

11-20-2019

Facial Expression Synthesis Using MALS-based Bilinear Factorization Model

Jiaci Guo

1. State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China; ;2. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China;

Shuling Dai

1. State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China; ;2. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research, Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Facial Expression Synthesis Using MALS-based Bilinear Factorization Model

Abstract

Abstract: Study on the facial expression synthesis is a constructive and creative subject in the field of virtual reality technology. The bilinear model is used for facial expression synthesis by separating the identity factor and the expression factor. The translation procedure in the bilinear model always requires a repetitive computation of matrix inverse operations to reach the identity factor and the expression factor. This computation may be instable when the observation data has the correlation or noisy information. *In order to increase the stability of the bilinear model for expression synthesis, the modified iterative least square (MALS) regression is introduced into the translation procedure to stabilize the computation.* Experimental results demonstrate the effectiveness of the method.

Keywords

facial expression synthesis, bilinear model, translation, MALS

Recommended Citation

Guo Jiacy, Dai Shuling. Facial Expression Synthesis Using MALS-based Bilinear Factorization Model[J]. Journal of System Simulation, 2019, 31(4): 671-678.

基于 MALS 双线性分解的表情合成

郭加赐^{1,2}, 戴树岭^{1,2}

(1. 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;

2. 北京航空航天大学自动化科学与电气工程学院, 北京 100191)

摘要: 表情合成研究是虚拟现实技术领域一项富有建设性和创造性的课题。使用双线性模型(Bilinear Model)分离出人的身份因素和表情因素来进行表情合成。在计算双线性模型的转移(Translation)问题时需要重复进行矩阵的逆运算来得到人的身份和表情信息, 而当观测数据具有相关性或存在噪声信息时, 该过程可能会不稳定。针对双线性模型在表情合成中的稳定性问题, 通过在双线性模型转移过程引入改进迭代最小二乘法(Modified Iterative Least Square, MALS), 使得该过程的计算具有更强的稳定性。设计实验进行验证, 实验结果证明了方法的有效性。

关键词: 表情合成; 双线性模型; 转移; 改进迭代最小二乘

中图分类号: TP391.9

文献标识码: A

文章编号: 1004-731X (2019) 04-0671-08

DOI: 10.16182/j.issn1004731x.joss.17-0144

Facial Expression Synthesis Using MALS-based Bilinear Factorization Model

Guo Jiaci^{1,2}, Dai Shuling^{1,2}

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

2. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China)

Abstract: Study on the facial expression synthesis is a constructive and creative subject in the field of virtual reality technology. The bilinear model is used for facial expression synthesis by separating the identity factor and the expression factor. The translation procedure in the bilinear model always requires a repetitive computation of matrix inverse operations to reach the identity factor and the expression factor. This computation may be instable when the observation data has the correlation or noisy information. *In order to increase the stability of the bilinear model for expression synthesis, the modified iterative least square (MALS) regression is introduced into the translation procedure to stabilize the computation.* Experimental results demonstrate the effectiveness of the method.

Keywords: facial expression synthesis; bilinear model; translation; MALS

引言

随着计算机技术的发展和虚拟现实设备的普及, 近几年国内外形成了对虚拟现实技术的又一

研究热潮, 其中表情合成技术的研究被许多学者所关注。表情合成就是将现实人脸表情在计算机中进行模拟再现, 是进行虚拟现实、人工智能、人机交互技术研究的一大方向。表情合成技术可以实现人机交互过程中的情感交流, 为人们提供更加友好的人机交互方式, 对于实现机器的智能化有着非常重要的意义。表情合成技术还有着较为广泛的应用前景, 在军事训练仿真、交互式游



收稿日期: 2017-03-24

修回日期: 2017-05-16;

作者简介: 郭加赐(1993-), 男, 山东郯城, 硕士, 研究方向为虚拟现实与飞行仿真; 戴树岭(1966-), 男, 回族, 河北, 博士, 教授, 研究方向为虚拟现实与飞行仿真。

<http://www.china-simulation.com>

• 671 •

戏、视频会议、辅助教学和科幻电影等众多领域都能找到表情合成的身影^[1]。

但是由于人脸表情的复杂性，对人脸表情进行建模一直没有形成比较完善的理论；同时，人的身份等因素也会对人脸表情的合成产生一定影响。因此，表情合成的研究仍然是虚拟现实技术领域一项富有建设性和创造性的课题。

