

# 基于表情分解-扭曲变形的人工表情合成算法

余重基 李际军

(浙江大学计算机学院, 杭州 310027)

**摘要** 为了能快速有效地生成任意强度的人脸表情图像,提出了一种鲁棒的可以生成带任意强度表情图像的人工表情合成算法,该算法首先通过施加高阶奇异值分解(HOSVD)来把训练集分解为个人、表情和特征3个子空间,并把它们映射到表情子空间中,用来合成任意人脸正面照片的任意强度、任意表情的图像;在生成图像时,不采用通常所使用的线性组合基图像生成法,而是对源图像进行扭曲变形,这不仅能使训练数据和计算量大为减少,还可以生成任意尺寸、任意背景、任意光照、任意色彩或任意姿势的表情图像,且通过二次插值,还可以得到任意强度的表情图像。实验证明,该算法效率较高,且生成的图像效果很好。

**关键词** 表情合成 HOSVD 表情分解 框架

**中图法分类号:** TP391.4 **文献标识码:** A **文章编号:** 1006-8961(2006)03-0372-07

## Facial Expression Synthesis Based on Facial Expression Decomposition and Warping

YU Zhong-ji, LI Ji-jun

(College of Computer Science, Zhejiang University, Hangzhou, Zhejiang 310027)

**Abstract** This paper proposes a robust algorithm of facial expression synthesis. Decompose a train set into person subspace, expression subspace and feature subspace by employing higher-order singular value decomposition (HOSVD), and then map them to expression subspace, which can be used to synthesize an expressive image of arbitrary strength, given any frontal facial image. By warping source image to generate the target image rather than by linear-combining the images in train set as usual, this approach not only lessens the train set data storage and the computation complexity, but also enables the system to cope with facial images of arbitrary size, background, illumination condition, coloration or pose. With square interpolation, expression facial images of any assigned strength can be obtained. Experiments prove that this algorithm has higher efficiency and can generate very excellent images.

**Keywords** expression synthesis, higher-order singular value decomposition (HOSVD), expression decomposition, shape

## 1 引言

目前基于人脸表情的应用正变得越来越重要,从简单的网络虚拟角色向导、远程虚拟通讯到复杂的人脸自动识别、人脸图像数据库的自动查询等,都会涉及面部表情。心理学家把人脸表情分为高兴、悲伤、愤怒、害怕、惊异和厌恶6种基本表情,其余表情是这6种表情的组合。

对于人脸图像,根据需要使之产生不同的表情,称为表情的人工合成。由于表情变化是脸部肌肉群伸缩的结果,因此逼真的表情合成必须符合表情变化时脸部各部位的位置移动规则。开发表情合成算法的实质就是寻找表情变化与各部位位移的映射关系,而鲁棒的算法则应该能把这种映射关系自动地适应各种条件,例如不同个体的同一种表情可能表现方式不尽相同,其最大强度也不一样,而且当背景环境不同、光照条件各异、姿势变化,具有彩色差异

收稿日期:2005-01-31; 改回日期:2005-05-09

第一作者简介:余重基(1961~),男,2005年获浙江大学工学硕士学位,现为杭州第九中学高级教师。主要研究方向为计算机图形学应用,曾开发过多个计算机图形学应用软件。E-mail: palms@yeah.net

时,鲁棒的算法应该仍能工作。

在计算机系统中,要达到上述要求目前还是非常困难的,因此如何以最小的代价近似模拟上述映射关系已成为评价表情合成算法优劣的标准。最早的研究者采用主观经验,对某幅具体的人脸图像人为地指定扭曲变形方案(warping path),以达到表情合成的目的<sup>[1]</sup>。Nur Arad 等人提出的基于径向基函数(radial basis functions,简称RBF)2维扭曲变形方法<sup>[2]</sup>,是根据主观经验插入控制点(anchor points)作为扭曲控制点实现的,但由于没有对如何有效确定这些控制点进行的研究,因此这类主观方法没有任何扩展性可言。为了克服主观性带来的缺点,Kobayashi 和 Hara 通过引入人工神经网络(artificial neural network,简称ANN)来模拟人脸上特征点(feature landmark)的移动与表情变化的映射关系<sup>[1]</sup>,它是通过在图像上加入30个特征点,这不仅成功地识别了6种基本表情(universal facial expression),并且还给出每种表情的强度(strength)计算方法<sup>[1]</sup>,但这个方法需要对人脸的大量特征点进行精确标注,这是一项强度很大的人工操作,该算法更大的缺陷是由于其必须对每个人的图像分别进行训练,所以实用性不强。为了能自动定位特征点,一些学者相继提出了一些自动定位算法<sup>[2-7]</sup>,试图自动定位眼睛或人脸上的所有器官。

