

MPEG - 4 标准视频编码初论

葛双全, 席传裕

(中国工程物理研究院计算机应用研究所 四川 绵阳 621900)

摘 要: MPEG - 4 视频部分(ISO/IEC 14496 - 3)是 MPEG - 4 标准核心内容之一, 它既提供了传统的基于帧的编码方法也提供了基于视频对象的编码方法, 为高效压缩与基于内容的交互提供了坚实的基础。文章对 MPEG - 4 标准视频编码部分做了详细的描述, 并对 MPEG - 4 视频的应用做了简要介绍。

关键词: MPEG - 4 视频; 视频对象; 基于对象

Abstract: MPEG - 4 video part(ISO/IEC 14496 - 3) is one of MPEG - 4 standard's core contents. It not only supplies the traditional frame - based coding way but also supplies coding way based on video object. Coding based on video object supplies strong base of highly efficient compression and content - based interaction. This article gives fully description on MPEG - 4 standard's video coding part and gives brief introduction to the application of MPEG - 4 video in multimedia.

Key words: MPEG - 4 video; video object; object - based

0 引言

MPEG - 4 是由运动图象专家组 MPEG (Moving Picture Experts Group) 建议的 ISO/IEC 标准。1998 年 10 月 ISO/IEC 公布了 MPEG - 4 标准草案, 并于 1999 年初成为国际标准, 即 ISO/IEC 14496。数年的发展表明, MPEG - 4 在数字电视、交互图象应用和交互多媒体方面等方面有着广泛应用。作为一个全新的编码标准, MPEG - 4 为数字化的视听数据的通信、访问和操作提供了新的方法, 而且还提供了一个灵活的框架和一套开放的工具来支持新的和传统的性能。

MPEG - 4 与 MPEG - 1、MPEG - 2 以及 H. 263 等标准最大的区别在于它是基于对象的编码方式, 以及它能对合成对象进行编码的能力。MPEG - 1、MPEG - 2 以及 H. 263 的压缩编码是基于帧的, 所采用的算

法在信源模型和块匹配等方面存在不足^[1], 与用户的交互以帧为单位。而 MPEG - 4 不再局限于传统的矩形视频对象, 而是可以对任意形状的视频图像编码, 因此为更有效的压缩、存储和传输数字视音频以及用户与视音频的交互提供了可能。

1 MPEG - 4 视频编码组织结构

从上世纪 90 年代前后至 90 年代末, 国际上先后制订了 H. 261、H. 263 以及 MPEG - 1、MPEG - 2 等众多国际标准, 覆盖了从低码率需求的可视电话、可视会议到高码率需求的数字演播室、高清晰度电视等应用领域。而 MPEG - 4 标准第 2 版本的带宽几乎囊括了以前制订的所有国际标准的带宽 (如图 1 所示), 也就是说, MPEG - 4 均可应用于上述领域。

