

动态上下文环境下的实时分布式多媒体系统框架

孙洛 赵彦钧 王垚 陶霖密 徐光祐

(清华大学, 计算机科学与技术系, 清华信息科学与技术国家实验室, 北京 100084)

摘要 伴随着多媒体采集和处理设备成本的不断降低, 以及处理目标的不断复杂化, 多媒体信息系统正朝着大规模数据分析和管理系统靠拢, 因此如何有效支持实时分布式多媒体信息所需的实时采集、压缩、传输、处理、描述、反馈与存档, 以及后期的浏览与检索, 已经成为多媒体计算研究的重要问题。为了有效支持实时分布式多媒体信息的实时采集、传输、处理、描述、存档等的需要, 在分析动态上下文环境下实时分布式多媒体系统对底层采集、压缩和传输以及信息的表达和存储的普遍需求的基础上, 提出并实现了一个通用的实时分布式媒体信息处理框架, 以便为实时分布多媒体系统提供基础服务支持。该框架包括软件平台和信息系统两部分。前者用于将逻辑功能与物理设备隔离开来, 完成实时采集、压缩、传输和解压等功能; 后者通过统一信息表达格式, 将数据与特定的采集处理模块隔离开来, 同时还提供了信息管理的功能, 以便于用户后期检索。这两部分共同将多媒体系统所需要的实时采集、处理、压缩、描述、反馈与存档以及离线浏览与检索等服务融入一致、统一的框架中。基于该框架的会议系统的相关实验表明, 该框架具有良好的开放性、可配置性和可扩展性。

关键词 框架 动态上下文 实时 分布式

中图法分类号: TP316.5 文献标识码: A 文章编号: 1006-8961(2008)04-0794-07

A Real-time Distributed Multimedia System Framework under Dynamic Context Environment

SUN Luo, ZHAO Yan-jun, WANG Yao, TAO Lin-mi, XU Guang-you

(Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084)

Abstract With the decreasing price of multimedia capturing and processing devices, as well as the capability to capture more complex, and sophisticated objectives, multimedia systems are facine the challenge of large data analysis and management. How to handle real-time capturing, compressing, transferring, decompressing as well as information representation and management in a multimedia system has become an important research topic. We propose a real-time distributed multimedia system framework for dynamic context environment based on a survey of the requirements of multimedia systems under dynamic context environment. It can be divided into two parts, the software infrastructure and the information system, fusing all needed functions in a real-time distributed multimedia system into a unique framework. The former is used to encapsulate the function of capturing, compressing, and decompressing and so on, isolating logical functions from physical devices. The latter is used for information representation, isolating information from special capturing or processing devices, and it also provides the capacity of information management, which means users can search for specified information later. The experiments of the meeting system based on this frame show the openness, configurability and extensibility of our framework.

Keywords framework, context, real-time, distributed

基金项目: 国家自然科学基金资助项目(60673189)

收稿日期: 2007-08-30; 改回日期: 2007-12-24

第一作者简介: 孙洛(1982 ~), 男。2004年获清华大学计算机系工学学士学位, 现为清华大学计算机系博士研究生。主要研究方向为计算机视觉、人机交互和普适计算。E-mail: sunluo00@mails.tsinghua.edu.cn

1 引言

近年来随着多媒体采集和处理设备成本的不断降低以及大容量存储设备的广泛应用,多媒体信息已经深入到人们日常生活的各个角落。多媒体系统中传感器的数量越来越多,其所需要的计算处理能力也随之迅速增加;日益庞大的可利用多媒体数据亦需要有效的表达和管理。这类应用不仅需求与一些重要的社会和经济问题相关,并且在许多专业和消费应用方面都是急需的,尤其是在网络技术高度发展的今天。因此,如何有效支持分布式多媒体信息的实时采集、压缩、传输、处理、描述、反馈与存档以及后期的浏览与检索,已经成为多媒体计算研究的首要问题之一^[1,2]。

