网络出版时间：2016-01-05 13:29:11
网络出版地址：http://www.cnki.net/kcms/detail/11.1826.TP.20160105.1329.002.html

利用几何度量的无监督实时面部动画生成算法

姜 那 刘少龙 石 峰 周 忠

(北京航空航天大学 虚拟现实技术与系统国家重点实验室，北京 100191)

摘 要 目前面部表情动画生成算法普遍具有捕捉设备昂贵、依赖用户表情数据预采集、需要用户具备专业知识等缺点，因此很难在普通用户群众进行推广。针对这些不足，本文选择价格适中、操作简单的 Kinect 作为采集设备，并提出一种无须预处理的面部表情捕捉算法。其首先从捕获的面部表情数据中提取面部特征点，利用几何度量建立低层面部特征点与高层表情语义之间的联系，根据权重和补偿策略建立几何度量样本集。然后采用无监督的方式自动分析样本分布，推测各表情单元的变化区间，实现表情参数的实时提取。最后利用表情参数驱动离线生成的通用表情基，生成能反映用户情绪的面部动画。在表情基生成过程中，首次引入控制点影响区域的概念来约束拉普拉斯变形算法，以提高通用 Blendshape 表情基的精度。实验结果表明，该方法简单易行，无需对每名用户进行表情数据预采集，即可在多人同时出现、部分遮挡等情况下实时、鲁棒地生成与用户近似的面部动画。主观评价中，该方法被证明具备优秀的采集灵活度、使用方便、实时性能良好，在普通用户群体中更具备推广价值。

关键词 Kinect；人脸跟踪；Blendshape 模型；表情动画；表演驱动
中图法分类号 TP391
论文引用格式
姜那，刘少龙，石峰，周忠，唐常杰，利用几何度量的无监督实时面部动画生成算法，2016，Vol.39：在线出版号 No.1
JIANG Na,LIU Shao-Long,SHI Feng,ZHOU Zhong, Unsupervised Algorithm of Real-Time Facial Animation by Geometric Measurements, Chinese Journal of Computers,2016, Vol.39: Online Publishing No.1

Unsupervised Algorithm of Real-Time Facial Animation by Geometric Measurements

JIANG Na LIU Shao-Long SHI Feng ZHOU Zhong

(State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China)

Abstract Most of current facial animation algorithms are difficult to be popularized among common users, because of the disadvantages of using expensive capture devices, depending on preprocess of expression data and needing special operator. To solve these problems, an affordable and convenient Kinect was chosen as capture device, and a non-preprocessing capture algorithm of facial expression was proposed in this paper. Firstly, the facial feature points was extracted from the RGBD data which was captured by Kinect, and at the same time the relationship between low-level facial feature points and high-level expressional semantics was built using geometric measurements. Meanwhile the sample group of geometric measurements was established according to the weight strategies and compensation strategies. Secondly, the distribution of sample was analyzed automatically by unsupervised method and then the range of expression unit was inferred, so that expression parameters can be extracted in real time. Finally, the universal Blendshape basis, which were generated offline,

本课题得到国家“八六三”高技术研究发展计划 项目基金(No.2015AA016403)、国家自然科学基金项目(No.61170188)资助. 姜 那(**通讯作者**)，女，1989年生，博士研究生，主要研究领域为3D面部动画、目标跟踪，E-mail: jiangna@buaa.edu.cn. 刘少龙，男，1988年生，硕士，主要研究领域为3D面部动画生成，E-mail: lsl642816419@gmail.com. 石 峰，男，1982年生，博士,主要研究领域为运动分割、计算机视觉，E-mail: supersf2008@hotmail.com. 周 忠(**通讯作者**)，男，1978年生，博士，副教授，中国计算机学会(CCF)高级会员，主要研究领域虚拟现实等，E-mail: zz@buaa.edu.cn.

were driven by the expression parameter to generated real-time facial animation of reflecting users' mood. In this process, in order to improve the accuracy of the universal Blendshape basis, the paper first introduced area of influence of control points to restrain the Laplace deformation algorithm. The results demonstrate that the proposed algorithm is a simple and convenient method to generate real-time and robust facial animation without preprocess of expression data collecting for every user, even in these case of appearing many people simultaneously and partial occlusion. It is proved that high flexible collection, easy operation and reliable real-time performance are provided by the method. Therefore, it is worth to generalize among ordinary users.

Keywords Kinect; face tracking; blendshape models; face animation; performance-driven

1 引 言

随着人们对非语言形式的人机交互关注程度的增加，实时面部表情动画生成技术在影视、游戏业内受到了广泛关注[1]。不仅如此，在计算机图形学领域，实时面部表情动画生成技术也逐渐成为了研究重点。以著名的 3D 特效电影猩球崛起Ⅱ为例，电影中角色的面部表情动画首先需要通过专用的设备获得演员的真实面部表情数据，然后借助计算机图形算法来再次表示面部表情，并通过合理的约束来保证生成的面部动画与真实面部表情一致。因此，现阶段面部表情动画生成过程中普遍存在以下三个困难：1.为保证动画的拟合精度，需要昂贵的采集设备和专业的数据预处理过程，使得建立系统的开销过大；2.面部表情的生理机制较为复杂，同时不同用户面部表情之间存在着难以简单量化的个性化差异，导致利用算法生成面部表情动画的做法适用范围有限；3.人类对不真实的面部表情非常敏感，对生成动画与真实面部表情一致性的要求很高。这些困难使得设计一个具有真实感的面部表情动画生成算法富有极大的挑战性。

为了克服上述困难，面部表情捕捉成为了表情动画生成算法的核心与关键。国内外的研究者们提出了大量的 3D 面部表情捕捉方法，例如侵入性的 3D 扫描法、基于 marker 点的捕捉系统以及非侵入性的结构光系统、基于图像的动作捕捉法等[2]。前者侵入性的方法普遍应用于有质量需求的影视制作行业，可以获得高质量的人脸模型，但是其设备昂贵、使用复杂，并且不能获得实时的结果。其中 3D 扫描法善于获得高清的面部细节，如皱纹等，但是只能处理静态姿态的人脸；基于 marker 点的捕捉系统最为常用、并且具有高时间分辨率，但是表情变化细节常常因为 marker 点的数量和位置而被忽略。后者非侵入性方法降低了对设备的要求，但是容易受到外界光照等条件变化的影响，依赖于大量数据的预处理，并需要用户具备一定的专业知识。其中，结构光系统可以捕捉动态的 3D 人脸，但是在时间分辨率上不如基于 marker 点的捕捉系统、在空间分辨率上比不过 3D 扫描法获得的效果；至于基于图像的动作捕捉法，由于输入数据不灵活，很难满足人类对面部动画生成技术的三项基本要求（基于动态姿态进行数据获取、实时获得、生成动画与真实表情相一致）。但是其工作原理随着采集设备的革新，成为了新算法改进的基础。2010 年微软推出了一种 RGBD 设备 Kinect 改变了原有的数据采集方式，推动了实时面部表情动画生成技术的发展。Kinect for Windows 能以每秒 30 帧的速率同步采集深度图像和彩色图像，其近景模式能够采集到最近 40cm 处物体的深度信息[3]，非常适合作为一种轻量级的表情捕捉设备。同时，其价格适中、操作简单，方便推广到消费级用户中。

针对上述分析，本文提出了一种基于 Kinect 的无监督面部表情捕捉算法，并在此基础上生成了实时表情驱动的面部动画。其主要分为在线和离线两个部分。离线部分负责通用 Blendshape 表情基的生成，过程中引入了控制点影响区域概念来约束拉普拉斯变形。在线部分则负责实时面部表情动画的生成，其主要分为特征点实时提取、表情参数实时提取以及表情动画生成三个阶段。第一阶段，首先利用 K-means 聚类算法对 Kinect 实时获得的深度图像进行背景剔除，得到用户面部区域的点云；然后根据相邻两帧的面部点云进行头部姿态追踪，其中使用迭代最近点（Iterative Closet Point，ICP）算法估计当前帧的头部姿态，进而利用 3D 主动外观模型（Active Appearance Model，AAM）算法从对应的彩色图像中提取用户的面部特征点。第二阶段，以上一阶段获得的面部特征点作为输入，根据面部表情编码系统（Facial Action Coding System，FACS）定义的表情单元（Action Units，AU）从特征点中

