但是将大量用户的评价平均之后，就能得到相对公平的评判结果。

在本实验中，我们一共让 20 位用户对这些表情变换结果进行了评价。首先，每位用户被要求独自观察这些结果，相互之间不能有交流。然后，我们要求他们给每种算法对应的结果打分，分数的范围是 1~5 分，1 分代表最差、5 分代表最好。最后，对所有用户的打分结果进行统计之后，显示在表 4-1 中。

从表格中看出，采用本文提出的基于弹性和肌肉分布模型得到的结果的平均分数最高，其次是文章[9]采用的“全局变换+ERI”的方法。效果最差的是仅仅在几何上进行全局变换的结果。从“用户学习”这个主观角度出发，证明了本文提出的新模型在带纹理的二维表情变换中的有效性。

表 4-1 基于“用户学习”的评价结果

表情编号	只全局变换	全局变换+ERI ^[9]	本文新方法
1	1.55	3.25	3.75
2	1.75	3.10	3.65
3	1.30	2.85	2.95
4	1.25	2.50	2.80
5	2.75	2.90	4.10
平均	1.72	2.92	3.45

4.3.2 客观评价标准

除了主观评价标准外，我们还采用了一种更客观的基于“特征人脸”(Eigen Face)的评判标准。“特征人脸”的概念与AAM模型(见第二章2.2.2)有些相似，都是通过“主成分分析法”(PCA)对人的脸部提取特征，并将提出的特征值组成一个“特征向量”(Eigen Vector)，用于接下来的人脸识别或表情识别过程。下面先对“特征人脸”作一个简要介绍^[19]。

特征脸方法是90年代初期由Turk和Pentland提出的目前最流行的算法之一，具有简单有效的特点，也称为基于主成分分析(principal component analysis, PCA)的人脸识别方法。该方法是先确定眼虹膜、鼻翼、嘴角等面像五官轮廓的大小、位置、距离等属性，然后再计算出它们的几何特征量，而这些特征量形成一描述该面像的特征向量。其技术的核心实际为“局部人体特征分析”和“图形/神经识别算法。”

特征脸还可被进一步分成特征子脸。特征子脸技术的基本思想是：从统计的观点，寻找人脸图像分布的基本元素，即人脸图像样本集协方差矩阵的特征向量，以此近似地表征人脸图像。这些特征向量称为特征脸。实际上，特征脸反映了隐含在人脸样本集合内部的信息和人脸的结构关系。将眼睛、面颊、下颌的样本集协方差矩阵的特征向量称为特征眼、特征颌和特征唇，统称特征子脸。特征子脸在相应的图像空间中张成子空间，称为子脸空间。计算出测试图像窗口在子脸空间的投影距离，若窗口图像满足阈值比较条件，则判断其为人脸。基于特征分析的方法，也就是将人脸基准点的相对比率和其它描述人脸脸部特征的形状参数或类别参数等一起构成识别特征向量，这种基于整体脸的识别不仅保留了人脸部件之间的拓扑关系，而且也保留了各部件本身的信息，而基于部件的识别则是通过提取出局部轮廓信息及灰度信息来设计具体识别算法。现在Eigenface算法已经与经典的模板匹配算法一起成为测试人脸识别系统性能的基准算法。而自1991年特征脸技术诞生以来，研究者对其进行了各种各样的实验和理论分析，FERET'96测试结果也表明，改进的特征脸算法是主流的人脸识别技术，也是具有最好性能的识别方法之一。

利用“特征人脸”算法，我们提出了一种基于特征向量间距离的评价标准。我们首先对模板表情和最后的移植结果分别提取特征向量，接着通过计算特征向量之间的欧几里德距离(Euclidean Distance)来衡量两幅图像的相似性，即特征向量间的距离越近，两幅图像越相似。最后，只要对比哪种算法得到的变换结果离模板表情的距离最近，就表明这种结果与模板最为相似，算法效果就最好。

