

多媒体视频编码研究^{*}

Research on Multimedia Video Coding

王相海 张福炎

(南京大学计算机软件新技术国家重点实验室 南京大学计算机与科学系 南京 210093)

Abstract More requirements for video coding are arisen from the rapid development of multimedia and Internet. In the paper, an overview of major techniques already used or under investigation for video coding is given, and the system, coding engine and features of the MPEG-4 are introduced. Finally, the future development of video coding is analyzed.

Keywords Video coding, Video object, MPEG-4, Content-based coding

1. 引言

在多媒体技术中,由于视听信息数据量的庞大、通信带宽和存储容量的限制,大量的信息必须进行编码处理。随着 Internet、有线和无线通信系统的迅猛发展,以及视频、音频数据综合服务应用的发展,对计算机多媒体数据编码技术的要求越来越高,不仅要求编码技术具有良好的压缩效果,而且要求它能够具有可分级性、鲁棒性以及能对码流进行直接操作等特性。90年代,国际标准化协会(ISO)、国际电子学委员会(IEC)和国际电信协会(ITU)等国际组织制定了三个重要的多媒体国家标准:JPEG 标准、H. 261 标准和 MPEG 标准,取得了辉煌的业绩,特别是新颁布的 MPEG-4 视频编码标准和正在制定的、即将成为静态图像编码的 JPEG-2000,它们以小波编码为基础,引进了基于对象(内容)等编码模式,明显具有新的时代特征,必将会对下个世纪多媒体技术的发展做出巨大的贡献。

本文对多媒体应用领域中已经使用的和正在研究的一些主要视频编码方法进行了综述,同时对 MPEG-4^[1]的编码机制和编码特性等进行了较为详尽的讨论,最后对视频编码技术的未来发展进行了展望。本文介绍了视频编码技术及其发展情况;对 MPEG-4 的系统、编码机制及编码特性进行了讨论;对视频编码技术的发展情况进行了展望。

2. 视频编码技术

视频信息的编码较静态图像编码必须要多处理一

维信息,即时间维信息。如果只是将每帧图像当作单独的图像进行处理,编码效率将会受到限制。因此,帧间编码方法除了要利用空间冗余、频谱冗余和心里视觉冗余外,还要利用相邻图像间存在的相似性而产生的时间冗余,来提供较优的编码效率。以下我们将从波形编码、基于对象的编码、基于模型的编码和分形编码四方面^[2]来讨论视频编码技术及其发展情况。

2.1 基于波形的编码技术

如果将视频数据看成具有三个维度的信息,那么用于静态图像中的波形编码技术均可被推广到视频编码领域。然而,由于在低比特率下用于视频编码的 3-D 变换方法^[3]通常会产生块效应,因而很少被采用。Karlsson 等人在文[4]中首次提出了视频的三维子带编码方法,该方法经过适当的子采样,将视频分解成各种不同的分量信号,其范围从一个低的空间和时间分辨率到各种较高频率的详细信号分量。对每个具有不同空间-时间频率的视频子带,根据其统计和视觉特点分别独立地进行编码。文[5,6]对该方案进行了改进。三维子带编码方法以其无块效应和具有可分级性(Scalability)而受到重视。然而它所存在的一个缺点是时间方向的滤波并不是沿着运动的方向,这通常会影响到时域冗余量的减少。减少帧序列时域冗余量的最有效和最流行的方法要数运动补偿法(MC),所谓运动补偿是指在编码端估计出图像中运动部分在相邻两帧间的位移,然后用类似静态图像预测编码的方法去编码。Dufaux 等人在文[7]中给出了几种运动估计技术,这些技术在保证预测效果的前提下,力求减少计算的复杂度。目前常见的 MC 技术通常利用过去帧(单向)

^{*} 南京大学计算机软件新技术国家重点实验室资助项目。王相海 博士后,主要研究领域为计算机图形学、图像及多媒体信息处理。张福炎 教授,博士生导师,主要研究领域为计算机图形学、多媒体技术。

或过去和未来帧(双向)^[18]进行运动估计或采用重叠式运动补偿方法^[19,21]。此外,基于时间分量的子带编码和MC相结合的编码方法^[10]以及利用LVQ(Lattice Vector Quantization)和半像素精度的MC对其进行改进的方案^[11]也被提出。在编码过程中,对预测误差图像(DFD)的编码通常采用DCT方法^[1,2]和小波变换方法^[1,2],前者在大多数视频编码标准中被建议采用,而后者则具有较好的发展前途。

