

# 数字媒体适配过程的一般框架、模型及应用研究

黄天云

(西南民族大学计算机科学与技术学院 成都 610041)

**摘 要** 丰富的媒体内容和格式、异构的网络以及多样的终端设备,成为通用媒体访问的巨大障碍,媒体适配成为必要.文中分析了 MPEG-21 数字媒体适配 DIA 各实体之间的抽象关系,讨论了数字媒体适配的一般框架,建立了基于混合变量的约束优化模型.该模型统一了现有的媒体适配应用研究,能够用一致的算法进行求解.同时,文中从图像理解和视频分析角度,建立了媒体适配的层次结构,对媒体适配应用进行分类,并以当前的主要应用研究如图像适配、视频转码、位率适配、视频对象适配等举例说明.文中也探讨了混合媒体的多模态适配,指出了今后的研究热点和难点,如媒体语义抽取和适配、用户主观测度和媒体访问体验最大化等.

**关键词** 通用媒体访问;数字媒体适配;MPEG-21 DIA;实体关系;抽象模型;视频转码;位率适配;模态转换  
**中图法分类号** TP391

## General Framework, Mathematical Model, Current Activities and Open Issues for Media Adaptation in MPEG-21 DIA

HUANG Tian-Yun

(School of Computer Science and Technology, Southwest University for Nationalities, Chengdu 610041)

**Abstract** A growing variety of terminals and heterogeneous networks with dynamic throughputs make it a challenge in multimedia contents consumption. Media adaptation has been introduced in order to maximize user experiences and ensure quality of service (QoS) in universal media access environment. Based on the relations among entities such as resources, adaptation operators and utility in MPEG-21 digital items adaptation (DIA), and the general media adaptation framework, a mixed-variable constrained optimization model is proposed in this paper. The model unifies researches in this field, brings together the problem formulation and solution in a consistent way. A hierarchical adaptation structure is also constructed, from the viewpoints of image understanding and video analysis, and applied to classify some representative media adaptation applications. Leading pursuits such as image adaptation, scalable video transcoding, bit rate adaptation, video objects (VOs) adaptation, multi-modality conversion and fusion of mixed medias, etc., are illustrated by examples. Hotspots such as semantic abstraction of media contents, measurements of subjective quality, and maximizing of user experiences, etc., are also addressed.

**Keywords** universal media access; media adaptation; MPEG-21 DIA; entity; abstract model; transcoding; bit rate adaptation; modality conversion

## 1 引 言

为满足不同用户和各种类型终端设备无缝访问

网络上分布广泛的媒体内容, Morhan 等引入通用媒体访问 UMA (Universal Media Access)<sup>[1-2]</sup>. UMA 的提出,促进了系列研究的开展和相应的标准化进程如 MPEG-7<sup>[3-5]</sup> 和 MPEG-21<sup>[6-7]</sup>. 然而,在

异构的网络环境下,如何根据用户偏好及终端设备能力,最大化媒体访问的用户体验并保证媒体服务质量,仍然存在巨大挑战。

同时,媒体格式和内容也在不断增加和丰富,对数字媒体进行适配操作已成为必要.媒体适配(media adaptation)<sup>[8]</sup>即为满足通用媒体访问的需要而提出,并在近年来得到广泛关注和研究。

在普适(pervasive)媒体环境中,媒体适配,特别是视频适配,在将视频序列转换为新的格式或内容表示时,综合众多因素如媒体内容特征、用户环境、用户偏好和访问历史、设备能力等,目的是在不同的约束条件下,最大化媒体表示的效用(utility).适配过程可以在信号级、结构级或语义级进行.通常,输入视频经适配引擎将产生新的编码视频甚至完全不同的媒体表示(模态转换)。

近年来,媒体适配研究已经取得一定进展,特别是针对不同终端访问技术和动态变化的网络带宽而进行的位率适配.典型的有空域转码(transcoding)如 DCT 系数再量化<sup>[9]</sup>或部分丢弃<sup>[10]</sup>、时域转码如帧丢弃<sup>[11]</sup>、对象转码如视频对象优先级丢弃<sup>[12-14]</sup>、可扩展编码如 MPEG-4 FGS 的动态速率适配<sup>[15-16]</sup>、速率-失真或速率-失真-复杂度(rate-distortion-complexity)<sup>[17]</sup>优化转码和传输等。

上述研究部分解决了不同约束条件下的媒体适配问题,尚无法提供统一框架和一般化模型及求解算法,以达到媒体内容的通用访问. MPEG-7 媒体描述方案 MDS<sup>[3]</sup>旨在通过元数据(metadata)对媒体资源进行有效描述,以便内容管理、组织、访问和交互等;而 MPEG-21 数字媒体适配 DIA (Digital Items Adaptation)<sup>[7,18]</sup>则提供从内容描述到内容适配的有效操作手段。

本文将完成如下工作:(1)分析 MPEG-21 数字媒体适配过程各实体间的抽象关系,给出具体的媒体适配框架;(2)建立针对该框架的混合变量约束优化模型,并讨论如何利用模式搜索法进行求解;(3)从图像理解和视频分析角度,对媒体适配应用进行分类,建立媒体适配的层次结构,并以当前的主要应用研究举例说明;(4)探讨混合媒体的多模态适配;(5)指出今后研究的热点和难点。

## 2 实体关系及媒体适配框架

### 2.1 实体关系

首先对媒体适配过程的相关要素进行抽象,以建立实体(entity)<sup>[8]</sup>之间的相互关系。

实体被定义为媒体适配过程的基本适配单元.实体可以在信号级存在,如像素、帧、视频对象;也可以在结构级或语法级存在,如 GoP、场景或镜头等;甚至按 MPEG-7 元数据表示的语义元素也可以作为实体进行适配.对不同类型的实体,可以定义相应的适配操作.视频序列的传输,可以通过降低分辨率、空域再量化或时域帧丢弃等适配操作来减少网络带宽需求.语义元素如视频序列的故事或情节,通过适配操作后,能以视频摘要或文字等形式传输到不同终端。

适配空间(adaptation space)定义对特定实体,相应适配操作的可行域(feasible region).多维坐标系(为简化起见,图 1 仅标出 3 维)中对应的点,代表对实体进行适配操作的一种组合.如对转码,可能包含 2 维:空域 DCT 系数丢弃 CD(Coefficient Dropping)和时域帧丢弃 FD(Frame Dropping)。

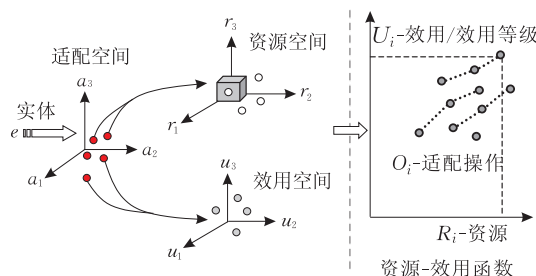


图 1 媒体适配及资源和效用空间的实体关系<sup>[8,19-20]</sup>

每个实体都与相应的资源(resource)和效用相关联.资源和效用给出适配操作的约束条件.媒体适配过程的资源空间包括传输网络如带宽、时延、位率等,用法环境(usage environment)如设备能力(分辨率、处理器、内存、电量)、信道容量等.效用空间则代表用户对媒体内容表示的满意程度.客观质量指标如 PSNR 适合计算,但无法反映用户的主观感受,特别是对媒体内容的语义级理解,它通常取决于用户的知识、经验、当前任务及相应背景等,目前尚无较好的方法对主观质量进行评价.同时,效用值也受用户偏好(preference)的影响.注意到,媒体适配过程更关注媒体内容的服务质量,网络服务质量如带宽、延迟和延迟变化等,通常作为资源约束应用到适配过程。

由此,媒体适配需要解决的问题可概括为:(1)给定内容实体和资源约束,寻求可行的适配操作,以最大化效用;(2)给定内容实体和用户期望的效用,寻求可行的适配操作,以最小化资源需求.问题(1)和(2)可转化为带约束的优化问题求解。

2.2 媒体适配框架

对媒体内容进行适配操作,以满足不同的终端需求(如彩色图像转换为灰度图像,MPEG 视频转换为 AVI 视频等,通常取决于终端设备能力和用户实际需求),在理论上是可行的.然而,限于服务器或中间节点的计算能力,实际操作将会遇到困难.因此,不必通过对媒体本身复杂的内容操作,仅需媒体内容包含的元数据,就能够依据网络和终端限制进行适配过程,将是更好的选择.

MPEG-7 MDS<sup>[3]</sup>提供对媒体数据的结构和语义级的统一描述工具(即元数据).MDS 按功能划分为:(1)基本元素:相关数据类型和基本工具;(2)内容描述:对媒体数据的结构或语义的感知;(3)内容管理和组织:对媒体数据的组织和分类;(4)导航和访问:内容摘要、层次、分解和重组;(5)用户交互:用户使用媒体数据的偏好和历史.MDS 使得媒体内容有效的分析和组织、访问和过滤、浏览和搜索成为可能.

MPEG-21 DIA<sup>[7]</sup>规范相关工具以便对数字媒体进行适配.数字媒体项(Digital Items)定义为结构化的数字对象,如标准的图像、视频数据表示、标识和元数据等.数字媒体项是 MPEG-21 DIA 框架进行适配的基本单元.

基于网络和终端约束,MPEG-21 DIA 获取相应的元数据如 MPEG-7 内容描述、通用约束(universal constraints)描述 UCD、用法环境描述 UED、适配元数据 AdaptationQoS,以进行快速、有效的适配操作过程(图 2).DIA 适配引擎分为两部分:(1)适配决策引擎 ADE(Adaptation Decision Engine),根据元数据和约束条件如网络、终端,进行合理的适配决策;(2)位流适配引擎 BAE(Bitstream Adaptation Engine),根据 ADE 提供的决策参数,进行实际的位流适配操作.

