

基于局部表情参数化的三维表情脸像合成

张申^{1*} 吴志勇² 蔡莲红¹

¹(清华大学计算机科学与技术系, 北京, 100084)

²(香港中文大学系统工程与工程管理学系, 香港)

摘要: 人脸表情合成是人机交互领域长期以来的研究热点。本文采用 MPEG4 人脸动画画框架, 提出了一种基于局部表情参数化的三维表情脸像合成方法。在人脸动画参数 (FAP) 的基础上, 定义了局部表情状态参数 (Partial Expression Parameter, 简称 PEP), 建立了 PEP 与 FAP 参数之间的关联规则。在此基础上, 实现了一个三维表情脸像编辑器, 利用 PEP 参数可以快速编辑合成人脸表情, 以适应可视语音、对话系统中人脸表情实时变化的需求。

关键词: 局部表情状态; 人脸动画参数; MPEG4

1. 引言

人脸表情是人机交互场景中的重要特征之一。随着情感计算等相关领域的研究深入, 对人脸表情的分析与合成, 实现具有丰富表现力的人脸动画不仅是研究的热点, 而且也是各种人机交互场景中的关键问题。

人脸表情合成主要有数据驱动和参数控制两种方法^[1]。数据驱动利用大量的视频图像进行拼接, 合成真实感很强的人脸表情; 参数化的方法利用动画参数驱动三维人脸模型, 实现复杂的人脸表情动作。目前参数化方法得到了广泛应用, MPEG4 的人脸动画框架^[2]定义了 84 个人脸特征参数和 68 个人脸动画参数, 这套参数广泛应用于人脸动画领域; 但如何利用 MPEG4 人脸动画框架实现复杂表情的合成, 依然是一个开放的领域。

本文在 MPEG4 人脸动画框架的基础上, 提出一种基于局部表情参数化的三维脸像合成方法; 实现了一个三维表情脸像编辑器, 可以对通用三维人脸模型进行编辑, 设定人脸动画参数, 生成多种复杂面部表情动作; 合成的三维表情脸像可应用于有表现力的可视语音合成系统^{[6][7]}。

2. MPEG4 人脸动画参数

MPEG4 人脸动画框架面向三维人脸模型, 定义了两组重要参数: 一是人脸定义参数 (Facial Definition Parameter), 定义了人脸特征点; 二是人脸动画参数 (Facial Animation Parameter), 定义了各种基本面部动作。MPEG4 人脸动画框架的思想之一是要实现模型

资助项目: 国家自然科学基金 (60418012, 60433030)

联系作者: 张申, E-mail: zhangshen05@mails.tsinghua.edu.cn

无关性、参数通用性，动画参数可以适用于不同的人脸模型和应用场景；但在参数设定和具体实现方面，MPEG4 并没有提供统一方案。

人脸定义参数（FDP）是对三维人脸模型拓扑结构的定义，实现了对人脸模型的校准归一；其中 84 个人脸特征点，主要集中在唇、眼、眉、下颌等器官。人脸动画参数（FAP）是对人脸表情动作的定义，共计 68 个参数；其中两个高层参数分别定义了人脸基本表情和视位^[6]（人发音时唇部的动作状态）；另外 66 个底层参数分别与FDP关联，共同定义了人脸关键部位的表情动作，比如眨眼、皱眉、嘴角上翘等。

MPEG4 人脸动画框架为合成人脸表情提供了很好的基础，但其中底层FAP只是定义了面部的各种基本动作，两个高层FAP又只是在语义层面定义了表情和视位状态；而如何利用FAP合成复杂人脸表情，尚没有统一标准。Raouzaïou等人^[3]利用FAP实现了基本表情的参数化描述，同时提出了基于参数统计的表情合成方法；Fabio等人^[4]对FAP相关性进行研究，提出了一种FAP插值方法。本文借鉴他们的工作，提出了一种基于局部表情参数化的人脸表情合成方法，实现了复杂人脸表情的参数合成。

