

一种基于运动信息描述的视频格式自适应方案

沈 秋 李厚强 王 毅 何 佳

(中国科学技术大学电子工程与信息科学系信息处理中心 合肥 230027)

摘 要 多媒体应用领域存在多种视频编码标准,因此实现不同视频格式之间的自适应转换具有重要意义.该文提出了一种新颖的基于运动信息描述的视频格式自适应解决方案.该方案利用作者前期提出的基于 H.264/AVC 标准的运动信息描述算法提取能够复用的详细完备的运动信息,根据需要快速生成不同视频编码格式、多种码率的码流;也可以利用此运动信息,指导不同视频格式码流之间的快速自适应转换.作为该方案的具体应用,文中提出了一种针对 H.264/AVC 标准和 MPEG-2 标准的视频自适应方法.实验结果表明,文中提出的基于运动信息描述的视频格式自适应解决方案不仅可以很好地保证编码性能,同时具有较低的计算复杂度.

关键词 视频自适应;格式转码;H.264/AVC 视频编码标准;MPEG-2 视频编码标准

中图法分类号 TP391

DOI号: 10.3724/SP.J.1016.2009.01162

A Framework of Video Format Adaptation Based on Motion Description

SHEN Qiu LI Hou-Qiang WANG Yi HE Jia

(Information Processing Center, Department of Electronic Engineering & Information Science,
University of Science & Technology of China, Hefei 230027)

Abstract Due to the existence of various video coding standards in multimedia applications, it is of great significance to implement video format adaptation among these standards. This paper presents a novel framework for video format adaptation, which is based on motion description and is originated from the authors' early work. Such motion description comprises abundant motion information, and can be obtained by pre-encoding a video sequence. Motion description can be used not only to fast generate a required bitstream, but also to guide the transcoding between different video formats. As an example, format adaptation between H.264/AVC and MPEG-2 which is on the basis of the proposed framework is illustrated in detail. A number of comparison experiments are performed to show that the proposed solution can greatly reduce the computational complexity meanwhile retain the coding efficiency, which manifestly demonstrates the proposed solution is superior to the conventional transcoding technique.

Keywords video adaptation; format transcoding; H.264/AVC video coding standard; MPEG-2 video coding standard

1 引 言

随着通信技术和多媒体技术的发展,人们对视

频应用的需求不断增长.由于视频数据量巨大,人们需要对其先编码压缩再存储、传输.针对不同应用和随着技术的发展,各标准组织、国家已经制定了多个视频压缩标准,目前主要包括 ITU 制定的 H.26x 系

列^[1-3]和 ISO/IEC 制定的 MPEG-x 系列^[4-5]以及由我国制定的具有自主知识产权的 AVS 标准. 这些标准充分利用视频编码的各种技术将具有固定时间、空间分辨率的视频信号压缩成一定的码流. 该码流主要由运动信息和残差信息两部分组成. 然而, 这种视频表达方式过于单调, 难以适应当前复杂的应用环境. 为了解决网络的异构性、终端设备的多样性、多媒体应用的复杂性给视频应用带来的难题, 文献[6]的作者提出了视频自适应(video adaptation)的概念. 视频自适应主要研究如何将已有的视频表达形式变换成为满足限制条件的某种形式; 同时, 尽可能提高变换后视频的有用性(utility, 即用户的满意度).

视频自适应技术涉及的范围很广, 包括码率的自适应调整、时空分辨率的自适应改变、格式的自适应转换、视频摘要(Video Summary)的自动生成、以及视频合成(如 Video Mosaics)等^[6]. 其中, 格式的自适应转换应用广泛, 具有非常重要的实用价值. 格式的自适应转换主要研究如何将某种格式的码流简单高效地转换为另一格式的码流. 目前多媒体应用中存在多种视频编码标准, 这些标准在各自的应用领域均占有一席之地, 却不能直接应用于其他领域. 以 H.264/AVC 和 MPEG-2 为例: 一般来说, 在同等级图像质量条件下 H.264/AVC 具有高于 MPEG-2 等其他早期标准的数据压缩比率, 并且其加强了对各种信道的适应能力, 具有良好的网络友好性. H.264/AVC 有着巨大的发展潜力, 其在视频编码所占的地位也越来越重要. MPEG-2 作为发展比较成熟的标准在很多应用领域占有一席之地, 包括存储媒体中的 DVD、广播电视中的数字广播电视(DBTV)和高清晰度电视(HDTV)、交互式的点播视频(VoD)和准点播视频(NVoD); 而且, 目前国内广电系统大都采用与 MPEG-2 标准兼容的设备. 因此, 短期内 MPEG-2 无法被 H.264/AVC 彻底取代, 两种标准共存的局面仍会持续一段时间, 这使得实现 H.264/AVC 和 MPEG-2 两种格式的自适应转换成为视频格式自适应中迫切的任务.

在原始视频可得的情况下, 一种最直接的解决方法就是将视频分别压缩成两种格式多种码率的码流. 然而, 由于现有这些视频编码的计算复杂度非常高, 对同一段视频进行多次压缩是一项艰巨的任务, 并且其存储所耗费的资源也随之大大增长. 在原始视频不可得的情况下, 就需要采用视频转码技术(video transcoding)^[7-9]来实现码流格式的转换. 视

频转码是对压缩视频从一种形式到另一种形式进行转换的操作, 这里的形式可以指码率、时间分辨率、空间分辨率, 也可以指编码语法、内容等等^[9]. 视频格式的转码主要研究如何实现不同编码格式之间快速高效的转换, 按照操作空间的不同大体可以分为两种, 一种是像素域转码, 一种是变换域转码.

