

# 基于人脸表情特征的情感交互系统\*

徐红, 彭力

(江南大学物联网工程学院, 江苏无锡 214122)

**摘要:** 设计了一套基于人脸表情特征的情感交互系统(情感虚拟人), 关键技术分别为情感识别、情感计算、情感合成与输出三个方面。情感识别部分首先采用特征块的方法对面部静态表情图形进行预处理, 然后利用二维主元分析(2DPCA)提取特征, 最后利用多级量子神经网络分类器实现七类表情识别分类; 在情感计算部分建立了隐马尔可夫情感模型(HMM), 并且用改进的遗传算法估计模型中的参数; 在情感合成与输出阶段, 首先采用 NURBS 曲面和面片相结合的算法, 建立人脸三维网格模型, 然后采用关键帧技术, 实现了符合人类行为规律的连续表情动画。最后完成了基于人脸表情特征的情感交互系统的设计。

**关键词:** 情感虚拟人; 二维主元分析; 多级量子神经网络; 隐马尔可夫情感模型

**中图分类号:** TP391      **文献标志码:** A      **文章编号:** 1001-3695(2012)03-1111-05

doi:10.3969/j.issn.1001-3695.2012.03.085

## Emotional interaction system based on characteristics of facial expression

XU Hong, PENG Li

(School of Internet of Things Engineering, Jiangnan University, Wuxi Jiangsu 214122, China)

**Abstract:** This paper presented a systematic design method for the emotion interactive system (virtual emotional human) based on the characteristics of facial expression. The key technologies involved emotion recognition, affective computing, emotion synthesis and output. In the part of emotion recognition, firstly, processed the static facial expression by the method of characteristic blocks, then the method 2DPCA was used to extract characteristics, finally applied the multi-level quantum neural network classifiers in achieving seven-class classification of expression. In the part of affective computing, built a hidden Markov model, and estimated the model parameters with improved genetic algorithm. In the part of emotion synthesis and output, establish a three-dimensional mesh face, and then used key frame techniques to achieve a continuous pattern of behavior consistent with human facial animation.

**Key words:** virtual emotional human; two dimension principal component analysis (2DPCA); multi-layer quantum neural network; hidden Markov model (HMM)

## 0 引言

情感处理系统在人机交互、电子教育、远程医疗、娱乐、行为科学测量以及智能体等方面都具有广泛的应用前景和极大的经济价值。赋予计算机某种性格从而能够适应人类情感是实现和谐人机交互环境的关键, 因此构造一个具有自主情感系统的智能虚拟人一直是人类追求的梦想。随着生物和信息科学的高速融合, 国内外对这个包含当今先进科学技术的跨学科、跨领域的新技术都投入了大量的精力, 成为了人工智能领域研究的热点<sup>[1]</sup>。情感虚拟人的关键技术在于情感处理系统。情感信息处理是认知科学和人工智能研究一直关注, 但始终未能彻底解决的问题<sup>[2]</sup>。性格及情感表现能力在人们的感知、计划、推理、学习、记忆、决策、创造性等方面扮演着极为重要的角色, 是人脑的重要功能之一。由于性格和情感具有随机性和不确定性, 使用目前的识别和建模方法效果不太理想, 为了提高识别和建模精度有必要研究新方法, 使用新技术。同时, 人的心理、情感和性格还有形成、成熟和不断进化的特点<sup>[3]</sup>, 将其建模过程和进化理论结合势在必行。可以预计在此研究基础上将能研究出具有各种性格的虚拟人和机器人, 包

括乐观向上的, 也包括悲观失落的, 甚至精神不正常的<sup>[4]</sup>。这不仅为研究个体机器人性格和情感提供平台, 也将为研究机器人人群打下基础。

本文在深入研究前人方法的基础上, 结合前人研究的各种分立理论, 经过改进后设计了一套交互平台进行情感交互实验, 有助于将具有某种性格的情感虚拟人系统嵌入到自主移动机器人中, 这使机器智能更加全面, 具有深远的科学意义和广泛的应用价值。

## 1 情感识别

本文训练和测试是以日本女性表情库(JAFFE)为实验数据库, 该数据库由 10 人的 213 幅图像组成, 每人有七种不同表情(生气、厌恶、恐惧、高兴、中性、悲伤、惊讶)的图像各 24 幅图像。本文首先采用特征块的方法对面部静态表情图形进行原始处理, 滤掉大部分表情干扰信息, 之后进行归一化处理。再利用比 PCA 更有优势的 2DPCA 提取特征。在识别部分, 本文采用了多级量子神经网络分类器实现串级识别, 将 2DPCA 提取后的表情特征输入到四级分类器中, 依次逐级精确分类。

