

多媒体同步*

宋军 顾冠群

(东南大学计算机科学与工程系, 南京 210096)

摘要 多媒体同步是分布多媒体应用的一个重要需求。本文从时间相关媒体入手,就多媒体同步的多样性、同步特性、同步服务质量参数、同步时间概念模型以及同步关系的抽象和描述等方面予以较全面的分析和阐述。

关键词 多媒体,多媒体同步,时间相关媒体,同步模型

1 引言

多媒体工作站、分布系统和高性能计算机网络技术的发展和融合促成了分布多媒体系统的出现和发展,可以预计在未来的信息高速公路上的应用绝大多数都将是分布多媒体应用。应用领域将遍及工业、商业、金融、医学、教育、出版、办公、旅游和科学与工程研究等各个方面。分布多媒体系统以能集成地产生、存贮、传输、处理和表现多媒体信息为其特征。由于多媒体对象的内在相关性(如音频和视频的“唇同步”),故媒体的集成处理是分布多媒体系统的一个重要特征。集成的关键技术是信息的数字化表示和多媒体同步。一般来说,多媒体同步是指对象、数据项或数据流之间的时间、空间或逻辑的关系。在分布多媒体系统中,多媒体同步是指多媒体对象间时间关系的同步,多媒体时间同步关系是多媒体应用语义的一个重要组成部分。

多媒体与传统的数据类型的主要不同是引入了时间相关的媒体类型(如音频和视频),因此,同步特性是多媒体与传统媒体之间最重要的差别。目前,分布多媒体系统的研究还刚开始不久,还缺乏对多媒体同步全面而深入的研究。本文从形式化地研究时间相关媒体入手,从多方位对多媒体同步予以分析和研究,希望能为今后的分布多媒体系统的研究提供有益的帮助。

2 时间相关媒体

多媒体是多种各异的信息媒体的集成。按时间特性分类,多媒体可分为时间无关媒体和时间相关媒体。时间无关媒体是指那些信息值不随时间变化的媒体,如文本、图形和图象等传统的信息媒体;时间相关媒体是那些信息值与时间有关,并随时间变化的媒体,如数字语音和视频、音乐、动画以及通过一些压缩编码标准(DVI、MPEG、JPEG、CD-I等)而生成的媒体流等。

时间相关媒体通常是一系列信息单元(称作媒体元素)的高度结构化、时间化的集合体,表现为一个时间化的媒体流。媒体元素可是视频帧、音频采样和音乐音节等。为了更好地分析研究丰富多样而又性能各异的时间相关媒体,本文定义了时间流(Timed Stream)的概念,作为时间相关媒体基本的形式化抽象和描述工具。

定义1 一个时间流是三元组 $\langle e_i, s_i, d_i \rangle, i = 1, \dots, n$ 的一个有限序列。 e_i 是媒体元素,每个媒体元素都有一个元素描述符以记录如媒体类型、编码、表现参数等媒体相关的元素属性; s_i 和 d_i 分别是媒体元素 e_i 在时间流中的开始时间和持续时间(时段),满足 $s_i + 1 \geq s_i$,和 $d_i \geq 0$, i 是三元组 $\langle e_i, s_i, d_i \rangle$ 在时间流中的序列号。

根据时间流数据模型的特征参数,时间相关媒

* 国家“863”高技术计划和国家自然科学基金资助课题(课题号:863-511-9511-005和69372018)
收稿日期:1996-11-27;收到修改稿日期:1997-05-04

体的主要类型如下:

(1)同构流:元素描述符是常值,一个例子是 CD 音频,每个元素都有一样的形式。

(2)异构流:元素描述符是变化的,例如压缩格式中的编码参数在录入时变化的媒体。

(3)连续流: $s_{i+1}=s_i+d_i, i=1, \dots, n-1$,在这种时间流中的每个时间值都对应于唯一一个元素(也就是,对于每个时间值 $t: s_1 \leq t \leq s_n + d_n$)。数字音频和数字视频是连续媒体最常见的例子。