X Wan 等利用拉普拉斯坐标的局部细节保留属性，使用拉普拉斯变换实现将表演者的面部表情克隆到 3D 面部模型上^[2]；Y Zhang 等使用一种新的弹性模型来平衡全局和局部变形效果，再结合肌肉分布模型实现了较好的表情克隆效果^[3]；H Liang 等使用耦合字典学习的方法，实现了输入一张二维人脸图片，合成不同表情的三维人脸模型^[4]。但上述方法没有对不同人脸的表情特征进行共性分析，难以消除身份因素对人脸表情合成的影响。Jiang-Long Chang 等使用双线性分解模型将人脸的表情信息和身份信息分离，从而实现三维人脸表情合成^[5]。双线性模型由 Tenenbaum 和 Freeman 提出^[6]，可以将观测值有效分离为两个独立的影响因素，在动作识别、人脸识别、以及语音转换^[7-9]等多个领域都有应用。在表情合成领域要用到双线性模型的转移(Translation)功能，该过程需要重复进行矩阵的逆运算，而当观测数据具有相关性或存在噪声信息时，该过程可能会不稳定。为了解决这个问题，Dongsoo Shin 等提出了岭回归双线性模型(Ridge Regressive Bilinear Model)，增强了“人脸”和“光照”因素的分离效果，提高了人脸识别率^[10]，Zhicai Ou 等将岭回归双线性模型应用到工件的姿态估计中^[11]。但是岭回归引入了一个偏差，当惩罚项权重增大时，这个偏差的影响也在增大，从而对双线性模型的应用效果产生影响。针对这一问题，本文将改进迭代最小二乘法(Modified Iterative Least Square, MALS)^[12]引入到双线性模型中，从而实现更好的表情合成效果。

本文第一节对双线性模型表情合成系统进行

概述，第二节介绍使用双线性模型进行表情合成的具体实现，第三节介绍如何将 MALS 引入到双线性模型中，第四节是实验设计方法及实验结果分析，第五节是结论。

1 系统概述

本文使用双线性模型进行表情合成，实现输入任意一张人脸，提取其特征向量，合成该人脸在目标表情下的特征向量，其总体框架如图 1 所示。

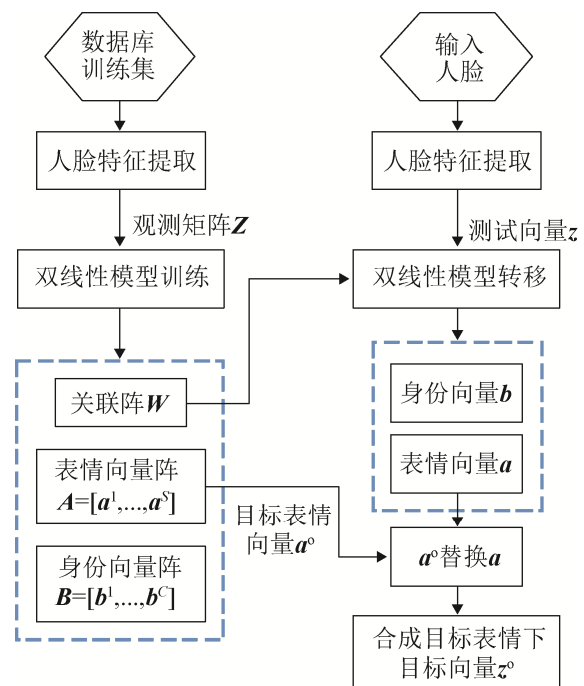


图 1 双线性模型表情合成总体框架图

Fig. 1 Framework of bilinear model for expression synthesis

在双线性模型转移模块需要重复进行矩阵的逆运算，而当观测数据具有相关性或存在噪声信息时，该过程可能会不稳定。针对这一问题，本文将改进迭代最小二乘(MALS)引入到双线性模型中。下面将详细介绍各个模块的具体实现以及如何将 MALS 引入到双线性模型中。

2 双线性模型表情合成

2.1 人脸特征提取

首先使用改进的 ASM(Active Shape Model)算

法^[13]对数据库人脸图像进行特征点定位, 获得不同人脸不同表情下的人脸特征点坐标, 如图 2 所示 (图片来自 KDEF 人脸表情数据库^[14])。

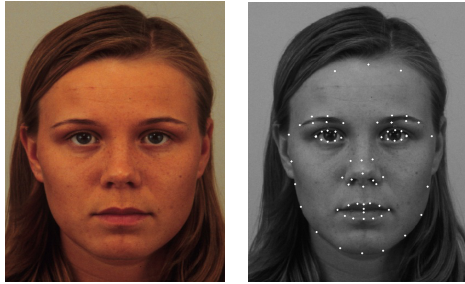


图 2 原始人脸图像及标定特征点图像
Fig. 2 Original facial image and the image with landmarks

以鼻尖为坐标原点, 将从人脸图像中提取的 N 个特征点坐标进行平移, 然后排列成列向量, 得到 $2N$ 维观测向量 $\mathbf{z} = (x_1, y_1, x_2, y_2, \dots, x_N, y_N)^T$, 作为人脸特征。

从人体的皮肤运动机理上说, 人脸有嘴巴、眼睛等相对独立的组织, 表皮的结构具有一定的差异, 因此人脸运动具备一定的区域性。对人脸进行分区, 这样对人脸表情进行统计分析时就容易发现人脸的局部细节特征, 而且还能更加有效地利用人脸样本信息。本文将人脸分为 4 个区域进行处理, 如图 3 所示。