基于3-D子带编码、运动补偿子带编码和运动补偿变换编码的波形编码算法已被大量用于低比特率的视频编码中^[14~16]。

2.2 基于对象的编码技术

基于对象的编码技术是在第二代静态图像编码技术的基础上发展起来的,比如,Willemin等人在静态图像四叉树分割编码^[21]的基础上提出了基于八叉树的分裂和合并的方法^[1,2],Salembier等人给出了基于数学形态学的方法^[19],Jozawa等人将对象定义为具有形状、纹理内容和运动三个参数的区域,给出了基于对象的编码技术^[20],其中三个参数可以通过基于动态2-D或3-D对象原模型的图像分析技术获得^[21]。对象的纹理内容可以采用基于变换的编码方法^[21,22];形状信息的编码可以采用位图编码方法^[21,23]、轮廓信息的链式编码方法^[24]、四叉树形状表示法^[25]或中轴变换方法^[26]。试验表明,形状编码通常付出较高的代价,因此一些更有效的形状编码方法被探讨,比如测地线的形态学骨架法^[27]和形状编码前的预处理操作技术^[28]等方法也已经被给出。

基于对象的编码技术不仅解决了在大压缩比的编码中所产生的块效应,实现高压缩比,而且可以实现基于图像内容(对象)的交互操作。为此MPEG-4标准^[1]和H.263+标准^[29]均采用了该编码技术。有关MPEG-4编码的一些情况可参见第2节内容。

2.3 基于模型的编码技术

所有的编码技术都得依靠一个确定的模型,然而我们这里所说的基于模型的编码是指利用预定义的2D或3D模型来表示一个3D场景的2D投影像图的方法,其目标是寻找一个带有对应参数的适当模型,该模型应尽可能地接近场景中要编码的对象。

基于模型的编码技术一般包含两个主要的步骤,即分析和合成。对自然场景的分析通常是比较困难的一项工作,近几年人们把注意力集中在一些简单场景的研究上,比如人肩头像序列^[30,31],取得了较好的结果;而在解码端对对象的合成相对比较容易,这是因为计算机图形学在这一领域已经积累了很多成熟的工具。基于模型编码的详细情况可参阅Pearson所给出

的综述性文章^[32],限于篇幅,此处不再赘述。ILMPFG-4标准中包含了基于人的面部模型和身体模型的编码技术^[1,2],具体可参见第2节。

2.4 基于分形的编码技术

分形视频编码技术也是在静态图像分形编码技术^[33,34]的基础上发展起来的。在过去的几年中已经出现了许多不同的编码方法,Beaumont给出了一种直接将2D延拓到3D情况的编码方法^[35];为了减少计算的复杂性,Li等人提出了一种不需要进行区域块搜索的编码方法^[36],但该方法通常会增加收缩变换的复杂性;Reusens和Lazar等人分别独立地给出了一种基于八叉树结构的序列自适应分割分形编码方法^[37,38],前者在方法中利用收缩变换或3D时间块匹配技术对分得的3D子块进行编码,而后者只采用了收缩变换技术。

分形图像编码是较有发展前途的编码方法之一,它具有许多好的性质,比如它突破了以往熵编码的界限,在编码过程中采用了类似描述的方法,而解码是通过迭代完成的,具有分辨率无关的解码特性等。然而就目前发展情况看,它还存在着许多不足的地方,比如压缩比较小,恢复图像中仍然有较为严重的方块效应,运算量较大等,究其原因主要在于在编码过程中没有考虑图像的内容和含义,只进行了盲目的分块。

3. MPEG-4 视频编码机制及其特性

MPEG-4^[1,2]和MPEG-2^[40]标准打开了视听设备的数字领域并允许高性能服务的开发,然而它们所支持的数据模型是有限的,且不能满足新兴的如超级连接、交互式操作以及自然和人工数据的合成等多媒体应用需求。新颁布的MPEG-4标准的目的就是提供一个视听表示的标准,它支持数字视听数据新的通信、获取及交互方式,并提供了不同的服务范围间进行电视通信、广播、交互所需的通用技术上的解决方案。限于篇幅,以下我们仅就MPEG-4的系统、视频编码机制和编码特性进行讨论,有关MPEG-4编码的现状及编码工具和算法的详细描述见文[1,41~43]。