如今多媒体系统正朝着实时、分布式的发展方向发展,对信息表达和管理的需求也日益增强,正朝着大规模数据分析和管理系统靠拢^[2]。近些年来研究人员通常将精力集中在如何更有效率地解决信息处理问题。迄今为止,仅有很少的研究着眼于整个多媒体信息处理的全过程,并尝试解决从数据采集、处理一直到结果表达和检索的全过程。其中最具有代表性的工作是美国加州大学圣地亚哥分校计算机系提出的 DIVA (distributed interactive video array) 框架^[3,4]和 IBM 的 S3 (smart surveillance system) 系统。

DIVA 是分布式监控系统结构,主要包括名为 NeST (networked sensor tapestry) 的数据融合平台^[5,6],其用途主要是提供一个虚拟的 3 维显示界面,用来集中地显示各种传感器数据及数据分析模块对传感器数据的分析结果。NeST 软件结构中的通信都由一个中央服务器 (server core) 控制,各种传感器数据被封装成 XML 格式,并通过中央服务器存储到知识数据库 (knowledge base database) 当中。DIVA 代表的结构及相关的研究是十分出色的,不过在构建系统的过程中, DIVA 也没有提出独立的中间件,这样虽然每个模块设计起来比较灵活,但是可移植性比较差。

S3^[2]是 IBM 提出的智能视频监控系統,其在设计上强调开放性和扩展性,以确保最终用户可以利用已较为成熟的开放标准的开发工具来方便快捷地为系统添加新组件,以增强系统功能。整个系统由智能监控引擎 (smart surveillance engine, SSE) 和大规模监控中间件 (middleware for large scale surveillance, MILS) 两个关键组件构成,前者为视觉处理模

块提供接口,并协调处理模块的运行顺序;后者用于对监控结果存档、索引,并提供检索服务。Brown 等人提出了一个通用的视频监控系统框架^[2],其虽具有较强的通用性和可扩展性,不过其中仍存在的问题,如缺少实时性支持、缺少多路数据的同步机制,以及没有将环境信息单独抽取出来组织在一起作为系统上下文信息的一部分。

一个通用的实时分布式多媒体信息处理框架不仅能够完成常见的多媒体系统具有的监控、查询等功能,还应该能够有效管理系统中出现的各种媒体信息,并具有良好的可扩展性,以方便后期进一步增强系统功能^[7]。本文正是希望能够深入探讨这个通用的实时分布式多媒体信息处理框架,并从技术上初步实现。该框架需要将系统所需要的实时采集、处理、描述、反馈与存档,和离线浏览与检索融合于一个统一的框架中。

2 框架需求分析

以人为本的计算环境 (human-centered computing) 中的多媒体系统不是盲目处理所有的多媒体数据,而是在当前上下文的引导下,有目的、有选择性地处理多媒体数据以及记录语义信息,并力求提高数据质量,以方便用户使用^[1]。完整的实时多媒体信息处理系统通常需要具有以下功能模块:

(1) 软件平台:分布式多传感器系统的支撑平台,除提供底层传输、压缩和解压等服务外,还需要为上层应用程序自动同步不同采样精度的数据,并通过逻辑传感器隔离物理传感器对系统的影响。

(2) 信息系统:该模块以统一的格式描述系统中涉及的各种信息,为系统提供存储和检索服务;同时该模块还需要支持场景描述,这关系到系统对环境的适应能力;

(3) 多媒体数据处理:该模块在软件平台和信息系统模块的支持下完成对数据的处理任务,其构成与多媒体系统的具体任务密切相关。

按照上述多媒体系统功能模块的划分,由于模块(3)和系统具体任务相关,且不同的任务需要不同的处理模块,因此通用框架的研究应该集中在模块(1)和模块(2)上,这是本文的主要研究内容。同时要确保框架可以通过增加不同的信息处理模块实现不同的处理任务。现将框架会面对的关键问题列举如下:

(1) 支持多任务的实时计算:由于多媒体系统

通常需要同时完成多任务的处理,因此在单机计算能力有限的情况下,多任务的实时处理需要有并行计算环境的支持。

(2) 多传感器的协同配合:在多传感器情况下,由于传输过程中的时延会造成多路视频数据之间同步的困难,因此框架需要确保从不同传感器得到的数据保持同步,并记录相应的时间关系。

