Architecture Design, System Implementation, and Applications of MPEG-4 Systems

Yi-Shin Tung, and Ja-Ling Wu, Senior Member IEEE

Communication and Multimedia Laboratory
Department of Computer Science and Information Engineering
National Taiwan University, Taiwan

ABSTRACT

MPEG-4, the newest generation of multimedia codec, is gifted with the ability of scene-based presentation, natural and synthetic object coding, powerful script and java-language support, and network transparent integration. Combined with these attractive features, MPEG-4 has the greatest potential to be applied to a large amount of applications. In this paper, we design an MPEG-4 compliant architecture 1) using pipeline-design to modularize decoding processes, 2) adopting a centralized control mechanism which enables interaction and synchronization, and 3) building a unique transparent-transmission unit for delivering data among the heterogeneous networks. We have implemented a prototype of streaming server, player component and deliver agent over the IP network to prove the workability of the designed architecture. In our opinion, whether MPEG-4 is successful will heavily depend on the appearance of killer applications. Specifically, good applications may lead MPEG-4 to become popularly accepted around the world. Therefore, we enumerate several interesting applications (News Report, Virtual Art Gallery, Personal Karaoke Room, and Live Sports Report) and make simple contents on our system for demonstration.

1. INTRODUCTION

The state-of-the-art technology, MPEG-4 [4], is an elaborate product of coding-field experts around the world. It has provided a lot of attractive features in its finalized versions 1 and 2 [2,3], and is believed to provide more in later versions. Each of those features can benefit certain applications. The most attractive features of MPEG-4 include:

(1) Scene-Based and Object-Oriented Representation: This is the most important feature in MPEG-4. It changes the representative style and adds more interactive ability.

(2) New Powerful Codec Support and Existing Contents Integration: New MPEG-4 compression standards cover natural and synthetic coding techniques, and many requirements like scalability and error resilience are also taken into account. These make it both intelligent and powerful. In addition, MPEG-4 also integrates existed compression standards, such as JPEG, MPEG-1 Video/Audio, MPEG-2 Video/Audio, etc, into it such that existing contents can be adapted to it easily.

(3) Script and Language Support for high Interaction: By java-script in scene streams or java-program in system streams, a mechanism can be set up to achieve server-side and client-side interaction.

(4) Transparent Network Interface: This feature enables the separation of data transfer task from applications.

(5) Streaming File Format Support: This feature makes streaming process independent of stored contents and employed codecs.

After aware of the above features, a well-devised architecture is required to be implementation reference. In this paper, we first address the architecture we proposed for MPEG-4 system in Section 2. With our design, most of the prescribed MPEG-4 features can be realized on it. Then, the proposed architecture is implemented over the IP network. Section 3 presents the implementation details and some related descriptions. To prove our design workable and to demonstrate the power of MPEG-4, we exemplify some practical applications based on our implementation prototype in Section 4. Finally, we conclude this write-up.

2. AN ARCHITECTURE DESIGNED FOR THE MPEG-4 SYSTEM

Based on our previous work [1], we refine the designed layout of MPEG-4 systems to fulfill aforementioned features and requirements [5,7] under a simple client-server environment. Fig.1 represents the three components in the design, which communicate among each other via DAI and DNI interfaces [1].

The first component is DMIF Server (c.f. Fig.2). It is uniquely sited on one machine. By building linked library (such as the DMIF/IP library described in Section 3) on using underlying protocols, the DMIF server connects to different network protocols. Thus, it may provide delivering services via DAI for streaming server and client applications.

The 'Player' is a client component (c.f. Fig.3), which consumes the incoming contents and displays the resulting scenes. This component provides two ways for user interaction. One is a programmers' interface provided in its library, and the other is an end-users' toolkit on the UI (the interactive control API and tools in Fig.2). Inside the player, the decoding process and controlling process are two major sub-components. The decoding process takes responsible for processing the incoming data, while the controlling process builds decoding graph and supervises its behavior by graph control API. The compositor renders the resulting display according to messages forwarded to it from the frame layout module via scene or object based API. To synchronize streams from different servers, the decoding process also maintains a clock reference table for the handlers to calculate the correct timestamp of each atom. Scene graph and object directory stored scene structure and properties of objects in the scene, respectively. Stream manager gives control commands and sends upstreams via Net API.