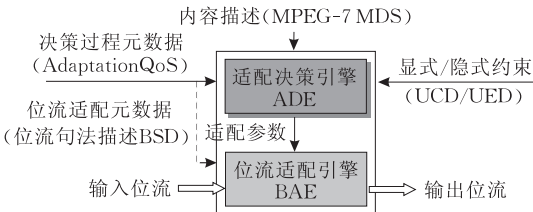


图 2 MPEG-21 DIA 数字媒体适配框架<sup>[8,18,21]</sup>

2.2.1 通用约束/用法环境描述 UCD/UED

UCD 给出媒体适配的显式(explicit)约束,而 UED 则给出用法环境的隐式(implicit)约束.例如:为满足特定终端的媒体需求和处理能力约束,需要

定义设备属性、编/解码能力,输入/输出特征等;为满足媒体服务的个性化需求,需要定义用户代理信息(Agent DS)(如偏好/访问历史)、消费习惯、访问限制以及位置特征(如移动、方向)等;为保证通过不同网络(如无线网络)的可访问性,需要定义网络特征(如容量、带宽、延迟、差错率、时戳等).

2.2.2 位流句法描述 BSD

位流句法描述 BSD(Bitstream Syntax Descriptions)在较高层次上,通过扩展标记语言 XML 对位流统计特征进行描述<sup>[22]</sup>.位流适配包括如下步骤(图 3):输入位流首先生成位流句法描述;通过标准的 XML 转换语言如扩展表格语言转换 XSLT 或 XML 流转换 STX 等对位流句法进行转换;转换过程可能包含基元的删除、增加和修改等.在转换后的位流句法上实施适配操作,以生成最终的输出位流.为方便 BSD 和位流的生成,DIA 定义了位流句法描述语言 BSDL(Bitstream Syntax Description Language).

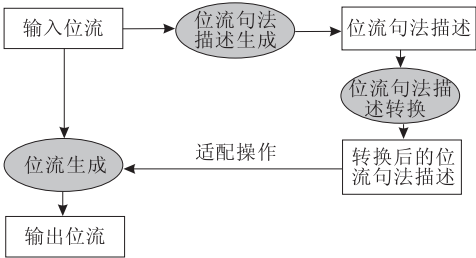


图 3 MPEG-21 DIA 位流句法描述 BSD 及适配<sup>[18,22]</sup>

此外,通用句法描述方案 gBSSchema 使得可以构造与媒体格式无关的 BSD,即 gBSDs. gBSDs 提供对语法基元的语义描述标记,因此与应用或领域相关的信息可以在 gBSDs 中标记.对位流描述的层次构造也使得随机访问和多重适配成为可能.为方便基于 BSD 的适配,BSDLink 工具建立了位流、gBSDs、一般 BSDs 之间的关联;位流适配的导引(steering)描述工具如 AdaptationQoS 则加快了适配过程.

2.2.3 决策过程元数据(AdaptationQoS)

为方便在多种约束条件下,选择最优的参数对媒体进行适配,人们定义了决策过程元数据 AdaptationQoS 工具. AdaptationQoS 建立了约束(或资源)、满足这些约束的适配操作以及相应的效用(或效用等级)三者之间的关联(参考图 1).

这种关联通过模块(Module)实现(图 4),以便根据应用要求选择最优参数,如适配操作的可行集、用法环境约束 UED、适配引擎参数等.

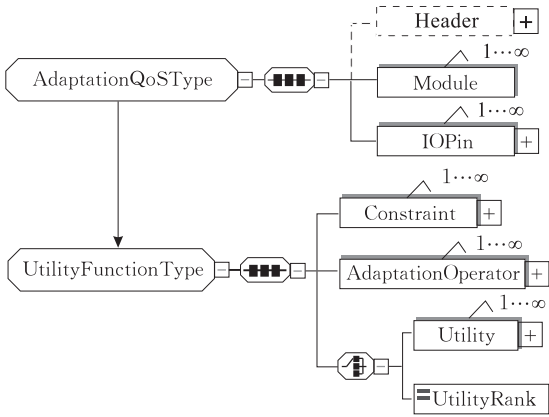


图 4 适配参数和效用函数的 XML 表示(参考例 1)<sup>[6,23]</sup>

模块分为 3 类:查找表(lookup table)、效用函数(utility function)、栈函数(stack function),分别对应着不同的数值或函数表示. 模块间的接口则通过输入/输出 Pins(IOPins)提供. 每个 IOPin 都是全局声明的唯一可标识变量,并可以从模块内部引用. IOPin 可以被解释为输入、输出或同时二者,这使得不同模块间的互连成为可能. IOPins 的变量值可直接声明为连续或离散值,或者为由语义确定的外部参数,如由 UCD 给出的约束(UCD 对用法或用法环境的限制或最优约束是通过栈函数实现的).

在多重适配情况下,其中一个 IOPin 作为后续适配单元的索引,其它 IOPins 都是该 IOPin 的函数(Dependent IOPins,参考图 5 右端). 注意到 UCD 或 AdaptationQoS 也可能从 UED 引用值,然而适配过程是由 UCD 和 AdaptationQoS 驱动,以保证操作的语义独立性.

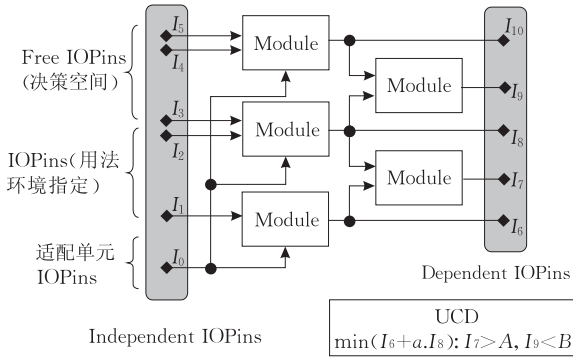


图 5 输入/输出 Pins(IOPins)及模块的关系<sup>[6-7,21]</sup>

由此,模块建立了 IOPins 的关联,即变量(约束)之间的依赖关系. 效用函数描述(图 4)则给出实体关系中,资源空间约束点对应的可行适配操作以及可能的效用(或效用等级). 效用值按离散索引的列表方式给出,以便根据实际的网络和终端 QoS 要求,选择最佳的适配操作.

图 6 给出一个媒体适配操作的应用场景(详细见文献[24]),图 7 是其对应的适配体系结构<sup>[8,23]</sup>.

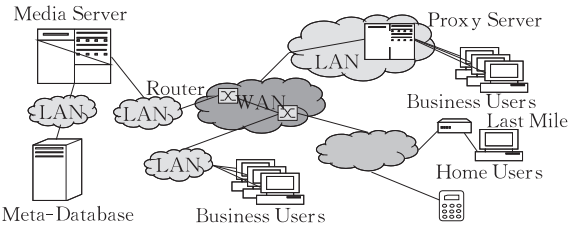


图 6 一个媒体适配操作的应用场景<sup>[24]</sup>

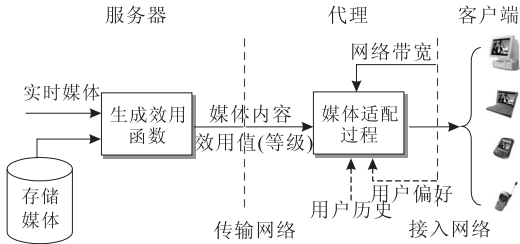


图 7 一种媒体适配的体系结构<sup>[8,23]</sup>

3 抽象模型及求解算法

基于实体关系和 DIA 适配框架,ADE 根据相应的约束条件(UCD 和 AdaptationQoS)和位流描述,选择可行的操作,对输入位流进行适配. 然而为避免依赖于特定的媒体类型、格式或对媒体内容的理解,需要将 ADE 抽象成与媒体无关的数学模型.

MPEG-21 DIA 的适配决策过程被映射为包含连续、离散或分类(Categorical)变量的混合变量(Mixed Variable)约束优化问题. 仅需对一个通用的数学模型进行求解,而不必考虑每个变量的具体含义,就可以得到相应的适配参数集,因此可以方便地对位流进行适配.

参考图 1,假设在  $n$  维约束空间对问题进行抽象,其中连续空间是  $R^c$ (如资源空间),离散空间是  $R^d$ (如适配空间,通常允许的适配操作是有限的,因此是离散的), $n=c+d$ . 分类变量通常可映射到整数离散变量集. 记适配参数集  $X=(x^c, x^d)$ .