像素域的转码是将源格式的码流解码至像素域, 再进行重新编码生成目标格式的码流. 其中最直接的方法是完全重建视频, 并以其为参考进行一次完整的再编码. 这种方法能够保证最优的编码性能, 但是一次完整的编码过程要进行复杂的运动估计、变换、量化、熵编码等过程, 其计算复杂度非常高, 不利于实时转码的应用. 因此, 研究人员提出了多种快速转码的算法, 通过从解码码流中提取有用参数来降低转码过程的计算复杂度. 以 MPEG-2 到 H.264/AVC 的转码为例, 文献[10-12]利用 MPEG-2 码流的解码运动矢量来指导 H.264/AVC 的模式选择和运动估计过程, 从而大大降低了 MPEG-2 到 H.264/AVC 格式转换的计算复杂度. 文献[13]介绍了一种 H.264/AVC 到 MPEG-2 格式的转码方法, 通过合并 H.264/AVC 的编码模式和运动矢量来生成 MPEG-2 的运动信息, 从而避免复杂的运动估计过程. 为了进一步降低计算复杂度, 文献[14-16]采取直接在变换域进行 MPEG-2 到 H.264/AVC 格式转码, 这样可以跳过量化反量化步骤.

这些方法虽然都在一定程度上降低了编码的计算复杂度, 但也存在很多不可避免的缺陷. 一方面, 已有码流中相关信息很单一, 往往针对转码前的标准、现有码率等, 这使得在转码时可用来降低算法复杂度的运动信息远远不够丰富; 即使对原有运动信息进行必要的转化也无法避免编码性能的损失, 这种损失只能通过进一步的运算来弥补. 因此, 现有转码方法只能在编码性能和计算复杂度之间取得某种折中, 无法两者兼顾. 另一方面, 由于各种编码格式采用的编码模式、变换方法和表达形式等都存在差异, 对于每一种格式的转码都要设计不同的算法, 难以用一种统一的框架解决, 这使得格式自适应转码的实现复杂而不具有普适性.

现有的视频编码标准框架中, 计算最复杂、对编码性能影响最大的部分是运动估计模块, 因此, 视频格式自适应中最关键的问题在于如何在降低运动估计的计算复杂度的同时生成高质量的目标码流. 我们提出了这样一种设想: 如果在进行某种格式的视

频编码时能有详细的运动信息做指导,那么运动估计的过程就可以变得非常简单快速,从而既确保了编码性能,又兼顾了算法复杂度.我们在文献[17]中提出了一种基于运动信息描述的 H.264/AVC 标准的码率自适应算法:通过预编码得到视频序列的运动特性描述,并以某种数据结构加以存储;当需要一定质量的码流时,可以直接从运动信息描述中抽取适当的模式和运动矢量进行编码.由于该算法不需要运动估计的过程,编码的复杂度非常低,可以实现 H.264/AVC 码流的快速生成.这种运动信息描述包括了丰富和精细的运动信息,因此本文将这一工作进行了拓展,将运动信息描述应用于格式自适应转换.我们选择 H.264/AVC 作为参考标准,因为 H.264/AVC 有多样化的分块模式、精细的运动矢量尺度以及最优的评价标准,从而保证得到的运动信息描述的详细与完备性.其他的标准在这些方面与其存在一定的差异.以 MPEG-2 为例: H.264/AVC 采用了七种分块模式,而 MPEG-2 只有一种分块模式; H.264/AVC 的运动矢量精细到 1/4 像素,而 MPEG-2 只精细到 1/2 像素.基于这种运动信息描述,我们可以建立一个统一的技术框架来解决不同视频格式间的自适应转换.当然,由于不同视频标准的运动信息存在差异,基于 H.264/AVC 生成的运动信息描述需要经过相应的处理才能应用于其他格式的编码,这是该技术框架下必须解决的关键问题.

我们的基本思路如下:首先,对于给定的原始视频采用文献[17]提出的生成算法得到视频序列的运动信息描述,并以一定的数据结构进行存储;其次,设计基于该运动信息描述生成所需视频编码标准运动矢量的方法.利用该方法,不需要进行运动估计,可以直接快速生成任何格式任何码率的码流.

本文第 2 节简单介绍 H.264/AVC 运动信息描述算法及其应用;第 3 节提出我们的格式自适应的框架;第 4 节以 H.264/AVC 和 MPEG-2 之间的格式转换为例,分析了运动信息描述转换成所需的运动信息时存在的问题,并介绍我们的运动信息映射算法;实验结果在第 5 节给出;最后,我们在第 6 节总结全文.

2 基于 H.264/AVC 标准的运动信息描述算法

由于 H.264/AVC 标准采用率失真优化(rate-

distortion optimization)技术和可变块尺寸(variable block size)技术,不同码率获得的运动矢量变化很大,这给码率自适应技术带来了巨大的挑战.我们在文献[17]中提出通过预编码来获取运动信息,并采用一定的数据结构生成视频序列运动特性的完整描述,我们称之为运动信息描述.基于这种描述我们可以实现快速码率自适应,即编码时不需要再做运动估计,只需从该描述中抽取一套合适的运动矢量直接进行快速编码.

运动信息描述采用一种基于模式的运动信息分层描述模型,在这个模型中运动矢量根据模式来组织而不是量化参数 QP(Quantization Parameter).如图 1 所示,一个宏块(Macroblock)对应多个模式,而每个模式可以保存一套或多套运动矢量用于描述该宏块在这一分块模式下的运动特性.基于文献[17]提出的分层模型,通过预编码可以获得所需要的运动信息描述.这个过程可以分为多量化参数运动估计(multi-QP motion estimation)和运动量化(motion quantization)两部分.多量化参数运动估计模块为每个宏块确定在不同的量化参数下最优的模式和运动矢量.运动量化根据分层模型组织运动矢量,生成对整个序列的运动信息描述.在多量化参数运动估计过程中,同一个模式可能被多次选定为最优模式,即不同的量化参数下最优的模式相同,但同一个模式对应的最优运动矢量可能是不一样的.运动量化模块的作用就是为了减少同一模式下运动矢量的数量,从而消除运动信息之间的冗余,减少了运动信息存储的开销.