Streaming server (c.f. Fig.4) includes three important elements. The request handler processes the client requests. If channel-relative requests are made, the handler will modify the request table, accordingly. All other user-commands are treated as events to the script-processing unit. The 'Scheduler' accesses stored contents in the MP4 file format, and sends requested data according to the policy of instruction rules given by the server designer. Script processing unit, works like a VM, read script instructions and processes what the script dictates.

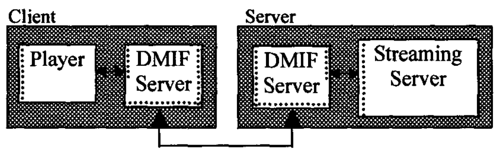

Figure 1: Components in the Proposed Architecture: There are three components in the proposed client-server environment. They are: (a) Streaming Server, (b) DMIF Server, and (c) Player.

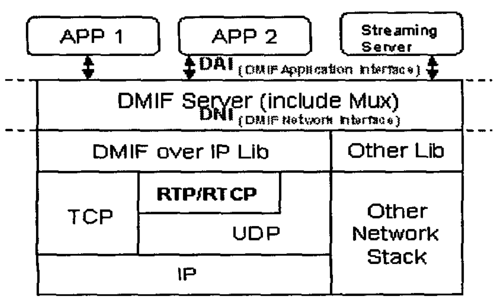

Figure 2: DMIF Server and Lapped IP-related Library: We implemented DMIF/IP library to transfer data using underlying RTP/TCP/UDP protocols.

3. SYSTEM IMPLEMENTATION

According to the architecture, we implement the prototype systems on PC over IP networks. In the streaming process, different data are of different properties. Loss of application request and system data will cause severe system failure, while loss of fixed textures that are alive during entire presentation period will degrade the representation a lot. So, TCP is adopted to prevent degradation. Media data like audio and video have real-time constraint but less priority. In this case, RTP can meet the requirement. Clock reference stream is sent casually by UDP connection. DMIF Server communicates with its clients by inter-process communication (IPC) techniques. A DMIF over IP library and multiplexing are also built in our system.

By our design, the player component can be fused to user applications seamless while retaining interactive ability. Using pipeline design helps us to integrate available codecs easily. Consequently, all MPEG-1/2/4 video/audio clips and JPEG images can be played in our systems. The client script is extended to support server-site interaction.

We use simple intuitive rules to instruct streaming process, and the server script is stored on user-defined track in the format of MP4 [3]. In the process of content creation, the editor needs to produce a set of corresponding scripts for server and client sites.

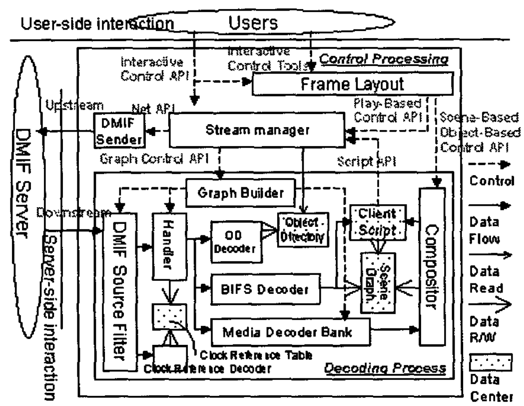

Figure 3: The Layout of Player Component: Modularized components and control mechanisms are shown here, several set of APIs are clearly defined to separate the development processes.

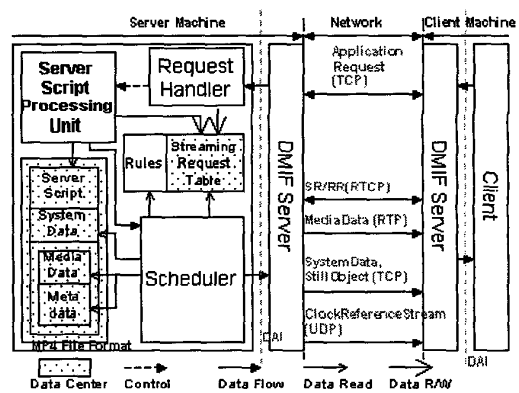

Figure 4: The Layout of Streaming Server: Three major components are devised: 1) request handler, 2) scheduler, and 3) script processing unit.