具体的表情人工合成算法总是要解决以下两个问题:(1)如何确定表情变化与特征点移动的映射关系;(2)如何生成目标图像。在确定表情变化与特征点移动的映射关系方面,WANG Hong-cheng, Ahuja Narendra 建立的表情分解(expression decomposition)技术<sup>[8]</sup>,从某种意义上说,已接近上述鲁棒性要求,不过其适用范围仍有较大限制,且存储及计算代价太高。该技术通过引入多重线性分析法,具体说就是基于高阶奇异值分解(higher-order singular value decomposition,简称HOSVD)来建立个人与表情的映射模型,用于对任意未知人脸的表情合成。在图像生成时,该算法采用的方法是把最终图像看成是训练集平均图像与其他所有训练图像的线性组合。因为此算法把灰度信息也作为训练的数据,所以只能对固定大小的图像进行操作,为了降低复杂性,不得不只实现灰度图像的表情合成。这个方法最大的问题是只能操作正常、高兴、悲伤、愤怒、害怕、惊异和厌恶7种表情,而没有表情强度的量值变化。

本文提出一种鲁棒的全自动人脸表情生成算

法——基于表情分解-扭曲变形的人工表情合成算法,该算法部分地借鉴WANG Hong-cheng 等人的技术<sup>[8]</sup>,首先采用活动框架模型(active shape model,简称ASM)算法在一幅任意照度、任意姿势和任意表情的人脸正面图像上自动匹配特征点;然后通过HOSVD把训练集分解成个人子空间、表情子空间和特征子空间,并将它们映射到表情子空间中来确定表情变化与特征点的映射关系,该方法实现的映射关系是因人而异的,即在合成表情时能够根据输入的人脸图像确定此人其他表情时特征点的位置。而生成目标图像时这些特征点将被用作图像扭曲变形过程所需的扭曲控制点(warping control points);最后应用二次插值算法生成目标扭曲控制点,就可以进行任意带强度变化的表情合成。

## 2 基于表情分解-扭曲变形的人工表情合成算法

根据人脸的解剖结构和对表情变化的观察,把最能反映表情变化的特定点定义为特征点,因为这些点反映了脸部的几何信息,所以这些点的集合称为几何框架(shape),简称框架。

基于表情分解-扭曲变形的人工表情合成算法(algorithm of decomposition and warping based facial expression synthesis)最基本的思路是把训练集分解后映射到表情子空间中去,以便在输入表情图像后,能生成此人所有的表情框架,再通过二次插值来获得指定强度的表情框架,并将其作为扭曲控制点对源图像进行扭曲变形以便产生指定表情图像。

### 2.1 算法概述

本算法通过引入活动框架模型技术来自动定位框架中的特征点。若每个框架中有 $n$ 个点,每个点有2个坐标,则一个框架将是 $2n$ 维的向量。为了减小计算和存储量,可选择主元分析(principal component analysis,简称PCA)技术降低维度,降维后的框架称为特征向量。

应用多重线性分析技术,把首先与训练中的图像对应的所有特征向量,采用HOSVD技术分解为个人子空间、表情子空间和特征子空间,然后再映射到表情子空间中,具体方法详见2.5节。最后由这个表情子空间,就可以计算出某人的全部种类的表情特征向量。

如果输入训练集以外任何人的正面照片,并指

出当前的表情种类,那么,通过表情子空间的映射,就能够计算出此人的个人特征矩阵,也就是这个人所有表情的特征向量,再把特征向量还原为框架,就可以当作扭曲控制点(warping control points)来对源图像进行扭曲变形,就可生成指定表情的图像。