提取相应的几何度量值；再根据样本权重和补偿策略将几何度量样本添加到几何度量样本集；针对不断更新的样本集，利用无监督的方式自动分析样本分布，推测出各个 AU 的变化区间，进而计算出当前帧各个 AU 的变化幅度，得到实时的表情参数。第三阶段，利用实时表情参数驱动离线生成的通用

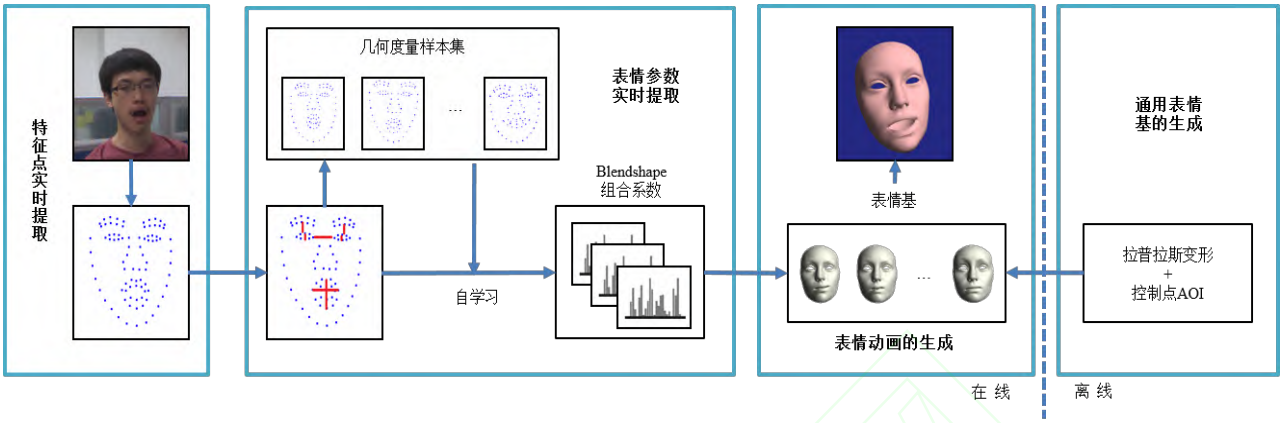


图 1　面部动画生成算法框图

Blendshape 表情基，生成与用户表情相似的面部动画。算法的详细框架如图 1 所示。

与现阶段依赖于昂贵专业设备、离不开用户表情预采集和需要用户具备专业知识的面部动画生成算法不同[2,4-6]。本文主要具备两点创新：1.算法主要利用几何度量值建立了低层面部特征点和高层表情语义之间的联系，并在此基础上对几何度量样本的分布情况进行自动分析，从而估计出用户的面部表情参数来驱动通用表情基生成与用户表情近似的面部动画。这种算法无需对每位用户的表情数据进行预采集，具有更好的普适性和易用性。2.本文设计的面部动画生成技术首次引入了控制点影响区域（Area of Influence，AOI）的概念来改进生成 Blendshape 表情基的拉普拉斯变形算法。改进后的算法能够有效地避免变形过程陷入局部最优解，显著提高了 Blendshape 表情基的生成精度，使得到的面部动画生成更符合人眼对表情差异辨别的要求。同时算法还能增强头部快速转动、室内光照条件变化以及多人同时出现等情况下的鲁棒性。

本文将分为 6 个部分：相关工作部分主要介绍近年来主要的面部表情捕捉技术和表情动画生成技术；无监督的面部表情捕捉、基于 AOI 的表情基生成两部分则是实时面部动画生成算法的核心，详细描述本文的创新和细节；实验结果分析部分对多组实验结果进行对比分析，体现本文算法性能及创新性；结论部分总结本文贡献、讨论算法的不足和限制，明确下一步工作的研究内容。

2 相关工作

自 1972 年 Parke[7]等第一次实现参数化的人脸模型以来，面部表情动画生成技术一直在不断提高。而近二十年的方法间虽然表达效果和实现形式不同，但其依照的基本原则十分类似，均需要首先利用面部表情捕捉技术捕捉使用者的面部表情并进行数据化，然后利用计算机图形学算法驱动虚拟角色生成与捕捉数据相一致的表情动画。根据这一基本原则，实时的面部动画生成技术需要重点研究面部表情捕捉与表情动画生成两个环节。

面部表情捕捉通常需要采集设备来完成，在影视制作中普遍使用基于标识（marker）点的面部表情捕捉系统[7-12]，此类表情捕捉系统是在被捕捉者的面部关键位置标记 marker 点，然后利用先进的运动捕捉设备直接获取这些 marker 点的三维位置。由于这些 marker 点处在面部关键位置，因此他们的坐标变化可以反映出人脸表情的变化。通过获取到的 marker 点的三维运动序列对一个预先准备好的面部模型进行变形就可以得到相似的表情动画。该类系统时间分辨率和鲁棒性极高，但是由于 marker 点的数量有限，会导致面部细节的丢失和较低的空间分辨率，从而失去了利用丰富的面部细节做更多处理的潜在机会。2011 年 Huang 等[2]进行了改进，提出了利用 Motion Capture 和三维扫描仪共同协作的方法，这类方法较以往仅使用三维扫描仪[13,14]或者仅使用基于 marker 点的面部表情捕捉系统来说，空间分辨率和时间分辨率均有提高。但是由于设备

昂贵、安装及操作复杂，很难在普通用户间推广使用。除此之外，还有一种结构光系统可以用来捕捉面部表情[15,16]，该类系统采用光流法或空间编码从图像序列中获得当前人脸的深度数据，但前者只能捕捉动态的人脸面部表情，后者只能捕捉静态面部表情，二者分辨率都很难提高。与之类似的还有多视角相机系统[17,18]，利用不同视角的相机获得人脸目标的深度信息，在处理时间上具有优势，然而各相机间存在干扰导致推断的深度数据不够准确、影响生成动画的效果。随着 Kinect 等多目设备的推出，深度数据被进一步的应用到了面部表情捕捉技术中。2011 年 Weise[4]等首次利用 Kinect 作为采集设备实现了实时面部表情捕捉，该算法以 Kinect 采集的深度图像和彩色图像作为输入，分别利用非刚性 ICP 算法和基于模型的光流法处理深度数据和彩色数据，然后通过混合概率主成分分析（Mixtures of Probabilistic Principal Component Analyzers，MPPCA）概率模型引入表情动画先验，将表情系数的优化转化为一个最大后验估计（Maximum A Posteriori，MAP）问题。但该方法需要针对不同用户表情数据进行预采集的缺点限制了方法的推广和使用。近三年相继出现了许多实时的面部表情捕捉系统和方法[6,19-21]，除使用 RGBD 作为输入的算法以为，还有部分使用单目的普通摄像头作为采集设备的方法[6,19]。该类方法的主要技术难点在于面部特征跟踪和头部姿态的估计。由于从图像中无法直接获得物体的原始三维信息，因此基于普通彩色相机的面部表情捕捉呈现为一个病态问题，解决这类问题一般需要给定足够的假设或者先验，或者配合使用多目相机来弥补信息的缺失。其中文献[19]与[6]均采用普通摄像机作为采集设备，由于不能直接获得深度信息，前者仍然需要预采集用户表情数据来训练针对不同用户的 3D 形状回归器；后者则需要根据用户的单张正面人脸图像训练特定用户的局部纹理模型。这些预处理操作不仅耗费时间，还需要用户具备特殊的使用技巧。同时，面对新用户的加入还需要重新系统设定。因此本文提出的无需预采集的表情捕捉算法十分必要，其省去了用户繁琐的预操作环节，体现了面部表情动画生成算法的普适性，并有助于在用户级群体中推广应用。