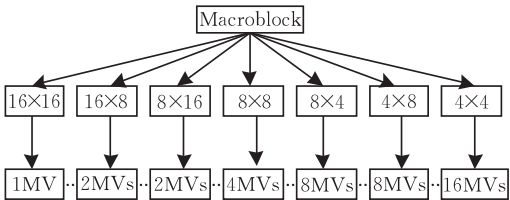


图 1 运动信息描述的分层模型

在预编码过程中,我们可以获取码率和量化参数的对应关系,通过保存这种对应关系,可以实现码率的自适应.在进行快速编码和转码时,可以根据给定的码率和已知的对应关系确定量化参数,并由此量化参数确定最优编码模式,再从运动信息描述中抽取最优的运动矢量.这种方法省略了运动估计的过程,大大降低了编码复杂度,同时,编码性能的损失也很少.

3 基于运动信息描述的视频编码格式自适应框架

运动信息描述包含了源视频丰富的运动信息,可以在一定程度上较精确地反映出源视频的运动特性.因此,无论采用何种视频编码格式,都可以在运动信息描述中找到合适的运动信息,从而省略复杂的运动估计过程,实现快速编码或快速的格式转换.

利用运动信息描述的以上特性,我们提出了一种新的视频格式自适应框架,如图2所示.首先,参考上一节的方法生成基于H.264/AVC标准的运动信息描述.其次,分析H.264/AVC编码的运动信息和其它编码格式运动信息的相关性,从而利用基于

H.264/AVC的运动信息描述生成适应其它格式的运动信息.最后,针对实际的应用环境对其加以利用,来实现快速的视频格式自适应转换.针对原始视频是否可知,实际应用包含两种情况.一方面,在原始视频可知,实际应用包含两种情况.一方面,在原始视频可知,实际应用包含两种情况.一方面,在原始视频可知,实际应用包含两种情况.一方面,在原始视频可知,实际应用包含两种情况.

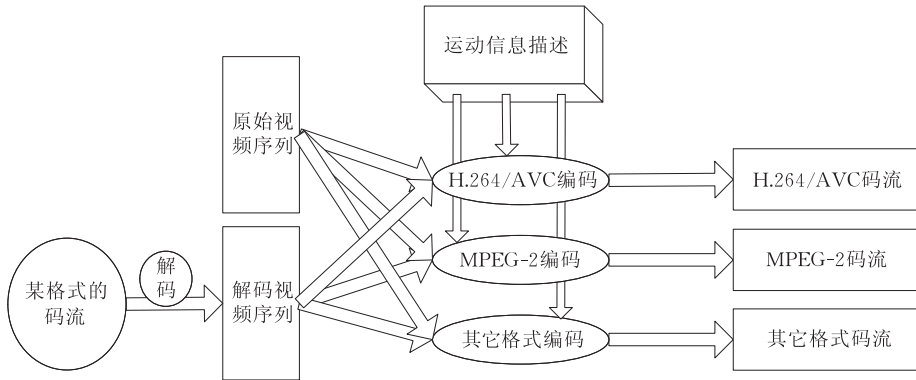


图2 运动信息描述应用于视频格式自适应框架

本文以H.264/AVC和MPEG-2的格式自适应转换为例,详细介绍该框架的实现过程和应用情况.其中,由于运动信息描述的格式遵从于H.264/AVC的编码模式,在进行H.264/AVC的快速编码或MPEG-2到H.264/AVC的格式转换时,不需要做任何附加的处理,直接按照上一节中概述的步骤完成即可.详细的操作以及计算复杂度和编码性能的结果可参考文献[17].但是在进行MPEG-2的快速编码或H.264/AVC到MPEG-2的格式转换时,由于编码模式和运动矢量精确度等等的差异,需要对运动信息描述进行必要的处理再加以利用.在下一节中,本文对存在的问题进行了详细的分析,并给出相应的解决方案.

4 基于运动信息描述的MPEG-2格式自适应算法

本节我们提出一种基于运动信息描述的MPEG-2格式自适应算法,作为本文所提方案的一

个具体应用实例.由于运动信息描述涵盖了多种编码质量下的运动特性,在对其进行充分利用情况下,我们不仅可以快速地编码出不同码率的MPEG-2码流,甚至可以在转码应用中也转成不同码率的MPEG-2码流.在生成码流的过程中,针对码率自适应的需求,可以参考已有的一些MPEG-2的码率控制算法^[18],根据码率来确定量化参数,再从运动信息描述中获取对应该量化参数的最优运动矢量.但是H.264/AVC和MPEG-2两种标准存在很多技术细节上的差异,使得运动信息的获取过程与基于H.264/AVC的方法有所区别.下面我们将分析两种标准运动信息之间的不同,并详细介绍我们的映射算法.

4.1 分块模式的选择

H.264/AVC采用了可变块尺寸这项关键技术来达到提高编码性能的目的.可变块尺寸指每个宏块可以分割为多个子块,其中每个子块拥有各自的运动矢量.H.264/AVC定义了7种分块模式:16×16,16×8,8×16,8×8,8×4,4×8和4×4(如

图 3)。一般来说,较大的分块模式可以用于静止区域和运动平缓的区域,而较小的分块模式则能更好地描述复杂的运动和纹理复杂的区域.但这些复杂的分块方式使得运动估计过程十分繁复,因此整个编码过程复杂度很高.相比而言,MPEG-2 只采用了 16×16 的分块模式,运动估计的过程很简单,编码复杂度低,但是对于运动复杂或纹理复杂的序列不能够取得很好的编码性能.

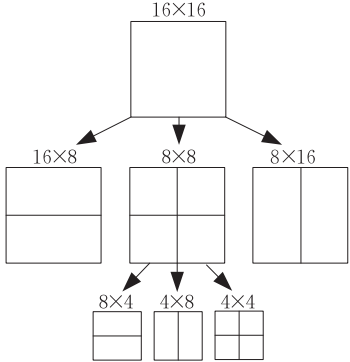
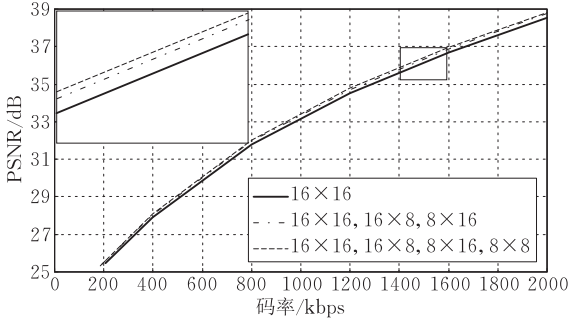


