MERL – A MITSUBISHI ELECTRIC RESEARCH LABORATORY
http://www.merl.com

Resource Adaptation Based on MPEG-21 Usage Environment Descriptions

Huifang Sun Anthony Vetro Kohtaro Asai

TR-2003-58 05 2003

Abstract

This paper addresses several aspects related to the distribution of content. The first aim is to provide an overview of the Universal Multimedia Access (UMA) concept. The primary function of UMA services is to provide the best Quality of Service (QoS) or User experience by either selecting/adapting the content format to meet the playback environment, or adapting the content playback environment to accommodate the content. The second purpose of this paper is to describe how the concept of UMA relates to the emerging MPEG standard, Digital Item Adaptation (DIA), which will become Part 7 of the MPEG-21 standard. An update on the standards activity in this area will be presented. Finally, the third purpose of this paper is to describe the impact that DIA will have on transcoding strategies, where we describe topics that we are currently working on and analyze some areas of future research.

Proc. IEEE Int'l Conference on Circuits and Systems, Bangkok, Thailand, May 2003.

RESOURCE ADAPTATION BASED ON MPEG-21 USAGE ENVIRONMENT DESCRIPTIONS

Huifang Sun[*], *Anthony Vetro*[*] *and Kohtaro Asai*[**]

[*]Mitsubishi Electric Research Laboratories, Murray Hill, NJ, USA
[**]Mitsubishi Electric Corporation, Kamakura, Kanagawa, Japan

ABSTRACT

This paper addresses several aspects related to the distribution of content. The first aim is to provide an overview of the Universal Multimedia Access (UMA) concept. The primary function of UMA services is to provide the best Quality of Service (QoS) or User experience by either selecting/adapting the content format to meet the playback environment, or adapting the content playback environment to accommodate the content. The second purpose of this paper is to describe how the concept of UMA relates to the emerging MPEG standard, Digital Item Adaptation (DIA), which will become Part 7 of the MPEG-21 standard. An update on the standards activity in this area will be presented. Finally, the third purpose of this paper is to describe the impact that DIA will have on transcoding strategies, where we describe topics that we are currently working on and analyze some areas of future research.

1. INTRODUCTION

During the past two decades, tremendous progress has been achieved in the areas of digital signal processing and communications. In the area of digital signal processing, the techniques for digital video compression, transmission, and storage have progressed at an astounding pace. For example, considering the video-coding standards developed by Moving Picture Expert Group (MPEG) [1] and International Telecommunication Union (ITU) [2], many successful applications from digital television and to streaming video can be realized. Of course, the major goal of these video coding standards is to greatly reduce the data amount by compression algorithms. However, one very important aspect of the video coding scheme is to standardize the binary data format. In this way, the compressed bitstream can be used by a wide variety of equipment including traditional hardware, such as television set-top boxes, as well as other multimedia devices, such as computers, personal digital assistants (PDA's) and mobile terminals.

With advances in the semiconductor industry, the processing speed of digital signal processors is rapidly increasing; also, the price of storage memory is becoming inexpensive for many applications. This allows many devices, not just TVs and PCs, but also many portable devices such as PDA's or mobile terminals, to handle the compressed video data under certain conditions. Similarly, in the communications area, Internet technologies, network technologies and mobile communications have also seen tremendous progress recently. These trends have become the major driving force for UMA.

The concept of UMA is to enable access to any multimedia content over any type of network, such as Internet, Wireless LAN or others, from any type of terminals with varying capabilities such as mobile phones, personal computers, and television sets [3]. The primary function of UMA services is to provide the best QoS or User experience by either selecting appropriate content formats, or adapting the content format directly, to meet the playback environment, or to adapt the content playback environment to accommodate the content.

Towards to the above goal, Part 7 of MPEG-21, referred to as Digital Item Adaptation (DIA) is defining a set of descriptors and tools to enable transparent and augmented use of multimedia resources across a wide range of networks and devices [4]. In the context of MPEG-21, a Digital Item is defined as a structured digital object with a standard representation, identification and description. This entity is also the fundamental unit of distribution and transaction within the MPEG-21 framework [5]. Although DIA will not specify the adaptation engines themselves, there are a variety of interesting factors to consider with respect to the resource adaptation engine given complete knowledge of a users environment.