考虑到多重适配的需要,如转码过程的帧率适配、连续视频序列的场景适配等,记已有决策的历史记录为  $H(m)=\{X(0), X(1), \dots, X(m-1)\}$ . 有如下混合变量约束优化模型:

$$\begin{aligned} & \text{Max } \{U_i[X(m), H(m)]\} \\ & \text{s. t. } O_j[X(m), H(m)] \leq 0 \text{ and/or} \\ & \quad R_j[X(m), H(m)] \leq 0, \\ & \quad i=1, 2, \dots, I; j=1, 2, \dots, J \\ & \text{Min } \{R_i[X(m), H(m)]\} \end{aligned} \tag{1}$$

$$\begin{aligned} \text{s. t. } & O_j[X(m), H(m)] \leq 0 \text{ and/or} \\ & U_j[X(m), H(m)] \leq 0, \\ & i=1, 2, \dots, I; j=1, 2, \dots, J \end{aligned} \quad (2)$$

式(1)为效用极大化模型,式(2)为资源极小化模型。以下仅解释式(1),式(2)可以同样理解。式(1)建立了  $U_i: R^{n \times m} \rightarrow R \cup \{+\infty\}$  的映射。模型的目标函数  $U_i (i=1, 2, \dots, I)$  给出  $I$  个适配操作的优化约束(optimization constraints);函数  $O_j/R_j$  给出  $J$  个等式或不等式的限制约束(limit constraints)。目标函数  $U_i$  可以是非线性、非连续,甚至不可微或随机的函数;而限制函数  $O_j/R_j$  既可以是线性、非线性函数,也可以是离散变量集,甚至黑盒函数,如随机仿真或程序代码生成的值集。模型求解的结果,是同时满足  $I+J$  个约束条件的适配参数集  $X(m)$ 。若不包含历史决策记录  $H$ ,则上述模型可简化为  $U_i: R^n \rightarrow R \cup \{+\infty\}$ 。

当  $I=0$  时,所有满足限制约束条件的解集都是可行解;当  $I=1$  时,通常只能得到一个可行解;而当  $I>1$  时,模型为多准则(Multi-Criteria)约束优化问题,在可行域的任一满足优化约束条件的 Pareto 优化解,都可以作为最终解。

综上,该抽象模型统一了现有数字媒体适配的相关研究<sup>[8-18, 21, 23]</sup>,能够用一致的框架进行描述和求解。鉴于模型的求解并不需要精确的媒体适配参数集,而传统非线性规划方法通常只能在强化的连续可微假设下,针对凸函数实施,无法直接应用到上述模型的求解;因此,考虑通过不依赖于方向导数计算的直接搜索法<sup>[25]</sup>(Direct Searches)求解。算法的详细分析将在后文给出<sup>[26]</sup>。

以下根据图 2、图 5 和上述模型,给出 ADE 适配操作的详细过程(算法 1)以及模型在不同约束条件下的求解方法(算法 2~算法 5)。对多目标优化,可选方案包括:仅对最重要的目标函数进行优化;对各目标函数加权,转化为单目标优化;各目标函数独立优化,取最终解的交集。

**算法 1.** ADE 适配操作过程(参考图 2 和图 5)。

1. 从 AdaptationQoS 获取决策过程元数据(决策空间 Free IOPins),这些数据可能是离散、连续或分类变量;
2. 从 MPEG-7 MDS 获取媒体内容元数据描述和表示;
3. 从 UCD 获取优化/限制约束条件(可能是线性、非线性约束函数,或离散变量集);
4. 从 UED 获取用法环境约束参数(用法环境 IOPins);
5. 已有适配参数作为适配单元 IOPins 输入(如多重适配);
6. 依据以上参数对当前适配单元进行约束优化,并获得适配参数集-混合变量约束优化模型;

7. 利用适配参数集,通过 BAE 对输入位流进行适配操作,得到最终的输出位流。

**算法 2.** 离散变量(所有 IOPins 均是离散变量)约束优化——穷尽搜索。

1. 设置初始候选点集为空;
2. 穷尽搜索离散变量点集
  - 2.1. 当前点是否满足优化和限制约束条件。若满足,则置入候选点集;否则丢弃;
  - 2.2. 检查当前点和候选点集中其它点的支配关系。若被支配(Dominated),则剔除当前点;若支配其它点,则剔除候选点集中的这些点;
3. 当所有点都搜索完毕后,任一在可行区域的解都可以作为最终的 Pareto 优化解。

**算法 3.** 连续变量(所有 IOPins 均是连续变量)约束优化,无限制约束,或限制约束为边界或线性约束——GPS<sup>[27-28]</sup>。

1. 在线性约束边界范围内,利用现有算法如模拟退火等,设定初始点、初始搜索方向(正支撑集)和搜索步长;
2. 从当前点开始寻求网格改善点;当到达线性约束边界时,根据边界几何特征,重新设定搜索方向和步长;
3. 迭代搜索直到步长小于某个上限,或搜索达到规定的次数,此时的点即为最终解。

**算法 4.** 连续变量(所有 IOPins 均是连续变量)约束优化,限制约束为非线性约束——GPS Filter<sup>[29]</sup>。

1. 由限制约束条件构造约束违背函数;
2. 在某个初始点集上选择有最小约束违背值或目标函数值的初始点,并构造初始覆盖集和过滤点集;
3. 对当前网格进行搜索和筛选(Poll),以寻求新的覆盖解,并构造新的过滤点集;
4. 重复上述过程直到步长收敛或到达规定的迭代次数。

**算法 5.** 混合变量(IOPins 可能是离散、连续或分类变量)约束优化——GMVP<sup>[30]</sup>。

1. 对确定的离散或分类变量,在连续空间采用 GPS 或 GPS Filter 搜索可行解;
2. 当网格搜索失败时,对离散或分类变量所在局部邻域的网格点集进行筛选和扩展筛选(extended poll);
3. 重复上述过程直到步长收敛或到达规定的迭代次数。

## 4 媒体适配层次结构

从图像理解角度看,图像数据包含多种类型的信息,如图标化数据(图像本身)、图像相关数据(分辨率、格式描述)、从处理的图像中提取的信息(数值、结构特征)、图像领域关系数据以及领域(应用程序)相关数据<sup>[31]</sup>。为了便于这些不同类型数据的表示,大多数的图像数据模型区分为逻辑和物理的图像表示<sup>[32]</sup>。物理语义表示在与应用语义无关的像素

层次上描述原始图像. 物理表示包括总体描述如颜色、纹理、直方图等, 几何描述如边缘、循环、线或连通图等. 逻辑表示与语义关联, 依赖于实际应用. 通常, 物理图像也包含原始图像, 而逻辑图像经过不同处理阶段获得多个层次的语义信息<sup>[33]</sup>. 对此, MPEG-7 MDS 提供了相关的视觉描述方案(彩色、纹理、形状、运动和位置)以及内容描述工具(结构和语义)<sup>[3]</sup>.

从视频分析角度看, 视频数据最重要的特征是它表征运动这一事实. 视频能在多个层次上表示, 表 1 给出不同表示层次上视频数据的结构粒度<sup>[34-35]</sup>.

表 1 不同表示层次视频数据的结构粒度		
层次	粒度	描述特征
视频	元	概念、制作者、导演…
情节	宏	事件描述…
剪裁或拍摄	小	活动、讲话、目标…
帧	微	对象及其空间关系…

在顶层, 视频可以看成单一原始对象, 该对象可以通过其概念、制作者、导演、参加者、类型等来总体描述. 情节层描述则沿着时间轴上发生的不同情节或事件来划分, 这是视频数据的一个宏表示粒度. 实现情节层描述的最好方式, 是首先将它划分成剪裁或拍摄的原子单位<sup>[35]</sup>. 一个视频序列的剪裁或拍摄涉及时间和空间描述连续视频帧相同活动的集合. 视频剪裁提供视频数据的最小粒度, 即视频分段. 视频剪裁包含场景的连续记录, 可以用多种特性和特征来索引. 此外, 相似索引的视频剪裁也可以聚集成

一个情节层描述<sup>[36]</sup>.  
视频结构的底层是单个视频帧, 它描述对象及其空间关系. 视频数据的对象索引可与事件描述结合起来, 以生成精确的对象和相关的活动索引.

实际的媒体应用可能包含复杂的要求<sup>[36]</sup>, 如以一个预定序列显示多个视频对象; 播放伴有特定音频对象的视频对象, 其中某个音频对象代表用某种语言讲述的故事; 同时播放两个视频对象, 其中每个代表从不同角度拍摄的同一景象. 因此, 需要有一种可以通过多个单一媒体对象的组合来产生复杂媒体对象的机制, 即组成模型<sup>[37-38]</sup>. 视频数据的组成模型包括空间和时间组成. 空间组成定义复合媒体对象成员的空间关系, 即一个显示画面中各成员的布局. 时间组成定义复合媒体对象成员的时间关系, 是生成复杂媒体对象的关键.

基于图像理解和视频分析, 可以建立数字媒体适配的层次结构(图 8). 首先, 因软、硬件限制或应用特定要求, 需要降低一个视频序列的分辨率或进行格式转换. 其次, 一个或多个视频序列可能需要进行分解和重组, 以产生新的视频序列或复合媒体对象, 例如在带限环境中, 一个视频序列的非重要分段可以用静止图像、标题文本或声音进行替换, 以降低带宽需求. 位率适配则涵盖了主要的视频传输研究, 可以通过 DCT 系数丢弃或再量化, 跳过部分 B 帧或 P 帧以降低帧率, 丢弃分级或对象可扩展编码视频的部分增强层等操作来进行位率控制.

