论 文

基于人脸运动捕捉的表情动画仿真研究

方小勇^{①*}, 魏小鹏^②, 张强^{②*}, 周昌军^②

① 湖南工学院计算机与信息科学学院, 衡阳 421002

② 大连大学先进设计与智能计算省部共建教育部重点实验室, 大连 116622

* 通信作者. E-mail: fangxiaoyong@dlu.edu.cn; zhangq@dlu.edu.cn

收稿日期: 2012-06-08; 接受日期: 2013-01-28

国家自然科学基金 (批准号: 60875046)、教育部科技研究重点项目 (批准号: 2009029, 211036)、辽宁省高等学校优秀人才支持计划 (批准号: LR201003)、辽宁省高校重大科研计划 (批准号: LS2010008)、辽宁省高校科研项目计划 (批准号: 2009S008, 2009S009)、辽宁省高校创新团队支持计划 (批准号: 2009T005, LT2010005)、辽宁“百千万人才工程”培养经费 (批准号: 2010921010) 和湖南省科技厅科技计划 (批准号: 2011FJ3108, 2012FJ3034, 2013FJ3095) 资助项目

摘要 基于运动捕捉的动画仿真技术是当前计算机动画的研究热点. 基于光学式运动捕捉系统, 文中提出了一种新颖的基于交叉映射的人脸表情动画重构方法. 针对 RBF 构建静态全局映射时嘴唇上下部分光标出现的伪关联问题, 文中构建了基于功能分区的 RBF 交叉映射方法. 在模型驱动过程中, 提出一种皮肤运动机理计算得出增强标记点运动, 基于增强标记点实现了基于人脸分区驱动的 RBF 插值方法, 改进了基于人脸分区模型驱动的仿真效果. 此外, 设计了预计算算法提升了交叉映射和动画仿真过程的实时效率. 实验证明该方法能够将同一演员的运动捕捉数据应用于任意不同人脸模型, 针对相应模型生成逼真的人脸表情动画.

关键词 表情动画 交叉映射 功能分区 辅助标记点 RBF

1 引言

人脸表情仿真研究当前计算机科学领域的一个热点课题^[1~3]. 首先, 作为动漫产业体系中的一个关键课题, 人脸 (或者拟人) 的表情仿真在动画、游戏及影视制作中往往扮演着不可或缺的重要角色. 其次, 在虚拟现实、认知科学、人机交互界面设计以及网络视频会议等研究领域, 人脸表情仿真的相关研究越来越受到关注. 人脸表情动画的真实性是表情仿真最为关键的评价指标之一, 在传统的表情动画制作过程中, 动画的逼真效果往往需要较大的人工干预作为代价. 因此, 真实性和高效性仍然是当前人脸表情仿真研究的主要指标.

运动捕捉 (motion capture: MoCap) 技术是近年来发展起来的一种新兴的数据获取手段, 能够实时记录和还原表演者的运动. 人脸运动捕捉技术是将运动捕捉技术运用于人脸表情动画制作, 相比于传统的表情动画制作, 真实性和高效性的优势都较为明显, 因此, 近年来运动捕捉技术在人脸表情仿真研究领域越来越受到关注.

在人脸表情动画仿真的过程中, 如何使用单个演员的表情数据驱动不同人脸模型, 是当前人脸仿真领域研究的一个重要课题. 在运动捕捉领域, 对不同人脸表情运动表情数据进行映射配准, 以保证

引用格式: 方小勇, 魏小鹏, 张强, 等. 基于人脸运动捕捉的表情动画仿真研究. 中国科学: 信息科学, 2013, 43: 982-995

同一演员的运动捕捉数据能够驱动其他人脸模型, 这样的技术被称之为交叉映射 (cross-mapping) 或运动配准 (motion retargeting). 例如在电影《金刚》、《指环王》等, 如金刚、咕噜等面部结构怪异的角色模型动画都是这类技术生成. 在学术界, 有许多学者对交叉映射方法进行了研究, 目前主要分为四类:

i) 基于局部几何结构的方法: Noh 和 Neumann^[4] 根据源模型与目标模型对应顶点的几何方位关系以及局部结构的不同确定该点运动的方向与幅度大小. 该方法直观性强, 但对于不同模型在实时处理过程中计算量是不确定的^[5].

ii) 基于 RBF 全局非刚性映射方法: Lorenzo 等^[6] 使用 RBF(radial basic function) 方法对运动序列的首帧模板与不同人脸的标记点位置信息建立了全局映射, 针对不同人脸模型生成了运动数据. Edge 等^[7] 同样采用了该方法进行人脸表情运动捕捉数据的映射配准.