The rest of this paper is organized as follows. In Section 2, we first give an overview of UMA. In Section 3, the objectives of DIA will be described and the current status of this part of the standard will be given. Then, we describe the resource adaptation engine for a DIA system in Section 4. Due to space limitations, we can only provide an overview of the areas we are working and do not describe the techniques in detail. Finally, concluding remarks will be given in Section 5.

2. OVERVIEW OF UMA

The history of information technology shows that during the recent two decades a tremendous number of new technologies have been developed to meet the growing needs for obtaining any information from anywhere and anytime. This is the major motivation to propose the concept and carry out projects for UMA by many researchers. The concept of UMA is illustrated in Figure 1.

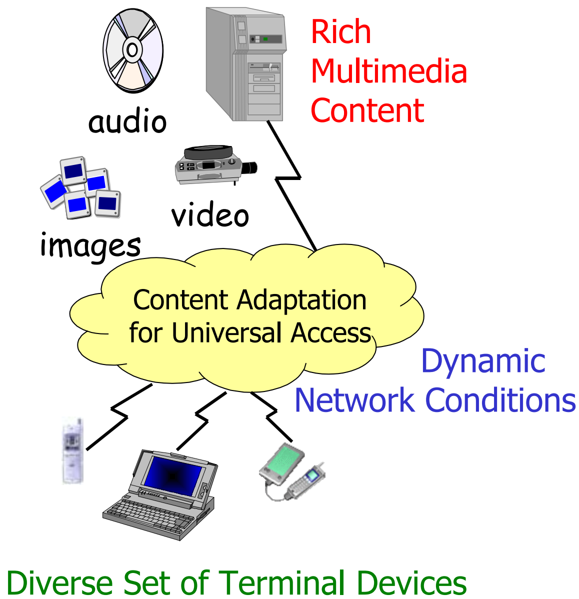

Figure 1. Concept of UMA.

The concept of the UMA has two aspects. From the user side, UMA allows users access to a rich set of multimedia content through various connections such as Internet, Optical Ethernet, DSL, Wireless LAN, Cable, Satellite, terrestrial broadcasting and others, with different terminal devices. From the content or service provider side, UMA promises to deliver timely multimedia contents with various formats to a wide range receivers that have different capabilities and are connected through various access networks.

Now, let us discuss what are today's driving forces for addressing the problems of UMA. There are several facts, which provide the evidence for the growing need of UMA. From the viewpoint of contents, one fact is that contents are available everywhere. This situation is created by the recent revolution of content format representation. The digitization of content format provides the possibility of letting different devices access the content. The advances of compression algorithms, especially digital video compression standards, greatly reduce the amount of data. This provides strong tools to reduce the difficulty for content-delivery. Another fact is the capability growth of the communication networks. The Internet provides a very convenient and powerful tool for content transmission, but its big limitation is the narrow bandwidth at present. However, with advances in wireless communication, video transmission becomes possible through the 3G and beyond 3G networks. In summary, mature content representation formats, improving communication technology and an increasing capability of terminals, support the possibility for UMA.

Given the above, what are the problems that have to be addressed today to achieve UMA? The major problem for UMA is to fix the mismatch between the content formats, the conditions of transmission networks, and the capability of receiving terminals. A mechanism for adaptation has to be created for this purpose and can be considered into the following two ways. One way is to adapt the content to fit the playback environment and other is to adapt playback environment to accommodate the existing input contents. In either of these scenarios, it should be noted that MPEG-7 plays a key role in providing a description of the content [6].

