

面向网上人际交流的便捷人脸动画

戴 鹏 徐光祐

(清华大学计算机科学与技术系 北京 100084)
(清华信息科学与技术国家实验室 北京 100084)
(daip02@mails.tsinghua.edu.cn)

摘 要 面向网上人际交流的实际需求,提出一种基于图像的便捷人脸动画方法.基于同一人脸的正面和半侧面图像,引入三维通用人脸模型估计其初始朝向,分别建立两图像的二维网格模型;然后采用图像插值和变形技术合成不同朝向人脸,并合理地解决遮挡问题.面部的唇动和表情则可将 MPEG4 三维模型的形变参数投影到相应的二维网格上.该方法无需重构对象人脸的三维模型和对摄像机进行标定,可同时合成面部表情与小范围朝向变化.实验结果证明了文中方法的便捷性、跨平台通用性,合成效果良好.

关键词 便捷人脸动画;基于图像绘制;姿态可变;MPEG4;以计算机为媒介的人际交流
中图法分类号 TP391

Facile Face Animation toward Computer-Mediated Communication

Dai Peng Xu Guangyou

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084)
(National Laboratory of Information Science and Technology, Tsinghua University, Beijing 100084)

Abstract To meet requirements of computer-mediated communication, an image based facile face animation approach is presented. Based on the two face images of the same person at front and half-profile view and an MPEG4 3D generic face model, we estimate the poses of the face at these two images and establish their 2D mesh models. Multiple views of the same face can be synthesized by image interpolation and warping, meanwhile occlusion problems are solved reasonably. Lip motion and facial expression are synthesized by projecting the FAP parameters of the MPEG4 3D model onto the current view. Our approach does not need to establish the 3D model of the individual face nor perform camera calibration. Experimental results demonstrate its efficiency, universality across platforms and acceptable image quality.

Key words facile face animation; image based rendering; pose estimation; MPEG4; computer-mediated communication

如何使计算机更好地服务于人际交流,提高以计算机为媒介的人际交流(computer-mediated communication, CMC)^[1]的效率和质量,使远距离的人与人之间能够自然、便捷、高效率地交流,是当前普适计算与人机交互领域的重要研究问题.人际交流是多模态的,非语言模式如表情体态等在其中占有重要地位.作为人际交流的媒介,计算机不仅应该传输语言信息,还应当传输交流对象的表情、动作等非语言信息.本文研究的面向以计算机为媒介的

人际交流的人脸动画,就是要在低网络带宽或是远方交流者不想暴露自己所处真实环境的情况下,扮演代表远方交流者的视觉窗口或替身的角色,以方便人们通过计算机或掌上设备进行远程的实时交互.具体的应用场景包括虚拟视频会议、远程教学、可视电话、即时通信等.

面向以计算机为媒介的人际交流的人脸动画技术需满足如下要求:1)唇动与语音同步,计算复杂度低,以满足交流的实时性和多种平台计算设备的

要求;2)虚拟人脸能做自然的姿态变化,以模拟说话人的自然头部运动;3)用户建模方便,人工干预少,以适应不同对象的应用需求。

现有的人脸动画方法主要有两大类:基于三维人脸模型的方法和基于图像的人脸动画方法。基于三维模型的人脸动画方法主要分为基于通用人脸模型的形变产生特定人脸模型^[2-3]、基于统计模型生成特定人脸模型^[4],以及直接通过立体图对或视频序列来重建特定人脸模型^[5-6]这几类。基于三维人脸模型的人脸动画方法虽能支持姿态的多角度变化,但是人脸建模过程较复杂,需要恢复人脸深度信息,且用户需进行较多的交互,使用不够便捷。基于图像的人脸动画方法虽然具有纹理真实感强的特点,但是不能妥善处理姿态变化。Ezzat等^[7]基于多幅人脸样本图像训练多维可变形模型用于合成唇动;Tekalp等^[8]在图像上定义符合MPEG4人脸动画标准的二维网格用于动画合成;但是它们均局限于正面姿态,无法实现深度信息的获取及遮挡处理,因此也无法合成说话人头部姿态的变化。

将基于图像的方法与三维人脸模型结合起来是解决姿态变化的一种途径,Liu等^[9]引入三维点分布模型来生成不同姿态下的人脸形状,并结合多姿态人脸样本集来合成纹理;但此方法需获取大量人脸图像,并需恢复人脸三维信息,同样不满足便捷性的要求。文献[10]通过人脸特征点将一张正面人脸照片映射到MPEG4通用三维人脸模型的网格上,实现通用形状特定纹理的动画;但是此方法并未保留特定人脸的形状信息,因此无法建立个性化人脸形状模型,并在动画中造成一定的人脸形状走样。