(3) 支持从数据采集到信息检索全过程的信息管理;信息管理既要保存原始的视频数据,同时还要建立和存储监控系统处理所得到的元数据(metadata)。这两类数据需要有机地组织在一起,并通过有效的方案描述出来,以便为用户提供方便的信息的浏览和检索。

(4) 存储和管理上下文信息:从信息处理的角度上讲,系统不仅仅需要整合现有的各种处理模块,更重要的是要建立动态上下文模型(dynamic context model),以感知当前上下文,而且要以当前上下文为基础,进一步引导信息处理和存档系统。另外框架必须要对动态上下文信息加以支持。

3 软件平台

3.1 分层服务结构

为了确保软件平台能够为高层应用提供更为通用和鲁棒的服务,首先需要了解依赖于软件平台运行的应用程序模块的特点。应用程序模块按照其处理任务的不同可以划分为 3 个类型^[8,9]:

(1) 传感器相关应用,该类型主要是底层传感器数据分析模块,如前景检测等,其特点是处理模块和多媒体流一一对应;

(2) 目标物体相关应用,该类型主要处理的是和物体相关的信息,如人的姿势等,其特点是和物体对应,通常依赖于其他模块提供的基本信息;

(3) 场景相关应用,该类型需要获得多路多模态传感器数据,并进行数据融合,如整个场景下的跟踪以及一些更高层的上下文信息的检测推理模块等。

根据应用程序模块的这种特点,本文将软件平台提供的服务划分成底层服务和上层服务两个层次(如图 1 所示)。底层服务只包括多媒体流相关服务和消息流相关服务,其中多媒体流相关的服务包括多媒体数据的采集、压缩与解压、传输、存储、缓冲和同步几个方面;消息相关的服务包括消息数据的转发、缓冲两个方面。底层服务主要针对摄像机相关的应用,其关注的是数据流的通道。上层服务的

实现需依赖于底层服务,例如“获取同步数据”的服务就依赖于底层传感器和消息数据的获取。有了这种服务的分层,就可以将整个软件平台的设计和实现更加条理化,其不仅可将底层服务和上层服务放在不同的阶段来实现,同时可在设计底层服务的同时,为上层服务预留接口等。

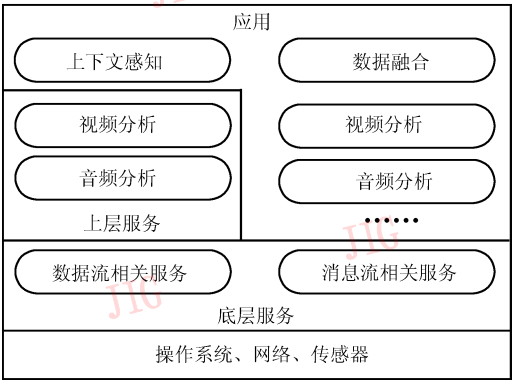


图 1 软件平台和上层应用之间的层次结构关系

Fig. 1 Relationship between software platform and application

3.2 后台服务进程

软件平台部分和其他部分的关系是构建整个框架的系统的两个基本的部分。本文将平台和每个特定功能的视频理解算法都封装成单独进程的方式。由于目前的研究并不针对某个特定的系统,而是针对一系列类似的系统,因此使用独立于上层应用的结构能够使平台具有更好的适应能力;此外,由于目前上层的视频内容理解等应用仍然处在研究或完善的阶段,结构互异,差别很大,不方便集成到一个程序当中,而将其安排成独立的进程则不仅可以在保持程序原有特点的情况下集成到系统当中,而且此时某个视频分析算法在运行时的不稳定也不会影响到底层服务的运行和其他分析算法,从而使得整个系统更鲁棒。

