

# 基于感知运动能量模型的 MPEG-21数字项适配\*)

赵 刚 杨宗凯 何建华

(华中科技大学电子与信息工程系 武汉430074)

**摘 要** 最新的 MPEG-21 标准定义了一个统一的多媒体框架,其目标是使多媒体资源能在异构网络和终端上进行透明传输和得到充分利用。多媒体适配(Multimedia Adaptation)技术作为实现上述目标的关键技术得到了越来越多的研究。本文将感知运动能量模型与 MPEG-21 数字项适配框架相结合,研究和实现了基于感知运动能量的 H. 264/AVC 视频丢帧适配技术。本文的研究工作有两个方面的优点:一方面,感知运动能量模型的使用将丢帧适配引起的运动抖动减至最小,保证了较高的用户视觉上的主观感知质量;另一方面,数据处理能力和资源受限的适配节点不需了解具体的 H. 264/AVC 编码语法结构和编解码算法,就能方便地执行丢帧操作,实现了独立于媒体编码格式的视频适配机制,提高了多媒体适配系统的灵活性和可扩展性。

**关键词** MPEG-21 数字项适配, H. 264/AVC 视频, 感知运动能量模型, 通用比特流语法描述

## MPEG-21 Digital Item Adaptation by Applying Perceived Motion Energy to H. 264/AVC Video

ZHAO Gang YANG Zong-Kai HE Jian-Hua

(Department of Electronics & Information Engineering, Huazhong University of Science & Technology, Wuhan 430074)

**Abstract** The new MPEG-21 standard defines a multimedia framework to enable transparent and augmented use of multimedia resources across heterogeneous networks and devices used by different communities. In this paper, we incorporate the Perceived Motion Energy (PME) Model into the proposed MPEG-21 Digital Item Adaptation Framework for frame dropping in H. 264 encoded video adaptation. There are two advantages of this work, one is the use of PME model to reduce the viewer's perceived motion jitter due to frame dropping to a minimum. The other is the adaptation nodes can easily apply frame dropping operations without knowledge of detailed encoding syntax of H. 264 videos.

**Keywords** MPEG-21 digital item adaptation, H. 264/AVC video, Perceived motion energy (PME) model, Generic bitstream syntax description (gBSD)

## 1 引言

随着多媒体技术的迅猛发展,人们对多媒体内容的需求不断增加,各种多媒体资源呈爆炸性的增长。同时,各种多媒体终端接入设备和传输网络不断涌现,使人们在自由使用多媒体信息方面有了更多的选择。然而,不同的多媒体内容编码格式、网络传输条件以及的终端设备处理和显示能力,给自由和方便地传输、使用多媒体资源带来了极大的困难。为实现任何地点、任何时间通过网络获取任意多媒体信息资源的目标,多媒体适配(Multimedia Adaptation)技术得到了越来越多的重视,成为国内外的研究热点之一。所谓适配,就是通过改变多媒体内容来适应不同的网络传输条件、终端设备处理和显示能力以及用户的偏好,例如降低视频传输码率以适应无线网络的低带宽特点,降低图像分辨率或视频编码帧率来适应处理和显示能力相对较弱的移动终端上的多媒体应用等。

在各种视频适配技术中,丢帧(Dropping Frame)是一种简单、有效的码率适配方法,即通过丢弃视频序列中一定数量的编码帧来降低视频的编码速率或帧率。但是,丢帧会造成视频回放过程中的运动抖动(Motion Jitter)现象,降低用户视觉上的主观感知质量。一般来说,运动抖动的大小在很大程度上

上取决于视频序列中的各种运动模式。为了有效降低运动抖动,有必要对视频中的运动特征进行相应的分析,找出对用户的主观感知质量影响较小的编码帧,作为丢帧操作的候选帧,从而最大限度地保证视频重建质量。

另一方面,执行多媒体适配操作的网络节点(如网关、代理、路由器等),其数据处理能力和计算资源通常十分有限,执行比较复杂的适配操作(如需进行重新编、解码的编码转换)受到一定的限制,很难支持所有的视频编码格式,尤其是新的视频编码标准,如 H. 264/AVC 标准。最新出现的 MPEG-21 多媒体框架标准的第七部分数字项适配 DIA(Digital Item Adaptation),定义了通用比特流语法描述 gBSD(Generic Bitstream Syntax Description)工具,提供了独立于媒体编码格式的通用多媒体资源适配机制。这一描述工具不仅可以对视频序列的数据结构进行描述,也能够用来表达和传递如运动模式分析结果等附加信息。与视频相关的 gBSD 将被传输到适配节点,用来协助其进行适配决策和执行适配操作,保证适配后的视频具有较高的重建质量。