收稿日期: 2011-07-30; 修回日期: 2011-08-29      基金项目: 国家自然科学基金资助项目(60973095)

作者简介: 徐红(1987-), 女, 江苏盐城人, 硕士研究生, 主要研究方向为图像处理与模式识别、情感虚拟人(qixin1227@126.com); 彭力(1967-), 男, 教授, 博士研究生, 主要研究方向为视觉传感器网络、人工智能、计算机仿真。

### 1.1 表情识别

表情识别主要步骤的流程如图 1 所示,包括表情图像预处理、表情特征提取和表情分类三个部分。预处理的优劣直接影响特征提取的准确度,特征提取的准确度对识别率产生重要的关键作用<sup>[5]</sup>,表情分类的准确度对表情输入起决定作用。对于表情识别来说,人脸形状、头发、饰品等都是噪声,为了减少此类噪声的影响,需要对直方图均衡化<sup>[6]</sup>处理后的图像进行切割。

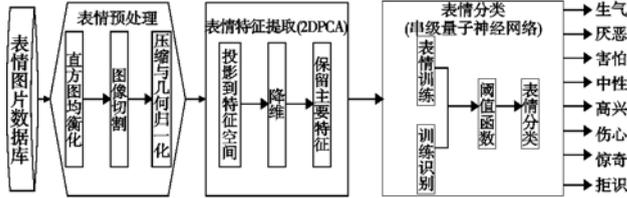


图1 表情识别主要步骤的流程

这里采用切割主要特征块的方法<sup>[7]</sup>,分别切割了嘴、眼睛、眉毛这三大特征区域<sup>[8]</sup>,实现单一感兴趣区向多个同级感兴趣区转换,如图 2 所示保留了特征信息,消除了大量的噪声信息。

切割后的图像大小不等且精度比较高,需将切割后的图像进行几何归一化和压缩<sup>[9]</sup>。计算机处理的图像都是用点阵表示并且是用图像文件的形式存放的。图像特性具有相关性,表现出空间数据的冗余。图像压缩消除特性之间的相关性,如图 3 所示。根据图片清晰度要求,处理后的图像大小为 20 × 28 像素。

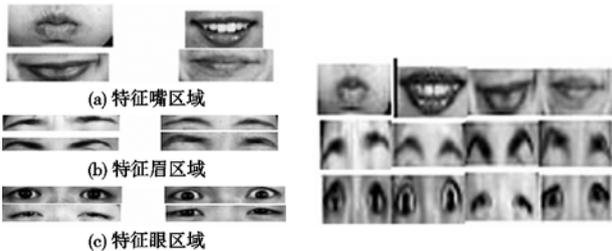


图2 切割后的表情特征块 图3 切割归一化处理后的图像

经过预处理后,已经去除了大量的噪声干扰。本文采用 2DPCA 对预处理后的特征块进行特征提取<sup>[10]</sup>。2DPCA 能克服 PCA 速度慢的特点,建立一个更小维数,只包含最基本元素的空间,其目标是最大化类间分布,同时最小化类内分布<sup>[11]</sup>。对于人脸面部表情识别而言,在 2DPCA 提取得到的特征空间将同一类的人脸表情图像移得更近,而不同类的人脸表情图像移得更远,更有利于接下来的表情分类识别。

2DPCA 算法首先计算训练样本集的总体散布矩阵。根据表情库图片表情类型,把表情分为七大类,分别设为  $e_1, e_2, e_3, e_4, e_5, e_6, e_7$ 。设第  $i$  类训练样本图像有  $n_i$  个,表示为  $A_{i1}, A_{i2}, \dots, A_{in_i}$ 。训练样本总数为:  $M = \sum_{i=1}^7 n_i$ 。每个样本图像表示为一个  $m \times n$  的矩阵。训练样本的均值可以表示为:  $\bar{A} = \frac{1}{M} \sum_{i=1}^7 \sum_{j=1}^{n_i} A_{ij}$ 。每幅图像的均值差为:  $\psi_k = A_{ij} - \bar{A} (i = 1, 2, \dots, 7; j = 1, 2, \dots, n_i)$ ,其中  $k = M$ 。令矩阵  $H = [\psi_1^T, \psi_2^T, \dots, \psi_k^T]^T$ ,则训练样本的总体散布矩阵可用下式进行估计:  $\Sigma = \frac{1}{M} H^T H$ ,其中  $\Sigma$  的维数为  $M \times M$ 。然后计算其特征值和特征向量,选取较大的特征值所对应的特征向量组成的矩阵作为投影空间,再把图像投影到投影空间中提取特征矩阵。根据定理最优投影向量组  $q_1, q_2, \dots, q_d$  是  $\Sigma$  的  $d$  个最大特征值 ( $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$ ) 所对应的标