iii) 基于模型参数化方法: Imageworks 公司动漫电影《怪兽屋》的制作过程中, 人脸表情运动捕捉数据与 FACS 的表情单元建立对应关系构成交叉映射^[8]. Curio^[9] 提出一种基于 FACS 原理的人脸表情运动交叉映射方法. Deng 等^[10] 利用 46 个变形融合参数表征人脸的表情变化, 针对精心挑选的多帧运动捕捉数据构建了与这些参数的匹配关系. 然而, 这种方法受限于多帧运动捕捉数据的挑选以及变形融合参数的定义. 姚健等人^[11] 提出一种基于球面参数化的交叉映射算法.

iv) 基于模板加权方法: Chai 等^[5] 提出一种实时处理过程中计算量恒定的方法, 利用一组变形基 (deformation bases) 线性组合任意表情运动数据. Chuang 和 Bregler^[12] 将模板对应的人脸模型与模板运动捕捉数据进行映射, 分解人脸表情运动捕捉数据并求得这些加权系数, 用来驱动人脸模型. 该方法主要受限于模板的分类精细程度, 模板过少容易造成失真, 模板过多易造成约束的过饱和, 最终造成帧间跳跃.

以上总结了人脸表情运动捕捉数据的交叉映射研究, 相应地针对人脸几何模型的表情交叉映射 (也叫表情克隆技术: expression clone) 也有许多学者进行过研究, 取得了丰硕的成果^[13~20].

在人脸仿真研究领域中, 模型驱动是生成动画的重要部分. 人脸表情动画仿真的方法多种多样. 逼真的人脸表情动画不仅需要构建真实感强的外观模型, 包括人脸网格模型及相应的纹理材质贴图, 还需要构建一个底层的运动机制, 用于处理运动捕捉数据、利用数据驱动模型产生动画. 当前基于运动捕捉数据的人脸模型驱动方法主要可以分为三类:

i) 基于面片模型的方法: 目前较多数方法使用一个几何面片或者网格模型构建人脸, 以实现表情仿真, 这类方法模型构建简单, 但驱动算法较为复杂. Edge 等^[7] 提出了一种基于基于 BIDs(Bezier-triangle induced deformation) 的人脸模型变形方法. 根据标记点构建了一个 Bezier 曲面模型, 将几何模型的点映射到曲面模型实现模型的驱动. Park 和 Hodgins^[21] 利用多步变形实现了体表皮肤的驱动. 首先基于二次变形方法^[22] 进行初步变形, 然后, 采用 RBF 插值方法完成余下空间的变形. Guenter 等^[23] 提出对相邻标记点运动进行加权的方法来驱动模型的顶点. 浙江大学的 Wang 和 Zhuang 等^[24,25] 提出了使用测地距替代欧式距离 (Euclidean distance) 来构建开放结构 (例如嘴的上下嘴唇) 的非连续性, 利用贝叶斯退化算法驱动人脸模型产生动画.

ii) 基于多层物理模型的方法: 这类方法将皮肤、肌肉和骨骼分开考虑, 模型构建复杂, 原理简单直观. Sifakis 等^[26] 根据人脸解剖结构, 建立了由骨骼、肌肉以及皮肤组织构建的人脸实体模型, 利用有限元方法驱动人脸模型生成动画. 然而, 该方法模型构建非常繁琐, 计算消耗很大. Somasundaram 与 Parent^[27] 使用一种简明的物理模型. 首先将标记点映射到相应肌肉, 再通过肌肉运动映射到人脸模型, 利用运动捕捉数据计算出肌肉的运动驱动模型产生动画.

图 1 算法框架流程图

Figure 1 The flowchart of the framework

iii) 基于表情模板的方法: 这类方法普遍使用现有模板, 包括运动数据模板、人脸表情模型模板与当前运动数据构建对应关系, 从而实现低维到高维的映射. Chai 等^[5] 利用视频捕捉的少数特征点运动信息与多维的运动捕捉数据信息建立关联, 实现了视频驱动人脸模型生成动画. Joshi^[28] 通过计算和调整模板的加权系数就可以合成其他表情. Kyung-Gun 和 Moon-Ryul^[29] 也使用类似方法实现了对不同对象模型的表情动画. Takamizawa^[30] 等人基于 PCA 构建人脸子空间方法实现了人脸表情动画的合成.

从皮肤运动机理上说, 人脸有嘴、鼻、眼、眉等具备相对独立的组织, 表皮的结构存在一定的差异, 造成人脸运动具备区域性, 这是人脸表情运动精妙复杂的重要原因之一. 在利用运动数据驱动人脸几何模型的过程中, 针对人脸表情运动的区域性进行考虑, 这对于实现真实感表情动画是非常必要的.

综上所述, 基于人脸表情运动捕捉的真实感表情动画仿真是一项具有挑战的课题. 人脸表情动画是一项系统工程, 在仿真的真实性和高效性方面, 仍然具有较大的研究空间. 基于此, 本文提出了一种新颖的基于人脸表情运动捕捉系统的表情动画重构方法. 该方法将捕捉的表情运动数据运用到不同五官特征的人脸模型, 并致力于实现一种快速高效的模型驱动方法, 从而实现一个基于运动捕捉的人脸表情动画实现系统.