4. MPEG-4 APPLICATIONS

Powerful tools are included within MPEG-4 in order to enable various applications [6]. In current progress, MPEG-4 has integrated advanced technology and combined efficient implementation for compression, transmission and display;

however, the most important thing for its success is the existence of killer applications.

If we overview MPEG-4, what is the major impact it can bring to us? In brief, MPEG-4 is a general framework to show scenes composed of all kinds of multimedia contents/objects. There are two types of scene structures supported in its scene description: 2D co-ordinate system/content and 3D co-ordinate one. Each may result in a lot of applications. We try to squeeze some appreciable ones from the pool, and consider possible connections with other technologies for further applications in the near future. In the following sections, we exemplify four examples and create simple contents to be used in our prototype system. Figs. 5~8 show some segments of our demonstration. The first one uses 2D systems and may be applied to the TV market, while the others are 3D systems. With the above-mentioned 3D systems, advanced devices such as head-mounted helm display, transparent-display glasses or globular projection may bring users to the VR (virtual-reality) world.

Figure 5: News Report and Video On-Demand for TV Applications: In this example, a pocket controller appeared on the left-down corner controls the layout of TV according to users' preference. Four lienable informative objects are presented: a list of headline news for user selection (left), information about selected headline (down), virtual reporter using text-to-speech technology (right-down), and the movable stock information to deliver update data (right-up). Users may customize the layout style by loading appreciative scene-stream from the server or from local storage.

4.1 News Report / Video On-Demand

Providing VoD services on next generation TV is demanded by the current market. MPEG-4 seems to be the key technology to make it possible. Supporting script language will make MPEG-4 become more powerful to achieve higher user interaction. In the system stream (scene stream), it can include java-script to enhance the ability to change scene structure, objects and even their corresponding attributes. The MPEG-4 engine may be the main processing unit in the settop box for future interactive TV. With MPEG-4, the content provider may give the indices of hot news for user selections, or offer additional instant financial information simultaneously. On the other side, users may customize the layout or its presentation style. The layout may change dynamically by triggering the sensors provided by VRML. To separate the control and display, multiple scenes supported in MPEG-4 will enable control elements to be centralized on the display of the remote control device such as the LED on a PDA. By integrated MPEG-4 facial coding and text-to-speech technology, we can simulate a virtual reporter when main display switches to prerecorded tapes. All these features will enrich TV representation.

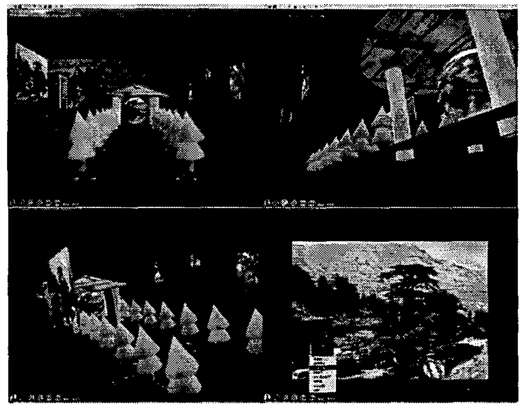

Figure 6: Virtual Art Gallery: By connecting to database, the visitor may see the interested paintings sited everyplace simultaneously. Different scenes may deliver various feelings to users. Progressive image coding tries to exhibit high-quality arts into your eyes. Besides, by VR technology the upcoming gallery tour may lead you into a fantastic world.

4.2 Virtual Art Gallery

Have you ever thought of seeing a painting in the sky or fine art in the wood? It seems impossible in the real world. Usually, we need to spend a whole day in the art museum and see the valuable paintings in a silent room. Due to the rare and expensive properties of exhibition items, we must visit different museums to see all Monet collections, and can only view his paintings from behind a handrail so as not to damage them. By using digitalization technology, all paintings located worldwide may be accessed via a network such as the Internet. MPEG-4 wavelet coding helps multi-resolution display in an efficient way, while scene-based representation makes the virtual art gallery come true. Besides, combined with the digital library, the server can filter users' selections according to their interest or preference. By integrating these technologies, virtual art gallery will be created and change certain aspects of our daily life.

4.3 Personal Karaoke Room / Singing Show

Singing pop songs in KTV rooms is a popular form of entertainment in Taiwan. Making people feel free and comfortable may be the reason why it has won wide acceptance.