表情动画生成是指利用计算机表示和生成连续变化的面部表情。面部表情动画可分为 2D 动画（如图像）和 3D 动画（如三维模型），现阶段相关的算法和系统主要集中在研究 3D 表情动画的生成。人脸的运动方式取决于面部肌肉的运动，为此早期研究者提出了基于生理的肌肉系统。1981 年 Platt 和 Badler 率先将质点弹簧系统应用到了基于生理的肌肉模型中[22]，这种方法将面部皮肤视为富有弹性的网格，面部下方的肌肉在收缩时将力作用于弹性网格上，从而使面部网格变形并产生表情。为了更逼真的进行面部物理仿真，Terzopoulos 等人[23]在此基础上根据人脸的解剖学结构又提出了一种三层可变形网格的模型。但是这类方法需要大量的物理结算，参数选择十分困难。Waters 等人则改变思路，使用向量模型对人脸肌肉系统进行建模[24]，在时间效率上有所提高，不过仿真效果不如前者；因此，该类方法基本已经不能满足现在面部表情动画生成算法在实时和保真方面的要求。与基于生理的肌肉模型不同，还有许多研究者在 Parke[7]参数化人脸模型的基础上进行改进，并假设任何表情都可以通过其他若干表情的组合进行近似表达，降低了计算复杂度。其中最基础的是基于 PCA 的线性模型[25-27]，该类模型计算简单，但是由于 PCA 维度的限制在表达不同个体间的表情差异时效果不佳。近几年基于 Blendshape 的混合模型[9,16,28,29]相对更为流行。与基于 PCA 的线性模型相比，Blendshape 混合模型则可以利用唯一的一组基来生成不同人的表情，这一特性十分适合将真人的表情转移到不同的角色上。然而 Blendshape 模型中表情基的质量将会直接关系到人脸表情动画的生成效果。本文对生成 Blendshape 表情基的拉普拉斯变形算法进行改进，引入控制点的影响区域(AOI)来克服变形过程极易陷入局部最优解的问题，提高了面部表情动画生成算法的准确性和鲁棒性。

3 无监督的面部表情捕捉

面部表情捕捉一般指获得用户的面部特征信息，而面部特征信息一般可以通过面部稀疏特征点来表示，所以提取面部特征点是提取表情信息的有效方式。但是由于不同人的面部形态存在差异，即使两个人表情相同也会得到位置不同的特征点数据。因此算法需要进一步分解面部特征信息为具有用户特色的面部形态信息和具有语义一致性的面部表情信息。现有的实时表情捕捉算法[4,19]大多通过精确的先验信息实现面部形态和表情的分解。而先验获取的方法则是在捕捉前要求用户做出一系列的特定表情，然后从这些表情中学习出用户相关

的表情先验。这类方法是一种监督式的学习方法，最大的缺点在于需要用户配合训练、训练质量依靠专业知识并且质量难以把握。这些问题直接导致基于此类算法的系统普适性和易用性极差。为解决这个问题，为解决这个问题，本文提出了一种无监督的面部形态和表情的分解方法，该方法不需要对被捕捉用户进行任何监督式的训练，提取面部特征点后可以自动提取出用户的面部表情参数。基于本方法的实时面部表情生成系统可以做到用户即来即用，普适性和易用性得到大大提高。

3.1 面部特征点的实时提取

面部特征点通常位于面部关键位置，例如眼睛周围、嘴巴周围等。当面部表情发生变化或者头部进行运动时，这些点的位置也会随之变化。前一类变化属于非刚性运动，蕴含了面部表情信息；后一类变化属于刚性运动，蕴含了头部姿态信息。算法首先将用户头部这两类运动解耦合，然后只根据其中的非刚性运动来提取表情。即首先基于 Kinect 深度图估计头部姿态；然后基于 Kinect 彩色图提取面部特征点。

3.1.1 头部姿态估计

头部姿态估计主要是为了计算出头部相对相机的平移和旋转。现有方法大都在彩色图像上进行[30]，直接在整幅图像中搜索人脸，进行了大量不必要的计算，忽略了场景的几何信息。我们的方法从深度图入手，在深度图中搜索人脸区域，因此能够充分利用场景几何信息，从而提高了运算效率。

首先进行深度图的背景剔除，采用 K-means 聚类算法分离场景的前景和背景，从而得到有效的头部区域。而在实时捕捉过程中，对每一帧进行背景剔除后，都将得到与之对应的只包含有效头部区域的深度图。该区域每一个像素点均带有深度信息，因此可以将得到的面部区域视为由三维点组成的点云。这样一来，头部姿态的跟踪就转化成了三维点云之间的匹配。Weise[4]等采用非刚性迭代最近点（Non-rigid Iterative Closet Point，Non-rigid ICP）算法匹配相邻两帧的面部点云，该方法不仅能够得到点云之间的匹配关系，还能计算出点云之间的非刚性运动。然而 Non-rigid ICP 算法需要的迭代次数较多，计算量较大，会成为实时应用的性能瓶颈。我们通过对人类头部运动进行大量分析后发现其中的刚性运动占主导地位，同时还发现用户面部在相邻两帧之间的非刚性运动通常不会过于剧烈。因此，为了快速获得头部姿态，采用刚性迭代最近点（Iterative Closet Point，ICP）算法[31]求解两个点云间的相对平移和旋转。

算法 1.　刚性迭代最近点算法.

将待匹配的两个点云分别记为 C_1 和 C_2：

步骤 1. 对于 C_2 中的每个点，在 C_1 中寻找距其最近的点。C_2 在 C_1 中的最近点集合记为 $C_2^{'}$，C_2 和 $C_2^{'}$ 中的点存在一一对应关系；

步骤 2. 计算协方差矩阵 M：

$$M=\frac{1}{n}\sum_{i=1}^{n}\left(C_{2i}^{'}-C_{2m}^{'}\right)^{T}\left(C_{2i}-C_{2m}\right)$$

其中：

$$C_{2m}^{'}=\frac{1}{n}\sum_{i=1}^{n}C_{2i}^{'}\quad,\quad C_{2m}=\frac{1}{n}\sum_{i=1}^{n}C_{2i}$$

步骤 3. 对 M 进行奇异值分解（SVD， Singular Value Decomposition）：$M=UWV^{T}$

步骤 4. 计算旋转矩阵 R 和平移向量 t：

$$R=UV^{T},\quad t=C_{2m}^{'}-C_{2m}\cdot R$$

步骤 5. 用 R 和 t 更新 C_2，并重复上述步骤，至收敛。

而用户面部在第一帧的三维点云则和一个标准模型进行匹配以得到该用户的初始头部姿态。由于 Kinect 深度图存在一定的误差，因此基于深度图的姿态估计结果也必然存在误差。在一般情况下，这个误差值会不断变化，造成平移和旋转存在抖动的现象，使用窗口平滑方法可消除因深度图误差带来的姿态抖动[32]。

3.1.2 特征点的提取

算法以 Kinect 彩色图像作为输入，采用主动外观模型（Active Appearance Model，AAM）算法[33]提取面部特征点。在头部不发生旋转的情况下，2D AAM 算法可以比较准确的提取面部特征点，然而实际情况中用户不可能始终保持头部正对相机的姿态，面对头部的转动 2D AAM 算法特征点提取的精度会大幅降低，获得的二维坐标也将无法正确反映出面部关键点的真实位置关系。因此，为了保证算法在用户头部发生旋转的情况下依然能够鲁棒得获取到该用户的面部特征点，需要将 2D AAM 扩展到 3D [34]，利用 3D AAM 算法来完成面部特征点的获取。3D AAM 不但需要用户的头部姿态信息，而且要求面部形状基是三维的。考虑人脸面部表情的变化为一种退化变形，我们选择了一种低秩的形状变形模型对一系列 2D 面部形状基进行三维重建[35]。这类算法称为运动恢复非刚性三维结构算法（Non-rigid Structure from Motion，NRSfM），可以有效的抑制噪音和丢失数据人脸 3D 结构重建的影

响。针对 150 帧的一组人脸表情序列，提取 68 个特征点，重建的 3D 结构效果如图 2 所示。

经典的 2D AAM 算法中任何形状 s 都可以表示为一个基本形状 s_0 和一系列形状基 s_i 的线性组合；任何外观也都可以表示为基本外观 $A_0(x)$与一组外观基 $A_i(x)$的线性组合：

$$s = s_0 + \sum_{i=1}^{n} p_i s_i \tag{1}$$

$$A(x) = A_0(x) + \sum_{i=1}^{m} \lambda_i A_i(x) \tag{2}$$

其中 $s = (x_1, y_1, x_2, y_2, \ldots, x_n, y_n)^T$，$x_i$，$y_i$ 分别是第 i 个面部特征点的 x 和 y 坐标，n 是面部特征点的个数。组合系数 p_i 称为形状参数。x 表示形状 s_0 中的所有像素，$A(x)$ 表示 x 的外观（即像素值）。