2 方法概述

本方法算法框架和流程如图 1 所示.

该方法的理论框架分为两个主要部分: 交叉映射构建及运动数据生成以及人脸表情动画 RBF 重构. 首先为构建一个基于针对运动捕捉数据的交叉映射, 结合人脸表情运动生理特点, 提出了一种基于表情功能分区的交叉映射算法. 基于 RBF 方法构建了目标人脸标记点与首帧模板数据的交叉映射, 计算出目标人脸标记点的运动数据. 在 RBF 重构表情动画过程中, 对人脸模型进行分区驱动. 针对相邻分区在分区结合处的驱动问题, 添加了增强标记点. 基于人脸运动机理计算求解增强标记点的运动,

图 2 人脸模型的表情功能分区

Figure 2 Expressional and functional partitions for the facial model

将增强标记点作为运动数据增强运动信息, 采用基于预计算的 RBF 插值方法驱动人脸模型. 在预计算处理模型驱动的过程中, 基于 RBF 构建了预计算方法; 另外, 讨论了模型法线同步计算方法, 实现了顶点法线的快速更新.

3 基于功能分区的交叉映射构建方法

运动捕捉技术应用不仅能够准确地捕捉和重构运动本身, 还能够运用统一的数据驱动不同模型产生相应的运动. 本文在人脸功能分区的基础上, 提出使用 RBF 网络训练及构建首帧模板与目标人脸标记点数据的映射关系, 并且将这种映射关系应用到其他帧的运动捕捉数据, 从而获取目标人脸的运动数据.

3.1 人脸表情功能分区及非连续性构建

人脸的表情运动具备区域性, 不存在运动关联的相邻区域 (如上下嘴唇分隔的两个区域) 的运动不存在连续性. 本文对人脸模型以及标记点进行分区构建人脸表情运动的非连续性.

事实上, 开放结构嘴和眼产生的区域性运动, 拓扑不相关的区域不大, 如图 2 所示. 根据区域性运动对人脸进行分区, 需要构建多个运动区域. 由于人脸模型的运动衍生自稀疏标记点的运动, 因此同样需要确认相应各个区域的对应稀疏标记点. 如果稀疏点个数不能满足要求, 各个相邻区域的拓扑关联处的运动连续性降低, 最终易导致模型驱动的失真. 考虑到运动捕捉数据驱动模型的实际情况, 分别可以采用两个分区构建嘴部和眼部的运动非连续性, 因此可将人脸区域分为三个不同功能分区, 如图 2 所示. 图中的三个区域按空间位置关系可以分为前额及眼上部功能分区、脸颊功能分区以及下颚功能分区. 本文方法针对嘴部运动的区域性, 将人脸分为两个主要分区: 下颚功能分区和脸颊及前额分区.

3.2 基于 RBF 构建分区交叉映射构建及数据生成

RBF 方法是被广泛用于人脸表情和皮肤模型动画的一种计算、插值和训练方法 [25,28,31~33]. 基于 RBF 构建的交叉映射从数学上是源数据与目标数据的高阶非线性插值关系; 本质上是使用 RBF 训练方法构建的一个固定的神经网络, 可以看做是一个固定的几何拓扑映射模型.

然而, 针对整个人脸构建交叉映射时, 需要考虑 RBF 方法的一个本质属性. RBF 方法是一种全局训练机制, 网络输入层中每个节点彼此都存在训练关系. 基于此, 如果针对整体人脸运动捕捉数据构建交叉映射, 人脸所有标记点都在本质上存在着相互关联, 这样就忽略了人脸开放性结构 (如嘴、眼等) 存在运动区域性的运动特性.

本文利用 RBF 方法构建交叉映射. 对人脸标记点根据 3.1 小节所述进行分区, 对每个分区的标记点集采用文献 [6] 中的方法构建交叉映射. 从而, 为下颚功能分区和脸颊及前额分区分别构建了相应的交叉映射.

我们根据下式构建映射关系:

$$\mathbf{m}_i^0 = \sum_{j=1}^n \lambda_j \phi(\|\mathbf{p}_i^0 - \mathbf{p}_j^0\|) + P \times \mathbf{p}_i^0, \quad (1)$$

其中 $\mathbf{m}_i^0$ 和 $\mathbf{p}_i^0$ 分别表示人脸运动捕捉数据模板和目标个性人脸模型的第 i 个标记点位置, n 为当前功能分区的映射标记点个数. P 为 RBF 的多项式项, 被认为是一个固定的仿射变换. 因此, 我们可以根据 (1) 式求出基函数权值矩阵, 仿射变换矩阵, 二者可以组合则可以构成 $(n+3) \times 3$ 的交叉映射矩阵:

$$\mathbf{T} = [\lambda_0, \lambda_1, \dots, \lambda_{n-1}, P]^T, \quad (2)$$

其中径向基函数选用逆 —MultiQuadric 基函数:

