Metadata Broadcasting for Personalized Service: A Practical Solution

Kyeongok Kang, Jae-Gon Kim, Heekyung Lee, Hyun Sung Chang, Seung-Jun Yang, Young-tae Kim, Han-kyu Lee, and Jinwoong Kim

As the number of broadcasting channels and programs increases rapidly, the importance of personalized service has been emphasized. In this paper, we propose a practical framework of metadata broadcasting to provide personalized service according to user preferences and various terminal/network conditions. First, we present an overall system architecture of a metadata broadcasting system and then propose several core technologies (particularly in the parts of metadata authoring, metadata encoding, and metadata-based personalized content consumption). For interoperability, the proposed solution is designed to be compliant with the relevant standards of the TV-Anytime Forum, MPEG-7 Systems, and MPEG-2 Systems. Considering a home network environment, we also propose a metadata-based content adaptation scheme. Each component technology has been implemented individually, integrated into an end-to-end prototype broadcasting system, and successfully tested with a set of personalized service scenarios that are also developed in this study.

Keywords: Metadata, personalized broadcasting service, electronic content guide, TV-Anytime, MPEG-7.

Manuscript received Apr. 9, 2003; revised May 31, 2004.

Kyeongok Kang (phone: +82 42 860 5521, email: kokang@etri.re.kr), Jae-Gon Kim (email: jgkim@etri.re.kr), Heekyung Lee (email: lhk95@etri.re.kr), Seung-Jun Yang (email: sjyang@etri.re.kr), Young-tae Kim (email: kytae@etri.re.kr), Han-kyu Lee (email: hkl@etri.re.kr), and Jinwoong Kim (email: jwkim@etri.re.kr) are with Digital Broadcasting Research Division, ETRI, Daejeon, Korea.

Hyun Sung Chang (email: hschang@ieee.org) was with ETRI.

I. Introduction

The advent of digital broadcasting services has increased the number of broadcasting channels explosively, especially via cable and satellite media. According to this change, a user can have a chance to view his or her preferred content among a large number of programs, and a broadcaster can provide more relevant programs to a user group of a specific interest in a customized way. In order to provide an environment for the user to consume customized broadcasting programs or content with his or her preferred methods at anytime, independent of broadcasting time, a set-top box (STB) with storage media, called a personal digital recorder (PDR), is needed.

On the other hand, the rapid increase of information according to the increase of broadcasting channels makes it difficult for a user to find and select what he or she wants, and it demands effective methods for the selection and management of content broadcasted, distributed, or locally stored. Therefore, it is necessary to develop technologies for more efficiently accessing and browsing stored content on a PDR, such as an agent-based recommendation of content, a selective filtering and recording of content based on user preferences, and a nonlinear navigation of content using metadata. To resolve the above issues, international standards and industrial consortiums such as MPEG-7 [1] or the TV-Anytime Forum [2]-[6] have developed generic or domain-specific metadata standards and their requirements for content description, respectively.

Metadata means ‘data about data’. In broadcasting environments, it means additional data to describe multimedia content, including audiovisual (AV) features for content-based retrieval as well as electronic program guide (EPG)

information for selecting a channel and its program. These metadata are used for searching, selecting, recording, and managing segments as well as programs in broadcasting environments. Therefore, application technologies using metadata play a very important role in providing intelligent and personalized broadcasting services in the digital era.

In emerging multimedia applications based on a new paradigm in which broadcasting and communication are converged together, broadcasting content will be delivered over heterogeneous networks (e.g., broadcasting channels, the Internet, wireless communication channels, wireless local area networks, home networks, etc.) to a multitude of devices having different capabilities (e.g., display size, computation power, etc.) and user preferences. This kind of integrated multimedia framework is often referred to as universal multimedia access, in which environment, multimedia content adaptation to heterogeneous resource conditions of terminals, networks and/or user preferences is another important technology to provide personalized broadcasting services based on metadata.

In this paper, technology-developing trends and related standardization activities are introduced in section II. Core technologies based on metadata and the implementation of a prototype system using them are described in sections III and IV, respectively. Finally, a conclusion is given in section V.

II. Related Works

1. Standardization

A. MPEG-7

With the increasing use of the Internet and digital broadcasting services, digitalized multimedia contents are overwhelming. It is becoming very important to efficiently search, transform, and deliver multimedia content in the way users want to. In this context, a multimedia expertise group of ISO/IEC, popularly known as MPEG, has been conducting the activity of specifying a multimedia description standard called MPEG-7. Until now, the first edition of the MPEG-7 standard has been released and the standardization for the second edition will be finalized soon.

MPEG-7 defines a set of descriptors and description schemes (DSs) for multimedia content, spanning also a system aspect (e.g., compression, decoder behavior, etc.), description definition language, and conformance testing [1]. The purpose of MPEG-7 is to provide a general framework for the multimedia description rather than be targeted to broadcasting environments.

B. TV-Anytime Forum

The TV-Anytime Forum is an association of organizations which seeks to develop specifications to enable audio-visual and other services based on mass-market high volume digital storage in consumer platforms such as an STB.

Since its formation in 1999, the TV-Anytime Forum has developed specifications for applications based on local persistent storage, independent of networks and various delivery mechanisms such as Advanced Television Systems Committee (ATSC), Digital Video Broadcasting (DVB) and the Internet. The TV-Anytime Forum also aims to define interoperable and integrated system structures with security equipment to protect the interests of all parties such as content creators, broadcasting service providers, STB manufacturers, and telecommunications companies.

As a result of continuous working with the above objectives, at the end of 2002, the TV-Anytime Forum announced its first specification series for Phase 1 [3]-[6]. The TV-Anytime Phase 1 specifications enable one to search, select, acquire, and rightfully use content on local and/or remote storage both from broadcast and online sources. In the year 2003, the TV-Anytime Phase 1 specifications have been adopted as European Telecommunications Standards Institute (ETSI) standards [7].

Figure 1 illustrates an example of dynamic behaviors of a TV-Anytime system [3], which consists of the following functional blocks: content creator (CC), content service provider (CSP), search and navigation (SN), location resolution (LR), local storage management (SM), and user interaction (UI).

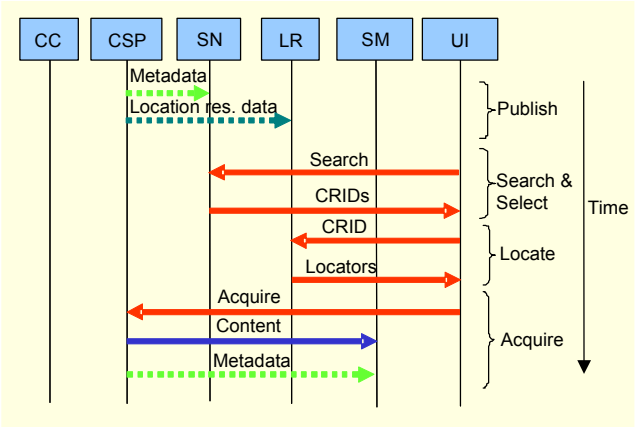


Fig. 1. TV-Anytime content flow model.

TV-Anytime metadata are instantiated as XML documents under a root element called "TVAMain." Each part of the metadata instances is associated together with the same AV program by the content referencing identification (CRID) as shown in Fig. 2 [4]. Content referencing is the process of associating a token to a piece of content that represents its location where the content can be acquired. The key concept in content referencing is the separation of the reference to a

content item, the CRID, from the information needed to actually retrieve the content item, the locator. From a system perspective, content referencing and resolution lies between search and selection, and actually acquiring the content. From the content referencing perspective, search and selection yields a CRID, which is resolved into either a number of CRIDs or a number of locators (the number may be one) by the process of location resolution.

Metadata may be distributed across many TV-Anytime documents, but it is always possible to relate appropriate pieces through CRIDs. In a TV-Anytime concept, a program is an editorially coherent piece of content and typically acquired by a PDR as a whole, and one or more programs are grouped together into a group. Programs can belong to groups, and groups can belong to other groups. This relationship is reflected in the metadata, again by linking program descriptions with group descriptions using CRIDs. TV-Anytime defines several types of program groups such as "series" and "program compilation".