针对现有的基于三维人脸模型和基于图像的人脸动画方法在面向人际交流应用中的局限性,本文提出一种基于图像的唇动、表情与姿态变化并存的人脸动画方法。该方法融合了这两类方法的优势,保持了个性化人脸的形状和纹理的真实感,较好地平衡了唇动、姿态以及便捷性这三者之间的矛盾,并开发了面向人际交流应用的多平台人脸动画系统。该系统仅需用户使用摄像机拍摄2个姿态下的人脸图像即可自动完成虚拟人脸建模,无需摄像机标定,也无需重构个性化三维人脸模型,因此方便、快捷。

1 方法概述

人脸动画需要合成两方面的人脸运动信息:反映头部姿态变化的刚体运动和反映面部表情和唇动的

非刚体运动。以计算机为媒介的人际交流系统需要将交流对象的信息通过网络进行传输,因此本文中考虑采用符合网络通信标准的MPEG4人脸动画^[11]。

传统MPEG4人脸动画以三维模型来描述人脸对象,会带来建模复杂的问题。为降低复杂度,考虑将三维模型和基于图像的方法结合起来,在图像上建立二维网格形状模型,并将MPEG4人脸动画标准投影到图像平面上,利用图像变形技术合成动画。但是,已有的处理方法^[8]并未能引入姿态的变化。如何在基于图像的MPEG4人脸动画中引入姿态变化就是本文研究的主要内容。

要在基于图像的人脸动画中解决姿态变化的问题,除了如文献[4]建立合适的二维网格形状模型以外,更为关键的是需要确定二维网络形状模型随姿态变化的关系,而姿态的变化又是相对于初始图像中的人脸姿态而言的。因此,归根结底需解决2个问题:1)初始人脸姿态估计;2)二维网络形状模型随姿态变化的关系。现有的方法正是因为未能解决这两方面的问题而无法实现姿态变化。本文处理的姿态变化包括水平方向上的变化及轻微点头动作,其中水平方向姿态变化较大。这是由于多人交互场景中虚拟人脸的视角变化主要发生在水平方向上,具体表现为注视不同对象,因此对水平方向姿态变化做重点处理,考虑以正面和半侧面姿态的人脸图像作为输入。同时,通过同步引入正面和半侧面人脸图像,能有效地降低头部姿态变化时人脸的形状畸变,并指导遮挡部分的纹理处理。本文中采用的正面和半侧面人脸图像在获取时需保持人脸与摄像机之间的距离近似相等,且人脸保持中性状态。

本文提出的便捷人脸动画方法分步解决了虚拟人脸姿态变化与面部动作的合成问题:1)为处理姿态变化问题,在初始两姿态图像上分别定位人脸特征,引入简化的MPEG4三维通用人脸形状模型,并根据特征约束将其映射到两姿态图像上,由此得到表示人脸形状的二维网格形状模型。基于网格模型的对应,通过图像插值技术合成中间姿态人脸。为解决遮挡问题,本文采用了一种近似方法来定位被遮挡的特征点。2)为处理面部动作合成问题,借助MPEG4人脸动画标准,将三维通用人脸动画规则映射到各姿态人脸,生成基于二维网格形状模型的人脸变形规则,在动画参数驱动下通过图像变形技术合成面部动作。图1所示为本文提出的姿态可便捷人脸动画的基本方法框架。

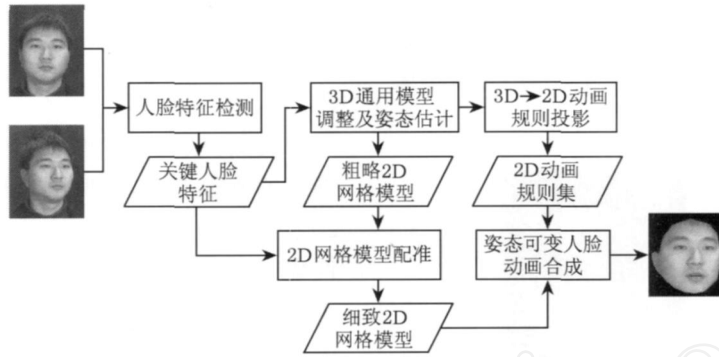


图1 本文方法框架

为简便起见,仅讨论基于正面人脸与单一方向半侧面人脸的处理问题,目前另一方向半侧面人脸依据人脸对称性假设镜像得到.当出现侧面光照时,这种简单的对称处理方法会造成镜像图像光照不真实的情况,为克服这个缺点,可增加另一方向人脸图像作为系统输入.