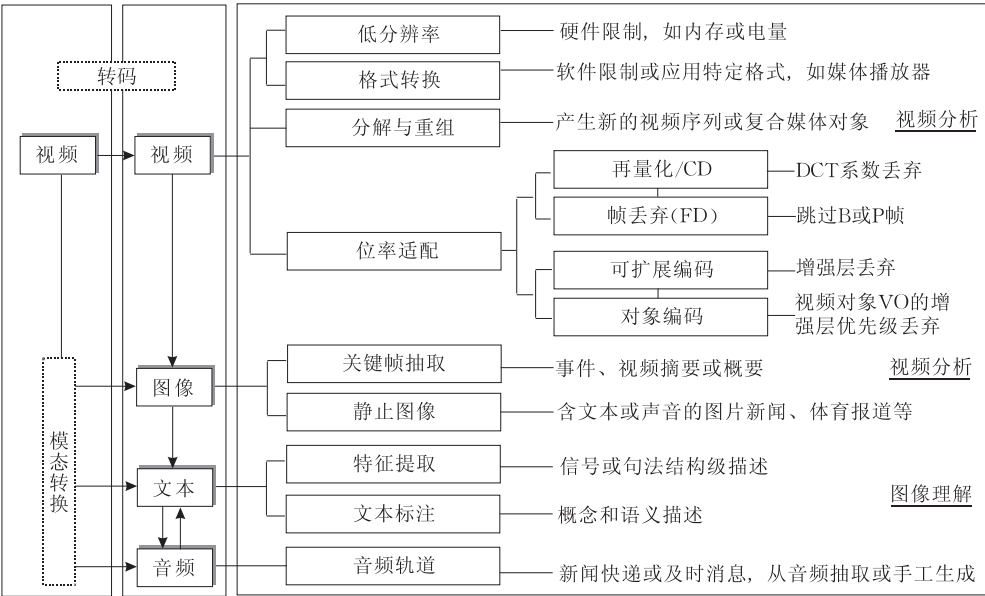


图 8 数字媒体适配的层次结构(从图像理解和视频分析角度进行分类)

如果上述适配操作尚不能满足应用要求,则需要进行媒体的模态转换,即从视频到图像、文本或音频的转换过程. 通过对视频序列的事件检测和关键帧抽取,可以形成视频摘要或概要(Video Abstract);并且其中的静止图像(如 I 帧)可以形成含文本或声音的图片新闻、体育报道等. 关键帧图像的特征提取可以在信号或句法结构级进行描述,而文本标注则形成对图像的概念和语义描述. 这可以通过自动或手工方式实现. 视频序列的音频轨道可以从音频抽取或通过文本手工生成,这在新闻快递或及时消息中可以得到广泛应用.

5 媒体适配应用研究

以下参照图 8 的适配层次结构,对近年来数字媒体适配的主要应用研究进行分析. 这些研究均是基于一定的限制约束条件,通过相应适配操作,以满足特定的优化约束条件. 因此,可以按媒体适配框架进行统一的描述、分析和求解.

5.1 图像适配<sup>[21,39-40]</sup>

JPEG 图像或 MPEG 的 I 帧,在码率和失真度(如 MSE)两个指标间的均衡,一般通过不同的量化因子实现<sup>[39]</sup>. 依据不同的 UCD 约束,决策适配引擎 ADE 搜索离散参数集{(量化因子,码率,MSE)};并由位流适配引擎 BAE 实施具体的适配过程.

可以看出,离散参数集中的 3 个参数是相互关联的. 因此,有两种不同指标的 JPEG 图像适配方式:(1)Min(MSE), s. t. 码率 $\leq r$ ,其中  $r$  为事先设定的码率上限;(2)Min(码率), s. t. MSE $\leq m$ ,其中  $m$  是允许的最大失真度. 因此,决策过程元数据如图像尺寸和分辨率(Free IOPins),UCD 约束如码率和失真度(Dependent IOPins),作为 ADE 的输入,以求得满足 UCD 约束的量化因子;并以此参数输入 BAE,进行实际的位流适配过程.

而对可扩展编码的 JPEG2000 图像<sup>[40]</sup>,因为同时包括空域、SNR 和色度(Component)扩展层等, ADE 需要计算出相应的可扩展层个数,以满足特定的 UCD 约束. 输入决策元数据(如分辨率、初始可扩展层数(Free IOPins))、UCD 约束(如码率和失真度(Dependent IOPins)),通过 ADE 决策,以求出满足约束的最终可扩展层数. 同样有两种不同的适配方式:(1)Max(图像分辨率), s. t. 图像分辨率 $\leq$ 显示设备分辨率,码率 $\leq r$ (在特定传输码率和终端分辨率限制下,最大化图像分辨率);(2)Min(MSE),

s. t. 图像分辨率 $\geq$ 期望的最小分辨率;码率 $\leq r$ (在特定码率和用户期望的最小分辨率下,最小化失真度).

因此,JPEG 图像的不同适配方式,既考虑了客观指标如 UCD 约束(码率、失真度),也引入了主观指标如用法环境参数 UED(终端设备能力、用户偏好等). 由此相应的适配参数集经 BAE 的具体适配操作,得到满足不同需要的输出图像.

5.2 FD/CD 视频转码

转码通过仅对视频进行部分解码,然后再编码成满足特定网络带宽的码率或终端用户需求的格式,因此可以有效降低服务器的存储容量(对比多描述编码 MDC)或处理能力要求<sup>[41-43]</sup>. 然而,传统的转码如 DCT 系数再量化<sup>[9]</sup>,可能会因参考帧的不匹配而导致错误漂移(Error Drift). 而对宏块 MB 高频 DCT 系数截断的动态速率整形 DRS<sup>[10]</sup>(Dynamic Rate Shaping),未考虑到帧率适配,在严格的带宽限制下可能会导致单帧较差的质量.

基于对媒体适配实体关系的分析,结合空域 DCT 系数丢弃 CD<sup>[10]</sup>和时域帧丢弃 FD<sup>[11]</sup>, FD-CD 转码模型在近年来得到极大关注<sup>[11,19]</sup>(图 9);并对传统的速率-失真 R-D 模型进行推广,提出资源-效用 R-U 模型(Resource-Utility)<sup>[19-20,23]</sup>.

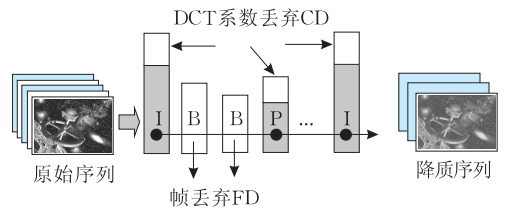


图 9 FD/CD 组合的视频转码示意图<sup>[8,19]</sup>

假设以视频序列的 GoP 为单元进行适配, $F$  和  $F'$  是输入视频和适配后视频的帧率,定义  $f=F/F'$  为 FD 适配因子;若  $R$  和  $R'$  分别是输入视频和适配后视频的位率,定义  $c=1-R'/R$  为 CD 适配因子,即参数  $c$  表示丢弃的高频 DCT 系数的比例;定义适配参数集  $a=(f,c)$ . 由此,参数集  $a$  确定了满足特定带宽限制或视频质量(如 PSNR)的适配操作. 适配过程对每个 GoP 按固定的模式,持续丢弃当前帧 DCT 系数和后续的  $f-1$  帧(图 10). 注意到,为避免出现参考帧位移,仅允许丢弃 B 帧和 P 帧;为保证适配后的视频和原始视频有尽可能一致的质量指标,所有的 CD 适配帧都采取相同的参数  $c$ . FD-CD 适配组合能在时间和空间粒度上取得较好的平衡,并且满足较大范围的动态带宽变化.

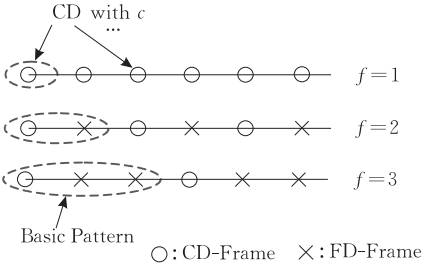


图 10 3 种模式的 FD-CD 适配示例<sup>[23]</sup>

由上述假设,可以建立特定适配操作  $a=(f,c)$  的 FD-CD 数学模型<sup>[23]</sup>,以最大化 PSNR,或最小化失真度.并且,试验表明:忽略帧间的累积差错和传播误差,基于块级的 CD 能取得较好的 PSNR<sup>[23]</sup>.

基于 R-U 模型和目标码率的限制,可以得到相应适配操作  $a=(f,c)$  的离散采样点;通过对离散采样点的分段线性插值,可以建立任意目标码率和效用函数关系的适配操作参数集(图 11).因此可以在媒体适配框架下,方便地对视频序列进行适配.

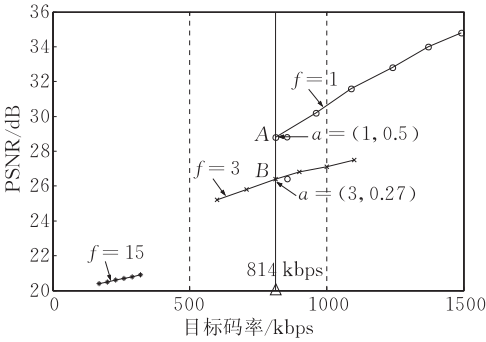


图 11 目标码率与效用函数关系<sup>[21,23]</sup>

**例 1.** 一个 AdaptationQoS 的 XML 实例,采用基于 PSNR 的效用函数表示(详细见文献[6], Annex 2.1: AdaptationQoSCS).

```
<DIA>
<DescriptionMetadata>
  <ClassificationAlias alias = "AQoS" href = "urn:mpeg:
    mpeg21:2003:01-DIA-AdaptationQoSCS-NS"/>
</DescriptionMetadata>
<Description xsi:type="AdaptationQoSType">
  <Module xsi:type="UtilityFunctionType">
    <Constraint iOPinRef="BANDWIDTH">
      <Values xsi:type="IntegerVectorType">
        <Vector>1510 1359 1200 1071 941 814 1000 952 909
          814 712 600 396 359 331 293 255 217</Vector>
      </Values>
    </Constraint>
    <AdaptationOperator iOPinRef="B_FRAMES">
      <Values xsi:type="IntegerVectorType">
```

```
<Vector>0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2</Vector>
</Values>
</AdaptationOperator>
<AdaptationOperator iOPinRef="P_FRAMES">
  <Values xsi:type="IntegerVectorType">
    <Vector>0 0 0 0 0 0 0 0 0 0 0 4 4 4 4 4 4</Vector>
  </Values>
</AdaptationOperator>
<AdaptationOperator iOPinRef="COEFF_DROPPING">
  <Values xsi:type="FloatVectorType">
    <Vector>0.0 0.1 0.21 0.3 0.4 0.5 0.0 0.08 0.15 0.27
      0.38 0.5 0.0 0.1 0.2 0.3 0.4 0.5</Vector>
  </Values>
</AdaptationOperator>
<Utility iOPinRef="PSNR">
  <Values xsi:type="FloatVectorType">
    <Vector>34.47 33.56 32.48 31.58 30.27 29.10 27.03
      26.49 26.21 26.08 25.87 25.12 21.44 21.36 21.29
      21.18 21.02 20.87</Vector>
  </Values>
</Utility>
</Module>
<IOPin semantics="":AQoS:1.1.1" id="BANDWIDTH"
  input="true" output="false"/>
<IOPin semantics="":AQoS:2.1" id="PSNR" input="
  false" output="true"/>
<IOPin semantics="":AQoS:3.1.1" id="B_FRAMES"
  input="false" output="true"/>
<IOPin semantics="":AQoS:3.1.2" id="P_FRAMES"
  input="false" output="true"/>
<IOPin semantics="":AQoS:3.1.3" id="COEFF_DROP-
  PING" input="false" output="true"/>
</Description>
</DIA>
```