3. 基于局部表情参数化的脸像合成

3.1 局部表情

人类表情主要是通过面部几个关键器官体现的，如眉、唇、眼、下颌等。MPEG4 按面部区域对FAP进行简单划分^[2]；Fabio等人^[4]在分析FAP相关性时，根据人脸局部表情动作选取FAP参数。Raouzaïou等人^[3]对六种基本表情的描述主要用到了 31 个FAP参数，其中 16 个与唇部动作相关，8 个与眉毛相关，4 个与眼部相关。由此可见，在合成人脸表情时，某些面部区域（如耳朵）并不起重要作用；而眼、眉、唇等区域则直接影响到表情合成效果。借鉴上述工作，本文提出基于局部表情参数化的脸像合成方法，其主要思想是按照表情动作划分面部关键区域，在关键区域内定义局部表情状态参数（Partial Expression Parameter,简称PEP），利用PEP控制合成局部表情；同时考虑局部表情动作的关联，对局部表情进行调整融合，合成最终表情脸像。

3.2 局部表情状态参数

分析表情与人脸器官状态变化的关系，本文选取眉、眼、唇、下颌 4 个面部区域作为表情合成基本单元，每个区域对应多个表情状态参数，分别控制该区域的人脸器官状态，从而合成不同表情。参数定义如表 1 所示：

表 1 局部表情状态参数定义表

局部区域	PEP 编号	表情动作描述	
		参数为正	参数为负
眉	1.1	眉毛上扬	眉毛下垂

	1.2	眉毛紧皱	眉毛舒展
	1.3	眉梢下垂、眉尖上挑	眉梢上挑、眉尖下垂
眼	2.1	眼皮张开	眼皮闭合
	2.2	眼球向左看	眼球向右看
	2.3	眼球向下看	眼球向上看
唇	3.1	嘴唇张开	嘴唇闭合
	3.2	嘴角向上弯曲	嘴角向下弯曲
	3.3	唇向前突 (撅嘴)	唇向后缩 (抿嘴)
	3.4	嘴角左右张开 (咧嘴)	嘴角左右收缩
下颌	4.1	下巴向下张	下巴向上收
	4.2	下巴向左撇	下巴向右撇

如表 1 所示，每个局部表情状态参数 (PEP) 描述了对应区域一组连续变化的表情状态。PEP 在 $[-1,+1]$ 归一化区间的连续变化，对应了该局部区域的状态变化。PEP 的数值决定了表情的变化幅度；PEP 的符号决定了表情的范畴；PEP 为 0 对应中性表情状态，为正对应愉悦度、激发度高的表情状态。以

PEP 参数 3.2 为例，它描述了嘴唇的弯曲变化，状态变化如图 1 所示。

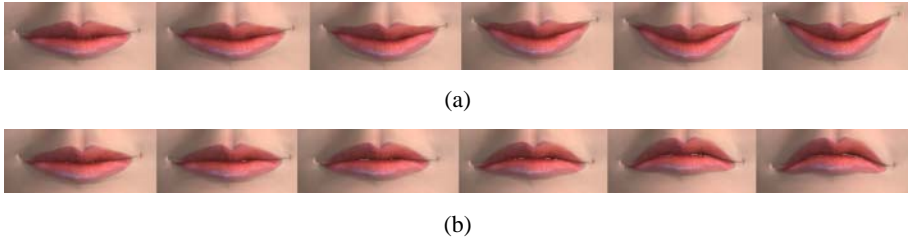


图 1. 唇部区域局部表情状态变化示意图：

(a) 状态参数 3.2 从 0 到 1 连续变化；(b) 状态参数 3.2 从 0 到 -1 连续变化。

与 FAP 相比，PEP 更为方便地描述了人脸常见表情状态。FAP 与人脸特征点(FDP)对应，定义了以特征点为中心局部范围的表情动作，包括动作类型和动作幅度；而 PEP 则直接跟人脸器官对应，刻画了该器官的表情状态；局部表情状态包含了多个 FAP 参数，因此 PEP 刻画的是一组相关 FAP 共同作用下的表情状态。在实现时，PEP 主要解决两个问题：一是 FAP 之间的相关性，如何协同多个 FAP 状态变化的关系；二是通过局部动作的组合，刻画了局部器官的表情状态，

3.3 PEP 与 FAP 的关联规则

对于三维人脸模型，局部表情状态的底层描述是 FAP 参数；当利用 PEP 参数合成局部表情时，还需要建立 PEP 与 FAP 的关联，包括 PEP 与哪些 FAP 关联，PEP 的变化如何影响 FAP 的变化。在实验中，本文借鉴 Fabio^[4] 提出的 FAP 参数插值方法，建立了 PEP-FAP 线性关联规则，得到了比较满意的表情合成效果。