本文首先使用感知运动能量模型 PME(Perceived Motion Energy)<sup>[1,2]</sup>来分析 H. 264/AVC 编码视频中不同的编码帧对于用户主观感知质量的重要性,并将此模型与 MPEG-21

\*)本文得到国家自然科学基金项目(编号60202005)资助。赵 刚 博士研究生,主要研究方向为多媒体处理、无线多媒体传输、远程教育,杨宗凯 教授,主要研究方向为远程教育、计算机网络。

DIA 框架相结合,通过通用比特流描述符来描述 H. 264/AVC 编码视频和表达 PME 分析结果,并将 gBSD 传输到网络中的适配节点上,在适配节点上的 gBSD 多媒体资源适配引擎将对 gBSD 进行解析,并根据提取的 PME 分析结果,丢弃不重要的编码帧来适应网络带宽或终端设备能力。值得指出的是,本文中的丢帧仅限于丢弃双向预测帧(B 帧),一方面因为丢弃 B 帧所需操作比较简单和实用,不会如丢弃 I、P 帧引起后续帧甚至整个 GOP(Group of Picture)的丢弃;另一方面,B 帧的数目通常占整个视频序列中总帧数的三分之二,编码码率占总编码码率的 30%~40%,丢弃一定比例的 B 帧可以满足一定的码率适配要求。

本文的研究工作有两个方面的优点:一方面,感知运动能量模型的使用将丢帧适配引起的运动抖动减至最小,保证了较高的用户视觉主观感知质量;另一方面,数据处理能力和计算资源受限的适配节点不需了解具体的 H. 264/AVC 码流语法结构和编解码算法,就能方便、高效地执行丢帧操作,实现了独立于媒体编码格式的视频适配,提高了多媒体适配系统的灵活性和可扩展性。

## 2 感知运动能量模型

在已有的研究文献中,人们提出了多种通过分析视频运动特征来进行视频检索的方法,如基于主导运动的方法<sup>[3]</sup>、基于摄像机运动的方法<sup>[4]</sup>等。然而,这些方法并不能为普通用户的视频应用和服务提供足够多的运动特征信息,也不适合于视频适配的应用场景。因此,我们使用感知运动能量(PME)模型来分析和评估由于丢帧操作所引起的运动抖动。下面我们将首先简要介绍一下 PME 模型,然后将给出针对 H. 264/AVC 编码视频的 PME 提取算法。

### 2.1 感知运动能量

在压缩视频流中,视频运动信息通常以运动矢量来加以表达。然而,运动矢量仅仅是真实的光流场(real optical flows)粗略和不连续的近似,其可靠性在一定程度上受到噪声的影响。一般来说,运动矢量的幅度比其方向对于噪声更加敏感。因此,要有效利用运动矢量信息来分析视频序列的运动特征,需要使用滤波器来对运动矢量的幅度进行滤波,以消除噪声影响。

首先,采用空间中值滤波器(Median Filter)对运动矢量的幅度进行滤波,定义为中值滤波器窗口的宽度,滤波后的运动矢量幅度可以通过(1)式计算得到:

$$Mag_{i,j} = \begin{cases} Mag_{i,j} & (\text{if } Mag_{i,j} \leq Max4th(Mag_k)) \\ Mag4th(Mag_{i,j}) & (\text{if } Mag_{i,j} > Max4th(Mag_k)) \end{cases} \quad (1)$$

其中  $Mag_{i,j}$  为宏块  $MB_{i,j}$  处运动矢量的幅度值,函数  $Max4th(Mag_k)$  返回滤波器窗口  $W_i$  内以降序排列的运动矢量幅度值序列中的第四个元素值。

经过中值滤波后的运动矢量幅度再通过一个时间域的能量滤波器来进一步消除噪声,可以采用如三维空时跟踪卷(3-D spatial-temporal tracking volume)内的截断率(alpha-trimmed)滤波器。设三维空时跟踪卷的空间窗口大小定义为  $W_i$ ,其持续时间长度为  $L_i$ ,则包括对象运动能量和摄像机运动能量的混合运动能量  $MixEn_{i,j}$  的计算式为:

$$MixEn_{i,j} = \frac{1}{(M-2 \lfloor aM \rfloor) \cdot W_i^2} \sum_{m=\lfloor aM \rfloor+1}^{M-\lfloor aM \rfloor} Mag_{i,j}(m) \quad (2)$$