准正交的特征向量。表情图像往该方向投影使得图像的总体的分散程度最大化,称为最佳投影轴。把最佳方向的向量组构成投影矩阵  $[q_1, q_2, \dots, q_d]$ ,也就是投影特征空间。再把图像矩阵  $A_{ij}$  投影到该空间中,得到图像  $A_{ij}$  的整体投影特征矩阵:

$$\begin{bmatrix} p_{11} & p_{12} & \dots & p_{1d} \\ p_{21} & p_{22} & \dots & p_{2d} \\ \dots & \dots & \dots & \dots \\ p_{m1} & p_{m2} & \dots & p_{md} \end{bmatrix} = A_{ij} [q_1, q_2, \dots, q_d]$$

这个投影矩阵就是所要提取的表情特征矩阵。表情识别过程中,每一个测试样本表情块经过 2DPCA 算法投影后得到一个投影特征矩阵。

### 1.2 串级识别系统

量子神经网络是一种能有效对不确定数据进行分类的多层前馈神经网络<sup>[12]</sup>。它的输入为经过 2DPCA 提取的表情特征向量,输出为表情分类向量,隐层神经元借鉴了量子理论中的量子态叠加的思想,激励函数为多个 Sigmoid 函数的线性叠加构成的多层激励函数,这样一个隐层神经元就能表示更多的状态和量级。叠加的每一个 Sigmoid 函数有不同的量子间隔,可以通过调整量子间隔,使不同类的数据映射到不同的量级或阶梯上,从而使分类有更多的自由度。因此,对复杂的表情进行分类效果比较好<sup>[13]</sup>。量子神经网络的量子间隔可以通过训练获得,只要通过适当的学习算法,采样数据中的不确定性可以被量子神经网络获得并量化。如果样本的特征向量是位于两类模式的边沿重叠处,那么量子神经网络会按一定的比例将其分配给所有相关联的类别,这使网络具有了一种固有的模糊性,能将决策的不确定性数据合理地分配到各模式中,从而减少模式识别的不确定度,提高模式识别的准确性。

对于串级组合来说,每一级就是一个量子神经网络分类器<sup>[14]</sup>。经过 2DPCA 特征提取后的特征数据输入到第一级量子神经网络分类器,得出分类结果。结果包括识别和拒绝两种,识别中包含正确识别和错误识别。被拒绝识别的特征数据输入到下一级量子神经网络分类器再次进行识别。每一级子系统将给出识别结果,拒绝识别的就输入到下一级,整个串级组合的拒绝识别由最后一级的拒绝识别得到。对于这个串级组合,当确定组合的级数为  $M$  级时,假定第  $i$  级量子神经网络分类器的正确识别率为  $C_i$ ,错误识别率为  $W_i$ ,拒绝识别率为  $R_i$ ,其中  $i = 1, 2, \dots, M$ 。正确识别率、错误识别率、拒绝识别率的总和为 1。当各级量子神经网络相互独立时,根据概率论里独立事件同时发生的概率  $P(AB) = P(A) \cdot P(B)$ ,则串级组合的正确识别率为

$$C = C_1 + C_2 \times R_1 + C_3 \times (R_1 \times R_2) + \dots + C_M \times (R_1 \times R_2 \times \dots \times R_{M-1}) \quad (1)$$

错误识别率为

$$W = W_1 + W_2 \times R_1 + W_3 \times (R_1 \times R_2) + \dots + W_M \times (R_1 \times R_2 \times \dots \times R_{M-1}) \quad (2)$$

拒绝识别率为

$$R = R_1 \times R_2 \times R_3 \times \dots \times R_M \quad (3)$$

其中:  $C + W + R = 1$ 。

从式(1)(3)中可以看出,减少错误识别率可以提高整体的正确识别率,而对于串级组合识别来说,有着较大的拒绝识别是可以接受的,不会影响整体的正确识别率。只要串联的级数足够多,并且假设每级的错误识别率  $W_i$  都为零,那么正确识别率可以达到很高。在这种情况下正确识别率为

$$C = 1 - R_1 \times R_2 \times \dots \times R_M$$

$$C = 1 - (1 - C_1) \times (1 - C_2) \times (1 - C_3) \times \dots \times (1 - C_M)$$

因为  $R_i \in (0, 1)$ , 当  $M$  趋向于无穷时,  $\lim_{M \rightarrow \infty} R_1 \times R_2 \times R_3 \times \dots \times R_M = 0$ , 故有  $\lim_{M \rightarrow \infty} C = 1$ 。