为了实现表情的强度变化,本文定义了以下 13 种基本表情:无表情、中度高兴、极度高兴、中度悲伤、极度悲伤、中度愤怒、极度愤怒、中度害怕、极度害怕、中度厌恶、极度厌恶、中度惊异和极度惊异。无表情对应的强度值为 0.0,中度表情对应的强度为 0.5,而极度表情则对应 1.0,其余表情强度则由这些表情二次插值得到。

本算法的特点是目标图像的生成不是采用训练集图像的线性组合,而是对源图像进行扭曲变形。若把待生成图像当作训练集中的全部图像(作为基图像)与平均图像进行加权组合,则不但要求训练集中的图像有较高且统一的分辨率,这不仅极大地增加了存储与计算量,而且生成的脸部图像窗口与源图像的光照、色彩方面的差异或发型不同也无法使之与源图像无缝融合,同时还存在着最大的弱点,即不能表现强度变化。而基于表情分解-扭曲变形的人工表情合成算法生成目标图像的手段是对源图像进行扭曲变形,则可以克服上述所有缺点。该算法在扭曲变形时,采用的变形算法是径向基函数扭曲变形算法(radial basis function warping algorithm)<sup>[9]</sup>。这种扭曲算法的优点是其变形的局部性,即与扭曲控制点距离越近的区域受控制点移动的影响越大。这个特点正好符合本算法对扭曲控制的要求。

## 2.2 框架中特征点的确定

框架中的特征点越多,反映的位移信息就越准确,同时运算代价也就越大,另外框架的确定必须符合解剖结构特征,而且在表情变化时,特征点位置移动必须具有代表性。

如图 1 所示,本算法的框架中有 78 个特征点,分别分布在眼睛、眉毛、鼻子、双唇和下颌边界。这些点有的是表情肌肉(群)的力点,在表情变化时,移动最大;有的是肉附着点,在表情变化时很少移动,可以作为参照点;有的则是脸形的边界,其对反映脸部姿势有重大帮助。但前额上方发际线没有被加,因为这样可以避免发型的干扰。

## 2.3 自动定位

采用 ASM 算法来自动定位,需要一定数量的图

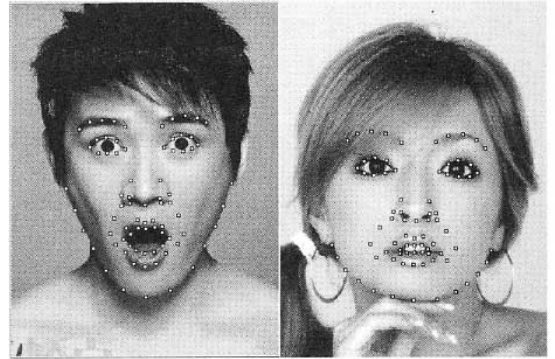


图 1 每幅图像<sup>[10]</sup>上标注 78 个特征点作为框架

Fig. 1 78 landmarks Labeled on each image<sup>[10]</sup> to form a shape

像进行训练。训练时,首先用人工在每幅训练图像上精确地定位相同数量的特征点作为框架,而训练的结果是获得了一个被称为特征点分布模型(point distribution model,简称 PDM)的统计模型和反映特征点周围灰度变化的统计模型。PDM 不仅包括平均框架(mean shape),还包含了框架中个特征点变化的方差。当匹配图像时,则对初始框架(本算法以平均框架为初始框架)的各个特征点根据 PDM 调整坐标,以适应当前图像的匹配。由于 ASM 把上述调整严格限制在 PDM 变化范围内,所以不但不需要像 Snake 那样用人工给出足够精确的初始框架,而且框架中各特征点的位置移动被严格限制在训练集的方差范围内,这就保证了框架的结构关系。也就是说,该算法只会按照训练的模式去匹配图像,而不会像 Snake 那样因初始框架不当而出现严重的错误匹配。

### 2.3.1 特征点分布模型

PDM 的获得分为:标定训练集(labeling the training set)、对齐训练集(aligning the training set)和统计训练集(capturing the statistics)3 个步骤。

标定训练集就是在每幅训练图像上标定框架,若训练集中有  $N$  幅图像,每个框架上有  $n$  个特征点,则标定第  $i$  幅图像上第  $j$  个特征点的坐标记为  $(x_{i,j}, y_{i,j})$ 。标定训练集就获得了  $N$  个  $2n$  维向量  $x_i, x_i$  表示框架中每个特征点坐标:

$$x_i = [x_{i,0}, y_{i,0}, x_{i,1}, y_{i,1}, \dots, x_{i,n-1}, y_{i,n-1}]^T, 1 \leq i \leq N \quad (1)$$

