维普资讯 http://www.cqvip.com

MPEG-4人脸动画技术和一个基于MPEG-4的人脸动画系统的设计*)

MPEG-4 Facial Animation Techniques and the Design of a MPEG-4 based Facial Animation System

王 淘[1,2] 董兰芳[1] 陈国良[1,2] 许胤龙[1,2]

(中国科学技术大学计算机科学技术系 合肥230027)[1](国家高性能计算中心 合肥230027)[2]

Abstract Being the first object-based multimedia compression standard in the world, MPEG-4 defines facial animation tools. This paper presents MPEG-4 facial animation techniques, including neutral state facial model, facial animation parameters and facial definition parameters. It also presents the design of a MPEG-4 based facial animation system we are realizing and the main problems we must solve.

Keywords MPEG-4, Facial animation, Facial animation parameters, Facial definition parameters

1 引言

运动图像专家组(Moving Picture Experts Group, MPEG)是国际标准化组织ISO下属的专门负责制定运动图像、音频及其组合的压缩、解压缩、处理以及编码表示的标准的工作组[1]。自1988年成立以来,MPEG相继制定了一系列国际标准,其中MPEG-1成为VCD和MP3产品的基础,MPEG-2成为DVD和数字电视产品的基础。MPEG-4版本1在1998年10月举行的MPEG会议上通过,并于1999年年初经ISO批准正式成为国际标准[2~4]。MPEG-4版本2在1999年12月举行的MPEG会议上通过,并于2000年年初正式得到ISO的批准[1]。

MPEG-4是世界上第一个基于对象的多媒体压缩标准[2],该标准可以对各种自然的或合成的音频/可视对象(audio-visual objects, AVO)分别独立编码,将它们有机地集成在同一个场景中。由于人脸动画在电影与广告、视频会议、可视电话、人机接口、虚拟环境、游戏娱乐等诸多领域中占据着极其重要的位置,因此MPEG-4版本1定义了专门的人脸动画(facial animation)工具,MPEG-4版本2更是进一步定义了专门的人体动画(body animation)工具,从而构成一套标准的完整的虚拟人物工具。

我们人脸动画课题组从1999年9月起开展人脸动画的研究工作,对国内外相关资料进行了深入细致的调研,设计并实现了一个基于弹性网格和肌肉模型的人脸动画系统,目前正在实现一个基于MPEG-4的人脸动画系统。

本文介绍MPEG-4人脸动画技术,包括中性人脸模型、人脸动画参数(facial animation parameters, FAP)和人脸定义参数(facial definition parameters, FDP);并介绍我们正在实现的一个基于MPEG-4的人脸动画系统的设计思想和需要解决的主要问题;最后是结束语。

2 MPEG-4人脸动画技术

2.1 中性人脸模型

如图1所示,MPEG-4定义了具有如下性质的中性人脸模型:目光注视着Z轴方向;所有脸部肌肉处于放松状态;眼睑与虹膜相切;瞳孔直径为虹膜直径的三分之一;上下嘴唇接触,唇线处于水平状态并且与两边嘴角处于同一高度;嘴巴关闭,上下牙齿接触;舌头平直,处于水平状态,并且舌尖接触上下牙齿的交界处。

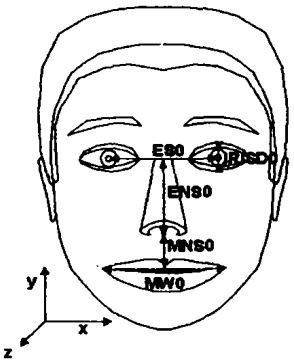

图1 中性人脸模型

表1 人脸动画参数单位

IRISD0	虹膜直径	IRISD=IRISD0/1024
ES0	两眼之间距离	ES=ES0/1024
ENS0	眼睛与鼻子之间距离	ENS=ENS0/1024
MNS0	嘴巴与鼻子之间距离	MNS=MNS0/1024
MW0	嘴巴宽度	MW=MW0/1024
AU	角度单位	10E-5弧度