(4)非连续流: $s_{i+1} \neq s_i + d_i, \exists i \in \{1, \dots, n-1\}$,在这种情况下,元素间有“间隙”或“重叠”。例如,在音乐表现中,音符媒体元素正在表现,这时一个和音元素要求重叠其上。

(5)事件流: $d_i=0, i=1, \dots, n-1$,这是非连续流的一个特殊例子,媒体元素是时段为零的“事件”。一个例子是媒体元素为音乐事件“起始符 X”和“终止符 Y”的 MIDI。

(6)常频流:连续流并且元素时段是固定的。这方面的例子包括帧速固定的数字视频和许多种音频。

(7)常速流:连续流并且元素大小(即元素数据量)与时段的比率是常数,一些时间流具有固定的数据速率(该速率是元素大小与时段的比率)。

(8)统一流:连续流并且元素的大小和时段都是固定的,这是常速流的一个常见特例。例子包括许多“原始”(未压缩)的数字音频和数字视频。

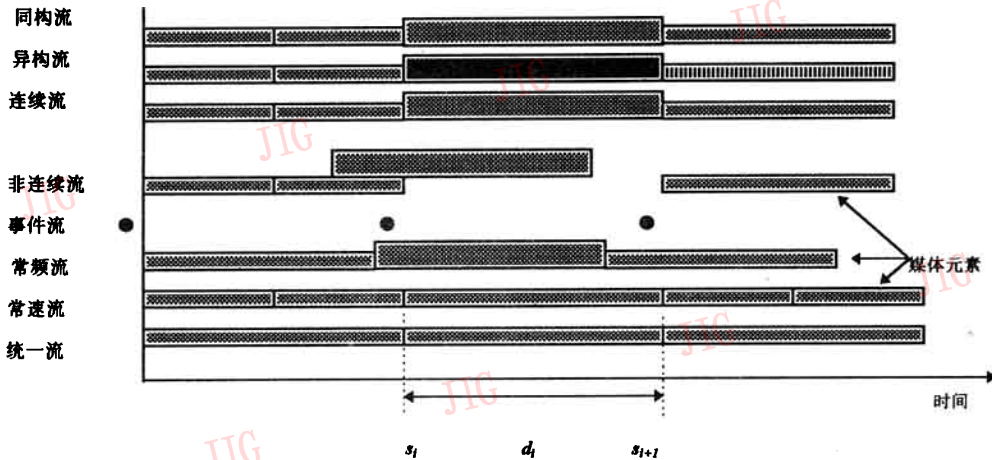


图 1 时间相关媒体类型示意图

Fig. 1 Schematic diagram of different types of time-dependent media

上述时间流的分类即包容又区分了众多的音频、视频、音乐和动画的数字化表示。从这点上说,它提供了时间相关媒体的一个非常有用的分类。例如,可以说音频和视频是连续媒体,因为音频和视频通常表示为连续流。上述媒体类是比较通用的时间相关媒体类型,通过以上分析也可以看出时间相关媒体的内部结构和时间关系的复杂性。

图 1 图示了每个媒体元素的开始时间、时段、大小(矩形面积)和元素描述符(矩形的填充图案)。图中,媒体元素用矩形(或圆点)表示。

3 多媒体同步的分类

3.1 基于同步对象的分类

按同步涉及的对象,多媒体同步分为三类:媒体

内同步、媒体间同步和人机交互同步。

(1)媒体内同步:指时间相关媒体内各个媒体元素间的时间同步关系。

上节已对此进行了系统、形式化的分析。

(2)媒体间同步:指不同媒体间的时间同步。

这种同步关系可能是不同的时间相关媒体间的,如音频与视频的“唇同步”;也可能是时间无关媒体与时间相关媒体通过后来建立的某种关系而产生的,例如在视频画面上加注释文字。