为了正确捕获训练集中框架间同一特征点的位

置变化方差,必须对齐训练集。对齐过程实际上就是调整框架的姿态(pose),即对整个框架进行必要的缩放、旋转或移动,使每个框架相互对齐,详细对齐算法参见 Ghassan Hamarneh 改进的方法<sup>[11]</sup>。

设平均框架的向量为  $\bar{x}$ , 对齐后第  $i$  框架的向量  $x_i$  与  $\bar{x}$  的差为  $d_{x_i}$ , 即

$$d_{x_i} = x_i - \bar{x} \quad (2)$$

应用主元分析可以获得一些主元,且每个主元都是原来变量的线性组合。由于主元相互正交,所以没有信息冗余。这些主元作为整体,就形成了数据空间的正交基:

$$d_{x_i} = b_{i,0}p_0 + b_{i,1}p_1 + \dots + b_{i,2n-1}p_{2n-1}$$

式中,  $p_k$  是第  $k$  个主元轴(即列矢量),且标准化为单位长度,也就是说  $p_k^T p_k = 1$ 。又因为主元都是相互正交的,所以它们也相互正交。 $b_{i,k}$  可以看成是缩放量,它决定了  $p_k$  在第  $i$  个框架中的权重。式(2)也可写成如下形式:

$$x_i = \bar{x} - d_{x_i} \quad (3)$$

记  $P = [p_0, p_1, \dots, p_{2n-1}]$ ,  $b_i = [b_{i,0}, b_{i,1}, \dots, b_{i,2n-1}]^T$ , 于是

$$x_i = \bar{x} - P b_i \quad (4)$$

因为  $P$  是正交矩阵,所以  $P^{-1} = P^T$ , 于是

$$b_i = P^{-1}(x_i - \bar{x}) = P^T(x_i - \bar{x}) \quad (5)$$

若设方差矩阵的特征值  $\lambda_1 > \lambda_2 > \dots > \lambda_r > \lambda_{r+1} > \dots > \lambda_n$ ,  $r$  为矩阵的秩,设前  $t$  个主元已经有足够的精度表示原数据了,  $t \leq r$ , 则  $P$  只需要  $t$  列( $b$  只需要  $t$  行), 即  $P = [p_0, p_1, \dots, p_{t-1}]$ ,  $b_i = [b_{i,0}, b_{i,1}, \dots, b_{i,t-1}]^T$ 。

对于任何  $x$  (包括训练集以外的情况), 式(3)都是成立的。于是 PDM 标准的写法就是

$$x = \bar{x} - P b \quad (6)$$

### 2.3.2 自动匹配过程

典型的 ASM 自动匹配过程是通过在框架法线上搜寻图像的边界来实现的,但由于这样有诸多弊端,故本算法通过灰度模型来实现。当前特征点  $p_i$  及其前后点(即  $p_{i-1}, p_i, p_{i+1}$ )所组成的曲线称为边界,描述边界在  $p_i$  点处法线上的像素灰度变化所获得的统计模型叫特征点  $p_i$  的灰度模型(gray level model)。

ASM 根据每个特征点的灰度模型,首先在法线方向上搜索当前图像,以求得姿态参数和框架参数的调整量,然后对这个调整量以 PDM 进行约束,使

之保持固有的几何特征。因为统计学理论认为绝大多数分布点都位于平均值的 3 个标准差以内,故本算法把  $b_j$  限制在  $-3\sqrt{\lambda_j} \leq b_j \leq 3\sqrt{\lambda_j}$  中。重复迭代上述过程,直到没有明显的参数修改为止,图像特征点自动匹配就完成了。