我们假设 \vec{E}_{e_t} 为目标人脸表情移植后的特征向量， \vec{E}_{n_t} 为目标人脸中性表情的特征向量， \vec{E}_{e_s} 为模板表情的特征向量。这样，变换结果到模板表情之间的归一化距离可以用公式(4-1)表示。

$$D_{norm} = \frac{\|\bar{E}_{e_t} - \bar{E}_{e_s}\|}{\|\bar{E}_{n_t} - \bar{E}_{e_s}\|} \quad (4-1)$$

计算由各个算法得到的变换结果的归一化距离，结果列在表 4-2 中

表 4-2 各变换结果特征向量的归一化距离

	全局变换+ERI ^[9]	文章[10]的方法	仅用弹性模型平衡	本文新方法
例子 1	0.613	0.609	0.575	0.531
例子 2	0.986	0.977	0.949	0.946
例子 3	0.954	0.901	0.912	0.898
平均	0.861	0.827	0.836	0.819

从表 4-2 所示的结果中可以看出，本文提出的新模型得到的变换结果与模板表情的距离最近，因此这种结果是最接近“真实表情”的表情变换结果。因此，从“特征人脸”的客观角度出发，同样可以证明本文提出的新算法在提升表情变换结果上的有效性和优越性。

4.4 表情变换的应用实例

提出这种新的表情变换模型后，最终的目的是要将其运用到生活中的应用实例中去，例如非常流行的动画电影制作领域。本节给出了利用真实的人脸对虚拟形象进行“表情控制”的实例。如图 4-9 所示，为作者本人的中性表情和其他几种不同的表情。利用这一系列表情模板，我们将其移植到图 4-10 (a) 所示的虚拟角色“阿凡达”的脸部上去。4-10 (b) ~ (e) 为对应表情模板的表情变换结果。

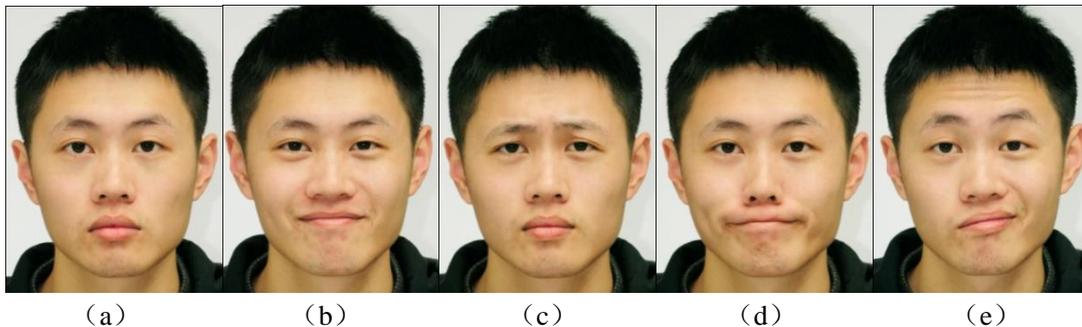


图 4-9 用于控制其余角色表情的模板表情

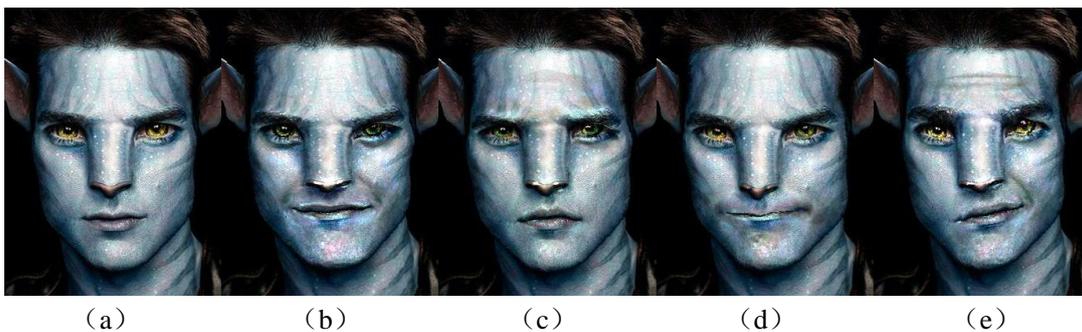


图 4-10 根据图 4-9 的模板得到的表情控制换结果

这种应用同样可以用到现在的电影制作当中。我们知道，目前的动画电影在制作时都需要真人带上一系列传感器来采集动作。传感器不仅价格昂贵，而且能在脸部放置的数量也不多。如果能够将本文讨论的方法提升到一个成熟的高度，就能投入到实际的动画电影的生产 and 制作中去了。