(3)人机交互同步:指不同媒体对象和交互对象之间的同步。

交互对象是指用户和计算机之间的通信,如用户用键盘或鼠标控制快进、倒放、暂停等。该类同步的主要特点是其同步发生时间的不确定性。

在分布多媒体系统中,这三类同步是密切相关的。媒体内同步的同步粒度最小,它是媒体间同步的

基础和前提,它们都由分布多媒体系统实现。人机交互同步则是在上述两种同步的基础上,用户和分布多媒体系统间交互的同步,它的同步粒度是最大的。

3.2 基于同步产生的分类

按同步关系的产生方法,多媒体同步分为两类:

(1)自然同步:指在数据采集时就有的内在时间同步关系。

一般存在于时间相关媒体内和时间相关媒体间,例如每秒25帧的视频流以及同时录入的音频、视频都有内在的时间同步关系,这种同步关系在表现时也要保持。

(2)合成同步:由应用程序后来所强加的时间同步关系。

经常通过合成同步将一些媒体片断组合成新的媒体。自然同步媒体也可以被用于构成新的合成同步媒体。由于实况数据(Live-data)是动态、实时的、无重排序性和预览性,因此合成同步只对存贮数据(Stored-data)有效。

4 同步特性与系统支持

多媒体同步具有动态特性,这表现在以下两方面:(1)多媒体对象的表现是允许在正常表现时间基础上有一定的时间变化(也就是说,在这一时段里多媒体对象的语义是有效的)。这意味着在一定范围内的流内延迟抖动和流间扭曲是能为人们所接受的。例如人们视觉可接受的视频最大延迟抖动为10ms,而可接受的视频与音频间的最大扭曲是80ms,并且当音频略超前于视频时同步效果反而是最好的。表1给出了人们视觉所能接受的媒体间扭曲值,这些值可被用作多媒体同步有关的服务质量参数。实际上,在分布多媒体系统中,异步系统(如计算机网络通信子系统)所引入的延迟是不可避免的^[1],所以多媒体体的动态同步特性是分布多媒体系统所固有的。(2)交互同步需要处理用户在多媒体表现过程中的输入,而用户操作的时间是不确定的、动态的。

因为同步特性是多媒体的内在特性和应用语法的组成部分,所以多媒体同步关系存在于多媒体的整个生命周期中。分布多媒体系统以能集成地产生、存贮、传输、处理、和表现多媒体信息为其特征,因此在整个媒体集成处理过程中,分布多媒体系统都要动态地保持多媒体同步。在分布多媒体系统中,影响多媒体同步的主要环节为存贮系统、操作系统和通

信系统。例如,从一个存贮设备存取多个同步媒体流,存取延迟、媒体流间的存取冲突和竞争都会破坏多媒体同步;当前分时操作系统不支持多媒体同步基于时限的实时处理要求;计算机网络通信系统具有很强的异步特性,表现在它有比较大的随机延迟和延迟抖动以及会丢失数据。这些都会导致多媒体同步异常,故需要研究出新的方法和技术以支持多媒体的同步存贮、同步处理和同步通信^[2]。

表1 多媒体同步服务质量参数

Table 1 QoS parameters for multimedia synchronization purposes

媒体		模式,应用	服务质量参数
视频	动画	有相互关系	+/- 120ms
	音频	唇同步	+/- 80 ms
	图形	重叠	+/- 240 ms
		不重叠	+/- 500 ms
文本	重叠	+/- 240 ms	
	不重叠	+/- 500 ms	
音频	动画	事件相关(跳舞)	+/- 80 ms
	音频	紧耦合(立体声)	+/- 11 μ s
		松耦合(多个参与者的对话)	+/- 120 ms
		松耦合(背景音乐)	+/- 500 ms
	图形	紧耦合(有评注的音乐) 松耦合(幻灯)	+/- 5 ms +/- 500 ms
	文本	文字注释	+/- 240 ms
鼠标	语音与所指项相关	-500/+700 ms	

所有的多媒体同步问题都可分为两类问题:多媒体同步关系的抽象和描述;提供同步机制来保持该同步关系。同步机制是为了克服延迟抖动和扭曲而采取的处理措施,在分布多媒体系统的不同子系统和层次中各不相同,而多媒体同步关系的抽象和描述则是实现多媒体同步的前提。