式中  $M$  为跟踪卷中的幅度值元素总数, $\lfloor aM \rfloor$  为不大于  $aM$

的最大整数, $Mag_{i,j}(m)$  为跟踪卷内幅度值序列的元素值,截断率参数  $a(0 \leq a \leq 0.5)$  则决定了被排除在运动能量累积计算过程以外的数据样本数目。

对于整个视频帧来说,可以使用该帧内所有宏块的运动矢量的平均幅度值  $Mag(n)$  来评估其运动强度,P 帧和 B 帧内的  $Mag(n)$  分别通过(3)式和(4)式计算而得:

$$Mag(n) = \sum MixFEn_{i,j}(n) / N \quad (3)$$

$$Mag(n) = (\sum MixFEn_{i,j}(n) / N + \sum MixBEn_{i,j}(n) / N) / 2 \quad (4)$$

式中  $MixFEn_{i,j}(n)$  和  $MixBEn_{i,j}(n)$  分别代表第  $n$  帧内的前向运动矢量和后向运动矢量的混合运动能量,为帧内的宏块总数。

由于人眼更倾向于跟踪和注意视频场景中的主导运动(dominant motion),因此在模型中引入主导运动方向的百分比  $\alpha(n)$  来更好地表征相关运动特性, $\alpha(n)$  可以由(5)式计算:

$$\alpha(n) = \frac{\max(AH(n,k), k \in [1, m])}{\sum_{k=1}^m AH(n,k)} \quad (5)$$

式(5)中角度空间  $2\pi$  被划分为  $m$  个角度区间,定义  $AH(n,k)$  为角度直方图,则  $\max(AH(n,k), k \in [1, m])$  即为落入主导运动方向区间的角度数目。

感知运动能量 PME 定义为运动矢量的平均幅度值与主导运动方向百分比的乘积:

$$PME(n) = Mag(n) \times \alpha(n) \quad (6)$$

其中  $Mag(n)$  反映了丢弃低运动强度的帧比丢弃高运动强度的帧较不容易被人眼察觉的事实。当视频序列中存在摄像机运动或其他主导对象运动时, $\alpha(n)$  将使得 PME 值有一个较大幅度的增加,这也反映出人眼对主导运动的敏感性。

### 2.2 H. 264/AVC 视频 PME 提取算法

本文中我们首次将 PME 模型运用到新的视频编码标准 H. 264/AVC 中。由于 H. 264/AVC 编码标准中引入了各种新的编码技术和方法,对感知运动能量的特征提取和分析产生了一定的影响,因此我们针对传统 MPEG 视频的 PME 提取算法进行了相应的修改和补充,提出了 H. 264/AVC 视频的 PME 提取算法。

H. 264/AVC 标准中一个新的编码方法是编码块大小可变的运动预测和补偿,  $16 \times 16$ 、 $8 \times 16$ 、 $16 \times 8$ 、 $8 \times 8$ 、 $4 \times 8$ 、 $8 \times 4$  和  $4 \times 4$  大小的编码块都可能拥有与之相对应的运动矢量。所以,传统基于宏块( $16 \times 16$ )级别的 PME 提取方法就显得不适合了。基于此,我们提出的 PME 提取算法则是基于最小的编码块级别—— $4 \times 4$  块上进行的,对于本身无对应运动矢量信息的  $4 \times 4$  块,取其运动矢量为包含此  $4 \times 4$  块的大级别编码块(如  $16 \times 16$ 、 $8 \times 16$ 、 $16 \times 8$  等)的运动矢量。

H. 264/AVC 标准中另一个非常显著的新特征就是多参考帧的运动预测和补偿。例如,某一编码帧可能拥有最多 5 个预测参考帧。如果某一参考帧与当前帧之间的距离较远,我们发现在基于该参考帧上的运动矢量对感知运动能量的贡献将比其他较近的参考帧上的运动矢量小,这是因为人眼通常对最近发生的运动抖动比较敏感。因此,我们基于参考帧与当前帧的距离为各个  $4 \times 4$  块的 PME 值进行加权,并以  $MixFEn_{i,j}$  来替代式(3)、(4)中的  $MixFEn_{i,j}(n)$ ,  $\overline{MixFEn_{i,j}}$  可由(7)式计算得到:

$$\overline{MixFEn_{i,j}(n)} = \beta(d) \cdot MixFEn_{i,j}(n) \quad (7)$$

式(7)中的  $\beta(d)$  为  $4 \times 4$  块  $(i, j)$  的加权系数,它与该编码块使

用的预测参考帧和当前帧之间的距离  $d$  有关。为简单起见，我们可以定义  $\beta(d)$  为：

$$\beta(d) = \frac{\lambda}{d} \quad (8)$$

在式(8)中， $\lambda$  为一平衡常数，它是用来避免对具有不同 I、P、B 编码模式的视频序列进行 PME 加权时，可能引起的 PME 值急剧下降的效果。