2 虚拟人脸建模

2.1 虚拟人脸模型

人脸建模是人脸动画系统的基础.为了实现基于 MPEG4 的人脸动画,需要在初始人脸图像上建立二维网格形状模型.为了进一步合成姿态可变的虚拟人脸动画,需要根据当前及初始姿态合成形状与纹理,即需要首先对初始人脸进行姿态估计.形状、纹理与姿态是本文方法的 3 个基本要素,我们提出一种新颖的虚拟人脸模型表示形式

$$m = M(s, I, \theta) \tag{1}$$

该模型由 3 部分构成,其中 $s = (u_1, v_1, u_2, v_2, \dots, u_N, v_N)^T$, 表示二维网格形状模型; I 表示纹理,具体表现为形状模型范围内的人脸图像; θ 表示当前姿态,具体表示为人脸相对于正面姿态绕 3 个坐标轴的旋转角度.

本文中所述的虚拟人脸建模是指在初始的两姿态人脸上分别构建如式(1)所示的人脸模型,而非特定人脸三维模型的重构.本文引入三维通用人脸的形状模型,通过三维模型与初始人脸图像的适当融合来同时解决估计姿态与建立形状模型两方面的问题.文中提到的三维人脸模型均特指三维形状模型,具体表示为 $S = (X_1, Y_1, Z_1, X_2, Y_2, Z_2, \dots, X_N, Y_N, Z_N)^T$,为降低运算复杂度,采用由 95 个顶点构成的简化三维模型,它们是 MPEG4 通用人脸模型^[11]顶点集合的子集.

在进行虚拟人脸建模处理时,选取人脸图像上合适的特征点集作为三维模型向人脸图像投影的依据,根据两姿态下的投影关系分别估计正面与侧面人脸的姿态 θ_1 与 θ_2 ;然后利用投影的结果在 2 幅初始图像上建立二维网格形状模型 s_1 与 s_2 .

2.2 人脸特征定位

人脸特征点是连接 MPEG4 三维形状模型与人脸图像纹理的纽带,也是基于图像建立二维网络形状模型的基础.人脸特征点主要是指那些具有一定语义信息的点,本文选取的 33 个特征点主要集中在眼、眉、口、鼻、脸颊以及面部边缘轮廓等区域,是 MPEG4 特征点^[11]的子集.图 2 所示为特征集合及本文采用的二维网络形状模型.

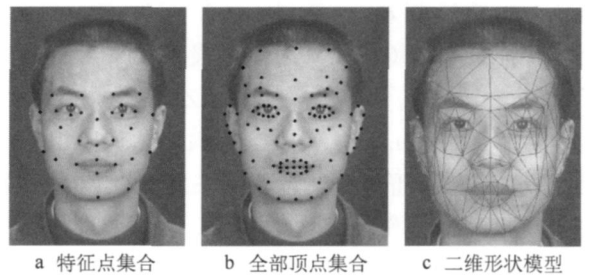


图2 特征点及二维形状模型

为减少用户交互,本文采用分层的直接表观模型方法 (hierarchical direct appearance model, HDAM)^[12]来自动检测面部特征集合.通过构建人脸的形状统计模型、纹理统计模型及映射预测模型,来达到直接表观模型根据纹理信息推断形状信息的目的.分层的直接表观模型利用两层直接表观模型实现由粗到精的人脸特征定位,降低了算法对初始值的敏感度,并提高了鲁棒性.

鉴于 HDAM 方法对于人脸姿态的敏感性,本文进行了分姿态的 HDAM 的训练,并分别用于不同初始姿态下的特征定位.图 3 所示为 HDAM 定位

结果与本文所用特征点之间的关系,HDAM 未确定的部分特征点通过相对插值来定位.对定位不够准确的个别特征点,系统允许用户进行简单的交互调整.

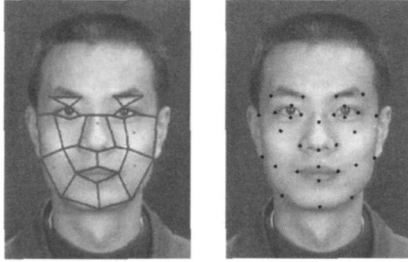


图3 基于 HDAM 的人脸特征定位

2.3 初始人脸姿态估计及建模

引入通用的三维模型以后,本文以一种合理的近似方法来进行初始图像中人脸的姿态估计.首先根据正面人脸图像对三维模型进行粗略的比例调整,以两眼间中点为原点建立参考坐标系,对模型上不同区域的顶点在 X, Y 和 Z 轴方向上进行局部尺度变换,变换标准包括正面图像上眼睛与嘴唇中心间距、两瞳孔间距以及嘴唇宽度等信息,这些信息根据人脸特征定位结果直接获得.

