

# 拟人 Agent 面部复杂运动合成的研究与实现

周芳芳<sup>1</sup> 赵颖<sup>1</sup> 杨斌<sup>2</sup>

(中南大学信息科学与工程学院 长沙 410075)<sup>1</sup>

(清华大学计算机科学与技术系 智能技术与系统国家重点实验室 北京 100084)<sup>2</sup>

**摘 要** 在多维面部运动的形式化描述方法的基础上,分析了面部肌肉的运动特征和控制约束,提出了面部运动单元的概念,采用正向运动学和示教再现两种方法实现拟人 Agent 的基本面部运动,并通过运动合成产生复杂的面部运动。实验表明,正向运动学方法能精确控制面部运动并降低非自然表情的出现;示教再现方法自动产生面部运动简化了操作细节;运动合成则通过运动重用提高了面部运动生成的效率。

**关键词** 面部运动, Agent, 运动学, 面部运动综合

中图分类号 TP301 文献标识码 A

## Research and Implement of Facial Action on Human-like Agent

ZHOU Fang-fang<sup>1</sup> ZHAO Ying<sup>1</sup> YANG Bin<sup>2</sup>

(School of Information Science and Engineering, Central South University, Changsha 410075, China)<sup>1</sup>

(State Key Laboratory of Intelligent Technology and Systems, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)<sup>2</sup>

**Abstract** Based on a uniform model for facial action, the characteristic and restriction of facial muscles were analyzed. The unit of facial action was proposed to generate facial action by using kinematical method and teaching method. The complex facial actions can be synthesized by the basic units of facial action. Experiments show that the kinematical method reduces the probability of the unnatural expression by controlling muscles accurately, and the teaching method generates the facial action without considering the details, and the synthesized facial actions are generated effectively by reusing the units of facial action.

**Keywords** Facial action, Agent, Kinematics, Synthesized facial action

人脸面部运动的研究始于 19 世纪,法国神经学家 Duchenne 用电流刺激肌肉,观察面部表情的生成机制<sup>[1,2]</sup>。后来,研究人员通过分析人脸表情照片或录像,来了解人的面部运动机制。1978 年,Ekman 和 Friesen 研究了 6 种基本表情(高兴、悲伤、惊讶、恐惧、愤怒和厌恶),系统地建立了上千幅表情图像库<sup>[3]</sup>。1983 年至 1992 年,Summerfield 和 Brooke 等用录像技术分析脸部标志点,对唇动进行分析研究<sup>[4-6]</sup>。现在对面部运动的研究,主要应用于仿生机器人和虚拟人领域<sup>[7-9]</sup>,但研究的问题集中在改善面部运动形态的视觉效果上,缺乏面部运动的生成过程控制和面部运动的标准,因此很难自动衡量面部运动是否符合客观实际以及运动的状态是否符合要求<sup>[10-12]</sup>。

人脸面部运动由肌肉控制产生,不同的人,肌肉数量、连接关系和作用机制相似。本文从生理解剖的角度出发,根据人面部肌肉的控制量,提出多维面部运动的形式化描述方法<sup>[13,14]</sup>。对面部运动的肌肉运动特征和控制约束的研究,采用正向运动学和示教再现两种方法实现拟人 Agent 面部运动,并将简单运动合成为复杂的动作。实验表明,本文提出的方法互相补充,既能精确生成面部运动,又能实现运动的重

用,是一种高效的设计方案。

## 1 面部运动模型的设计

### 1.1 面部形态形式化描述及控制约束

定义面部形态空间 FS,空间中每个点  $r_i$  对应一个面部形态。 $r_i$  为 39 维向量,每个维度表示肌肉的控制参数,即  $FS = \{r_i\}, i \in R^d; r = \{FM, FE, FL, FJ, FT, FH\}$ ,它们分别表示脸部、眼部、嘴、下巴、舌以及头 6 个部分的控制参数<sup>[9]</sup>。其中  $FM = \{\Delta m_1, \dots, \Delta m_{23}\}$ ,  $\Delta m_i =$  肌肉伸缩强度/肌肉长度,表示肌肉的归一化伸缩强度。这 23 条脸部辐射肌有 4 个约束条件,其中 3 个是空间限制,分别是影响最大角、衰减的起始距离和结束距离;另一个是辐射肌的收缩范围,它与肌肉特性、肌肉长度等因素相关。人脸部肌肉收缩一般不会超过肌肉长度的 30%,因此可用 30% 作为一个控制上限。