5.3 可扩展编码视频适配

对分级可扩展编码的 MPEG 视频在动态网络环境下的位率适配,已进行了诸多研究.其核心是如何在带宽有限且动态变化的网络环境下,采取有效的分层丢弃策略(时域、空域或 SNR),并保证视频传输的整体质量.

**例 2.** 基于 GoP 的可扩展编码视频位率适配. ADE 以 GoP 为适配单元,根据网络带宽和终端能力,动态决定每个 GoP 需要传输的不同扩展层次的个数. BAE 则直接丢弃整个子层(时、空或 SNR),不对位流进行任何解析. IOPins 由 AdaptationQoS 定义和声明,并由 UCD 根据模块关联的约束条件来进行具体的适配过程(参考图 5).

适配单元:GoP(连续适配过程必须考虑历史数据,后续 GoP 的空间分辨率应与第 1 个 GoP 保持一致);Free IOPins: NTEMP(时域层数),NSPATIAL(空域层数),NSNR(SNR 层数);Dependent IOPins: 帧速率、位率、视频质量、空间分辨率等. 适配操作的实例请参考文献[21].

MPEG-4 的精细粒度可扩展(FGS)视频因采用位平面编码技术,特别适合动态变化的 IP 流传输. 因此以下分析 FGS 的位率适配<sup>[15-16,44]</sup>.

按视频序列分解的层次结构,通过对相关统计信息如分段个数、每个分段的场景个数和场景持续时间的分析,在每个分段基础上对视频序列进行适配操作. 在基层无丢失假设下,若视频序列当前分段  $m$  共有  $N$  个场景,  $r_e(n)$  是当前分段的场景  $n$  增强层传输速率,则可以定义分段最优传输策略<sup>[16]</sup>:

$$\begin{aligned} \Omega_m^* = & \text{Min} \sum_{n=1}^N \left[ \frac{1}{N_n} \sqrt{\sum_{i=1}^{N_n} |\Phi_{n,i}(r_e(n)) - \bar{\Phi}_n(r_e(n))|^2} \right] \\ \text{s. t. } & \begin{cases} B(\mathcal{R}) = N_1 \cdot r_e(1) + N_2 \cdot r_e(2) + \dots + \\ N_N \cdot r_e(N) \leq B_{\max}; \\ r_e(n) \leq r_{\max}^e, n=1, 2, \dots, N \end{cases} \end{aligned} \quad (3)$$

其中,  $\bar{\Phi}_n(r_e(n))$  是场景  $n$  的平均质量指标,  $\Phi_{n,i}(r_e(n))$  是场景  $n$  第  $i$  帧的质量指标;  $N_n$  是当前分段第  $n$  个场景的长度,  $B_{\max}$  是分段增强层编码带宽限制;  $r_{\max}^e$  是增强层允许的最大码率.

因此,分段最优传输策略的求解,就是要在带宽限制条件下,使得当前分段中任意场景与原始场景有最为接近的平均质量指标,并且使该分段整体质量变化指标最小化.

该模型可以利用分段状态迁移图和动态规划进行求解(图 12),得到当前分段每个场景的增强层最佳传输速率  $r_e(n)$  和客户端最优缓冲时间  $\Delta_i$ <sup>[16]</sup>. 然而考虑到  $r_e(n)$  通常取至离散点集(例如文献[44]的 400~2000kbps,间隔为 200kbps),该模型可利用模

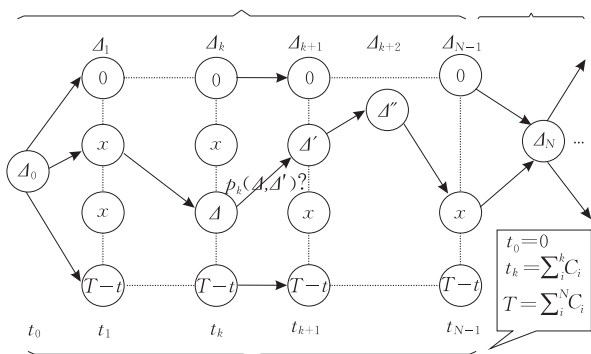


图 12 视频序列分段状态迁移图及最优传输策略<sup>[16]</sup>

式搜索求解,并可通过线性插值得到在不同  $r_e(n)$  下的视频质量指标.

### 例 3. 基于场景的 MPEG-4 FGS 视频位率适配.

首先,任意 MPEG-4 FGS 视频的速率迁移矩阵  $\mathbf{R}$  和缓冲控制矩阵  $\mathbf{B}$  可以通过上述调度算法,在服务器上离线获得并事先存储<sup>[16]</sup>. ADE 以场景为适配单元,根据网络带宽和终端能力,动态决定每个场景的增强层速率(通常 MPEG-4 FGS 视频的基层按较低码率进行编码,因此可假设基层数据无丢失). BAE 根据增强层码率决定相应的位平面个数(FGS、第 1 个 FGST、第 2 个 FGST),不对位流进行任何解析.

适配单元:场景; Free IOPins: FGS 的位平面个数、第 1 个和第 2 个 FGST 的位平面个数; Dependent IOPins: 场景个数、场景增强层速率等.

## 5.4 视频对象适配

基于对象的 MPEG-4 编码,使得可以在视频对象 VO 的不同扩展层上对视频序列进行适配. 场景中的不同 VO 根据其重要程度,被赋予不同的初始优先级值,并映射为不同的传输优先级. 在传输过程中根据网络状况动态丢弃不重要的扩展层,甚至整个 VO<sup>[14]</sup>. 试验表明,相比基于帧的可扩展层丢弃算法,VO 扩展层丢弃策略能取得更好的终端表示质量.

### 例 4. 基于视频对象扩展层的位率适配.

例 2 的思想可以进一步推广到视频对象 VO,如图 13,该方案可以对不同优先级的 VO 扩展层进行动态丢弃,从而达到视频回放的优雅降质<sup>[14]</sup>. 由此我们也可以建立一个场景中各视频对象的优先级树,并据此实施具体的适配操作.

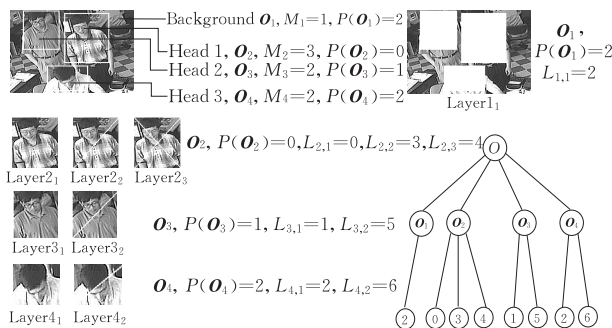
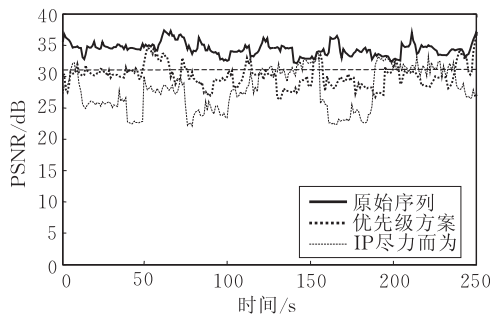


图 13 视频对象扩展层的多优先级树结构表示<sup>[14]</sup>

图 14 给出上述视频序列在不同传输策略下的性能指标比较. 与原始序列(PSNR: 34.2dB)相比, IP 的尽力而为服务仅为 26.3dB 且有较大的质量波动,而通过视频对象扩展层的优先级丢弃,平均质量则达到 30.8dB.

图 14 不同传输策略下的性能指标比较<sup>[14]</sup>

因为 VO 层的粗粒度控制,上述方案只能消极适配网络状况.通过对 VO 层选择适合的量化参数和时间分辨率(帧率),可以在速率约束下最小化失真度<sup>[12]</sup>.假设  $Q$  为量化参数,  $s$  为空域复杂度参数,  $(x_1, x_2)$  为一阶和二阶模型参数;  $r$  为 VO 纹理编码需要的比特数.则有如下 VO 速率控制纹理模型<sup>[45-46]</sup>:

$$r = s \cdot \left( \frac{x_1}{Q} + \frac{x_2}{Q^2} \right) \quad (4)$$

给定  $r$  和  $s$ , 模型参数  $(x_1, x_2)$ , 可以求解量化参数  $Q$ ; 并且在已知  $r, s$  和  $Q$  的情况下, 连续帧的模型参数  $(x_1, x_2)$ , 可以通过线性回归来估计和更新. 同时, 考虑到量化过程对  $ac$  系数的影响, 引入 VO 的 DCT 复杂度参数  $\bar{s}$ , 以替代空域复杂度参数  $s$ . 定义:

$$\bar{s} = \frac{1}{M_c} \sum_{m \in M} \sum_{i=1}^{63} \rho(i) \cdot |B_m(i)|^2 \quad (5)$$

其中,  $B_m(i)$  为块的  $ac$  系数;  $m$  为 VO 的编码块集  $M$  的宏块索引,  $M_c$  为 VO 的编码块个数;  $\rho(i)$  为不同  $ac$  系数的权重, 通过类似 MPEG 的量化矩阵对  $ac$  系数进行高频加权.