本文采用弱透视投影模型来描述三维模型与各姿态人脸图像之间的映射关系,无需摄像机参数标定,也无需获取确切的人脸深度信息.具体地说,三维模型上的顶点 $P_i = (X_i, Y_i, Z_i)^T$ 与当前姿态图像上的特征点 $p_i = (u_i, v_i)^T$ 之间的投影关系可表示为

$$[u_i, v_i]^T = M[X_i, Y_i, Z_i, 1]^T \quad (2)$$

其中矩阵 $M = \{m_{ij}\} (i = 1, 2, 3, 4; j = 1, 2)$ 既含有摄像机的内参数信息,又包括了三维模型与摄像机之间的相对旋转和平移运动参数.我们分别用 $R = \{r_{ij}\} (i, j = 1, 2, 3)$ 和 $T = (t_x, t_y, t_z)^T$ 表示三维人脸模型与摄像机参考坐标系之间的相对旋转和平移运动;用 f 表示摄像机焦距, k_u 和 k_v 表示水平和垂直方向上单位长度的像素数, p_u 和 p_v 表示成像平面中点的纵横坐标, s 表示比例放缩系数, Z_0 表示三维模型的整体深度.则根据弱透视投影原理,投影矩阵可表示为

$$M = \begin{bmatrix} k_u f s \frac{r_{11}}{Z_0} & k_u f s \frac{r_{12}}{Z_0} & k_u f s \frac{r_{13}}{Z_0} & k_u f s \frac{t_x}{Z_0} + p_u \\ k_v f s \frac{r_{21}}{Z_0} & k_v f s \frac{r_{22}}{Z_0} & k_v f s \frac{r_{23}}{Z_0} & k_v f s \frac{t_y}{Z_0} + p_v \end{bmatrix} \quad (3)$$

而在投影生成正面姿态人脸图像时,我们认为三维人脸模型不作旋转而仅保持平移运动,投影矩阵可简化为

$$M = \begin{bmatrix} k_u f s \frac{1}{Z_0} & 0 & 0 & k_u f s \frac{t_x}{Z_0} + p_u \\ 0 & k_v f s \frac{1}{Z_0} & 0 & k_v f s \frac{t_y}{Z_0} + p_v \end{bmatrix} \quad (4)$$

我们分别以 M 和 M 表示半侧面姿态与正面姿态下的投影矩阵,选取眼、眉、鼻、嘴等标志性区域的 20 个特征点作为对应特征点代入式(2),分别得到两姿态下的线性方程组.通过矩阵奇异值分解,可以求出投影矩阵中各项的最优解.由于在我们的应用中摄像机无需标定,因此式(3)中的摄像机参数是未知的.但第 1 节已提到 2 幅初始图像中的人脸需满足深度近似相等的条件,基于此假设,2 个不同姿态下的比例系数 s 与 s ,以及人脸整体深度 Z_0 与 Z_0 均可近似看作相等.结合式(3)和(4),可以通过矩阵对应项的比例关系求解出两姿态之间的相对旋转参数^[13].由于采用比例关系求解,因此无需对深度信息进行初始化,只要初始两姿态下拍摄的人脸到摄像机的距离近似相等即可.

二维形状模型的构建需要解决 2 个关键问题:确定二维形状模型的初始位置和基于图像特征对其进行精细化调整.初始两姿态下的二维形状模型由三维通用模型在 2 个初始图像上投影而得,图 4a 所示为根据投影矩阵的求解结果将三维模型近似投影在两姿态图像上的效果.由于投影后得到的仅是粗略的网格模型,因此需要对其做精细化调整.

第 2.2 节中自动定位了一组人脸特征,我们的目标是通过这些特征点和粗略二维网格模型的融合,得到描述特定人脸形状的个性化二维网格模型.解决这类散乱数据插值问题的典型方法就是基于径向基函数(radial basis function, RBF)的数据插值方法^[14].本文根据面部特征集合,采用多步骤、紧支持集的 RBF 方法进行平面网格模型的个性化调整,但有别于以往的应用,本文中是针对二维点集进行数据插值.具体来说,就是给定通用人脸模型直接投影产生的二维点集 s 、人脸特征定位产生的特征点集 $\{y_i, 1 \leq i \leq N\}$ 以及特征点集在通用人脸模型上对应的集合 $\{x_i, 1 \leq i \leq N\}$,确定合适的插值函数

$$F(x) = \sum_{j=1}^N C_j \phi(x - x_j); \text{使得 } F(x_i) = y_i, 1 \leq i \leq N.$$

根据插值函数可以完成通用人脸模型上其他非特征点的个性化调整.本文使用紧支持集 RBF 作为插值函数的基函数 ϕ ,对稀疏点集使用平滑性较好的基函数及较大的支持半径,对稠密部分则使用平滑性减弱的基函数并减小支持半径,分步完成插值调整. RBF 插值的结果如图 4b 所示.