$$\phi(x) = \frac{1}{\sqrt{x^2 + r^2}}, \quad (3)$$

对于每个径向基函数, r 设置为当前标记点到其最邻近标记点的距离:

$$r_i = \min\|x_i - x_j\|, (i \neq j). \quad (4)$$

为使目标人脸模型产生运动捕捉数据所捕捉运动对应的表情动画, 目标人脸上标记点的运动是必须首先求解的. 根据 (2) 式的交叉映射矩阵, 我们利用当前帧 (第 t 帧) 运动捕捉数据构建系数矩阵:

$$\mathbf{A}^t = \begin{bmatrix} \phi_{0,0}^t & \cdots & \phi_{0,n-1}^t & \mathbf{p}_0^t \\ \vdots & \ddots & \vdots & \vdots \\ \phi_{n-1,0}^t & \cdots & \phi_{n-1,n-1}^t & \mathbf{p}_{n-1}^t \\ [\mathbf{f}\mathbf{p}_0^t]^T & \cdots & [\mathbf{p}_{n-1}^t]^T & 0 \end{bmatrix}. \quad (5)$$

令 $\mathbf{M} \in \mathbb{R}^{n+3,3}$ 为当前帧目标人脸标记点位置矩阵, 那么

$$\mathbf{M} = \mathbf{A}^t \mathbf{T}. \quad (6)$$

从 $\mathbf{M}$ 中可以提取出人脸标记点当前帧的位置信息: $\mathbf{m}_i^t = [\mathbf{M}_{(i,:)}]^T$, $\mathbf{M}_{(i,:)}$ 为矩阵的第 i 行向量.

图 3 相邻功能分区结合部出现运动不连续

Figure 3 The motion discontinuity at the junction area of two adjacent functional partitions

4 基于增强标记点的人脸表情动画 RBF 重构

在人脸表情动画仿真过程中, 如何通过稀疏的标记点运动信息驱动密集点几何模型产生真实连续的运动是必须克服的问题. 本文基于表情功能分区, 根据生成的运动数据对人脸模型进行 RBF 重构. 首先, 对标记点进行划分, 将可能影响各个运动功能分区的标记点都划分到相应功能分区的驱动标记点序列中. 其次, 添加增强标记点增强运动信息. 再次, 使用基于预计算的 RBF 方法驱动模型生成动画.

4.1 增强标记点

4.1.1 增强标记点的设置

设置添加增强标记点主要是出于两点考虑: 一方面, 由于嘴唇内侧及内部等地方无法进行捕捉, 因此我们对每个嘴唇添加三个增强标记点; 另一方面, 由于相邻功能分区存在具备拓扑关联的结合部, 且由于影响相邻功能分区的标记点不同, 结合部的运动容易出现不连续, 如图 3 所示.

综上所述, 我们对上下嘴唇分别添加三个增强标记点; 为确保结合部的运动连续性, 在相邻分区的结合部我们添加 8 个增强标记点, 总计添加增强标记点 11 个, 如图 4 所示. 图中所示增强标记点出现在嘴唇内侧以及脸颊与下颚功能分区的结合部, 索引前加文字标识“V”, 其余标记点为运动捕捉时贴在人脸上的相应标记点.

4.1.2 基于人脸运动机理的增强标记点运动求解

本文根据当前目标人脸的运动数据 (由交叉映射生成) 计算增强标记点的运动. 人脸标记点运动是一种非刚性运动, 人脸皮肤运动机理是: 相邻两点的运动关系主要取决于相互之间的距离, 距离越短, 运动越接近, 接近程度与距离的关系是一种复杂的非线性关系. 基于此, 本文构建了一个非线性模型, 采用一定数量临近点的位移来预测增强标记点位置:

$$\mathbf{v}_i^t = \mathbf{v}_i^0 + \sum_{j \in K_i} \lambda_j (\mathbf{m}_j^t - \mathbf{m}_j^0), \text{ where } \lambda_j = \frac{\frac{1}{\|\mathbf{v}_i^0 - \mathbf{m}_j^0\|^\tau}}{\sum_{j \in K_i} \frac{1}{\|\mathbf{v}_i^0 - \mathbf{m}_j^0\|^\tau}}, \quad (7)$$

图 4 增强标记点的配置

Figure 4 The configuration of enhanced markers

该权值构建了临近点运动对当前点运动的非线性影响关系 (距离越远, 影响越小), $\sum_{j \in K_i} \lambda_j = 1$, $\tau = 2$ (为经验值), K_i 为以当前增强标记点为圆心以搜索半径 r_i 的球内所有标记点的索引集合. 半径大小设置存在差别, 对于上下嘴唇上的增强标记点, 搜索半径为嘴唇宽度的一半, 对于其他增强标记点, 搜索半径为嘴唇宽度.

