

基于 DVB 的视频关键帧浏览系统的实现

郑 鹏 薛海峰

(武汉大学计算机学院, 武汉 430072)

摘 要 随着计算机技术和电视技术的发展, 数字视频节目日益增多. 为了迅速了解视频节目的内容, 提出了建立基于数字视频广播(DVB)的视频关键帧浏览系统的方案. 该系统首先利用 MPEG 压缩视频提供的压缩参数, 直接将视频序列划分成以镜头为单位的视频片段; 然后提取每个镜头中的第一个 I 帧作为关键帧, 重构其 DC 图象; 最后, 根据 DVB 标准, 通过扩展 SI 表实现对这些视频关键帧进行封装的数据结构, 从而实现电视台对视频关键帧的发送和接收. 文中给出了前端系统结构示意图和基于关键帧 DC 图象的快速浏览实例. 由于该系统直接利用了压缩参数, 从而减少了解压缩的开销, 具有计算代价小, 浏览速度快等特点.

关键词 图象处理(510·4050) 数字视频广播 关键帧 机顶盒

中图法分类号: TP391.41 **文献标识码:** A **文章编号:** 1006-8961(2003)12-1462-05

The Implementing of Browsing Key-frames of Video Based on DVB

ZHENG Peng, XUE Hai-feng

(School of Computer, Wuhan University, Wuhan 430072)

Abstract With developing of computer technology and TV technology, the amount of digital video program is larger and larger. In order to find out the content of the video program rapidly, a system of browsing video key-frames based on DVB(digital video broadcasting) is presented in the paper. Firstly, the video sequences are segmented into shots directly by making use of compressed parameters provided by MPEG compressed video, such as DCT coefficients, motion vector, the types of macroblock and so on. Secondly, the first I-frame is selected as key-frame, and DC image of the I-frame is reconstructed. Finally, based on the standard of DVB, the SI tables are extended to realize the data structure of encapsulating the key-frames so that the key-frames of video in TV stations can be picked-up and transferred. The system diagram and a real example of rapid browsing prototype based on DC image of key-frame are offered in the last of the paper. Because the method that presented in the paper makes use of the compressed parameters provided by MPEG compressed video and reducing the cost of decompressed, the method has the feature of lower computing cost and faster browsing.

Keywords Digital video broad casting, Key-frame, Set-top box

0 引 言

现代信息技术和通讯技术的飞速发展, 带动了电视数字化的发展, 大量高质量的数字电视节目通过通讯卫星下传到各级电视台或家庭用户, 这些数字电视节目所采用的数字视频广播标准主要有: 欧洲的 DVB(Digital Video Broadcasting)、美国的 ATSC(Advanced Television Systems Committee) 和日本的 ISDB(Integrated services digital

broadcasting). 这些数字视频广播标准中, 欧洲的 DVB 标准使用得比较广泛, 我国采用的也是这种数字视频广播标准. DVB 标准是建立在 MPEG-2 协议标准之上的. 数字电视是电视的发展方向, 今后将涌现大量的数字电视节目, 但电视台以及个人用户都很难对其进行有效的检索和管理, 其主要原因是数字视频节目容量大, 不易存储和索引.

现在, 随着多媒体存储和检索技术的发展, 上述问题可以基本得到解决. 解决上述问题, DVB 标准本身没有现成可利用的方法, 为此提出了一套系统

基金项目: 国家自然科学基金(60173045/F020204)

收稿日期: 2003-02-13; 改回日期: 2003-06-30

的解决方案,此方案以 DVB 标准作为基础,对其进行了功能扩充,同时采用实时的视频关键帧提取技术,实现了前端电视台对数字电视节目进行视频关键帧的提取和存储,然后,根据扩充后的 DVB 标准将其传出去,用户终端使用电视机机顶盒(STB)接收这些视频关键帧,快速浏览后,即可确定自己收看的电视节目。

1 视频分段算法

视频序列一般由多个镜头(shot)连续拼接而成,镜头作为视频序列中的天然片段,可作为视频检索的基元。这样,视频分段就转化为对镜头变换的检测。

镜头的变换有突变(cut 或 break)和渐变(gradual transition)两种类型,对渐变的检测比对突变的检测要困难一些,因为两个连续的镜头之间的变化是经过若干帧的变化逐渐完成的。渐变主要是在制作视频时,采用了诸如淡入(fade in)、淡出(fade out)、溶暗(dissolve)、划出(wipe)等一些特技手法而形成的一种镜头切换。淡入是指一个镜头亮度从暗到明逐渐出现;淡出是指一个镜头亮度从明到暗逐渐消失;溶暗是指前一个镜头在淡出的同时后一个镜头在淡入;划出是指通过一个边界的移动,前一镜头逐渐消失同时后一镜头逐渐出现,这一移动边界可以有多种形状。