为了能够在任意人脸模型上定义人脸动画参数,MPEG-4定义了人脸动画参数单位(facial animation parameter units, FAPU)。FAPU被定义为关键脸部特征的分数(fraction)(见图1),这些关键脸部特征(如两眼之间的距离)是在中性人脸模型上定义的。FAPU使得人脸动画参数在任意人脸模型上具有一致的解释,能够在任意人脸模型上产生合理的表情与口型。表1为MPEG-4定义的人脸动画参数单位。

如图2所示,MPEG-4在中性人脸模型上定义了84个特征点(feature point),这些特征点的主要作用是为定义人脸动画参数提供空间参考。有一些特征点(如11.1、11.2、11.3等)并不会受到人脸动画参数的影响,但是它们在校准私有人脸模

*)本课题得到中国科学技术大学青年基金资助。王 淘 副教授,主要研究计算机图形学、并行与分布计算。董兰芳 讲师,主要研究计算机图形学、软件体系结构。陈国良 教授、博导,主要研究并行/分布/网络计算。许胤龙 副教授,主要研究并行与分布计算。

维普资讯 http://www.cqvip.com

型(proprietary face model)的时候需要用到(见2.3节)。这84个特征点被分为若干组(如脸颊、眼睛、嘴巴等),所有符合MPEG-4的人脸模型都应当了解这些特征点的位置,并且一般说来,相关特征点的位置之间存在着一定的约束关系(表2列出了其中一部分,更详细的描述请参阅文[3])。

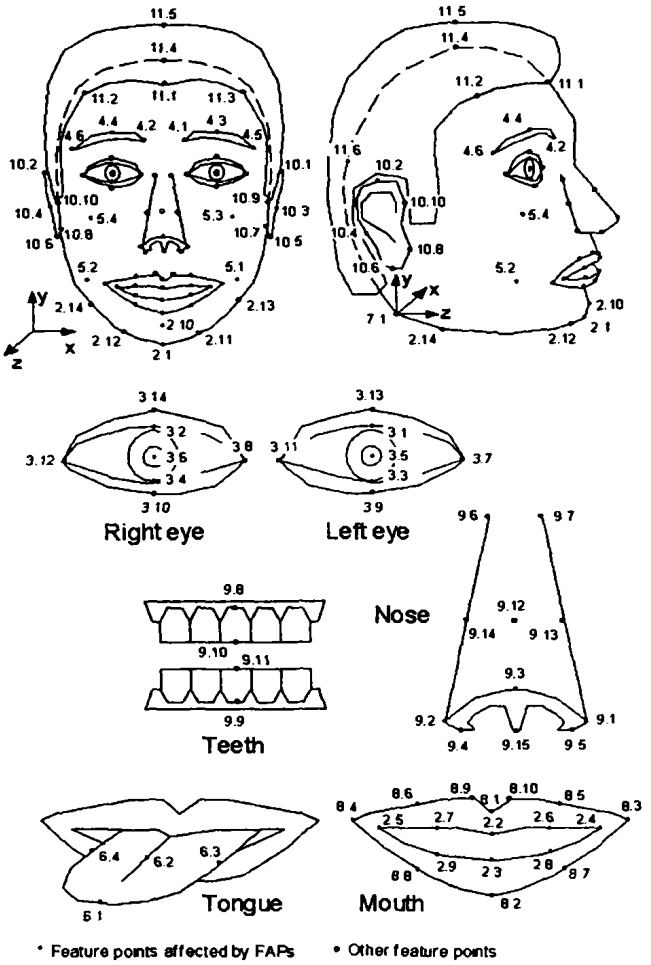

图2 特征点

表2 相关特征点的位置之间的约束关系

特征点		位置之间的约束关系		
号	文字描述	X	Y	Z
2.1	下巴底端点	7.1.x		
2.2	上嘴唇内侧中点	7.1.x		
2.3	下嘴唇内侧中点	7.1.x		
2.4	嘴唇内侧左角点			
2.5	嘴唇内侧右角点			
2.6	上嘴唇内侧2.2与2.4的中点	(2.2.x+2.4.x)/2		
2.7	上嘴唇内侧2.2与2.5的中点	(2.2.x+2.5.x)/2		
2.8	下嘴唇内侧2.3与2.4的中点	(2.3.x+2.4.x)/2		
2.9	下嘴唇内侧2.3与2.5的中点	(2.3.x+2.5.x)/2		
2.10	下巴突出点	7.1.x		
2.11	下巴左角点	>8.7.x且<8.3.x		
2.12	下巴右角点	>8.4.x且<8.8.x		
2.13	颚骨左角点			
2.14	颚骨右角点			