图 2 显示了软件平台在组织上的基本思路,图 中最里面的部分是物理传感器;中间一圈是服务进程,它们通过局域网相互连接形成一个协作的服务平台;最外层是上层应用,它构建在平台之上,通过平台获取传感器数据和操作传感器,并完成相互之间的协调。在上层应用的工作过程中,仅需要和本地的服务进程通过一套接口通信。由于底层服务和上层服务同时包含在服务进程内,所以对于应用而言,服务进程的接口也是在不断扩展的,但是这种扩展不会影响到原先的应用。

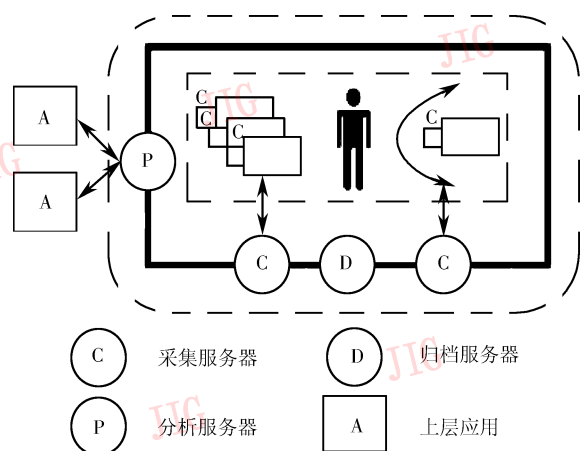


图2 软件平台的结构

Fig. 2 Logical connection of software platform

由于每个参与系统的主机有且仅有一个服务进程,因此服务进程对应的是计算资源。在系统的组织过程中,可根据服务进程需要完成的任务将其分为以下3类:采集服务器、分析服务器和归档服务器。

(1) 采集服务器,其运行在那些和视频传感器相连接的主机上,负责和摄像机或采集卡接口连接,并把数据分发到其他主机上面,采集服务器同时还负责视原始视频数据的保存任务;

(2) 分析服务器,该服务器主要提供视频内容理解算法所需要的运行环境,同时接收传感器数据,并提供给上层应用,最后把结果发送到归档服务器或其他分析服务器,其意义在于计算资源的扩展,当系统中运行的算法比较复杂时,则必然需要更多的计算资源来满足实时性的要求,因此引入了分析服务器之后,参与视频内容分析的主机就可以根据计算的情况动态地增减;

(3) 归档服务器,主要负责汇集分析服务器的处理结果,并建立档案,而基于归档服务器的应用则可以为用户提供视频文件的检索,同时由于归档服务器的默认会保持和其他所有服务器的连接,所以可以设计相应的调度程序。

4 信息系统

本系统框架中的信息系统是以 MPEG-7 作为基本存储结构。MPEG-7^[10] 是 MPEG 组织提出的多媒体描述语言,正式名称为多媒体内容描述接口 (multimedia content description interface)。由于该标准以 XML 作为基本存储格式,从而确保了通用性,其

能够为各类多媒体信息提供一种标准化的描述。基于 MPEG7 的信息描述结构 (THVS) 如图 3 所示,该结构是在 MPEG7 标准描述方案的基础上,增加了适合于多媒体系统的元素,并采用了底层语义层 (LLevel)、高层语义层 (HLevel)、上下文信息 (Context) 3 层结构,不同层次之间以图连接 (Graph Link)。其中上下文信息主要用于描述监控系统中的静态上下文信息;底层语义层用来描述多媒体数据经过底层特征抽取之后得到的信息;高层语义层则用来描述底层特征经过进一步抽取、推理后得到的高层语义信息。