3.1 MPEG-4 系统

MPEG-4系统的任务是开发一个视音频对象及其相关的时变数据编码流的表示法,并描述它们是如何结合的。图1是基于对象的MPEG-4系统结构图。在发送端,与各个对象相关的几个基本流被复合传输;在接受端,基本流被多路分离,各个媒体对象被解码,并使用场景描述信息来合成整个场景。

MPEG-4系统不处理视音频的编码,而是处理与流相关的多路技术,比如如何将视音频对象有机地结

台来创建音频场景及实现流的同步、存储和传输等。

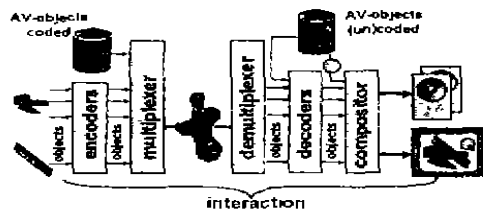


图1 MPEG-4系统结构图

3.2 MPEG-4 视频编码机制

MPEG-4 之前的视频标准,无论是 MPEG-1、MPEG-2 还是 H.261、H.263,都是把视频序列按时间先后分为一帧帧,每一帧图像又分成 16×16 的宏块来进行运动补偿和编码。这种方案在编码过程中不考虑图像的具体结构和内容,通常会带来一些问题,比如在高压比下具有严重的块效应,不能对图像的内容进行查询、编辑、回放等操作以及没有利用人的视觉系统的特性等等。MPEG-4 编码采用了基于对象的编码方案,即根据图像的内容将图像分割成不同的视频对象,并对视频对象的形状、大小、纹理等特征进行编码;同时还将基于合成的编码方案(模型编码)和基于网格的编码技术结合在编码中。

(1) MPEG-4 编码解码的数据结构 MPEG-4 按照五个层次组织要编码的图像:视频段(VS)、视频对象(VO)、视频对象层(VOL)、视频对象层组(GOV)和视频对象平面(VOP)。每个 VS 由一个或多个 VO 构成;每个 VO 由一个或多个 VOL 构成;每个 VOL 代表一个层次,表示某一种分辨率;在每个 VOL 上,都有时间上连续的一系列 VOP(见图 2)。

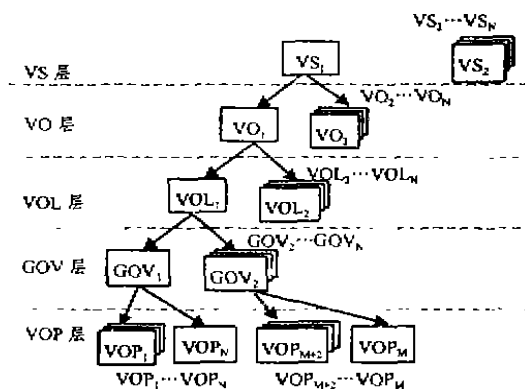


图2 MPEG-4 编码解码数据结构

(2) MPEG-4 视频编码结构 图 3 给出了 MPEG-4 视频编码的基本结构图,整个结构体现了基于对象的编码思想。MPEG-4 中的主要视频对象包括:自然视频对象、Sprite 对象、二维网格对象、三维人脸和身体

对象、静态纹理对象以及三维网格对象,针对不同的对象,MPEG-4 通常有不同的编码方法,以便达到高的压缩比,并实现不同应用环境下所需实现的交互操作。

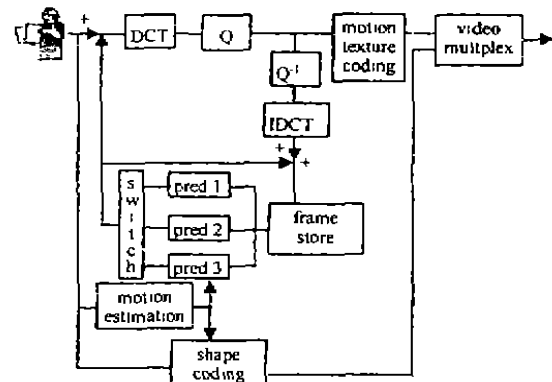


图3 MPEG-4 视频编码器基本框图

(3) MPEG-4 视频解码结构 MPEG-4 码流的解码过程是编码的逆过程,图 4 给出了一个解码示例。除了 IDCT 外,解码过程的定义必须满足所有的解码器都必须能够生成在数值上完全一致的结果,只要能做到这一点所有的解码过程都是可以采用的^[4]。由于 IDCT 是基于统计方法定义的,所以允许有该功能的不同实现。