图 3 人脸分区示意图
Fig. 3 Diagram of the facial partitions

则区域 $k(k=1,2,3,4)$ 的人脸特征用 \mathbf{z}_k 表示, \mathbf{z}_k 为该区域内的 M 个特征点坐标排列成的观测向量, 表示为 $\mathbf{z}_k = (x_1, y_1, x_2, y_2, \dots, x_M, y_M)^T$ 。

2.2 双线性模型训练

双线性模型由 Tenenbaum 和 Freeman 提出,

可以将观测值有效分离为两个独立的影响因素^[6]。在双线性模型中, 两个相互影响的因素被称为“形式”和“内容”, 它们以乘积的方式共同作用于观测集, 对称性双线性模型表示为^[6]:

$$\mathbf{z} = \sum_{i=1}^I \sum_{j=1}^J \mathbf{w}_{ij} \mathbf{a}_i \mathbf{b}_j \quad (1)$$

式中: “形式”因素用 I 维向量 \mathbf{a} 表示; “内容”因素用 J 维向量 \mathbf{b} 表示; 共同作用于 K 维观测向量 \mathbf{z} , \mathbf{w} 表示连接“形式”和“内容”的关联项。

我们使用奇异值分解(SVD)来进行对称性双线性模型的训练。假设我们具有 $S \times C$ 个训练样本, 按照“形式”和“内容”的行列顺序将 K 维观测向量排列成观测矩阵 \mathbf{Z} , 则观测矩阵 \mathbf{Z} 的维数为 $SK \times C$ 。引入向量转置的概念, 记为 \mathbf{V}^T , 向量转置就是将矩阵里的 K 维观测向量看作一个整体, 再对矩阵进行转置, 则 \mathbf{Z} 和 $\mathbf{Z}^{\mathbf{V}^T}$ 见公式(2)^[6]:

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}^{11} & \dots & \mathbf{z}^{1C} \\ \dots & \ddots & \dots \\ \mathbf{z}^{S1} & \dots & \mathbf{z}^{SC} \end{bmatrix}, \quad \mathbf{Z}^{\mathbf{V}^T} = \begin{bmatrix} \mathbf{z}^{11} & \dots & \mathbf{z}^{S1} \\ \dots & \ddots & \dots \\ \mathbf{z}^{1C} & \dots & \mathbf{z}^{SC} \end{bmatrix} \quad (2)$$

对称性双线性模型也可以表示为^[6]:

$$\mathbf{Z} = (\mathbf{W}^{\mathbf{V}^T} \mathbf{A})^{\mathbf{V}^T} \mathbf{B} \text{ 或 } \mathbf{Z}^{\mathbf{V}^T} = (\mathbf{W} \mathbf{B})^{\mathbf{V}^T} \mathbf{A} \quad (3)$$

式中: \mathbf{A} 由形式向量组成; \mathbf{B} 由内容向量组成; \mathbf{W} 由关联项组成^[6]:

$$\mathbf{A} = [\mathbf{a}^1, \dots, \mathbf{a}^S], \mathbf{B} = [\mathbf{b}^1, \dots, \mathbf{b}^C], \quad \mathbf{W} = \begin{bmatrix} \mathbf{w}^{11} & \dots & \mathbf{w}^{1J} \\ \dots & \ddots & \dots \\ \mathbf{w}^{I1} & \dots & \mathbf{w}^{IJ} \end{bmatrix} \quad (4)$$

则对称性双线性模型的训练算法如下:

- (i) 首先对矩阵 \mathbf{Z} 进行奇异值分解 $\mathbf{Z} = \mathbf{U} \mathbf{S} \mathbf{V}^T$, 取 \mathbf{V}^T 的前 J 行作为 \mathbf{B} 。
- (ii) 由于 \mathbf{B} 正交, 对公式(3)进行转换的到 $(\mathbf{Z} \mathbf{B}^T)^{\mathbf{V}^T} = \mathbf{W}^{\mathbf{V}^T} \mathbf{A}$, 对 $(\mathbf{Z} \mathbf{B}^T)^{\mathbf{V}^T}$ 进行奇异值分解 $(\mathbf{Z} \mathbf{B}^T)^{\mathbf{V}^T} = \mathbf{U} \mathbf{S} \mathbf{V}^T$, 取 \mathbf{V}^T 的前 I 行作为 \mathbf{A} 。
- (iii) 由于 \mathbf{A} 正交, 再对公式(3)进行转换的到

$(\mathbf{Z}^{\text{VT}}\mathbf{A}^{\text{T}})^{\text{VT}}=\mathbf{WB}$, 对 $(\mathbf{Z}^{\text{VT}}\mathbf{A}^{\text{T}})^{\text{VT}}$ 进行奇异值分解 $(\mathbf{Z}^{\text{VT}}\mathbf{A}^{\text{T}})^{\text{VT}}=\mathbf{USV}^{\text{T}}$, 取 \mathbf{V}^{T} 的前 J 行作为 \mathbf{B} 。