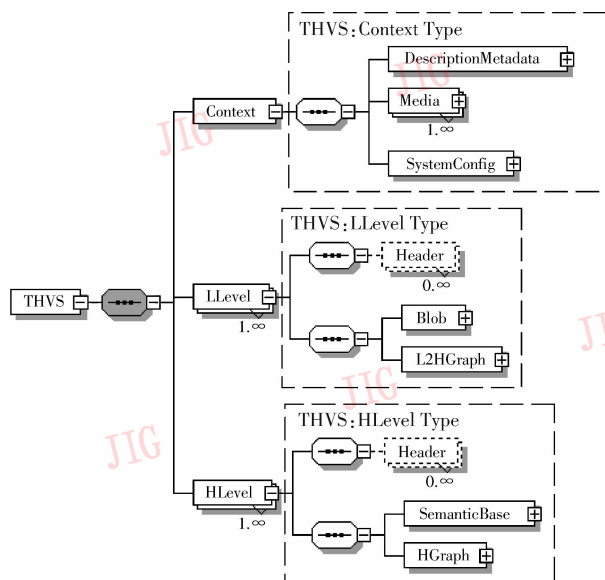


图3 信息系统描述结构

Fig. 3 Description structure of information system

底层特征与高层语义之间的关系是通过高层底层语义图连接 (L2HGraph) 和高层语义图连接 (HGraph) 来表示,例如某一前景物体属于人物甲的身体,即是底层与高层之间的关系。同时高层语义图 (HGraph) 连接还可以表示高层语义之间的内部关系,例如人物甲与人物乙在某时刻同时进入大厅,即是高层语义之间的联系。这种以图的形式来连接底层特征和高层语义的方法具有以下优点:①图的表示方法可以在不改变已有数据内容的情况下进行纠错;②可扩展能力强,易于扩展到概率描述。

4.1 底层特征表达

由于底层特征的数据量较大,因此选取表述方法的时候尤其需要考虑到系统的分布性和实时性。底层特征通常和传感器数据紧密结合,通常具有一定的采样间隔。对于底层特征的表述通常采用基于片段 (segment) 或者基于区块 (blob) 的方法。本文

采用片段和区块相结合的方法,片段作为大粒度单位,而以区块描述精细粒度特征。需要注意的是,片段中并不一定需要包含区块信息,这样的片段可以用来描述一段时间内数据的整体特征。这对于音频数据尤其重要,因为音频数据并不存在空间中区块的概念,故需要处理的仅有瞬时或片段数据。

4.2 高层特征表达

MPEG-7 多媒体描述语言不仅提供了丰富的高层语义描述,同时通过 MPEG-7 内嵌的数据定义语言 (data definition language) 支持,用户也可以自定义新的标记,以确保系统的扩展性,并可以满足视频监控系统的需要。高层语义层采用了 MPEG-7 标准中的语义描述子 (SemanticBase) 是通过高层语义图来描述高层语义物体与底层特征之间的联系,以及高层语义之间的联系。

底层特征与高层语义之间的联系采用高层语义图进行连接,例如:

```
<Relation xsi:type = "SegmentSemanticBaseRelationType"
name = "hasMediaPerceptionOf" source = "#Body_1"
target = "#Blob_1"/>
<Relation xsi:type = "SegmentSemanticBaseRelationType"
name = "hasMediaPerceptionOf" source = "#Body_1"
target = "#Blob_3"/>
```

上述描述即表示高级语义实体 Body_1 在视频图像中的区块为 Blob_1 和 Blob_3。这种表示方法还可以在不改变已有数据的情况下进行纠错,例如已经过一段时间的处理发现,由于 Body_1 和 Body_5 事实上是同一物体,因此可以通过添加一句描述语句来使二者等价:

```
<Relation xsi:type = "SemanticBaseSemanticBaseRelation"
name = "equivalentTo" source = "#Body_2" target = "#Body_5"/>
```

高层语义之间的联系同样由高层语义图来连接:

```
<Relation xsi:type = "ObjectEventRelationType"
name = "agentOf" source = "# Person-3" target = "#Run"/>
```

4.3 上下文表达

当前框架下的上下文信息包含如图 4 所示三大部分。首先是元数据生成信息,其用于描述当前元数据的采集时间、地点、人物,并直接采用了 MPEG-7 标准中定义的元数据描述类型 (Description Metadata);其次是多媒体信息 (Media),即各摄像机