Koprinska 等各种分段算法进行了分析比较^[1],从中可以发现设计识别镜头切分点算法的主要依据是:一段镜头序列内部,相邻帧间有很大的连续性和相似性,而一旦发生了镜头切换,这种连续性和相似性将被破坏。但在实际中会遇到很多问题。首先,对于渐变来说,尽管发生了切换,但由于是平滑衔接,相邻两帧之间连续性仍保持得很好,按照检测相邻帧间的相似性从而判断有无切变的思想就会漏掉所有的渐变切换点。对于 MPEG 压缩视频,目前只有镜头突变检测获得了较高的准确率,而且大部分工作是利用对 DCT 系数部分解压来进行视频分析的,其计算成本有待进一步降低。对镜头渐变的检测,其准确率有待于进一步提高。

在 DVB 标准中,对数字电视节目进行编解码用的是 MPEG-2 标准。对视频分段采用检测场景变化来实现。直接在压缩视频上进行场景变化检测,可减少解压开销,提高处理速度。MPEG-2 的编码码流分为 6 个层次^[2],从上至下依次为:视频序列层

(video sequences), 图象组层(GOP: groups-of-pictures, SGOP; 子图象组), 图象层(pictures), 像条层(slice), 宏块层(MB: macroblock)和像素块层(block)。

图象组层下是图象层,分为 I、P、B 3 类。在 MPEG 中,视频序列 I 帧之间的帧序列称为一个图象组(不包括后一个 I 帧)。MPEG 并没有对图象组作出严格规定,使得人们可以用相当灵活的方式来组织视频,但 I、P、B 帧序列遵循如下规律:即 I 帧之间由 0 个或多个 P 帧组成, I 帧和 P 帧之间,或 P 帧与 P 帧之间由 0 个或多个 B 帧组成。典型的图象组结构如图 1 所示。

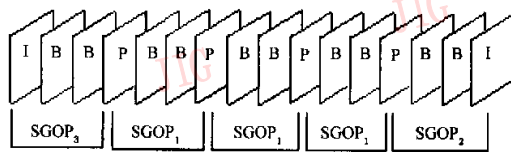


图 1 典型的图象组结构

I 帧中的所有宏块都是帧内编码的,只有一种宏块类型,此类宏块只反映了 I 帧图象中空间压缩关系,没有反映帧间的时序关系。P 帧中的宏块有两种类型,帧内编码和前向预测,如果 P 帧中的某一宏块可以从前面最近的 I 帧或 P 帧中进行预测,说明此宏块变化不大,与前面帧中的宏块具有较大的相关性,其类型为前向预测,若不能进行预测,说明此宏块与前面的 I 帧或 P 帧无关,其类型为帧内编码。B 帧中的宏块有 4 种类型,帧内编码、前向预测、后向预测和双向预测,当宏块与前后最近的 I 帧或 P 帧无关、与前相关、与后相关、双向相关时分别对应上述 4 种类型。

Pei 提出利用 P 帧和 B 帧中宏块类型的统计值检测镜头突变,获得了较高的准确率^[3]。通过实验,得到检测突变的阈值为 0.60。

在溶暗渐变检测中,仍然用 P、B 帧中宏块类型进行预判。虽然渐变会引起 P 帧中块内编码的宏块数增加,但由于帧间差异较小,块内编码的宏块数远小于突变时块内编码的宏块数。为此,采用的阈值比突变小。在实验中检测渐变的阈值为 0.05。

Pei^[3]指出在溶暗视频序列中的 B 帧的运动补偿倾向于 FB 模式,双向预测的宏块数在 B 帧中占多数。在实验中,我们发现,虽然在溶暗渐变中 B 帧的双向预测宏块数占多数,但当镜头运动或帧图象中有运动对象时,B 帧的双向预测宏块数也占多数,

这样就降低了渐变检测的精确度.为提高精确度,我们在检测到双向预测宏块数占多数的B帧时,再检测相应SGOP中P帧的运动矢量.因为当镜头出现运动时,P帧中的运动矢量较大,当出现溶暗渐变时,主要是亮度变化,运动矢量较小.通过对运动矢量大小的判断可以区分溶暗渐变和镜头运动.