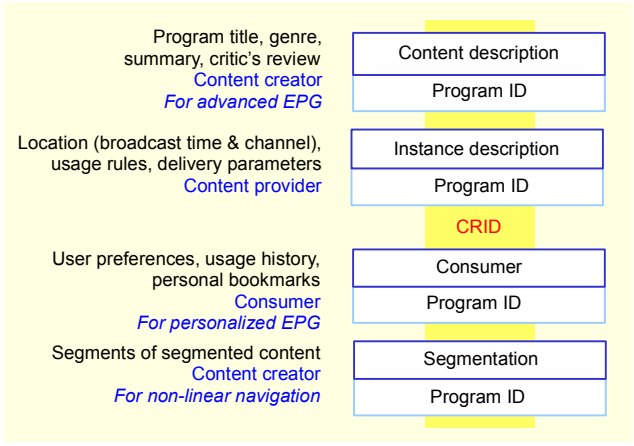


Fig. 2. TV-Anytime metadata and CRID.

The basic kinds of TV-Anytime metadata are content description metadata, instance description metadata, segmentation metadata and consumer metadata. Content description metadata are general information about a piece of content that does not change regardless of how the content is published or broadcast. Instance description metadata describe a particular instance of a piece of content, including information such as the content location, usage rules, and delivery parameters (e.g., video format). Consumer metadata, borrowed from MPEG-7, include usage history data (logging data), annotation metadata, and user preferences for a personalized content service. Segmentation metadata describe a segment or groups of segments. A segment is a continuous portion of a piece of content. For example the part of a news program describing a particular news topic can be defined as a segment.

In addition to Phase 1, the TV-Anytime Forum has defined the requirements and business models for Phase 2 [8], which mainly deals with the sharing and distribution of rich content among local storage devices and/or network digital recorders in home network environments. The definition of the requirements and business models of Phase 2 was finished at the end of 2003. Now, the forum has been working on the normative specifications of Phase 2 metadata including the following issues: packaging, targeting, synchronization, remote programming, etc. [9].

2. Technology Development

The *myTV* project, one of the Information Society Technologies (IST) programs in Europe, had a target to provide personalized service for digital TV among broadcasters and consumer electronics [10]. To provide personalized program services on the platform of users' STBs under the convergence of digital broadcasting and broadband communications, it had a final goal to develop, standardize, and evaluate technologies for accessing broadcasting content at anytime users want, using their favorite methods independently of the scheduled time of the programs. In other words, the goal of this project is to provide a DVB-compliant system with TV-Anytime functionalities.

By the end of 2001, at the completion of the *myTV* project, a new project, *Share it!*, was launched [11]. Compared to the fact that the *myTV* project contributed to make the TV-Anytime specifications in Phase 1 (local STB environments), the *Share-it!* project has a goal to develop, standardize, and evaluate technologies needed for the TV-Anytime specifications in Phase 2 (home-to-home network or virtual family network environments).

In the US, standardization bodies such as ATSC, Society of Motion Picture and Television Engineers (SMPTE), Motion Picture Association (MPA), and universities have driven the development of metadata-related technologies for analysis, classification, description, and retrieval of multimedia content. Service providers such as TiVo and ReplayTV have provided EPG services based on their ad-hoc solutions for their personal recording devices. In March 2001, the T3/S8 group of ATSC developed requirements for advanced EPGs to provide search and navigate functionalities at the segment level: they made a draft revision of the Program and System Information Protocol to reflect the above aspects based on the TV-Anytime specifications in 2002 [12].

A standardization entity in Japan, Association of Radio Industries and Business (ARIB), has developed technologies to provide interactive enhanced TV services and TV-Anytime

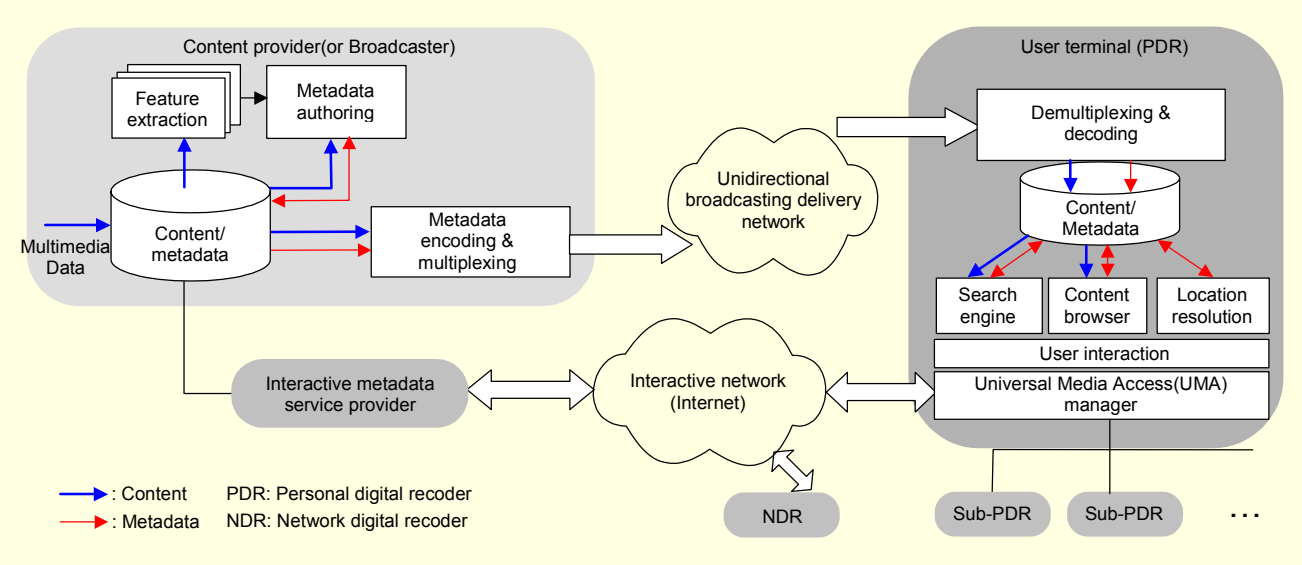


Fig. 3. The overall configuration of a metadata broadcasting system on a conceptual level.

services based on the TV-Anytime specifications. Some members of ARIB have developed technologies to provide Integrated Service TV (ISTV) services to give multimedia information as well as digital broadcasting.

III. Core Technologies

In this section, we present a practical framework of a metadata broadcasting system for personalized TV services. Figure 3 illustrates the overall configuration of the metadata broadcasting system on a conceptual level. As shown in Fig. 3, it has a server-client structure. Metadata authored at the server side are coded as a binary format, multiplexed with AV content (programs) into an MPEG-2 transport stream (TS), and then delivered to the client via a unidirectional broadcasting channel. The client terminal (i.e., a PDR) is equipped with various kinds of functional modules for accessing, searching and browsing content, and enables users to consume the content in a personalized way. In a home network environment, multiple devices (i.e., sub-PDR's like mobile devices) can also be connected with the main PDR. Then, content may need to be adapted, and delivered to diverse types of sub-PDR's according to their capabilities.

A bi-directional return channel (i.e., the Internet) provides on-line metadata services, in which a user could acquire additional metadata related to programs received from a broadcasting channel. Furthermore, it allows targeted services according to user content consumption behaviors and/or preferences by collecting user-centric information through the return channel.

In the development of the metadata broadcasting system based on this configuration, the following metadata related technologies should be integrated into an end-to end system in a unified way:

- Metadata authoring (generation and editing)
- Metadata delivery (coding, encapsulation, multiplexing)
- Metadata processing (parsing, user preference extraction, etc.) and metadata-based content accessing/browsing
- Metadata-based content adaptation.

In the subsequent sub-sections, we present some novel core technologies in the area of the above technologies.

1. Metadata Authoring

In this sub-section, we present a novel scheme for authoring content descriptive metadata compliant to the TV-Anytime metadata [4] in a "What You See Is What You Get" environment. The proposed scheme includes a variety of element technologies (visualization and interactive editing of metadata, metadata fragmentation and compression, media access via a timeline and key-frames) and their integration into a standalone tool.

A. Segment Metadata Generation

Automated extraction of a segment metadata (e.g., shot boundary detection, video highlight detection, etc.) has been an active research issue for the past decade [13], [14]. In this sub-section, we briefly present a shot boundary detection method used for the generation of a segment metadata in the proposed scheme.

We use color and edge features for the shot boundary detection and perform all the operations in a compressed domain without an inverse discrete cosine transform. Denoting

the color dissimilarity between two images F_i and F_j by $d_c(F_i, F_j)$ and their edge dissimilarity by $d_e(F_i, F_j)$, the overall distance metric between two consecutive frames can be computed as

$$d(F_k, F_{k+1}) = w_c d_c(F_k, F_{k+1}) + w_e d_e(F_k, F_{k+1}),$$

where w_c and w_e are weighting factors for color and edge features, respectively.