采集的视频数据参数 (MediaInstance)、摄像机标定数据 (CameraCalibration) 和 PTZ 摄像机参数,其中视频数据的参数是采用 MPEG-7 标准的时间定位 (MediaTime) 和文件定位 (MediaLocator),摄像机标定和 PTZ 摄像机参数可以通过计算机视觉算法标定得到;最后一部分是监控计算机系统参数 (System Config),以主机 (Host) 为基本描述单位,由于本文的信息管理框架面对的多路视频监控系统是分布式的,因此需要给出相关的分布式计算环境的配置。

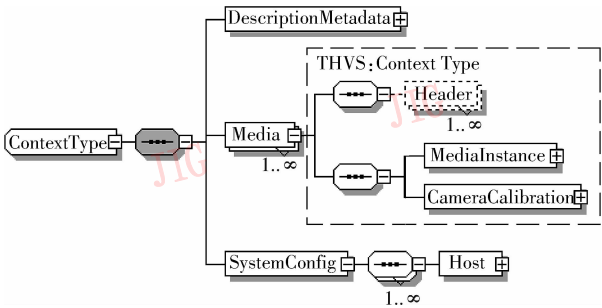


图 4 基本上下文信息描述方案
Fig.4 Basic context description

5 实例系统和实验结果

为验证本文提出的系统框架的实用性,利用本框架搭建了会议室环境来对框架的功能进行检验,并对性能作出评价。这是一个用于小规模学术讨论的会议室,并配有丰富的计算和采集设备,其物理设备布置如图 5 所示。另外还有若干普通 USB 摄像

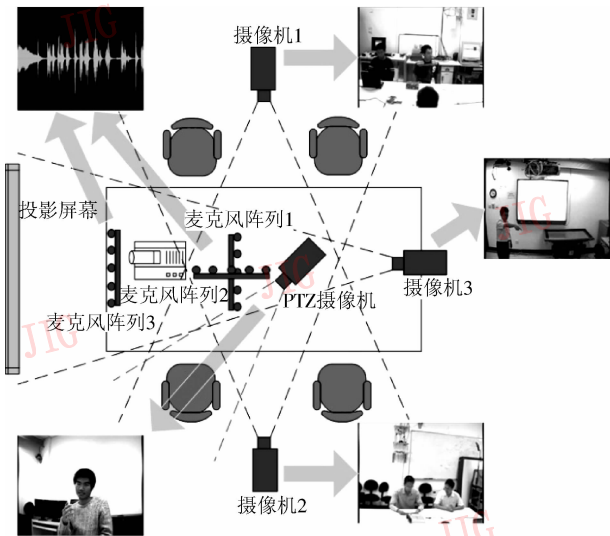


图 5 会议系统设备布置
Fig.5 Equipment setting in the meeting room

头和 CCD 摄像机以及云台。整个系统可以分为在线和离线两部分,前一部分用于从实时多媒体数据中提取语义化信息,并加以存储,同时对 PTZ 摄像机进行实时控制;后者用于浏览和管理已存储的信息。

该会议系统支持动态上下文的推理,并使用动态贝叶斯网作为推理模型。上下文信息(Context)以时间长短划分为数个层次,整体结构如图 6 所示。其中每个层次都有较强的语义特征,其中最顶层为会议场景自身,其表明整个上下文的变化过程都是在会议场景下;第 2 层用于从较高的语义层次描述当前状况,其主要包括讨论、演讲等;第 3 层用于表示与会者如何与环境或者其他与会者交互;

对于实时性需求较高的应用而言,延迟是一个十分重要的指标,延迟实际上包括了采集延迟、压缩延迟、传输延迟和解压缩延迟以及缓冲数据的延迟几个方面;系统延迟的测试方法是采用文献[11]中提出的测试方案,对于每种情况测试 10 次取平均值,得到的实测数据如表 1 所示。测试结果表明,对

于服务器而言,传输不同规格的视频流的传输延迟并没有很大变化。

表 1 媒体流传输性能测试
Tab. 1 Latency of video transferring

协议	分辨率	比特率(Kbps)	数量	延迟(ms)
UDP	352×288	500	1	228
UDP	352×288	500	3	225
UDP	352×288	500	3	237
UDP	640×480	2 000	1	227
UDP	720×576	2 000	1	212
TCP	352×288	500	3	215
TCP	640×480	1 500	3	257