2.2 人脸动画参数

MPEG-4的人脸动画参数基于对最小可感知动作(Minimal Perceptible Actions,MPA)的研究,MPA与脸部肌肉动作密切相关。人脸动画参数定义了一个完整的基本脸部动作集合(包括头部、舌头、眼睛和嘴巴的动作等),人脸动画参数也可以用于定义Ekman和Friesen提出的脸部动作编码系统(Facial Action Coding System,FACS)[5]中的动作单元(Action Unit,AU)。人脸动画参数既可以用来表示人类的自然的脸部动作,也可以用来表示卡通角色的夸张的脸部动作。

表3 FAP组

组	FAP数	组	FAP数
1:可视音素和表情	2	6:舌头	5
2:颚、下巴、嘴唇内侧、嘴角	16	7:头部旋转	3
3:眼球、瞳孔、眼睑	12	8:嘴唇外侧	10
4:眉毛	8	9:鼻子	4
5:脸颊	4	10:耳朵	4

表4 可视音素及相关音素

可视音素号	音素	例子
0	none	Na
1	p, b, m	put, bed, mill
2	f, v	far, voice
3	T,D	think, that
4	t, d	tip, doll
5	k, g	call, gas
6	tS, dZ, S	chair, join, she
7	s, z	sir, zeal
8	n, l	lot, not
9	r	red
10	A:	car
11	e	bed
12	I	tip
13	Q	top
14	U	Book

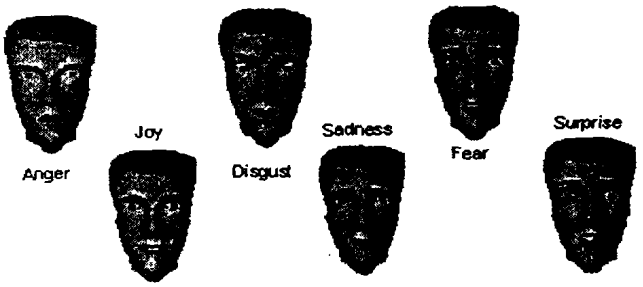

图3 脸部表情

MPEG-4共定义了68个人脸动画参数,这68个人脸动画参数被分为10组(见表3)。组1(FAP1与FAP2)定义了2个高层人脸动画参数,FAP1包括14个可以明显区分开的静态可视音素(static visemes)(见表4),用于生成语音口型同步动画。FAP2包括喜悦(joy)、悲伤(sadness)、愤怒(anger)、恐惧(fear)、厌恶(disgust)、惊讶(surprise)共6种常用脸部表情(见图3)。组2-10(FAP3-68)定义了66个底层人脸动画参数(如张颚、放低上嘴唇内侧中点等)。人脸动画参数的完整描述请参阅文[3]。

2.3 人脸定义参数

2.3.1 私有人脸模型 每一个能够接受人脸动画参数并产生人脸动画的MPEG-4解码器应当具有符合MPEG-4的人脸模型,该模型通常是一个私有人脸模型,编码器并不了解

维普资讯 http://www.cqvip.com

该模型的构造和形状。可以通过以下四个步骤建立一个私有人脸模型：

1. 建立中性人脸模型，根据2.1节和图2在中性人脸模型上定义特征点的位置。

2. 对于每一个底层人脸动画参数，定义特征点是如何移动的。MPEG-4在人脸动画参数的功能描述中只给出了特征点在某一维方向上的运动。例如，FAP54只给出了嘴唇外侧右角点在水平方向上的运动，而嘴唇外侧右角点在向右移动的同时通常也向后移动。因此，MPEG-4解码器需要考虑人脸动作的这些特点，才能得到比较真实自然的效果。