$FL = \{fl\}$ , 嘴部口轮匝肌的椭圆收缩参数,  $fl > 0$ 。

$FE = \{P_{leyelid}, P_{nyelid}, \theta_{leye}, \theta_{ryeye}, \theta_{lseye}, \theta_{rseye}, fl_{leye}, fl_{ryeye}\}$ , 左、右眼皮的开合度  $P_{leyelid}, P_{nyelid} \in [0, 1]$ ; 左、右眼球的垂直旋转角度和水平旋转角度  $\theta_{leye}, \theta_{ryeye}, \theta_{lseye}, \theta_{rseye} \in [-45^\circ, 45^\circ]$ ; 眼部的环状眼轮匝肌椭圆收缩参数  $fl_{leye} > 0$ 。

到稿日期:2011-06-01 返修日期:2011-11-10 本文受国家自然科学基金项目(60433030)资助。

周芳芳(1980-),女,博士,讲师,主要研究方向为虚拟现实与可视化、人工智能和情感计算,E-mail:zff@mail.csu.edu.cn。

$FJ = \{J_1, J_2, J_3\}$ , 控制下巴颏下颌关节的 3 个方向的运动强度。  $J_1 \in [-0.3, 1]$ ,  $J_2 \in [-0.1, 0.1]$ ,  $J_3 \in [-0.3, 0.3]$ 。

$FT = \{t\}$ , 控制舌头上下运动的参数,  $t > 0$ 。

$FH = \{H_1, H_2, H_3\}$ , 控制头颅绕 X、Y、Z 3 个方向的旋转参数,  $H_1 \in [-45^\circ, 90^\circ]$ ,  $H_2 \in [-90^\circ, 90^\circ]$ ,  $H_3 \in [-45^\circ, 45^\circ]$ 。

由时间序列  $\{t_i\}$ ,  $0 \leq i \leq n$  和对应的面部形态序列  $r_i = [f_{s1i} \dots f_{s39i}]$  共同构成了动态面部运动的形式化描述。此模型可以生成符合人脸实际运动规律的、容易控制的面部运动过程; 并且它与人脸几何模型无关, 可重定位应用于不同的人脸几何模型上。

## 2 面部运动生成方法研究

### 2.1 面部运动的肌肉运动特征

肌肉的运动存在拉伸和放松两个过程。拉伸由肌肉收缩形成, 而放松是肌肉弹性的复位, 两者互为镜像, 都是非均匀变化。肌肉的运动类似于弹簧作用, 其速度曲线可表示为单峰钟型曲线, 如图 1 所示。本文采用两个不同方差的高斯曲线组成新的速度曲线, 如式(1)所示:

$$v(t) = \begin{cases} Ae^{-\frac{(t-\epsilon T)^2}{2\sigma_1^2}} - b_1, & t \leq \epsilon T, 0 < \epsilon < 1 \\ Ae^{-\frac{(t-\epsilon T)^2}{2\sigma_2^2}} - b_2, & t > \epsilon T, 0 < \epsilon < 1 \end{cases} \quad (1)$$

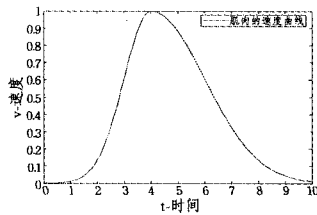


图 1 肌肉的速度曲线

### 2.2 基于正向运动学的面部运动生成

基于正向运动学的方法通过建立控制量与面部特征点的关系生成面部运动, 其由于操作简单、直观, 因此被广泛应用于三维建模软件。本文以面部运动的物理规律为基础, 通过减少控制对象(肌肉等)与被控制对象(面部特征点)的关系, 来降低非自然表情的出现概率。