(iv) 重复(ii)、(iii)步直至收敛, 得到 \mathbf{A} 和 \mathbf{B} 。

(v) 计算 $\mathbf{W}=(\mathbf{Z}\mathbf{B}^{\text{T}})^{\text{VT}}\mathbf{A}^{\text{T}})^{\text{VT}}$ 。

2.3 基于双线性模型转移功能的表情合成

使用双线性模型进行表情合成时将“表情”因素作为“形式”, “身份”因素作为“内容”, 在进行表情合成时要用到双线性模型的“转移”功能。

双线性模型的转移问题数学描述为: 输入一个未知的测试向量 \mathbf{z} , 使训练好的双线性模型同时适应于新形式 \mathbf{a} 与新内容 \mathbf{b} , 同时保证从训练集得到的关联项矩阵 \mathbf{W} 保持不变。实际上就是以 \mathbf{a} 和 \mathbf{b} 为自变量的代价函数^[6]:

$$f(\mathbf{a}, \mathbf{b}) = \|\mathbf{z} - (\mathbf{W}\mathbf{b})^{\text{VT}}\mathbf{a}\|^2 \quad (5)$$

的最小化问题。对(5)式进行求导, 得到优化解, 如下面两个方程所示^[6]:

$$\mathbf{a} = ((\mathbf{W}\mathbf{b})^{\text{VT}})^{-1}\mathbf{z} \quad (6)$$

$$\mathbf{b} = ((\mathbf{W}^{\text{VT}}\mathbf{a})^{\text{VT}})^{-1}\mathbf{z} \quad (7)$$

$(\bullet)^{-1}$ 表示广义逆, 交替使用公式(6)和公式(7)来迭代计算 \mathbf{a} 和 \mathbf{b} , 直至 \mathbf{a} 、 \mathbf{b} 收敛。

则使用双线性模型进行表情合成的具体实现如下: 提取数据库训练集人脸图片的特征点坐标组成观测向量, 按照“表情”和“身份”的行列顺序将观测向量排列成观测矩阵 \mathbf{Z} , 进行双线性模型的训练; 对于任意输入人脸, 提取特征点坐标组成测试向量 \mathbf{z} , 使用双线性模型转移功能得到其表情向量 \mathbf{a} 和身份向量 \mathbf{b} , 将训练得到的表情向量 $[\mathbf{a}^1, \dots, \mathbf{a}^S]$ 进行线性组合得到目标表情向量 \mathbf{a}^0 , 使用目标表情向量 \mathbf{a}^0 替换其原表情向量 \mathbf{a} , 合成输入人脸在目标表情下的目标特征向量 \mathbf{z}^0 。由 2.1 节可知人脸分为了 4 个区域, 对每一个人脸分区分别进行双线性模型的“训练”和“转移”, 进行表情合成, 然后再把各个人脸区域组合在一起。每一个人脸分区的观测向量为各个区域的特征点坐标组成的数据向量 $\mathbf{z}_k(k=1,2,3,4)$ 。

当观测数据具有相关性或存在噪声时, 计算双

线性模型转移问题可能会不稳定, 为此将改进迭代最小二乘(MALS)^[12]引入到双线性模型中。

3 MALS 双线性模型

当观测数据具有相关性或含有噪声时, 交替使用(6)、(7)式计算双线性模型的“转移”问题可能会不稳定, 为了克服这个问题来稳定计算, 可以引入岭回归算法, 代价函数(5)变为^[11]:

$$f(\mathbf{a}, \mathbf{b}) = \sum_{k=1}^K (\mathbf{z}_k - \sum_{i=1}^I \sum_{j=1}^J w_{ijk} \mathbf{a}_i \mathbf{b}_j)^2 + \lambda_1 \sum_{i=1}^I \mathbf{a}_i^2 + \lambda_2 \sum_{j=1}^J \mathbf{b}_j^2 \quad (8)$$

式(8)也可以表示为^[11]:

$$f(\mathbf{a}, \mathbf{b}) = (\mathbf{z} - (\mathbf{W}\mathbf{b})^{\text{VT}}\mathbf{a})^{\text{T}}(\mathbf{z} - (\mathbf{W}\mathbf{b})^{\text{VT}}\mathbf{a}) + \lambda_1 (\mathbf{a}^{\text{T}}\mathbf{a}) + \lambda_2 (\mathbf{b}^{\text{T}}\mathbf{b}) = (\mathbf{z} - (\mathbf{W}^{\text{VT}}\mathbf{a})^{\text{VT}}\mathbf{b})^{\text{T}}(\mathbf{z} - (\mathbf{W}^{\text{VT}}\mathbf{a})^{\text{VT}}\mathbf{b}) + \lambda_1 (\mathbf{a}^{\text{T}}\mathbf{a}) + \lambda_2 (\mathbf{b}^{\text{T}}\mathbf{b}) \quad (9)$$