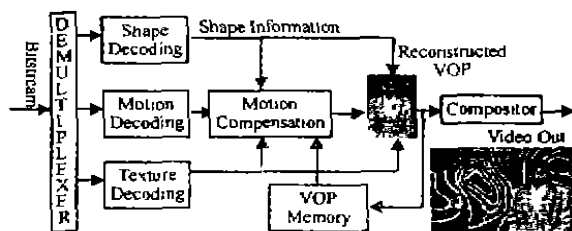


图4 MPEG-4 视频解码示意图

3.3 MPEG-4 视频编码的主要特性

MPEG-4 视频编码通常具有以下特性^[4]:

(1) 采用了基于对象的编码方法,并能实现基于对象(内容)的编辑和交互操作。MPEG-4 视频标准能描述组成一幅画的各个任意形状视频对象之间的空间和时间关系,并允许随机访问每个视频对象,即能以 0.5 秒的时间间隔访问对象,能单独解码对象的形状信息而不解码对象的纹理信息,能实现对视频对象进行剪贴、平移、旋转等编辑操作。

(2) 可以实现多输入流多视点的编码和解码。MPEG-4 能有效地编码来自不同视频源的图像,并能根据配置文档在解码端把这些对象的码流同步地解出来组成一幅图。在多视点的立体视频编码的应用中,MPEG-4 可利用同一场景不同视点的信息冗余性提高

编码效率,一般对同一场景可支持4个视点的编码。

(3)具有可分级性(Scalability)。一个对象比特流的可分级性是指利用该比特流的一部分(或者说子集)就可以产生该对象的一个有用表示的能力。MPEG-4支持各种不同传输速率下的编码,也支持固定码率和编码率的编码,即具有码率可分级性;此外它也支持不同分辨率编码,比如时间分辨率、空间分辨率,即具有时间可分级和空间可分级性;同时它还支持对解码图像信噪比的控制。

(4)具有在噪音环境中的鲁棒性。MPEG-4对每个对象的关键数据如对象的头信息和形状信息提供更高的容错保护,并提供了一个新的变长编码码表,能检验传输中产生的误码和部分纠正误码。

(5)采用了模型和网格编码。MPEG-4定义了人脸模型参数(FDP)和人脸运动参数(FAP),也定义了人体模型参数(BDF)和人体运动参数(BAP)。在解码器中的人脸模型能通过传来的运动参数产生面部的运动,也可以通过传来的脸的模型参数由一个通用的人脸模型生成一个特定的人的脸部。有关人体运动的规范正在研究之中。此外,MPEG-4可以对需要编码的二维或三维视频对象生成各种网格模型,并能有效编码基于网格的视频对象的运动信息,包括编码网络的拓扑形状和每个结点的运动矢量。

(6)可以实现全局运动估计和Sprite图像编码。MPEG-4定义了二维和三维的全局运动模型,包括平移、旋转、映射和投影等操作,以补偿基于摄像头的运动所带来的不足。如果一段视频序列的背景是固定的,但每帧图像的背景是这个大的固定背景的一部分,MPEG-4可以静态或动态地生成这样的背景图像,该图像被称为Sprite图像。对这样的背景视频对象的运动估计和补偿是参照Sprite图像进行的。

4. 视频编码技术展望

随着Internet和远程通讯的迅猛发展,对视频编码的一些更高要求被提了出来,MPEG-4中所体现的基于对象(内容)的编码、可分级编码、对错误的鲁棒性以及直接对码流进行操作等特征已成为新一代编码技术所追求的目标。尽管预测未来视频编码技术能达到怎样的程度是比较困难的,但随着多媒体应用领域的发展和视频编码工具的不断完善,我们认为未来视频编码有可能在以下几个方面会有更好的发展:

(1)在完善MPEG-4中基于对象编码技术的基础上,利用更多的智能图像分析技术,使基于内容的视频编码技术逐渐向语义层发展。

(2)基于模型的编码技术会逐步完善,朝着更加一般化、实用化方向发展,相信它与基于对象的编码方法

一道,会成为未来视频编码方法的主流。

(3)用户对视频编码信息的操作和访问更具开放性和灵活性。对编码信息的可操作性是未来视频编码技术所追求的一个重要目标,它不仅可以节省存储开销、减少运算的复杂度,而且在操作的过程中可以大大降低噪音的引入。开放和灵活的视频编码信息操作将大大拓宽多媒体的应用领域。