Currently, multimedia content is encoded with various schemes including MPEG-1, MPEG-2, MPEG-4, Wavelets, JPEG, AAC, AC-3, etc.; all these encoding formats at various spatial resolutions, frame rates and bit rates. The communication networks also have different characteristics such as bandwidth, bit error rate, latency, and packet loss rate depending on the network infrastructure and load. Likewise, the receiving terminals have different content playback capabilities and different user preferences that affect the type of content that can be played on the terminals. The mismatch between the rich multimedia content and the content playback environment is the primary barrier for the fulfillment of the UMA promise. The adaptation engine is the entity that bridges this mismatch by either adapting the content to fit to the content playback environment or adapting the content playback environment to accommodate the content. For content adaptation, the nature of content determines the operations involved in the actual adaptation. For example, if the mobile terminal with MPEG-4 decoding capability wants to receive DTV signals which is encoded with MPEG-2, the adaptation needs to perform the conversion for several things: syntax conversion from MPEG2 to MPEG-4, spatial resolution conversion from SDTV or HDTV to QSIF or even lower resolution, bit rate conversion to reduce the bit rate for wireless network and other necessary conversion. Adapting the content playback environment involves acquiring additional resources to handle the content. The resources acquired can be session bandwidth, computing resources at the sending and receiving terminals, decoders in the receiver terminals, or improving the network latency and packet loss. The playback environment can change dynamically and content adaptation should match the changing environment to deliver content at the best quality possible.

3. MPEG-21 DIA

As we discussed in the previous section, the key problem of UMA is to fix the mismatch between rich multimedia contents, network and terminals. To address this problem, the mechanism of adaptation is one of the most important issues for UMA. The current situation is that there exist a lot of standards for content representations such as audio/video coding standards developed by MPEG and ITU; also, a lot of standards for communications and a lot of protocols for networks. All elements for building an infrastructure of UMA exist and have been standardized or their standards are under development. However, there is no "big picture" how these elements relate to each other. In other words, the way to fix the mismatch between elements has not been defined. As we described previously, there are exist a large number of receiver terminals; the networks or medias have very different natures; also, the multimedia contents are represented with different formats.

To accomplish the promise of the UMA, we need to develop new standards these meet these needs. Now, the emerging standard MPEG-21, especially Part 7 of MPEG-21, Digital Item Adaptation, aims at fixing these gaps between elements. The

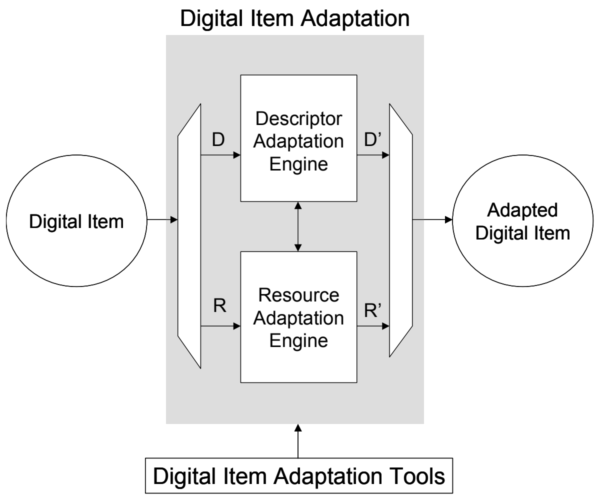

Figure 2. Illustration of MPEG-21 Digital Item Adaptation.

conceptual architecture of Digital Item Adaptation is illustrated in Figure 2.

From this architecture it can be found that the adapted Digital Item that meets the content playback environment is obtained from the original input Digital Item through either descriptor adaptation engine or resource adaptation engine. DIA aims at providing the standardized descriptions and tools that can be used by these adaptation engines [4].

Four major areas of *usage environment descriptions* have been considered until now. All of these descriptions may be used to support future UMA services. These areas include user characteristics, terminal capabilities, network characteristics and natural environment characteristics. User characteristics include five aspects: user type, content preferences, presentation preferences, presentation preferences, accessibility characteristics and mobility characteristics. The descriptions in the terminal capabilities currently include encoding and decoding capabilities, display and audio output capabilities, as well as power, storage and input-output characteristics. The items in the network descriptions are network capabilities and network conditions. The descriptions in the natural environment characteristics are currently defined by location, time and audio-visual environment characteristics.

In addition to the usage environment descriptions, MPEG-21 DIA is also targeting the specification of tools that support format-independent manipulations, tools to help make trade-offs between coding parameters and content characteristics, tools that support metadata adaptation, and tools that enable session mobility.