MPEG压缩参数提供了运动矢量的值.为了判断运动矢量的大小,引入一个规范因子 p ,对运动矢量进行规范,如果规范化后,大多数宏块的运动矢量为0,则肯定是溶暗渐变.因为帧图象中编码的最小单位是块(block),块是由 8×8 的像素组成,所以令规范因子 p 的值为4,即将位移距离小于半个像块的运动宏块,当作静止宏块处理.实验表明规范因子能够提高算法的鲁棒性.

溶暗渐变是线性的,而划出渐变则不具备此特性.相对于溶暗渐变,划出渐变的检测更困难,相关的文献也更少. Alattar^[4]和 Fernando等人^[5]提出了基于非压缩视频的划出检测算法,尽管可以获得较为满意的结果,但对于不同的划出形状、方向、和模式,计算成本很大. Wu等人^[6]提出了一个基于DC序列的划出检测算法,但只能识别水平和垂直划出.

虽然划出的形式各种各样,但通过观察分析,发现在划出中,移动边界的存在可分为3种类型:

(1)前一镜头被后一镜头逐渐推出,两个镜头的连接处形成一个移动边界;

(2)前一镜头逐渐缩小,后一镜头逐渐扩大,两个镜头的连接处形成一个移动边界;

(3)通过一条扫描线,前一镜头被后一镜头逐渐覆盖,扫描线形成一个移动边界.

在第3种类型的划出渐变过程中,移动边界所占据的宏块变化较大,P帧中这些宏块的前向预测成本较高,所以MPEG在编码时往往采用块内编码,而边界之外的宏块变化较小,也无运动,所以MPEG编码时,采用前向预测,相应宏块的运动矢量很小.此类划出特征与溶暗类似.

对于前两种类型的划出渐变,由于全局宏块发生变化,少量宏块变化较大,采用块内编码,大量宏块采用前向预测,而且运动矢量的大小反映了划出的速度.这一特征与镜头的摇动或推拉运动类似.当摄像机摇动或推拉时,P帧中也会有一定数量的帧内编码宏块,多数宏块前向预测,其运动矢量的大小反映了镜头运动的速度.为了与镜头运动区分开来,考察P帧前向预测宏块中的亮度误差值.在前两种

类型的划出渐变中,由于大多数宏块的亮度变化较大,导致P帧中预测宏块亮度误差较大,所以总的预测误差也较大,而镜头运动中宏块亮度变化较小,所以P帧中预测宏块亮度的误差也较小.为此,提出利用P帧中的亮度误差来区分镜头运动与划出.

P帧的亮度误差定义为

$$DC_{err} = \frac{\sum |DC_{ij}^p|}{S_p} \quad (1)$$

式中, DC_{ij}^p 表示P帧中前向预测型宏块的亮度误差, S_p 表示P帧中前向预测型宏块的块数, i, j 分别表示宏块的行、列坐标.

若 $DC_{err} > Th$,则说明存在移动边界,检测到划出渐变,否则未检测到划出渐变.其中, Th 是为区分渐变和镜头运动所采用的阈值,对80000多帧视频序列进行渐变检测实验,当 DC_{err} 的值为100时,是渐变和运动的分水岭.这样就构成了一个比较完整的镜头分段算法.

2 视频关键帧提取

I帧只包含实际像素数据的DCT系数,无运动矢量,P、B帧包含从参考帧预测后留数误差生成的DCT系数和与参考帧间的运动矢量.因为I帧包含实际像素的DCT系数,所以可直接生成DC图象,而P或B帧却需要利用重构算法构造DC图象^[7,8],因此最好的候选帧是I帧.关键帧可以很好地描述视频内容,同时也是建立视频索引的基础.由于通常的镜头长度至少超过1秒,所以镜头中至少包含一个I帧.对于极短的镜头,在分段时已将其去掉,因为极短的镜头无法向观众传递信息.在进行视频分段之后,为了加快关键帧的提取速度,采用提取每个镜头中的第1个I帧作为该镜头的关键帧,利用关键帧的DC图象来进行视频的快速浏览.