当然，本文曾提到过，ERI 图像对脸部色彩基调不同的人脸兼容性较差。图 4-10 确实反映了这个问题。最明显的问题体现在嘴和眼部的轮廓上，出现了一些蓝色的噪声。这也是本文今后需要改进的地方。

4.5 本章小结

本章展示了利用弹性和肌肉分布模型对人脸进行表情变换的实验结果，并且提出了主观和客观两种评价标准来评判算法的好坏。实验结果证明，本文提出的新模型能在几何和纹理细节上对传统的二维表情变换结果作出提升。在本节的最后，还给出了利用这种表情移植方法进行电影创作的实例，效果生动有趣。

第五章 总结与展望

5.1 论文总结

近年来,随着计算机技术的飞速发展,对这个真实的世界进行模拟和仿真成了越来越多人的梦想。人们试图建立一个虚拟世界,在那里可以实现一些现实生活中做不到的事情,让自己的感官和身心得到更大的满足感,同时使生活也变得丰富多彩,不再局限于以往单一的生活方式中。因此,计算机人工智能技术得到了极大的重视,其中,基于计算机图形学的人脸建模和动画技术也得到了飞速发展。对人脸的模型进行仿真建模,并且制作成动画,可以在电影和游戏等娱乐行业发挥极大的作用,同时在教育、通信、人机交互以及医学心理学领域也有用武之地。本文主要介绍了目前几种较为流行的人脸建模和动画方法,重点分析了带纹理细节的二维表情变换的算法依据,并在实践过程中将其实现了出来。

传统带纹理的二维表情变换算法确实能给表情移植带来更逼真、更生动的结果,让这种变换的结果更能让心信服。但是其中存在的问题导致了这种算法依然有很大的提升空间,包括几何形状上的修正的纹理细节程度的提升。基于这些问题,本文又提出了一个基于人脸弹性和肌肉分布模型的新算法,对传统的表情移植结果进行优化。具体地,本文的创新点和贡献可总结如下:

(1) 提出了人脸的弹性模型。在这个模型中,我们利用人脸特征点之间的弹力作用,对几何变形过程中的“全局”和“局部”变换结果进行了平衡,使特征点在这种情况下达到受力平衡。实验结果证明,人脸的弹性模型确实能够对表情变换的几何结构进行合理的修正。

(2) 提出了基于肌肉分布的纹理增强模型。在这个模型中,我们通过对传统人脸建模领域中脸部肌肉模型的分析,提出了“高斯模型”对原始 ERI 图像进行局部增强,从而提升纹理细节的表现效果。

另外,本文还通过几个例子,证明了新的模型能比传统的几种算法都取得更好的效果。同时,我们提出了基于“用户学习”的主观评价标准和基于“特征人脸”的客观评价标准两种方式对我们的算法进行评估。结果证明,无论是从主观角度还是客观角度,本文提出的新模型都能给表情移植带来更好的效果。因此,本文主要的创新点和贡献是有价值的。

5.2 工作展望

本文提出的算法虽然相比以往的方式能够取得更好的表情移植效果,但是在特定的应用中,算法本身仍然存在某些问题,或者应用条件有很大的限制。总结下来,本文在今后的研究中,还能在如下方面对算法进行改进:

(1) 本文在标定人脸特征点时采用的是手动标定方法,因此不能做到实时处理,而是需要大量前期准备工作。本文今后研究的一个方向可以是增强这种算法的实时性,即利用 AAM 模型等某些自动的标点方式,实现实时的表情控制,并将其真正运用到人际交互甚至是电影制作的过程中去。