For color features, we use a 256-bin histogram in a Cb-Cr color space that is extracted from DC frames, and an L_2 metric for $d_c(\cdot)$. Meanwhile, for edge features, we compare the block edges extracted from two successive frames by

$$d_e(F_k, F_{k+1}) = \max(B_{in}, B_{out}),$$

where B_{in} and B_{out} denote the number of entering edge blocks and the number of exiting edge blocks, respectively.

In [15], we showed how block edges can be extracted from a discrete cosine transform (DCT) coefficient domain. The fast algorithm proposed in [15] is based on the examination of the quantitative degree with which each DCT coefficient contributes to block edge patterns.

B. Metadata Authoring Tool

Although some kinds of metadata are extractable, as discussed in section III.1.A, there are still the needs of human intervention for refining results (especially, in terms of subjective criteria) and manually filling in the fields such as keyword and synopsis that are hard to extract in automated ways. In this sub-section, we propose an architecture of a metadata authoring tool. The purpose of the metadata authoring tool is to provide users (i.e., metadata scriptwriters) a visual and intuitive environment for authoring content descriptive metadata.

The overall structure of the proposed metadata editing tool is illustrated in Fig. 4. It is largely comprised of four internal components, each for metadata visualization, metadata updating, content browsing, and I/O interface, respectively.

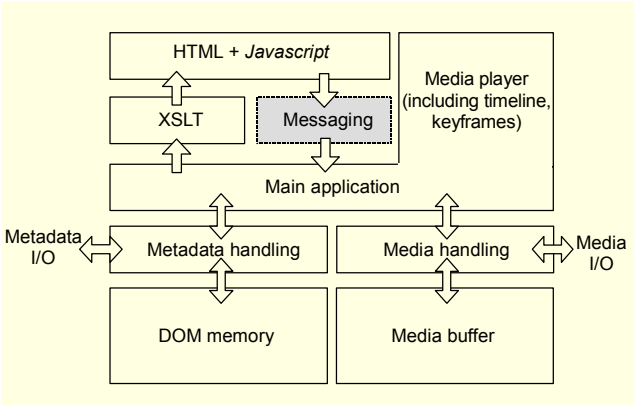


Fig. 4. The overall structure of the proposed metadata authoring tool.

1) Metadata visualization

Three types of visualization are possible. The first one is a pure text, the most naive form of visualization, where XML tags and names are treated the same as content. The second is a tree-style visualization for the hierarchical structure of metadata or a spreadsheet style presentation for the content of a selected element that may be advantageous for browsing a metadata structure and/or searching for particular elements. The last one is a template-based visualization which provides the most intuitive editing condition. In this scheme, each metadata element is transformed into a part of an HTML document according to the template defined in a style sheet document. If we make use of an extensible style sheet language transformation, which is a generic web application tool for transforming XML documents [16], and define proper templates for metadata elements, this form of visualization can be quite straightforward. Common to all three types of visualization, metadata elements related to particular locations of media data (e.g., highlights, preview description, etc.) are mapped to a time-line and used for playing and browsing segments.

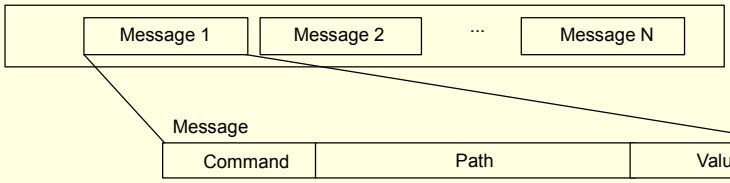
2) Metadata updating (editing)

In order to support metadata editing in the stylized HTML sheets, the transformed document should have interactivity: It should be editable and contain information of reverse paths from each visualized component to the original metadata. The interactivity can be achieved by embedding commands in a script language into the proper parts of the HTML document. Based on the scripts, the visualized documents can be browsed in a self-contained manner or communicate with the main application to request operations regarding media control or metadata updating. For the latter case, if a user event occurs, the document sends messages, whose formats are depicted in Fig. 5, to the main application to request appropriate operations. In our scheme, we define six kinds of messages according to their purposes. Most of the messages are used to update metadata and others to request media specific processing (e.g., image capture, media playing, etc.).

As a message arrives from the visualized document, the main application parses the message and calls the appropriate modules to act accordingly. For example, the metadata updating module updates the metadata loaded in memory. Once the metadata are updated, the windows displaying the visualized document are also updated to display the updated information consistently.

3) Content browsing

The interactive HTML documents can work in a self-



Command	Path	Value	Meaning
replace	*<P: XPath>*	*<V1: simple value>*	Change the value of the node referenced by *<P>* as *<V1>*
add	*<P: XPath>*	*<V2: metadata fragment>*	Add the metadata fragment *<V2>* as a child of the node referenced by *<P>*
delete	*<P: XPath>*	--	Delete the node referenced by *<P>* and its subtree.
update-image	*<P: XPath>*	--	Capture the currently displayed video frame, and set the value of the node referenced by *<P>* as its time code.
play	--	*<V3: sequence of time codes>*	Play the video segment whose time codes are concatenated in *<V3>* in a sequence
apply-xsl	--	*<V4: XSL name>*	Apply *<V4>* to visualize the metadata

Fig. 5. Message formats delivered from editing sheet to main application.

contained manner. For example, a table-of-contents browser can be browsed without interrupting the main application. In cases where some actual operations for media data (e.g., image capture, key frame extraction, segment playing, timeline display and update) are needed, the document sends an appropriate message such as "play" or "update-image."

4) I/O interface

Generated or edited metadata may be stored as a textual file. Metadata can be compressed for the efficient use of bandwidth and multiplexed with media data for delivery over a broadcasting channel, as will be explained in detail in the next sub-section.

2. Metadata Coding and Delivery

This sub-section deals with metadata delivery including metadata encoding, encapsulation, and multiplexing. In particular, the architecture of a developed binary format for multimedia descriptions (BiM) codec and multiplexing scheme is presented. We basically follow the standard technologies for the delivery of metadata: MPEG-7 Systems for metadata encoding (BiM) [17]; the MPEG-2 Digital Storage Media Command and Control (DSM-CC) for metadata encapsulation as a data carousel [18], and MPEG-2 Systems AMD 1 for metadata multiplexing into a transport stream [19].

A. Metadata Coding: BiM Codec

For bandwidth efficiency, metadata that is an XML document to be delivered are encoded as a binary format. In addition, in order to allow a more flexible and efficient manipulation of the updating, metadata are decomposed into self-consistent units of data, called fragments when they come to be binarized.

MPEG-7 Systems [17] defines two kinds of packet structures: the first is the access unit (AU), an atomic unit of metadata to which a time stamp can be attached, and the second is the DecoderInit (DI) for initializing the configuration parameters required for the decoding of subsequent AUs.

According to MPEG-7 Systems, metadata are basically structured into a sequence of AUs. As illustrated in Fig. 6, an AU is composed of one or more fragment update units (FUUs), each of which represents a small part of a metadata description. Therefore, the AU may convey updates for several distinct parts of the description simultaneously. Each FUU consists of a fragment update command (FU Command), a fragment update context (FU Context), and a fragment update payload (FU Payload) [17], [20], [21].

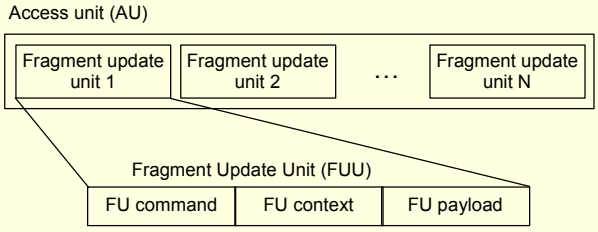


Fig. 6. Structure of the MPEG-7 access unit.

Figure 7 illustrates a functional block diagram of a developed BiM encoder. At first, the encoder generates DI initializing configuration parameters required for the decoding of a

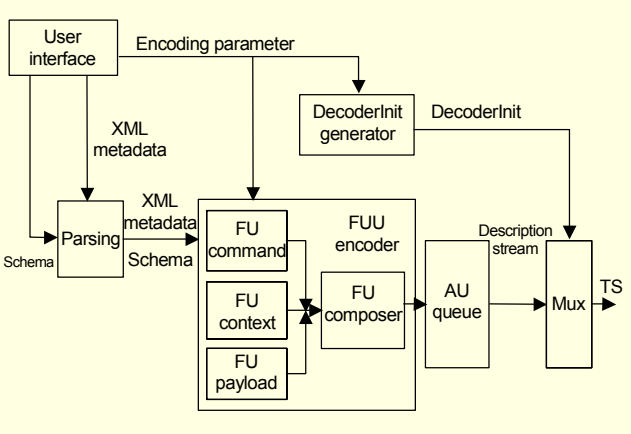


Fig. 7. Functional block diagram of a BiM encoder.