4.2 基于预计算的人脸模型驱动

4.2.1 基于 RBF 的预计算方法

本文采用 RBF 方法对人脸模型的标记点 (包括增强标记点) 进行训练, 通过目标模型相应标记点的当前帧位置信息驱动模型产生表情动画. 和交叉映射使用的 RBF 方法不同, 我们通过每一帧训练一个运动关系, 将这个运动关系运用到目标模型计算每一个顶点的运动.

首先构建一个驱动模型:

$$\mathbf{s}_i^t = \sum_{j=1}^n \lambda_j^t \phi(\|\mathbf{s}_i^0 - \mathbf{t}_j^0\|), \quad (8)$$

其中 $\mathbf{s}_i^t$ 表示 t 帧时刻第 i 个模型顶点的位置, $\mathbf{t}_j^t$ 表示目标模型 t 帧时刻第 j 个标记点和增强标记点位置, n 为当前功能分区内标记点及增强标记点的总数.

从求解过程来看, 首先根据 (8) 式计算径向基权值矩阵 $\mathbf{T}^t = [\lambda_0^t, \lambda_1^t, \dots, \lambda_{n-1}^t]^T$, 然后利用训练的径向基网络与权值矩阵, 计算得出当前帧模型顶点的位置. 在实时计算时, 求解径向基权值矩阵由于涉及到矩阵的逆阵运算, 计算量往往较大. 因此, 为减少实时运算消耗, 采用预计算方法, 主要分为两个步骤. 第一, 预计算系数矩阵的逆阵 (9) 式, 保存逆阵, 用于求解每帧的径向基权值矩阵:

$$\mathbf{A}^{-1} = \begin{bmatrix} \phi_{0,0}^t & \cdots & \phi_{0,n-1}^t \\ \vdots & \ddots & \vdots \\ \phi_{n-1,0}^t & \cdots & \phi_{n-1,n-1}^t \end{bmatrix}. \quad (9)$$

第二, 预计算模型顶点与标记点及增强标记点位置的径向基网络信息. 由 (8) 式可知, 每一帧的模型顶点的运动求解都需要调用该点到所有标记点及增强标记点的径向基网络信息, 即 $[\phi(\|\mathbf{s}_i^0 - \mathbf{t}_0^0\|), \dots,$

$\phi(\|\mathbf{s}_i^0 - \mathbf{t}_{n-1}^0\|)$, 而模型的顶点数目庞大, 因此每帧耗费大量的计算. 将所有这些径向基网络信息预存到 $N \times n$ 阶矩阵 $\mathbf{B}$, 则每帧的细节顶点的位置可以由下式求解:

$$\mathbf{s}_i^t = \sum_{j=1}^n \lambda_j^t \mathbf{B}_{(i,j)}. \quad (10)$$

同理, 我们在交叉映射构建时, 也对首帧目标个性人脸模型上的点所构建径向基训练网络 (见 (1) 式) 进行预计算. 从理论角度上, 本文的预计算方法从本质上分析了 RBF 计算过程, 区分开了冗繁的重复计算部分, 并在实时计算之前进行预计算并存储计算结果, 从而节省了一定的计算消耗.

4.2.2 顶点法线快速更新

人脸模型的顶点法线计算是模型渲染过程中需要考虑的问题. 当前顶点法线计算方法多种多样, 针对离散面片构建网格模型, 顶点法线计算普遍是针对相邻面片的法线进行加权求和^[34]. 对于网格曲面法线计算主要可以分为两个步骤: 第一步, 计算当前每个顶点的相邻面元索引; 第二步, 根据相邻面元的法线加权求解当前顶点的法线. 对于成千上万个顶点所组成的模型而言, 面元数量往往更多, 因此获取每个顶点的相邻面元的计算过程是比较耗时的, 通常的快速求解方法采用 K 邻近方法缩减搜索空间来提高计算速度. 在人脸模型运动的过程中求解模型的顶点法线, 本文基于 K 临近方法求解顶点法线, 并且在首帧保存了每个模型顶点的相邻面元信息, 实现了在求解过程中顶点法线的快速更新. 法线求解如下式:

$$\mathbf{n}_i = \frac{\sum_{k \in F_i} f_k \mathbf{f}_k}{\|\sum_{k \in F_i} f_k \mathbf{f}_k\|}. \quad (11)$$

令 F_i 为第 i 个标记点的相邻三角面片集合, f_k 和 $\mathbf{f}_k$ 分别表示 F_i 的第 k 个三角面片的面积和法线 (法矢长度为 1), 则对相邻面片法线根据面积大小进行加权计算每个标记点的临时法向量.

Botsch 等^[35] 提出使用对基函数进行微分的方法, 求解每个顶点的 Jacobi 矩阵来更新法线, 基于 GPU 图形硬件加速技术实现了顶点的快速更新. 这种计算方法相比于要每帧更新每个顶点的相邻面元索引的法线求解方法要更快速. 而事实上, 如果在运动前保存相邻面元索引, 则计算效率比计算顶点的 Jacobi 矩阵更高, 另外, 利用 Jacobi 矩阵更新变形曲面法线信息存在着一定的局限性^[36].