(2) 本算法的另一个局限性是必须保证模板人脸和目标人脸的图像具有相似的光照强度,而不同的颜色基调将会导致一定程度的噪声。本文在 4.4 节给出的“阿凡达”应用实例也说明了这个问题。因此,本文今后研究的另一个方向可以是找出一种新的 ERI 图像,或

者提出新的对 ERI 图像进行滤波的方法，来增强表情移植过程对图像色彩和光照的兼容性。总之，算法仍有许多需要改进的方面。但不管怎样，我们有理由相信，只要今后研究者们继续努力，在表情变换领域提出更多更新、更好的方法，那么一定可以有所突破，再一次改变这个世界的生活方式。

参考文献

- [1] MEHRAB IAN A. Communication without words [J]. *Psychology Today*, 1968, 2 (4) : 53-56.
- [2] PARKE F. Computer generated animation of faces[C]. *Proc of ACM Annual Conference Boston, Massachusetts*:[sn], 1972:451-457
- [3] K. Waters, "A Muscle Model for Animating Three-Dimensional Facial Expression," *SIGGRAPH*, vol. 21, pp. 17-24, 1987.
- [4] P. Ekman and W. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto, 1978
- [5] T. Weise, S. Bouaziz, H. Li, and M. Pauly, "Kinect real-time performance-based facial animation," *SIGGRAPH*, pp. 1-8, 2011.
- [6] T. Beier and S. Neely, "Feature-based image metamorphosis," *SIGGRAPH*, pp. 35-42, July 1992.
- [7] T. Igarashi, T. Moscovich and J. F. Hughes, "As-rigid-as-possible shape manipulation," *ACM Trans. Graphics*, vol. 24, no. 3, pp. 1134-1141, 2005.
- [8] M. Alexa, D. Cohen-or and D. Levin, "As-Rigid-As-Possible Shape Interpolation," *SIGGRAPH*, pp. 157-164, 2000.
- [9] Z. Liu, Y. Shan and Z. Zhang, "Expressive expression mapping with ratio images," *SIGGRAPH*, pp. 271-276, 2001.
- [10] M. Song, Z. Dong, C. Theobalt, H. Wang, Z. Liu, H. Seidel, "Generic framework for efficient 2D and 3D facial expression analogy," *IEEE Trans. Multimedia*, vol. 9, no. 7, pp. 1384-1395, 2007.
- [11] S. M. Seitz and C. R. Dyer, "View morphing," *SIGGRAPH*, pp. 21-30, 1996.
- [12] R. W. Sumner and J. Popović, "Deformation transfer for triangle meshes," *ACM Trans. Graphics*, vol. 23, no. 3, pp. 399-405, 2004.
- [13] G. Edwards, C. Taylor, T. Cootes, "Interpreting face images using active appearance models," *International Conf. Face and Gesture Recognition*, pp. 300-305, 1998.
- [14] Timidsmile. AAM (Active Appearance Model)主动外观模型[OL]. 2011-09-17[2012-5-15]. <http://blog.csdn.net/timidsmile/article/details/6785744>.
- [15] SoRoMan. Delaunay 三角剖分 (Delaunay Triangulation) 相关知识 [OL]. 2007-05-17[2012-5-15]. <http://www.cnblogs.com/soroman/archive/2007/05/17/750430.html>.
- [16] K. Kahler, J. Haber, and H. P. Seidel, "Geometry-based muscle modeling for facial animation," *Graphics interface*, pp. 37-46, 2001.
- [17] S. Basu, N. Oliver, A. Pentland, "3D Modeling and Tracking of Human Lip Motions," *ICCV*, pp. 337-343, 1998.
- [18] PLATT S, BADLER N I. Animating facial expression[J]. *Computer Graphics*, 1981, 15(3):245-252.
- [19] 自由的风. 人脸识别主要算法原理 [OL]. 2011-12-21[2012-5-15]. <http://lovefreewind.blog.163.com/blog/static/178346357201112133030727>.
- [20] J. Noh and U. Neumann, "Expression cloning," *SIGGRAPH*, pp. 277-288, 2001.