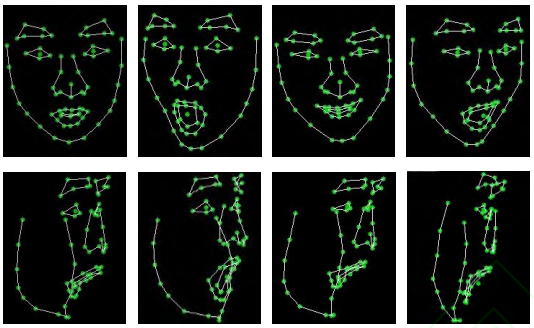


图 2 NRSfM 算法下人脸形状重建

然而为了正确处理头部的三维运动，我们需要将 2D AAM 扩展到 3D AAM。过程中，首先要根据 2D AAM 的基本形状 s_0 和形状基 s_i 恢复出各个形状基代表的面部特征点的二维坐标，记为矩阵 W 。然后，将其每一列均减去平均列向量，以获得均值化的测量矩阵 $\hat{W}$，并利用低秩形状变形模型分解测量矩阵 $\hat{W}$ 为三维形状矩阵 S 与摄像机投影矩阵 R 的乘积。其中投影矩阵的正交性可用来估计表面和摄像机在每帧的相对位置，实现求解矩阵 S。而求解出的三维形状矩阵 S 则可以应用 PCA 分解计算出所需的三维基本形状和形状基。

$$\hat{W} = RS = \begin{pmatrix} R_1 & & \\ & \mathrm{O} & \\ & & R_F \end{pmatrix} S \tag{3}$$

与 2D AAM 算法只有一个优化项不同，3D AAM 的优化项不但包含输入图像和 AAM 模型重建之间的误差，而且还包含了面部区域的 3D 重投影误差。但是由于 3D 形状基是由 2D 形状基根据 NRSfM 算法生成的，在三维形状参数和二维形状参数之间存在一一对应关系。因此 3D AAM 算法并没有增加未知量，使用期望最大化 EM 算法即可对其进行迭代求解。

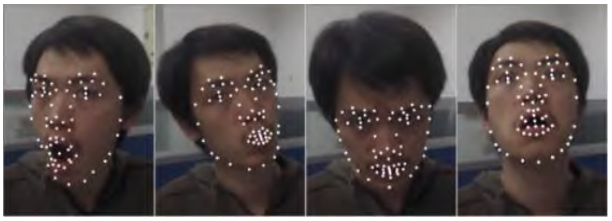


图 3 面部特征点提取效果

如图 3 所示，结合了头部姿态信息的 3D AAM 算法可以应对不同的头部姿态和面部表情，提取特征点的位置比较准确。

3.2 面部表情参数的实时获取

带有语义信息的表情参数关联着面部特征点和人脸表情单元，因此实时获取表情参数是驱动表情基生成动画的关键。现有的面部表情参数获取算法一般需要从用户预采集的表情序列中学习先验知识，以实现用户头部姿态和面部表情的解耦合。因此，普适性和易用性较差。且使用者必须具备采集表情、设定系统等专业知识。而采用无监督的方式对用户的面部表情进行实时捕获的方法，最大的优点在于不需要对待捕捉用户进行任何监督式的训练，通过自动数据分析即可提取出用户的面部表情参数。这使得在此基础上实现的面部动画生成算法具有更加良好的普适性和易用性。

3.2.1 几何度量样本集

面部表情编码系统 FACS 由 P. Ekman 等于 1978 年提出，其根据人脸各部分肌肉功能的不同将面部表情划分为若干个相互独立的表情单元 AU。通过选取不同的表情单元进行组合，就可以得到不同的表情。AU 可以通过面部特征点之间的位置关系进行度量。因此，通过几何度量值将面部特征点的坐标位置和 FACS 的表情单元 AU 关联起来，从而在获取的面部特征点和表情语义之间建立了联系。经过分析，选取如下几何度量值：嘴部：上下嘴唇高度差、嘴巴宽度、上下嘴唇水平距离；眼部：眉眼高度差（左、右）、上下眼皮高度差（左、右）、双眼内眼角间距；鼻子：鼻孔内眼角高度差（左、右）。为避免缩放带来的误差，将这些特征点间的绝对距离进行归一化，分别除以双眼内眼角间距：$g'=g/w$，其中 g 表示某个几何度量值，w 为双眼内眼角间距（见图 4）。

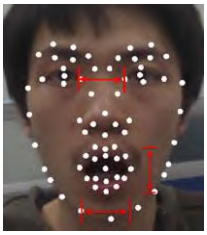


图 4　几何度量示意图

将输入视频中每帧所对应的几何度量值组合，形成一个几何度量样本；并将每一帧所对应的几何度量样本缓存起来形成样本集，为面部表情参数的实时获取提供分析数据。而该过程主要存在两个问题：第一，由于 AAM 算法存在一定的误差，因此无法保证几何度量值的准确。带有误差的样本添加到样本集会对自学习产生负面的影响；第二，样本集中样本的数量随着在线捕捉时间的增加会不断地增多。尽管样本的增加会使得学习越来越准确，但是由于存储空间的限制，如果不控制存储样本的数量，样本集会发生溢出现象。

针对第一个问题，经过反复实验观察，发现 AAM 算法获取的面部特征点的误差存在一定的规律：当面部基本正对相机且距离适中时，特征点的稳定性较好，此时误差很小；随着面部的转动，或是与相机距离过近或过远时，误差逐渐增大。因此，可通过估计样本潜在的误差大小，并设置匹配的权重值，来降低误差对自学习过程的负面影响。设样本权重为 ω，则有 $\omega=\omega_t\cdot\omega_R$，其中 ω_t 为平移权重，ω_R 为旋转权重，并有：

$$\omega_t=1-\left|t_z-z_{opt}\right|\times 0.2 \tag{4}$$

$$\omega_R=\left(\max\left(\vec{n}_c\cdot\vec{n}_f,0\right)\right)^{\alpha} \tag{5}$$

其中 t_z 是当前面部到相机的距离，z_{opt} 是面部到相机的最佳距离，$\vec{n}_c$ 是面部当前的单位法向量，$\vec{n}_f$ 是面部正对相机时的单位法向量，α 是旋转权重衰减因子。分别取 $z_{opt}=0.5$，$\alpha=0.5$。同时，引入补偿样本的概念，目的是用来弥补被错误信息掩盖的正确信息。假定样本误差为高斯误差，样本真实值服从以观测值为期望，以某一与权重相关的值为方差的高斯分布：$s_{v,\omega}\sim N\left(\mu,\sigma^2\right)$，其中 $s_{v,\omega}$ 表示观测值为 v，权重为 ω 的样本的真实值，$\mu=v$，$\sigma=-\ln\omega$。当样本权重 $\omega<1$ 时，从该样本观测值的左右两侧各取一个补偿样本，使其权重为 $(1-\omega)/2$，然后将补偿样本一起加入到样本集。设补偿样本的值为 v'，满足 $\omega/\left((1-\omega)/2\right)=f(v)/f(v')$，其中 f 为高斯分布的概率密度函数，通过求解即可计算出 v' 的值。值得注意的是，当样本权重 $\omega<1/3$ 时，补偿样本的权重高于观测样本的权重，其潜在误差过大，被视为无效样本，需要删除。

针对第二个问题，与直接丢弃新增样本或者丢弃包含新样本在内的最低权重样本的方法不同，本文使用样本合并策略来避免样本溢出。该策略可以保持样本总数不变、保证样本集的自我完善能力并能够反映当前样本的分布密度。以插入一个新样本 $s_1(v_1,\omega_1)$ 为例，从样本集中找到它的最近邻样本 $s_2(v_2,\omega_2)$，将这两个样本合并为 $s_3(v_3,\omega_3)$ 替换 s_2 即可，其中：

$$v_3=\frac{v_1\omega_1+v_2\omega_2}{\omega_1+\omega_2} \tag{6}$$

$$\omega_3=\omega_1+\omega_2 \tag{7}$$

3.2.2　单/双向表情参数提取

根据样本集提取表情参数是无监督面部表情捕捉算法的核心。首先对样本集中样本的变化空间进行估计，然后再对面部表情参数进行提取。注意到在众多表情单元 AU 中，有些表情单元构成了单向变化的表情，例如张嘴、闭眼；而有些表情单元则构成互为反向变化的表情例如撅嘴和咧嘴。因此，对于不同类型的表情，将采取不同的方式来提取参数。