5 多媒体同步关系的抽象和描述

5.1 多媒体同步的时间概念模型

多媒体复杂的同步关系是通过其复杂的时间关系来表示的,因此多媒体同步关系的抽象和描述需借助于时间模型和形式化描述工具。

时间是可模型化的真实世界的一个方面,同一个时间序列可用不同的时间模型描述。根据所采用的基本时间单位,当前用于描述时间的概念模型主要有两类:第一类,时间是通过一维时间域上的点来

表示的;第二类则是用“时间片”(Interval)作为时间域上的原子单位。

5.1.1 基于时刻的时间模型

该模型的基本单位是事件(Event),每个事件是时间域上的一个零长瞬间点(时刻)。例如“下午四点”。该模型是通过描述事件的发生时刻来描述事件间的时序,事件间共有三种基本时序关系:一个事件可以“先于(<)”、“同时(=)”或“后于(>)”另一个事件。

一个时间模型不仅要能描述过去事件间的时序关系,而且要能描述未来事件间的时序关系。与过去事件相比,未来事件间的关系可能是非确定的。例如我们知道事件 e_1 不能在事件 e_2 前发生。显而易见,非确定的时序关系集是基本时序关系的幂集,共有 $8(=2^3)$ 种, $\emptyset, <, =, >, \neq, \geq, \leq, \Phi$ 。其中 $<, =$ 和 $>$ 是基本时序关系, Φ 是基本时序关系的全集 $\{<, =, >\}$, \emptyset 是空集,其它的如其符号所示,是两种基本时序关系的组合。

5.2.2 基于时间片的时间模型

时间片是该模型的基本时间单位。一个时间片是由两个时刻或它们间的时段来定义的,例如“100

毫秒”或“上午九点到下午五点”。

定义2 设 $[S, \leq]$ 是一个偏序集, a 和 b 是集合 S 的两个元素,且 $a \leq b$ 。那么,称集合 $\{x | a \leq x \leq b\}$ 为 S 的一个时间片,记作 $[a, b]$ 。 S 的任一个时间片 $[a, b]$ 有以下特性

- (1) $[a, b] = [c, d] \Leftrightarrow a = c \wedge b = d$
- (2) 如果 $c, d \in [a, b] \wedge e \in S \wedge c \leq e \leq d$, 那么 $e \in [a, b]$
- (3) $\# ([a, b]) > 1$

该模型用时间片来描述事件的发生时间,用事件间的相对位置来表示事件的时序。两个时间片间相互共有十三种基本时序关系^[3],可分为串行和并行两类时序关系。表2综述了基本的时间片关系集(除“相等”关系外,其它六对时间片关系是互逆的),其中 X 和 Y 表示两个时间片, Bx 和 Ex 分别是时间片 X 的开始和结束时刻。同时表2还表明,13种基本的时间片关系可用3种基本的时刻关系的交集来表示。

和时刻关系相似,非确定的时间片关系集是基本时间片关系的幂集,共有 $2^{13} = 8192$ 种关系。

表2 两个时间片间的基本关系

Table 2 Basic relationships between two intervals

关系	表示符	逆关系	时刻表示	示例	分类
X before Y	<	>	$Bx < Ex < By < Ey$		串行
X meets Y	m	mi	$Bx < Ex = By < Ey$		串行
X overlaps Y	o	oi	$Bx < By < Ex < Ey$		并行
Y finishes X	f	fi	$Bx < By < Ex = Ey$		并行
Y during X	d	di	$Bx < By < Ey < Ex$		并行
X starts Y	s	si	$Bx = By < Ex < Ey$		并行
X equals Y	=	=	$Bx = By < Ex = Ey$		并行

5.2 多媒体同步关系的抽象和描述

多媒体同步时序关系是通过时间模型来抽象和描述的。它是媒体元素间一种相对的时间关系,每个媒体元素都有其表现时段,而基于时间片的时间模型更适于描述相对时序关系,并且基于时间片的时间模型比基于时刻的时间模型具有更强的时间描述