图 3 H.264/AVC 分块模式

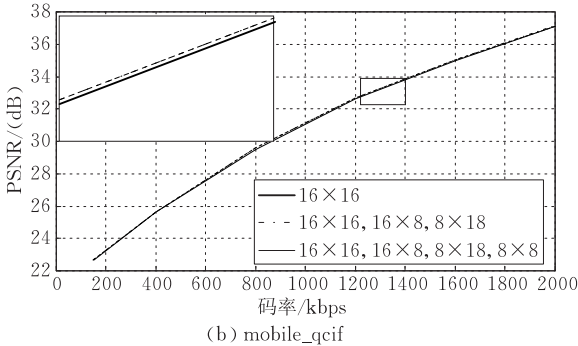
MPEG-2 只采用 16×16 的分块模式来进行运动估计,但是在有些情况下,如纹理复杂的区域或运动分散的区域,视频的运动不具备整块内部的一致性,这时,H.264/AVC 运动估计模块往往选择更小的分块模式,特别是在量化参数很小的时候.传统的格式转码只能依靠这些小分块模式的运动矢量来合成 16×16 的运动矢量.文献[13]中的一系列实验表明,取中值是将小模式运动矢量合成为大模式运动矢量的一种较有效的方法,但是合成的运动矢量与最优运动矢量仍然存在偏差;分块越细,合成的效果越差.而此时,运动信息描述中的一些大的分块模式虽然是在不同的量化参数下生成的,但可能比合成的结果要更接近期望值.因此,在整个运动信息描述中寻找合适的运动矢量比在确定的模式中效果更好.

我们通过实验来衡量不同的模式对寻找 16×16 模式下合适的运动矢量的贡献大小.我们比较了三种方法下利用 H.264/AVC 的运动信息直接合成 MPEG-2 运动信息的编码性能,并给出了序列 Football 和 Mobile 在 QCIF 文件格式和 15Hz 帧率时的结果(见图 4,左上角为图中方块处的放大显示):第一种(用实线代表)是在所有 16×16 模式的运动矢量中选择最优的;第二种(用虚线代表)是加入 $16\times 8/8\times 16$ 模式的运动矢量的中值作为候选运动矢量;最后一种方法(用点划线)在第二种的基础

上再加上 8×8 模式的运动矢量的中值,其中每个 8×8 子块根据量化参数来确定进一步的分块模式.测试结果显示,第三种方法的曲线几乎和第二种重合,即 8×8 以及更小的模式运动矢量的加入对编码性能的提高几乎没有帮助,这说明这些小模式的运动矢量过于分散,导致合成的 16×16 模式的运动矢量很不精确.而 $16\times 8/8\times 16$ 模式的加入可以在一定程度上提高编码性能.



(a) football_qcif



(b) mobile_qcif

图 4 采用不同模式的运动矢量来进行 MPEG-2 编码的性能比较

经过以上分析,运动信息描述至 MPEG-2 运动矢量的映射算法不是先根据量化参数确定模式,再从选定模式下的几个候选运动矢量中选择最优者,而是对于任何量化参数都是将运动信息描述中 16×16 或者 $16\times 8/8\times 16$ 模式的运动矢量作为候选,更小模式的信息则不加利用.

4.2 运动矢量代价函数的计算

另一项使 H.264/AVC 编码性能提高的关键技术就是率失真优化,根据拉格朗日代价函数(1),确定每个宏块最优的模式和运动矢量.式中 D 是失真, R 是编码运动信息和残差所需要的比特数, λ 是拉格朗日乘子,由量化参数唯一确定^[19].

$$J = D + \lambda \times R \tag{1}$$

MPEG-2 没有这些复杂的准则,而是直接根据运动搜索之后的绝对误差和 SAD(Sum of Absolute Difference)大小来选取运动矢量,不考虑编码比特

数. 我们的运动信息描述在生成的时候是根据式(1),在选取的时候如果只是根据 SAD 就可能存在误差,从而不能够真正地选出最优的运动矢量. 因此,需要有一种比 SAD 更有效的衡量准则,但是完全按照式(1)的计算要增加过多的计算复杂度. 在综合考虑编码复杂度和 MPEG-2 编码模式比较单一等特性的前提下,我们提出一种简单的 J 计算方法来作为选择运动矢量的代价函数,计算仍按照式(1),但此时 D 为 SAD 值, R 为编码运动信息所需要的比特数. 在这里只是忽略了残差的影响,但这个影响相比于模式差异带来的影响几乎可以忽略.

4.3 运动矢量精度的转化

为了达到更高的编码性能, H.264/AVC 相比于其它的标准在运动矢量计算的一些细节上做了改进. 一方面, H.264/AVC 的运动矢量精确到了 1/4 像素级, 另一方面, H.264/AVC 采用了边界延拓, 使得边界块也有完整的搜索范围, 这使得运动估计更加精细更加完整, 在复杂的编码模式前提下进一步提高了编码性能. 但是 MPEG-2 只将运动矢量精确到 1/2 像素, 也没有进行边界延拓. 因为这些区别, 由 H.264/AVC 生成的运动信息描述的运动矢量经过了模式合并之后仍然要进行以下处理才能作为 MPEG-2 的运动矢量来使用.

图 5 显示了 1/4 像素和 1/2 像素运动矢量的具体关系, \times 为 1/4 像素运动矢量的终点, \circ 为 1/2 运动矢量终点, \bullet 为最终选择的 1/2 像素点. 1/4 像素运动矢量投影到 1/2 像素运动矢量域中可能刚好有对应点(如图 5(a)), 也可能没有对应完整的点(如图 5(b), (c), (d)), 因此, 我们建议对于与 1/4 像素运动矢量距离最近的几个 1/2 运动矢量分别计算相应的 J , 取能最小化 J 值的运动矢量作为最终结果.