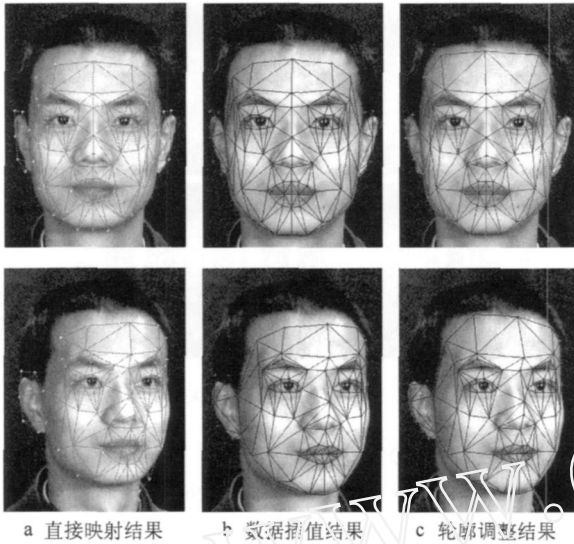


图4 二维网格模型调整

基于 RBF 方法并未妥善处理额头与耳朵等边缘区域的调整,本文参考文献[15]中提出的改进主动轮廓线算法,使用梯度向量流作为主动轮廓线的外在拉动力,同时进一步将人脸肤色信息引入到轮廓线的外在拉动力上,采用多线索融合的方法进行主动轮廓线的调整,以提高精度.主动轮廓线的能量函数 Euler-Lagrange 方程可表示为

$$F_{\text{int}} + F_{\text{ext}} = 0 \quad (5)$$

式(5)可看成对应内部能量的力 F_{int} 与对应外部能量的力 F_{ext} 之间平衡的结果.外部力 F_{ext} 可以融合进多线索的图像信息,我们将反映图像边缘的梯度向量场^[9]和反映图像颜色分布的区域分割信息一起作为外部力, $F_{\text{ext}} = [(\alpha_1 F_{\text{GVF}} + \alpha_2 F_{\text{Seg}}) \cdot N] / N$;其中, F_{GVF} 是梯度向量流的力, F_{Seg} 是根据轮廓线附近的前景区域的肤色概率置信度量所确定的力, N 是在轮廓采样点处朝内的单位法线向量, α_1 和 α_2 是 2 种力的调节系数.肤色模型通过从 RGB 空间变换到 HSV 空间,并建立肤色区域颜色直方图的方式来表示.对当前图像区域中的像素计算其属于肤色区域的概率置信度,再加以阈值化,即可得到肤色前景区域分割的结果.肤色区域分割结果与梯度向量流结合起来共同确定主动轮廓线.

本文对 3 段固定端点的轮廓线进行调整,分别是左、右耳以及额头的轮廓线.固定轮廓线的两端点可以使主动轮廓线算法不至于发生较大的位置偏移或过度收敛.基于改进主动轮廓线的轮廓定位结果如图 4c 所示.

3 人脸动画合成

本文方法首先根据初始图像及形状模型通过插值或图像变形合成姿态变化后的人脸,然后将 MPEG4 规则投影到当前姿态人脸以合成面部的唇动或表情.

3.1 合成姿态变化

多姿态的人脸合成是通过对对应形状模型的插值或图像变形完成的,其关键是建立不同姿态下网格模型顶点之间的对应.根据人际交流的实际需求,本文主要研究水平方向上一定范围内的姿态变化以及轻微的点头动作.由于多人交互情景中水平方向姿态变化范围相对较大,会产生面部区域的遮挡问题,且不同姿态下的网格模型不能建立起完全的顶点对应,因此需要重点处理姿态变化时的遮挡问题.

在基于三维模型的人脸动画中,可以根据顶点的深度信息来判断遮挡,而对基于图像的人脸动画而言,三维信息的缺乏使得系统无法根据深度来进行遮挡判断.文献[16]通过建立多姿态形状统计模型,将正面姿态下参考序列中的人脸表情动作复制到多个不同离散姿态下的虚拟人脸上,但不适合应用于针对特定人的实时动画系统,并且未处理姿态连续变化的问题.