4. RESOURCE ADAPTATION ENGINE

Resource adaptation includes a variety of conversions or transformations, including conversion between video coding formats, such as from MPEG-2 to MPEG-4, and coding parameter conversion, such as bit rate, spatial resolution and frame rate. This section deals primarily with video transcoding, and covers the basic design goals that motivate research on video transcoding, some recent advances in this area, as well as new challenges related to the video transcoding process.

4.1 Design Goals & Issues

The key design goals of transcoding include two aspects: 1) to maintain the video quality during the transcoding process, and 2) to keep complexity as low as possible. The most straightforward solution is to simply decode the video signal, perform any post-processing if necessary, and re-encode the modified signal according to any new constraints. Although the quality is good, the complexity of this approach is relatively high. To avoid such a conversion, some researchers have propose scalable coding schemes that can be easily scaled down according to the requirements of terminal or network. MPEG-4 Fine Granular Scalability is one of such scalable coding scheme [?]. However, in the entertainment industry, many contents exist with a fixed single-layer representation format. For example, the contents of Digital Television and DVD are encoded with an MPEG-2 format. In order for receivers, such as mobile terminals, to receive this signal, we must convert the bitstream from MPEG-2 to MPEG-4. Therefore, there are instances in which transcoding is certainly needed; the application scenarios are very clear. Such conversion must be done with low complexity and high quality.

In most of video coding standards, the temporal redundancy is removed by predictive coding techniques, where the current field/frame is predicted with the previously coded field/frame. The predictive difference is then quantized and coded with variable run-length codes. During transcoding, the reference used for prediction is typically transcoded to a lower quality or spatial resolution; therefore, the error in this reference frame will accumulate. This accumulation of error is referred to as drift. Minimizing the amount of drift incurred during the transcoding process, while keeping complexity low, is a major goal of the transcoding design.

4.2 Transcoding Background

Reducing the drift with a low-complexity architecture is a significant challenge for transcoding and a great deal of research effort has been dedicated to this task. In the context of transcoding to a lower bit-rate, several papers have described architectures that avoid full decoding and re-encoding, e.g., [7] and techniques that rely on the on the efficient reuse of macroblock layer information, e.g., [8]. More recently, several architectures and techniques for spatial resolution reduction have been presented in [9]. Most notable may be the use of intra-block refresh to combat the effects of drift errors incurred by transcoding to a lower spatial resolution. Also, new work in the area of temporal resolution reduction has been presented [11]. For all the above methods, motion vector refinement can also be applied for improved quality without a significant increase in complexity [12].

4.3 Transcoding QoS

Existing transcoding techniques provide a powerful tool for the resource adaptation engine, where changes in bit-rate, frame-rate and spatial resolution can easily be controlled. However, controlling these parameters to achieve the best perceptual quality for any user, or group of users, is a challenging problem. In this subsection, we introduce relevant work in this area and discuss possibilities for further work.

The concept of a utility function to measure the users satisfaction of a coded video bitstream was introduced in [12]. Based on features extracted from the video, machine learning and classification techniques were used to estimate the subjective quality of the video coded according to different scaling profiles, e.g. drop B-frames, scale DCT coefficients, drop color components. Based on the quality estimate, some decision regarding the best way to scale the content could be made.

In [13], a video coding algorithm that considered the trade-off between spatial and temporal quality was presented. This algorithm was based on analytical models that estimated actual MSE and rate for a set of possible frameskip factors and quantization parameters. A similar problem was considered in [14], where the authors sought to optimize the coding across bit-rate, frame-rate and spatial resolution. In this work, various reconstruction patterns in the receiver were considered. Since the emphasis was on an optimal framework and ways to solve the multi-dimensional problem, actual rates and distortions were used rather than estimates obtained from a model.

More recently, rate-distortion models to estimate the rate and distortion resulting from specific transcoding operations, such as re-quantization, spatial scaling and temporal scaling, have been developed [15]. These models allow one to consider the choices between various scaling operations directly from data contained within the compressed bitstreams.

The above works indicate that we can optimize the quality of a single transcoded output with some confidence. Given constraints on the bit rate and spatial resolution specified by the capability of a terminal and the current network condition, the transcoder can be designed to automatically skip more frames in the low motion area, thereby using more bits for coded frames, to guarantee the spatial quality. One the other hand, in the fast motion area, more frames will be coded and less bits will be assigned to each of coded frame.