能力,所以当前都采用该模型作为多媒体同步关系的时间概念模型,时间片表示多媒体元素的表现时段,它们的相互位置表示多媒体元素的时间相关性,即同步关系。在实际应用中,多媒体的同步关系是多个媒体流内众多媒体元素间复杂的时序关系,故相应的多媒体同步时间模型也就比较复杂。因此,为了便于计算机应用和更好地研究多媒体同步关系及内

在机制,需要采用形式化描述工具来建立形式化的多媒体同步模型。当前比较有效、可供选择的形式化建模工具主要有 Petri 网、有限自动机和同步编程语言等。因为 Petri 网既是一个图形模型又是一个数学模型,具有很强的描述能力和建模能力,所以基于 Petri 网的模型提供了一种描述多媒体同步关系的好方法。对象复合 Petri 网(OCPN)模型^[4]是一个增强型时间 Petri 网模型,每个具有表现时段的多媒体对象用一个时间化库所表示,而每个变迁则表示一个同步点。一个有解说、图象和视频段的多媒体信息片断的 OCPN 模型如图 2 所示。

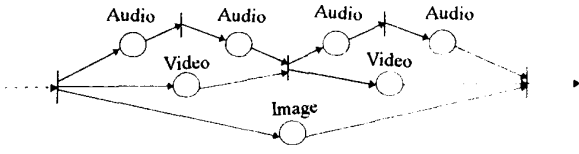


图 2 一个 OCPN 的多媒体应用实例

Fig. 2 A multimedia application example using OCPN

6 结束语

多媒体同步是指多媒体对象间时间关系的同

步,是分布多媒体应用的一个重要需求。多媒体与传统数据类型的主要不同是引入了时间相关的媒体类型(如音频和视频)。因此,本文提出了一个“时间流”数据模型,对时间相关媒体的特性予以形式化的描述和分类分析。多媒体同步具有多样性、复杂性和动态同步特性,因此多媒体同步关系的抽象和描述是实现多媒体同步的前提。基于时刻的时间模型和基于时间片的时间模型是当前两个主要的时间模型,其中基于时间片的时间模型更适合作为多媒体同步的时间概念模型。为了便于计算机应用和更好地研究多媒体同步关系及内在机制,需要采用形式化描述工具来建立形式化的多媒体同步模型。多媒体同步对计算机系统(特别是网络通信系统)提出了一系列新的研究课题。

参考文献

- 1 顾冠群,宋军. 高速多媒体信息通信. 电信科学,1995,11(11):35~38.
- 2 宋军,顾冠群. 多媒体通信媒体间同步技术综述. 电信科学,1996,12(9):3~9.
- 3 Allen J F. Maintaining knowledge about temporal intervals. Communications of the ACM, 1983, 26(11): 832~843.
- 4 Little T D C, Ghafoor A. Synchronization and storage models for multimedia objects, IEEE journal on selected areas on communications, 1990,8(3): 413~427.



宋军,东南大学计算机系博士生,研究方向为计算机网络与通信,现在的主要研究兴趣是多媒体同步通信、高性能计算机网络和 CSCW。曾参与科技攻关课题、国家“863”课题和国家自然科学基金课题的科研工作,其中江苏省“七五”重大科技攻关项目“异种大中型计算机远程 OSI 网络”荣获 1995 年度国家科学技术进步三等奖;现已发表论文十多篇。

Multimedia Synchronization

Song Jun, Gu Guanqun

(Department of Computer Science and engineering, Southeast University, Nanjing 210096)

Abstract Multimedia synchronization is a significant requirement for distributed multimedia applications. Beginning with time-dependent media, many aspects are studied comprehensively in this paper, such as the multiformity of multimedia synchronization, synchronization characteristics of multimedia, quality of service for synchronization purposes, the conceptual time models for multimedia synchronization, the abstraction and description for the temporal relations of multimedia synchronization.

Keywords Multimedia, Multimedia Synchronization, Time-dependent media, Synchronization model.