

Architecture Design of an MPEG-4 System

Yi-Shin Tung¹, and Ja-Ling Wu^{1,2}, Senior Member IEEE, Ho Chia-Chiang¹
Communication and Multimedia Lab., Department of Computer Science and Information
Engineering, National Taiwan University, Taiwan¹, Cyberlink Inc., Taipei, Taiwan²

Abstract

MPEG-4, the newest generation of MPEG compression standards is the first object-oriented coding standard in the world. It integrates scene and object descriptions into the bitstream to provide abilities of object reusability. From technology point of view, MPEG-4 includes coding objects of natural and synthesis content types so that wider spectrum of applications can be developed. With respect to media transmission, it communicates with networks via standardized interfaces, and therefore can adapt to varieties of network protocols. To design an architecture for a generic standard that covers broad areas, the designers should have knowledge-base in various fields, so that the solution can be efficient and complete. In this paper, we use Microsoft multimedia architecture, Directshow, to devise a total solution for MPEG-4 systems. Three major components (MPEG-4 server, player control and realtime encoder) are addressed; moreover, the modules required in each component and the corresponding functionality and mechanism are also discussed. The proposed architecture provides a useful framework for MPEG-4 software and hardware designers to follow.*

Summary

This paper prescribes the five layers and four interfaces (Fig.1) defined in MPEG-4 systems [1-5]. From application to network, the five layers are composite layer, compression layer, sync layer, deliver layer and transmux layer. Each layer does its own work and cooperatively completes the full MPEG-4 system functions. The four in-between interfaces behave in a standard way to communicate with successive layers. The functionality of each component will be described in detail. The design and implementation of the two most important components, MPEG-4 server and MPEG-4 player control, are the major concern of this paper. The MPEG-4 Server (Fig.2) should be embedded in operating systems. Based on "push" technology of streaming applications, different considerations should be taken. There are five modules in the server: MP4 file parser gets media data and time related information from MPEG-4 storage file archive; Streaming generating unit produces sync-layer packetized elementary streams (SPS's) according to time constraints; Flexmux unit combines different SPS's with the same destination and QoS parameter into one Flexmux stream;

* DirectShow is a trademark of Microsoft Corporation.

Network monitor supervises the session, report network delay, loss rate, error rate, etc., to applications or streaming unit; Clock synchronization unit periodically communicates with cooperating server to guarantee the head-end synchronization [8]. MPEG-4 player control (Fig.3) will be treated as a necessary component for any application that claims to be MPEG-4 related. The player control is realized based on Microsoft DirectShow [7] (the industry de facto standard in PC multimedia), and extended to provide more interactive abilities for MPEG-4 operations. From software viewpoints, the kernel of DirectShow is a modularized pluggable system, based on the usage of the so-called filters. The most significant advantage of DirectShow comes from its ability to make multimedia application design more clear and easy. By carefully dividing the work into connected filters in the DirectShow architecture, each filter can be implemented by different program developers. Another advantage of DirectShow is the filter re-usage, which fastens the developing of new multimedia applications. In our design, there are three types of filter modules for MPEG-4 system filtergraph: DMIF (Delivery Multimedia Integration Framework) source filter gets SPS via DAI (DMIF Application Interface) from MPEG-4 Server. All kinds of handler filters take responsible for collecting elementary elements, controlling the synchronization of decoding process and adapting to user environment to drop less important data. Decoding filter comprises the decoding bank such that the player control can handle all bitstreams supported by MPEG-4. Finally, object compositor and audio mixer filters generate the expected effect according to the user defined scene structure. Figs. 4 represents the system decoding subgraph of the MPEG-4 system.

Conclusions and Future Work

We have realized a prototype of MPEG-4 system and proved a practical framework. By the design, any individual MPEG-4 developer can integrate its work into the proposed architecture (Figs. 2 and 4). Since the first version of MPEG-4 standard is just finalized and the second version of it is still under developing, some new features need to be integrated into the proposed architecture for supporting full MPEG-4 system functions. The new features include IPMP (Intellectual Property Management and Protection), body animation, 3D-mesh, and complex scene solutions. The future work of us will focus on taking these new features into account, implementing the refined system, and devising the

realtime encoding components.

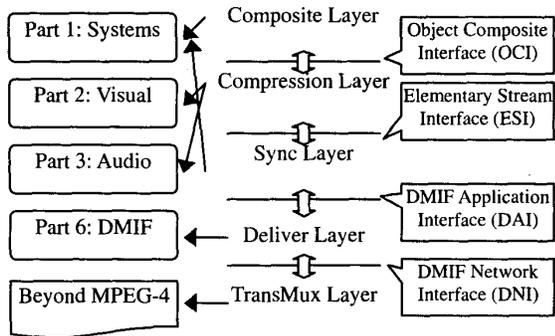


Figure 1: The Five Layers, Four In-between Interfaces and Their Relationships Defined in MPEG-4.