收稿日期: 2002 - 10 - 15

作者简介: 葛双全(1975 -), 研究方向为多媒体信息处理; 席传裕(1968 -), 研究方向为计算机网络。

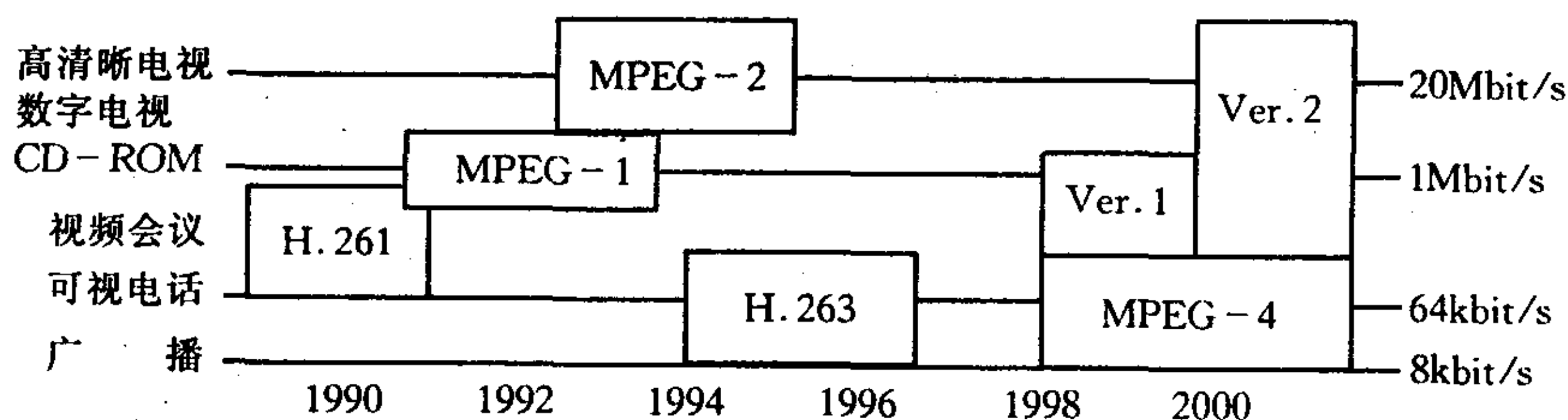


图1 视频压缩标准及其应用

由于 MPEG-4 的应用范围如此广泛,不同的应用的要求又有很大的差别,为此在 MPEG-4 标准中通过定义类(Profile)、级(Level)、算法(Algorithm)和工具(Tool)四层结构来描述 MPEG-4 在某一具体应用中的编码方案。工具被定义为通过 MSDL (MPEG-4 Syntactic Description Language) 得到的一种方法,例如如何运动估计或轮廓描述, MPEG-4 为编码视音频对象提供了大量有效的工具。算法则被定义为可提供一个或多个功能的工具集合,例如 MPEG-1 音频, MPEG-1 视频或 MPEG-2 系统等等。类被定义为解决特定编码或功能需求的标准方法,用于确定 MPEG-4 元素(工具和算法)如何配置、组合为一个压缩器或解压器,以完成特定的应用。MPEG-4 包含视频、音频、图形、场景描述、MPEG-J 与对象描述六大类,每个类下面又分为一个或多个级,这种分级的方法与 MPEG-2 类似。

采用这样的分层结构有许多好处,标准并不规定一个具体的编码和解码过程,在具体应用和中只需根据标准的规定组织该应用的层,搭配哪些工具和算法就能运用到某一具体的编码应用,并且所产生的码流能被其他的应用所回放。其次,随着技术的发展,许多新的有效的编码工具可以方便地以算法或工具的形式添加到标准中。

2 MPEG-4 视频编码算法结构

2.1 MPEG-4 视频数据结构

2.1.1 视频对象

视频对象 VO (Video Object) 是可视场

景中景物的抽象描述,从用户的角度,它代表画面中任何有意义的物理实体。视频对象是 MPEG-4 中编码的独立单位。场景由一个或多个视频对象组成,每个视频对象的形状、运动和纹理用其相关的时间和空间信息给出。VO 的生存期为一个片段 (Session)。MPEG-4 采用视频对象来表示图像内容,可以组合已有的视频对象来生成复合的视频对象,并由此生成视频场景^[2],允许对视频对象的数据灵活地多路合成与同步,以便选择合适的网络来传输这些对象数据。在接收端允许用户在场景中对视频对象进行交互操作。

2.1.2 视频对象平面

MPEG-4 引入了视频对象平面 VOP (Video Object Plane) 概念。假设输入的视频序列的每一帧都被分割成多个任意形状的 VOP (在 MPEG-1、MPEG-2 及 H.263 中,被处理的图像总是矩形),每个 VOP 定义场景中特定的视频内容。各个 VOP 的形状和位置可随帧变化。属于场景中同一实际物体连续的有着任意形状任意位置的 VOP 序列就是视频对象 VO。

VOP 是 MPEG-4 中编码的基本单位。每个 VOP 可以独立地编码,也可以使用运动补偿技术相互依赖地编码。传统的视频帧也可以用矩形 VOP 表示。VOP 包含了视频对象的运动参数、形状信息和纹理等数据。VOP 既是一个空间概念也是一个时间概念。从另一个角度说,VOP 是视频对象 VO 在特定时刻的取样。因此,属于场景中同一个物理对象的连续的 VOP 就代表了一个视频对象 VO。从编码的角度来看,VO 实际上是由一

组同一实体的任意形状和位置的 VOP 序列组成的。VOP 包括主体对象、背景对象以及文字图形三类。

2.1.3 视频对象层

属于同一 VO 的 VOP 的形状(shape)、运动(motion)和纹理(texture)信息可被编码为一个单独的视频对象层(VOL, Video Object Layer)。对同一个 VO,可以用不同的空间或时间分辨率编码为多层结构,我们能从一个基础层开始,用增加一些增强层的方法,以分层的方式重建视频。每个视频对象可以编码成可伸缩的(多层)或不可伸缩的(单层)。用哪一种方式编码取决于应用。

2.1.4 场景

场景是一个或多个声视频对象的组合。场景的逻辑结构可以用一棵树表示,树中的节点是声视频对象。MPEG-4 系统用二进制场景格式 BIFS 描述场景中声视频对象的空间和时间位置及它们之间的关系。MPEG-4 的视频比特流提供了对场景的分层描述。在比特流中,表示场景的层是可视对象序列 VS(Video Object Sequence),它是一个完整的 MPEG-4 场景,其中可能包含自然对象或合成的对象以及它们的增强层。

