

## 摘要

综合媒体并不是一种具体的媒体形式，而是文本、动画、音频、视频、交互性信息的综合表达，是由多种媒体信息有机构成的动态可交互集合，支持按照指定的时间、空间布局同步呈现各种媒体，可以描述各媒体元素时间、空间场景布局等信息，并且支持媒体处理以及与用户的交互。这些综合媒体内容是通过综合场景来组织与管理的。

用综合场景表达综合媒体环境，首先需要考虑两个问题，一是数据的组织表示，也就是综合场景的表达方法；二是场景的解析管理与呈现，也就是综合场景的呈现技术。

综合场景的表达需要包含各种元数据信息（例如关于音视频资源的信息、图形图片信息、文本信息等等）以及呈现信息（比如场景表示所采用的描述语言）。在数据信息的组织管理上本文采用了 MPEG-21 的数字内容项（Digital Item）相关标准来对这些数据进行表示，呈现信息则采用 MPEG 轻量级应用场景表示标准 LAsER（Lightweight Application Scene Representation），通过将二者结合，本文设计并实现了 MPEG 结构化信息呈现（Presentation of Structured Information）的标准，提出了综合场景的表达方法。相关提案被 MPEG 标准组接收，该标准将以 MPEG-21 Digital Item 标准补篇的形式发布。

在此基础上，本文同时设计并实现了一个综合场景引擎，可以进行综合场景解析以及对场景中多个对象和用户交互的管理以及呈现，实现了综合场景的呈现技术。在综合场景引擎中，本文提出了并应用了基于对象的时序模型来处理场景中各元素的时间管理，据此所写的论文也被 PCM 2008 会议所接收。相关的理论在北京奥运 WIM TV 试验系统中也得到了有效的验证。

本文将综合场景与视频分析引擎相结合，提出了一个增量场景模型来处理非预期的场景内容，解决了处理动态场景的技术问题，实现了综合场景的一种高层次应用。在本文提出的系统设计中，首先利用视频分析引擎分析来自各种流媒体服务系统的视频数据，从中解析出一些语义线索（例如新闻中的字幕、人脸信息或者关键镜头的切换等等），这些语义线索信息以增量场景方式传送给用户终端。在用户终端上，视音频与动态增量信息以综合场景的方式进行集成，即增量信息

以叠加的方式添加到用户终端初始场景之中，从而支持用户与视频的动态交互，此即本文设计的增量场景处理模型。利用这一模型，可以支持用户在视频中点击其感兴趣的内容，并将这些用户互动信息反馈给服务端，服务端对海量用户点击数据进行分析处理，生成一些记录用户敏感区域的虚拟场景，从而支持用户在一种内容丰富的可交互环境中消费数字媒体内容。这就是本文所提出的基于用户标注的综合场景系统。

在标准研究和模型设计的基础上，本文进行了软件系统开发和可行性验证。

关键词： 综合媒体，综合场景，结构化信息处理，轻量级应用场景表示，增量场景。