对于控制命令传输性能的测试,本文参考文献[12]中的测试方法,即通过发送和接收到数据的时间差计算出延迟。为了和真实情况更加接近,测试过程中在传送消息数据的同时,在系统中传输一些视频流,其测试结果如表 2 所示。

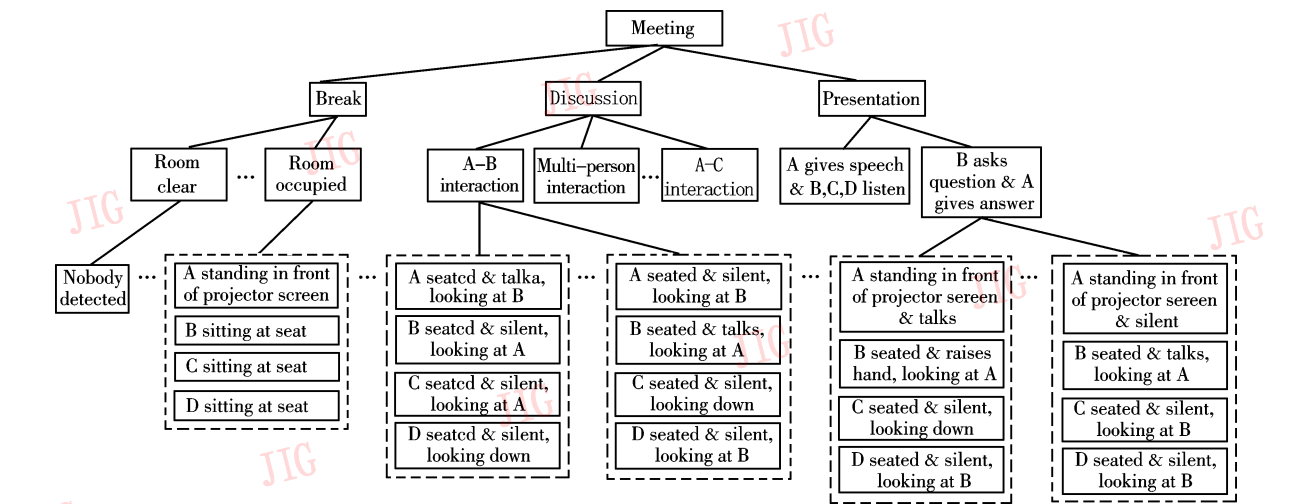


图 6 会议系统上下文层次结构

Fig. 6 Context hierarchy in the meeting system

表 2 消息传输性能测试

Tab. 2 Latency of message transferring

视频流数量	视频流码率(Mbps)	UDP 传输延迟(ms)
4	1	5
2	2	4
4	2	60
6	2	130

图 7 是系统对真实数据运行结果的一部分,图 7

表明,Person-3 是行走这个动作的执行者,Person-3 的名字为 Wang Yan,行走时间是从视频数据的 3 分 15 秒开始,延续 10s;同时 Person-3 具有语义实体 Body_1 和 Head_1。

实验表明,本框架不仅能够有效屏蔽不同硬件设备和网络结构的差异,不但可为上层应用提供统一的数据流获取和消息数据发送与接收接口,还提供了统一的信息模型,这就极大地简化了不同设备、

处理模块之间的通信协议,同时能够有效支持分布式多媒体信息的实时采集、压缩、传输、处理、描述、反馈与存档以及后期浏览与检索的全过程。

```
< SemanticBase xsi:type = "AgentObjectType" id = "Person - 3" >
< Agent xsi: type = "PersonType" >
< MediaOccurrence >
  < MediaLocator xsi: type = "TemporalSegmentLocatorType" >
    < MediaTime >
      < MediaTimePoint > T00:03:15 </MediaTimePoint>
      < MediaDuration > PT10S </MediaDuration>
    </MediaTime >
  </MediaLocator >
</MediaOccurrence >
< Name >
  < GivenName > Yan </GivenName>
  < FamilyName > Wang </FamilyName>
</Name >
</Agent >
<Object id = "Body_1"/>
<Object id = "Head_1"/>
</SemanticBase>
< Relation xsi: type = "ObjectEventRelationType" name = "agentof"
source = "#Person_2" target = "#Run"/>
```