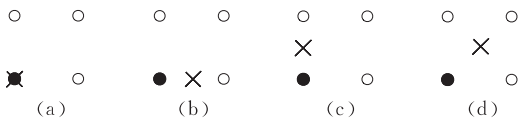


图 5 运动矢量精度转换:1/4 像素转化为 1/2 像素

而对于从运动信息描述中得到的因为边界延拓越出了图像区域的运动矢量, 我们通过边界截取直接强制将其约束在图像区域内.

4.4 阈值的计算

由于部分模式的运动矢量在运动信息合并过程中被舍弃, 因此映射算法搜索的运动信息并不完备, 这会导致在某些特殊情况下, 特别是运动复杂的情况下, 通过映射算法提取出的运动信息有效性很差.

所以, 我们给定一个阈值来衡量提取的运动信息的有效性.

图 6 给出了 Bus 序列的运动矢量分布图, 我们可以看到, 运动矢量的分布存在几个峰值, 其中最重要的峰值是 (0, 0), 这表明了零运动矢量的出现概率很高. 另外, 根据运动的空间相关性, 任一宏块的运动可以通过周围宏块的运动大致推断出来. 这样就可以在零运动矢量和根据 H.264/AVC 的空间预测算法得到的运动矢量两者中选取能最小化 J 的值作为最终的预测运动矢量, 而相应的 J 值就是所需的阈值.

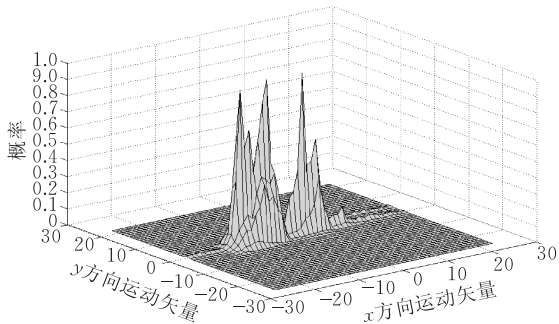


图 6 Bus 序列的运动矢量分布图

4.5 运动矢量映射算法

经过以上的模式选择、运动矢量精度换算以及阈值确定之后, H.264/AVC 运动信息描述到 MPEG-2 运动矢量的投影算法就可以确定了, 具体流程如图 7 所示. 在这里 vop 为要输出的最优运动矢量 (MV) 对应的 J 值, 初始值设为预测运动矢量 (MV_pred) 对应的 J 值 (vop). 对于任何量化参数的 MPEG-2 编码都可以按照此算法来快速地获取合适地运动矢量.

算法具体步骤如下.

1. 按顺序读取 16×16 模式中的一个 1/4 像素精度的 MV , 直接除以 2 取整作为候选 MV , 计算运动补偿后对应的 J . 如果 $J < vop$, 则 $vop = J$, 直接进入步 4; 如果 $J \geq vop$ 而且 16×16 模式下还有其它没有读到的 MV , 则重复步 1; 否则, 进入步 2.
2. 按顺序读取 16×8 模式中的一个 1/4 像素精度的 MV , 直接除以 2 取整作为候选 MV , 计算运动补偿后对应的 J . 如果 $J < vop$, 则 $vop = J$, 直接进入步 3; 如果 $J \geq vop$ 而且 16×8 模式下还有其它没有读到的 MV , 则重复步 2; 否则, 进入步 3.
3. 按顺序读取 8×16 模式中的一个 1/4 像素精度的 MV , 直接除以 2 取整作为候选 MV , 计算运动补偿后对应的 J . 如果 $J < vop$, 则 $vop = J$, 直接进入步 4; 如果 $J \geq vop$ 而且 8×16 模式下还有其它没有读到的 MV , 则重复步 3; 否则, 进入步 4.

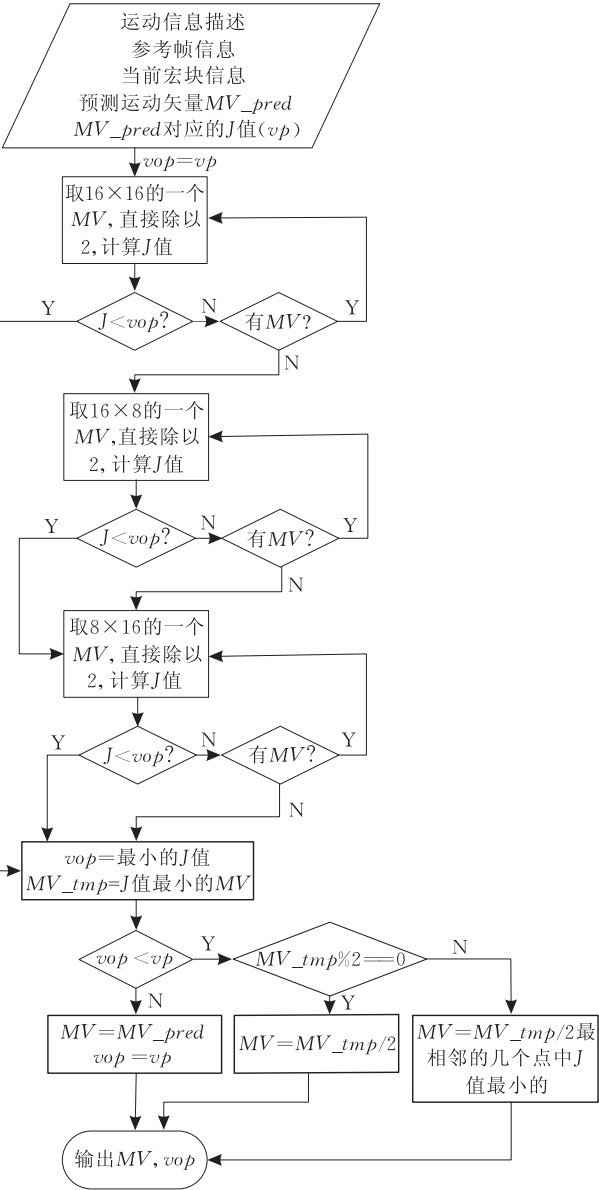


图 7 运动矢量计算步骤

4. 如果 $vop \geq vp$, 则 $MV = MV_pred$, $vop = vp$, 输出 MV 和 vop ; 如果 $vop < vp$, 则由以上步骤得到的对应最小 J 的 $1/4$ 像素 MV 计算得到 $1/2$ 像素 MV , 具体过程如 4.3 节中所述, 输出 MV 和 vop .