首先将面部运动沿时间轴划分为简单的几段。然后根据面部形态, 选择相应的控制对象(辐射肌、轮匝肌等)逐一调整。当面部形态达到设计者要求时, 将控制量作为面部形态的形式化描述参数。重复上述方法, 可以得到静态面部形态序列。任意两个静态面部形态之间都可以插值获得新的面部形态和参数, 插值计算步骤如下:

第 1 步 在两段静态面部形态  $(r_{i-1}, r_{i+1})$  中间插入新的时间点  $t_i$ , 插值点处的面部形态参数  $r_i$  按式(2)计算:

$$r_i = (r_{i+1} - r_{i-1}) \times \frac{t_i - t_{i-1}}{t_{i+1} - t_{i-1}} + r_{i-1}, 0 < i < n \quad (2)$$

第 2 步 通过正向运动学方法, 将面部形态向量  $r_i$  展现在人脸几何模型上。

本文方法能精确地控制几何模型的空间形态以及转换时间, 生成的面部运动非常逼真, 比传统的动画制作方法更简单有效。

### 2.3 基于示教再现的面部运动生成

示教再现是一种典型的机器人控制方法, 本文将它应用

到拟人 Agent 面部运动生成中。通过设计者设定示教标记点, 让拟人 Agent 记住这种面部运动的行为特性, 实现自然的运动过程。运动生成过程分为以下 5 步。

#### 第 1 步 示教运动特征点和轨迹点

该步骤主要解决特征点的选择和轨迹点的示教两个问题。由于面部形式化模型对特征点没有特殊要求, 因此采用 MPEG-4 中的 FDP。如图 2 所示, 图中包含 36 个特征点, 基本覆盖了整个脸部。轨迹点可在模型所在的三维空间上标出。此方法适合某些运动幅度比较大的运动, 比如头部、颞大肌的运动等。

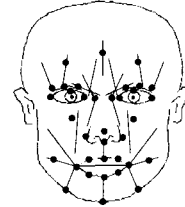


图 2 用于示教再现的面部特征点

#### 第 2 步 逆运动学方法求解控制量

问题转化为在时间  $t_n$ , 根据特征点  $P$  及其位移  $\Delta P$ , 应用逆运动学计算对于肌肉的控制量。

对于眼球、眼皮、下巴、舌、头部等较独立的部位, 通过正向运动方程的逆运算计算出面部形态控制参数。对于辐射肌, 如果特征点  $P$  仅落入一条肌肉内, 则计算方法同上。而如果特征点落入多条肌肉的作用范围, 则需找出最相近的两条肌肉  $M_{j1}$  和  $M_{j2}$ , 并计算各自的位移分量, 计算式如下:

$$\vec{\Delta P} = \sum_{i=1}^2 \lambda_i \Delta P_{mj_i} \vec{M}_{j_i} \quad (3)$$

式中,  $\Delta P_{mj_i}$  是点  $P$  在肌肉  $M_{j_i}$  的作用下沿肌肉  $M_{j_i}$  方向的位移, 如图 3 所示; 系数  $\lambda_i = 1.0$ 。然后通过逆运算得到两条肌肉各自的控制量。

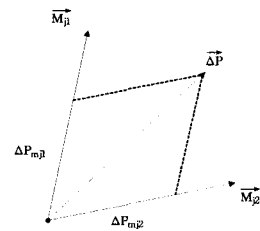


图 3 多条肌肉的共同作用

如果特征点被轮匝肌和辐射肌共同作用, 需要判断特征点运动方向。若位移方向靠近轮匝肌, 则通过其它特征点确定辐射肌的位移, 然后计算出轮匝肌的位移, 最后利用逆运算得到轮匝肌的控制量。如果特征点的位移方向远离轮匝肌, 此时只考虑辐射肌的作用, 直接采用逆运算计算辐射肌运动参数。依此类推, 将所有同时段发生的特征点变化全部用逆运动学方法计算一遍。综合所有控制量的值, 就可得到新的面部形态空间向量  $r_i$ 。