### 2.4 对训练集的组织

训练后,框架的维数就降低了,称为特征向量(即第  $i$  个框架被表示为  $b_i$ ,  $b_i$  称为特征向量)。把训练集中所有的特征向量组织成 3 维张量(tensor)  $A \in \mathbf{R}^{m \times g \times l}$ , 其中  $m$  为训练集中的人数,  $g$  是表情种类的数目,而  $l$  则是特征向量的长度。应用 HOSVD<sup>[12]</sup>,  $A$  分解为个人子空间  $U^{\text{person}}$ 、表情子空间  $U^{\text{expression}}$  和特征子空间  $U^{\text{feature}}$  3 个子空间。即

$$A = S \times_1 U^{\text{person}} \times_2 U^{\text{expression}} \times_3 U^{\text{feature}} \quad (7)$$

式中,  $S$  是核心张量,用于刻画个人、表情和特征子空间之间的相互作用,  $U^{\text{person}}$ 、 $U^{\text{expression}}$  和  $U^{\text{feature}}$  3 个相互正交的矩阵分别代表个人、表情和特征子空间。这些矩阵中的行代表某个特定模式的向量。例如个人子空间  $U^{\text{person}} = [u_1^p, u_2^p, \dots, u_n^p, \dots, u_l^p]^T$  (上角  $p$  代表 person) 中的行向量  $u_n^p$  表示第  $n$  个人的所有特性(characteristics), 而表情子空间  $U^{\text{expression}} = [u_1^e, u_2^e, \dots, u_n^e, \dots, u_l^e]^T$  (上角  $e$  代表 expression) 中的行向量  $u_n^e$  表示第  $n$  个表情的所有特性。对于列向量而言也类似,也就是说,列向量代表其他模式的贡献,如在个人子空间中的列向量是表示特征表情和特征框架的贡献(the contributions of eigenexpressions and eigenshapes)。

$S$  不是对角矩阵,但是其相似于矩阵奇异值分解的对角奇异值矩阵。

为了更好地进行陈述,下面先对多重线性分析技术的有关部分作个回顾。

#### 2.4.1 张量的展开

张量是多重线性分析中的基本对象,也是对向量、矩阵等概念的扩展。向量与矩阵都是张量的特例,即 1 维的张量叫向量,2 维的张量叫矩阵,3 维的张量叫盒(box)。在张量中,维(dimension)称作模式(mode)。把张量  $A$  沿  $n$  模式展开为矩阵,  $A$  的操作叫展开(unfold), 记作  $uf(A, i)$ 。给定  $N$  阶张量  $A \in \mathbf{R}^{I_1 \times I_2 \times \dots \times I_N}$ , 则  ${}_n A \in \mathbf{R}^{I_n \times (I_1 I_2 \dots I_{n-1} I_{n+1} I_{n+2} \dots I_N)}$  的第  $i_n$  行, 第  $(i_{n+1} - 1)I_1 I_2 \dots I_{n-1} I_{n+2} I_{n+3} \dots + \dots + (i_{N-1} - 1)I_1 I_2 \dots I_{n-1} + (i_1 - 1)I_2 I_3 \dots I_{n-1} + \dots + i_{n-1}$  列的元素为  $a_{i_1, \dots, i_N}$ 。

### 2.4.2 张量与矩阵的积

张量  $A$  与矩阵  $U \in \mathbf{R}^{J_n \times I_n}$  的第  $n$  模式积记为  $A \times_n U$ , 定义为

$$(A \times_n U)_{i_1 \dots i_{n-1} i_{n+1} \dots i_N} = \sum_{i_n} a_{i_1 \dots i_n} u_{i_n i_n} \quad (8)$$

### 2.4.3 张量的高阶奇异值分解

HOSVD 是奇异值分解 (singular value decomposition, 简称 SVD) 在张量中的推广。

**定理** 任何张量  $A$  可以被写成以下积的形式:

$$A = S \times_1 U^{(1)} \times_2 U^{(2)} \times \dots \times_N U^{(N)} \quad (9)$$

式中, 第  $n$  个子空间为  $U^{(n)} = (u_1^{(n)}, u_2^{(n)}, \dots, u_{I_n}^{(n)})$ ,  $n = 1, 2, \dots, N$  ( $n$  代表子空间名 person, expression, feature) 是一  $I_n \times I_n$  大小的矩阵

$S$  是个  $I_1 \times I_2 \times \dots \times I_N$  大小的张量, 其子张量  $S_{i_n} = \alpha$  具有全正交性质, 且以  $F$  范数  $\|S_{i_n}\|_F$  为序排列。 $S$  被称为核张量, 是个全矩阵 (不是伪对角矩阵)。