基于网络带宽限制  $R$ , 可以给出任意时刻最大允许的视频对象平面 VOP 的目标码率  $r$ . VO 适配的目的是保证在新的量化参数  $Q'$  下, 每个 VO 相比于原始 VO, 只有尽可能小的失真<sup>[12]</sup>:

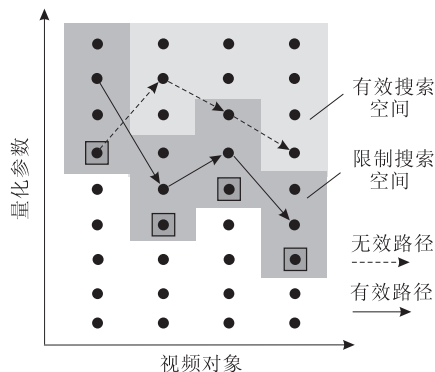
$$\begin{aligned} \text{Min}\{\Delta D(Q'|Q)\} &= \text{Min}\left\{\sum_{k \in \kappa} \alpha_k [D(Q'_k) - D(Q_k)]\right\} \\ \text{s. t. } r &\leq R; Q_k \leq Q'_k \leq Q_{\max} \end{aligned} \quad (6)$$

其中,  $k$  为 VOP 集  $\kappa$  的索引,  $\alpha_k$  为不同 VO 的优先级或重要程度; 取决于对象复杂度, 也可能由内容提供者事先标记<sup>[14]</sup>.  $Q_{\max}$  为允许的最大量化参数, 通常取 31. 该问题可以通过 Lagrange 乘子转化为无约束问题, 也可以利用动态规划求解.

**例 5.** 基于视频对象量化参数 QP 的位率适配<sup>[12]</sup>.

图 15 的矩形方框表示每个视频对象 VO 的原始量化参数  $Q_k$ , 灰色区域样本点表示可行的量化参

数  $Q'_k$ . 其速率可以用上述纹理模型近似, 而失真可以用  $\Delta D(k, Q'_k) = Q'^2_k - Q^2_k$  近似. 注意到图 15 允许的量化参数空间是有限点集, 因此 ADE 可以通过离散搜索, 得到满足最小失真度准则的每个 VO 的量化参数  $Q_k$ , 从而 BAE 可以对每个 VOP 位流进行适配.

图 15 量化参数 QP 的搜索空间<sup>[12]</sup>

然而, 量化参数  $Q'_k$  和  $Q_k$  不能相差太多, 否则会造成部分 VO 较差的空域质量. 因此, 引入参数  $\Delta Q_k$ , 通过限制可行解空间, 在空域和时域之间折中:  $Q_k \leq Q'_k \leq \text{Min}\{Q_k + \Delta Q_k, Q_{\max}\}$  (图 15 的深灰色区域). 可以对所有 VO 采用相同的  $\Delta Q_k$  (所有 VO 均匀分布), 也可以根据不同的 VO 空域复杂度来决定 (高复杂的 VO 通常使用较小的  $\Delta Q_k$ , 如前景对象; 而不重要的 VO 则使用较大的  $\Delta Q_k$ , 甚至是丢弃整个 VO, 如背景对象). 整个 VO 的丢弃可能会在场景合成时造成空洞现象, 此时基于元数据的 Shape Hints 能更好地解决该问题<sup>[13]</sup>.

## 5.5 多模态转换和融合

基于内容可伸缩 (content scaling) 的媒体适配, 已经可以满足大多数的媒体访问需求. 然而对无线网络等带宽非常有限的环境, 为适应较大范围的资源约束, 媒体数据的模态转换<sup>[47]</sup> (modality conversion) 更为有效. 模态转换依据相应的资源约束, 将媒体内容从一种表示形式转换为另一种表示形式, 如无线网络中的视频访问, 相比于较差的视频表示质量, 关键帧传输能取得更好的用户主观感受. 因此, 模态转换能保证更好的媒体访问 QoS.

模态转换取决于以下因素<sup>[48]</sup>: 媒体语义、资源约束、终端能力和用户偏好. 是否需要进行模态转换以及转换的最终模态, 主要取决于资源约束和用户偏好. 模态转换的一般顺序是: 视频 → 图像 → 声音 → 文本, 如表 2 所示.

表 2 媒体内容的模态转换顺序

→	视频	图像	音频	文字
视频	1	2	3	4
图像	—	1	2	3
音频	—	—	1	2
文字	—	—	2	1

注：整数表示转换优先级，‘1’最高，‘—’表示通常这种转换不允许。

类似资源-效用 R-U 模型,图 16 给出混合媒体的模态曲线及相应的效用函数关系. 图 16(a)各曲线表示媒体内容处于不同模态下时,对应的效用值(客观效用值如 PSNR 和用户定义的主观质量指标的综合). 给定资源约束,媒体所处的模态及最佳效用值可以通过离散采样有效估计;并且也能方便地确定不同模态之间的转换点,如图 16(b)的垂直分隔线(剔除了声音模态曲线).

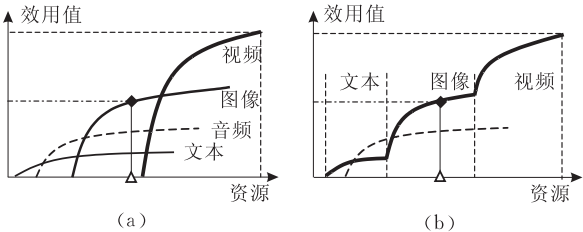


图 16 混合媒体模态曲线及对应的效用值<sup>[47-48]</sup>

可以得到媒体对象  $i$  在特定资源限制点,不同模态曲线间的距离  $d_{ij} (j=1,2,3,4)$ . 标记  $w_{ij} (j \leq i)$  为将媒体对象  $i$  转换到模态  $j$  的加权,则可以得到加权模态曲线如图 17,其中  $d'_{ij} = w_{ij} \cdot d_{ij}$  为加权距离. 较大的加权将使得对应模态所在的曲线变宽(加权或用户偏好取决于用户选择,如对新闻视频,可能更愿意选择关键帧浏览,甚至仅访问文本消息). 对加权距离进行归一化即可得到最终的归一化距离  $\hat{d}_{ij}$ :

$$\hat{d}_{ij} = (w_{ij}d_{ij} \cdot \sum_j d_{ij}) / (\sum_j w_{ij}d_{ij}) \quad (7)$$

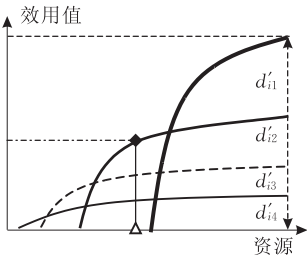


图 17 混合媒体加权模态曲线

因此对媒体对象  $i$ ,可以在资源约束  $r_i$  下,根据用户偏好  $p_i$ 、最大化内容值  $U$  来优化计算<sup>[47]</sup>:

$$U = \text{Max} \sum_i U_i = \text{Max} \sum_i f_i(r_i, p_i), \text{ s. t. } \sum_i r_i \leq R \quad (8)$$

同样地,该模型可以通过动态规划或搜索算法求解.由此,内容可伸缩的适配、结合媒体对象的多模态转换,极大地扩展了媒体适配的应用,能够满足较大范围的资源约束;这也使得不同环境下多样的媒体访问成为可能.

例 6. 混合媒体对象的多模态表示与传输.

考虑一多媒体文档(如网页),其中包含视频、图像、音频和文本等媒体对象,如表 3. 规定每种类型媒体的最低传输率要求,若分配的带宽低于该下限,只能转换为更低级别的模态类型. 每个对象在其标准传输率下可以获得最高分,否则按模态曲线递减.

表 3 每个媒体对象的标准和最低传输率及相对重要程度(最高分为 10 分,手工标定)

	重要程度	标准传输率/kbps	最低传输率/kbps
$V_1$	10	347	32
$I_2$	5	121	12
$I_3$	5	105	12
$I_4$	5	112	12
$A_5$	3	64	4
$T_6$	2	4	1

表 4 给出不同带宽限制下,为达到最大效用值  $U$  (不超过 30),各媒体对象可能的带宽分配及所处模态(用字母表示). 表 5 则引入用户加权(偏好),令视频→视频的权重为 1.2,得到新的分配方案和效用值.