#### 第 3 步 拟合控制量变化曲线

假设示教再现的面部运动从  $t_1$  开始到  $t_n$  结束, 对应的面部形态向量为  $\{r_{11}, \dots, r_{n1}\}$ 。若每组有  $l$  个特征点, 则第  $i$  组的特征点  $P_{ji}$  的位移  $\Delta P_{ji}$  和控制量  $f_{sk}$  之间的关系为:

$$f_{sk} = f_{sp_{k,j}}(\Delta P_{ji}), 1 \leq k \leq 39, 1 \leq j \leq l, 1 \leq i \leq n \quad (4)$$

根据特征点  $P_j$  在  $n$  个示教时间点上得到的特征点运动

位移序列  $\{\Delta P_{j1}, \dots, \Delta P_{jn}\}$ , 可得到控制量  $f_{S_k}$  的变化序列  $\{f_{S_{k1}}, \dots, f_{S_{kn}}\}$ 。控制量  $f_{S_k}$  的拟合变化曲线为:

$$f_{S_k} = (f_{S_k} - f_{S_{k(i-1)}}) \frac{t - t_{i-1}}{t_i - t_{i-1}} + f_{S_{k(i-1)}}, 1 \leq k \leq 39, 1 \leq i \leq n, t_{i-1} \leq t \leq t_i \quad (5)$$

#### 第4步 生成运动模型

根据运动的精细度, 可生成等时间间距的时间序列  $\{t_i\}$ 。应用式(5)可得到面部形态向量的控制量, 将同样时间点  $t_i$  的控制量数据进行组合, 可得到该时间点对应的面部形态向量  $r_i$ 。综合时间序列  $\{t_i\}$  和对应的面部形态向量序列  $\{r_i\}$ , 可得到运动模型。

#### 第5步 再现示教动作所生成的模型

使用正向运动学方法, 按时间顺序播放面部形态向量序列, 可再现示教动作所生成的模型。

通过使用示教再现方法, 设计者可以不用考虑细节的操作, 大大提高了面部运动生成的效率。

### 3 面部运动合成方法研究

无论是基于正向运动学还是基于示教再现的运动生成方法, 都需要研究运动规律, 模拟运动过程。在现实生活中, 人的面部运动非常丰富, 如果都采用这两种方法, 即使是经验丰富的动画设计人员, 也需要耗费大量的时间。

本文提出面部运动单元的概念, 即通过组合由正向运动学和示教再现方法产生的多个基本面部运动单元, 生成更多更复杂的运动, 而且对面部运动的重用, 降低了重复劳动, 简化了面部运动的生成。

#### 3.1 面部运动单元

本文定义面部运动单元为动作序列及其结束状态。动作序列表示了面部形态变化的过程, 对应面部形态空间中的某段曲线; 结束状态表示运动变化的程度, 可认为是某个静态的面部形态。出现静态面部形态时, 面部运动曲线的一阶导数  $f_x'(t) = 0$ 。定义  $\Delta f_x(t) = \|f_x(t) - f_x(t - \Delta t)\|$ , 如果  $\Delta f_x(t) < 0.05$ , 则认为是一个运动的结束。从运动起始时间到  $t$ , 面部运动经历了一个面部运动单元。

#### 3.2 面部运动合成方法

面部运动合成包括两类操作: 乘法和加法。

##### 3.2.1 乘法 MUL

乘法 MUL 是实数与面部形态空间向量之间相乘, 一般用于运动强度, 且每个控制量的最终值要遵守肌肉运动的空间限制。设强度系数为正实数  $\lambda$ 。面部运动的时间序列为  $\{t_i\}$ , 对应的面部形态向量序列为  $\{r_i\} = \{[f_{S_{1i}}, \dots, f_{S_{39i}}]\}$ , 则强度系数与某个面部形态向量相乘, 过程如下:

- 1) 强度系数乘以面部形态向量, 等同于强度系数与面部形态向量中的 39 个控制量相乘;
- 2) 每个控制量的值的运算结果如果超过控制量的最大值, 则强制控制量的值为最大值。

由此, 可得新的面部运动, 其时间序列为  $\{t_i\}$ , 面部形态向量序列为  $\{\lambda r_i\} = \{[\lambda f_{S_{1i}}, \dots, \lambda f_{S_{39i}}]\}$ 。