3 发送视频关键帧时扩充的数据结构及终端用户的接收

DVB按照ISO13818^[9]标准对节目源进行量化和编码.把节目中的视频音频各自转变为ES(Elementary stream),然后加入头信息打包成PES(Packetized elementary stream),最后再切分打包成TS(Transport stream)进行传送,并在TS中插入PSI(Program service information)和DVB的SI

(Service information), 其中 PSI 用于节目的复用, SI 为解码器构成电子节目指南 (EPG) 及频道自动搜索提供各种各样的信息. 针对不同的应用, DVB 数据广播规范提供了以下 8 种可选择的方法: 数据通道 (Data piping)、数据流 (Data Stream)、异步数据流 (Asynchronous data streaming)、同步数据流 (Synchronous data streaming)、被同步数据流 (Synchronized data streaming)、多协议封装 (Multi-protocol encapsulation)、数据轮放 (Data carousels) 和对象轮放 (Object carousels). 对于每种方法, 都提供了相应的 API (Application program interface). 为了实现视频关键帧的发送, 选用对象轮放的方法, 这种方法为用户提供的应用, 独立于最底层的 MPEG-2TS 流的接口, 比较固定, 而且其提供的 API 与平台无关, 用于支持在 DVB 广播网中周期广播的业务.

在 DVB 标准中, 随着新业务的不断出现和增长, DVB 标准中的 SI 表也在不断地扩充. 在发送视频关键帧时, 现有的 SI 表不能有效发送这些视频关键帧, 所以我们需要对 DVB 的 SI 表进行扩充, 以此进行发送编码.

在 MPEG-2 的 TS 流中, 可以包含多个节目, 每个节目又可以包含多个基本码流, 基本码流和其他的控制数据都被打成固定长度的包, 每个包都有一个包标识符 PID (Packer identifier). 对于 PSI 和 SI 的大部分表都有确定的 PID. 为了实现发送视频关键帧, 需要通过扩充 SI 表. 在 DVB 标准中, 一些表的 PID 值是确定的, 0x0003 到 0x0007 的 PID 值保留未用^[10], 用值为 0x0004 的 PID 来标识新表 PVT, PVT 表的内容就是要发送的视频帧, 规定 PVT 表的 table_id 值 0x80 (table_id 值从 0x80 到 0xFE, 用户可以自定义). 表的内容有时很大, 由于在 DVB 中, 表只是一个逻辑上的概念, 其具体实现是通过段 Section 来进行的. 下边给出 PVT section 语法的主要数据结构内容:

```
pvt_section() {
    table_id;          /* 值为 0x80 */
    section_length;    /* 段长 */
    network_id;        /* 网络标识, 指明发送此表所用的
                       网络参数 */
    section_number;    /* 总段数 */
    delay_time;        /* 每个视频帧的延时时间 */
    for(i=0; i<N; i++){
        descriptor();
    }
}
```

```
CRC_32;              /* CRC 校验 */
}
```

对于其中的 descriptor(), 定义其 Tag 值为 0xf0, 其内容为所发送的每一个视频帧的内容, 数据结构的主要内容如下:

```
pvt_descriptor() {
    descriptor_tag;    /* 值为 0xf0 */
    descriptor_length; /* descriptor 数据的长度, 即为下
                       边循环中关键帧数据总长 */
    for(i=0; i<N; i++){
        key_frame_data; /* 关键帧数据 */
    }
}
```

依据上述数据结构对视频关键帧进行编码后, 就可以把编码数据输出到传送流复用器进行输出. 前端系统示意图如图 2 所示.

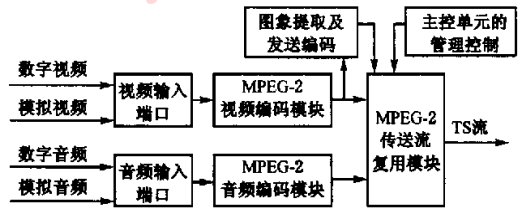


图 2 前端系统示意图

当终端用户接收到这些关键帧时, 机顶盒中的数据过滤器, 根据 PID 值, 通过软件控制过滤器硬件寄存器值进行解包以及数据的自动装配. 此时, 在软件中, 只需设定寄存器接收的 PID 值为 0x0004, 这样就可以把视频关键帧数据解包, 然后依据 table_id 和其他的数据结构就可以把视频关键帧重构出来. 将关键帧重构后, 通过 PVT 中的 delay_time 进行相应的帧时延, 然后依次接收并显示这些关键帧. 对于以上任务, 接收端机顶盒中的软件只需做少量的修改. 修改后的机顶盒中的软件可以采用 DVB 提供的 Downloader 机制进行空中升级, 非常容易实现.