3. 对于每一个底层人脸动画参数，定义了特征点的移动之后，还需要定义特征点周围的顶点是如何移动的。这种从特征点的移动到其周围顶点的移动的映射可以使用基于肌肉的变形[6~8]、散列点插值[9~11]或人脸动画表(facial animation table, FAT)[12]来实现。

4. 定义高层人脸动画参数，即14个静态可视音素(FAP1)和6种常用脸部表情(FAP2)。

2.3.2 *私有人脸模型校准* MPEG-4编码器可以给出中性人脸模型中全部或部分特征点的位置，这就要求解码器能够对私有人脸模型进行变形，使得私有人脸模型的特征点与编码器给出的特征点重合。该过程称为私有人脸模型校准。

除了特征点之外，MPEG-4编码器还可以给出纹理图和特征点的二维纹理坐标，解码器可以利用纹理映射方法显著地提高人脸动画的真实感。

2.3.3 *场景图* 私有人脸模型校准只能保证私有人脸模型的特征点与编码器给出的特征点重合，私有人脸模型的其余顶点如何变化是由解码器确定的。如果编码器希望能够完全控制解码器端的人脸模型，那么就可以使用场景图下载一个完整的人脸模型到解码器端。

MPEG-4编码器使用场景二进制格式(Binary Format for Scenes, BIFS)来描述人脸模型，BIFS提供了类似于VRML[13,14]的节点。BIFS和VRML用一系列节点来描述场景，这些节点按照场景图进行组织。在定义人脸模型时使用到以下三种节点：组节点是包含一系列子节点的容器，利用组节点可以建立起场景的层次结构；作为一组共同移动的对象应当包含在同一个变换组中，变换节点定义了施加在其子节点上的三维仿射变换(如平移、缩放、旋转等)，当变换节点中又包含其它变换时，其变换效果是累积的，利用嵌套的变换节点可以建立起变换的层次结构；索引人脸集合(IndexedFaceSet)节点定义了多边形对象的几何形状(三维网格)和表面属性(颜色、纹理)，纹理图使用MPEG-4静止图像编码器中的小波编码器进行编码[3]。

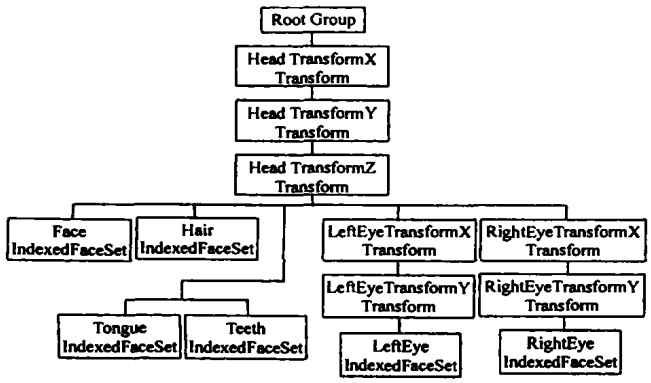

图4 简化的人脸模型的场景图

图4表示简化的人脸模型的场景图，嵌套的变换节点依次表示绕着x轴、y轴和z轴的旋转，左眼和右眼的旋转嵌套在头部的整体运动中。不同的索引人脸集合分别定义了脸部、头发、舌头、牙齿、左眼和右眼的几何形状和表面属性，因而可以使用不同的纹理图。

除了以上介绍的中性人脸模型、人脸动画参数和人脸定义参数之外，MPEG-4在人脸动画方面还定义了人脸动画表、FAP算术编码器(FAP arithmetic coder)、FAP离散余弦变换编码器(FAP discrete cosine transform coder)、FAP插值表(FAP interpolation table, FIT)、文本-语音合成器接口(text-to-speech interface, TTSI)、能力测验图(profile)等，限于篇幅，本文对这些内容不一一介绍，有兴趣的读者可以参阅文[2~4]。