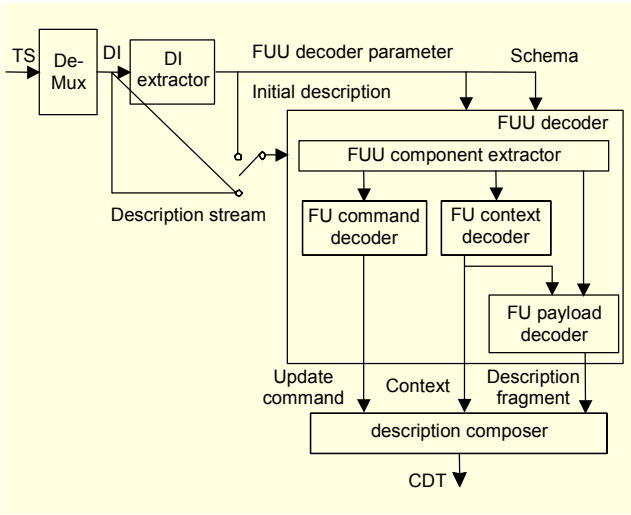


Fig. 8. Functional block diagram of a BiM decoder.

sequence of AUs such as the identifiers of an instantiated schema, an initial description. In an FUU encoder, each module generates a respective FUU component referring to metadata and user parameters: The FU command code word specifies a command to be executed on a description tree to be maintained at a BiM decoder; the FU context specifies the node of the description tree on which the FU Command should be executed; the FU payload performs the task of compressing the structural information of the description and encoding each data type. Finally, the FUU components are composed into an FU Composer and passed to an AU queue.

As illustrated in Fig. 8, the BiM decoder receives the DI and metadata stream and reconstructs the delivered metadata. First, a FUU decoder is initialized by the initial parameters and schemas coming from the DI. After initialization by the initial description, a description tree is updated by subsequent AUs from the description stream. Then, each FUU component is extracted from a given AU in the FUU decoder. An FU command decoder refers to a simple look-up table to decode the update command and passes it to a description composer. Decoded FU context information is passed both to the description composer and to an FU Payload decoder. Aided by the FU context information, the FU Payload decoder decodes an FU Payload to yield a description fragment. The FU Payload is composed of a flag defining certain decoding modes and a payload content which can be either an element or a simple value. In particular, a complex type with complex content is decoded by a finite state automation decoder.

B. Metadata Encapsulation and Multiplexing

To deliver metadata over a broadcasting channel (here, terrestrial ATSC is assumed), we first encapsulate an encoded description stream as a data carousel and carry it over an MPEG-2 TS [22] by multiplexing with the main AV content according to MPEG-2 Systems AMD 1 [19].

Information on the version and metadata format is included during encapsulation. For efficient multiplexing, we consider metadata as an elementary stream, so we carry it in a separate transport stream at the first step. In the TS carrying metadata only, the "stream_type" to be specified in the program map table for signaling a receiver that the MPEG-2 TS contains metadata has a value in the range of "0×16" to "0×20" [19]. Several descriptors, which are used to inform the receiver of the metadata location and the association between metadata and the main AV content, and to configure the metadata decoder, such as "content labeling descriptor," "metadata pointer descriptor," "metadata descriptor," and "metadata system target decoder descriptor," are also included in the program map table.

Finally, two individual transport streams carrying the main AV content and metadata are multiplexed into a single transport stream to be transmitted in an injector, as shown in Fig. 9. In the injector, the null TS packets of the TS containing the main AV content are replaced with TS packets of the metadata stream, in which the location of the injection in a final stream and the amount of metadata injected can be controlled by specifying the injection information. For instance, asynchronous metadata without the value of a unique decoding

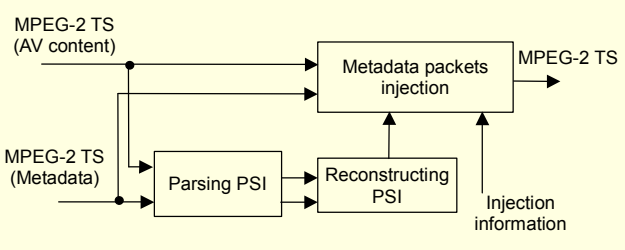


Fig. 9. Injection-based multiplexing of metadata into an MPEG-2 TS.

time can be transmitted repeatedly by controlling the injection. In addition, Program Specific Information is regenerated to contain all information about both AV content and metadata.

3. Metadata-Based Personalized Consumption

Metadata take an important role for personalized TV applications. In order to provide personalized services based on metadata, a feasible architecture of a client terminal and proper user interaction methods need to be developed. We have developed a metadata processing engine based on the TV-Anytime system model [3] emulating an STB and built several personalized broadcasting service scenarios using the implemented metadata processing engine, in particular, *personal channel* and *personal program*. Personalized services mean content consumption in a customized way according to user conditions and/or user preferences (e.g., user preferences on program genre, terminal capabilities, etc.).

In this section, we present a high-level architecture of the metadata-processing engine and the basic features of personalized services implemented based on the designed architecture.

A. Architecture of the Metadata Processing Engine

Figure 10 illustrates the overall architecture of functional modules on a conceptual level, which is based on the TV-Anytime system model [3]. In Fig. 10, a content service provider (i.e., broadcasting service provider) transmits content and associated metadata to a client. SN and LR modules access metadata and content location information, respectively, and return certain parts of them in response to a user interaction module. Note that the SN and the LR modules can also be located on the outside of the client terminal, which means that external service providers who serve metadata or content location information may be connected to the client terminal with interactive channels such as the Internet. In addition, the terminal needs functional modules for content storage, content presentation, and user interaction.

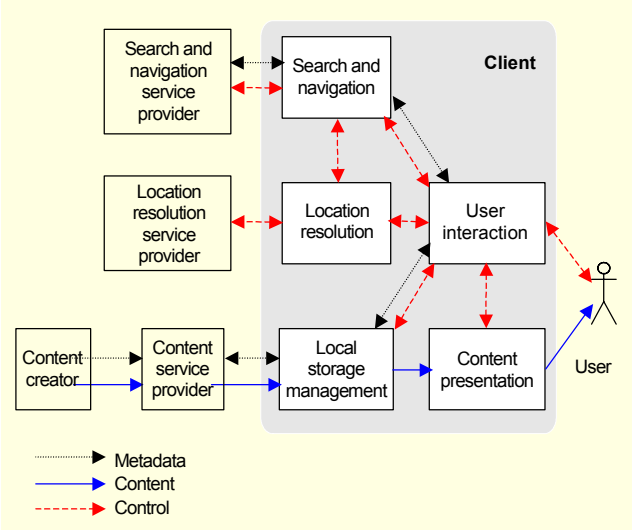


Fig. 10. A conceptual block diagram for an STB.

B. Basic Features of the Metadata Processing Engine

To provide an STB with advanced features exploiting metadata, effective use cases and interaction schemes for content access and consumption should be developed. Figure 11 shows a set of screen shots of the implemented functions: an advanced content guide (ACG) that is an advanced EPG employing an automatic program recommendation based on user preferences, table-of-content (ToC) based browsing, and an event-based summary.

When a large number of channels are available, ACG is one of the most basic functions since it efficiently lets a user know where his or her wanted program is located in a broadcasting schedule. Figure 11(a) shows the ACG function implemented on the metadata-processing engine. It shows a weekly timetable in which preferred programs can be represented by coloring each time slot according to the degree of preference on program genre. In addition, the programs available in the same time slot are listed at the lower half of the screen in the order of user's genre preference.