(4)编码技术应能有效地适应基于内容检索的需求。基于内容的检索是多媒体和Internet技术发展过程中所提出的一项重要的实用技术,如何使视频编码技术能有效地适应基于内容检索的需求同样是视频编码技术所追求的一个重要目标。

(5)与人体视觉特性相结合的视频编码技术有更好的发展。纯粹从像素值出发的MSE准则与人眼的视觉判断通常有一定的误差,尽管目前已有一些与人体视觉特性相结合的编码方案,但更加有效的编码技术仍然是进一步需要解决的课题。

结束语 视频编码技术作为多媒体技术的重要组成部分随着Internet和远程通讯的发展越来越受到重视,同时这些领域的发展也对视频编码技术提出了更高的要求。本文首先以波形编码、基于对象的编码、基于模型的编码和分形编码四方面为主线讨论了视频编码技术及其发展情况,同时介绍了新颁布的MPEG-4视频标准的系统、视频编码机制和编码特性,最后对视频编码的未来发展进行了讨论。

参考文献

- 1 ISO/IEC JTC1/SC29/WG11. Overview of the MPEG-4 standard. N3536, Beijing, 2000. 7
- 2 Ebrahimi T, Kunt M. Visual data compression for multimedia applications-an overview 1997
- 3 Baskurt A, Goutte R. 3-Dimensional Image Compression by Discrete Cosine Transform. In: Proceedings of the European Signal Processing Conference-EUSIPCO, 1988, 4: 79~82
- 4 Karlsson G, Vetterli M. Three Dimensional Subband Coding of Video. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP, 1988, 2: 1100~1103
- 5 Ebrahimi T, Kunt M. Image Sequence Coding Using a Three Dimensional Wavelet Packet and Adaptive Selection. In: Proceedings of SPIE Conference on Visual Commun. and Image Processing-VCIP, Boston, USA, 1992, 1818: 222~232
- 6 Chang E, Zakhori A. Scalable Video Coding Using 3-Dsubband Velocity Coding and Multirate Quantization. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP, 1993, 5: 575~577
- 7 Dufaux F, Moscheni F. New Perspectives in Motion Estimation Techniques for Digital TV. In: Proceedings of the