对于局部区域 R 内的第 i 个表情状态参数 P_i^R ，首先选取 R 内的一组 FAP 参数与之关联，进而确定关键 FAP。关键 FAP 是指与 P_i^R 所描述的表情状态最相关的 FAP；例如状态参数 3.1 刻画了嘴唇的张开（闭合）状态，我们选定 FAP5 (raise_b_midlip, 下唇中

部的上下运动)作为关键 FAP。表 2 定义了 PEP 与 FAP 的对应关联。定义的 12 个 PEP 参数对应关联了 40 个 FAP 参数；换句话说，我们用 12 维的 PEP 替代 40 维的 FAP，从而降低了人脸表情参数化合成的复杂度。

表 2. 局部表情状态参数 PEP 与 FAP 参数的关联表

关键区域	PEP 编号	FAP 个数	关键 FAP	关联的 FAP 参数
眉	1.1	6	F33	[F33, F31, F35, F34, F32, F36]
	1.2	2	F37	[F37, F38]
	1.3	6	F31	[F31, F35, F33, F32, F36, F34]
眼	2.1	4	F19	[F19, F21, F20, F22]
	2.2	2	F23	[F23, F24]
	2.3	2	F25	[F25, F26]
唇	3.1	12	F5	[F5, F10, F11, F52, F57, F58, F4, F8, F9, F51, F55, F56]
	3.2	4	F12	[F12, F13, F59, F60]
	3.3	2	F16	[F16, F17]
	3.4	4	F6	[F6, F7, F53, F54]
下颌	4.1	1	F3	[F3]
	4.2	1	F15	[F15]

给定局部表情状态参数 P_i^R ，设其所关联的 FAP 向量为 $[F_1^R, F_2^R, \dots, F_m^R]$ (m 为关联的 FAP 个数)，设关键 FAP 为 F_1^R 。对于关键 FAP，满足如下的分段线性映射关系：

$$F_1^R = f(P_i^R) = \begin{cases} P_i^R \cdot F_+^R & (P_i^R \in [0, +1]) \\ P_i^R \cdot F_-^R & (P_i^R \in [-1, 0]) \end{cases} \quad (1)$$

其中 F_+^R 和 F_-^R 是 P_i^R 所刻画的局部表情分别在正向和负向状态下 F_1^R 对应的最大幅度值。对于非关键 FAP 参数，

我们借鉴 Fabio 对 FAP 相关性的研究，建立如下映射关系：

$$F_j^R = \alpha_k^j \cdot F_k^R \quad (\alpha_k^j \in [-1, +1], j \in [2, \dots, m], k \in [1, 2, \dots, m], k \neq j) \quad (2)$$

其中 α_k^j 是 FAP 参数 F_j^R 和 F_k^R 的线性相关系数， α_k^j 的设定满足如下规则：

- (1) 如果 F_j^R 和 F_k^R 互不相关，那么 $\alpha_k^j = 0$ ；
- (2) 如果 F_j^R 和 F_k^R 相关，且具有相同的变化方向，那么 $\alpha_k^j > 0$ ；
- (3) 如果 F_j^R 和 F_k^R 相关，且具有相反的变化方向，那么 $\alpha_k^j < 0$ ；

在实验中， F_k^R 一般取关键 FAP 参数，即非关键 FAP 直接与关键 FAP 相关。我们以 PEP 参数 1.3 (记为 P13, 参见表 1) 为例，具体说明上述设定规则。参数 P13 刻画了眉毛“内八字”(眉尖上挑、眉梢下垂)和“外八字”(眉梢上挑、眉尖下垂)的表情状态(参见图 2(b))。以左眉为例，P13 关联的 FAP 参数有 [F31, F33, F35]，分别对应于左眉眉尖、眉中、眉梢的动作，我们选定眉尖参数 F31 作为关键 FAP。根据 P13 对表情状态的描述，眉尖和眉梢应具有相反的变化方向，那么 F35 和 F31 的相关系数为负；而无论眉尖上扬还是下垂，眉中始终保持上扬状态，那么当 F31 为正向时，F33 和 F31 的相关系数为负，反之两者的相关系数为正。相关系数取值参考 Fabio 的工作^[4]，F33 与 F31 的相关系数取