A ToC-based browser provides non-linear navigation of selected program content. From the ToC, users could figure out the overall story-structure of the underlying content and have access to the content in a segment level randomly. As shown in

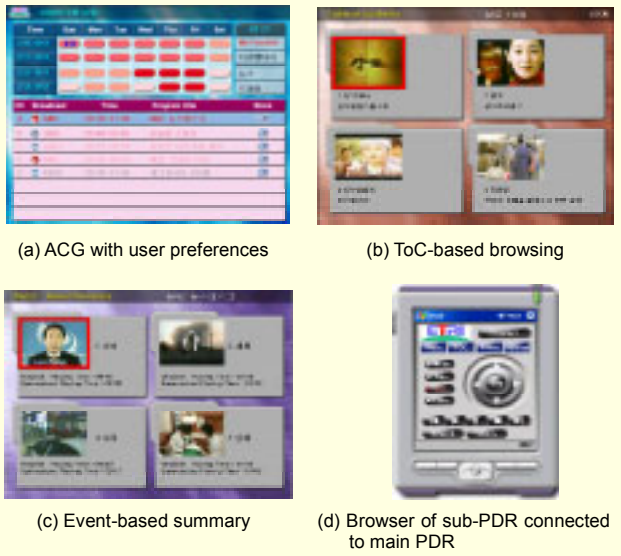


Fig. 11. Representative screen shots for personalized consumption of content.

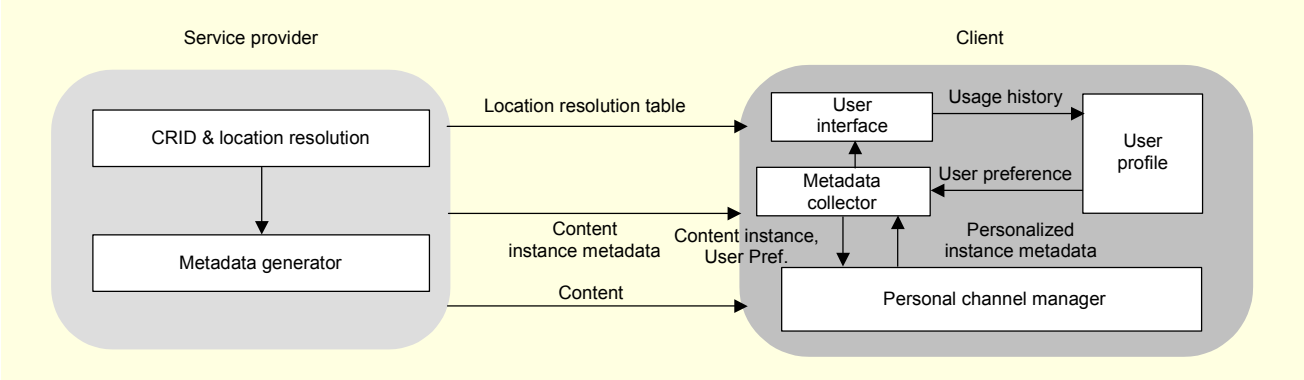


Fig. 12. System interaction for personal channel service.

Fig. 11(b), users can select and watch a specific segment after browsing the hierarchical structure of the content.

The event-based summary in Fig. 11(c) provides a function similar to the index found at the end of a book. In particular, a news or sports video, where the event is well defined (a news theme, sports event such as a scored goal, etc.), is quite suited for browsing with the event-based summary. In the validation of this feature, we selected a news program and built up an event-based summary in which each segment is classified into political, international, social, and economic items according to the theme.

In addition, we have developed some functions for home network environments where multiple PDRs exist and content delivery among them is assumed. In the near future, user environments for consuming broadcasting content would become constructed with multiple PDRs rather than a single PDR. As a first step, we integrated a remote control function on a sub-PDR (similar to a PDA), as shown in Fig. 11(d). After the sub-PDR is connected to the main PDR, a user may log into the main PDR with his or her own information, receive the ACG information based on his or her preference, and control all the functions of the main-PDR using the sub-PDR.

C. Personal Channel

Personal channel service is a kind of program rescheduling according to various user preferences in order to support a more efficient selection of relevant programs, which happens dynamically in a PDR after the ACG information has been collected from a number of service providers. The rescheduled programs can be provided via a personal channel at the preferred time.

Figure 12 shows the overall operations between a service provider and a PDR for the personal channel service. The service provider offers content referencing information (CRID and location resolution [5]) along with content description metadata and instance description metadata which are specified in the TV-Anytime standard [4].

All the received metadata are stored and managed in a metadata collector of the PDR. Then, an ACG is generated and rendered via a user interface using the collected metadata. A user can browse and select a program from the ACG and watch the program at the scheduled time. Usage history of interactive user actions such as browsing, playing, recording, etc., can be tracked and kept as parts of a user profile and subsequently used for extracting user preferences on time, genre, program title, and so on. The extracted user preference metadata are also stored in the metadata collector and referred to by a personal channel manager to build a virtual channel comprising only the preferred programs rescheduled to the desired time. The tailored personal channel is rendered to the ACG screen of Fig. 11(a) along with other (actual) channels. In this way, an automatic recommendation of preferred programs is provided as a virtual channel (called personal channel here) based on the user-centric metadata of the usage history and user preference, each of which are described in an interoperable manner using the UsageHistory DS and the UserPreference DS, respectively, defined in the TV-Anytime borrowed from the MPEG-7 Multimedia Description Scheme [1].

D. Personal Program

In home network environments where a multitude of user devices (e.g., PDA, MP3 player, DVD player, etc.) are connected by networks, as mentioned in section III.3.B, users may still want to access and consume broadcasting content without much effort at anytime and anywhere. In order to support this scenario, we have developed a notion of a personal program, which is a content service customized to user preferences and/or a consuming environment.

Figure 13 shows the proposed system architecture for a personal program service. The operation proceeds as follows: A content service provider provides content and metadata to a personal program service provider. Then, a personal program analyzer in the personal program service provider determines

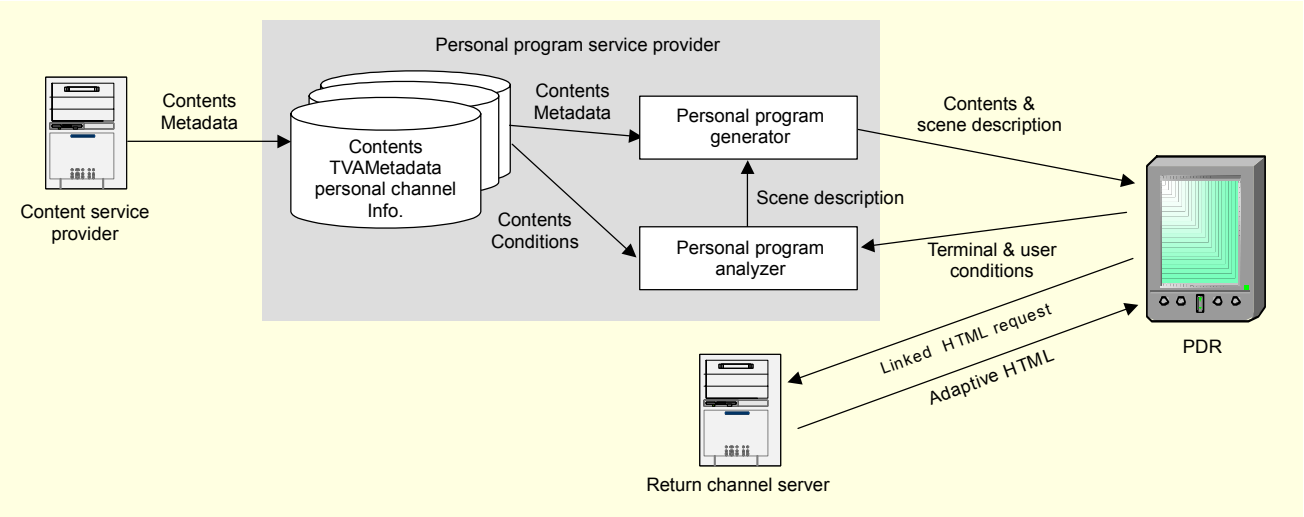
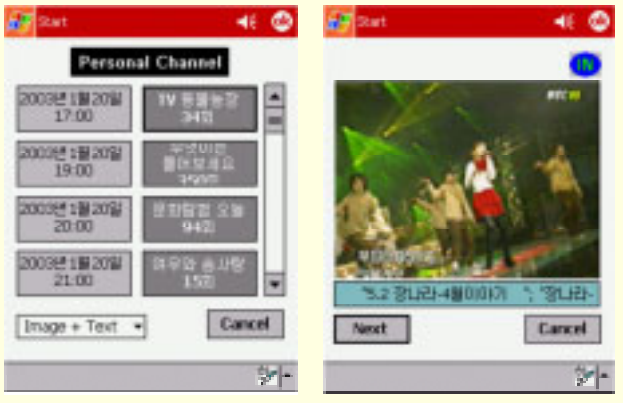


Fig. 13. The architecture of a system for a personal program service.

how an instance of a scene description for certain content should be generated based on given conditions received from metadata storage and a PDR device. Here, the conditions include the characteristics of content, a user, and a terminal. A personal program generator makes a real instance of the scene description and customizes content resources such as images, audio, text, and HTML according to the determined scene description. Both scene descriptions and content resources are sent to the PDR, and a player capable of handling this scene description, called a personal program player, renders the personal program. The personal program player also requests a return channel server for linked material if necessary, and the return channel server provides the requested material.