为了解决这一矛盾,本文采用一种便捷的方法来近似处理遮挡问题.将各姿态下的二维形状模型 s 分为 2 个区域:一部分定义为内部区域 s_{I} ,各姿态下均可见;另一部分为外部区域 s_{E} ,随姿态变化会发生遮挡.图 5 所示为人脸的区域划分法,其中灰色三角区域表示可能发生遮挡的外部区域.为使图像插值方法有效,需要对半侧面姿态下被遮挡的顶点进行定位,以保证不同姿态下网格模型的各项点均能在半侧面姿态下找到对应点.

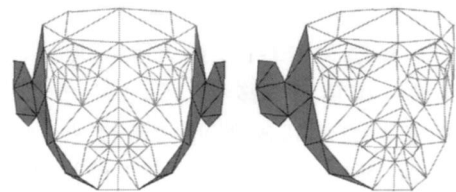


图5 面部遮挡处理

本文中,两初始姿态人脸之间的转动主要是横向旋转运动,并且拍摄 2 幅图像时的人脸整体深度近似相同.基于这样的假设,我们根据一定的位置比例关系,在半侧面姿态下的面部边缘确定与正面

姿态人脸边缘顶点对应的被遮挡顶点. 实际情况下, 被遮挡的顶点并不是恰好处在面部边缘, 因此本文确定被遮挡顶点的方法是一种近似方法, 但是使用便捷, 且实验效果较为理想. 建立起两初始姿态下完整的顶点对应关系后, 摇头过程中各姿态下的二维形状模型由两初始人脸模型插值确定, 即 $s = f(s_1, s_2, \theta)$. 其中, s, s_1 与 s_2 分别表示当前姿态以及初始两姿态图像中人脸的二维形状, θ 表示当前姿态. 在此基础上可进行进一步的纹理合成

$$I(u, v) = f_I(I_1, I_2, \theta)(u, v) \quad s_I,$$

$$I(u, v) = f_w(I_1, \theta)(u, v) \quad s_E;$$

即人脸内部区域 s_I 中的纹理在合成时需由两初始姿态纹理经中间插值 $f_I(\ast)$ 得到; 而外部区域 s_E 中的纹理在合成时仅根据正面姿态纹理进行图像变形 $f_w(\ast)$. 针对内外区域交界处可能出现的纹理突变, 采用适当的像素平滑处理进行调整. 图 6 所示为按照从正面姿态转向半侧面姿态的顺序给出了摇头动作序列中的部分帧.



图 6 摇头序列人脸采样

由于输入人脸图像仅有正面与单个侧面姿态, 为实现头部的双向转动, 本文基于人脸对称性假设, 镜像得到另一方向的半侧面姿态人脸形状 s_3 与纹理 I_3 , 这样可以将初始的两姿态扩展为三姿态, 头部扩展为双向转动.

在处理点头动作时, 由于点头幅度较小, 因此无需考虑面部遮挡问题. 在点头过程中, 简化通用三维形状模型 S 根据 MPEG4 三维人脸动画规则生成不同姿态下的三维顶点位移 $P_i = (X_i, Y_i, Z_i)^T$, 则当前姿态下的形状模型可以通过 $s = s_1 + M \cdot h(P_i)$ 得到, 其中 s_1 为正面姿态人脸形状模型. 因为本文研究工作中的头部上下偏转幅度较小, 投影矩阵 M 可近似用正面姿态投影代替, 所以直接将因点头引起的三维网格变化投影到正面姿态下. 得到形状模型以后, 根据正面人脸变形生成当前姿态人脸纹理 $I(u, v) = f_w(I_1, \theta)$. 图 7 所示为一个点头动作合成序列中的部分帧.



图 7 点头序列人脸采样

3.2 合成面部动作

MPEG4 标准中定义的人脸动画技术是围绕人脸对象和人脸动画参数来进行的, 人脸对象通常用三维网格模型表示, 动画合成则是通过人脸动画参数 (facial animation parameter, FAP) 的驱动及人脸模型的变形来实现的. MPEG4 人脸动画标准中以分段线性函数形式来描述人脸动画规则^[11].

本文运用二维形状模型及相应的动画规则来合成动画, 但有别于传统方法, 这里根据三维通用模型与不同姿态图像之间的投影关系, 将 FAP 引起的人脸形状变化投影到当前姿态人脸, 生成基于二维形状模型的动画规则 $[u_i, v_i]^T = M [X_i, Y_i, Z_i, 1]^T$. 进一步利用 FAP 驱动二维网格模型发生变化, 利用图像变形技术动态地合成唇动和表情等各种面部动作 $s = s + \Delta s, I(u, v) = f_w(I)$.