值 0.9008, F35 和 F31 的相关系数取值 0.7584。

给定局部表情状态参数, 根据公式(1)(2)我们就可以计算得到相关 FAP 的值, 从而合成局部表情。实验中利用 PEP 合成的局部表情结果如图 2 所示。

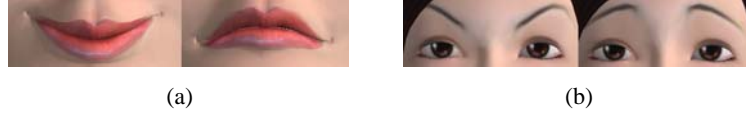


图 2. 参数合成的局部表情状态: (a)局部表情参数 3.2 (b)局部表情参数 1.3

3.4 PEP 参数之间的关联规则

在 3.3 小节的关联模型中, 我们假设各局部表情状态参数相互独立, 即某个区域的表情变化不会影响其他区域。实际上, 人脸在表情动作时, 各区域的表情状态往往是相互关联的; 如眉与眼, 唇与颌的动作都会互相影响。同时, 由于人脸的对称性, 表情状态往往也是左右对称的。因此在上面的关联模型中, 我们又引入了 PEP 参数之间的关联规则。我们认为, PEP 参数之间的关联分为两类, 一是左右对称关联; 二是主次动作关联。具体定义如表 3 所示。

表 3 局部表情状态参数之间的关联

关联关系	关联区域	状态参数	关联表情
左右对称	左眉	1.1 - 1.1	双眉同时上扬/压低
		1.2 - 1.2	双眉同时向眉心皱紧/向两侧舒展
		1.3 - 1.3	双眉呈“内八字”或“外八字”形状
	右眉	2.1 - 2.1	两眼睁开/闭合
		2.2 - 2.2	向左看/向右看
		2.3 - 2.3	向上看/向下看
主次动作	眉-眼	1.1 - 2.1	扬眉-睁眼/低眉-闭眼
	眉-眼	1.3 - 2.1	眉倒竖-眼睁大/眉低垂-眼微闭
	眼-眼	2.3 - 2.1	向上/下看时, 眼球带动眼皮运动
	唇-唇	3.3 - 3.4	噘嘴/咧嘴时嘴角与唇突的关联
	唇-颌	3.1 - 4.1	嘴张开闭合带动下颌骨

由于人脸对称性, 左右关联的表情动作往往具有同样的幅度和状态变化; 而主次关联, 则是由主要表情动作引发次要表情动作, 状态参数之间存在主次关系。

对于左右关联的状态参数 P_L 和 P_R , 我们令其具有相同的幅度变化: 即若局部表情从状态 1 变化到状态 2, 设 P_L^1 (P_R^1)为 P_L (P_R) 在表情状态 1 下的数值, P_L^2 (P_R^2)为 P_L (P_R)在表情状态

2 下的数值, 那么两个参数幅度变化保持一致:

$$|\Delta P_L| = |\Delta P_R| \quad (\Delta P_L = P_L^1 - P_L^2, \Delta P_R = P_R^1 - P_R^2) \quad (3)$$

对于主次关联参数 P_{main} 和 P_{slave} , 我们令其幅度变化满足线性关系,

$$|\Delta P_{slave}| = \alpha \cdot |\Delta P_{main}| \quad (\alpha \in [0, 1], \Delta P_{main} = P_{main}^1 - P_{main}^2, \Delta P_{slave} = P_{slave}^1 - P_{slave}^2) \quad (4)$$

其中 α 是两个关联区域关键 FAP 的相关系数。根据关联规则(3)(4)合成的局部表情

b	喜悦	0.78	0	0	0	0	0	-0.02	0.74	-0.34	0.34	0	0
c	厌恶	-0.74	0.9	0	0	0	0.74	0.08	-0.34	0	0	-0.04	0
d	愤怒	0	1.0	-0.76	0	0	0	0.36	-0.34	0	0	-0.44	0
e	悲伤	0	0.28	0.88	0	0	0	0.14	-0.82	0	0	-0.06	0
f	恐惧	0	1.0	0.8	0.8	0	0	0.28	-0.88	0	0	0.34	0
g	惊喜	0.74	0	0	0	0	0	0.88	0.62	0	0.18	-0.04	0
h	焦虑	0	0.92	0.28	-0.1	0	0.28	-0.74	-0.14	0	-0.74	0	0