Considering the work on utility functions, this kind of control can also be applied according to user preferences instead of objective quality measurements. For example, for surveillance video, the users may not care about overall quality of the transcoded output, but be very interested in a certain time period of video or requires to have a minimum level of frame quality to view certain details in the scene.

Finally, one emerging research area could be on the transcoding of multiple streams. In other words, given a single processing engine that is required to transcode several bitstreams, one may consider algorithms to allocate the quality among those streams. Also, algorithms to allocate the processing resources as well may be considered. The optimization of quality and processing complexity may also be considered in a joint manner.

5. CONCLUDING REMARKS

In this paper, an overview of UMA has been presented. The purpose of UMA is to deliver or to use rich multimedia contents with a range of network and receiver capabilities. The emerging MPEG-21 standard, Digital Item Adaptation, which aims to solve the mismatch between components in the end-to-end delivery chain, has also been reviewed. Additionally, we have described the current status of usage environment descriptions being considered by the standard. Finally, in this paper, current areas that we are working on and the key technical challenges for transcoding have been discussed. Furthermore, an important research direction, Transcoding QoS, has been proposed.

6. REFERENCES

[1] ISO/IEC 14496-2:2001, "Coding of Audio-Visual Objects – Part 2: Visual," 2nd Edition, 2001.

[2] ITU-T Recommendation H.263++, "Video Coding for Low-Bit-Rate Communication," ITU-T Standardization Sector (Geneva), 3, 2000.

[3] R. Mohan, J.R. Smith, and C.S. Li, "Adapting Multimedia Content for Universal Access," *IEEE Trans. Multimedia*, vol. 1, no. 1, pp. 104-114, March 1999.

[4] "MPEG-21 Digital Item Adaptation Working Draft v2.0," ISO/MPEG N4944, Klagenfurt, Austria, July 2002.

[5] "MPEG-21 Overview v.4," ISO/MPEG N4801, Fairfax, USA, May 2002.

[6] "MPEG-7 Overview v.8," ISO/MPEG N4980, Klagenfurt, Austria, July 2002.

[7] P. Assuncao and M. Ghanbari. "Post-processing of MPEG-2 coded video for transmission at lower bit-rates," *Proc. IEEE Int'l Conf. Acoustics, Speech and Signal Processing*, vol. 4, Atlanta, GA, May 1996.

[8] H. Sun, W. Kwok, and J. Zdepski, "Architectures for MPEG compressed bitstream scaling," *IEEE Trans. Circuits Syst. Video Technol.*, April 1996.

[9] P. Yin, A. Vetro, B. Liu, and H. Sun, "Drift compensation for reduced resolution transcoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 11, pp. 1009-1020, Nov.2002.

[10] K.T. Fung, Y.L. Chan and W.C. Siu, "New architecture for dynamic frame-skipping transcoder," *IEEE Trans. Image Processing*, vol. 11, no. 8, pp. 886-900, Aug. 2002.

[11] J. Youn, M.T. Sun, and C.W. Lin, "Motion vector refinement for high performance transcoding," *IEEE Trans. Multimedia*, vol. 1, no. 1, pp. 30-40, March 1999.

[12] R. Liao, P. Bocheck, A. Campbell, and S.F. Chang, "Content-aware Network Adaptation for MPEG-4", *Proc. NOSSDAV*, New Jersey, USA, June 1999.

[13] A. Vetro, Y. Wang and H. Sun, "Rate-distortion optimized video coding considering frameskip," *Proc. IEEE Int'l Conf. Image Processing*, Thessaloniki, Greece, Oct. 2001.

[14] E.C. Reed and J.S. Lim, "Optimal multidimensional bit-rate control for video communications," *IEEE Trans Image Processing*, vol. 11, no. 8, pp. 873-885, Aug. 2002.

[15] P. Yin, A. Vetro and B. Liu, "Rate-Distortion Models for Video Transcoding," *Proc. SPIE Conf. Image and Video Commun. Processing*, San Jose, CA, Jan 2003.