谢辞

首先感谢我的指导老师林巍峤老师给我的细心指导，时刻关心我的毕业设计的进度，并且在某些问题的分析上给予了我极大的帮助，使我在遇到障碍时得到了不小的启发，指明了我在毕业设计中的研究方向，不断给我指点迷津。尤其是林老师严谨的治学态度给了我十分深刻的印象，他的言传身教将使我终身受益。可以说，林老师不仅是我学术上的良师，更是我生活中的益友。同时感谢实验室的其他老师，感谢他们在毕业设计的细节上给予了我详细的指导，也感谢平时生活中细致入微的关怀。

感谢范波、高鸿煜、陈远哲等师兄，在很多基本概念上给予了我指导，与你们的讨论同样使我在相关问题上有了更深刻的认识，使我受益匪浅。

感谢同班的同学在平时生活中给予我的支持。虽然我们毕业设计研究的课题不一样，但是在大家都要涉及的算法上，你们给予了我充足的指导。没有与你们的讨论，我不会经历一个这么有意义并且快乐的毕业设计过程。

最后要感谢我的父母，你们的忠厚善良、勤劳正直是我学习的榜样，你们在我成长过程中付出的心血一直以来是我前进的最大动力。

RESEARCH ON FACIAL EXPRESSION MODELING BASED ON ELASTISITY MODEL

In this paper, a new algorithm is proposed for facial expression mapping. The proposed algorithm first introduces a new elastic model to balance the global and local warping effects such that the impacts from facial feature differences between people can be avoided, thus more reasonable geometric warping results can be created. Furthermore, a muscle-distribution-based (MD) model is also proposed. The proposed MD model utilizes the muscle distribution information of the human face to evaluate and strengthen the facial illumination details. By this way, the impacts from human face difference as well as the effects of unsuitable noise filtering can be effectively alleviated. In addition, we also propose a new distance-based metric for objectively evaluating the performance of expression mapping results. Experimental results show that our proposed algorithm can create obviously better facial expression results than the existing methods.

Facial expression synthesis is of increasing importance in many applications such as movie making, video conferencing, and video games. In this area, expression mapping (or cloning) is one of the most effective techniques for synthesizing facial expressions. Basically, the target for facial expression mapping is to transfer one person's expression (i.e., the source person) to another person's neutral face (i.e., the target person), thus the second person's facial expression can be synthesized. Several algorithms have been proposed for expression mapping. Many algorithms use geometry warping or motion mapping on the identified face feature positions or the triangulated meshes to map facial expressions. For example, Sumner and Popović^[12] utilize nonlinear deformation transfer to map the 3D motions or expressions from one source object to the target. Song et al.^[10] use vertex tent coordinate transfer for performing geometric warping based on 3D models. However, since these methods perform warping globally while the local facial feature differences between people (e.g., face shape differences, mouth or eye differences) are not well considered, their facial expression mapping results are less satisfactory in some cases. Noh and Neumann^[20] aim to refine the local parts of the face based on the muscle model of human's face^[3]. Although they can create good facial feature motions for the same person, their methods are less effective when applied to synthesize the expression of another person as the facial feature differences between people are still neglected.

Also, since the facial illumination details will also change with different expressions, these detailed illumination information also needs to be transferred to the target person for creating more vivid expression results. However, most existing algorithms only focus on the geometry warping of the face while the illumination details are ignored. Although some researchers introduced expression ratio image (ERI)^[9] or mesh image^[10] to transfer the illumination details, their methods still have limitations due to face feature difference, unsuitable noise filtering, or detail importance differences. Furthermore, almost all existing algorithms evaluate the facial expression mapping results in subjective ways while objective assessment methods are still

lacking. Therefore, it is also important to develop objective metrics for evaluating facial expression mapping results.