##### 3.2.2 加法 ADD

加法 ADD 实现了两个面部动作单元之间的叠加, 其前提条件是两动作单元的时间序列一致。

假设 动作 1: 时间序列为  $\{t_i\}$ , 面部形态向量为  $\{r_i\}$ ; 动

作 2: 时间序列为  $\{t_i\}$ , 面部形态向量为  $\{r_i'\}$ 。

加法的实施过程如下:

- 1) 两个在等时间长度内的动作曲线相加, 可分解为对应时间的两个向量相加;
- 2) 向量相加可以分解为对应控制量的值相加;
- 3) 每个控制量的值的运算结果如果超过该控制量的最大值, 则强制该控制量的值为最大值。

拟人 Agent 可以通过有限的面部动作组合实现近乎无限的面部运动表达。

### 4 面部复杂运动的实现

本文采用 OpenGL 和 Visual Studio 2003 实现了人脸面部运动合成。首先采用正向运动学方法对人脸模型进行操作, 生成“高兴”、“悲伤”和发音“哦”3 面部运动单元。对于“高兴”, 面部各部分形态为嘴角向上翘起、抬额和眉毛扬起, 那么对应发生作用的肌肉有颧大肌和额肌。通过调节这些控制量生成的“高兴”的面部形态如图 4 所示。对于“悲伤”, 面部各部分形态为左右两眉靠紧, 向下眉间出现反八字形的皱纹, 眼睛一部分或全部闭上, 鼻子两侧绷紧, 变细稍稍变长, 唇角向下, 对应发生作用的有皱眉肌、眼皮、鼻翼内肌以及降嘴角肌。通过调节这些控制量生成的“悲伤”的面部形态如图 5 所示。对于“哦”, 它的发音对应的是元音“O”, 所以只对应一个面部形态, 相应的控制量为下巴和口轮匝肌。通过调节这些控制量生成的发音“哦”的面部形态如图 6 所示。



图 4 “高兴”



图 5 “悲伤”



图 6 发音“哦”

打呵欠这个动作虽然不复杂, 但是用正向运动学方法去控制生成还是有一定的难度, 因为这是一个连续的非线性运动, 它的特点是前快后慢。其用示教再现的方法比较容易实现, 因为这个运动只涉及到下巴。手工示教人打呵欠行为的下巴轨迹点, 分别在 0.75s、1.25s、1.5s、2.25s 以及 2.5s 标定 5 个关键帧, 求出下巴、眼皮控制量为:

下巴控制量  $H_1$  的值: 0.4, 0.8, 0.95, 1, 0

左眼皮闭合度  $P_{leyelid}$ : 0, 1, 1, 1, 0

右眼皮闭合度  $P_{reyelid}$ : 0, 1, 1, 1, 0

生成的面部序列如图 7 所示。



图 7 基于示教再现生成的打呵欠面部运动序列

人在不同的情感状态下与别人交互, 即使是发同样的音, 由于情感的影响, 面部表达也不一样。比如人在交往的过程中, 有时候为了方便, 答应别人的时候通常说“哦”。以生成的 3 个面部运动单元为例实现多运动合成。

1) 高兴的时候说哦。由于这两个面部运动都只有一帧, 因此合成运动为这两个对应面部形态向量的叠加。合成运动

(下转第 216 页)

Q 学习算法能够很好地适应动态的环境,更高效地收敛于最优策略。

## 参考文献

- [1] 徐昕. 增强学习与近似动态规划[M]. 北京: 科学出版社, 2010
- [2] Sutton S. Learning to predict by the methods of temporal difference[J]. Machine Learning, 1998(3):9-44
- [3] Watkins C J C H, Dayan P. Technical note: Q-learning[J]. Machine Learning, 1992, 8(3/4): 279-292
- [4] 张芳. 面向多移动机器人系统的再励学习方法研究[D]. 上海: 上海交通大学, 2002
- [5] 王长缨, 尹晓虎, 鲍翎平, 等. 一种共享经验元祖的多 agent 协同强化学习算法[J]. 模式识别与人工智能, 2005, 18(2): 234-239
- [6] 焦殿科, 石川. 共享经验的多主体强化学习研究[J]. 计算机工