5 实验结果

为了验证本文所提方法的有效性, 我们依然以 H.264/AVC 和 MPEG-2 的格式自适应转换为例, 进行一系列实验. 运动信息描述的生成采用 H.264/AVC 编码标准, 其中量化参数变化范围为 $20 \sim 40$, 因此快速编码和转码的应用可以覆盖很广的码率范围. 对于 H.264/AVC 码流的快速生成和快速格式转换, 文献[17]给出了详细的实验结果, 结果显示其

损失不超过 0.3dB. 因此, 本节的实验主要验证 MPEG-2 的快速编码和 H.264/AVC 到 MPEG-2 的快速格式转换的编码性能和计算复杂度. 6 个标准序列 News, Paris, Mobile, Bus, Foreman 和 Football 用于本次测试. 其中, News 和 Paris 运动相对平缓; Mobile 有规则的水平 and 垂直运动; Bus 主要是水平运动但相对比较剧烈; Foreman 则具有丰富的脸部旋转运动; 而 Football 的运动则复杂零乱而且很剧烈. 用这些具有不同运动特性的序列作为测试样本, 可以更充分验证算法的有效性. 这些序列都是 CIF 大小, 输入帧率为 30Hz, 这种格式普遍适用于网络传输和电视传播. 实验采用 IPPPPP... 格式的编码, 每帧只允许参考前一帧, 取每个序列的前 75 帧作为样本.

5.1 基于运动信息描述的 MPEG-2 快速编码

我们将基于运动信息描述进行 MPEG-2 快速编码的方法, 与直接用 MPEG-2 全搜索编码进行比较, 搜索窗口尺寸设为 11. 图 8 分别显示了 6 个序列编码性能的比较结果. 由图可见, 在非常大的码率范围内, 运动平缓的序列性能基本上没有损失, 而对于运动复杂或剧烈的序列性能会有一些损失, 但这些损失也控制在 0.3dB 之内.

在 MPEG-2 编码过程中, 运动估计占用了绝大部分的编码时间. 而在运动估计中, 最耗时的操作就是搜索运动矢量时 SAD 的计算. 因此, 我们用平均每个宏块使用 SAD 的操作次数来衡量不同方法的计算复杂度. 表 1 给出了基于运动信息描述编码与 MPEG-2 编码的计算复杂度的比较结果. 从表中可以看出, 我们的方法将 SAD 操作的次数降到了 5 以下, 而全搜索的编码在搜索窗口为 11 时 SAD 操作的次数达到 400 以上, 而随着搜索窗口变大计算次数也会剧烈增加.

表 1 基于运动信息描述与全搜索的编码复杂度比较

单位: SAD 次数/宏块

序列	复杂度	
	运动信息描述编码	全搜索编码
News	1.41	482.99
Paris	1.89	482.99
Mobile	3.69	482.99
Bus	3.89	483.17
Foreman	3.59	483.07
Football	4.23	483.01