再对式(9)求导, 得到优化解^[11]:

$$\mathbf{a} = (((\mathbf{W}\mathbf{b})^{\text{VT}})^{\text{T}}(\mathbf{W}\mathbf{b})^{\text{VT}} + \lambda_1 \mathbf{I})^{-1}((\mathbf{W}\mathbf{b})^{\text{VT}})^{\text{T}}\mathbf{z} \quad (10)$$

$\mathbf{b} =$

$$(((\mathbf{W}^{\text{VT}}\mathbf{a})^{\text{VT}})^{\text{T}}(\mathbf{W}^{\text{VT}}\mathbf{a})^{\text{VT}} + \lambda_2 \mathbf{I})^{-1}((\mathbf{W}^{\text{VT}}\mathbf{a})^{\text{VT}})^{\text{T}}\mathbf{z} \quad (11)$$

交替使用公式(10)、(11)来迭代计算 \mathbf{a} 和 \mathbf{b} , 该方法称为岭回归双线性模型(Ridge Regressive Bilinear Model)。岭回归虽然增加了计算的稳定性, 但也引入了偏差。

对公式(10)进行变换, 得到:

$$((\mathbf{W}\mathbf{b})^{\text{VT}})^{\text{T}}(\mathbf{W}\mathbf{b})^{\text{VT}}\mathbf{a} + \lambda_1 \mathbf{a} = ((\mathbf{W}\mathbf{b})^{\text{VT}})^{\text{T}}\mathbf{z} \quad (12)$$

可以看出岭回归引入了偏差 $\lambda_1 \mathbf{a}$, 因此岭回归要求惩罚项权重不可以取得太大。但是为了消除偏差 $\lambda_1 \mathbf{a}$ 带来的影响, 我们也可以在式(12)右边加入相同项 $\lambda_1 \mathbf{a}$ ^[12]:

$$((\mathbf{W}\mathbf{b})^{\text{VT}})^{\text{T}}(\mathbf{W}\mathbf{b})^{\text{VT}}\mathbf{a} + \lambda_1 \mathbf{a} = ((\mathbf{W}\mathbf{b})^{\text{VT}})^{\text{T}}\mathbf{z} + \lambda_1 \mathbf{a} \quad (13)$$

但此时式(13)的求解就变得非常复杂, 最简单的解决方法就是使用上一次迭代得到的表情向量结果 \mathbf{a}^{old} 代替式(13)等号右面的 \mathbf{a} :

$$((\mathbf{W}\mathbf{b})^{\text{VT}})^{\text{T}}(\mathbf{W}\mathbf{b})^{\text{VT}}\mathbf{a} + \lambda_1 \mathbf{a} = ((\mathbf{W}\mathbf{b})^{\text{VT}})^{\text{T}}\mathbf{z} + \lambda_1 \mathbf{a}^{\text{old}} \quad (14)$$

则 \mathbf{a} 可以通过下式求得:

$$\mathbf{a} = (((\mathbf{W}\mathbf{b})^{\text{VT}})^{\text{T}}(\mathbf{W}\mathbf{b})^{\text{VT}} + \lambda_1 \mathbf{I})^{-1}(((\mathbf{W}\mathbf{b})^{\text{VT}})^{\text{T}}\mathbf{z} + \lambda_1 \mathbf{a}^{\text{old}}) \quad (15)$$

尽管 \mathbf{a}^{old} 和 \mathbf{a} 存在偏差, 但在迭代过程中 \mathbf{a} 是趋向收敛的, 因此 \mathbf{a}^{old} 和 \mathbf{a} 的偏差随着迭代的进行是逐渐减小的, 直至小到可以忽略。

同理, \mathbf{b} 可以通过下式求得:

$$\mathbf{b} = (((\mathbf{W}^{\text{VT}} \mathbf{a})^{\text{VT}})^{\text{T}} (\mathbf{W}^{\text{VT}} \mathbf{a})^{\text{VT}} + \lambda_2 \mathbf{I})^{-1} ((\mathbf{W}^{\text{VT}} \mathbf{a})^{\text{VT}})^{\text{T}} \mathbf{z} + \lambda_2 \mathbf{b}^{\text{old}} \quad (16)$$

交替使用公式(15)、(16)来迭代计算 \mathbf{a} 和 \mathbf{b} , 直至 \mathbf{a} 、 \mathbf{b} 收敛, 该方法称为 MALS 双线性模型。

则使用 MALS 计算双线性模型的“转移”问题算法如下:

- (i) 输入观测向量 \mathbf{z} , 设定 \mathbf{a} , \mathbf{b} 初始值, 令 $\mathbf{a}^{\text{old}} = \mathbf{a}$ 。
- (ii) 使用公式(15)计算出 \mathbf{a} , 并令 $\mathbf{b}^{\text{old}} = \mathbf{b}$ 。
- (iii) 使用公式(16)计算出 \mathbf{b} , 并令 $\mathbf{a}^{\text{old}} = \mathbf{a}$ 。
- (iv) 重复(ii)、(iii)步直至达到结束条件, 得到 \mathbf{a} 和 \mathbf{b} 。

对于惩罚权重 λ 的设置方式非常灵活, 可以设置为常值, 也可以随着迭代而改变。但需要注意的是, 由于岭回归引入了偏差, 因此权重不可以设置的过大, 而 MALS 双线性模型则没有该限制, 因为该方法在迭代过程中逐渐消除了偏差影响。

4 实验设计与分析

实验所用人脸图片来自 KDEF 人脸表情数据库^[14], 该数据库包含 70 人(35 男 35 女)7 种表情(中性、恐惧、生气、厌恶、高兴、悲伤、惊讶)的人脸图像。选择数据库中表情展现较好的图像进行训练和测试, 选择 12 人(6 男 6 女)7 种表情的正面人脸图像作为训练集, 选择训练集外 6 人(3 男 3 女)5 种表情(中性、恐惧、高兴、悲伤、惊讶)的正面人脸图像进行测试。提取人脸图像的特征点坐标, 以鼻尖为坐标原点进行平移, 以中性表情下不同人脸图像内眼角点间距统一为标准, 对人脸图像特征点坐标进行缩放(不同表情下同一人脸缩放一致), 将人脸划分为 4 个区域, 每个区域特征点坐标排列成列向量作为观测向量(忽略脸边缘特征点), 分别进行对称性双线性模型的训练和转移。实验中权重 λ

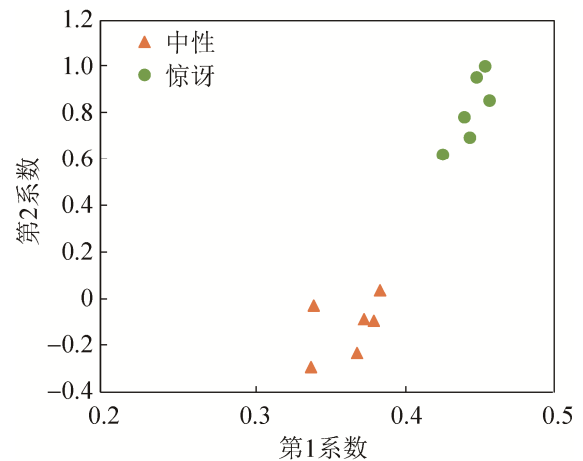
设置为常值, 岭回归双线性模型权重设置为 $\lambda_1 = \lambda_2 = 15$, MALS 双线性模型权重设置为 $\lambda_1 = \lambda_2 = 750$, 由于训练时使用样本较少, 为进一步防止计算双线性模型转移问题时的过拟合现象, 设置迭代次数限制, 当各自代价函数没有明显下降时迭代停止, 设定实验初值, 运行程序验证使用不同方法的效果。

在计算出测试人脸图像的表情和身份表示后, 使用得到的表情和身份向量重新合成观测向量。使用特征点均方差与内眼角点间距的比值作为重建误差, 衡量重建人脸与原人脸的相似度:

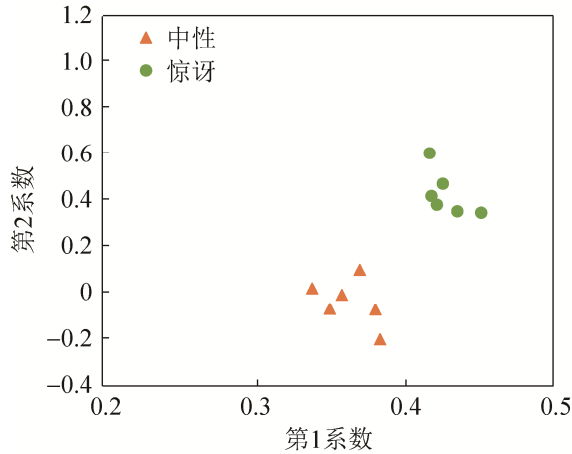
$$ER_{FP} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n \|\mathbf{v}_{i,\text{syn}} - \mathbf{v}_{i,\text{inp}}\|^2}}{d} \quad (17)$$

式中: n 为图像中提取的特征点个数; $\mathbf{v}_{i,\text{inp}}$ 表示从原测试人脸图像提取的第 $i(i=1,2,\dots,n)$ 个特征点坐标向量; $\mathbf{v}_{i,\text{syn}}$ 表示重新合成的第 $i(i=1,2,\dots,n)$ 个特征点坐标向量; d 表示原测试人脸图像中性表情下两个内眼角点的间距。