For the scene description, MPEG-4 XMT-O (Extensible MPEG-4 Textual Format) [23], [24] is used under the consideration of the following aspects: synchronization among resources, human readability, compatibility with other standard scene description technologies such as MPEG-4 Binary Format for Scene (BIFS) and Synchronized Multimedia Integration Language (SMIL).

The personal program player supports functions for parsing and rendering XMT-O instances, as shown in Fig. 14. Using the personal channel browser to be rendered on the PDA, as shown in Fig. 14(a), a user can select a program customized to his or her personal PDR device. In addition, the user can choose a scene description type that is one of the scene description types supported by the personal PDR device. Upon being selected, the program can be presented in a preferred and appropriate way as shown in Fig. 14(b). This is a typical scenario of our personalized content consumption assuming a home network environment with a sub-PDR.



(a) A personal channel on a PDA (b) A personal program rendered on a PDA

Fig. 14. Snapshots of the personal program player: personalized consumption based on the personal channel and the personal program.

4. Metadata-Based Content Adaptation

In this section, we present a utility-based adaptation scheme to provide systematic solutions for selecting an appropriate adaptation method in a given resource condition and content entity [25]. The scheme extends a conventional rate-distortion model to represent relations among feasible adaptation methods satisfying given constrained resources and associated utilities in an interoperable functional form. It allows for the formulation of various adaptation problems as a resource-constrained utility maximization. We apply this scheme to the practical case of adapting MPEG-4 video content to a time-varying bandwidth.

A. Utility-Based Adaptation

Video adaptation problems involve the identification of a video content entity and three important spaces – *adaptation,*

resource, and *utility* (ARU). We use the term "space" in a loose sense here to indicate the multiple dimensionalities involved. Given a video entity (e.g., a compressed video segment), for example, there can be various adaptations such as re-encoding, re-quantization, frame dropping (FD), DCT coefficients dropping (CD), and resolution reduction. In general, all types of adaptations that can be applied in an underlying application are represented by a multidimensional adaptation space. Resources are constraints from terminals or networks like bandwidth, computation capability, power, and display size, etc. In some applications, resource constraints may include several types of resources. For example, in order to provide a video streaming service to handheld devices, besides the bandwidth, other factors like spatial resolution and/or power consumption should be taken into account simultaneously. Therefore, multidimensional resource space in general represents all types of resources which should be met. Utility is the quality of the video that has undergone adaptations. Utility can be measured in an objective or subjective manner. Similarly, utility space represents multiple quality measures to be defined in certain applications.

Figure 15 illustrates the ARU spaces. The adaptation space represents the conceptual space of all possible adaptation operations. If we restrict feasible adaptations to FD and CD only, the adaptation space becomes two-dimensional space. Each dimension represents a specific adaptation operation of FD and CD. Given a video segment, a particular combination of FD and CD adaptation operations corresponds to a point in the adaptation space, and this point is mapped to specific resources that are required terminals or networks and to a certain utility value that the adapted segment preserves. Note that mapping among points in the ARU spaces are often multiple-to-one since different adaptation operations may give the same resource values or induce the same utility value. The interesting point of an adaptation problem lies in choosing an optimal adaptation operation among multiple choices.

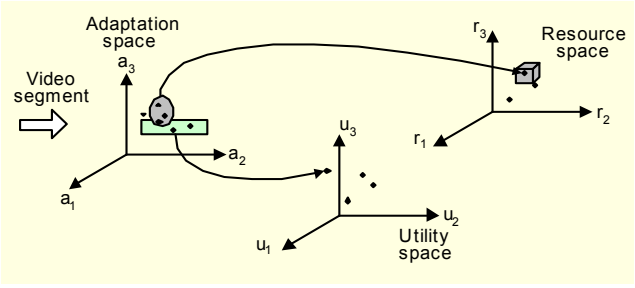


Fig. 15. General scheme for modeling ARU spaces in a video adaptation.

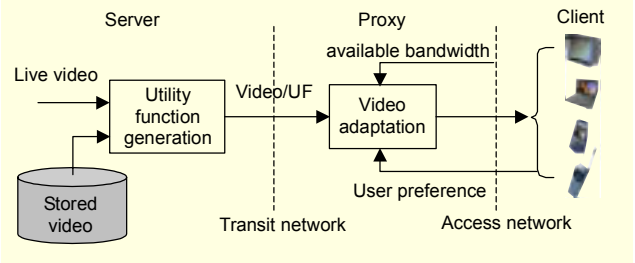


Fig. 16. A three-tiered adaptation architecture using the utility-based framework.

A three-tiered (server-proxy-client) architecture, depicted in Fig. 16, is considered as a promising solution to meet diverse resources and preferences in the Universal Multimedia Access framework. Instead of re-generating a large number of videos with the same content but with different formats or qualities at the server, an adaptation engine can be deployed in the proxy to dynamically adapt the video. Information about ARU relations is very useful for the adaptation engine to select an optimal operation.

B. Case Study: Bit Rate Adaptation

Let us illustrate the formulation and representation of a utility-based scheme by considering a practical use case scenario. Here, we present a use case in the application of adaptive video streaming over resource-limited networks characterized with a time-varying available bandwidth (e.g., heterogeneous access networks associated with a variety of terminals or wireless mobile networks). In this scenario, we assume the architecture as shown in Fig. 16. The server generates a utility description based on the adaptation method of a combination of FD and CD, which will be consumed by an adaptation tool in the gateway.

In this use case, the bit rate of an MPEG-4 video stream to be delivered is adapted to a time-varying bandwidth by an adaptation engine placed in a proxy or a base station in real-time. The description directly tells the adaptation engine what are feasible adaptations satisfying the target rate and associated qualities in a given incoming content. Additionally, in the case where several adaptation operators are available for the target rate, the description enables the adaptation engine to select a more appropriate one according to user preference on spatio-temporal quality or other considerations.

The combination of FD and CD is used as an adaptation method under the consideration that these are efficient and effective ways of rate adaptation with low computational complexity. The operations just truncate parts of the compressed bitstream and thus can be implemented to manipulate video frames and DCT coefficients directly in a compressed domain.

1) FD: FD is a common method of temporal adaptation that adjusts the frame rate by skipping frames in a given video segment. We consider a simple FD operation that allows the dropping of unreferenced frames only within each group of

pictures (GOP). FD can meet only the coarse level of a target bit rate since the smallest unit of data that can be removed is an entire frame. For instance, in the case of GOP (N = 15, M = 3) (N is the size of GOP and M is the distance between two successive anchor frames), we have a set of discrete operations in the FD dimension: "no dropping," "one B frame dropping in each sub-GOP," "all B frames dropping," and "all B and P frames dropping."

2) CD: CD adjusts the spatial quality of each frame by truncating a subset of DCT run-length codes at the end of each block. Lagrangian optimization [28] was used to minimize the overall distortion by optimally allocating the number of DCT coefficients to be dropped in different blocks within a frame. Unlike FD, CD provides an ability to meet the available bandwidth with a finer granularity by adjusting the amount of dropped coefficients. We can define the level of CD by specifying the percentage of rate reduction to be achieved. For example, CD = 10 % represents an operation that reduces 10% of the bit rate in each frame.

3) Combination of FD and CD: For significant bit rate reduction, FD or CD alone is not sufficient to accommodate the target rate. Moreover, the saving of only a few bit-rate values can be achieved by FD, while a finer rate adaptation is possible using CD. Therefore, the FD-CD combination enables us to expand the dynamic range of the achievable bit rate reduction while meeting the target rate at a finer level. Moreover, the combined FD-CD adaptation space provides freedom in balancing the trade-offs between spatial and temporal quality.