3 一个基于MPEG-4的人脸动画系统的设计

MPEG-4只给出了人脸动画的功能描述，并没有规定人脸动画的具体实现方法，这就使得我们可以设计一个基于MPEG-4的人脸动画系统，在该系统中研究人脸动画技术。

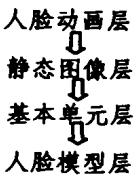

图5 人脸动画系统的四层结构

我们认为，一个人脸动画系统从逻辑结构上可以分为四层(见图5)，每一层通过下一层实现其功能。下面我们以"笑"为例来说明每一层的功能以及层与层之间的关系：人脸动画层为任务级的抽象描述，在该层只需要给出"笑"的强度(大笑还是微笑?)、开始时间和结束时间。由于一个完整的脸部动作过程通常分为开始(onset)、增强(climax)和复原(relax)等三个阶段[15]，因此，"笑"的动画可以由一系列强度不断变化的反映"喜悦"的静态图像构成。静态图像层是人脸动画过程中某一个时刻的图像，"喜悦"可以通过"眉毛放松"、"嘴巴张开"、"嘴角向耳朵方向拉"等一系列基本单元来实现。基本单元层描述了脸部的一系列基本动作单元，可以通过对人脸模型进行变形来实现。人脸模型层反映了人脸的几何形状和表面属性，一般用多边形网格和纹理图来表示。

在我们提出的人脸动画系统的四层结构中，人脸模型层可以采用MPEG-4的人脸定义参数；基本单元层可以采用MPEG-4的66个底层人脸动画参数；静态图像层除了采用MPEG-4的2个高层人脸动画参数(包括14个静态可视音素和6种常用脸部表情)之外，还需要根据实际应用领域定义一些静态可视音素(如中文可视音素)和脸部表情(如复合表情)；由于人脸动画层在MPEG-4中没有对应的功能描述，因此我们也需要根据实际应用领域定义一些常用的人脸动画任务，使得用户可以在较高的层次上使用人脸动画系统，无需关心其实现细节。

在实现方法上需要解决以下几个主要问题：1)如何采用纹理映射的方法增强人脸动画的真实感?2)如何通过对人脸模型进行变形来实现基本动作单元?3)如何通过基本动作单元来实现高层人脸动画?4)当多个高层人脸动画同时影响某一个基本动作单元时，如何解决冲突才能得到真实的、自然的

维普资讯 http://www.cqvip.com

动画效果?

结束语 作为世界上第一个基于对象的多媒体压缩标准,MPEG-4首次定义了专门的人脸动画工具,使得人脸动画可以更加广泛地应用在电影与广告、视频会议、可视电话、人机接口、虚拟环境、游戏娱乐等诸多领域中。由于MPEG-4只给出了人脸动画的功能描述,并没有规定人脸动画的具体实现方法,这就使得我们可以设计一个基于MPEG-4的人脸动画系统,在该系统中研究人脸动画技术。除了人脸动画技术之外,进一步的研究包括基于MPEG-4的人体动画技术,最终形成一套完整的虚拟人物系统。

参 考 文 献

1 The MPEG Home Page. http://www.cselt.it/mpeg/

2 ISO/IEC IS 14496-1 Systems. 1999

3 ISO/IEC IS 14496-2 Visual. 1999

4 ISO/IEC IS 14496-3 Audio, 1999

5 Ekman P,Friesen W V. Facial Action Coding System. Consulting Psychologists Press, Palo Alto, CA, 1978

6 Kala P,et al. Simulation of facial muscle actions based on rational free form deformations, Proc. Eurographics'92, 1992. 59～69

7 Waters K. A muscle model for animating three-dimensional facial expression. Computer Graphics (Proc. SIGGRAPH'87),1987,21(4):17～24

8 Lee Y C, Terzopoulos D, Waters K. Realistic modeling for facial animation. In:Proc. SIGGRAPH'95, 1995. 55～62

9 Lavagetto F, Pockaj R. The facial animation engine: toward a high-level interface for the design of MPEG-4 compliant animated faces. IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY,1999,9(2):277～289

10 Lavagetto F,Pockaj R,Costa M. Smooth surface interpolation and texture adaptation for MPEG-4 compliant calibration of 3D head models. Image and Vision Computing,2000,18:345～353