由上面可以得出, 当串级组合达到无穷大的级数时, 正确识别率  $C$  会趋向于 100%。但是, 实际情况中错误识别率不可能为 0, 所以在实际做的时候减少错误识别率是关键问题。本文采用自适应设定阈值的方法提高拒绝阈值, 使得错误识别率尽可能地小, 拒绝识别率尽可能地高, 同时不会影响整体的正确识别率<sup>[15]</sup>。

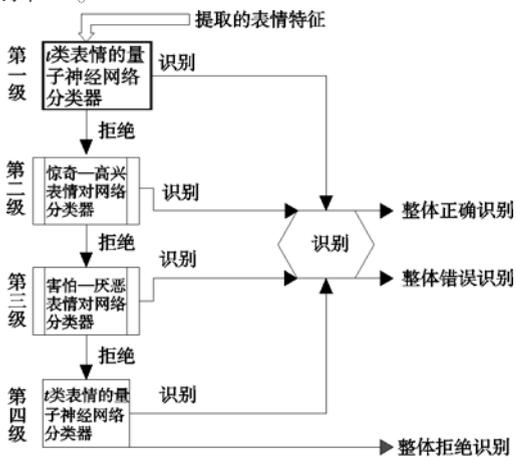


图4 串级识别系统

在串级组合的时候, 最一般的方法是每一级都是一个完整的量子神经网络分类器, 都可以输出识别七类不同的表情。但在实际的测试中发现, 惊奇与高兴、害怕与厌恶是比较容易混淆的。把这两对表情对分别设立一个分类器将会大大提高正确识别率。如图 4 所示, 采用四级串级组合方式: 第一级为七类表情的量子神经网络(表 1 中用 No. 1-7-QNN 表示), 第二级为惊奇—高兴表情对网络分类器(表 1 中用 No. 2-S-HNet 表示), 第三级为害怕—厌恶表情对网络分类器(表 1 中用 No. 3-F-DNet 表示), 第四级为七类表情的量子神经网络(表 1 中用 No. 4-7-QNN 表示)。每一级网络的正确识别率( $C$ )、错误识别率( $W$ )和拒绝识别率( $R$ )如表 1 所示。

表 1 各级网络的识别情况 %

串级组合方式	$C$	$W$	$R$
No. 1-7-QNN	81.43	0	18.57
No. 2-S-HNet	30.77	7.690	61.54
No. 3-F-DNet	33.33	16.67	50.00
No. 4-7-QNN	60.00	20.00	20.00

根据式(1)(3), 由表中的数据可以得到整个串级组合的正确识别率为 94.38%, 错误识别率为 4.48%, 拒绝识别率为 1.14%。在第一级量子神经网络分类器里, 通过自适应算法得到的拒绝识别阈值为 0.91, 能够最大限度地拒绝识别任何可能会产生错误判决的样本, 提高串级系统的可靠性。第二级、第三级阈值为 0.6, 这样就可以把不属于那两类表情的样本拒绝识别, 不至于产生错误分类。在最后一级得到的阈值是 0.15, 为的是能降低拒绝识别, 因为最后一级将没有后续的分类, 因此应该使拒绝识别的可能降到最低。

## 2 情感虚拟人建模

经过 2DPCA 的特征提取和串级识别系统的分类, 已经能够成功识别输入表情。下面则需要通过建立情感模型得到输出值。这里采用 HMM 情感模型<sup>[16]</sup>, 与文献[16]不同的是本

文利用 HMM 建立的模型中输入采用的是图像而不是声音或者语言, 输入转换的激励由“奖励”和“惩罚”两种简单的激励信号变为七种激励信号, 输出信号中观察值也与输入相对应有七种表情, 隐含的心情为“好、中、差”三种。

HMM 是一种用参数表示的、用于描述随机过程统计特性的概率模型<sup>[17]</sup>。它是一个双重随机过程, 其中之一是 Markov 链, 这是基本随机过程, 它描述状态的转移。另一个是随机过程描述状态和观察值, 不像 Markov 链模型中的观察值和状态一一对应, 因此, 不能直接看到状态, 而是通过一个随机过程去感知状态的存在及其特性。这与情感的特性非常相似, 其中的观察值对应可见的表情, 隐含的状态对应人类的心情。