C. Utility Function Generation and Description

We need a scheme to represent permitted points in the adaptation space and relationships between the associated utility and resource values. If we consider only a single measure of resource, e.g., bit rate, and utility, e.g., peak signal-to-noise ratio (PSNR), as in the case of an FD-CD combination, a scheme similar to a conventional R-D curve can be used. We call such representation a *utility function* (UF). An example of a UF is shown in Fig. 17. Different adaptation points (representing specific combinations of FD-CD) are plotted on the resource (i.e., bit rate)–utility (i.e., PSNR) plane. In Fig. 17, the adaptation points with the same FD level are grouped with the same curve. Different points on the same curve represent different amounts of rate reduction achieved by CD, e.g., CD: 0 to 50%.

The generation of a UF description can be done for each video stream stored in a server in advance in the case of on-demand streaming applications since computation time may not be a serious issue. In the streaming of live video, a UF description could be generated by a prediction-based approach in real-time [27].

We also developed a description scheme to represent the UF information in an interoperable way. The scheme has been adopted as a part of the *AdaptationQoS* descriptor in MPEG-21 Digital Item Adaptation [28], [29].

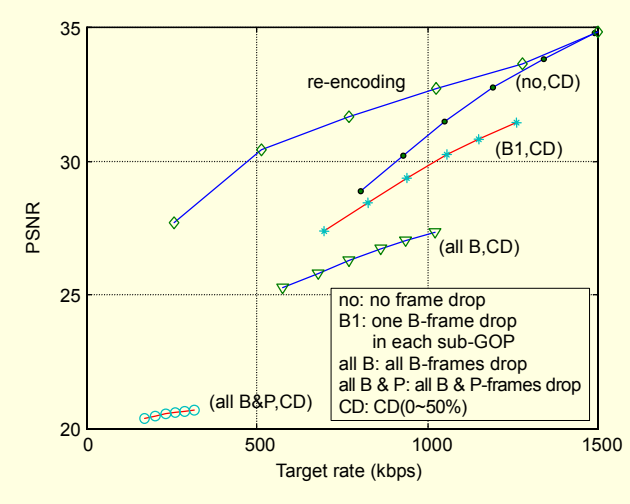


Fig. 17. Utility function using FD-CD operations on the "coastguard" video at 1.5 Mbps, GOP (N = 15, M = 3).

D. Performance Evaluation

We have developed a prototype of a proposed utility-based video adaptation scheme that uses an FD-CD combination to adapt MPEG-4 video streams in response to dynamic bandwidth constraints. The prototype consists of a metadata parser, an adaptation engine, and a user interface. The parser extracts a UF descriptor associated with a video segment. The adaptation engine chooses an optimal operator according to the descriptor and target bit rate, and transcodes the incoming video clip. The user interface allows the monitoring of the adapted video, the UF, the chosen operator, and a trace of the adapted bit rate to the time-varying target rate.

For live videos requiring a dynamic prediction of the UF, our statistical predictor based on content characteristics [27] achieves an accuracy of more than 80% on average in estimating the optimal combination of FD-CD over a wide range of bandwidths, which is regarded as very promising. In terms of the overall performance, the utility-based adaptation achieves a promising gain (0.83 dB in PSNR and noticeable improvement in subjective quality) over the conventional system which simply uses the most frequent adaptation operation for each given target rate.

IV. Prototype System Implementation

A prototype system of metadata broadcasting for personalized TV services has been implemented, integrating each component technology presented in the previous sections into a unified end-to-end system. The proposed technologies and adopted standard technologies have been validated on the prototype system with a set of use-case scenarios of metadata-based personalized content consumption.

Figure 18 illustrates an overall configuration of the end-to-end prototype system.

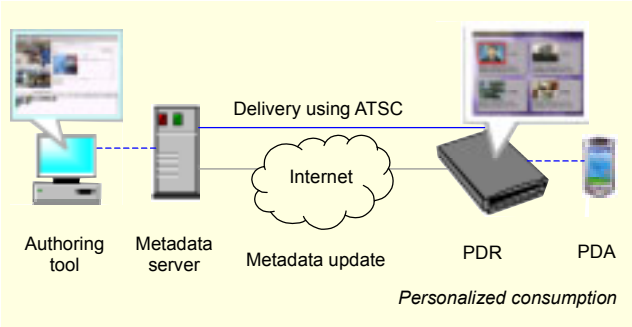


Fig. 18. An overall configuration of the prototype system.

The prototype system consists of four parts: metadata authoring, metadata server, metadata delivery, and terminal platform (a kind of prototype PDR). Metadata are generated off-line, compliant to the TV-Anytime specifications, as a binarized form (BiM) at the authoring tool. In the metadata server, the coded metadata are encapsulated as a data carousel and multiplexed into an MPEG-2 TS with the main AV content to be delivered and then stored in advance. In a test using the prototype, the metadata server transmits metadata over the ATSC channel by pumping the stored MPEG-2 TS at 19.39 Mbps (according to the SMPTE 390M) in the form of an RF signal. The terminal platform receives the delivered MPEG-2 TS with a sequence of demodulation, demultiplexing, and decapsulation in real time. The metadata of a BiM format are decoded into an XML form and manipulated (i.e., parsed, populated into a DB, etc.) allowing for personalized content consumption. The terminal platform is a Linux-based prototypical STB with an ATSC tuner and a 60 GB HDD. Additionally, we also implemented bi-directional metadata services through the Internet, in which a user can get supplemental information about a program, especially segment metadata, based on the TV-Anytime specification [6]. In our implementation, the metadata server also plays the role of a return channel server that may be provided by a content provider or a third party, as shown in Fig. 18.

Based on the prototype system, we have tested a set of use-case scenarios of personalized services. Here, we used a large amount of real metadata (EPG data for 7,500 programs corresponding to 2 weeks/24 hours) and several HDTV video sequences along with a rich segment metadata. We have tested the following use cases in term of feasibility and computational efficiency:

- ACG-based efficient searching for and accessing a specific program from a local storage and/or broadcasting schedule
- Automatic recommendation of user-preferred programs according to user preferences which is automatically extracted using usage history information or manually set by a user
- Efficient searching, browsing, and non-linear navigation in a segment level by new schemes of ToC-based or event-based summary browsing utilizing segment metadata stored in a local storage
- Enabling various consumption modes including watching live TV, EPG surfing, and browsing stored content in a dynamic manner
- Personal channel and personal program services in a PDA
- Dynamic update of segment metadata through a return channel of the Internet, etc.

It has been validated that the use cases are reliably performed in real time on the prototype STB employing a common MIPS® CPU (400 MHz).

V. Conclusions

In this paper, we proposed a practical solution of metadata broadcasting to make broadcasting services personalized according to user preferences and various terminal and network conditions. First, we presented the overall system architecture of a metadata broadcasting system and then proposed several core technologies including metadata authoring, metadata coding and multiplexing, and consumer-side metadata processing for personalized consumption. For achieving interoperability, we followed the standard specifications of the TV-Anytime Forum metadata, MPEG-7 Systems, and MPEG-2 Systems. Considering home network environments where heterogeneous devices are connected to each other, we also proposed a utility-based content adaptation scheme.

We have integrated the proposed functional components on an end-to-end prototype system and successfully validated the feasibility and practicability of the system with a set of personalized service scenarios (in particular, a personal channel and personal program) that were also developed in this study. In particular, at the IBC 2003 exhibition held in Amsterdam, we verified the interoperability of our prototype system with other TV-Anytime compliant systems. Promisingly, the terminal platform could receive and parse the metadata delivered by different broadcasting servers in an exchanged

manner, and it also could update some parts of the metadata acquired from a metadata server through the Internet.

We are still making progress to employ a more realistic broadcasting server, which performs in a sequence, metadata encapsulation, multiplexing, and transmission in real time rather than just pumping ready-made streams. In order to handle a number of metadata in a more intelligent way, metadata archiving and agent technologies are also required, which are in the scope of our ongoing research.

References

[1] Shif-Fu Chang et al., "Special Issue on MPEG-7,", *IEEE Trans. Circuits and Syst. For Video Techno.,* vol. 11, no. 6, June 2001, pp. 685-772.

[2] S-1 (version 1.2: SP001v12), *Phase 1 Benchmark Features*, The TV-Anytime Forum, Feb. 2003; ftp://tva:tva@ ftp.bbc.co.uk/pbu/Plenary/.

[3] S-2 (version 1.3: SP002v13), *System Description,* The TV-Anytime Forum, Feb. 2003; ftp://tva:tva@ftp.bbc.co.uk/pbu/Plenary/.