11 Pighin F,Hecker J,Lischinski D,Szeliski R,Salesin D H. Synthesizing realistic facial expressions from photographs. In: Proc. SIGGRAPH'98, 1998. 75～84

12 Ostermann J,Haratsch E. An animation definition interface: Rapid design of MPEG-4 compliant animated faces and bodies. In:Proc. Intl. Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging, 1997. 216～219

13 Hartman J,Wernecke J. The VRML Handbook, Addison-Wesley, Reading, MA, 1996

14 ISO/IEC 14772-1. Information Technology-Computer Graphics and Image Processing-The Virtual Reality Modeling Language-Part 1: Functional specification and UTF-8 encoding, 1997

15 Murakami S,Tadokoro H. Generation of facial expressions based on time functions. In: Proc. Intl. Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging,1997. 212～215

(上接第74页)

OC上使用只读的传统的文件系统来保存所占比重很大的那部分不变的信息。而另外划分一块相对较小的区域来使用JFFS存储经常修改的数据。这样做的好处是:对JFFS来说,在一定程度上,文件系统的容量越小,它的效率会越高。而相对较大的使用传统文件系统的区域,由于它的属性是只读的,非正常的断电也不可能对它造成破坏。这样组合起来使用系统的效率就会提高很多。

但是对于DOC来说,我们如果要使用传统的文件系统,必须使用NFTL层,NFTL需要使用整个硬件设备,而JFFS也需要使用一个硬件设备。DOC驱动在检测DOC硬件时,一块DOC只是当作一个逻辑设备加入到系统。那么,我们如何才能让JFFS和正常文件系统在一块DOC上共存呢?

上述问题的解决办法是:在原理上也不复杂,我们可以修改DOC的驱动,使DOC驱动在检测硬件时,把一块DOC作为两块逻辑设备加入到系统中来,同时修改驱动接口函数,让这两个逻辑设备对上层来说,好像是一个独立的设备,这样我们就能够在其中一块上通过NFTL来使用正常的文件系统,而在另外一块上使用JFFS。鉴于篇幅,我们略去具体的实现。

4.4 DiskOnChip上的JFFS和正常文件系统的性能比较

①通常的性能比较:经过我们的测试,DOC上的JFFS和DOC上的传统文件系统在读写速度上它们没明显的差异。但是当垃圾清除的时间到来时,JFFS会锁住整块设备,会有数秒的等待时间。只是垃圾清除动作发生的频度不高,整体上对JFFS系统的系统没有太大的影响。

②防断电性能比较:DOC上的传统文件系统只要有一次非正常断电就会报告文件系统错误。如果不经过修复,连续非正常断电数次就可能导致文件系统的严重破坏。即便是每次断电以后都进行修复,连续非正常的断电后的次数大于10次以上就肯定会导致文件系统的严重破坏。

DOC上的JFFS在经历了几天断电试验,非正常断电几百次后系统还能够保持完好。

从上面可以看出,JFFS在防断电方面有优良的性能,而且相对于传统的文件系统来说,性能并没有明显的降低,是嵌入式存储设备中用来存放经常修改的数据的理想选择。

小结 JFFS是一个正在发展的文件系统。它能够在嵌入式存储设备上做到防止非正常断电导致的文件系统破坏和数据丢失。但是目前它还有很多的不足,比如说垃圾清除耗时过长,系统还不是足够稳定等等。在DOC上JFFS的实现也仅仅是刚刚完成,还需要长时间的测试和错误修复。不过随着时间的推移,JFFS也会日益成熟,也必将会在嵌入式领域广泛使用。

参 考 文 献

1 http://www.linux-mtd.infradead.org/ (MTD主页)

2 http://oss.software.ibm.com/developerworks/opensource/jfs/ (JFS主页)

3 http://www.freebsd.org/docs/en/books/design-44bsd/ (4.4BSD的网上技术文档)

4 Leffler S J,McKusick M K,Quarterman J S. The Design and Implementation of the 4.3BSD Unix Operating System, Addison-Wesley, Reading, MA, 1989