此外，在 H. 264/AVC 编码中允许多种不同的编码块预测模式，如帧内预测 (Intra Prediction)、前向帧间预测 (Forward Prediction)、后向帧间预测 (Backward Prediction) 和双向帧间预测 (Bi-direction Prediction)。这些不同的预测模式对运动抖动的产生及其影响需要仔细地加以研究。通常，帧内预测块编码的是新的信息，该块应被视为具有比较大的运动量。由于帧内预测块不拥有实际的运动矢量，因此在 PME 提取算法中我们假设其运动矢量的大小与最大的运动预测搜索范围相等，例如(16,16)。由于视频帧播放顺序的关系，我们发现前向预测块的丢失往往比后向预测块的丢失引起的运动抖动更大，因此必须对前向预测块和后向预测块的正负计算进行区别对待，即为前向预测运动矢量赋予较大的权值，而为后向运动矢量赋予较小的权值。我们可以简单地定义不同块预测模式的权值  $\gamma(t)$  为：

$$\gamma(t) \in \{1, 1, \frac{1}{2}, \frac{1}{4}\} \quad (9)$$

式(9)表示对于帧内预测、前向帧间预测、后向帧间预测、双向帧间预测模式， $\gamma(t)$  将分别取值为 1、1、1/2、1/4。

由(7)、(8)、(9)式，我们可以将式(3)、(4)分别改写为 H. 264/AVC 视频 PME 计算式：

$$\overline{Mag(n)} = \sum d_i \cdot \frac{\lambda}{d} \cdot \gamma(t) \cdot MixEn_i(n) / N \quad (10)$$

式(10)中的  $d_i$  为一常数，由于前向帧间预测模式中的多帧预测加权可能引起实际 PME 值的大幅度减少，为使前向帧间预测模式拥有可以与其他预测模式相比较的 PME 值，因此使用  $d_i$  来作为平衡因子。

### 3 MPEG-21数字项适配

最新的 MPEG-21 标准定义了一个统一的多媒体框架，来实现使多媒体资源在异构网络和终端上进行透明传输，并得到充分利用的最终目标<sup>[6]</sup>。在这一多媒体框架中，分发和交易过程中的基本单元被称为数字项 DI (Digital Item)，它是具有标准的表达、标记和相关元数据的数据化数字对象。在实际应用中，数字项是资源、元数据信息和结构信息的组合。资源主要指各种多媒体资源，如文本、图片、音频、视频；元数据信息是描述整个数字项实体或数字项内包含的资源的相关信息；而结构信息则说明了数字项实体的组成结构以及各组成部分之间的关系，如资源和元数据的关系。

为了更好地实现跨越多种网络和终端对多媒体资源的透明利用，MPEG-21 标准针对多媒体资源的适配问题，特别制定了第七部分——数字项适配 DIA (Digital Item Adaptation)。这一部分定义和集成了多媒体资源适配所需的各种工具及技术，主要包括三个方面的内容：应用环境描述工具 UED (Usage Environment Description Tools)、数字项资源适配工具 (Digital Item Resource Adaptation Tools)、数字项声明适配工具 (Digital Item Declaration Adaptation Tools)。

在数字项资源适配工具中一个重要的工具就是比特流语法描述工具，它用来实现以不依赖具体媒体格式，即媒体格式

独立 (format-independent) 的方式来进行多媒体比特流的适配机制。尤其是当执行数字项资源适配操作的网络节点不清楚特定的多媒体资源编码格式时，比特流语法描述工具显得极其重要和有效。其基本思想是先通过 XML 技术对多媒体资源比特流的结构或语义进行描述，随后多媒体资源适配器就基于此描述信息执行相应的适配操作。描述信息被称为比特流语法描述符 (Bitstream Syntax Description)，它并非用来取代原始的二进制编码格式信息，而是作为原始二进制数据流的一个增强层信息来使用，类似于元数据信息。比特流语法描述符不是关于比特级的描述信息，它更多地是描述比特流的高层结构，例如比特流如何由不同质量层或数据包构成。比特流语法描述本身具有可伸缩性，能在不同的语义层次对比特流进行描述，如精细描述或粗糙描述，这取决于具体应用。