5.2 基于运动信息描述的 H.264/AVC 到 MPEG-2 格式转换

这里我们将基于运动信息描述进行 H.264/AVC 到 MPEG-2 格式转换的方法, 与两种传统的

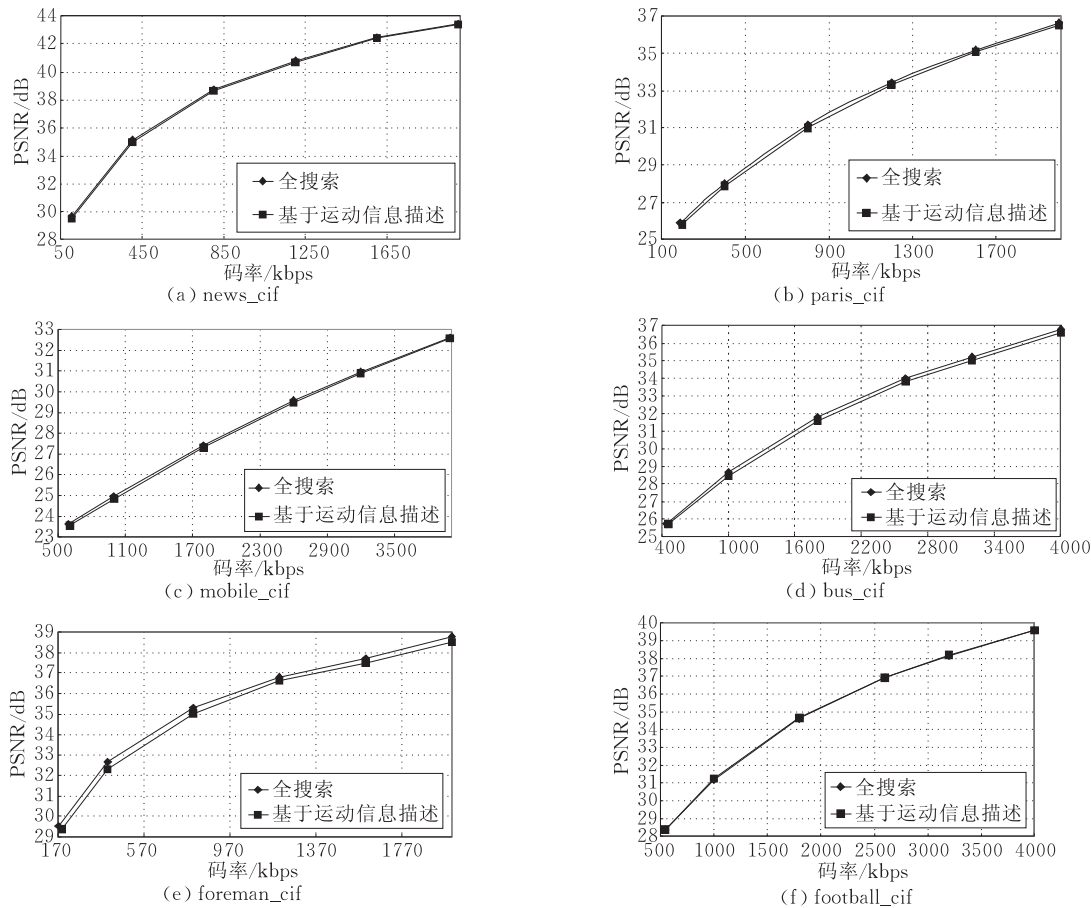


图 8 基于运动信息描述的快速编码与 MPEG-2 全搜索编码性能比较

转码方法进行比较. 一种方法是全解全编, 即先将码流解码再对解码视频直接用 MPEG-2 编码器进行编码, 编码过程则要通过全搜索来得到运动矢量; 另一种方法就是文献[13]中所述的利用 H.264/AVC 码流中的运动矢量经过简单的换算来形成 MPEG-2 的运动矢量, 在编码过程中省略了运动估计的过程; 而本文的方法是通过上述的方法利用运动信息描述来生成 MPEG-2 的运动矢量, 同样也避免了运动估计的过程. 输入的高质量视频是由 H.264/AVC 编码器在量化参数为 20 的码流解码得到的重建序列. 图 9 分别显示了 6 个序列编码性能的比较结果. 结果显示, 我们的方法在性能上没有很大的损失, 一般都低于 0.3dB, 对于运动复杂的 football 序列其损失也小于 0.5dB; 而简单的运动信息复用的编码性能会损失 1~3dB^[13].

下面给出这三种方法的计算复杂度的比较, 如表 2 所示, 全解全编的方法虽然编码效率最高, 但是计算复杂度也最高, 而且是其他两种方法的几百倍. 而运动矢量复用的方法虽然计算复杂度很低, 但性能的损失也非常严重. 相比而言, 本文提出的方法

不仅计算复杂度低, 而且编码性能也保持的很好.

表 2 3 种转换方法的计算复杂度比较
单位: SAD 次数/宏块

序列	复杂度		
	运动信息描述	运动矢量复用	全解全编
News	1.41	2.25	482.99
Paris	1.88	2.48	482.99
Mobile	3.69	3.32	482.99
Bus	3.89	3.84	483.17
Foreman	3.54	4.18	483.07
Football	4.24	2.46	483.01

6 结 论

本文提出了一种新颖的通用的格式自适应转换框架, 该框架利用预编码生成的运动信息描述来实现多种视频格式之间的快速转换. 文中以 H.264/AVC 和 MPEG-2 的自适应转换为例, 展示了该框架的具体实现过程, 给出了详尽的运动信息映射算法. 文章通过实验对本文方法与传统转码方法做了对比, 实验结果验证了我们所提出的方法在快速编码和快速格式转码的应用中保持了良好的编码性能

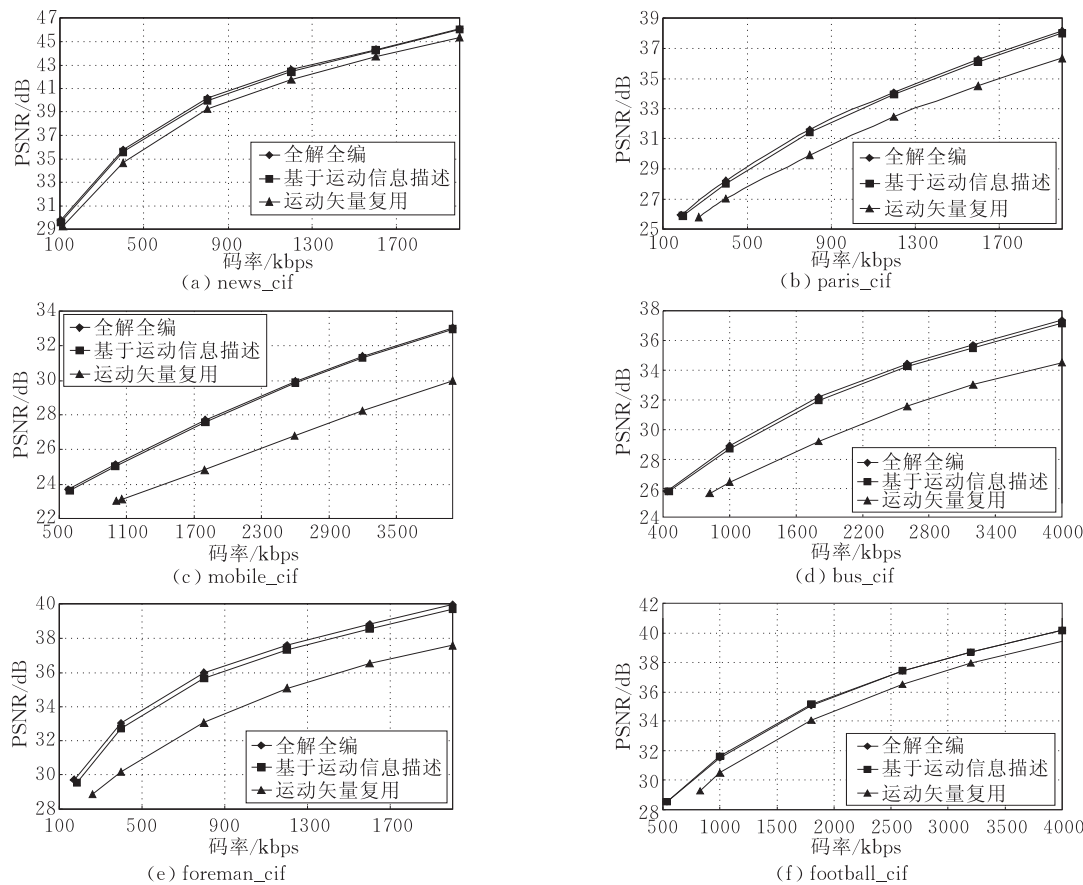


图 9 3 种 H.264/AVC 到 MPEG-2 格式转码方法的性能比较

的同时,大大降低了计算复杂度。

本文采用的基于 H.264/AVC 标准的运动信息描述不仅可以应用于视频格式自适应、码率自适应,还可以应用于视频时间分辨率自适应、空间分辨率自适应等其他多种视频自适应技术,这是我们目前正在研究的工作;我们的目标是研究基于视频信号描述的统一普适的视频自适应技术框架,这是我们今后的研究重点。