表 4 不同带宽下各媒体对象可能的带宽分配、所处模态(用字母表示)及最大效用值  $U$

带宽/kbps	$V_1$	$I_2$	$I_3$	$I_4$	$A_5$	$T_6$	$U$
700	347V	110I	91I	100I	48A	4T	28.67
600	300V	95I	75I	86I	40A	4T	27.08
500	250V	80I	62I	72I	32A	4T	24.32
400	200V	65I	51I	57I	24A	3T	19.83
300	180V	40I	28I	34I	16A	2T	16.85
200	100V	40I	20I	30I	8A	2T	14.30
100	64V	20I	4A	6A	4A	2T	10.66
70	36V	20I	4A	4A	4A	2T	7.61
50	24I	8A	5A	7A	4A	2T	5.38
30	10A	6A	4A	5A	4A	1T	3.84
20	6A	4A	1T	4A	4A	1T	2.41
10	4A	1T	1T	1T	2T	1T	1.13

表 5 引入视频→视频的权重为 1.2 后,新的带宽分配、所处模态(用字母表示)及最大效用值  $U$

带宽/kbps	$V_1$	$I_2$	$I_3$	$I_4$	$A_5$	$T_6$	$U$
200	100V	40I	20I	30I	8A	2T	14.30
100	64V	20I	4A	6A	4A	2T	10.66
70	40V	16I	4A	4A	4A	2T	7.58
50	32V	6A	4A	4A	2T	2T	5.24
30	12I	6A	4A	5A	2T	1T	3.75
20	8A	4A	1T	4A	2T	1T	2.32
10	4A	1T	1T	1T	2T	1T	1.13

**例 7. 基于视频对象分解的视频序列多模式转换和融合.**

考虑图 18 的视频序列 News, 初始编码成 128kbps, 从中抽取 4 个视频对象 ( $O_1$ : 背景;  $O_2$ : 主持人 1;  $O_3$ : 主持人 2;  $O_4$ : 舞蹈表演画面). 通过基于对象的丢弃可以将码率降到最低 32kbps.



图 18 视频序列 News 的对象抽取和标记<sup>[12]</sup>

对应的音频(A)和文本(T)可从原始序列抽取, 图片新闻(Picture News-P)也可由该序列构造. 表 6 给出不同对象的平均传输率和手工标定的效用值. 表 7 给出不同带宽限制下, 可行的传输策略及获得的复合效用值.

表 6 每个对象的平均传输率和效用值(最高 10 分)

对象	平均速率/kbps	效用值
Full	128	10
$O_1$	16	4
$O_2$	24	5
$O_3$	24	5
$O_4$	32	6
P	16	7
A	8	3
T	2	1

注: 各对象效用值标定不包含对应音频或文本.

表 7 不同带宽限制下的传输策略及复合效用值  $U$

带宽/kbps	Full	$O_1$	$O_2$	$O_3$	$O_4$	P	A	T	$U$
140	✓	—	—	—	—	—	—	—	10.0
100	—	—	✓	✓	✓	—	✓	—	9.5
70	—	—	—	✓	✓	—	✓	—	8.5
50	—	—	—	—	✓	—	✓	—	8.0
40	—	—	—	—	✓	—	✓	—	8.0
	—	—	—	—	—	✓	—	—	7.0
35	—	—	—	—	—	✓	—	—	8.0
	—	—	—	—	✓	—	—	✓	7.0
10	—	—	—	—	—	—	✓	—	3.0
5	—	—	—	—	—	—	—	✓	1.0

注: 表中“✓”表示传输; “—”表示不传输或不需要传输.

6 当前研究热点

媒体适配其它方面的研究也应引起关注, 如视频序列的语义抽取和语义驱动的适配; 媒体访问过程的用户主观感受难以度量; 如何以最大化用户体验的方式, 满足用户的媒体访问需求等.

6.1 语义抽取和适配

从图像处理角度, 对视频序列在信号级、结构级的分析已进行了多年研究. 基于语义级的视频分析和语义抽取, 也在图像检索、视频浏览中得到应用.

视频分析可以识别序列的场景和对象、文字等, 并通过对事件(Event)的边界检测, 来形成相应的语义标注(Annotation). 统计模式识别方法如支持向量机 SVM、概率 Bayes 网络、层次隐 Markov 模型 HHMM 等, 均可以应用到该过程. MPEG-7 MDS 也定义了相应的语义描述方案(Semantic DS).

一旦抽取出视频序列的语义标注, 就可以应用到视频适配的优化转码和传输. 如: 对视频摘要(video summarization)的非重要分段, 可以用关键帧、文本描述、声音等替换, 也可以考虑用户喜好等因素, 因此可以更好地满足约束环境下的主观质量需求. 而基于特征一致性(图像帧特征或统计特征)和语法规则的关键帧提取, 取决于提取哪些帧以及提取比例, 往往由约束条件如带宽、终端能力等决定, 在视频序列的滑动浏览(Slide Show)和导航访问(Navigation)中可以得到应用. 提取过程也可能导致重要视频场景信息的丢失以及音、视频内容不同步, 甚至音频无法识别. 因此提取过程应以分段为基础, 对其中的每个句法结构单元进行适配; 同时引入相应的用户心理模型, 以最大化相关质量指标.

然而, 视频序列的语义抽取和标注, 尚存在诸多难点. 同时, 对抽取出的语义如何作为元数据应用到媒体适配过程, 相关的研究和标准化过程也尚在进行之中.

6.2 用户主观测度

媒体的适配操作过程可能会导致用户主观上的感知、语义、理解等的变化. 而且, 即使是对同一媒体内容, 不同背景、任务等用法环境的用户感受, 也可能会完全不同. 然而, 现有的基于信号级的客观度量指标如 PSNR 等, 无法很好地衡量用户对媒体访问的主观感受.

因此, 需要引入相应的用户心理模型, 并通过对用户偏好和访问历史等数据的分析, 将客观质量指标如 PSNR 和主观质量指标如平均主观尺度 MOS (Mean Opinion Scale)或时间平滑度等进行关联. 这样, 才能在一定程度上预测和衡量适配过程的效用, 从而选择满足资源限制的最优适配策略. 相比而言, 这方面的工作才刚开始.

6.3 最大化用户体验

人为因素(human factor)如访问内容、任务、用

法环境、情感等,都将影响用户媒体访问的主观体验(experience)。然而,现有媒体适配研究主要集中在媒体 QoS 和网络 QoS 等客观指标的最优化,没有统一测度来衡量用户访问过程的主观感受,不同用户对媒体访问的体验难以定量分析和预测。

一种可行的方式是在媒体访问过程中,由用户给出其主观偏好值。适配引擎以此综合客观质量指标如 PSNR,寻求使得用户理解级的效用最大化的适配策略,并针对不同用户动态调整适配过程。问题是该过程需要用户交互,并且用户难以定量给出对不同媒体内容具体的偏好值。

媒体内容的个性化是增加用户体验的有效方法,例如:语义抽取得到的语义标注等,可以在语义级转码和内容个性化中得到应用。这方面已经进行了一定的研究<sup>[49-50]</sup>。

对媒体适配过程中影响用户体验的主观因素的定量分析以及将主观因素和客观质量指标相结合从而最大化用户体验等方面的研究,对通用媒体访问和数字媒体适配的发展,将起到积极的促进作用。

## 7 结 论

对通用媒体访问和媒体适配的详细分析使得可以建立其适配框架,提出统一的数学模型,并用一致的算法进行求解。

对媒体适配相关应用的分析表明,现有媒体适配的主要研究可以归结到该框架。因此,本文的研究统一了之前的主要工作,并对后续的媒体适配研究能起到一定的参考作用。

**致 谢** 作者对审稿人的宝贵意见及编辑的辛勤劳动表示衷心感谢!

## 参 考 文 献

- [1] Morhan R, Smith J R, Li C S. Adapting multimedia Internet content for universal access. *IEEE Transactions on Multimedia*, 1999, 1(3): 104-114
- [2] Vetro A, Christopoulos C, Ebrahimi T eds. *IEEE Signal Processing Magazine, Special Issue on Universal Multimedia Access*, 2003, 20(2)
- [3] ISO/IEC. Information technology-multimedia content description interface — Part 5, multimedia description schemes (15938-5). MPEG MDS Group, 2001
- [4] Manjunath B S, Salembier P, Sikora T eds. *Introduction to MPEG 7: Multimedia Content Description Language*. New York: Wiley, 2002
- [5] van Beek P, Smith J R, Ebrahimi T, Suzuki T, Askelof J. Metadata-driven multimedia access. *IEEE Signal Processing Magazine*, 2003, 20(2): 40-52
- [6] ISO/IEC. Text of ISO/IEC 21000-7 FCD — Part 7: Digital Item Adaptation. ISO/IEC JTC1/SC29/WG11/N5845, July 2003
- [7] ISO/IEC. Information technology: Multimedia framework — Part 7: Digital Items Adaptation (21000-7). 2004
- [8] Chang S F, Verto A. Video adaptation: Concepts, technologies and open issues//*Proceedings of the IEEE — Special Issue on Advances in Video Coding and Delivery*, 2005, 93(1): 148-158
- [9] Nakajima Y, Hori H, Kanoh T. Rate conversion of MPEG coded video by requantization process//*Proceedings of the IEEE International Conference on Image Processing*. Washington, DC, 1995, 3: 408-411
- [10] Eleftheriadis A. Dynamic rate shaping of compressed digital video [Ph. D. dissertation]. New York: Department of Electronic Engineering, Columbia University, 1995
- [11] Fung K T, Chan Y L, Siu W C. New architecture for dynamic frame-skipping transcoder. *IEEE Transactions on Image Processing*, 2002, 11(8): 886-900
- [12] Vetro A, Sun H, Wang Y. Object-based transcoding for adaptive video content delivery. *IEEE Transactions on Circuits and Systems for Video Technology*, 2001, 11(3): 387-401
- [13] Vetro A, Sun H, Divakaran A. Adaptive object-based transcoding using shape and motion-based hints. ISO/IEC M6088. Geneva, Switzerland, 2000
- [14] Huang T Y, Zheng C. Prioritized MPEG-4 audio-visual objects streaming over the DiffServ. *Journal of Electronic Science and Technology of China*, 2005, 3(4): 314-320
- [15] van der Schaar M, Radha H. A hybrid temporal-SNR fine granular scalability for Internet video. *IEEE Transactions on Circuits and Systems for Video Technology*, 2001, 11(3): 318-331
- [16] Huang T Y. The quality definition and an optimal algorithm for MPEG-4 FGS video streaming. *Chinese Journal of Computers*, 2006, 29(5): 751-759 (in Chinese)  
(黄天云. 一类 MPEG-4 FGS 流视频传输的质量定义及算法. *计算机学报*, 2006, 29(5): 751-759)
- [17] van der Schaar M, Andreopoulos Y. Rate-distortion-complexity modeling for network and receiver aware adaptation. *IEEE Transactions on Multimedia*, 2005, 7(3): 471-479
- [18] Vetro A, Timmerer C. Digital item adaptation: Overview of standardization and research activities. *IEEE Transactions on Multimedia*, 2005, 7(3): 418-426
- [19] Chang S F. Optimal video adaptation and skimming using a utility-based framework//*Proceedings of the International Workshop on Digital Communications (IWDC'02)*. Capri Island, Italy, 2002
- [20] Kim J G, Wang Y, Chang S F. Content-adaptive utility-based video adaptation//*Proceedings of the IEEE International Conference on Multimedia & Expo (ICME-2003)*. Baltimore, Maryland, 2003, 3: 281-284