MPEG-21 DIA 提供了两种比特流语法描述工具：比特流语法描述语言 BSDL (Bitstream Syntax Description Language) 和通用比特流语法方案 gBS Schema (generic Bitstream Syntax Schema)。BSDL 用来设计针对各种具体编码格式的比特流语法方案 BS Schema，而 gBS Schema 可以被看作 BSDL 应用的一种特例，即事先定义的独立于编码格式的一种特殊 BS Schema。使用 BS Schema 和 gBS Schema 生成的比特流描述信息分别被称为比特流语法描述符 BSD (Bitstream Syntax Description) 和通用比特流语法描述符 gBSD (Generic Bitstream Syntax Description)。然而，BSD 必须与相应的针对具体编码格式的 BS Schema 配合才能使用，限制了其在处理能力和计算资源受限的远程设备 (如网关、代理、路由器等) 上的应用，而 gBSD 则不受此限制，它不仅可以对比特流中的各种层次结构进行有效的描述，还能实现简便、高效的存取和处理多媒体比特流的机制；gBS Schema 也为比特流或其片段提供了语义级的处理接口，有利于实现基于语义的多媒体资源适配。

有关文献提出了基于 BSD 和 gBSD 的多媒体资源适配引擎<sup>[7]</sup>。该引擎既实现了多媒体资源的适配，也实现了 BSD 和 gBSD 本身的适配。其基本适配过程包括：首先，根据一定的 BS Schema 对多媒体比特流进行描述，生成 BSD (gBSD)；然后，适配引擎将根据适配策略 (即满足适配约束条件，如网络带宽、显示分辨率等的最佳适配方案) 对原始的比特流描述符进行转换，例如选择适当的 XSLT 对描述符 XML 文件进行转换；最后，适配引擎基于转换后的新 BSD (gBSD) 描述，对原始的多媒体比特流进行对应的适配操作，生成原始比特流的适配版本 (Adapted Bitstream)。

### 4 基于感知运动能量的 H. 264/AVC 视频适配

基于 gBSD 的数字项适配不需要额外的 BS Schema，使得 gBSD 较 BSD 更适合应用在处理能力和计算资源受限的网络节点上，因此它获得了广泛的关注和研究。HP 实验室提出了一套针对 gBSD 数字项适配的 MPEG-21 数字项适配引擎模型和软件系统<sup>[8]</sup>，作为 MPEG-21 的参考实现软件之一。该软件能够根据网络状况、终端能力以及数字项内的 gBSD 信息来进行适配决策，动态地进行多媒体内容的适配。然而，作为适配决策和适配操作中关键辅助信息的 gBSD，其自动生成问题在这一软件实现中并未予以考虑。

因此，我们基于 gBSD 自动生成模块<sup>[7]</sup>和 H. 264/AVC 编解码器的参考实现软件，结合 H. 264/AVC 感知运动能量模型，开发了包含 PME 信息的 gBSD 自动生成工具，并在 HP