此系统建立在充分使用已有设备基础之上, 只是在前端电视台添加了一台视频关键帧提取存储以及对这些帧进行发送编码的服务器. 这种处理可大大节省资金, 并且可为将来建立视频数据库奠定基础.

4 实验结果

利用本算法对新闻、体育、电影等几种不同类型的 MPEG 视频序列进行分段实验. 实验的总帧数为

8 万多帧;实验的图象帧尺寸有 3 种: 352×288 , 352×240 和 160×120 ; 图象帧中宏块数量分别为 396, 330 和 80.

兹定义识别率(R)和准确率(P)来作为评价算法的标准.

$$R = \frac{C}{C + M} \quad (2)$$

$$P = \frac{C}{C + F} \quad (3)$$

其中, C 为正确识别的镜头变换个数, M 为没有识别出的镜头变换数, F 为错误识别的个数. 实验表明, 对渐变的识别率和准确率分别达到 85.7% 和 84.4%, 均高于目前的同类算法.

建立基于关键帧 DC 图象的快速浏览器原型系统, 图 3 是对一段电影视频进行快速浏览的实例.

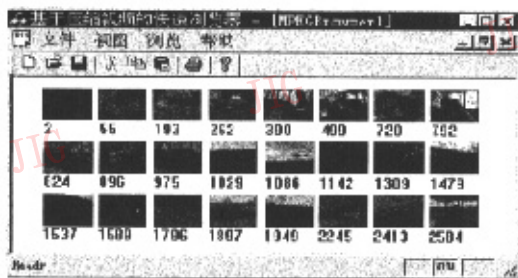


图 3 关键帧 DC 图象快速浏览实例

5 结 语

数字化时代的到来, 数字媒体也越来越多, 对于数字媒体的检索和管理日益迫切. 对于电视台大量的数字电视节目, 没有一个有效的管理, 将严重影响其业务, 也很难拓展新的业务. 我国的机顶盒功能比较单一, 基本上只是单一地接收电视节目, 几乎没有交互性. 交互式电视是未来电视的一个发展方向, 其要求电视台能够为用户提供电视节目的初步浏览, 而以上系统解决了电视台的视频检索和提供用户对节目的浏览, 进而可以实现视频数据库, 具有实际应用价值.

参 考 文 献

- 1 Koprinska I, Carrato S. Temporal video segmentation: A survey [J]. Signal Processing: Image Communication, 2001, 16(5): 477~500.
- 2 钟玉琢, 王琪, 赵黎等. MPEG-2 运动图象压缩编码国际标准及

MPEG 的新进展[M]. 北京: 清华大学出版社, 2002.

- 3 Pei S C, Chou Y Z. Efficient MPEG compressed video analysis using macroblock type information[J]. IEEE Transactions on Multimedia, 1999, 1(4): 321~333.
- 4 Alattar A M. Wipe scene change detector for segmenting uncompressed video sequences [A]. In: Proceedings of IEEE International Symposium Circuits and Systems[C], California, Monterey, 1998, 4: 249~252.
- 5 Fernando W A C, Canagarajah C N, Bull D R. Fade, dissolve and wipe production in MPEG-2 compressed video[J]. IEEE Transactions on Consumer Electronics, 2000, 46(3): 717~727.
- 6 Wu M, Wolf W, Liu B. An algorithm for wipe detection[A]. In: Proceedings of the IEEE International Conference on Image Processing[C], Chicago, 1998, 1: 893~897.
- 7 Shen K, Delp J. A fast algorithm for video parsing using MPEG compressed sequences [A]. In: Proceedings of the IEEE International Conference on Image Processing[C], Washington DC, 1995, 2: 252~255.
- 8 Yeo B L, Liu B. On the extraction of DC sequence from MPEG compressed video[A]. In: Proceedings of the IEEE International Conference on Image Processing[C], Washington DC, USA, 1995: 260~263.
- 9 ISO/IEC 13818 (MPEG 1 DIS). Generic coding of moving picture and associated audio[S], 1994.
- 10 Reimers U. Digital video broadcasting [J]. IEEE Communications Magazine, 1998, 36(6): 104~110.



郑 鹏 1965 年生, 副教授, 1986 年、1992 年于武汉大学计算机系获学士、硕士学位, 2002 年于武汉大学计算机学院获博士学位. 主要研究方向为视频信息处理.



薛海峰 1975 年生, 1997 年于河南师范大学计算机系获学士学位, 2001 年进入武汉大学计算机学院攻读硕士学位. 主要研究方向为视频信息处理.