$U^{(i)}$  提供了沿着第  $i$  个模式的能量方向, 由于求  $U^{(n)}$  的方法就是把  $A$  沿第  $n$  个模式展开, 进行 SVD 后的左奇异值矩阵 (其列矢量为特征矢量), 因此  $N$  阶张量的 HOSVD 就是计算  $N$  个  $I_1 I_2 \dots I_{n-1} I_{n+1} \dots I_N$  不同矩阵的 SVD, 即

$$S = A \times_1 U^{(1)} \times_2 U^{(2)} \times \dots \times_N U^{(N)} \quad (10)$$

### 2.5 根据训练集生成待测表情框架

如前所述, 训练集张量  $A$  虽已被 HOSVD 分解为 3 个子空间。但为了实现表情分解, 还要把它们映射为表情子空间, 这是模拟表情变化与框架调整对应关系的关键。

对应于表情子空间的表情张量  $T^{\text{expression}[8]}$  定义为

$$T^{\text{expression}} = S \times_2 U^{\text{expression}} \times_3 U^{\text{feature}} \quad (11)$$

从本质上讲, 表情张量  $T^{\text{expression}}$  定义了一个与所有表情相关联的所有脸部特征基矩阵系列, 其与  $A$  及  $S$  有相同的维度。这样就把 3 个子空间映射到了表情子空间。

若有一张表现第  $i$  ( $1 \leq i \leq 13$ ) 种表情的人脸图像, 且要求此人的其余表情, 则由于待测张量  $T_{\text{test}}$  是个  $1 \times 1 \times I_3$  张量 (使用第 3 模式), 所以个人向量  $u^p$  表示如下<sup>[8]</sup>:

$$u^p = uf(T_{\text{test}}, 1)^T \cdot (uf(T^{\text{expression}}(i), 1))^{-1} \quad (12)$$

式中,  $uf(T, 1)$  表示张量  $T$  的第 1 模式 (个人) 展开, 于是合成的表情张量为

$$T_{\text{syn}} = (S \times_2 U^{\text{expression}} \times_3 U^{\text{feature}}) \times_1 u^p \quad (13)$$

有了合成表情张量  $T_{\text{syn}}$ , 此人所有表情的特征向量  $b$  就得到了, 然后根据合成框架  $s_{\text{syn}} = \bar{x} + Pb$  就可以还原出所需的表情框架, 并以此用来形成扭曲控制点, 最后即可通过扭曲变形输入图像来产生相应的表情图像。

### 2.6 实现细节

输入 1 个训练集以外的任何人脸图像, 并指出图像中的表情类型, 若要求合成某强度的另一种表情, 则算法要首先以 ASM 对其进行自动定位, 来获得当前图像的框架, 但这个框架并不能直接当作待测张量  $T_{\text{test}}$  用于求个人向量  $u^p$ , 还必须先与平均框架进行对齐 (缩放、旋转或平移) 后, 才能转换成特征向量作为待测张量  $T_{\text{test}}$ 。设源框架为  $s_s$  (下角  $s$  代表 source), 对齐后为  $\hat{s}_s$ , 平均框架为  $\bar{x}$ , 则作为待测张量的特征向量为

$$T_{\text{test}} = P^T (\hat{s}_s - \bar{x}) \quad (14)$$

首先应用式 (12) 计算出此人的个人向量  $u^p$ , 再根据式 (13) 计算出此人的合成表情张量  $T_{\text{syn}}$ 。如果要求合成的表情为第  $e$  种表情, 其强度为  $s$  ( $0 < s \leq 1$ ), 则从表情张量  $T_{\text{syn}}$  中取出无表情、第  $e$  种中度表情和第  $e$  种极度表情等 3 个特征向量, 并还原为框架, 分别记为无表情框架  $s_0$ 、中度表情框架  $s_m$  和极度表情框架  $s_f$ 。对这 3 个框架进行二次插值, 就可方便地获得目标框架  $s_d$ 。二次插值的计算公式如下:

$$s_d = \frac{1}{4} [2(s - 0.5)(s - 1)s_0 - s(s - 1)s_m + 2(s - 0.5)s_f] \quad (15)$$