模型定义: 一个基于 HMM 的情感模型可以由下列参数描述。

a)  $N$ : 模型中 Markov 链状态数目, 即心情的种类数。记  $N$  个状态为  $\theta_1, \theta_2, \dots, \theta_N$ , 记  $t$  时刻所处的心情状态为  $q_t$ , 则  $q_t \in (\theta_1, \theta_2, \dots, \theta_N)$ 。

b)  $M$ : 每个状态对应的表情种类数。记  $M$  个观察值为  $V_1, V_2, \dots, V_M$ , 记  $t$  时刻观察到的表情状态为  $O_t$ , 则  $O_t \in (V_1, \dots, V_M)$ 。

c)  $\pi$ : 初始状态所处的心情概率矢量。 $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ , 其中初始心情属于某种心情的概率  $\pi_i = P(q_1 = \theta_i), 1 \leq i \leq N$ 。

d)  $A$ : 心情状态转移概率矩阵。 $A = (a_{ij})_{N \times N}$ , 其中  $a_{ij}$  表示  $t$  时刻所处的心情状态转移到  $t+1$  时刻所处的心情状态的概率:  $a_{ij} = P(q_{t+1} = \theta_j | q_t = \theta_i), 1 \leq i, j \leq N$ 。

e)  $B$ : 表情的条件概率矩阵。 $B = (b_{jk})_{N \times M}$ , 其中  $b_{jk}$  表示在  $\theta_j$  心情状态的情况下出现的表情概率,  $b_{jk} = P(O_t = V_k | q_t = \theta_j), 1 \leq j \leq N, 1 \leq k \leq M$ 。这样, 可以记一个 HMM 为  $\lambda = (N, M, \pi, A, B)$ , 或简写为  $\lambda = (\pi, A, B)$ 。

基于上述 HMM 的情感模型, 可以根据个人的喜好或者应用的要求, 创建具有不同性格特征的情感交互系统。这里以乐观向上的性格为例建立情感模型。下面着重分析如何确定每个参数的取值。首先, 需要确定情感的每个层次上的分类数, 也就是说, 要确定有多少种心情——对应隐马尔可夫模型中的状态数  $N$ , 如表 2 所示, 心情分为中、好、差三种状态, 分别编号, 便于下面的计算。

表 2 心情分类编号

心情状态编号	心情状态描述
1	中
2	好
3	差

接着要确定多少种表情——对应隐马尔可夫模型中的观察值的个数  $M$ , 根据输入的七种表情激励, 如表 3 所示, 定义了对应的七种表情输出作为观察值序列。

对于一个乐观向上的人来说, 他大多数时间是处于好的心态下的, 在之前未受过任何激励的时候, 他应该处于一种好的心态。由此分析, 本文所设计的虚拟人对象起初的概率矢量  $\pi$  应为:  $\pi = (\pi_1, \pi_2, \pi_3) = (0, 1, 0)$ 。当虚拟人接受到外界表情刺激时, 它根据该信号作用下的情感模型会得到一个情感输出结果, 包括此时的心情和表情。把前面分类得到的七种表情作为七种意识刺激。这样, 针对每种刺激分别有一个状态转移概率矩阵及可见符号概率矩阵与之对应。实际上, 具有不同个性特征的人, 可见符号概率矩阵, 以及在同样激励下的观察值序列是不相同的, 在同样的激励下的状态转移矩阵也是不同的。

矩阵中各行元素值随着情绪的变化而变化,情绪越稳定,各行元素分布得越平均。另外,就某一个个体而言,其情感模型中的状态转移概率矩阵的值与意识刺激的类型有关,在不同的意识刺激下,状态转移概率矩阵取值不同;它还与个体所接受刺激的历史状况及自身内部的需求机制的作用等因素有关。用于模型训练的观察序列,应该根据所塑造的角色性格特征来给出,它是在某一种特定情感刺激下的情感表层表现——表情序列。每一时刻,心情都会处在不同的状态,并且在下一时刻会向不同的状态转移,三种心情状态的 Markov 链如图 5 所示。

表 3 表情分类编号

表情编号	表情描述
000	生气
001	厌恶
010	害怕
011	中性
100	高兴
101	伤心
110	惊奇

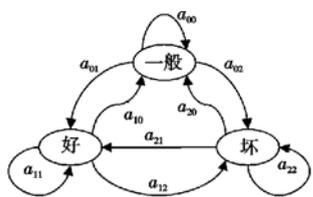


图5 心情变化的三状态Markov链

设在接收“生气”信号时,心情的状态转移概率矩阵和表情的条件概率矩阵分别为  $A_0[3][3]$ 、 $B_0[3][7]$ ;在接收到“厌恶”信号时,心情的状态转移概率矩阵和表情的条件概率矩阵分别为  $A_1[3][3]$ 、 $B_1[3][7]$ ;依此类推,接收到其他五种表情激励信号时也对应着两个矩阵。这里,心情的状态转移概率矩阵指的是由初始心情向不同心情状态转移的概率,表情的条件概率矩阵指的是在某种心情条件下所表现出来的表情的概率。这样,就可以由以上的初始数据,对情感模型进行训练。由于要解决的问题是确定一个角色的心理模型参数,再由该参数模型产生情感表现的观察序列,即怎样调整模型参数  $\lambda = (\pi, A, B)$ ,使  $P(O/\lambda)$  最大的问题。一般采用 Baum-Welch 算法进行的参数估计,由于 Baum-Welch 算法在参数训练过程中容易出现下溢现象,导致得不到最优解<sup>[18]</sup>。所以本文采用一种改进的遗传算法进行模型<sup>[19]</sup>参数的估计。对于一个虚拟人来说,它可能受到外界不同的刺激,基于前面对激励信号的阐述,定义激励信号为  $t$ ,  $t$  的取值与表 3 中输出信号一致。对于同一个虚拟人,可以给他连续不同的刺激,这样可以得到不同的激励信号序列。根据经验人为定义实际输出表情的概率矩阵为  $C$ ,当输入激励信号序列不同时,  $C$  值也不同。