图 7 会议系统信息表达示例

Fig. 7 Representation sample in the meeting system

6 结 论

如何有效支持分布式多媒体信息所需的实时采集、压缩、传输、处理、描述、反馈与存档以及后期的浏览与检索,已经成为多媒体计算研究的重要问题。针对该问题,本文通过分析实时分布式多媒体系统对底层采集、压缩和传输以及信息的表达和存储的普遍需求,提出并实现了一个通用的实时分布式媒体信息处理框架。在该框架的支持下,处理模块可以自由地访问系统中所提供的各种资源和服务,而无需关注整个多媒体系统的物理结构;同时用户可以自由装配处理模块,以完成不同的处理任务。通过基于该框架的会议系统的相关实验表明,该框架具有良好的开放性、可配置性和可扩展性。

参考文献 (References)

1 Maja Pantic, Alex Pentland, Anton Nijholt, *et al.* Human computing and machine understanding of human behavior: a survey[A]. In: Proceedings of International Conference on Multimodal Interfaces

[C], Banff, Alberta, Canada, 2006; 239 ~ 248.

2 Brown L, Hampapur A, Connell J, *et al.* IBM smart surveillance system (S3): An open and extensible architecture for smart video surveillance[A]. In: Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance[C], Como, Italy, 2005; 318 ~ 323.

3 Trivedi M M, Gandhi T L, Huang K S. Distributed interactive video arrays for event capture and enhanced situational awareness [J]. IEEE Intelligent Systems, 2005, 20(5) :58 ~ 66.

4 Trivedi M M, Huang K S, Mikic I. Dynamic context capture and distributed video arrays for intelligent spaces[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part A, 2005, 35(1) :145 ~ 163.

5 Fidaleo D A, Nguyen H A, Trivedi M. The networked sensor tapestry (NeST): a privacy enhanced software architecture for interactive analysis of data in video-sensor networks [A]. In: Proceedings of ACM International Workshop on Video Surveillance and Sensor Networks[C], New York, USA, 2004: 46 ~ 53.

6 Fidaleo D A, Schumacher R E, Trivedi M M. Visual contextualization and activity monitoring for networked telepresence [A]. In: Proceedings of ACM International Workshop on Effective Telepresence [C], New York, USA, 2004: 31 ~ 39.

7 Stanford V, Garofolo J, Galibert O, *et al.* The NIST smart space and meeting room projects: signals, acquisition, annotation and metrics [A]. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing[C], Hong Kong China, 2003: 736 ~ 739.

8 Wang Yan. Mug Shot Acquisition Based on in Formation Fusing of Distributed Cameras[M]. Beijing: Tsinghua University Press, 2006. [王焱. 基于分布式摄像机信息融合的人脸跟踪和获取[M], 北京: 清华大学出版社, 2006.]

9 Wang Yao, Tao Lin-mi, Liu Qiang, *et al.* A flexible multi-server platform for distributed video information processing [A]. In: Proceedings of International Conference on Computer Vision Systems [C], Bielefeld, Germany, 2007: 102 ~ 111.

10 MPEG-7 overview[EB/OL] . <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>.

11 Francois R J, Medioni G G. A modular middleware flow scheduling framework[A]. In: Proceedings of IEEE International Conference on Multimedia and Expo[C], Lusanne, Switzerland, 2002: 429 ~ 434.

12 Xie Wei-kai, Shi Yuan-chun, Xu Guan-you, *et al.* Smart platform: a software infrastructure for smart space [A]. In: Proceedings of IEEE International Conference on Multimodal Interfaces[C], Washington DC, USA:IEEE Computer Society, 2002: 429 ~ 434.