By using MPEG-4, we can construct well-decorated virtual rooms in a California style, Hawaii style, etc. A group of singers may hold their private concert in their favorite room and invite their friends to join them equipped with virtual reality devices. By taking advantage of interactivity and scripting ability, the program may connect to a music database and intelligently list available songs for user selections. Thus, the users can choose the latest or even their favorite songs, while the service provider can charge the copyright fee when requests are made.

Figure 7: Karaoke and Personal Concert Shows: Multiple users may hold a private show of singing in a synthetic room of their favorite style. The interactive tool enables the simulation of walking or singing around the room. The ability of scripting makes it possible to retrieve music movies they want. In this example, the controller is also equipped for selection of the playing movies (c.f. the right-down corner).

4.4 Live Sports Report / Sports Game

To show or report a sports game in an intelligible way is one important consideration to improve representation. In general, due to quick movement and heavy body deformation of sportsmen, it is much harder to model their behavior synthetically in live sports. At the other extreme, sports computer games use simple human models to simulate the behavior of sportsmen with a robot-like movement. The technology gap between these two still exists. However, the application gives the direction what technology we are seeking for.

5. CONCLUSIONS

The MPEG group that constitutes the worldwide leaders in the field of data compression, has taken seven years to finalize the new standard, MPEG-4. The elaborate MPEG-4 possesses attractive properties for most of the applications. However, such a general standard will, by nature, unfortunately lose its focus. Therefore, proper and well-advertised applications are the keys to making MPEG-4 successful. In this paper, some contributions are made: 1) we design the architecture to fully meet the requirements of the MPEG-4 standard and the modularized design adopted in our work will ease the burden of development, 2) a prototype implementation is realized according to the proposed architecture that proves our design workable, and 3) we also show some possible applications based on our implementation. We build a simple tool to make exhibitive contents that are in MPEG-4 file format. The demonstrations show the great power of MPEG-4 and how it can benefit from using the most advanced technology currently available.

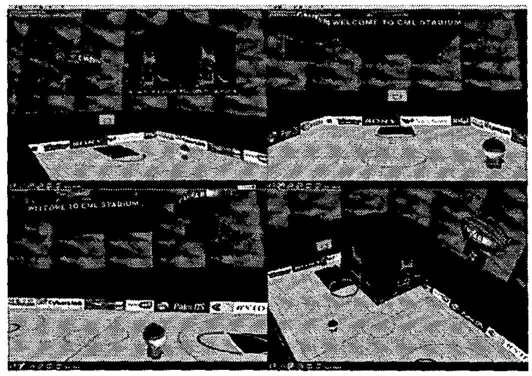

Figure 8: Sports Report and Sports Game: In live sports report, multiple views of live video may display on the top panel, and synthetic version can be produced by sensors equipped with people and balls. Oppositely, in sports games, a model-based march is celebrated virtually, and simulating sports video can be generated and display on the top panel.

6. ACKNOWLEDGEMENT

The modules assembled in our systems have been created by many programmers. The author would like to thank Communication and Multimedia Lab. members, Kuo-Luen Perng, Meng-Jyi Shieh, Deng-Rung Liu, Wei-Sho Huang, Li-Shien Chen, Chi-Hui Huang, and Wei-Teh Wang for their active involvement in system implementation and content creation.

7. REFERENCES

[1] Y.S. Tung, J.L. Wu, and C.C. Ho, "Architecture Design of an MPEG-4 System", ICCE 2000 Digest of Technical Papers.

[2] ISO/IEC IS 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.

[3] ISO/IEC 14496 AM I, Part 1: Systems, Part2: Visual, Part 3: Audio, Part 6: DMIF, International Organization for Standardization, 2000.

[4] MPEG Requirements Group, "MPEG-4 Overview", Document ISO/IEC/JTC1/SC29/WG11/N3536, Beijing MPEG meeting, July 2000.

[5] MPEG Requirements Group, "MPEG-4 Requirements Document", Document ISO/IEC JTC1/SC29/WG11/N3534, Beijing MPEG meeting, July 2000.

[6] MPEG Requirements Group, "MPEG-4 Applications Document", Document ISO/IEC JTC1/SC29/WG11/N2724, Seoul MPEG meeting, March 1999.

[7] MPEG Systems Group, "Multi-users technology: Requirements and Applications", Document ISO/IEC JTC1/SC29/WG11/N3386, Geneva MPEG meeting, June 2000.