(上接第 197 页)

也是只有一帧的运动, 其面部形态如图 8 所示。

2) 悲伤的时候说哦。这两个运动的合成与上一个实验类似, 合成运动的面部形态如图 9 所示。

3) 对于人的复杂情感, 比如悲喜交加, 这种情况下很难描述出它的面部形态, 也就很难应用正向的运动学控制生成方法和示教再现的面部运动生成方法。这种情况下, 将两个运动合成, 是一种较方便生成对应人复杂情感的面部运动的方案。定义悲喜交加为高兴与悲伤的混合, 其合成运动是只有一个关键帧的运动。其关键帧的面部形态如图 10 所示。



图 8 高兴地说“哦” 图 9 悲伤地说“哦” 图 10 “悲喜交加”

**结束语** 本文在多维面部运动的形式化描述模型中, 提出了运动的控制约束条件, 然后采用了正向运动学和示教再现两种生成面部运动的方法。正向运动学方法分段控制运动, 较精确地生成了各种静态面部形态以及动态面部运动, 减少了非自然表情出现的概率。示教再现的方法根据面部运动的轨迹点, 采用逆运动学方法计算目标点对应的面部形态向量, 拟合出向量运动曲线; 再通过正向运动学的方法进行展现, 使设计者可以不用考虑操作细节; 最后, 将产生的基础运动单元合成复杂面部运动生成, 简化了面部运动的生成过程, 提高了面部运动生成效率。实验表明, 本文方法不仅有助于生成逼真的表情, 还有助于弥补过程控制缺乏的问题, 帮助人们生成了面部运动过程的中间细节, 实现了运动的重用。

## 参考文献

- [1] Duchenne G B, Cuthbertson R A. The Mechanism of Human Facial Expression[M]// Cuthbertson R A, ed. Cambridge: Cam-

bridge University Press, 1990

- [2] Gray H, Lewis W. Anatomy of the Human Body [M]. New York: Bartleby, 2000
- [3] Ekman P, Friesen W V, Hager J. Facial Action Coding System [M]// Alto P. CA: Consulting Psychologists, 1978
- [4] Brooke N M, Summerfield Q. Analysis synthesis and perception of visible articulatory movements [J]. Journal of Phonetics, 1983, 11(1): 63-76
- [5] Summerfield Q. Roles of the Lips and Teeth in Lipreading Vowels [C]// Proceedings of the Institute of Acoustics, 1984
- [6] Summerfield Q. Lipreading and Audio-Visual Speech Perception [J]. Philosophical Transactions on Biological Sciences, 1992, 335: 71-78
- [7] Breazeal C, Scassellati B. How to build robots that make friends and influence people [C]// International Conference on Intelligent Robots and System. Piscataway: IEEE Press, 1999: 858-863
- [8] Breazeal C, Fitzpatrick P. Social amplification of animate vision [C]// Proceedings of the AAAI Fall Symposium on Society of Intelligence Agents—The Human in the Loop. AAAI Press, 2000
- [9] Hegel F, Muhl C, Wrede B. Understanding Social Robots [C]// Second International Conferences on Advances in Computer-Human Interactions, 2009: 169-174
- [10] 李旭东, 张振跃. 人脸表情的形变线性拟合方法 [J]. 自动化学报, 2008, 34(5): 593-597
- [11] Lu Xiao-guang, Jain A K. Deformation Modeling for Robust 3D Face Matching [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(8): 1346-1357
- [12] Caramero L, Fredslund J. I show you how I like you—can you read it in my face? [J]. IEEE Transactions on Systems, Man and Cybernetics, 2001, 31(5): 454-459
- [13] 吕梦雅, 张志刚, 唐勇, 等. 基于肌肉模型的眼睛动画研究 [J]. 系统仿真学报, 2008, 20(20): 5573-5580
- [14] Yang B, Jia P. A Facial Expression Model for Human-like Agent [C]// Proceeding of 2006 IEEE Conference on Robotics, Automation and Mechatronics, 2006: 1-6