2.2 VOP 的形状、运动和纹理信息编码

MPEG-4 视频 VOP 基本的编码结构包括形状编码(对任意形状的视频对象)、运动补偿以及基于 DCT 的纹理编码(使用标准 8×8 DCT 或形状自适应 DCT)。编码器首先对输入的任意形状的 VOP 序列用基于块的混合编码技术编码。先帧内 VOP,后帧间 VOP 和双向预测 VOP。在对 VOP 的形态状信息编码之后,取得任意形状 VOP 的采样,每个 VOP 划分成各不相交的宏块,每个宏块含有四个 8×8 的像素块,以进行运动估计和补偿及纹理编码。已编码的 VOP 帧保存在 VOP 帧存储器中。运动矢量在当前 VOP 帧和已编码 VOP 帧之间计算。对将被编码的块或宏块,计算其运动补偿预测误差。运动补偿预测后的 VOP 及误差用 8×8 块 DCT

编码,并进行 DC 系数的量化,然后进行行程编码和熵编码。最后,形状、运动和纹理信息合成符合 MPEG-4 标准的比特流输出^[3]。纹理编码、运动估计与补偿基本采用传统的编码方法,而形状编码是首次引入的视频编码技术。

对输入的矩形或任意形状的图像序列,图 2 描述了 MPEG-4 进行压缩编码的基本算法。

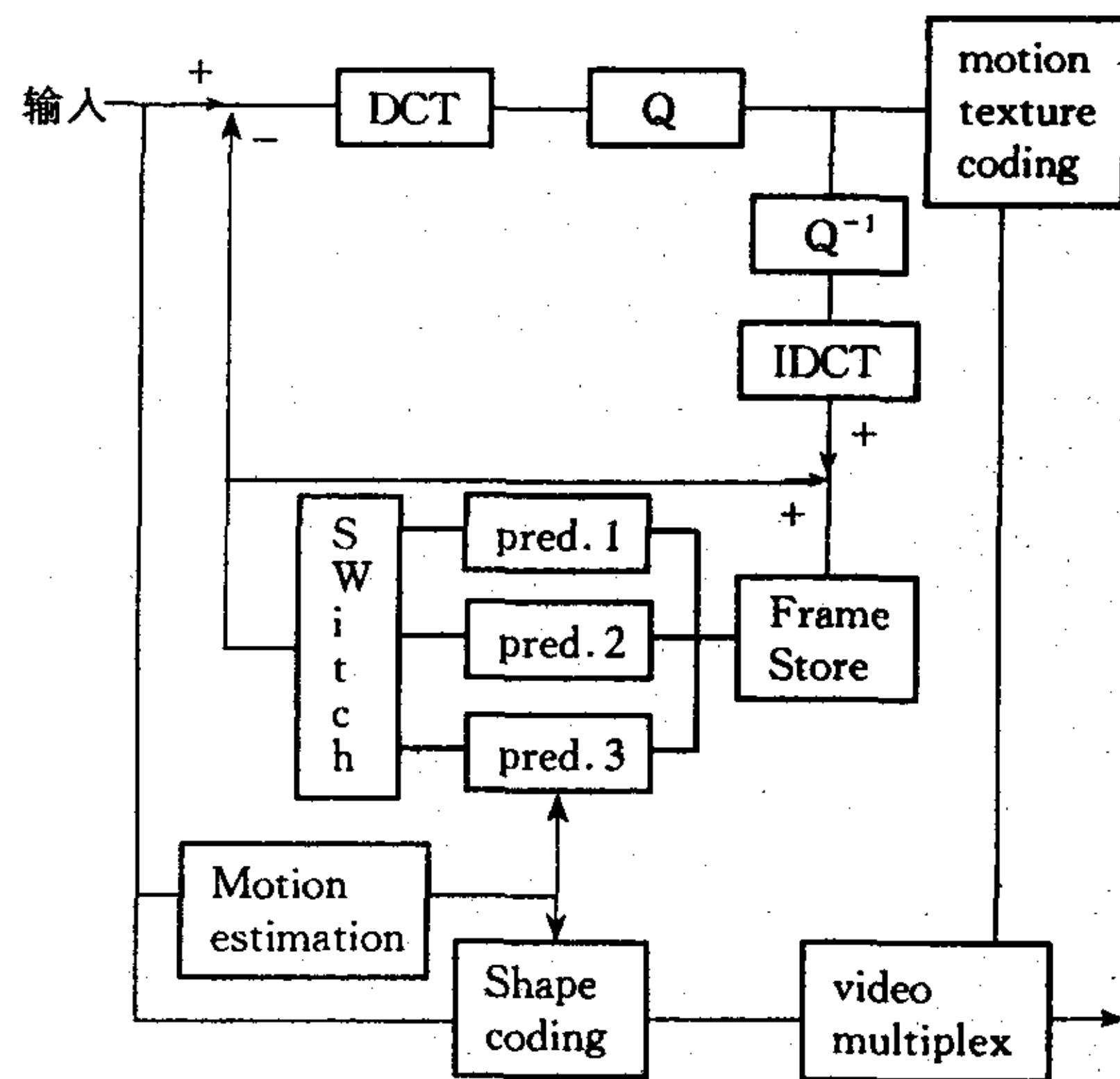


图 2 MPEG-4 编码器基本框图

2.2.1 形状编码

MPEG-4 标准中视频对象 VO 的形状信息有两类:二值形状信息和灰度形状信息。二值形状信息通常表示为与 VOP 的边界框相同大小的矩阵。矩阵的每个元素可以取两种可能的值 0、1,这取决于像素是在视频对象内还是在视频对象外,0 表示该像素在 VOP 区域之外,1 表示在 VOP 区域之内。对二值形状信息进行编码时可以采用基于块的运动补偿技术,可以是无损或有损编码。灰度形状信息与二值形状信息有相同的结构,用 $0 \sim 255$ (8 比特)之间的数值表示该像素的透明度,0 表示完全透明,255 则表示完全不透明,中间值对应用像素的相应的透明度,这与计算机图形学中的阿尔法平面的概念大致相同。编码时采用基于块的运动补偿 DCT 方法,属于有损编码。

2.2.2 运动信息编码