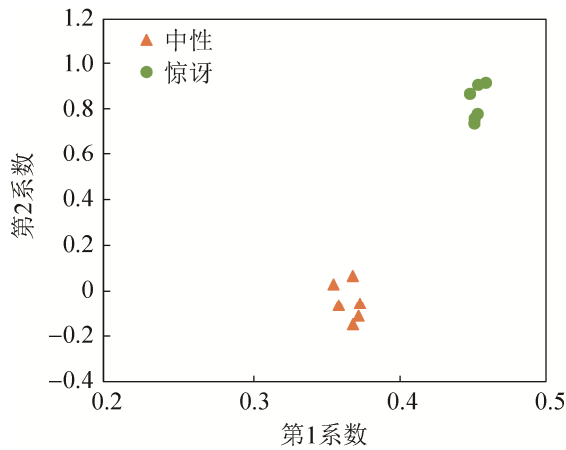
首先测试 3 种模型对身份和表情的分离效果, 为了显示的比较清晰, 选取 6 个不同人脸嘴部区域分别在中性、惊讶两种表情下的表情向量的前两个系数进行展示, 分离效果对比见图 4。



(a) 使用双线性模型得到的表情因子分布



(b) 使用岭回归双线性模型得到的表情因子分布



(c) 使用 MALS 双线性模型得到的表情因子分布

图4 使用3种模型得到的表情因子分布对比

Fig. 4 Comparison of expression factor distributions of three models

从图4可以看出,不同人脸同一表情的表情向量聚集在一起,不同表情的表情向量则差别较大,双线性模型实现了较好的身份和表情的分离效果。而相比于双线性模型,岭回归双线性模型和 MALS 双线性模型使得不同人脸同一表情的表情向量相差进一步减小,为了更加直观的说明上述情况,我们可以比较3种模型的类内散布矩阵 S_w 。假设有 c 个类别,每个类别向量集合 D_i 中含有向量的个数为 n_i ,则 S_w 计算公式见公式(18)、(19)^[15]:

$$S_w = \sum_{i=1}^c S_i \quad (18)$$

$$S_i = \sum_{x \in D_i} (x - \frac{1}{n_i} \sum_{x \in D_i} x)(x - \frac{1}{n_i} \sum_{x \in D_i} x)^T \quad (19)$$

将表情向量的前两个系数作为行向量 x ,则使用双线性模型得到的表情因子类内散布矩阵为 0.190,而使用岭回归双线性模型和 MALS 双线性模型得到的表情因子类内散布矩阵分别为 0.103 和 0.065,岭回归双线性模型和 MALS 双线性模型使得不同人脸同一表情的表情向量相差明显减小,说明身份因素对表情因素的影响进一步减弱,而 MALS 双线性模型由于权重 λ 的设置更加灵活,在本次实验中取得了更好的身份表情分离效果。

然后测试3种模型的稳定性,由于3种不同的方法对因子向量不重要的成分影响更为剧烈^[10],选取同一人脸嘴部区域分别在中性、恐惧、高兴、悲伤、惊讶5种表情下的身份向量的第6和第8个系数进行展示,效果对比见图5。

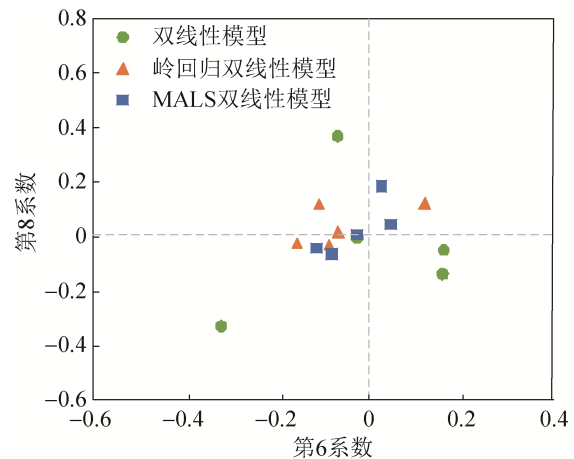


图5 使用3种模型得到的身份因子分布对比

Fig. 5 Comparison of identity factor distributions of three models

从图5可以看出,相比于双线性模型,在引入岭回归和 MALS 后,同一人脸不同表情下的身份向量系数以及相互之间方差明显减小,这解释了岭回归双线性模型和 MALS 双线性模型比双线性模型具有更强的稳定性的原因。我们使用上述3种模型得到该人脸嘴部区域在5种表情下的身份和表情向量后,再重新进行合成,比较3种模型的重建误差 ER_{FP} ,衡量3种模型重建的人脸与原人脸的相似度,对比见表1。

表 1 3 种模型重建误差对比

方法	重建误差 $ER_{FP}(\times 10^{-3})$				
	中性	高兴	恐惧	悲伤	惊讶
双线性模型	4.25	3.24	1.68	2.47	3.44
岭回归	7.26	6.48	5.75	4.74	8.32
双线性模型	5.33	4.76	2.98	4.43	4.41