单向表情变化区间的估计相当于对其左右两个节点值 $g_{\min}$ 和 $g_{\max}$ 进行估计（如图 5 所示）。首先从样本集中最左侧样本开始依次向右扫描并计数，若相邻样本距离大于样本集宽度的 l%，则移除左侧的样本，并重新计数；若计数达到样本总数的 k%，或已扫描的样本总数达到 m%，则终止算法。算法终止后，最外侧的样本为区间左右节点，其中条件值的选择根据经验决定。双向表情变化区间的估计相当于对其左中右三个节点值 $g_{\min}$、g_{rest} 和 $g_{\max}$ 进行估计（如图 6 所示）。由于多了中间节点，估计难度有所增加，因此需要在以下三点假设下对表情区间的节点进行估计：1.样本足够充分；2.g_{rest} 附近样本相对较多；3.夸张表情较少。在估计过程中，不断地对最近邻样本进行合并，直至剩下三个样本。

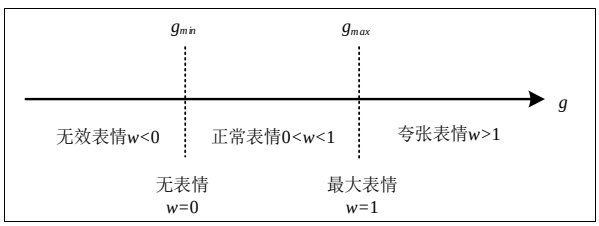


图 5　单向表情变化区间

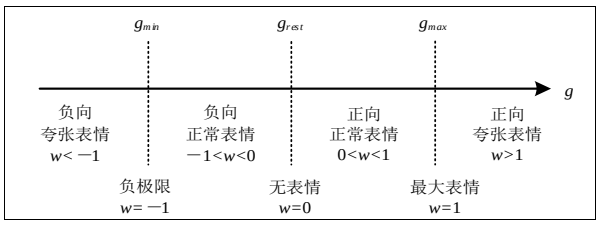


图 6 双向表情变化区间

节点的估计值确定后可知表情的分布区间，即可计算各个 AU 的组合系数 w，从而完成面部表情参数的提取。其中单向表情根据公式（8）计算，双向表情根据公式（9）计算：

$$w=\frac{g-g_{min}}{g_{max}-g_{min}} \tag{8}$$

$$w=\begin{cases}\dfrac{g-g_{rest}}{g_{max}-g_{rest}} & g\geq g_{rest}\\ \dfrac{g-g_{rest}}{g_{rest}-g_{min}} & g<g_{rest}\end{cases} \tag{9}$$

4 基于 AOI 的表情基生成

实时面部表情捕捉技术是表情动画生成算法的基础。而在实现实时表情驱动的面部动画过程中，表情基的质量也非常关键，将直接影响到面部动画最后的生成效果。由于 Blendshape 模型是一种线性模型，具有求解方便、数据量小、与面部复杂度无关等优点，非常适合用于表情存储、识别、动画驱动或远程传输等应用场合。因此，采用 Blendshape 模型来描述面部表情，用 FACS 表情单元作为 Blendshape 表情基。对于一个自然表情的三维模型来说，通常需要变形才能生成 Blendshape 表情基（如图 7 所示）。

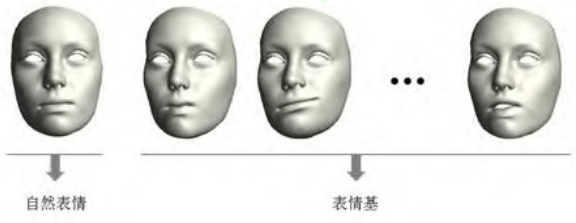


图 7 Blendshape 表情基示意图

其中拉普拉斯变形算法的第一步是进行坐标变换。以某三角网格为例（见图 8），v 是网格中的某个顶点，v_i 是 v 的邻居顶点（图中 $i=1,2,\mathrm{L},5$），以 vv_i 为公共边的两个三角形的相对内角分别记为 α_i 和 β_i，若用 l 表示顶点 v 的拉普拉斯坐标，则有：

$$l=\sum_{i=1}^{n}\omega_i\left(v_i-v\right) \tag{10}$$

其中 v_i 和 v 均默认表示对应顶点的欧氏坐标，权重 ω_i 的计算方式为：

$$\omega_i=\frac{1}{2}\left(\cot\alpha_i+\cot\beta_i\right) \tag{11}$$

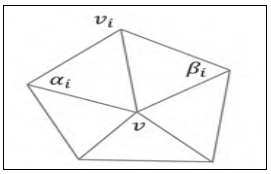


图 8 三角网格示例

为了进一步控制变形，引入控制点的 AOI 来限制变形，从而更好的抑制了变形失真。处理过程中，当一个控制点发生移动时，受其影响的区域大小通常与控制点的移动距离成正比，位于控制点附近的顶点在变形中将会改变其原有的局部特征，而远离控制点的顶点则可以保持局部特征。因此，算法将控制点的运动假想为产生面部变形的力，通过模拟力在面部网格上的传播来计算控制点的影响区域。记 d_v 为顶点 v 在变形中的位移，控制点 d 的初值为其位移，非控制 d 的初值为 0。记 v_i 表示顶点 v 的相邻顶点，对于非控制点来说，通过一次拉普拉斯平滑可以计算出顶点 v 的新位移：

$$d_v^{'}=\frac{\sum_i\omega_i d_{v_i}/l_i}{\sum_i\omega_i/l_i} \tag{12}$$

其中 $l_i=\|v-v_i\|$，ω_i 是顶点的权重。对控制点使用较大的权重，非控制点使用较小的权重，从而加强控制点对临近顶点的影响作用。同时，计算各个顶点的形变因子 $\delta_v=\min\left(\left|d_v\cdot s\right|,1\right)$，该因子值的大小决定了顶点在变形中局部特征的变化程度。不在 AOI 中的顶点形变因子值为 0，处于 AOI 中的顶点的形变因子值在 $[0,1]$ 范围内。将三维模型 n 个顶点变形因子写成对角阵，则得到三维模型的形变矩阵 D：

$$D_{n\times n}=\begin{pmatrix}\delta_1 & 0 & 0\\ 0 & \mathrm{O} & 0\\ 0 & 0 & \delta_n\end{pmatrix} \tag{13}$$

得到 AOI 之后，需要加速变形求解，预先计算好模型的拉普拉斯坐标变换矩阵 M，并在求解过程中保持 M 不变。具体过程如下：

过程 1. 变形求解

步骤 1. 计算模型的拉普拉斯坐标变换矩阵 M 和形变矩阵 D；

步骤 2. 用模型变形前各顶点的欧式坐标初始化 V；

步骤 3. 计算拉普拉斯变形坐标 $L = M * V$，记 $L_0 = L$；

步骤 4. 迭代以下步骤直至收敛：

a. 已知 L 和 M，以控制点目标位置的欧氏坐标为约束反求 V；

b. 计算新的拉普拉斯坐标 $L' = M * V$；

c. 将 L' 的长度缩放至 L_0 的长度，$L'' = \frac{L_0}{L'} \cdot L'$；

d. 引入影响区域约束，$L = DL'' + (I - D)L_0$；

通过上述方式计算的表情基变形结果可以有效抑制形变因子值较低的顶点带入局部特征变化，这些顶点通常位于远离控制点的位置，因此变形的局部性得到了进一步的加强。本文使用 67 个 Blendshape 表情基，提取 68 个特征点，引入 AOI 后的局部基础变形见图 9。