$s_d$  之所以仍不能直接用于扭曲变形, 可能是由于以下原因: (1) 经二次插值后的框架不符合框架约束; (2) 没有与源图像对齐, 所以还要进行约束, 使之  $-3\sqrt{\lambda_j} \leq b_j \leq 3\sqrt{\lambda_j}$ , 然后把它对齐于源框架, 才能用作扭曲变形过程的控制点。

## 3 实验

为了验证本文算法的有效性, 选取了 6 个人的照片作为训练集进行了表情合成实验。6 人中 3 位成年男性, 2 位成年女性和 1 位 11 岁女孩。每个人的照片有 13 张, 分别表现 13 种基本表情, 即无表情、中度高兴、极度高兴、中度悲伤、极度悲伤、中度愤怒、极度愤怒、中度害怕、极度害怕、中度厌恶、极度厌恶、中度惊异和极度惊异。

由于每个特征点有  $x$  和  $y$  两个坐标值,78 个特征点意味着框架向量是  $78 \times 2 = 156$  维向量,因此应采用 PCA 进行降维处理来得到 68 维的向量(PDM 中的  $b$  向量),于是张量  $A \in \mathbf{R}^{6 \times 13 \times 68}$ 。

图 2(a) 是一张训练集以外的人脸照片<sup>[13]</sup>,人工判断表情为中度高兴,若用算法自动合成此图像其余表情的框架,则经扭曲变形,就可以产生任意表情的图像。图 2 显示了其中的 5 种表情图像。

从图 2 中可以看到,原始图像的表情为中度高兴,合成的极度高兴图像眼睛眯得更拢,嘴角更为上提。0.8 悲伤的合成图像中,的确体现了悲伤时的特征,如嘴角向下向外,鼻翼向外扩展;惊异时,眼睛睁大,下巴下沉,嘴上翘,眉上提;厌恶时,嘴变小了。这些都说明本算法生成的目标图像符合表情变化的

规律。

图 3 是另一组表情合成的实验结果图像,各行的第 1 张是原始图像<sup>[10]</sup>,其余都是合成的图像。由图 3 可以看出,表情变化与框架变化对应关系符合逻辑地出现因人而异变化。

3 张原始图像尺寸(分辨率)不同,不仅背景和光照不一样,表情种类也不尽相同,但都能被正确处理,体现了算法的鲁棒性。另外 HOSVD 的应用不但使得算法对相机的远近、光照条件,原始图像的表情种类及强度等条件不敏感,而且扭曲变形技术引入使得算法对原始图像的尺寸和色彩不敏感。

上述实验结果证明基于表情分解-扭曲变形的人工表情合成算法是有效的和鲁棒的。算法可以产生各种表情强度的目标图像。

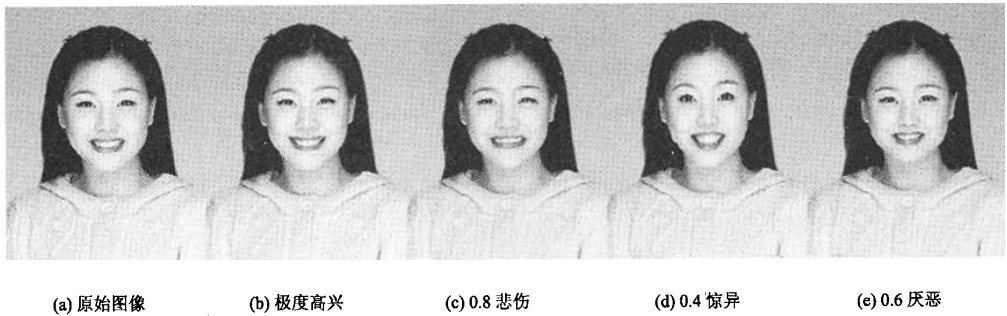


图 2 表情合成的效果

Fig. 2 The results of the synthesis



图 3 另一组表情合成效果

Fig. 3 Another group of synthesis results

## 4 结 论

本算法的突出优点是能够对任意尺寸,任意分辨率的图像进行有效处理,实验结果显示,合成图像不但能保持源图像的基本特征、光照效果和色彩,而且可以生成任意强度的任意表情。