参 考 文 献

- [1] Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 14 496-10 AVC). Joint Video Team(JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-G050, 2003
- [2] Video codec for audiovisual services at px64 bit/s ITU-T recommendation H.261, Version 1. ITU-T, ITU-T Recommendation H.261 Version 1, 1990
- [3] Video coding for low bit rate communication. ITU-T, ITU-T Recommendation H.263 Version 1, 1995
- [4] Generic coding of moving pictures and associated audio information—Part 2: Video. ITU-T and ISO/IEC JTC 1, ITU-T Recommendation H.262 and ISO/IEC 13 818-2(MPEG-2), 1994
- [5] Coding of audio-visual objects—Part 2: Visual. ISO/IEC 14 496-2(MPEG-4 Visual Version 1), Apr. 1999
- [6] Chang S, Vetro A. Video adaptation: Concepts, technologies, and open issues. *Proceeding of the IEEE*, 2005, 93(1): 148-158
- [7] Vetro A, Christopoulos C, Sun H. Video transcoding architectures and techniques: An overview. *IEEE Signal Processing Magazine*, 2003, 20(2): 18-29
- [8] Ahmad L, Wei X, Sun Y, Zhang Y. Video transcoding: An overview of various techniques and research issues. *Multimedia, IEEE Transactions*, 2005, 7(5): 793-804
- [9] Xin J, Lin C, Sun M. Digital video transcoding. *Proceedings of the IEEE*, 2005, 93(1): 84-97
- [10] Chen G, Zhang Y, Lin S, Dai F. Efficient block size selection for MPEG-2 to H.264 transcoding//*Proceedings of the 12th Annual ACM International Conference on Multimedia*. New York, USA, 2004: 300-303
- [11] Zhou Z, Sun S, Lei S, Sun M T. Motion information and coding mode reuse for MPEG-2 to H.264 transcoding//*Proceedings of the ISCAS2005*. Kobe, Japan, 2005: 1230-1233
- [12] Lu X, Tourapis A, Yin P, Boyce J. Fast mode decision and motion estimation for H.264 with a focus on MPEG-2/H.264 transcoding//*Proceedings of the ISCAS2005*. Kobe, Japan, 2005: 1246-1249
- [13] Chu J, Lu W, Liu Y, So Y. H.264/MPEG-2 transcoding

based on personal video recorder platform//Proceedings of the 9th International Symposium. Macau, China, 2005: 438-440

- [14] Xin J, Vetro A, Sun H. Converting DCT coefficients to H.264/AVC transform coefficients//Proceedings of the Pacificrim Conference on Multimedia (PCM). Tokyo, Japan, 2004: 939-946
- [15] Sun Y, Xin J, Vetro A, Sun H. Efficient MPEG-2 to H.264/AVC intra trnascoding in transform domain//Proceedings of the ISCAS2005. Tokyo, Japan, 2005: 1234-1237
- [16] Tang Q, Nasiopoulos P, Ward R. An efficient re-quantization error compensation for MPEG-2 to H.264 transcoding//Proceedings of the 2006 IEEE International Symposium on

Signal Processing and Information Technology. Vancouver, Canada, 2006, 530-535.

- [17] Wang Y, Sun X, Wu F, Li S, Li H, Liu Z. Off-line motion description for fast video stream generation in MPEG-4 AVC/H.264//Proceedings of the International Conference on Multimedia and Exposition. Torondo, Canada, 2006: 685-688
- [18] Test model 5. ISO/IEC JTC1/SC29/WG11/No 400, MPEG-93/457, Apr. 2003
- [19] Ma S, Gao W, Lu Y. Rate-distortion analysis for H.264/AVC video coding and its application to rate control. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(12): 1533-1544



SHEN Qiu, born in 1982, Ph. D. candidate. Her research interests include video coding and video adaptation.

WANG Yi, born in 1979, Ph. D. His research interests include video coding and video analysis.

LI Hou-Qiang, born in 1969, Ph. D. , professor. His research interests include video coding, image processing and computer vision.

HE Jia, born in 1976, Ph. D. His research interests including image processing and video surveillance.

Background

The work of this paper is supported by National Natural Science Foundation of China (NSFC) Key Program under contract No. 60632040, NSFC General Program under contract No. 60672161.

Currently, there exist various video coding standards in multimedia applications, including ITU-T H.26x, ISO/IEC MPEG-x and national standardization initiatives, for example AVS codec standardized by Audio and Video Coding Standard Workgroup in China. These standards play an important role in their respective fields, but cannot be directly applied to other fields. To adapt the application environment with multiple standards, video signal should be coded into several different formats, which would bring a great computational complexity and consume a large amount of storage capacity. Format transcoding tries to transform a video bit-stream from one format to another in an effective and rapid manner. Generally speaking, conventional method should make trade-off between computational complexity and coding efficiency, meanwhile different algorithms should be designed distinctively for different format transcoding, which bring a great burden to real-time transcoding system.

This paper attempts to give a novel solution of realizing format adaptation in a general framework. The basis of the

framework is an intermediate motion description of video, which is obtained by pre-encoding the original video signal. As the description can provides abundant motion information of the original video signal, while encoding or transcoding a video signal into one format, the complex process of motion estimation can be substituted by extracting motion information from the description with a simple process. Considering H.264/AVC and MPEG-2 are two widely used video coding format, this paper illustrates format adaptation between these two formats in detail. Furthermore, a mapping algorithm of motion information required for MPEG-2 encoding and/or transcoding is described particularly. Experimental results demonstrate that the proposed solution can support video format adaptation rapidly while maintaining coding efficiency.

In the future, more mapping algorithms of motion information should be designed to support more format adaptation. Also, the intermediate description can be enhanced by importing new features of video signal, so that multiple video adaptations rather than format adaptation can be supported, for example bit-rate adaptation, temporal and/or spatial resolution adoption, etc.