5. 结论及展望

本文提出了一种基于局部表情参数化的表情脸像合成方法。在 MPEG4 人脸动画参数(FAP)的基础上, 定义了局部表情状态参数(PEP), 建立了 PEP 和 FAP 之间的关联规则, 利用 PEP 更为有效地合成人脸表情。同时实现了一个三维表情脸像编辑器, 合成结果可以应用于可视语音等人机交互场景。PEP 搭建了底层 FAP 与表情状态之间的桥梁。在合成表情时, 只需关注面部关键区域(器官), 设定对应的 PEP 参数值, 而不必关心底层 FAP 参数。

为了实现更为有效的表情合成, 需要进一步探索局部表情状态参数与高层表情语义之间的关联。

参考文献

- [1] Cosatto, E. Ostermann, J.Graf, H.P. Schroeter, J. "Lifelike talking faces for interactive services". In: PROCEEDINGS OF THE IEEE, VOL. 91, NO. 9, SEPTEMBER 2003 pp 1406~1429
- [2] Gabriel Antunes Abrantes, Fernando Pereira."MPEG-4 Facial Animation Technology: Survey, Implementation, and Results". In: IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 9, NO. 2, MARCH 1999
- [3] A.Raouzaiou, N. Tsapatsoulis, K. Karpouzis and S. Kollias. "Parameterized facial expression synthesis based on MPEG-4". In: EURASIP Journal on Applied Signal Processing 2002:10, pp1021~1038
- [4] Fabio Lavagetto and Roberto Pockaj. "An Efficient Use of MPEG-4 FAP Interpolation for Facial Animation at 70 bits/Frame". In: IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 11, NO. 10, OCTOBER 2001 :1085-1097
- [5] Koray Balci. "XFace: Open Source Toolkit for Creating 3D Faces of an Embodied Conversational Agent". <http://xface.itc.it>, 2004
- [6] 王志明. "汉语视位建模及可视语音的研究". 清华大学工学博士学位论文, 2003
- [7] Zhiyong Wu, Shen Zhang, Lianhong Cai, Helen M. Meng, "Real-time Synthesis of Chinese Visual Speech and Facial Expressions using MPEG-4 FAP Features in a Three-dimensional Avatar", In: Proc. Int. Conf. on Spoken Language Processing (ICSLP), 2006:9, to Appear

Region-based Facial Expression Synthesis on a Three-dimensional Avatar

Shen Zhang¹⁺, Zhiyong Wu², Lianhong Cai¹

¹(Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, China)

²(Department of Systems Engineering and Engineering Management,

The Chinese University of HongKong, Shatin, N.T., HongKong SAR, China)

⁺ Corresponding author: Phn: +86-10-6277-1587, E-mail: zhangshen05@mails.tsinghua.edu.cn

Key words: Face Region, Facial Animation Parameter, MPEG4

Abstract: Facial Expression is one of the most important features in Human-Computer Interaction. MPEG4 Facial Animation Framework provides Facial Animation Parameter (FAP) for parameterized facial expression synthesis. However the implementation of FAP is still an open area. This paper presents a region-based method for fast-generating facial expressions on a three-dimensional avatar, which can be used as talking head in Text to Visual Speech System. A Facial Expression Editor is also implemented for user to generate complex expressions using high-level parameters proposed in this paper.

MPEG4 provides an efficient way to animate face model using FAP, but the low-level FAPs just define the basic facial actions corresponding to feature points; on the other side the high-level FAPs only describe expression and viseme state semantically. Thus it is still a complicated work to generate expressions using FAP directly. In order to fill the semantic gaps, we proposed the region based method: First, the face is divided into regions and 4 key regions are selected which play important role in human expressions. In each key region, Partial Expression Parameter (PEP) is defined, which describes the emotional state of the corresponding region. Based on FAP, PEP considers the correlation of FAPs in generating expressions. A relation between PEP and corresponding FAPs is established by a linear function. There are totally 12 PEPs defined to generate expressions instead of using 40 or more FAPs. Considering the symmetric character of human face and the inter-dependences of face region, this paper proposes “left-right” and “primary-secondary” relation between PEPs, which makes the facial expression more lively and comfortable.

A facial expression editor is also implemented using the region-based techniques proposed in this paper. The input of the editor is an original 3D avatar (in vml form), user can define FDP and FAP on the model manually, and set the 12 PEPs value to get partial facial expression automatically. Each synthesized facial expression corresponds to FAP vector, so that the result expression can be utilized in many scenarios of Human-Computer Interaction, such as dialog system.