- IEEE, Special Issue on Digital Television-Part I: Enabling Technologies, 1995, 83(6): 858~878
- 8 Musmann H G, Pirsh P, Gralert H. Advances in Picture Coding. In: Proceedings of the IEEE, 1985, 73: 523~548
- 9 ITU-T SG XV. DRAFT H. 263, Video Coding for Low Bitrate Communication, 1996
- 10 Ohm J R. Temporal Domain Subband Video Coding with Motion Compensation. In: Proceedings of the IEEE International Conference on Acoustics, Speech and signal Processing-ICASSP, 1992, 3: 229~232
- 11 Ohm J R. Three-Dimensional Subband Coding with Motion Compensation IEEE Trans. on Image Processing, 1994, 3(5): 559~571
- 12 Stiller C, Lappe D. Laplacian Pyramid Coding of Prediction Error Images. In: Proceedings of SPIE Conference on Visual Communications and Image Processing-VCIP, 1991, 1605: 47~57
- 13 Zhang Y Q, Zafar S. Motion-Compensated Wavelet Transform Coding for Color Video Compression. In: Proceedings of SPIE Conference on Visual Communications and Image Processing-VCIP, 1991, 1605: 301~316
- 14 Ngan K N, Chooi W L. Very Low Bit Rate Video Coding Using 3D Subband Approach. IEEE Trans. on Circuits and Systems, 1994, 4(3): 309~316
- 15 Dachike K, et al. Motion Compensation Subband Extra/Interpolative Prediction Coding at Very Low Bit Rate. In: Proceedings of Very Low Bitrate Video Workshop-VLBV, Essex, UK, 1994
- 16 Katto J, et al. A Wavelet Codec With Overlapped Motion Compensation for Very Bit-Rate Environment. IEEE Trans. on Circuits and Systems for Video Technology, 1994, 4(3): 328~338
- 17 Kunt K, Ikononopoulos A, Kocher M. Second-Generation Image Coding Techniques. In: Proceedings of the IEEE, 1985, 73(4): 549~574
- 18 Willemin P, Reed T R, Kunt M. Image Sequence Coding by Split and Merge IEEE Trans. on Communications, 1991, 39(12): 1845~1855
- 19 Salembier P, Pardo M. Hierarchical Morphological Segmentation for Image Sequence Coding. IEEE Trans. on Image Processing, 1994, 3(5): 639~651
- 20 Hotter M. Object-Oriented Analysis-Synthesis Coding Based on Moving Two-Dimensional Objects, Signal Processing, Image Communication, 1990, 2(4): 409~428
- 21 Jozawa H, Watanabe H. Video Coding Using Segment-Based Affine Motion Compensation. Proceedings of the Picture Coding Symposium-PCS, Sacramento, USA, 1994: 283~341
- 22 Lavagetto F, Cocurullo F, Suringa S. Texture Approximation Through Discrete Legendre Polynomials. In: Proceedings of Very Low Bitrate Video Workshop-VLBV, Essex, UK, 1994
- 23 Bussen F, Ebrahimi T. A Simple and Efficient Binary Shape Coding Technique based on Bitmap Representation. In: Proceedings of the IEEE International Conference on Image-ICIP, 1997
- 24 Freeman H On the Encoding of Arbitrary Geometric Configurations. IRE Transactions on Electronic Computers, 1961, EC-10: 260~268
- 25 Samet H Region Representation: Quadrees from Boundary Codes. Commun. of the ACM, 1980, 23(3): 163~170
- 26 Blum H, Nagel R N. Shape Description Using Weighted Symmetric Axis Features. Pattern Recog., 1978, 3: 167~180
- 27 Brigger P, Ayer S, Kunt M. Morphological Shape Representation of Segmented Images based on Temporally Modeled Motion Vectors. In: Proceedings of the IEEE International Conference on Image Processing-ICIP, 1994, 3: 756~760
- 28 Gu C, Kunt M. Contour Simplifications and Motion Compensation for Very Low Bit-Rate Video Coding In: Proceedings of the IEEE International Conference on Image Processing-ICIP, 1994, 2: 423~427
- 29 ITU-T SG XVI. DRAFT H. 263 + Q15-A-60, Video Coding for Low Bitrate Communication, 1997
- 30 Huang T S, Reddy S C, Aizawa K. Human Facial Motion Modeling, Analysis, and Synthesis for Video Compression. In: Proceedings of SPIE Conference on Visual Communications and Image Processing, 1991, 1605: 234~241
- 31 Li H, Forchheimer R. Two-View Facial Movement Estimation. IEEE Trans. on Circuits and Systems for Video Technology, 1994, 4(3)
- 32 Pearson D. Developments in Model-Based Video Coding. In: Proceedings of the IEEE, Special Issue on Digital Television-Part I: Enabling Technologies, 1995, 83(6): 892~906
- 33 Jacquin A E. A novel Fractal Block-Coding Technique for Digital Images. In: Proc. ICASSP, 1990: 2225~2228
- 34 赵耀, 王红星, 袁宝宗. 分形图像编码研究的进展. 电子学报, 2000, 28(4): 95~101
- 35 Beaumont J M. Image Data Compression Using Fractal Techniques. British Telecommunication Technology Journal, 1991, 9(4): 93~109
- 36 Li H, Novak M, Forchheimer R. Fractal-Based Image Sequence Compression Scheme. Optical Engineering, 1993, 32(7): 1588~1595
- 37 Reusens E. Sequence Coding based on the Fractal Theory of Iterated Transformations Systems. In: Proceedings of SPIE Conference on Visual Commun. and Image Processing-VCIP, 1993, 1: 132~140
- 38 Lazar M S, Bruton L T. Fractal-Based Image Sequence Compression Scheme. IEEE Trans. on Circuits and Systems for Video Technology, 1994, 4(3): 297~308
- 39 CCITT SG VIII, ISO/Moving Pictures Experts Group, Committee Draft Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s 1990
- 40 CCITT SG VIII, ISO/Moving Pictures Experts Group, Committee Draft. Coded Representation of Picture and Audio Information, 1993
- 41 高文, 刘峰, 黄铁军. 数字图书馆. 北京: 清华大学出版社, 2000
- 42 高文, 吴枫. MPEG-4 编码的现状和研究. 计算机研究与发展, 1999, 36(6): 641~652
- 43 王相海, 张福炎. MPEG-4 自然视频编码工具和算法. 计算机科学, 2001, 28(1): 1~7