- [21] Mukherjee D, Delfosse E, Kim J G, Wang Y. Optimal adaptation decision-taking for terminal and network quality of service. *IEEE Transactions on Multimedia*, 2005, 7(3): 454-462
- [22] Devillers S, Timmerer C, Heuer J, Hellwagner H. Bitstream syntax description-based adaptation in streaming and constrained environments. *IEEE Transactions on Multimedia*, 2005, 7(3): 463-470
- [23] Kim J G, Wang Y, Chang S F, Kim H M. An optimal framework of video adaptation and its application to rate adaptation transcoding. *Journal of Electronics and Telecommunications Research Institute, Korea*, 2005, 27(4): 341-354
- [24] ISO/IEC. Requirements and applications for scalable video coding. ISO/IEC JTC1/SC29/WG11/N6025, October 2003
- [25] Lewis R M, Torczon V, Trosset M W. Direct search methods: Then and now. *Journal of Computational and Applied Mathematics*, 2000, 124(1-2): 191-207
- [26] Huang Tian-Yun. Research advances on pattern searches in constrained optimization. *Chinese Journal of Computers*, 2008, 31(7): 1200-1215(in Chinese)  
(黄天云. 约束优化模式搜索法研究进展. *计算机学报*, 2008, 31(7): 1200-1215)
- [27] Torczon V. On the convergence of pattern search algorithms. *SIAM Journal on Optimization*, 1997, 7(1): 1-25
- [28] Audet C, Dennis J E. Analysis of generalized pattern searches. *SIAM Journal on Optimization*, 2003, 13(3): 889-903
- [29] Audet C, Dennis J E. A pattern search filter method for non-linear programming without derivatives. *SIAM Journal on Optimization*, 2004, 14(4): 980-1010
- [30] Audet C, Dennis J E. Pattern search algorithms for mixed variable programming. *SIAM Journal on Optimization*, 2000, 11(3): 573-594
- [31] Grosky W I, Mehrotra R. Image database management. *IEEE Computer*, 1989, 22(12): 7-8
- [32] Jagadish H V, O'Gorman L. An object model for image recognition. *IEEE Computer*, 1989, 22(12): 33-41
- [33] Chang S K, Hsu A. Image information systems: Where do we go from here? *IEEE Transactions on Knowledge and Data Engineering*, 1992, 4(5): 431-442
- [34] Hampapur A. Designing video data management systems [Ph.D. dissertation]. University of Michigan, 1994
- [35] Smoliar S W, Zhang H J. Content-based video indexing and retrieval. *IEEE Multimedia*, 1994, 1(2): 62-72
- [36] Grosky W I, Jain R, Mehrotra R. *The Handbook of Multimedia Information Management*. Upper Saddle River, NJ: Prentice Hall PTR, 1997
- [37] Gibbs S, Breiteneder C, Tsichritzis D. Data modeling of time-based media//*Proceedings of the ACM SIGMOD*. Minneapolis, MN, 1994, 23(2): 91-102
- [38] Hamakawa R, Rekimoto J. Object composition and playback models for handling multimedia data//*Multimedia Systems*. Berlin: Springer-Verlag, 1994, 2(1): 26-35
- [39] Bauschke H H, Hamilton C, Macklem M S, McMichael J S, Swart N R. Recompression of JPEG images by requantization. *IEEE Transactions on Image Processing*, 2003, 12(7): 843-849
- [40] Taubman D S, Marcellin M W. *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Pubs, 2002
- [41] Dogan S, Sadka A H, Kondoz A M. Efficient MPEG-4/H.263 video transcoder for interoperability between heterogeneous multimedia networks. *IEEE Electronic Letters*, 1999, 35(11): 863-864
- [42] Shanableh T, Ghanbari M. Heterogeneous video transcoding to lower spatial-temporal resolutions and different encoding formats. *IEEE Transactions on Multimedia*, 2000, 2(2): 101-110
- [43] Xin J, Lin C W, Sun M T. Digital video transcoding. *Proceedings of the IEEE*, 2005, 93(1): 84-97
- [44] de Cuetos P, Ross K W. Adaptive rate control for streaming stored fine-grained scalable video//*Proceedings of the NOSS-DAV'02*. Miami, Florida, USA, 2002, 12(14): 3-12
- [45] Vetro A, Sun H, Wang Y. MPEG-4 rate control for multiple video objects. *IEEE Transactions on Circuits and Systems for Video Technology*, 1999, 9(1): 186-199
- [46] Chiang T, Zhang Y Q. A new rate control scheme using quadratic rate-distortion modeling. *IEEE Transactions on Circuits and Systems for Video Technology*, 1997, 7(1): 246-250
- [47] Thang T C, Jung Y J, Ro Y M. Modality conversion for QoS management in universal multimedia access//*IEEE Proceedings—Vision, Image, and Signal Processing*, 2005, 152(3): 374-384
- [48] Thang T C. CE report on modality conversion preference: Part-I. ISO/IEC JTC1/SC29/WG11/M9495. Pattaya, 2003
- [49] Pereira F, Burnett I. Universal multimedia experiences for tomorrow. *IEEE Signal Processing Magazine*, 2003, 20(2): 63-73
- [50] Tseng B L, Lin C Y, Smith J R. Using MPEG-7 and MPEG-21 for personalizing video. *IEEE Multimedia*, 2004, 11(1): 42-53



**HUANG Tian-Yun**, born in 1973, Ph.D., associate professor. His current research interests include video flow analysis, optimal video transmission, quality of streaming video, and media adaptation, etc.

## Background

In order to meet the requirements of different users and all kinds of terminals, to access media contents widely distributed over networks, the concept of universal media access (UMA) was introduced in recent years, which had greatly enhanced research activities in this field, and accelerated standardization progresses such as MPEG-7 and MPEG-21. However, in heterogeneous network environment, how to maximize user experiences on media consumption and guarantee quality of service (QoS) based on user preferences and terminal capacities? There still exist many challenges.

Meanwhile, media formats and contents are continuously increasing and becoming more abundant, media adaptation has become necessary. Digital items adaptation (DIA) has been proposed in MPEG-21 in order to satisfy the requirements of UMA, and has attracted more attentions in recent years.

People have devoted in this field for years, such as JPEG images or MPEG I frames adaptation in size and color space, MPEG video transcoding to different formats (e. g. , MPEG-2 to MPEG-4, Real media to AVI) or into smaller size (e. g. , DCT coefficient requantization or dropping, frame dropping), etc. Probably the most salient work is bit rate adaptation for adaptive delivery of media content according to the network conditions, especially streaming of MPEG-4 finer granularity scalability (FGS), MPEG-4 video objects (VOs) over Internet and wireless, etc. But notice that almost all researchers are concentrated on constructing their own mathematical models and try to find the optimal solutions in different ways, e. g. , by dynamic programming or some elaborate heuristics. These researches only partly solved media adaptation problem under different constraints, there is yet no consistent way to describe, analysis and solve the problem in the

media adaptation framework.

Aimed at a unified solution for media adaptation, the author proposes a mixed-variable constrained optimization model in this paper, based on analysis of the relations among entities in MPEG-21 DIA. The model unified current researches in media adaptation, and can be solved by the direct searches. The author also proposes a hierarchical media adaptation structure, from the viewpoints of image understanding and video analysis. Leading research activities such as FD/CD transcoding, full scalable and object encoding video adaptation, multi-modality conversion and fusion, etc. , are illustrated by examples. Hotspots in media adaptation such as semantic abstraction of media contents, measurements on subjective quality, etc. , are also addressed in this paper. A thorough analysis of algorithms to solve the model had been conducted in related paper.

The author became interested in video traffic analysis, optimal video transmission and quality of streaming video since he was pursuing his Ph. D. degree in UESTC in 1999. He had published one book (Video traffic analysis and QoS management) and more than ten papers in journals and conferences. He holds a digital media lab in SWUN, and made some contributions to media adaptation with his students in recent years, such as rate control of FGS video, prioritized discarding of EL layers in video objects over the Diffserv, energy-aware FGS video streaming over wireless, etc. Now, he is dedicated in the assessment and measurement on subjective quality of scalable videos, and tries to apply it to wireless video streaming. This work is partly supported by the research fund for the State Ethnic Affairs Commission of China (05XN09)—"Research on the Key Issues of Video Streaming over Wireless".