用上述算法对虚拟人的情感模型进行参数估计。随着人机交互的不断进行,虚拟人所接受的激励情况也随之变化,从而,在某些情况下情感模型的初始概率矢量会有所变化。比如:上一个激励是一个“生气”信号,那么,根据该信号作用下的情感模型会得到一个情感输出结果,包括此时的心情和表情,并且程序通过计算同样得到了出现其他可能出现的心态及表情的概率;如果下一次是一个“惊奇”信号作为情感激励,那么此时的初始概率矢量应该由已经计算出的可能出现的各种心情的概率决定,并且选用“生气”信号下的模型初值  $A_0[3][3]$ 、 $B_0[3][7]$ ;而如果接下来仍旧是“生气”信号,则不需要重新估计模型参数,而是按照上一刻计算得到的模型来产生下一个表情概率。根据输入激励信号序列,可以得到  $E$  个表情概率矩阵,其中  $E = \sum_{i=1}^n 2^i$ ,  $n$  为激励信号序列中包含的刺激信号的个数。为了训练模型,文中取  $n = 7$ ,此时,  $E = 2^1 + 2^2 + 2^3 + 2^4 + 2^5 + 2^6 + 2^7 = 254$ ,即有 254 个表情概率矩阵。

用改进的遗传算法进行模型参数的估计,这里定义误差函数分别为  $ERROR_1$  均方误差,  $ERROR_2$  绝对误差。其中,

$$ERROR_1 = \|C_0 - C\|^2 + \|C_1 - C\|^2 + \dots + \|C_{253} - C\|^2 \quad (4)$$

$$ERROR_2 = |C_0 - C| + |C_1 - C| + \dots + |C_{253} - C| \quad (5)$$

分别把误差函数代入到适应度函数  $f = C_{\max} - ERROR$  中, MATLAB 仿真结果如图 6 所示。图中适应度函数的曲线变化是越来越高,说明由该算法估计的模型参数计算出来的表情输出概率是越来越接近于经验定义的表情输出概率。图 6 中 33 为零误差曲线。从图 6 中可以看出,两条误差曲线均不超过零误差曲线,且两者的趋势都是越来越接近于零误差曲线,说明误差是越来越小的。因为本文选择了 11 组训练数据,每一组有三个元素,每个元素都在  $[0, 1]$  之间,所以误差总和不超过 33。又因为对于平方和误差是小于绝对值误差的,所以均方误差的起点是高于绝对误差的。

### 3 情感虚拟人表情输出

由于利用隐马尔可夫模型作为情感引擎的输出结果为各心情和表情可能出现的概率,而要实现和谐的人机交互,输出结果应该以表情画面的形式体现出来。因此,结合 VC++ 与 OpenGL 技术,实现了情感交互虚拟人系统平台。OpenGL 是一个功能强大的 3D 图形库<sup>[20]</sup>,它包括将近 120 个绘制点、线和多边形等 3D 基本图形的绘制命令,并且可在这些基础上构造更复杂的 3D 物体。此外,OpenGL 还支持雾化、纹理映射、反走样、光照以及动画等效果。所以,本文选择在 VC++ 环境下使用 OpenGL 并采用 NURBS 曲面和三角形面片相结合的方式<sup>[21]</sup>进行人脸表情动画建模。如表 4 所示,为了表达面部表情根据心情的变化,本模型参考了 Ekman 的 FACS 系统中面部表情的描述。

表 4 面部表达系统设计

输出表情	特性描述
生气	外侧眼皮翘起,眼皮下垂,鼻子皱起,上嘴唇抬起
厌恶	外侧眼皮抬起,眼皮下垂,鼻子皱起
害怕	内侧眼皮抬起,眼皮下压,嘴角紧绷
中性	无
高兴	内侧眼皮上翘,脸颊抬起,嘴角紧绷
伤心	内侧眼皮抬起,眉毛下压,嘴角紧绷
惊奇	眼皮上扬,上嘴唇抬起,嘴巴打开