运动估计与运动补偿技术通过对已知图像的一块像素值重新定位,来预测当前图像中相应块的像素值。MPEG-4 与 MPEG-1、MPEG-2 一样利用运动估计与补偿来减少帧间的时间冗余度。主要区别在于后两者的压缩基于块,而 MPEG-4 则是基于任意形状的 VOP。VOP 有 3 种编码模式:帧内编码模式(I-VOP)、帧间编码模式(P-VOP)和帧间双向预测编码模式(B-VOP),I-VOP 与其它标准的 I 帧一样只采用帧内压缩,与其它任何 VOP 无关;P-VOP 和 B-VOP 编码时需要运动估计与运动补偿,P-VOP 基于另一个先被解码的 VOP 作出预测,B-VOP 则基于当前 VOP 的前面和后面的 VOP 作出预测,故 B-VOP 称为双向插值 VOP,是基于 I-VOP 或 P-VOP 的插值帧。

如果一个宏块全部位于一个 VOP 内,则采用通常的基于 16×16 像素宏块或 8×8 像素块运动估计,如果一个宏块只有一部分位于 VOP 内,则需要采用图像填充(Image Padding)技术和多边形匹配(Polygon Matching)技术^[4]。图像填充技术利用 VOP 内部的像素值来外推 VOP 外的像素值,以此获得运动预测的参考值。多边形匹配技术则将 VOP 的轮廓宏块的活跃部分包含在多边形之内,以此来增加运动估值的有效性。这样在 VOP 的边界上就可以搜索预测所需的候选像素了。

2.2.3 纹理编码

纹理编码的对象可以是帧内编码模式的 I-VOP,也可以是帧间编码模式 B-VOP 或 P-VOP 运动补偿后的预测误差。对纹理信息编码时,采用标准的基于 8×8 像素块的 DCT 方法。VOP 内部的 8×8 块可直接编码。跨出边界的块成为边界块,而对于横跨 VOP 边界的块,先采用图像填充技术得到 VOP 之外的像素值,再进行编码。

2.3 魔景(Sprite)编码

MPEG-4 中引入了魔景的概念是基于提高视频的压缩效率的目的。从摄像机得到的图像主要分为为主体和背景两类。主体可以千变万化,背景却是相对静止的。魔景就是由视频对象出现在背景中的那些区域组成,这些区域在整个视频段中都有可能出现,通过图像的镶嵌技术把整个序列的背景图像拼接成一个大的完整的背景图像,这个图像叫做 Sprite 图像。我们只对主体这一可能产生任意动作的视频对象单独编码;而将背景图像一次性地传送给接收端。以后对于同一个镜头中的每一帧图像,就只需要传输 Sprite 信息,接收端可根据这些信息自行生成背景内容;然后将主体叠加上去,就能得到完整的图像。由于整个过程中只需传送一次背景图像,因此 MPEG-4 的编码效率得到了提高。