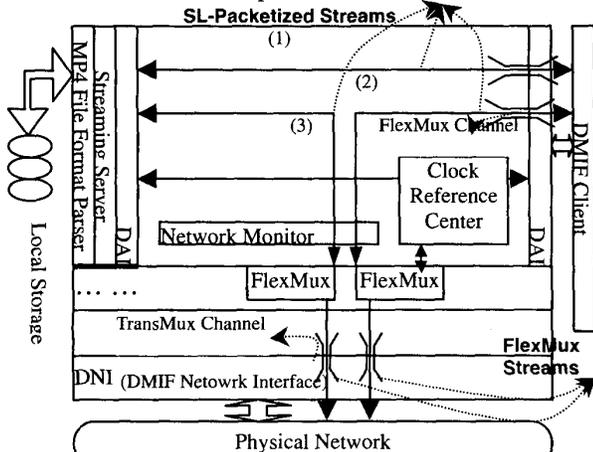


Figure 2: The Proposed Architecture of MPEG-4 Server (also known as DMIF/Streaming server): MPEG-4 Server should be treated as a service provided for any PC consuming MPEG-4 contents. (1) DMIF for local files read/write, storage-based applications. (2) DMIF for remote service read/write, storage-based, and real-time applications. (3) DMIF providing file read/write services of remote access, storage-based applications.

Reference

- [1] ISO/IEC IS 11172, Information Technology, International Organization for Standardization, 1991.
- [2] ISO/IEC IS 13818, Information Technology – Generic Coding of Moving Pictures and Associated Audio Information, International Organization for Standardization, 1994.
- [3] CCITT SG XV, Recommendation H.261, International Telecommunication Union, August 1990.
- [4] ITU-T IS, Recommendation H.263 – Video Coding for low bit-rate communication, International Telecommunication Union, November 1995.
- [5] ISO/IEC FDIS 14496, Information Technology – Generic Coding of Audio-Visual Objects – Part 1: System, Part 2: Visual, Part 3: Audio, Part 6: DMIF, International Organization for Standardization, 1998.

- [6] Video For Window, Microsoft Corporation
- [7] DirectX Media: Multimedia Services for Microsoft Internet Explorer and Microsoft Windows, Microsoft Corporation, October 1998.
- [8] PC Video Synchronization and Playback, Microsoft Corporation, November 1998.

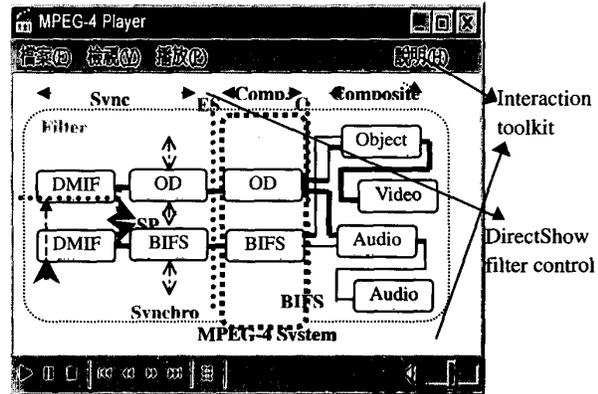


Figure 3: The Player Control Unit: The player control includes DirectShow filter control and some interaction tools. It can be integrated into any application that consumes MPEG-4 encoded contents.

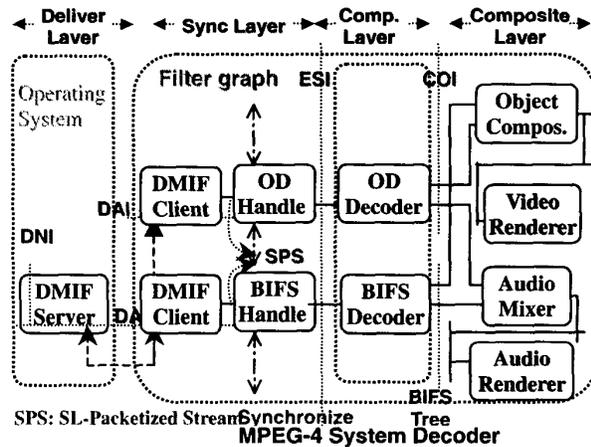


Figure 4: The major Filtergraph of MPEG-4 System: This figure represents the system decoding graph. DMIF client waits for incoming packets and feeds them to the corresponding handler. Synchronization and data collection are processed between different handlers. OD decoder and BIFS decoder interpret the object information and scene structure and then transmit the results to object compositor and audio mixer.