对于部分静态的脸部模型,采用 NURBS 曲面进行构造。对于脸部表情变化的部分,根据人脸的生理特点,将人脸表面分成若干块。将人脸表情特征部位,如眼睛、嘴巴等处的 NURBS 曲面用面片构成的曲面代替。当特征面片块体位移和形变都很小时,可认为块体内部个体位移是其坐标  $(x, y, z)$  的一阶线性组合(高阶忽略),即

$$\begin{cases} u = a_0 + a_1x + a_2y + a_3z \\ v = b_0 + b_1x + b_2y + b_3z \\ w = c_0 + c_1x + c_2y + c_3z \end{cases} \quad (6)$$

人脸出现某种特定表情时,每一个子块的形变比较小,因此,完全可以采用弹性形变模型来模拟脸部表情动画。以嘴巴为例,嘴巴外形近似于一个椭圆,椭圆和抛物线在空间曲线的统一表示中其形状因子最相近,所以用两段抛物线模拟一个嘴巴的外形。其外形轮廓近似符合下面公式:

$$y = a(x - b)^2 + c \quad (7)$$

对上嘴唇:设嘴唇闭合是嘴唇的一个弹性平衡态,且此时下嘴唇无弹性形变,则当上嘴唇的抛物线与纵轴焦点值为  $c$ ,水平方向上的拉动物位移  $\Delta x = 0$ ,垂直方向上的位移  $\Delta y$  (抛物线拟合上嘴唇线)时,所有上嘴唇三角形定点的位移可表示为(以四个特征点的坐标系原点建立坐标系):

$$\Delta y = a(x_2^2 - x_1^2 - 2b(x_2 - x_1)) \Delta x = \sqrt{(y_2 - c)/a} - \sqrt{(y_1 - c)/a} \quad (8)$$

因此,定点的坐标变为 $(x', y', z')$ ,且满足

$$(x', y', z') = (x, y, z) + (\Delta x, \Delta y, \Delta z) \quad (9)$$

结合式(7)可以得到顶点的新坐标。变动  $a$  与  $c$  值,就可以得到上嘴唇运动的一系列新坐标,从而模拟出上嘴唇的整个运动过程。同理可得下嘴唇运动过程。采用同样方法,还建立了眼睛、眉毛等许多动作的弹性形变模型。由于保存的是弹性形变模型的数学公式而非动作功能网格模型,所以在空间耗费上同以往的方法相比要经济多了;每个表情动作,只有少数区域的点参与形变,形变过程完全由弹性形变模型确定,因而计算量较小。弹性模型建立好后,再利用关键帧技术<sup>[22]</sup>,通过系统在关键帧之间进行样条插值产生中间帧,进而产生动画帧序列。然后再加上纹理,纹理是使用纹理映射,人脸表面的纹理具有颜色和几何纹理的双重属性。在贴纹理前,必须要确定纹理与人脸的相对关系,即在场景中生成实体时要同时用到纹理坐标和几何坐标。对于一个二维纹理来说,纹理坐标在两个方向上都是从 0.0 到 1.0,所以要注意使其左右、上下各边能够很好地拼为一体,即要实现纹理图像的无缝连接。经这样的处理后就可以实现具有真实感较强的人脸表面表情。通过 OpenGL 生成表情如图 7 所示,从左到右依次是高兴、生气、愤怒、厌恶的表情。

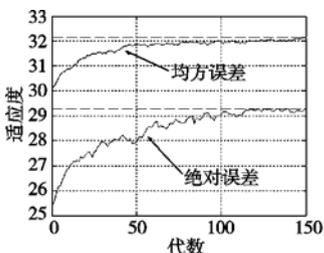


图6 适应度函数曲线

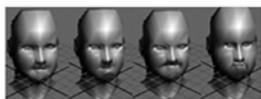


图7 生成的面部表情

#### 4 结束语

本文综合使用了多级量子神经网络分类器实现串级识别,将 2DPCA 提取后的表情特征输入到四级分类器中,中间两级采用了表情对的分类器,有效地解决了易混淆表情被误判的问题。在输入表情分类好后,建立了基于隐马尔可夫的情感模型,实现了一个心理模型实时驱动表情动画的情感虚拟人系统。最后利用 VC++ 和 OpenGL 实现了人机交互界面的设计。本系统的开发,为虚拟人生成应用、情感计算、情感机器人和友好人机界面等领域提供了一个基础平台。系统登录界面如图 8 所示。点击“选择表情”后弹出表情选择对话框,然后依次点击“特征提取”“串级分类”。在串级分类成功后,在输入激励识别结果中显示结果。