5 实验及分析

如图 5 所示, 本文的运动数据使用了 8 台摄像机成布置, 其中六台摄像机高度为 2.0 m, 其余两台高度为 1.7 m, 数据采集场地长宽各为 2 m. 摄像机设置采集频率为 60 帧/秒 (fps), 人脸标记点 (反射球) 直径为 2~3 mm. 本文的人脸表情重构实验的平台为 VC++ 构建的程序系统, 其中, 利用 OpenGL 构建三维图形环境, 嵌入 Matrix(LIB) 进行矩阵运算. 程序运行机器配置为 Intel®Core™2 Quad CPU 2.4 GHz (4 CPUs)、3GB 内存.

首先我们使用本文方法与全局 RBF 交叉映射方法所构成真实感进行比较. 由于全局 RBF 交叉映射方法忽略了嘴唇上下部分运动的不连续性, 因此, 在映射过程中表情运动容易造成一定程度的失真, 如图 6 所示. 由于伪关联存在, 会造成人脸表情运动仿真的失真. 本文试验中, 占据少数光标的下颌部分的运动失真相对较为严重, 嘴唇张合程度越大, 上下颌部分的运动受到伪关联的影响更明显, 如图 6 中 (b) 和 (d) 栏的图片显示 (下巴尖部的光标被上拉, 下巴变尖), 相比之下, 使用本文方法下颌部

图 5 人脸表情运动捕捉环境

Figure 5 Facial MoCap environment (a) and the marker configuration in this paper (b) and (c)

图 6 基于 RBF 人脸全局映射的运动失真

Figure 6 The distortion derived from global RBF cross-mapping method. (a) the motion data; (b), (d) the animated face using global RBF cross-mapping method; (c), (e) the animated face using partitioned RBF cross-mapping method

分的仿真效果则相对真实. 另外, 比较方法由于没有使用增强标记点, 形成了分区结合部的运动不连续.

接下来我们对本文的人脸表情运动重构真实感进行实验. 实验选取 5 个个性人脸模型, 如图 7 所示, 将图中的 5 个个性人脸从左至右标记为个性人脸 1~5, 个性人脸 1~5 分别表征了不同特征的人脸, 包括不同脸型 (圆脸、长脸、方脸等) 以及不同种族 (亚洲黄色人种、欧洲白色人种、非洲黑色人种等). 选择由同一个表演者所捕捉的不同表情的运动捕捉序列, 驱动不同个性人脸模型产生动画, 如图 8 所示.

从图 8 可以看出, 针对单个演员不同的表情运动数据, 本文采用的人脸表情重构方法能够有效地驱动不同人脸模型, 并且生成比较满意的动画效果. 本文的方法不仅适应于驱动不同个性人脸, 还能适应各种变化迥异的表情运动数据而不至于产生局部失真, 因而具备一定的普适性.

最后我们通过实验检测使用预计算方法的计算效率. 为对人脸表情运动捕捉数据交叉映射及表情重构的算法运行时间进行评估, 我们随机抽取了 3 组运动序列, 利用这些序列数据驱动人脸模型实时

图 7 不同中性表情人脸模型及其标记点设置

Figure 7 Different experimental neutral facial models and their marker configurations

MpCap data	Description	Face model 1	Face model 2	Face model 3	Face model 4	Face model 5
	Happy					
	Angry					
	Afraid					
	Surprise					

图 8 不同人脸模型的表情动画

Figure 8 The expressional facial animation for different models

生成动画. 运动序列的长度为 300 帧, 实验用的人脸模型采用个性人脸 1 (见图 7), 模型顶点和三角面元个数分别为 2968 及 23022 个.

表 1 对是否使用预计算方法的人脸表情重构算法运行效率比较

Table 1 The efficiency comparison between the facial animation construction methods with pre-computing and without pre-computing

Random mocap sequence	Frame size	With pre-computing			Without pre-computing	
		Pre-computing(ms)	Running time		All	Average
			All(ms)	Aver(ms)	running	running
					time(ms)	time(ms)
Sequence 1	300	1141	47328	157.760	86516	288.667
Sequence 2	300	1094	47563	158.543	86516	288.667
Sequence 3	300	1109	47125	157.083	86437	288.123

从表 1 可以看出, 对于 300 帧的运动捕捉序列而言, 使用预计算方法能够节省将近 40 s 的计算消耗, 平均每帧减少计算量约 130 ms, 计算效率相比非预处理计算提升了约 45%.

6 总结和讨论

本文提出了一种新颖的基于运动捕捉的人脸表情仿真方法. 针对人脸开放结构如嘴、眼等产生区域性运动的表情运动机理, 提出了基于功能分区构建交叉映射的思想. 本文主要有如下贡献: 针对交叉映射的构建, 对现有全局 RBF 映射方法进行了改进, 提出使用分区映射解决了全局 RBF 映射时嘴唇上下部分运动出现伪关联的问题; 针对模型驱动本文有两个学术贡献: 第一, 提出了一种根据人脸运动机理计算加强光标的方法, 运用于分区模型驱动; 第二, 分析了 RBF 的模型驱动问题, 提出了使用预计算方法提高动画实时效率.