实际处理中, 将摇头与点头时姿态变化的角度与相应的旋转 FAP 建立起对应关系. 在合成姿态变化时, 仅生成新姿态下的中性人脸形状模型, 然后将面部动作带来的形状变化叠加上去, 生成当前姿态下带有唇动或表情的人脸形状, 再通过纹理插值或变形合成人脸图像.

3.3 人脸动画合成系统流程

根据面向人际交流的计算机系统应用需求, 本文将人脸动画技术为核心的多模态人机交互接口与网络技术紧密结合起来, 其流程如图 8 所示. 系统通过网络接收自然语言文本或音频流, 通过文语转换系统或语音识别系统将其转换为音素流, 再利用音素—FAP 转换器将音素流转换为符合 MPEG4 标准的动画参数流, 最后利用自定义的同步方法在动画参数流和音频流的基础上合成动画. 为增强人脸动画的生动性和真实感, 系统可动态地加入表示姿态变化或面部表情的参数, 以合成伴有姿态、唇动及表情变化的人脸动画.

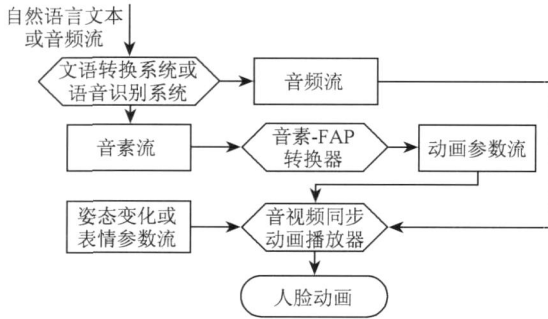


图 8 人脸动画合成系统流程

4 实 验

本文分别在正面和半侧面姿态下采集人脸图像对系统进行测试,摄像机未经标定.合成动画时,将牙齿作为补丁动态地贴入嘴部区域,同时根据输入语音的频率调整动画的播放速度,使之与音频流同步.对于 270 ×300 大小的人脸,本文提出的人脸动画系统在 P 1.29 GHz 处理器的 PC 上能够以 20 ~ 30 帧/s 的正常帧速运行.针对不同目标人脸的动画合成结果如图 9 所示.

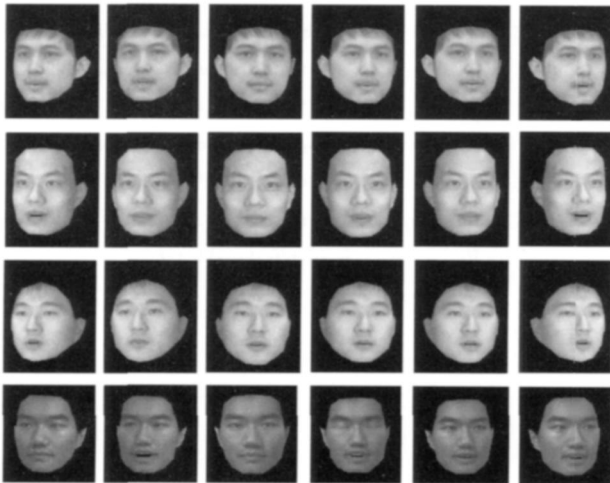


图 9 动画合成结果

本文方法不仅能处理一定范围内的姿态变化,同时能更真实地体现实际空间中的光照变化情况,更好地满足人际交流系统对临场感的要求.从图 9 中不难看出,本文方法在人脸姿态变化过程中,能够较好地呈现出现场面部光照效果,在脸颊、眼角及前额区域显得较为明显.由于目前采用单边姿态图像及镜像方法来合成另一方向姿态的人脸图像,因此仅单向人脸的光照反映实际场景中的光照,而镜像后的人脸不能体现真实光照方向,后续工作考虑通

过引入另一个方向半侧面姿态的初始图像来进行改进.

除了在 PC 及互联网上的应用之外,便捷人脸动画方法还能够被广泛应用于移动设备及无线网络等计算性能相对较差及传输码流相对较低的环境.我们将虚拟人脸建模过程在 PC 上完成,并在移动设备上开发出基于本文方法的人脸动画引擎.经过一系列算法优化处理,本文提出的便捷人脸动画系统在处理速度为 130 MHz OMAP 310 处理器及 Symbian 操作系统的智能手机上能够以 10 ~ 15 帧/s 的速度合成大小约为 90 ×100 的人脸,基本满足流畅播放人脸动画并用于辅助人与人之间交流的实际需求.移动设备上的人脸动画结果如图 10 所示.