如图 9 所示,由于此系统所建立的虚拟人是乐观向上的,虽然图中输入表情识别结果是中性,识别结果为 0001000,但是当点击“心理计算”按钮后,可以看到此时的心情仍是较好,输出表情仍为高兴。表情强度略减轻,从初始的 1 变为 0.805。

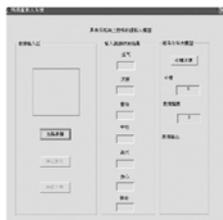


图8 情感交互系统的登录界面



图9 情感交互系统演示

在实际应用中还可以对本系统进行改进,首先部分输入可以连接摄像头采集图像,同时加入人脸识别的一些算法,输出部分的三维头像可以尝试改为动画,使得系统更加贴近实际生活。

#### 参考文献:

- [1] 王国江,王志良,陈锋军,等. 基于 Markov 决策过程的交互虚拟人情感计算模型[J]. 计算机科学, 2006, 33(12): 135-138.
- [2] 谷学静,王志良,魏哲华,等. 基于人工心理理论的情感模型构建方法研究[J]. 微计算机信息, 2006, 22(2): 264-266.
- [3] JIANG Q. Medical psychology, 3rd version [M]. Beijing: Chinese People's Medical Publishing House, 2002.
- [4] KIM K H, BANG S W, KIM S R. Emotion recognition system using short-term monitoring of physiological signals [J]. Med Biol Eng Comput, 2004, 42(3): 419-427.
- [5] 杨静宇,金忠,郭跃飞. 人脸图像有效鉴别特征提取与识别[J]. 南京理工大学学报, 2000, 24(3): 193-198.
- [6] 尹立敏,刘艳滢,贺琳,等. 基于多尺度灰度直方图的图像增强研究[J]. 电子器件, 2006, 29(2): 431-433.
- [7] 李俊华,彭力. 基于特征块主成分分析的人脸表情识别[J]. 计算机工程与设计, 2008, 29(12): 3151-3153.
- [8] 郑芳颖,赵杰煜. 基于改进型嵌入式隐马尔可夫模型的表情识别方法[J]. 模式识别与人工智能, 2008, 12(6): 836-842.
- [9] 黄永明,章国宝,董飞,等. 基于 Gabor, Fisher 脸多特征提取及集成 SVM 的人脸表情识别[J]. 计算机应用研究, 2011, 28(4): 1536-1539.
- [10] LI Jun-hua, PENG Li. Human expression recognition based on feature block 2DPCA and Manhattan distance classifier [C]//Proc of the 7th World Congress on Intelligent Control and Automation. 2008: 5941-5945.
- [11] NHAT V D M, LEE S Y. Two-dimensional weighted PCA algorithm for face recognition [C]//Proc of IEEE International Symposium on Computational Intelligence in Robotics and Automation. 2005: 219-223.
- [12] 盖怀存,张小锋,江泽涛. 基于量子神经网络的人脸识别技术研究[J]. 计算机工程与应用, 2010, 46(8): 187-190.
- [13] 李俊华,彭力. 量子神经网络在表情识别中的应用 [C]//中国控制与决策学术年会论文集. 沈阳: 东北大学出版社, 2007: 328-330.
- [14] 吴丽芸,王文伟,张平,等. 手写混合字符识别的多特征多级分类器设计[J]. 计算机应用, 2005, 25(12): 2948-2950.
- [15] KARAYIANNIS N B, XIONG Yao-hua. Training reformulated radial basis function neural networks capable of identifying uncertainty in data classification [J]. IEEE Trans on Neural Networks, 2006, 17(5): 1222-1234.
- [16] 谷学静,王志良,刘冀伟,等. 基于 HMM 的人工心理建模方法研究[J]. 计算机应用研究, 2006, 23(12): 30-32.
- [17] RABINER L R. A tutorial on hidden Markov models and selected applications in speech recognition [M]. San Francisco: Morgan Kaufmann Publishers, 1990.
- [18] 贾宾,朱小燕,罗子频,等. 消除溢出问题的精确 Baum-Welch 算法 [J]. 软件学报, 2000, 11(5): 707-710.
- [19] 朱飒飒,彭力,王巍. 人工情感进化模型及其学习算法的研究 [J]. 计算机工程与应用, 2008, 44(32): 45-47.
- [20] 乔林,费广. OpenGL 程序设计 [M]. 北京: 清华大学出版社, 2000: 59-150.
- [21] 谷学静,许广宾,熊伟,等. 人工心理模型驱动的人脸表情动画合成 [J]. 计算机工程与应用, 2008, 44(8): 107-109.
- [22] 李连东,樊养余,雷涛,等. 一种可编辑的三维人体蒙皮网络动画合成方法 [J]. 计算机应用研究, 2010, 28(3): 1176-1179.