可以看出, MALS 双线性模型的重建误差比岭回归双线性模型的重建误差明显减少, 说明使用 MALS 双线性模型重建的人脸要比使用岭回归双线性模型重建的人脸更加相似原人脸, 使用 MALS 双线性模型进行表情合成可以取得更好的效果。在使用 MALS 双线性模型时, 由于当其代价函数没有明显下降时便停止迭代, 因此 MALS 双线性模型的重建误差还是比双线性模型的重建误差要大一些。

从以上实验结果可以看出, MALS 双线性模型可以在基本不引入偏差的情况下增加双线性模型的稳定性, 并且增强双线性模型的分离效果, 使用 MALS 双线性模型进行表情合成可以取得更好的合成效果。

5 结论

本文使用双线性模型来进行人脸表情的合成。双线性模型能够有效分离人的身份信息 and 表情信息, 有效消除身份因素对人脸表情合成的影响。然而当观测数据具有相关性或存在噪声信息时, 身份和表情信息的分离过程可能会不稳定。本文的贡献在于将改进迭代最小二乘法(MALS)引入到双线性模型中来稳定计算过程, 改善了岭回归双线性模型由于引入偏差而对表情合成产生的影响, 提高了重建人脸与原人脸的相似度。实验结果表明, 使用 MALS 双线性模型具有更强的稳定性, 有效增强了双线性模型的分离效果, 同时该方法有效改善了直接将岭回归引入双线性模型所产生偏差的影响, 进行表情合成时可以取得

更好的效果。但是该方法还有一些不足之处, 如选择合适的权重需要大量的对比实验, 需要日后更深入的研究。

参考文献:

- [1] S Agianpuye, J L Minoi. 3D Facial Expression Synthesis: a Survey[C]// International Conference on Information Technology in Asia, USA: IEEE, 2013: 1-7.
- [2] X Wan, X Jin. Data-Driven Facial Expression Synthesis via Laplacian Deformation[J]. Multimedia Tools and Applications (S1380-7501), 2012, 58(1): 109-123.
- [3] Yihao Zhang, Weiyao Lin, Bing Zhou, et al. Facial Expression Cloning with Elastic and Muscle Models[J]. Journal of Visual Communication & Image Representation (S1047-3203), 2015, 25(5): 916-927.
- [4] H Liang, R Liang, M Song, et al. Coupled Dictionary Learning for the Detail-Enhanced Synthesis of 3-D Facial Expressions[J]. IEEE Transactions on Cybernetics (S1083-4419), 2016, 46(4): 890-901.
- [5] Chang Jianglong, Zheng Ying, Wang Zengfu. Facial Expression Analysis and Synthesis: a Bilinear Approach[C]// IEEE International Conference on Information Acquisition, USA: IEEE, 2007: 457-464.
- [6] J B Tenenbaum, W T Freeman. Separating Style and Content with Bilinear Models[J]. Neural Comput (S0899-7667), 2000, 12(6): 1247-1283.
- [7] MS Cheema, A Eweawi, C Bauckhage. Human Activity Recognition by Separating Style and Content [J]. Pattern Recognition Letters (S0167-8655), 2014, 50: 130-138.
- [8] HAÏFA NAKOURI, MOHAMED LIMAM. A New Bilinear Approach for Incremental Visual Learning and Recognition[J]. International Journal of Pattern Recognition & Artificial Intelligence (S0218-0014), 2013, 27(5): 487-496.
- [9] X Sun, X Zhang, J Yang, et al. Voice Conversion using Bilinear Model Integrated with Joint GMM-based Classification[C]// IEEE Third International Conference on Information Science & Technology, USA: IEEE, 2013: 1225-1228.
- [10] Dongsoo Shin, Hyung-Soo Lee, Daijin Kim. Illumination-Robust Face Recognition using Ridge Regressive Bilinear Models[J]. Pattern Recognition Letters (S0167-8655), 2008, 29: 49-58.
- [11] Zhicai Ou, Peng Wang, Jianhua Su, et al. Sub-Pattern Bilinear Model and its Application in Pose Estimation of

- Work-Pieces[J]. Neurocomputing (S0925-2312), 2012, 83: 176-187.
- [12] Wang Jihong, Philip K Hopke, Thomas M Hancewicz, et al. Application of Modified Alternating Least Squares Regression to Spectroscopic Image Analysis[J]. Analytica Chimica Acta (S0003-2670), 2003, 476: 93-109.
- [13] S Milborrow, F Nicolls. Active Shape Models with SIFT Descriptors and MARS[C]// 2014 International Conference on Computer Vision Theory and Applications (VISAPP), Portugal: SCITEPRESS, 2014: 380-387.
- [14] Lundqvist D, Flykt A, Öhman A. The Karolinska Directed Emotional Faces - KDEF, CD ROM from Department of Clinical Neuroscience[M]. Psychology section: Karolinska Institutet, 1998.
- [15] Duda R, Hart P, Stork P. Pattern Classification[M]. New York: Wiley, 2001.