需要注意的是, 在考虑采用 RBF 方法对人脸分区构建人脸映射时, 要考虑所属于各个分区的标记点的规模. 如果某个分区的标记点个数过少, 在使用 RBF 构建交叉映射和模型驱动时, 由于 RBF 训练网络单元过少, 映射的鲁棒性也会因此而降低, 从而容易造成结果产生失真.

实验证明了本文方法的有效性, 能够将同一人脸的运动数据应用不同个性人脸模型并生成逼真的人脸表情动画, 能够快速更新人脸模型的非刚性变形以及顶点法线信息.

参考文献

- 1 Parke F I, Waters K. Computer Facial Animation. Massachusetts: A K Peters Ltd., 1996
- 2 Noh J, Neumann U. A Survey of Facial Modeling and Animation Techniques. Technical Report. California: University of Southern Californis, 1998
- 3 Deng Z, Noh J. Computer facial animation: a survey. In: Data-Driven 3D Facial Animation. London: Springer, 2007. 1-28
- 4 Noh J, Neumann U. Expression cloning. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM, 2001. 277-288
- 5 Chai J, Xiao J, Hodgins J. Vision-based control of 3D facial animation. In: Eurographics/SIGGRAPH Symposium on Computer Animation. San Diego: Eurographics Association Aire-la-Ville, 2003. 193-206
- 6 Lorenzo M S, Edge J D, King S A, et al. Use and re-use of facial MoCap data. In: Proceedings of Vision, Video and Graphics. Bath: Eurographics Association Aire-la-Ville, 2003. 135-142

- 7 Edge J D, Sánchez M A, Maddock S. Animating Speech from Motion Fragments. Technical Report CS-04-02. Department of Computer Science, University of Sheffield, 2004
- 8 Havaldar P. Performance-driven facial animation. In: Proceedings of SIGGRAPH 2006 (course notes). New York: ACM, 2006. 23–42
- 9 Curio C, Breidt M, Vuong Q, et al. Semantic 3D motion retargeting for facial animation. In: ACM International Conference Proceeding Series. New York: ACM, 2006. 77–84
- 10 Deng Z, Chiang P, Fox P, et al. Animating blendshape faces by cross-mapping MoCap data. In: Proceedings of ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games. New York: ACM, 2006. 43–48
- 11 Yao J, Wang Y S, Ding B, et al. Facial animation remapping based on spherical parametrization. J Image Graph, 2009, 14: 1406–1412 [姚健, 王阳生, 丁宾, 等. 基于球面参数化的人脸动画重映射. 中国图像图形学报, 2009, 14: 1406–1412]
- 12 Chuang E, Bregler C. Performance Driven Facial Animation Using Blendshape Interpolation. Computer Science Technical Report. Stanford University, 2002
- 13 Pighin F, Lewis J P. Facial motion retargeting. In: Siggraph 2006 Course Notes: Performance-Driven Facial Animation. New York: ACM, 2006. 1–9
- 14 Blanz V, Vetter A. Morphable model for the synthesis of 3d faces. In: SIGGRAPH'99: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press/Addison-Wesley Publishing Co., 1999. 187–194
- 15 Blanz V, Basso C, Poggio T, et al. Reanimating faces in images and video. Comput Graph Forum, 2003, 22: 641–650
- 16 Kalberer G A, Gool L V. Realistic face animation for speech. J Visual Comp Animat, 2002, 13: 97–106
- 17 Pandzic I S. Facial motion cloning. Graph Model, 2003, 65: 385–404
- 18 Pyun H, Kim Y, Chae W, et al. An example-based approach for facial expression cloning. In: Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation. San Diego: Eurographics Association Aire-la-Ville, 2003. 167–176
- 19 Yano K, Harada K. A facial expression parameterization by elastic surface model. Int J Comput Game Technol, 2009, 2009, doi: 10.1155/2009/397938
- 20 Li H, Weise T, Pauly M. Example-based facial rigging. ACM Trans Graphic, 2010, 29, doi: 10.1145/1778765.177876912
- 21 Park S I, Hodgins J K. Capturing and animating skin deformation in human motion. ACM Trans Graphic, 2006, 25: 881–889
- 22 Muller M, Heidelberger B, Teschner M, et al. Meshless deformations based on shape matching. ACM Trans Graphic, 2005, 24: 471–478
- 23 Guenter B, Grimm C, Wood D, et al. Making faces. In: International Conference on Computer Graphics and Interactive Techniques, ACM SIGGRAPH 2005 Courses. New York: ACM, 2005. 55–66
- 24 Zhuang Y T, Pan Y H, Xiao J. A Modern Approach to Intelligent Animation: Theory and Practice. Berlin/Heidelberg: Springer, 2008
- 25 Wang Y S, Zhuang Y T, Xiao J, et al. A piece-wise learning approach to 3D facial animation. In: Lecture Notes in Computer Science, Advances in Web Based Learning-ICWL 2007. Berlin/Heidelberg: Springer, 2007. 119–184
- 26 Sifakis E, Neverov I, Fedkiw R. Automatic determination of facial muscle activations from sparse MoCap marker data. In: ACM SIGGRAPH 2005. New York: ACM, 2005. 417–425
- 27 Somasundaram A, Parent R. A facial animation system for expressive audio-visual speech. OSU-CISRC-4/06-TR46. Columbus, OH: Department of Computer Science and Engineering, The Ohio State University, 2006
- 28 Joshi P, Tien W C, Desbrun M, et al. Learning controls for blend shape based realistic facial animation. In: Eurographics/SIGGRAPH Symposium on Computer Animation. New York: ACM, 2003. 187–192
- 29 Kyung-Gun N, Moon-Ryul J. Weighted local shape blending for facial motion retargeting. Comp Animat Virt World, 2010, 21: 255–265
- 30 Takamizawa R, Suzuki T, Kubo H, et al. Expressive facial subspace construction from key face selection. In: SIGGRAPH 2009. New York: ACM, 2009. 3–7
- 31 Pighin F, Hecker J, Lischinski D, et al. Synthesizing realistic facial expressions from photographs. In: Proceedings of SIGGRAPH 98. Computer Graphics Proceedings, Annual Conference Series. New York: ACM, 1998. 75–84
- 32 Enciso R, Li J, Fidaleo D, et al. Synthesis of 3D faces. In: International Workshop on Digital and Computational Video, 2000. 8–15