引入扭曲变形过程可使得训练集不必保留灰度信息,而只需记录框架信息,这不仅降低了训练集的信息量,有助于快速计算,更重要的是,此举可以保证算法的灵活性以及生成的图像的真实性。若采用基图像线性组合,则存在以下 3 个缺陷:首先就必需使训练集图像分辨率足够高,如果要产生彩色图像,则原始信息量将还要扩大 3 倍,尽管可使用 PCA 降维,但由于仍然需要很大的容量,因而不利于计算;其次,通过基图像线性组合,只能产生一个脸形窗口,且这个窗口很难保证与原始图像具有相同的光照效果,即很难保证相同的光照角度、相同的照度等,另外也很难保证与当前图像的拥有相同的色彩,即很难保证有相同的色温和相同的色饱和度;第 3 个缺陷是不能进行表情强度的量值变化,灵活性欠佳。正是由于本算法引入了引入扭曲变形过程,所以才克服了上述缺陷。

在本研究前,尽管笔者已经研究过表情变化规律,得知表情强度变化不是线性的,但二次插值能较好地近似变化规律。限于篇幅,本文没有对此进行深入讨论。实验表明,本算法采用二次插值计算目标扭曲控制点既保证了精度,也很好地降低了系统开销。笔者还将进一步研究开口问题:若给定的源图像不开口,但有些表情是会张开嘴的,如大喜过望,会开口大笑,则此时合成的图像有欠失实,而模拟生成牙齿和舌头将有助于改善目标图像;另一个问题是褶皱,如表情变化时,在脸部表情肌群的作用下,会在一些部位产生褶皱,而解决褶皱问题还需要更多地研究解剖学。

## 参考资料 (References)

- 1 Leung Marry Y Y, HUNG Yen-hui, King Irwin. Facial expression synthesis by radial basis function network and image warping [A]. In: Proceedings of IEEE International Conference on Neural Networks [C], Washington, DC, USA, 1996:1400 ~ 1405.
- 2 GENG Xin, ZHOU Zhi-hua, CHEN Shi-fu. Eye location based on hybrid projection function [J]. Journal of Software, 2003, 14(8): 1394 ~ 1400. [耿新, 周志华, 陈世福. 基于混合投影函数的眼睛定位 [J]. 软件学报, 2003, 14(8):1394 ~ 1400.]
- 3 FENG Guo-can, YUEN Pong-c. Multi cues eye detection on gray intensity image [J]. Pattern Recognition, 2001, 34(5): 1033 ~ 1046.
- 4 ZHOU Zhi-hua, GENG Xin. Projection functions for eye detection [J]. Pattern Recognition, 2004, 37(5): 1049 ~ 1056.
- 5 Vladimir Vezhnevets, Anna Degtiareva. Robust and accurate eye contour extraction [A]. In: Proceedings Graphic on 2003 [C], Moscow, Russia, 2003: 81 ~ 84.
- 6 Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models [A]. In: Proceedings, First International Conference on Computer Vision [C], London, England: IEEE Computer Society Press, 1987: 259 ~ 268.
- 7 Cootes T, Taylor C, Cooper D, et al. Active shape models-their training and application [J]. Computer Vision and Image Understanding, 1995, 61(1): 38 ~ 59.
- 8 WANG Hong-cheng, Ahuja Narendra. Facial expression decomposition [A]. In: ICCV2003 [C], 2003: 958 ~ 965.
- 9 Jonas Gomes, Lucia Darsa, Bruno Costa, et al. Warping and morphing of graphical objects [M]. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998: 235 ~ 237, 45 ~ 48.
- 10 壁纸网站 [EB/OL]. [http://photo.myrice.com/class/class\\_nent10.html](http://photo.myrice.com/class/class_nent10.html)
- 11 Ghassan Hamarneh. Active shape models, modeling shape variations and gray level information and an application to image search and classification [EB/OL]. <http://www.ae.chalmers.se/~jessi/>
- 12 Alex M, Vasilescu O, Demetri Terzopoulos. Multilinear analysis of image ensembles: Tensor Faces [A]. In: The Proceedings of the European Conference on Computer Vision [C], Copenhagen, Denmark, 2002: 447 ~ 460.
- 13 Fanta Morph [EB/OL]. <http://www.abrosoft.com/>