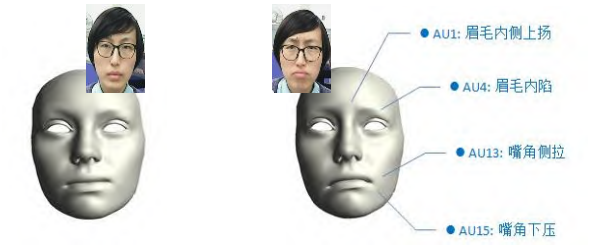
人脸图像　　三维模型变形效果

图 9　引入 AOI 后的基础变形效果

5 实验结果分析

本文实验使用 Kinect 作为实验采集设备，在装有 win7 64 位操作系统、主频 3.7GHz、内存 8GB，配有 Intel(R) Xeon(R)系列 CPU、NVIDIA GeForce GTX750 显卡的台式机上进行实验。实验主要包括三部分，分别是针对算法的鲁棒性分析、实时性分析以及与其它算法的对比分析。

5.1 鲁棒性分析



自然表情　　悲伤

喜悦　　生气

图 10 反映情绪的表情动画效果

单个用户在 Kinect 前方转动头部并展示不同表情以形成实时视频序列时，算法不仅可以针对不同的头部姿态稳定地捕捉用户的表情，还可有效地展现表情细节（如挑眉、皱鼻子等）。观察图 10 可以发现，用户表情驱动表情基所激活的 AU 均可以准确的体现用户情绪（如悲伤、喜悦、生气等），并生成与驱动者相似的表情动画（见图 11）。

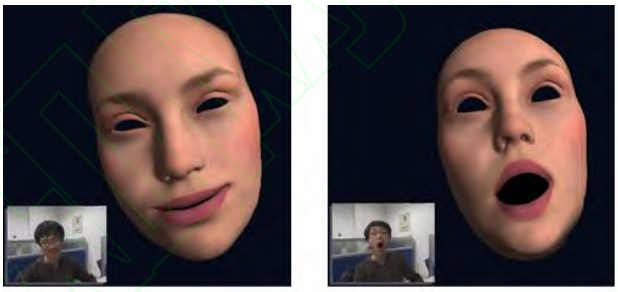


图 11　单人面部表情动画生成效果

图 12　多用户同时出现的面部表情生成效果

本文算法采用无监督的面部表情捕捉算法，无需预采集每名用户的表情数据以训练先验知识。因此在 Kinect 多用户模式下可以鲁棒地处理多名用户同时或交替出现的情况（见图 12）。用户交替过程中，未出现跟踪丢失，并迅速生成与当前被采集者近似的表情动画。

除此之外，本文算法还能够驱动不同的目标模型，鲁棒地生成与用户相似的高精度表情动画。在不同的光照条件下，驱动结果也不会受到影响（见图 13）。这进一步体现了本文面部表情捕捉算法所具有的较强鲁棒性。

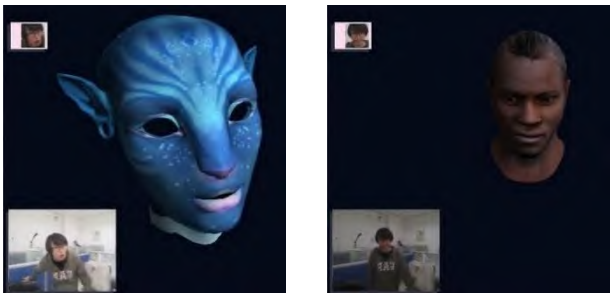


图 13 不同光照下驱动不同目标模型的面部动画效果

在上述实验基础上，本文对遮挡情况下算法的鲁棒性进行了测试，用手对面部进行部分遮挡，如图 14 所示。实验显示在小范围遮挡面部的情况下，算法能够实时捕捉面部数据，并生成与实时视频序列相接近的面部表情动画。但是当控制点区域被遮挡或者 20%的面部特征点被覆盖时，算法将出现跟踪丢失或面部表情动画不准确的现象。造成该现象的根本原因在于，大量或者关键特征点的丢失导致几何度量的失效，以至于表情参数的提取失败。总体评价，本文提出的算法在常规、光照改变、多用户等情况下是足够鲁棒的，并且满足普通用户的使用需求。

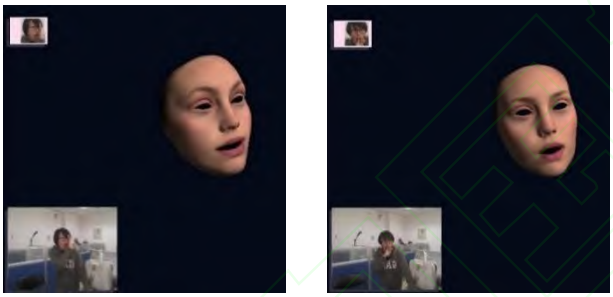


图 14 部分遮挡情况下面部动画生成效果

5.2 实时性分析

自动记录不同实验情况下的每帧耗时及即时帧率。随机选取连续的 1000 帧数据，帧数与每帧耗时情况如图 15 所示，帧数与即时帧率之间的关系如图 16 所示，数据中的最大值最小值及平均值情况见表 1。

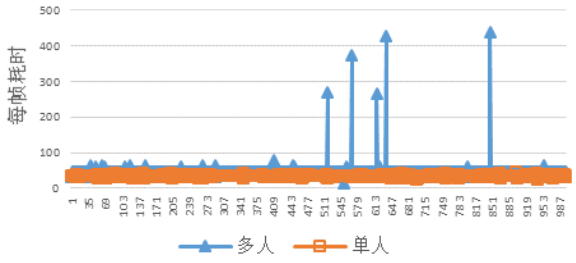


图 15 帧数与每帧耗时之间的关系

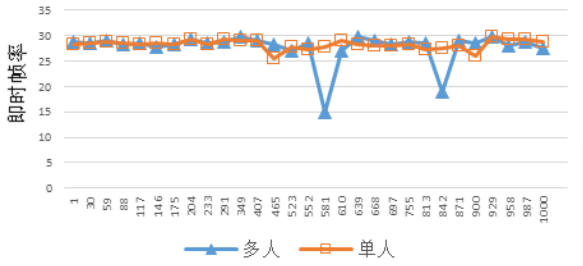


图 16 帧数与即时帧率之间的关系

由图 15 可知，在多人情况下的每帧耗时略高于单人情况，但差别不大，造成差异的主要原因是渲染消耗的时间不同。而图中偶尔出现的耗时很高的帧，是由于在捕捉过程中用户发生了变化，重新检测人脸花费了较长的时间，但算法仍然可以达到实时。由图 16 可知，单人情况下即时帧率基本保持水平，多人情况下帧率基本与单人情况下的数据一致，当新用户出现或者跟踪丢失时，即时帧率会出现临时低谷。如表 1 所示，实验的帧率基本稳定在 28fps 附近，每帧耗时平均不超过 50ms，达到了人眼对实时性的辨别要求。

表 1 每帧耗时及即时帧率的基本情况

	用户数量	最小值	平均值	最大值
每帧耗时 (ms)	单人	25	35.01	45
	多人	15	48.33	437
即时帧率 (fps)	单人	21.33	28.15	29.81
	多人	15.79	27.78	28.82

5.3 对比与评价

本文采用 Kinect 作为数据采集设备，与使用普通摄像头的方法[6]和方法[19]相比，其红外探测器比可见光传感器更能适应不同光照变化。同时 Kinect 还可以获得用户不同站姿、坐姿情况下的数据，使面向普通用户的面部动画生成更具实际意义（见图 11、12、13）。邀请 50 名志愿者对使用两种不同采集设备的面部表情生成算法进行主观评价，每项满分 10 分。其中文献[4]与本文方法使用 Kinect 作为采集设备，文献[6]与文献[19]使用普通摄像头作为采集设备，四种方法的平均得分情况见表 2。

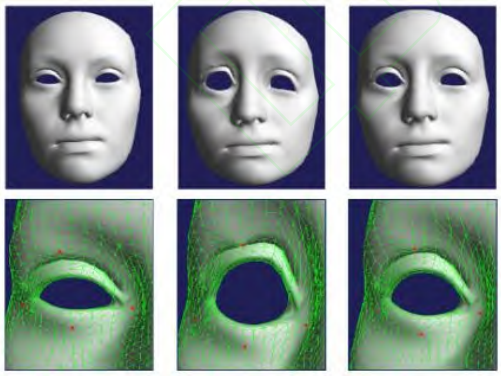
表 2 主观评价结果

	采集灵活度	显示效果	使用方便性	实时性	鲁棒性
文献[4]	9.41	9.39	7.02	9.50	8.79
文献[6]	7.99	8.57	8.76	9.47	6.85