图 10 移动设备上的应用

5 结 论

本文面向以计算机为媒介的人际交流的实际需求,提出了一种新颖的基于图像的姿态可变人脸动画方法.基于正面和半侧面人脸图像,引入三维通用人脸模型估计初始人脸姿态并建立多平面形状及纹理模型,借助图像插值和变形技术以及合理的遮挡判断来合成连续的姿态变化,并利用投影关系来合成各姿态下面部的唇动和表情.实验证明,本文提出的人脸动画方法在实际应用中具备三方面的优势: 1) 无需建立特定人脸三维模型,无需摄像机标定,用户交互少; 2) 系统通用性强,计算效率较高,适用于 PC 及移动设备等多种平台; 3) 动画中虚拟人脸的形状保持较好,图像纹理真实感强,同时能妥善处理一定范围内的姿态变化及局部遮挡.

进一步的研究包括: 1) 考虑在初始输入中剔除正面人脸,而直接根据左右 2 个方向上的半侧面姿态人脸合成正面人脸,并据此合成姿态可变的人脸动画; 2) 考虑增加完全侧面姿态的初始人脸图像,将水平姿态变化的范围扩大到完全侧面的情况.

参 考 文 献

- [1] Walther J. Computer-mediated communication: impersonal, interpersonal, and hyperpersonal interaction [J]. *Communication Research*, 1996, 23(1): 3-43
- [2] Pighin F, Hecker J, Lischinski D, *et al.* Synthesizing realistic facial expressions from photographs [C] // *Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH, Orlando, Florida, 1998*: 75-84
- [3] Park I K, Zhang H, Vezhnevets V, *et al.* Image-based photorealistic 3-D face modeling [C] // *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, 2004*: 49-54
- [4] Blanz V, Vetter T. A morphable model for the synthesis of 3D faces [C] // *Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH, Los Angeles, California, 1999*: 187-194
- [5] Mei Li, Bao Hujun, Zheng Wenting, *et al.* Realistic reconstruction of human face based on images [J]. *Chinese Journal of Computers*, 2000, 23(9): 996-1002 (in Chinese)
(梅 丽, 鲍虎军, 郑文庭, 等. 基于实拍图像的人脸真实感重建[J]. *计算机学报*, 2000, 23(9): 996-1002)
- [6] Shen Yehu, Liu Jilin. From stereo pair to 3D face model [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2006, 18(12): 1904-1910 (in Chinese)
(沈晔湖, 刘济林. 利用立体图对的三维人脸模型重建算法[J]. *计算机辅助设计与图形学学报*, 2006, 18(12): 1904-1910)
- [7] Ezzat T, Geiger G, Poggio T. Trainable videorealistic speech animation [C] // *Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH, San Antonio, Texas, 2002*: 388-398
- [8] Tekalp A M, Ostermann J. Face and 2D mesh animation in MPEG4 [J]. *Signal Processing: Image Communication*, 2000, 15(4/5): 387-421
- [9] Liu Y H, Xu G Y, Wang Q. Realistic multi-view face animation with aid of 3D PDM [C] // *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, 2004*: 511-516
- [10] Song Mingli, Zhao Qi, Bu Jiajun, *et al.* An efficient method of face texture mapping directed to portable devices [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2005, 17(12): 2670-2675 (in Chinese)
(宋明黎, 赵 琦, 卜佳俊, 等. 面向移动设备的快速人脸纹理映射[J]. *计算机辅助设计与图形学学报*, 2005, 17(12): 2670-2675)
- [11] Pandzic I S, Forscheimer R. MPEG-4 facial animation—the standard, implementation and applications [M]. New York: John Wiley & Sons, 2002
- [12] Song G, Ai H Z, Xu G Y. Hierarchical direct appearance model for elastic labeled graph localization [C] // *Proceedings of SPIE, Beijing, 2003*, 5286: 139-144
- [13] Dai P, Xu G Y, Riegel T, *et al.* Automatic face modeling and synthesis based on image pairs [C] // *Proceedings of the 4th IEEE International Conference on Computer Vision Systems, New York City, NY, 2006*: 18
- [14] Lavagetto F, Pockaj R. The facial animation engine: toward a high-level interface for the design of MPEG-4 compliant animated faces [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 1999, 9(2): 277-289
- [15] Xu C Y, Prince J L. Snakes, shapes, and gradient vector flow [J]. *IEEE Transactions on Image Processing*, 1998, 7(3): 359-369
- [16] Zalewski L, Gong S G. Synthesis and recognition of facial expressions in virtual 3D views [C] // *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, 2004*: 493-498