实验室模型的基础上建立了基于 PME 的 H. 264/AVC 视频适配框架,其结构如图1所示。

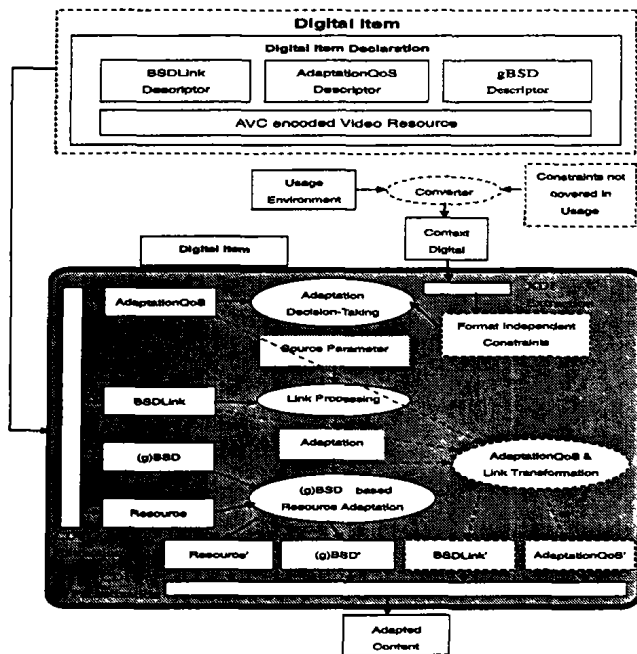


图1 基于 PME 的 H. 264/AVC 视频适配系统结构

图1中的 gBSD 自动生成工具首先根据 H. 264/AVC 感知运动能量模型计算各个视频帧的感知运动能量值,然后依据感知运动能量值的大小将视频的各个编码帧分配到不同的感知运动质量层。此处的感知运动质量层代表了对于视觉主观感知质量具有不同重要性的视频帧集合,属于不同感知运动质量层的视频帧具有不同的运动特性,而属于相同感知运动质量层的视频帧则具有相同或相近的运动特性。当丢弃不同的感知运动质量层时,用户感受到的运动抖动的大小将有所不同。例如,设感知运动质量层数为10,则可以将所有视频帧对应的 PME 值按降序排列等分为10个集合,按从高到底标记依次为集合10,9,8,...,1。如果某一视频帧对应的 PME 值属于集合1,则该帧被分配到最低的感知质量层1(Quality-Layer1);如果该帧的 PME 值属于集合2,则该帧被分配属于较高的感知质量层2(QualityLayer2)。gBSD 自动生成工具将这种分层信息嵌入到最终生成的 gBSD 中,即通过 gBSD 中的“Marker”元素将各个视频帧标记为属于对应的感知质量层。

gBSD 自动生成工具除了生成 gBSD 之外,还生成适配服务质量描述符(AdaptationQoS Descriptor)。适配服务质量描述符说明了网络或终端的限制条件、满足这些限制条件的可用适配操作以及适配后的服务质量之间的关系。它让适配引擎清楚了解满足适配限制条件的可用适配操作以及采取相应操作后的服务质量,以便于适配引擎进行适配决策。在基于 PME 的 H. 264/AVC 视频适配系统中,适配服务质量描述符描述了丢弃不同的质量层后传输该段视频所需的网络带宽。上述两种描述符和视频比特流一起被打包组合成一个完整的数字项,作为数字项适配引擎的输入之一。

数字项适配引擎解析输入的数字项及其包含的描述符信息,基于当前网络可用带宽  $BW$  和当前视频传输码率  $R_k$  来决定最佳的适配操作。此处  $k$  表示当前传输的最低感知质量层的标号,即从感知质量层  $k$  到最高的感知质量层  $n$  之间的所有层内的视频帧正被传输,而标号小于  $k$  的感知质量层数据则已被丢弃。 $R_1$ 无疑是最大的视频传输速率,此时所有感知质

量层的数据都被传输。这一适配过程的最终目的就是在满足网络可用带宽的约束条件下,尽可能地丢弃低感知质量层的视频帧,而尽可能多地保留高感知质量层的视频帧,从而最优化视频传输后的主观感知质量。具体适配决策算法如下:

- 当  $R_k < BW < R_{k+1}$  时,即当前网络可用带宽仅支持传输感知质量层  $k, k+1, \dots, n$ 。如果将感知质量层  $k-1$  加入当前的传输过程中,将超过网络当前的容量。因此,此时不必采取任何的适配操作。

- 当  $R_{k+1} < BW < R_k$  时,即当前网络可用带宽降低到仅支持传输感知质量层  $k+1, k+2, \dots, n$ 。因此,适配引擎应当将低于感知质量层  $k+1$  的各层中的视频帧丢弃,以降低视频传输码率。

- 当  $BW > R_{k-1}$  时,即当前网络可用带宽增加到可以支持更高的视频传输码率,因此标号小于  $k$  的感知质量层也可以被传输,适配引擎将把更多的视频帧加入到当前的传输序列中,即传输感知质量层  $k-1, k, \dots, n$ 。

## 5 实验结果和分析

为了验证我们提出的 H. 264/AVC 视频感知运动能量模型及适配算法的有效性,我们采用了用户主观评价的方法,来对丢弃不同比例的 B 帧后得到的视频进行主观质量的评价。测试环境和条件完全遵循 ITU 组织制定的视频质量主观评价方法建议 P. 910<sup>[9]</sup>,实验中所选择的视频测试序列为 H. 264/AVC 标准站点推荐的测试视频序列。在实验中,十位测试者被要求为所测视频进行评分。依据 P. 910 中的规定,评分分为五个等级:5. 非常好(Excellent),4. 好(Good),3. 一般(Fair),2. 较差(Poor),1. 差(Bad)。具体实验结果如表1所示。