[4] S-3 (version 1.3: SP003v13), *Metadata*, The TV-Anytime Forum, Dec. 2002; ftp://tva:tva@ftp.bbc.co.uk/pbu/Plenary/.

[5] S-4 (version 1.2: SP004v12), *Content Referencing*, The TV-Anytime Forum, June 2002; ftp://tva:tva@ftp.bbc.co.uk/pbu/Plenary/.

[6] S-6 (version 1.0: SP006v10), *Metadata Services over a Bi-directional Network,* The TV-Anytime Forum, Feb. 2003; ftp://tva:tva@ftp.bbc.co.uk/pbu/Plenary/.

[7] ETS TS 102 822, *Broadcast and On-line Services: Search, Select and Rightful Use of Content on Personal Storage Systems*, ETSI, Sophia-Antipolis, France, 2003; http://www. etsi.org.

[8] R-1 (version 2.0: RQ001v20), *Requirements Series: R-1 (Phase 2) on Business Models*, The TV-Anytime Forum, Aug. 2003; ftp://tva:tva@ftp.bbc.co.uk/pbu/Plenary/.

[9] TV179r3, *Phase 2 Call For Contributions*, The TV-Anytime Forum, Aug. 2003; ftp://tva:tva@ftp.bbc.co. uk/pbu/Plenary/.

[10] The myTV project; http://www.extra.research.philips.com/euprojects/mytv

[11] The Share-it project; http://www.extra.research.philips.com/euprojects/share_it

[12] Working Draft T3/S8 470, *Metadata for Advanced EPG Functionality*, ATSC, Jan. 2002.

[13] R. Zabih, J. Miller, and K. Mai, "A Feature-Based Algorithm for Detecting and Classifying Scene Breaks," *Proc. ACM Multimedia95,* Nov. 1995, pp. 189-200.

[14] J.-G. Kim, H. S. Chang, Y. Kim, K. Kang, M. Kim, J. Kim, and H.-M. Kim, "Multimodal Approach for Summarizing and Indexing News Video," *ETRI J.,* vol. 24, no. 1, Feb. 2002, pp. 1-11.

[15] H. S. Chang and K. Kang, "A Compressed Domain Scheme for Classifying Block Edge Patterns," accepted for the publication in *IEEE Trans. Image Processing (Published in 2004)*.

[16] W3C Recommendation 16, *XSL Transformations (XSLT) Version 1.0.*, W3C, Nov. 1999 ; http://www.w3.org/ Style/XSL.

[17] ISO/IEC 15938-1, *Information Technology-Multimedia content Description Interface-part 1 Systems*, ISO/IEC, 2002.

[18] ISO/IEC 13818-1, *Information Technology-Generic Coding of Moving Pictures and Associated Audio Information-part 10 Digital Storage Media Command and Control (DSM-CC),* ISO/IEC, 1998.

[19] N5270, *Text of ISO/IEC 13818-1:2000/PDAM1-Transport of Metadata*, ISO/IEC JTC1/SC29/WG11 (MPEG), Nov. 2002.

[20] S.-J. Yang, H. S. Chang, Y. Kim, K. Kang, and N. N. Thanh, "Design of TeM Codec for Delivering MPEG-7 Descriptions," *Proc. IEEE ICACT-2003*, Jan. 2003, pp. 117-120.

[21] AN 282, *A Streamable XML Binary Encoding for TV-Anytime Metadata*, The TV-Anytime Forum MD WG, June 2001.

[22] ISO/IEC 13818-1, *Information Technology-Generic Coding of Moving Pictures and Associated Audio Information-part 1 Systems*, ISO/IEC, 2002.

[23] M. Kim, S. Wood, and L.-T. Cheok, "Extensible MPEG-4 Textual Format (XMT)," *Proc. ACM Multimedia 2000,* Oct. /Nov. 2000, pp. 71-74.

[24] K.-Y. Kim, H.-C. Kim, W.-S. Cheong, and K. Kim, "Design and Implementation of MPEG-4 Authoring Tool," *Proc. IEEE ICCIMA 2001,* Oct. 2001, pp.348-351.

[25] J.-G. Kim, Y. Wang, and S.-F. Chang, "Content-Adaptive Utility-Based Video Adaptation," *Proc. ICME-2003*, Baltimore, Maryland, July 2003.

[26] A. Eleftheriadis and D. Anastassiou, "Constrained and General Dynamic Rate Shaping of Compressed Digital Video," *Proc. IEEE ICIP,* Oct. 1995, pp. 396-399.

[27] Y. Wang, J.-G. Kim, and S.-F. Chang, "Content-Based Utility Function Prediction for Real-Time Video Transcoding," *Proc. IEEE ICIP-2003*, Barcelona, Spain, Sept. 2003.

[28] N5353, *MPEG-21 Digital Item Adaptation CD,* ISO/IEC JTC 1/SC 29/WG 11, Dec. 2002.

[29] ISO/IEC 21000-7 FCD – Part 7: *Digital Item Adaptation,* ISO/IEC JTC1/SC29/WG11/N5933, Brisbane, Australia, Oct. 2003.

Kyeongok Kang received the BS and MS degrees in physics from Pusan National University, Busan, Korea, in 1985 and 1988, and his PhD degree in electrical engineering from Hankuk Aviation University, Seoul, Korea, in 2004. He has been with Electronics and Telecommunications Research Institute (ETRI) since 1991, and he is now a Principal Member of Engineering Staff and the Leader of 3D Media Research Team. His research interests are in personalized broadcast technologies based on MPEG-7 and TV-Anytime, and audio signal processing including 3D audio.

Jae-Gon Kim received the BS degree in electronics engineering from Kyungpook National University, Korea, in 1990, the MS degree in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Korea, in 1992. Since 1992, he has been a Senior Member of Research Staff in the Broadcasting Media Research Group of ETRI, Korea. He is currently the Team Leader of the Broadcasting Content Research Team. From 2001 to 2002, he was a Staff Associate at Columbia University, New York. His research interests include video processing, networked video, multimedia applications, MPEG-7 and MPEG-21. He is a Member of IEEE.

Heekyung Lee received the BS degree in computer engineering from Yeungnam University, Korea in 1999, and the MS degree in engineering from Information and Commuications University (ICU), Korea, in 2002. She is currently a Member of Research Staff, in the Broadcasting Media Research Group of ETRI, Korea. Her research interests include personalized service for user terminals, content adaptation, and metadata generation/consumption.

Hyun Sung Chang received the BS and MS degrees in electrical engineering from Seoul National University, Korea, in 1997 and 1999. Since 1999, he has been with the Electronics and Telecommunications Research Institute, Korea, as a Member of Research Staff. His research interests include image analysis and understanding, multimedia systems, digital broadcasting, and artificial intelligence.

Seung-Jun Yang received the BS degree in computer science from Sunchon University, Korea in 1999, and the MS degree in computer science from Chonnam University, Korea, in 2001. He is currently a Member of Research Staff in the Broadcasting Media Research Group of ETRI, Korea. His research interests include MPEG-7 BiM, interactive broadcast systems, and metadata generation.

Young-tae Kim received the BS, MS, and PhD degrees from the Department of Electronics & Telecommunications Engineering from Kwangwoon University, Seoul, Korea, in 1991, 1993, and 1998. He joined ETRI in 1998 as a Postdoctorial Fellow. Since 1999, he has been with Broadcasting Media Research Group of ETRI as a Senior Member of Research Staff. His current research interests include multimedia indexing and retrieval and interactive broadcasting technologies.

Han-kyu Lee received the BS and MS degrees from Kyungpook National University, Daegu, Korea, in 1994 and 1996. Since 2003, he has studied as a PhD student at Information and Communications University (ICU), Daejeon, Korea. After he received his MS degree in 1996, he has served as a Member of Research Staff for ETRI, Korea. He has been engaged in the development of an MPEG-2 video encoder, elemenentary stream multiplexer, HDTV encoder system, MPEG-7, and TV-Anytime technologies. His research interests are digital signal and image processing in the fields of video communications, multimedia systems, and interactive broadcast systems.

Jinwoong Kim received the BS and MS degrees from Seoul National University, Seoul, Korea, in 1981 and 1983, and the PhD degree from the Department of Electrical Engineering from Texas A&M University, United States, in 1993. Since 1983, he has been on the Research Staff in ETRI, Korea. He has been engaged in the development of the TDX digital switching system, MPEG-2 video encoder, HDTV encoder system, and MPEG-7 technology. His research interests include digital signal processing in the fields of video communications, multimedia systems, and interactive broadcast systems.