- 33 Ulgen F. A step toward universal facial animation via volume morphing. In: 6th IEEE International Workshop on Robot and Human Communication, 1997. 358–363
- 34 Jin S, Lewis R R, West D. A comparison of algorithms for vertex normal computation. Visual Comput, 2005, 21: 71–82
- 35 Botsch M, Kobbelt L. Real-time shape editing using radial basis functions. Eurograph, 2005, 24: 611–621
- 36 Fernando R. GPU Gems: Programming Techniques, Tips and Tricks for Real-Time Graphics. America: Pearson Higher Education, 2006. 491–497

On the simulation of expressional animation based on facial MoCap

FANG XiaoYong^{1*}, WEI XiaoPeng², ZHANG Qiang^{2*} & ZHOU ChangJun²

1 *School of Computer and Infomation Science, Hunan Institute of Technology, Hengyang 421002, China;*

2 *Key Laboratory of Advanced Design and Intelligent Computing (Dalian University), Ministry of Education, Dalian 116622, China*

*E-mail: fangxiaoyong@dlu.edu.cn; zhangq@dlu.edu.cn

Abstract MoCap(motion capture)-based animation is a hot issue in computer animation research currently. Based on the optical MoCap system, this paper proposes a novel cross-mapping based facial expression simulating method. To overcome the problem of the false upper and lower jaw correlation derived from the facial global RBF-based cross-mapping method, we construct a functional partition based RBF cross-mapping method. During model animating, enhanced markers are added and animated by our proposed skin motion mechanism. In addition, based on the enhanced markers, an improved RBF-based animating approach is raised to derive realistic facial animation. Further-more, a pre-computing algorithm is presented to reduce computational cost for real-time simulation. The experiments proved that the method can not only map the MoCap data of one subject to different personalized faces but generate realistic facial animation.

Keywords facial animation, cross-mapping, function partition, enhanced marker, RBF

FANG XiaoYong was born in 1982. He received the Ph.D. degree in Dalian University of Technology, Dalian, in 2010. Currently, he is a Lecturer at Hunan Institute of Technology. His research interests include computer animation, computer vision and artificial intelligence.

ZHANG Qiang was born in 1971. He received the Ph.D. degree in Xidian University, Xi'an, in 2002. Currently, he is a Professor at Dalian University. His research interests include neural networks, DNA computing and computer animation. Dr. Zhang is a council member of the China Graphics Society, as well as Vice Chairman of the Graphics Society Liaoning, and an academic committee member of CCF Young Computer Scientists & Engineers Forum Dalian.

WEI XiaoPeng was born in 1959. He received the Ph.D. degree in Dalian University of Technology, Dalian, in 1993. Currently, he is a Professor at Dalian University. His research interests include neural networks, DNA computing, mechanical design and CAD&CG. Dr. Wei is a Vice Chairman of the China Graphics Society, as well as Chairman of a Council of Graphics Society Liaoning.

ZHOU ChangJun was born in 1977. He received the Ph.D. degree in Dalian University of Technology, Dalian, in 2008. Currently, he is a Professor at Dalian University. His research interests include pattern recognition, artificial intelligence and computer animation. Dr. Zhou is a member of CCF, and an academic committee member of CCF Young Computer Scientists & Engineers Forum Dalian.