文献[19]	8.37	**9.45**	7.88	9.54	**9.14**
本文	**9.56**	9.43	**9.43**	**9.58**	8.92

由表 2 可知，使用 Kinect 作为采集设备的方法采集灵活度得分更高，而使用普通摄像头的方法，只有摄像头正向面对用户头部时，才能进行动画生成，降低了采集的灵活度，用户体验相对较差。同时，与文献[4]、文献[19]中提出的在使用前必须对每一名用户进行表情数据的预采集以训练得到先验信息的方法相比。本算法可以直接通过无监督的学习获得实时表情参数，无需通过专业操作对用户表情数据进行预采集。不仅如此，本文通过简单的离线表情编辑即可生成不同三维模型的通用 Blendshape 表情基，存储后更方便用户切换使用。因此，本文算法使用方便性得分最高，进一步证明了其更容易在普通用户群中推广。关于实时性方面，四组方法基本都达到了用户的实时性要求，得分近似。而鲁棒性与显示效果两个方面，本文算法仅次于文献[19]中的方法，主要原因在于 Kinect 获得的深度误差处理的还不够完善，在下一步的工作中将重点研究。

与 Huang[2]等使用拉普拉斯变形加最小二乘约束的表情基生成办法相比。我们的表情基生成算法更加鲁棒，能够有效抑制部分变形失真。以眼部和嘴部添加控制点为例（见图 17），结果显示，带有影响区域（AOI）的拉普拉斯变形算法能够很好地实现局部变形，而且极大地抑制了变形失真。而 Huang 等提出的方法在脱离 marker 点获取设备和 3D 扫描仪后则容易产生失真（见图 17b）。



眼部

嘴部

a 变形前　　b 无 AOI　　c 有 AOI

图 17　AOI 变形效果对比图

通过实验分析，发现文献[6]中基于特定用户局部纹理模型的动画生成方法非常依赖特征点的跟踪结果，当头部转动角度较大时，则不能鲁棒地生成表情动画。与之相比，本文方法则可以鲁棒地处理 [–25°,25°] 范围内的头部转动和 [–15°,15°] 范围内的头部俯仰，动画显示效果更加逼真（见图 18）。

如图 18 所示，第一列为实时采集的用户数据帧，第二列为未渲染纹理的实时效果，第三列为添加渲染纹理后的实时效果，第四列为实时生成的 Avatar 面部表情动画效果，图中红色虚线表示铅垂线方向，用于辅助判断头部转动幅度。对比四列效果可知，本文算法可以鲁棒地生成不同人体姿态中转头、侧头等情况下的面部动画，并且动画效果与实际用户表情基本一致，渲染后的动画效果依旧逼真有效。

6 结　语

本文提出了一种利用几何度量的无监督实时面部动画生成算法，首先基于 Kinect 深度图像和彩色图像实现了面部特征点实时提取。然后对特征点进行几何度量并存储为几何度量样本集，采用无监督的方式自动分析样本分布，推测各表情单元的变化区间，实现实时的表情参数提取。算法无需对每一位用户表情数据进行采集和训练，从根本上摆脱了对先验知识和预处理所需专业技巧的依赖。最后提出基于控制点影响区域的拉普拉斯变形算法来生成通用 Blendshape 表情基，提高了表情基的精度，使得实时生成的面部动画更加逼真。实验结果表明本文动画生成算法可以准确地捕捉常规表情并生成高近似度的表情动画；面对部分遮挡或者光照条件变化的情况具有高鲁棒性；并且当多名用户交替使用或同时使用时依旧可以保证实时性。由于

Kinect 自带红外传感器对光照变化适应性强，提供的骨骼跟踪技术支持用户站、坐等姿势的数据获取，使得本文实时的高鲁棒性面部动画生成算法更方便普通用户使用。

尽管如此，本文方法还存在不足之处：1.所选用的采集设备 Kinect，虽然提供了有效的深度信息，但是为捕捉全身数据而设计所使用的宽角度镜头也给算法本身带来了一定的限制。其捕捉的脸部面积占整幅捕捉图片的 10%左右，导致面部细节特征（如皱纹）的丢失。2.本文算法目前在特征点被大面积遮挡情况下会发生跟踪丢失，因此不能生成相应的表情动画。在后续的研究和实验中，我们将尝试相邻帧表情平滑猜测、RGBD 恢复面部细节等思路来提高面部表情捕捉精度和适用范围。

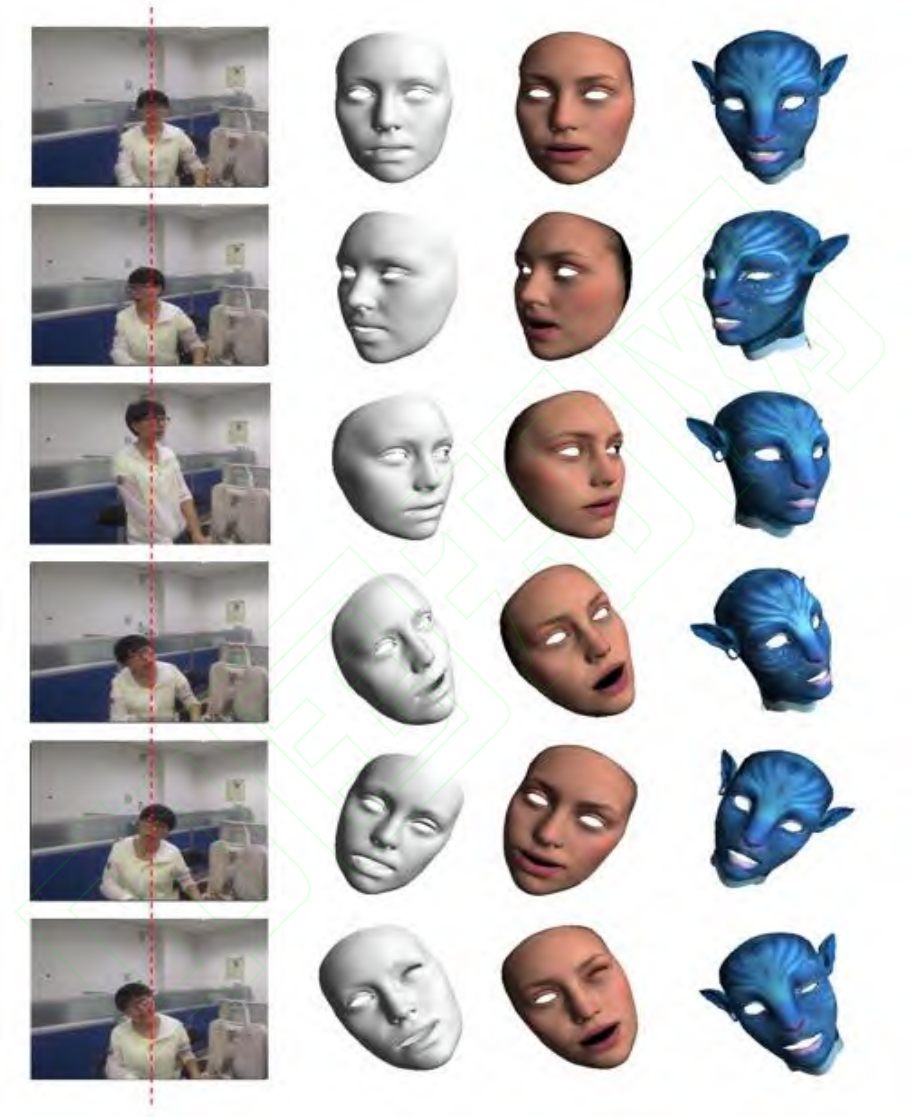


图 18 本文算法实时生成表情动画效果图

致 谢 感谢本文审稿专家和编辑老师所提出的宝贵意见和建议！感谢实验室成员对本文实验测试的大力支持！

参 考 文 献

[1] Wan Xianmei, Jin Xiaogang. Realistic 3D Facial Expression Synthesis: a Survey. Journal of Computer-Aided Design & Computer Graphics, 2014, 26(2): 167-178(in Chinese)
(万贤美, 金小刚. 真实感 3D 人脸表情合成技术研究进展. 计算机辅助设计与图形学学报, 2014, 26(2): 167-178)