表1 主观用户评价值与丢帧率

	Excellent	Good	Fair	Poor	Bad
Foreman	0%	10-20%	30-60%	70-80%	90-100%
Akiyo	10%	20-30%	40-60%	70-90%	100%
Container	0%	10%	20%	30-60%	70-100%
Hall	0%	10-20%	30-60%	70-80%	90-100%
Mad	0%	10-20%	30-50%	60-80%	90-100%

从表1可以看出,测试视频中大部分序列在丢帧率平均达到50%~60%的时候,其用户主观感受质量评分仍可以达到可接受的程度3. 一般(Fair),而视频编码速率平均有24%的减少。这就意味着基于 PME 的 H. 264/AVC 视频适配系统丢弃了对感受质量影响较小的视频帧,而保留了对视觉感受质量比较重要的视频帧,从而在最大限度上保证了最终的用户视觉感受质量的同时,显著地减少了传输视频所需的网络带宽,达到了适配的要求。

此外,我们还用客观评价方法考察了 H. 264/AVC 视频的适配效果。由于丢帧是对整个视频序列进行适配,因此在实验中我们采用视频序列中所有帧的平均 PSNR 作为评价指标来比较基于 PME 的丢帧适配与随机丢帧的效果。实验结果如图2所示,其中横坐标表示随机丢帧实验次数,纵坐标表示视频序列的平均 PSNR,丢帧率为30%。从图2可以看出,基于 PME 的丢帧适配始终优于随机丢帧的方法,其 PSNR 可以获得平均0.4dB 的提高。值得指出的是,PSNR 并不能完全刻画视频主观感知质量的高低,尤其是视频的运动特征,然而

(下转第158页)

性能评估的因素,我们使用了10交叉验证法。下文中的表格列出了对同样的188首豪放风格的宋词和210首婉约风格的宋词,使用不同方法得出的诗词风格判别的准确度以及作为特征的字数。

我们上述研究中采用的分类性能评估的方法是机器学习领域里比较常用的10交叉验证法,即将训练集随机划分为10个不相交的相等规模的组,每组有  $n/10$  个宋词样本,  $n$  为样本总数( $n$  在实验中为398),分类器要训练10次,每次留出1组作为验证集。

表1 10交叉验证结果

方法序号	判别模型	模型准确度	特征数(汉字数)
1	Na lve Bayes 模型	0.502	1327
2	用爬山法改进的 Bayes 模型	0.744	13 to 33
3	用遗传算法改进的 Na lve Bayes 模型	0.885	55

一种重要的分类问题的统计分析工具是接受特性(Receiver Operating Characteristic)曲线,通常简称ROC曲线,可用于比较各个分类方法的优劣,ROC曲线下包围的面积,代表了对于“二中选一”的问题正确回答的概率,ROC曲线下包围的面积越大说明对应的分类方法越好,综上所述,根据ROC曲线下包围的面积判断,基于遗传算法改进的朴素贝叶斯方法是较好的诗词风格判别方法(见表2)。

**结果及分析** 文学性语言的分析在自然语言处理中一直是一个难题,对诗词的计算机分析更难,本文在此作了初步的尝试。通过我们的研究表明,对古典诗词的采用以字为单位的向量空间表示模型能较好地表现诗词的内容和感情,较好地避免了诗词语句的分词难题,同时汉字字符的总数控制了问

题的规模,不会使处理的无限扩大。用基于字的向量空间模型表示诗词文本后,我们将古典诗词作者判别问题转化为了文本分类的问题,利用已有的成熟理论和技术,我们建立了多种机器学习和分类器,并且成功地用朴素贝叶斯方法结合遗传算法达到了较好的风格判别结果,准确率近似达到88.5%。我们的试验系统在输入诗词文本后,其中的诗词文本风格判别系统,能给出所输入诗词的风格是豪放或婉约的评价及其风格倾向的概率,我们下一步的研究将关注建立诗词风格量化的评价准则和方法。

表2 ROC曲线下面积的比较

判别模型	ROC曲线下包围的面积	特征数(汉字数)
Na lve Bayes 模型	0.394	1327
爬山法改进的 Bayes 模型	0.743	13 to 33
用遗传算法改进的 Na lve Bayes 模型	0.8968	55