魔景编码通常适用于合成对象

2.4 MPEG-4 视频编码特点

基于对象的编码是 MPEG-4 视频编码最重要的特点之一,在 MPEG-4 中对象既可以是自然的图像,也可以是合成的各种动画图像^[5],既可以是静止图像也可以是运动图像。MPEG-4 视频标准可以描述组成一幅画的各个任意形状视频对象之间的时间和空间关系。基于对象的编码使解码器能单独解出对象的形状、运动和纹理信息三者之一。由此可见用户就可以随机访问每个视频对象,能对视频对象进行位移、剪贴甚至对视频对象置换或清除等各种操作。

高效压缩是 MPEG-4 压缩编码的又一重要特点。MPEG-4 编码采用分层结构,允许每个视频对象有一个基层和最多四个增强层,在需要时再编码传输增强层来提高传输图像的质量和细节,如果仅仅对用户感兴趣的内容压缩编码而去掉某些无关信息,MPEG-4 视频的压缩效率是非常高的。

MPEG-4 还可实现基于内容的伸缩性,就是指分级编码后,纹理、图像和视频基于内容的伸缩性。视频序列中时间、空间及质量的伸缩性,时间伸缩是指在带宽允许或用

户要求时在基本层之上的增强层中增加码率,在带宽窄或用户不需要时在基本层中减小码率,以达到充分利用带宽,提高压缩效率的目的;空间伸缩是指对基本层中的图像采样插值,增加或减小空间分辨率。

3 MPEG-4 视频应用

MPEG-4 是一种通用的编码标准,能满足各种应用需要,适应于不同的传输带宽、图像尺寸和分辨率,提供不同的服务。MPEG-4 的应用非常广泛,是一种可用于通信、广播和计算机等诸多领域的通用多媒体编码方式,如视音频编辑、多媒体互联网、视频点播(VOD)、交互式视频游戏、视频会议等。

国际上有很多公司已经开发了众多 MPEG-4 应用产品。DivXNetworks 公司根据 ISO MPEG-4 标准开发的压缩解压缩器 DivX 提供基于宽带 IP 网视频点播解决方案。Toshiba 公司已经开发出一种支持 MPEG-4 标准的低功耗芯片,可在 QCIF 屏幕上每秒产生 15 帧图像。Toshiba 还展示过采用 MPEG-4 标准的蜂窝视频电话样机。国内著名的流媒体技术公司上海汉唐科技于

2000 年 6 月成功的开发了 MPEG-4 视频编解码器,代表了中国在国际宽带流媒体技术领域的领先地位。

4 结束语

MPEG-4 从其提出之日起就引起了人们的广泛关注,MPEG-4 的最大创新在于赋予用户针对应用建立系统的能力,而不是仅仅使用面向应用的固定标准。借助于 MPEG-4,我们第一次有可能建立个性化的多媒体系统。我们有理由相信,随着对 MPEG-4 研究的不断深入,越来越多的 MPEG-4 产品将进入人们的生活。

参考文献:

- [1] 吴乐男. 数据压缩. 电子工业出版社,北京,2000.
- [2] Rob Koenen,KPN Research. MOEG-4 Multimedia for our time. IEEE Spectrum February 1999,36(2):
- [3] ISO/IEC 14496-2:2001,Coding of Audio-Visual Objects-Part 2:Visual,Ver. 2,2001.
- [4] Thomas Sikora, The structure of MPEG-4 video coding algorithm, Image Processing Department, Heinrich Hertz Institut Berlin,November 11,1997.
- [5] 杨长生. 图象与声音压缩技术. 浙江大学出版社,杭州,2000.

《电脑与信息技术》投稿须知

- (1)稿件应具有创新性(在国际、国内同行业中处于领先水平)或实用性(紧密结合生产、科研、教学需要),本刊尤其欢迎介绍超前信息或前沿课题的导向性稿件。
- (2)来稿请勿一稿多投,稿件请用打印件(包括程序清单),插图应构图合理、清晰,并有图题(图名),参考文献著录项目须齐全。
- (3)稿件力求短小精悍,简报简讯 500 字左右,论文一般不超过 6 000 字。
- (4)国家重点项目或各类基金资助项目成果请附相关证明的影印件。
- (5)稿件请注明作者详细地址、邮编及电话号码等,以便联系。
- (6)稿件中题目、摘要、关键词请译成英文;文章的第一作者和第二作者须有作者简介。
- (7)稿件超过 10 个月未刊登者,作者可自行处理。由于编辑部人员所限,未录用稿件一律不退稿。