In this paper, a new elastic-plus-muscle-distribution-based (E+MD) algorithm is proposed for facial expression mapping. Our paper has the following three contributions: (1) We propose a new elastic model to balance the global and local warping effects such that the effects from facial feature differences between people can be avoided. Thus more reasonable geometric warping results can be created. (2) We propose a muscle-distribution-based (MD) model which utilizes the muscle distribution information of the human face to evaluate and strengthen the facial illumination details. By this way, the effects from human face difference as well as the impacts of unsuitable noise filtering can be effectively reduced. (3) We also propose a new distance-based (DB) metric for evaluating the performance of expression mapping results together with the method of “User Study”. The proposed DB metric evaluates the facial expression mapping results by calculating the normalized expression distance between the synthesized target person’s expression face and the source person’s expression face. As far as we know, this is the first objective metric for evaluating the facial expression mapping results. Experimental results demonstrate the effectiveness of our proposed method.

As aforementioned, geometric warping can be used for creating the target person’s expression face. In the geometric warping process, the face feature positions are first identified either automatically or manually for each face. Then, triangulation can be performed for creating the triangle meshes according to these feature positions. Based on the triangulated mesh information, the geometric warped result of the target expression face can be achieved. Based on these methods, we can synthesize expressions such as “smile” from one person to another. However, the traditional warped expression result is unsatisfactory since it may look far from the expression “smile”. This is because most of the existing geometric warping algorithms are performed “globally” where the feature positions in the target expression image are moved “relatively” according to the movements of their neighboring feature positions as well as the movements of the source person’s feature positions. Since people may often have very different local features on their faces, these “relative” movements may often fail to create vivid expression results. For example, if the eyes of the source person are much smaller than those of the target person, it will be very difficult to “close” the target person’s eyes by global warping as the relative movements of the source person’s eyes are small.

In order to overcome the problem of global warping, we propose a “local” warping method which includes the following two rules: (a) Each organ on the target person’s face (e.g., eye and mouth) is warped independently without considering their relationship with the rest of the face. (b) The feature positions for each organ are warped in a way such that its absolute shape is close to the organ in the source person’s expression face. Compared with global warping which try to make the relative neutral-expression feature-position movement to be the same, the local warping try to equalize the “absolute” organ shapes between the source and the target expression faces. By this way, even if organ appearances are different between people, the organ in the target person’s expression face can still be warped similar to that of the source person. However, since the local warping method does not consider the relationship among different organs or the neutral-expression relativeness among organs, the local warped results still look unnatural. For example, the mouth may be too small or too big and incoherent with the target person’s face. Therefore, new methods need to be proposed to combine both the local and global warping results.

In this paper, we propose a new elastic model to balance the effects of global and local warping. The proposed elastic model can not only create more vivid results for each local organ, but also keep the reasonable global relationship among organs. Furthermore, the detailed illumination information is also important in creating good target expression results. Liu et al.^[9] propose expression ratio image (ERI) which transfers the illumination detail information based on the relative pixel-value ratio between the neutral and the expression face images. Although this simple method can create effective results, it has the following limitations: (a) Since the facial features are different for different people, simply applying the relative pixel-value ratio from one face to another may not be able to transfer details efficiently. (b) Since ERI may include noise, the transferred illumination details have to be filtered. However, since the filtering in ERI does not consider the structural characters of the face, significant illumination details may be weakened after filtering. Although Song et al.^[10] improve ERI by mapping all details into a mesh image to reduce the facial feature differences, the facial structure characters are still not considered when evaluating illumination details, thus leading to unsatisfactory results in some cases. Therefore, in this paper, we propose a muscle-distribution-based (MD) model which utilizes the muscle distribution information to evaluate the importance of the facial illumination details. Since the MD model includes the muscle distribution to reflect the facial structure, the illumination details transferred by our MD model are more precise than the previous methods in [9,10]. Based on the above discussions, we can propose a new elastic-plus-muscle-distribution-based (E+MD) algorithm which uses the elastic model for geometric warping and the MD model for transferring illumination details.

In the experimental parts of this paper, we showed a lot of expression mapping examples and proved that our new E+MD algorithm really performs well. It can not only transfer the illumination details from one person to another, but also improve the mapping result both geometrically and texturally.