### 参考文献

- 1 Aas K, Eikvil L. Text Categorisation: A Survey. June 1999. 3
- 2 Yao Tiansun. Natural Language Understanding. Tsinghua University Press, 2002. 373
- 3 Han J, Kamber M. Data Mining: concepts and techniques. 2001
- 4 Mitchell T M. Machine Learning. McGraw-Hill Companies
- 5 Wen Cai. Extension Engineering Method. Science Press, 2000. 99
- 6 Marques de sa J P. Pattern Recognition: Concepts Methods and Applications. Springer-Verlag Berlin Heidelberg New York, 2002
- 7 Duda R O, Hart P E, Stork D G. Pattern Classification (Second Edition). John Wiley & Sons, Inc., 2001
- 8 Yi Yong, He Zhongshi, Li Liangyan. Studies on Traditional Chinese Poetry Style Identification. In: the Proc. of ICMLC04. Shanghai, Aug. 2004. 2936~2939
- 9 唐圭璋. 全宋词. 中华书局, 1965

(上接第134页)

它仍然在一定程度上提供了辅助评价的支持。

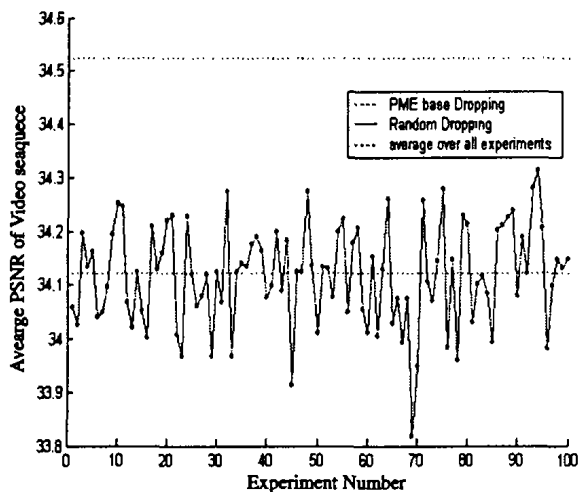


图2 视频序列平均 PSNR

**总结和展望** 本文中感知运动能量模型的应用使我们能够根据视频序列中运动活动的水平及视觉感知质量来对视频帧进行标记。本文将此模型与新制定的 MPEG-21 数字项适配框架相结合,提出了基于 PME 的 H. 264/AVC 视频适配模型及系统,实现了独立于媒体编码格式的视频适配机制。本文工作提供了一种标准化和灵活的方法来使 H. 264/AVC 视频适应实际的应用环境,同时视频的主观感知质量尽可能地得到了保证。

今后的研究工作应是进一步改进 H. 264/AVC 视频感知能量模型来更加准确地描述视频中的运动特性,从而更好地保证视频主观感知质量。此外,分布式视频适配,如在分布式的数个网络节点上协同地执行连续的适配操作,也将是我们的研究重点之一。

### 参考文献

- 1 Ma Y, Zhang H J. A New Perceived Motion based Shot Content Representation. 2001 IEEE Intl. Conf. on Image Processing (ICIP 2001), Greece, Oct. 2001
- 2 Liu T, Zhang H J, Qi F. Perceptual Frame Dropping in Adaptive Video Streaming. 2002 IEEE Intl. Symposium on Circuits and Systems (ISCAS2002), Arizona, USA, May 2002
- 3 Ardizzone E, et al. Video Indexing Using MPEG Motion Compensation Vectors. Multimedia Computing and Systems, 1999. 725~729
- 4 Ardizzone E, La Casica M, Molinelli D. Motion and Color-Based Video Indexing and Retrieval. 13th International Conference on Pattern Recognition (ICPR96), Vienna, Austria, Aug, 1996
- 5 Wang H, Divakaran A, Vetro A, Chang S-F, Sun H. Survey of Compressed-Domain Features Used in Audio-Visual Indexing and Analysis. Journal of Visual Communication and Image Representation, June 2003, 14:150~183
- 6 Bormans J, Gelissen J, Perkis A. MPEG-21: The 21st century multimedia framework. IEEE Signal Processing Magazine, March 2003, 20:53~62
- 7 Panisa G, Huttera A, Heuera J, et al. Bitstream syntax description: a tool for multimedia resource adaptation within MPEG-21. Signal Processing: Image Communication, EURASIP, Sept. 2003, 18:721~747
- 8 Mukerjee D, Kuo G, Liu S, Beretta G. Motivation and Use cases for Decision-wise BSDLink, and a proposal for Usage Environment Descriptor-Adaptation QoSLinking. ISO/IEC JTC 1/SC 29/WG 11, Hewlett Packard Laboratories, April 2003
- 9 ITU-T Recommendation P. 910. Subjective video quality assessment methods for multimedia applications. Geneva, 1996