[2] Huang H, Chai J, Tong X, et al. Leveraging motion capture and 3d scanning for high-fidelity facial performance acquisition. ACM Transactions on Graphics (S0730-0301), 2011, 30(4): 74:1-74:10

[3] Jana A. Kinect for Windows SDK Programming Guide. UK: Packt Publishing Ltd, 2012

[4] Weise T, Bouaziz S, Li H, et al. Realtime performance-based facial animation. ACM Transactions on Graphics , 2011, 30(4): 77:1-77:9

[5] Cao C, Hou Q, Zhou K. Displaced dynamic expression regression for

real-time facial tracking and animation. ACM Transactions on Graphics , 2014, 33(4): 43:1-43:10

[6] Luo Changwei, Jiang Chen, Li Rui, Yu Jun, and Wang Zengfu. 3D Virtual Facial Animation for General Users. Journal of Computer-Aided Design & Computer Graphics, 2015, 27(3): 492-498(in Chinese)
(罗常伟, 江 辰, 李 睿, 於 俊, 汪增福. 面向普通用户的 3D 虚拟人脸动画. 计算机辅助设计与图形学学报, 2015, 27(3): 492-498)

[7] Parke F I. Computer generated animation of faces//Proceedings of the ACM Annual Conference. New York, USA, 1972: 451-457

[8] Williams L. Performance-driven facial animation// Proceedings of the ACM SIGGRAPH Computer Graphics. Dallas, USA, 1990: 235-242

[9] Bermano A H, Bradley D, Beeler T, et al. Facial performance enhancement using dynamic shape space analysis. ACM Transactions on Graphics , 2014, 33(2):13:1-13:12

[10] Guenter B, Grimm C, Wood D, et al. Making faces//Proceedings of the 25th annual conference on Computer graphics and interactive techniques. Orlando, USA, 1998: 55-66.

[11] Bickel B, Botsch M, Angst R, et al. Multi-scale capture of facial geometry and motion. ACM Transactions on Graphics , 2007, 26(3): 33:1-33:10

[12] Le B H, Zhu M, Deng Z. Marker optimization for facial motion acquisition and deformation. IEEE Transactions on Visualization and Computer Graphics, 2013, 19(11): 1859-1871.

[13] Borshukov G, Piponi D, Larsen O, et al. Universal capture-image-based facial animation for The Matrix Reloaded// Proceedings of the ACM SIGGRAPH 2005 Courses. Los Angeles, USA, 2005: 16-26.

[14] Alexander O, Rogers M, Lambeth W, et al. The Digital Emily project: photoreal facial modeling and animation// Proceedings of the ACM SIGGRAPH 2009 Courses. New Orleans, USA, 2009: 12-22.

[15] Zhang L, Snavely N, Curless B, et al. Spacetime faces: High-resolution capture for modeling and animation. ACM Transactions on Graphics , 2004, 23(3): 548-558

[16] Weise T, Li H, Van Gool L, et al. Face/off: Live facial puppetry//Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation. New Orleans, USA, 2009: 7-16

[17] Beeler T, Hahn F, Bradley D, et al. High-quality passive facial performance capture using anchor frames. ACM Transactions on Graphics , 2011, 30(4): 75:1-75:10

[18] Valgaerts L, Wu C, Bruhn A, et al. Lightweight binocular facial performance capture under uncontrolled lighting. ACM Transactions on Graphics , 2012, 31(6): 187:1-187:11

[19] Cao C, Weng Y, Lin S, et al. 3D shape regression for real-time facial animation. ACM Transactions on Graphics , 2013, 32(4): 41:1-41:10

[20] Chen Y L, Wu H T, Shi F, et al. Accurate and robust 3D facial capture using a single RGBD camera// Proceedings of the International Conference on Computer Vision. Sydney, Australia, 2013: 3615-3622

[21] Wang K, Wang X, Pan Z, et al. A Two-Stage Framework for 3D FaceReconstruction from RGBD Images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(8): 1493-1504.

[22] Platt S M, Badler N I. Animating facial expressions. ACM SIGGRAPH computer graphics. 1981, 15(3): 245-252

[23] Terzopoulos D, Waters K. Physically‐based facial modelling, analysis, and animation. Journal of Visualization and Computer Animation (S1049-8907), 1990, 1(2): 73-80

[24] Waters K. A muscle model for animation three-dimensional facial expression. ACM SIGGRAPH Computer Graphics, 1987, 21(4): 17-24.

[25] Blanz V, Vetter T. A morphable model for the synthesis of 3D faces//Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. New York, USA, 1999: 187-194

[26] Matthews I, Xiao J, Baker S. 2D vs. 3D deformable face models: Representational power, construction, and real-time fitting. International journal of computer vision, 2007, 75(1): 93-113.

[27] Cootes T F, Ionita M C, Lindner C, et al. Robust and accurate shape model fitting using random forest regression voting// Proceedings of the Computer Vision–ECCV 2012. Firenze, Italy, 2012: 278-291.

[28] Bouaziz S, Wang Y, Pauly M. Online modeling for realtime facial animation. ACM Transactions on Graphics , 2013, 32(4): 40:1-40:10

[29] Garrido P, Valgaerts L, Wu C, et al. Reconstructing detailed dynamic face geometry from monocular video. ACM Transactions on Graphics (S0730-0301), 2013, 32(6): 158:1-158:10

[30] Abate A F, Nappi M, Riccio D, et al. 2D and 3D face recognition: A survey. Pattern Recognition Letters (S0167-8655), 2007, 28(14): 1885-1906.

[31] Besl P J, McKay N D. Method for registration of 3-D shapes// Proceedings of the Robotics-DL tentative. International Society for Optics and Photonics. San Diego, USA, 1992: 586-606.

[32] Roberts S W. Control chart tests based on geometric moving averages. Technometrics , 1959, 1(3): 239-250.

[33] Cootes T F, Edwards G J, Taylor C J. Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(6): 681-685.

[34] Xiao J, Baker S, Matthews I, et al. Real-time combined 2D+ 3D active appearance models// Proceedings of the CVPR. Washington, USA, 2004: 535-542.

[35] Zhou Z, Shi F, Xiao J, et al. Non-Rigid Structure-From-Motion on Degenerate Deformations With Low-Rank Shape Deformation Model. IEEE Transactions on Multimedia, 2015, 17(2): 171-185.

JIANG Na, born in 1989, Ph.D. candidate. Her research interests include 3D facial animation and object tracking.

LIU Shao-Long, born in 1988, master. His research interests include 3D facial animation.

SHI Feng, born in 1982, Ph.D.. His research interests include

representation motion segmentation and computer vision.

ZHOU Zhong, born in 1978, Ph.D., associate professor. His research interests include virtual reality and so on.

Background

The problem of 3D facial animation is the important research subject in computer graphics, can dating back to the 1970's. It's the core problem in the field of video production, game and social networking services etc. Its goal is to generate 3D facial animation consistent with expression of users. In general, facial animation need capture device to interact with users. Capture system based marker point is the most common in the field of film and television production. Kinect and monocular equipment have been growing in popularity through online games. Therefore, methods using the three devices are the hotspot of research. In the process, relevant algorithms mainly include two steps: capture facial expressions and generate facial animation. Researchers are committed to improve the robustness and accuracy of the technology and have made great progress in recent years. Some of current algorithms with monocular equipment can generate facial animation in real time. However, the kind of approaches still need preprocess the data of expressions as input. Meanwhile, other state-of-the-art approaches depend on expensive instrument and apparatus. It's difficult to expend to ordinary consumers. Therefore, it's still a huge challenge that design a realistic algorithm of facial animation and expand universality.

In this paper, we have researched on 3D facial animation and proposes an unsupervised real-time algorithm of facial animation by geometric measurements. In the first stage, the algorithm uses geometric measurements to deal with input come from different users, then build sample dataset according to the strategies of weight and compensation. Thereby, expression parameters can be extracted in real time. In the second stage, we first introduces area of influence of control points (AOI) to further improve the accuracy of the universal Blendshape expression base. Finally, realistic facial animation without any preprocessing can be generated in real time. Experimental results show that the algorithm is more robust and accurate than some traditional methods. And it can be expanded to the level of ordinary user.

This work is supported by the High Technology Research and Development Program (863) of China under Grant of 2015AA016403, and the National Natural Science Foundation of China under Grant of 61170188.